

УДК 577.29

ПОИСК клТ-ДНК В РАСТИТЕЛЬНОМ ГЕНОМЕ С ПОМОЩЬЮ МЕТОДА ПОЛНОГЕНОМНОГО ВЫРАВНИВАНИЯ

Г.В. Хафизова, П.В. Добрынин, Т.В. Матвеева

Санкт-Петербургский государственный университет, Санкт-Петербург, Россия, galina.khafizova@gmail.com

Полногеномное выравнивание, примененное для последовательностей плазмиды pRi1724 агробактерии *Agrobacterium rhizogenes* Riker и генома *Nicotiana tomentosiformis* Linnaeus, позволило выявить в растительном геноме несколько вставок клТ-ДНК (ТА и ТС), а также отдельные фрагменты последовательностей, гомологичных агробактериальным (*orf20/mis* и др.). Для сборки данных фрагментов в единый регион необходимо использование дополнительных методов, таких как ПЦР, однако использование полногеномного выравнивания в качестве прескрининга позволяет значительно сократить ресурсозатраты на проведение дальнейшего анализа, поскольку оно позволяет при наличии минимума исходных данных получить информацию не только о наличии в растительном геноме последовательностей, гомологичных агробактериальным, но и о взаиморасположении фрагментов данных последовательностей, и в некоторых случаях – идентифицировать сайт локализации вставки в растительном геноме, что важно для анализа данных вставок.

Ключевые слова: горизонтальный перенос генов, агробактерии, Т-ДНК, *Nicotiana*.

клТ-ДНК – это последовательности, гомологичные Т-ДНК агробактерий, приобретенные растениями в результате горизонтального переноса генов. Первая клТ-ДНК была найдена в геноме *Nicotiana glauca* Graham [White et al., 1983], позже подобные последовательности нашли в геномах ряда других видов рода *Nicotiana* [Intrieri M.C., Buiatti M. 2001; Suzuki et al., 2002], а также родов *Linaria* Miller [Matveeva et al., 2012], и *Ipomoea* Linnaeus [Kyndt, Jarret, et al., 2015]. В геномах разных видов вставки клТ-ДНК различаются по составу, копииности, и по сайтам локализации [Chen et al., 2014]. В данном исследовании проведена оптимизация метода анализа данных геномного секвенирования для исследования клТ-ДНК.

Для проведения полногеномного выравнивания использовали программу LASTZ. Выравнивали скафолды *Nicotiana tomentosiformis* L. и *Nicotiana sylvestris* Speg. et Comes против последовательности плазмиды pRi1724 штамма *Agrobacterium rhizogenes* 1724 Riker (AP002086.1). Все растительные последовательности брали с ftp сервера сайта <http://www.plantgdb.org/>). Выравнивание производили со стандартными параметрами: командой вида – lastz Plasmid.fasta Genome.fasta –chain --format=maf. Результат выравнивания хранился в файле формата MAF. Для работы с MAF файлами был написан скрипт на Python. Заданный параметр минимального сходства – 65%. Визуализацию результатов производили с помощью программы Circos.

В результате проведенного выравнивания для последовательностей *Nicotiana sylvestris* Speg. et Comes и pRi1724 не было обнаружено значимых совпадений, что согласуется с данными об отсутствии клТ-ДНК в геноме *Nicotiana sylvestris* Speg. et Comes [Intrieri, Buiatti, 2001]. В результате выравнивания последовательностей *Nicotiana tomentosiformis* L. на плазмиду был найден участок вставки, который включает последовательности, гомологичные агробактериальным генам *orf18(orf13 in T-DNA)*, *orf19(orf14 in T-DNA)*, *orf20(mis)* – по центру вставки, и их же в инвертированной ориентации. Данная вставка совпадает по строению со вставкой ТА. Считается, что часть ТА была приобретена от штамма 1724. Также была найдена вставка, полностью совпадающая по своему составу с описанной вставкой ТС. В ее центре находятся последовательности, гомологичные гену *rolB* в их прямой и инвер-

тированной ориентации относительно того, как они расположены на плазмиде. Далее следуют последовательности, гомологичные агробактериальным генам *rolA*, *orf13(orf8 in T-DNA)*, *orf12(orf3 in T-DNA)*, и часть последовательности, гомологичной гену *orf11(orf2 in T-DNA)*. Предполагается, что вставка ТС была получена в результате горизонтального переноса от штамма *Agrobacterium rhizogenes* A4 Riker, однако благодаря высокому сходству нуклеотидных последовательностей онкогенов штамма А4 и штамма 1724, нам удалось идентифицировать вставку ТС в геноме *Nicotiana tomentosiformis* L. при выравнивании на плазмиду штамма 1724. Помимо фрагментов, объединенных в протяженные регионы, также были найдены отдельные фрагменты, объединить которые в единый регион можно с помощью таких методов молекулярной биологии, как ПЦР. Анализ данных фрагментов указывает на наличие в растительном геноме последовательностей, гомологичных гену *orf20(mis)*, помимо тех, что вошли в состав вставки ТА. По-видимому, это участки вставки ТВ. Считается, что часть этой вставки, включающая ген микимопинсинтазы, происходит от штамма 1724. Группой Chen и др. для поиска вставок был использован другой подход – применение алгоритма локального выравнивания BLAST, с дальнейшим подбором праймеров и восстановлением структуры вставки из фрагментов с помощью метода ПЦР. С использованием метода полногеномного выравнивания было найдено 2 вставки клТ-ДНК, несущие последовательности, и набор отдельных фрагментов, для сборки которых в единый регион также потребуется использование ПЦР. Однако использование полногеномного выравнивания позволяет получить больше информации о взаиморасположении фрагментов, гомологичных агробактериальным, в геноме растения, что сильно облегчает дальнейшую задачу сборки последовательности вставки клТ-ДНК.

Итак, использование метода полногеномного выравнивания для поиска клТ-ДНК в растительном геноме позволяет проводить скрининг при наличии минимальных исходных данных – набор скафолдов растительного генома и последовательность плазмиды агробактерии. Предложенный нами метод позволяет сэкономить время и другие ресурсы, расходующиеся на проведение молекулярно-генетических экспериментов. С помощью метода полногеномного выравнивания можно идентифицировать

последовательности, гомологичные агробактериальным, и для некоторых случаях – взаимолокализовать их, собирая в единый регион, что было показано на примере вставок ТА и ТС в геноме *Nicotiana tomentosiformis* L.

Данная работа выполнена при поддержке гранта РФФ 16-16-10010.

Библиографический список (References)

- Chen K., Dorlhac de Borne F., Szegedi E., Otten L., Deep sequencing of the ancestral tobacco species *Nicotiana tomentosiformis* reveals multiple T-DNA inserts and a complex evolutionary history of natural transformation in the genus *Nicotiana*, *Plant Journal*, 2014. V.80. N.4. P.669–682.
- Intrieri M.C., Buiatti M., The horizontal transfer of *Agrobacterium rhizogenes* genes and the evolution of the genus *Nicotiana*, *Molecular Phylogenetics and evolution*, 2001. V.20. P. 100–110.
- Kyndt T., Quispe D., Hong Zhai, Jarret R., Ghislain M., Liu Q., Gheysen G., and Jan F. Kreuze, The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: An example of a naturally transgenic food crop, *PNAS*, 2015. V.112. N.18. P.5844–5849.
- Matveeva T.V., Bogomaz D.I., Pavlova O.A., Nester E.W., Lutova L.A. Horizontal Gene Transfer from Genus *Agrobacterium* to the Plant *Linaria* in Nature, *MPMI*, 2012. V.25. N.12. P.1542–1551.
- Suzuki K, Yamashita I., Tanaka N., Tobacco plants were transformed by *Agrobacterium rhizogenes* infection during their evolution, *Plant Journal*, 2002. V.32. N.5. P.775–787.
- White F.F., Garfinkel D.J., Huffman G.A., Gordon M.P., and Nester E.W., Sequence homologous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants, *Nature*, 1983. N.301. P.348–350.

Plant Protection News, 2016, 3(89), p. 177–178

SEARCH OF cT-DNA IN PLANT'S GENOMES THROUGH WHOLE-GENOME ALIGNMENT

G.V. Khafizova, P.V. Dobrynin, T.V. Matveeva

Saint Petersburg State University, galina.khafizova@gmail.com

Using the method of whole-genome alignment applied to the sequences of plasmid pRi1724 of *Agrobacterium rhizogenes* Riker and *Nicotiana tomentosiformis* Linnaeus, several inserts of cT-DNA (TA, TC) were detected in the plant genome, as well as fragments of sequences homologous to *Agrobacterium*'s (*orf20/mis* et al.). To assemble the fragments into a region it is necessary to use other methods, such as PCR. Application of whole-genome alignment as pre-screening method significantly reduces resource consumption for further analysis. Having minimum of input data it allows to obtain information not only about the presence in the plant genome sequences homologous to *Agrobacterium*, but also the mutual arrangement of fragments of these sequences, and in some cases it lets to identify the site of localization of the insert in the plant genome that is important for the analysis of the inserts.