

УДК 578.5:578.89

**И. Э. Рубель¹, О. Ю. Баранов¹, С. В. Пантелеев¹, О. А. Разумова¹,
В. А. Гущин², В. В. Макаров²**

¹ Институт леса Национальной академии наук Беларуси

² Московский государственный университет
имени М. В. Ломоносова

МОЛЕКУЛЯРНО-ГЕНЕТИЧЕСКИЙ АНАЛИЗ ВИРУСОПОДОБНЫХ ЭЛЕМЕНТОВ В ГЕНОМЕ ХВОЙНЫХ

В статье описывается разнообразие вирусоподобных генетических элементов хвойных. Данные для последующего анализа и поиска вирусоподобных последовательностей в растительном геноме были получены в результате высокопроизводительного секвенирования на основе полупроводниковой детекции протонов (технология Ion Torrent). Приведены общие сведения о схеме секвенирования по этой технологии. В качестве исходного материала использовались геномные и транскриптомные ДНК- и кДНК-библиотеки сосны обыкновенной и ели европейской. Поиск среди полученных протяженных последовательностей проводился путем сопоставления с записями базы данных GenBank (National Center for Biotechnological Information – NCBI), с использованием онлайн-программы BLAST. Был аннотирован ряд последовательностей, имеющих сходство с ранее описанными вирусоподобными генетическими элементами. Среди них – ретротранспозоны групп Ty3/gypsy и Ty1/copia, содержащие длинные концевые повторы (LTR) и широко встречающиеся в геноме ели, а также ретротранспозоны, не содержащие LTR и относящиеся к семейству LINE. При этом все обнаруженные последовательности были уникальными, а большая их часть содержала последовательности различных генов первичного и вторичного метаболизма хвойных, что может косвенно свидетельствовать об их роли в формировании адаптивной изменчивости растений.

Ключевые слова: вирусоподобные генетические элементы, ретротранспозоны, высокопроизводительное секвенирование, хвойные.

**I. E. Rubel¹, O. Baranov¹, S. V. Panteleev¹, O. A. Razumova¹,
V. A. Gushchin², V. V. Makarov²**

¹ Institute of Forest of the National Academy of Sciences of Belarus

² Lomonosov Moscow State University

MOLECULAR GENETIC ANALYSIS OF VIRUS-LIKE ELEMENTS IN THE GENOME OF CONIFERS

This article describes a variety of virus-like genetic elements of conifers. Data for later analysis and search of the virus-like sequences in the plant genome were obtained by next-generation sequencing based on the semiconductor detection of protons (Ion Torrent technology). General scheme of this sequencing technology was described. Genomic and transcriptomic DNA and cDNA libraries of scotch pine and norway spruce were used as the starting material. Search among the obtained contigs was performed by comparing with the GenBank NCBI database records, using BLAST online tool. A series of sequences with similarity to the previously described virus-like genetic elements was annotated. We have identified the long terminal repeat (LTR)-containing retrotransposons belonging to the Ty3/gypsy and Ty1/copia families, that are ubiquitous throughout the spruce genome, and non-LTR retrotransposons belonging to the LINE family. In this case all detected sequences were unique and most of them contained the sequences of various genes of the primary and secondary metabolism of conifers, which may be indirect evidence of their role in the formation of adaptive variability of plants.

Key words: virus-like genetic elements, retrotransposons, next-generation sequencing, conifers.

Введение. Вирусы растений являются возбудителями инфекционных заболеваний, способны значительно ослаблять растение, ухудшать его хозяйственно ценные признаки [1]. Однако по сравнению с другими возбудителями инфекционных заболеваний фито-вирусы (особенно древесных видов) изучены слабо.

Вирусоподобные генетические элементы (ВГЭ) представляют собой нуклеотидные по-

следовательности, интегрированные в геномную ДНК растительных клеток и имеющие участки (локусы), гомологичные определенным вирусным генам.

Наиболее распространенным типом ВГЭ в геноме как хвойных, так и растений в целом являются ретротранспозоны [2, 3], которые согласно современным представлениям имеют общего предшественника с ретровирусами [4–6].

Однако несмотря на утрату инфекционных свойств, вирусоподобные генетические элементы имеют существенное негативное значение для лесного хозяйства, поскольку результатом их транспозиции могут стать различные нарушения структурно-функциональной организации генома: инсерции, приводящие к образованию транскрипционных разрывов в экзонных участках генов растения или увеличению размеров интронов, а также делеции и транслокации. Это является причиной возникновения нежелательных фенотипических эффектов [7]. ВГЭ являются также одной из причин возникновения выраженной инбредной депрессии у хвойных.

В связи с этим изучение разнообразия вирусоподобных генетических элементов в геноме хвойных представляет определенный интерес.

Основная часть. На начальном этапе исследований проводилось выделение из растительного материала суммарной ДНК модифицированным СТАВ-методом и суммарной РНК при помощи набора реагентов GeneJET Plant RNA Purification Mini Kit (Thermo Scientific). Для последующего создания ДНК-библиотек все образцы РНК были переведены в стабилизированную форму в виде кДНК.

Библиотеки ДНК-фрагментов для последующего секвенирования готовили согласно протоколу фирмы-производителя с использованием набора реактивов Ion Plus Fragment Library Kit (Ion Torrent – Life Technologies). Шесть библиотек были получены с использованием суммарной ДНК сосны обыкновенной (*Pinus sylvestris* L.) и три – с использованием кДНК, синтезированной на матрице суммарной РНК ели европейской (*Picea abies* (L.) H. Karst.). При подготовке библиотек также использовался набор реагентов Ion Library Equalizer Kit, благодаря чему их ожидаемая концентрация должна составлять ≈ 100 рМ.

Качество библиотек фрагментов проверяли с применением методов электрофоретического фракционирования, спектрофотометрии и ПЦР (полимеразной цепной реакции) в режиме реального времени (real-time PCR).

Real-time ПЦР проводили на амплификаторе Stratagene Mx3000P (Agilent Technologies) с использованием готовой смеси Maxima SYBR Green/ROX qPCR Master Mix (Thermo Scientific) и праймеров для амплификации библиотек Library Amplification Primer Mix из набора Ion AmpliSeq Library Kit 2.0 (Ion Torrent – Life Technologies). Данные эксперимента были представлены в виде графика зависимости интенсивности флуоресценции от количества прошедших циклов ПЦР, при этом резкое возрастание интенсивности флуоресценции началось

с 6–10-го цикла, что свидетельствует о достаточной концентрации исходных библиотек.

С целью установления соответствия реального размера полученных библиотек рекомендованному (согласно протоколу E-Gel Size Select Agarose Gels Quick Reference) был проведен электрофоретический анализ. Для него использовался 2%-ный (масса/объем) агарозный гель на основе TBE-буфера, напряженность электрического поля составляла 6 В/см. Гель окрашивался бромистым этидием, оценку данных проводили визуально при помощи УФ-трансиллюминатора. Проведенное электрофоретическое фракционирование выявило отсутствие деградации ДНК в препаратах. Было установлено, что амплифицированные фрагменты библиотек имеют размер более 300, но менее 400 п. н., что соответствует ожидаемому значению, приведенному в протоколе фирмы-производителя (≈ 330 п. н.). Фрагменты размером около 600 п. н., которые также были обнаружены в ходе визуального анализа, являлись конкатемерами, образовавшимися в ходе ПЦР.

Был проведен спектрофотометрический анализ библиотек для секвенирования, который показал, что средняя концентрация полученных препаратов нуклеиновых кислот библиотек находилась в пределах 4,5–7,5 нг/мкл образца. Соотношение экстинкций A_{260}/A_{280} находилось в диапазоне 1,83–2,10, что указывает на высокую степень очистки препаратов ДНК.

Далее готовили матрицу для секвенирования с использованием полученных библиотек. Секвенирование выполнялось на полногеномном анализаторе Ion PGM System (Life Technologies). Все этапы подготовки проводили в соответствии с инструкцией фирмы-производителя, с использованием наборов реагентов, поставляемых вместе с оборудованием.

На следующем этапе проводили отмывку микрочастиц, не несущих амплифицированной ДНК (стадия обогащения). Данная операция осуществлялась с использованием системы автоматической пробоподготовки Ion OneTouch ES и наборов реагентов Ion PGM Template OT2 Solutions 200 Kit, Ion PGM Enrichment Beads (Life Technologies).

Секвенирующая реакция протекала на чипе Ion 314 Chip v2. Финальную пробоподготовку и загрузку чипа проводили в соответствии с протоколом фирмы-производителя.

Первичная обработка данных, поступающих от геномного анализатора Ion PGM System, осуществлялась в автоматическом режиме при помощи программного обеспечения Ion Torrent Suite, поставляемого вместе с оборудованием.

В результате первичного анализа были получены следующие показатели для различных

образцов: плотность загрузки чипа – от 60 до 92%, доля моноклональных микросфер – 70–87%, доля пустых микросфер – 1–5%, а большая часть полученных прочтений (reads) имела длину от 206 до 269 п. н.

Количество полученных в результате секвенирования парноконцевых чтений варьировалось от 695 467 до 923 775. Оценка качества полученных последовательностей, а также удаление из сиквенсов последовательностей праймеров и адаптеров осуществлялись автоматически в ходе постобработки первичных данных программным комплексом Ion Torrent Suite. Из массива данных удалялись последовательности короче 20 п. н., а также с показателем качества $Q < 20$ (т. е. с вероятностью ошибки секвенирования более 1%). Таким образом, для каждого образца было получено 273 181–394 954 высококачественных парноконцевых чтений или 45,3–65,1 млн п. н.

Сборка парноконцевых чтений в протяженные последовательности (контиги) осуществлялась при помощи программы DNASTAR SeqMap NGen 12. Для различных библиотек число полученных контигов составляло от 13 711 до 27 712.

Поиск перспективных последовательностей для дальнейшего анализа проводился путем сопоставления всех контигов с последовательностями базы данных GenBank [8] с использованием программного комплекса BLAST [9], размещенного на этом сервере, при этом поиск осуществлялся по базе данных нуклеотидных последовательностей (nucleotide blast), а в качестве определяющего параметра поиска с целью расширить выборку была выбрана опция «поиск по частичному сходству» (somewhat similar sequences (blastn)).

В результате был выявлен 81 контиг – потенциальные кандидаты в вирусы и вирусоподобные генетические элементы.

Предварительно стоит отметить, что ретротранспозоны подразделяются на две группы: элементы, содержащие длинные концевые повторы (long terminal repeat – LTR), и элементы, не содержащие их [10, 11]. Среди последних различают длинные диспергированные повторы (LINE – long interspersed elements) и короткие диспергированные повторы (SINE – short interspersed elements) [12]. В свою очередь, среди

всех ретротранспозонов в геномах эукариот наибольшее распространение получили мигрирующие генетические элементы, содержащие длинные концевые повторы.

Проведенный анализ выявил большое количество контигов, имеющих сходство с различными типами мигрирующих генетических элементов. При этом 74 из идентифицированных последовательностей относились к LTR-содержащим ретротранспозонам группы Ty3/gypsy (семейство Metaviridae), а 2 контига – к группе Ty1/copia (Pseudoviridae).

Также было выявлено 5 контигов, относящихся к ретротранспозонам семейства LINE, представляющих собой альтернативную группу мигрирующих генетических элементов, не содержащих в своей нуклеотидной структуре длинные концевые повторы.

Стоит обратить внимание, что наибольшая часть (91%) контигов, идентифицированных как участки LTR-ретротранспозонов, относилась к семейству Ty3/gypsy.

Также следует отметить, что все выявленные последовательности являлись уникальными, что указывает на их самостоятельность. При этом большая часть ретротранспозонов ели европейской содержала последовательности различных генов первичного и вторичного метаболизма: 3-каренисинтазы, энт-каурен-синтазы, цитохрома P450 подсемейства CYP720B и др., что указывает на их определенную роль в формировании адаптивной изменчивости хвойных.

Заключение. В результате была аннотирована 81 последовательность, среди них было выявлено сходство с LTR-содержащими ретротранспозонами групп Ty3/gypsy и Ty1/copia, описанных ранее для ели, а также ретротранспозонами, не содержащими LTR и относящимися к семейству LINE. При этом все выявленные последовательности оказались уникальными, а большая их часть содержала участки различных генов первичного и вторичного метаболизма хвойных, что может свидетельствовать об их роли в формировании адаптивной изменчивости растений.

Полученные результаты могут представлять интерес для генетики и селекции лесных древесных видов, расширить знания о вирусоподобных генетических элементах хвойных и их значении.

Литература

1. Gonthier P., Nicolotti G. Infectious Forest Disease. Croydon (UK): CPI Group Publ., 2013. 641 p.
2. Copia-like retrotransposons are ubiquitous among plants / D. F. Voytas [et al.] // PNAS USA. 1992. Vol. 89. P. 7124–7128.
3. Suoniemi A., Tanskanen J., Schulman A. H. Gypsy-like retrotransposons are widespread in the plant kingdom // Plant Journal. 1998. Vol. 13. P. 699–705.
4. Xiong Y., Eickbush T. H. Origin and evolution of retroelements based upon their reverse transcriptase sequences // EMBO Journal. 1990. Vol. 9. P. 3353–3362.

5. Retrovirus phylogeny and evolution / R. F. Doolittle [et al.] // *Curr. Top. Microbiol. Immunol.* 1990. Vol. 157. P. 1–18.
6. Schulman A. H., Flavell A. J., Noel Ellis T. H. The Application of LTR Retrotransposons as Molecular Markers in Plants // *Methods in Molecular Biology*. 2004. Vol. 260. P. 145–173.
7. Vandenbussche M., Gerats T. TE-Based Mutagenesis Systems in Plants // *Methods in Molecular Biology*. 2004. Vol. 260. P. 115–127.
8. The National Center for Biotechnology Information (NCBI) [Electronic resource]. URL: <http://www.ncbi.nlm.nih.gov/genbank/> (date of access: 12.01.2016).
9. Basic Local Alignment Search Tool (BLAST) [Electronic resource]. URL: <http://blast.stva.ncbi.nlm.nih.gov/Blast.cgi> (date of access: 13.01.2016).
10. Grandbastien M.-A. Retroelements in higher plants // *Trends Genet.* 1992. Vol. 8. P. 103–108.
11. Bennetzen J. L. The contributions of retroelements to plant genome organization, function and evolution // *Trends Microbiol.* 1996. Vol. 4. P. 347–353.
12. Schmidt T. LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes // *Plant Mol. Biol.* 1999. Vol. 40. P. 903–910.

References

1. Gonthier P., Nicolotti G. Infectious Forest Disease. Croydon (UK), CPI Group Publ., 2013. 641 p.
2. Voytas D. F., Cummings M. P., Konieczny A. K., Ausubel F. M., Rodermel S. R. Copia-like retrotransposons are ubiquitous among plants. *PNAS USA*, 1992, vol. 89, pp. 7124–7128.
3. Suoniemi A., Tanskanen J., Schulman A. H. Gypsy-like retrotransposons are widespread in the plant kingdom. *Plant Journal*, 1998, vol. 13, pp. 699–705.
4. Xiong Y., Eickbush T. H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO Journal*, 1990, vol. 9, pp. 3353–3362.
5. Doolittle R. F., Feng D. F., McClure M. A., Johnson M. S. Retrovirus phylogeny and evolution. *Curr. Top. Microbiol. Immunol.*, 1990, vol. 157, pp. 1–18.
6. Schulman A. H., Flavell A. J., Noel Ellis T. H. The Application of LTR Retrotransposons as Molecular Markers in Plants. *Methods in Molecular Biology*, 2004, vol. 260, pp. 145–173.
7. Vandenbussche M., Gerats T. TE-Based Mutagenesis Systems in Plants. *Methods in Molecular Biology*, 2004, vol. 260, pp. 115–127.
8. The National Center for Biotechnology Information (NCBI) [Electronic resource]. Available at: <http://www.ncbi.nlm.nih.gov/genbank/> (accessed 12.01.2016).
9. Basic Local Alignment Search Tool (BLAST) [Electronic resource]. Available at: <http://blast.stva.ncbi.nlm.nih.gov/Blast.cgi> (accessed 13.01.2016).
10. Grandbastien M.-A. Retroelements in higher plants. *Trends Genet.* 1992, vol. 8, pp. 103–108.
11. Bennetzen J. L. The contributions of retroelements to plant genome organization, function and evolution. *Trends Microbiol.*, 1996, vol. 4, pp. 347–353.
12. Schmidt T. LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. *Plant Mol. Biol.*, 1999, vol. 40, pp. 903–910.

Информация об авторах

Рубель Илья Эдуардович – младший научный сотрудник лаборатории генетики и биотехнологии. Институт леса Национальной академии наук Беларуси (246001, г. Гомель, ул. Пролетарская, 71, Республика Беларусь). E-mail: illiarubel@yahoo.com

Баранов Олег Юрьевич – кандидат биологических наук, доцент, ведущий научный сотрудник лаборатории генетики и биотехнологии. Институт леса Национальной академии наук Беларуси (246001, г. Гомель, ул. Пролетарская, 71, Республика Беларусь). E-mail: betula-belarus@mail.ru

Пантелеев Станислав Викторович – кандидат биологических наук, старший научный сотрудник лаборатории генетики и биотехнологии. Институт леса Национальной академии наук Беларуси (246001, г. Гомель, ул. Пролетарская, 71, Республика Беларусь). E-mail: pukidesu@gmail.com

Разумова Ольга Александровна – кандидат биологических наук, старший научный сотрудник лаборатории генетики и биотехнологии. Институт леса Национальной академии наук Беларуси (246001, г. Гомель, ул. Пролетарская, 71, Республика Беларусь). E-mail: o-kovalevich@mail.ru

Гущин Владимир Алексеевич – кандидат биологических наук, научный сотрудник кафедры вирусологии биологического факультета МГУ имени М. В. Ломоносова (119991, г. Москва, ул. Ленинские горы, д. 1, стр. 12, Российская Федерация). E-mail: wowaniada@gmail.com

Макаров Валентин Владимирович – кандидат биологических наук, старший научный сотрудник отдела физических методов измерений. НИИ физико-химической биологии имени А. Н. Бело-

зерского МГУ имени М. В. Ломоносова (119991, г. Москва, ул. Ленинские горы, д. 1, стр. 40, Российская Федерация). E-mail: makarovvalentine@gmail.com

Information about the authors

Rubel' Il'ya Eduardovich – Junior Researcher, the Laboratory of Genetics and Biotechnology. Institute of Forest of the National Academy of Sciences of Belarus (71, Proletarskaya str., 246001, Gomel, Republic of Belarus). E-mail: illiarubel@yahoo.com

Baranov Oleg – PhD (Biology), Associate Professor, Leading Researcher, the Laboratory of Genetics and Biotechnology. Institute of Forest of the National Academy of Sciences of Belarus (71, Proletarskaya str., 246001, Gomel, Republic of Belarus). E-mail: betula-belarus@mail.ru

Panteleev Stanislav Victorovich – PhD (Biology), Senior Researcher, the Laboratory of Genetics and Biotechnology. Institute of Forest of the National Academy of Sciences of Belarus (71, Proletarskaya str., 246001, Gomel, Republic of Belarus). E-mail: pukidesu@gmail.com

Razumova Ol'ga Aleksandrovna – PhD (Biology), Senior Researcher, the Laboratory of Genetics and Biotechnology. Institute of Forest of the National Academy of Sciences of Belarus (71, Proletarskaya str., 246001, Gomel, Republic of Belarus). E-mail: o-kovalevich@mail.ru

Gushchin Vladimir Alekseevich – PhD (Biology), researcher, the Department of Virology. Faculty of the Biology of the Lomonosov Moscow State University (1-12, Leninskiye Gory str., 119991, Moscow, Russian Federation). E-mail: wowaniada@gmail.com

Makarov Valentin Vladimirovich – PhD (Biology), Senior Researcher, Physical Measurement Methods Department. Belozersky Research Institute of Physico-Chemical Biology of Lomonosov Moscow State University (1-40, Leninskiye Gory str., 119991, Moscow, Russian Federation). E-mail: makarovvalentine@gmail.com

Поступила 29.03.2016