

УДК 575.23

## ЭФФЕКТИВНЫЙ И ЧУВСТВИТЕЛЬНЫЙ МЕТОД ДЕТЕКЦИИ ИНСЕРЦИЙ ТРАНСПОЗОНОВ РАСТЕНИЙ НА ОСНОВЕ ЦЕЛЕВОГО НАНОПОРОВОГО СЕКВЕНИРОВАНИЯ АМПЛИКОНОВ

© 2024 г. П. Ю. Меркулов<sup>1, 2</sup>, М. А. Серганова<sup>1, 2</sup>, М. В. Дудников<sup>1, 2</sup>, И. В. Киров<sup>1, 2, \*</sup><sup>1</sup>Московский физико-технический институт, Московская область, Долгопрудный, 141701 Россия<sup>2</sup>Всероссийский научно-исследовательский институт сельскохозяйственной биотехнологии, Москва, 127550 Россия

\*e-mail: kirovez@gmail.com

Поступила в редакцию 01.11.2024 г.

После доработки 11.11.2024 г.

Принята к публикации 20.11.2024 г.

Транспозиция мобильных элементов — это важный источник генетического разнообразия растений. Быстрый скрининг популяции растений на наличие новых инсерций и идентификация мест их локализации в геноме необходимы при характеристике коллекций инсерционных мутантов. Нами разработан новый метод onTEI-seq, который основан на специфичной амплификации сайтов инсерций мобильных элементов и последующем нанопоровом секвенировании продуктов ПЦР. Разработанный метод апробирован при секвенировании эндогенных транспозонов *EVD* и *ONSEN Arabidopsis thaliana*. В результате за 2 ч получено высокое покрытие целевых транспозонов нанопоровыми ридами (>1000×). Технология onTEI-seq также использована для поиска инсерций *ONSEN* в популяции инсерционных мутантов *A. thaliana*, содержащих дополнительные копии LTR-ретротранспозонов этого семейства. Показано, что чувствительности метода onTEI-seq достаточно для того, чтобы идентифицировать инсерции в популяции растений, где доля их носителей составляет всего 2.5% (соотношение носителей инсерций к неносителям составляло 1 : 39). На основании полученных результатов можно сделать вывод, что onTEI-seq — чувствительный и быстрый метод для идентификации новых инсерций и определения их локализации в геноме.

**Ключевые слова:** LTR-ретротранспозоны, нанопоровое секвенирование, инсерции, *Arabidopsis thaliana*

**DOI:** 10.56304/S0234275824060097

Мобильные элементы (МЭ) занимают значительную долю генома большинства видов растений. Транспозиционная активность МЭ в клетках растений находится под строгим эпигенетическим контролем [1]. Именно поэтому, чтобы преодолеть “хозяйский надзор”, для исследования МЭ получают мутантные растения, дефектные по ключевым генам систем эпигенетического контроля [2]. Такие растения фертильны и отличаются повышенной активностью МЭ, что важно для их изучения. Однако геном растений содержит МЭ, которые способны к транспозиции только в определенных условиях у представителей с активной системой сайленсинга МЭ (растения “дикого типа”). Такие МЭ и условия их активации уже известны для ряда растений, включая *Arabidopsis thaliana*

(например, МЭ *ONSEN*). Для функциональной геномики особый интерес представляют коллекции растений дикого типа, содержащие инсерции МЭ (ИМЭ) в различных генах. Использование коллекций инсерционных мутантов требует детальной характеристики и каталогизации инсерционных сайтов целевых МЭ. Учитывая, что число растений в таких коллекциях может варьировать от сотен до десятков тысяч, остро стоит проблема идентификации ИМЭ в рамках первичной характеристики коллекций. Идентификация ИМЭ в геноме — задача сложная, в первую очередь потому, что большая часть генома растений представлена разными типами повторяющейся ДНК (сателлитные повторы и диспергированные повторы, включающие мобильные элементы). Эту “темную сторону” генома трудно исследовать, используя в качестве инструмента технологии секвенирования, основанные на коротких прочтениях, которые широко распространены для идентификации ИМЭ в растениях [3–5]. Секвенирование длинными прочтениями с использованием технологий PacBio или Oxford Nanopore позволяет кратно увеличить

**Список сокращений:** ИМЭ — инсерции мобильных элементов; МЭ — мобильные элементы; CANS (Cas9-targeted nanopore sequencing) — Cas9-направляемое нанопоровое секвенирование; СТАВ (cetyltrimethylammonium bromide) — цетилтриметиламмония бромид; GW (genome walking) — метод “прогулки по геному”; LTR (long terminal repeat) — длинный концевой повтор.

точность поиска инсерций в высокоповторяющихся регионах генома [6–8]. Хотя подход полногеномного секвенирования (WGS) теоретически позволяет идентифицировать все ИМЭ в отдельном растении, его применение в исследованиях больших геномов все еще затратно. В связи с этим было предложено несколько подходов для целевого секвенирования ИМЭ без “обращения” к WGS. Эти подходы основаны на обогащении фрагментами ДНК с инсерциями при помощи биотинилированных зондов (метод TE-capture) [3], амплификации (метод *Mu*-Illumina) [9] или микрофлюидной технологии и проточной цитометрии (метод Xdrop) [10]. Недавно разработан метод CANS (Cas9-targeted nanopore sequencing), который основан на Cas9-направляемом нанопоровом секвенировании ИМЭ [8]. Хотя CANS позволяет идентифицировать наследуемые и соматические ИМЭ, этот метод не подходит для массового скрининга больших коллекций растений.

На основании вышеизложенного целью проведенного нами исследования была разработка нового метода, подходящего для идентификации ИМЭ в коллекциях инерционных мутантов растений, используемых для функциональной геномики.

Принцип метода, который назван onTEI-seq, представлен на рис. 1а. Метод включает несколько этапов: (1) фрагментация ДНК, (2) удаление коротких фрагментов ДНК, (3) “затупление” и дефосфорилирование концов ДНК, (4) лигирование адаптеров GW, (5) амплификация с использованием адаптер- и транспозонспецифичных праймеров, (6) очистка ПЦР-продуктов, (7) подготовка библиотеки и нанопоровое секвенирование ампликонов, (8) детекция инсерций при помощи инструмента nanotei [7].

ДНК *Arabidopsis thaliana* выделяли из молодых листьев по СТАВ-методу [11], после чего фрагментировали путем многократного (30 раз) пропускания раствора ДНК (200 мкл, 5 нг/мкл) через иглу калибра G30 (SFM Hospital Products GmbH, Германия). Для удаления коротких фрагментов и концентрирования к раствору ДНК добавляли магнитные частицы Ampure XP (Beckman Coulter, США) в соотношении 0.7 : 1 (частицы : образец). После двукратной промывки частиц с иммобилизованной ДНК 70%-ным раствором этанола ДНК элюировали 30 мкл деионизированной воды, свободной от нуклеаз. Далее проводили “затупле-

ние” концов с использованием ДНК-полимеразы I *Escherichia coli* (“СибЭнзим”, Россия). Дефосфорилирование ДНК проводили щелочной фосфатазой Quick CIP (New England Biolabs, Великобритания). Для лигирования GW-адаптеров два олигонуклеотида (5'-GTAATACGACTCACTATAGGGCACGCGTGGTTCGACGGCCCGGGCTGGT-3'; PO<sub>4</sub>-5'-ACCAGCCC-3'-NH<sub>2</sub>) смешивали в равной молярной концентрации (50 мкМ). Реакцию лигирования GW-адаптеров к дефосфорилированной ДНК проводили с использованием T4 ДНК-лигазы (“Евроген”, Россия) в течение 16 ч при 14°C.

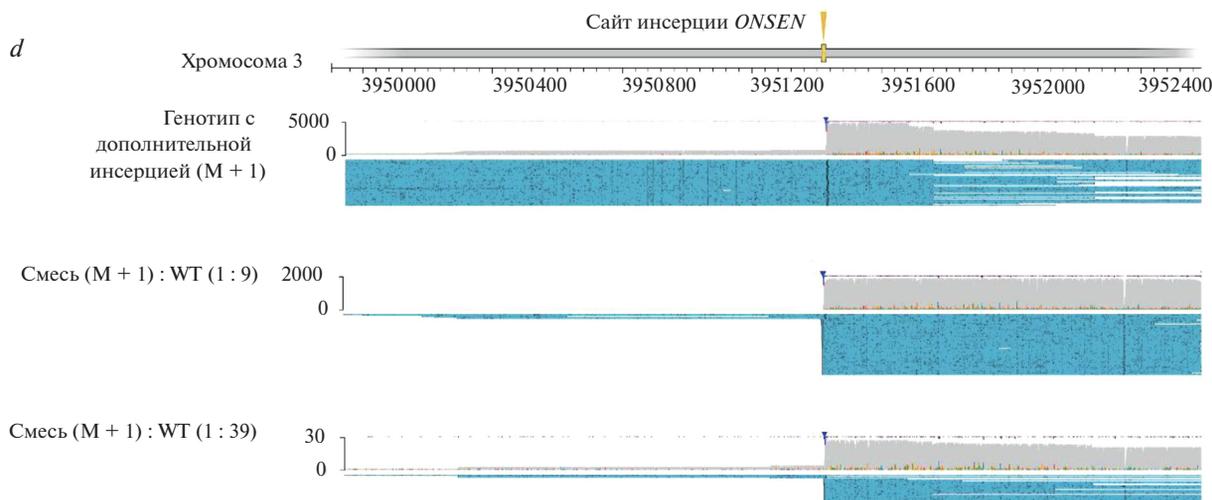
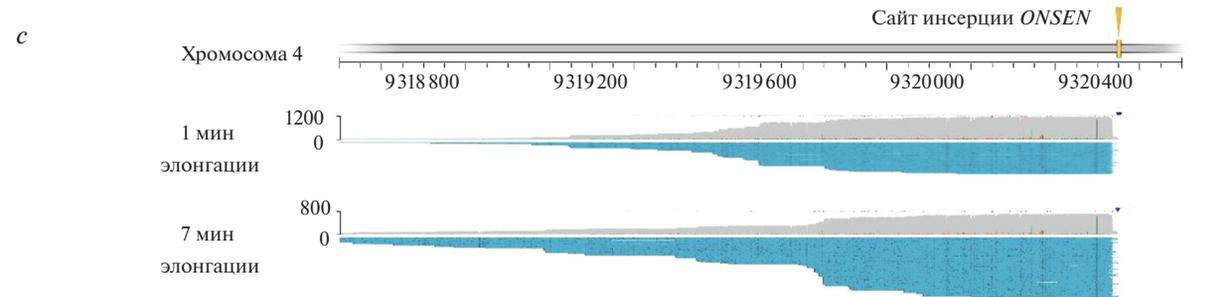
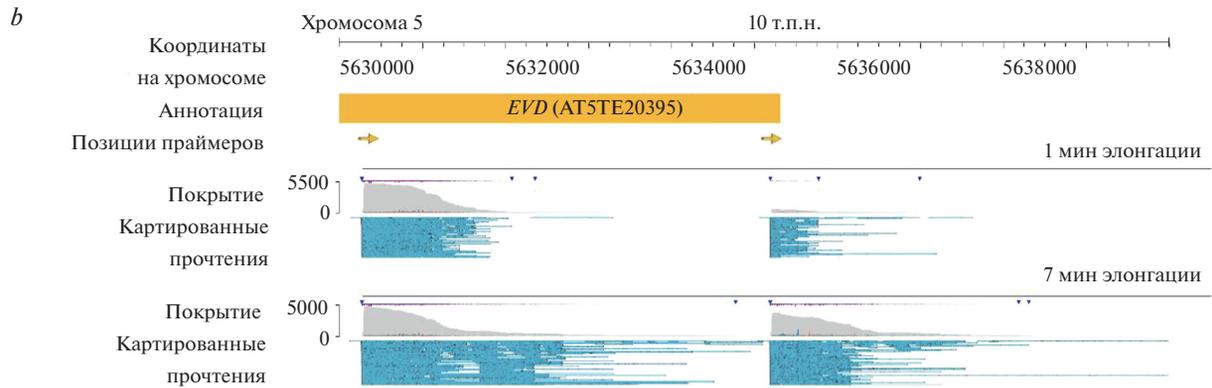
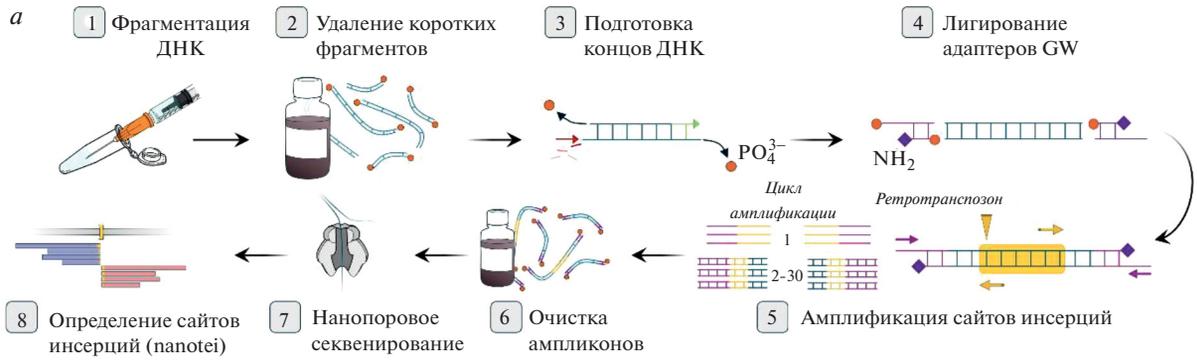
Для тестирования метода onTEI-seq использовали адаптерспецифичный праймер (PO<sub>4</sub>-5'-GTAATACGACTCACTATAGGGC-3') в комбинации с одним праймером к *EVD* (PO<sub>4</sub>-5'-CACAAAGGTACTATСТААТТСТАТСААТ-3') или *ONSEN* (PO<sub>4</sub>-5'-GCCTCCAAACTACAAAАТАТСТААА-3'), или с двумя к *ONSEN* (PO<sub>4</sub>-5'-CGGGACTTGGGAAGGGAACAT-3' и PO<sub>4</sub>-5'-TCAAGTTTAAAGAGGGG-GTGTG-3') – транспозонспецифичными праймерами, расположенными на последовательности длинных концевых повторов (LTR). Амплификацию проводили с использованием набора БиоМастер LR HS-ПЦР 2× (“БиоЛабМикс”, Россия). Полученные ампликоны очищали на магнитных частицах Ampure XP в соотношении 0.7 : 1 (частицы : образец).

Баркодирование образцов и подготовку библиотек для нанопорового секвенирования проводили в соответствии с рекомендациями производителя при помощи наборов NBD104 и LSK110 соответственно (Oxford Nanopore Technologies, Великобритания). Нанопоровое секвенирование выполняли на секвенаторе MinION с проточной ячейкой R9.4.1. Бейсколлинг выполняли с использованием программы Guppy 6.4.6 (Oxford Nanopore Technologies, 2019).

Для разработки и апробации метода использовали растения *A. thaliana*; для секвенирования – LTR-ретротранспозон *EVD* (AT5TE20395) и семейство LTR-ретротранспозонов *ONSEN*. Для увеличения длины прочтений нанопорового секвенирования критически важный этап – это амплификация. Мы сравнили длины прочтений onTEI-seq для *EVD* после амплификации в течение 1 и 7 мин и за 1 ч получили 48000 и 40000 ридов соответственно; при этом их длина была значительно больше при 7-минутной амплификации (рис. 1b). Важно

**Рис. 1.** Принцип метода onTEI-seq, и покрытие целевых транспозонов и инсерционных сайтов прочтениями. *a* – Схема метода onTEI-seq; *b* – покрытие концевых участков ретроинвертированного транспозона *EVD* ридами onTEI-seq после 1 и 7 мин элонгации; *c* – покрытие псевдоинсерции ретроинвертированного транспозона *ONSEN* ридами onTEI-seq после 1 и 7 мин элонгации; *d* – детекция инсерции *ONSEN* после onTEI-seq на ДНК растения с инсерцией (M + 1), разведенной ДНК растения дикого типа (WT) в разных соотношениях.

**Fig. 1.** Principle of the onTEI-seq method, and coverage of target transposons and insertion sites by reads. *a* – Scheme of the onTEI-seq method; *b* – coverage of the terminal regions of the *EVD* retrotransposon by onTEI-seq reads after 1 and 7 min of elongation; *c* – coverage of the pseudo-insertion of the *ONSEN* retrotransposon by onTEI-seq reads after 1 and 7 min of elongation; *d* – *ONSEN* insertion detection after onTEI-seq on the plant DNA with insertions (M + 1), diluted wild-type plant DNA (WT) in different ratios.



отметить, что покрытие концевых фрагментов *EVD* составило  $5100\times$  и  $4600\times$ .

С целью определить возможность использования данных onTEI-seq для детекции инсерции мы использовали разработанный метод onTEI-seq для секвенирования ретротранспозона *ONSEN* и картирования ридов в геноме *A. thaliana* TAIR10 (The Arabidopsis Information Resource версии 10: [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_000001735.3/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001735.3/)). Ранее нами показано, что в данной версии генома не включена одна копия *ONSEN1*, локализованная на хромосоме-4 (позиция 9320524–9320579) и представляющая псевдоинсерцию, подходящую для тестирования onTEI-seq [7]. В результате проведенного картирования подтверждено, что обнаруженная с помощью данных onTEI-seq псевдоинсерция имела покрытие  $1174\times$  и  $691\times$  в вариантах с 1- и 7-минутной элонгацией соответственно (рис. 1с).

На следующем этапе, используя разработанный метод onTEI-seq, мы провели секвенирование генома растений *A. thaliana* с реальными инсерциями. Это исследование проведено на ранее созданной нами коллекции растений *A. thaliana*, в геном которых были включены новые инсерции *ONSEN* [8]. Для идентификации сайтов инсерции использовали onTEI-seq с праймерами на оба концевых фрагмента *ONSEN* растений из созданной коллекции. В результате получено 130767 ридов. Анализ данных с помощью программы NanoCasTE [8] показал наличие одной инсерции на хромосоме-3: 3951391–3951484 – с покрытием ридами  $776\times$  и  $4872\times$  для 5'- и 3'-участков соответственно (рис. 1d). Разница в покрытии указывает на разную эффективность амплификации подобранных праймеров.

Скрининг коллекций растений на наличие инсерций включает этап их идентификации в пуле растений, а для этого необходим высокочувствительный метод детекции. Для оценки чувствительности метода onTEI-seq проведено секвенирование ДНК растения с инсерцией, предварительно смешанной с ДНК “дикого типа” в соотношениях 1 : 9 и 1 : 39. В результате было получено 121756 и 89574 прочтений. Целевая инсерция имела покрытие  $1883\times$  и  $27\times$  соответственно. Таким образом, соотношение 1 : 9 (т.е., доля мутантной ДНК составляет 10%) обеспечивало достаточное покрытие сайта инсерции, но при разведении в 40 раз (1 : 39; доля мутантной ДНК – 2.5%) наблюдалось падение покрытия почти в 1.5 раза.

Заметим, что разработанный метод имеет и некоторые ограничения. Во-первых, по данным onTEI-seq невозможно определить ДНК-метилирование; во-вторых, это первичная оптимизация праймеров и, в-третьих, неспецифичная амплификация.

Тем не менее, на основании полученных результатов можно заключить, что метод onTEI-seq пригоден для скрининга инсерций целевых транспозонов в популяциях растений. onTEI-seq имеет

ряд преимуществ в идентификации инсерций целевых транспозонов по сравнению с другими технологиями: (1) высокая скорость определения инсерций (весь протокол от выделения ДНК до получения результатов занимает несколько суток); (2) возможность определять инсерции в смеси популяции растений (bulk-анализ); (3) возможность одновременной детекции инсерций для набора транспозонов (мультиплексирование); (4) низкая стоимость инсталляции и запуска метода благодаря особенностям нанопорового секвенирования; (5) небольшое число пор на ячейки для достижения высокого покрытия сайта инсерции (все эксперименты проведены на ячейках с 300–500 пор), что позволяет проводить более 5 запусков на одной ячейке.

#### ФИНАНСИРОВАНИЕ

Работа выполнена при поддержке Российского научного фонда (грант № 22-64-00076).

#### ЭТИЧЕСКИЕ НОРМЫ

В данной статье не содержится исследований с участием животных или людей.

#### КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

#### СПИСОК ЛИТЕРАТУРЫ

1. Liu B., Zhao M. How transposable elements are recognized and epigenetically silenced in plants? *Curr. Opin. Plant Biol.*, 2023, 99, 102428. <https://doi.org/10.1016/j.pbi.2023.102428>
2. Sigman M.J., Panda K., Kirchner R., McLain L.L., Payne H., Peasari J.R., Husbands A.Y., Slotkin R.K., McCue A.D. An siRNA-guided ARGONAUTE protein directs RNA polymerase V to initiate DNA methylation. *Nat. Plants*, 2021, 7(11), 1461–1474. <https://doi.org/10.1038/s41477-021-01008-7>
3. Quadrana L., Silveira A.B., Mayhew G.F., LeBlanc C., Martienssen R.A., Jeddloh J.A., Colot V. The Arabidopsis thaliana mobilome and its impact at the species level. *eLife*, 2016, 5, e15716. <https://doi.org/10.7554/eLife.15716>
4. Vendrell-Mir P., Barteri F., Merenciano M., González J., Casacuberta J.M., Castanera R. A benchmark of transposon insertion detection tools using real data. *Mobile DNA*, 2019, 10(1), 53. <https://doi.org/10.1186/s12864-024-10314-1>
5. Cai X., Lin R., Liang J., King G.J., Wu J., Wang X. Transposable element insertion: a hidden major source of domesticated phenotypic variation in *Brassica rapa*. *Plant Biotechnol. J.*, 2022, 20(7), 1298–1310. <https://doi.org/10.1111/pbi.13807>
6. Shahid S., Slotkin R.K. The current revolution in transposable element biology enabled by long reads. *Curr. Opin. Plant Biol.*, 2020, 54, 49–56. <https://doi.org/10.1016/j.pbi.2019.12.012>
7. Kirov I., Merkulov P., Dudnikov M., Polkhovskaya E., Komakhin R.A., Konstantinov Z., Gvaramiya S., Ermo-

- laev A., Kudryavtseva N., Gilyok M., Divashuk M.G., Karlov G.I., Soloviev A. Transposons hidden in *Arabidopsis thaliana* genome assembly gaps and mobilization of non-autonomous LTR retrotransposons unravelled by nanotei pipeline. *Plants*, 2021, 10(12), 2681. <https://doi.org/10.3390/plants10122681>
8. Merkulov P., Gyaramiya S., Dudnikov M., Komakhin R., Omarov M., Kocheshkova A., Konstantinov Z., Soloviev A., Karlov G., Divashuk M., Kirov I. Cas9-targeted nanopore sequencing rapidly elucidates the transposition preferences and DNA methylation profiles of mobile elements in plants. *J. Integr. Plant Biol.*, 2023, 65(10), 2242–2261. <https://doi.org/10.1111/jipb.13555>
9. Williams-Carrier R., Stiffler N., Belcher S., Kroeger T., Stern D.B., Monde R.A., Coalter R., Barkan A. Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy mutator lines of maize. *Plant J.*, 2010, 63(1), 167–177. <https://doi.org/10.1111/j.1365-313X.2010.04231.x>
10. Madsen E.B., Höijer I., Kvist T., Ameer A., Mikkelsen M.J. Xdrop: targeted sequencing of long DNA molecules from low input samples using droplet sorting. *Hum. Mutat.*, 2020, 41(9), 1671–1679. <https://doi.org/10.1002/humu.24063>
11. Pucker B., Kleinbölting N., Weisshaar B. Large scale genomic rearrangements in selected *Arabidopsis thaliana* T-DNA lines are caused by T-DNA insertion mutagenesis. *BMC Genomics*, 2021, 22(1), 599. <https://doi.org/10.1186/s12864-021-07877-8>

## A Sensitive Method Based on Targeted Nanopore Amplicon Sequencing for Detection of Transposon Insertions in Plants

P. Yu. Merkulov<sup>a, b</sup>, M. A. Serganova<sup>a, b</sup>, M. V. Dudnikov<sup>a, b</sup>, and I. V. Kirov<sup>a, b, #</sup>

<sup>a</sup>Moscow Institute of Physics and Technology, Dolgoprudny, 141701 Russia

<sup>b</sup>All-Russia Research Institute of Agricultural Biotechnology, Moscow, 127550 Russia

<sup>#</sup>e-mail: kirovez@gmail.com

**Abstract**—Transposition of mobile elements is an important source of genetic diversity in plants. Rapid screening of plant populations for new insertions and identification of their localization sites in the genome are essential for characterizing collections of insertional mutants. We have developed a new onTEI-seq method based on specific amplification of mobile element insertion sites and subsequent nanopore sequencing of PCR products. The developed method was validated for sequencing endogenous transposons from the *EVD* and *ONSEN* families of *Arabidopsis thaliana*. As a result, high coverage of target transposons with nanopore reads (>1000×) was obtained within 2 h. The onTEI-seq technology was also used to search for *ONSEN* element insertions in the population of *A. thaliana* insertional mutants containing additional copies of LTR retrotransposons of this family. The sensitivity of the onTEI-seq method was sufficient to detect insertions in the plant population where the proportion of their carriers was only 2.5% (the ratio of insertion carriers to non-carriers was 1 : 39). The results indicate that onTEI-seq is a sensitive and rapid method for identifying new insertions and determining their localization in the genome.

**Keywords:** LTR retrotransposons, nanopore sequencing, insertions, *Arabidopsis thaliana*

Свидетельство о регистрации средства массовой информации  
 ПИ № ФС77-75946 от 13 июня 2019 г., выдано Федеральной службой по надзору в сфере связи,  
 информационных технологий и массовых коммуникаций (Роскомнадзор)

Подписано к печати 20.08.2024 г. Дата выхода в свет 02.09.2024 г. Формат 60 × 88<sup>1</sup>/<sub>8</sub> Усл. печ. л. 14.76 Уч.-изд. л. 15.0  
 Тираж 27 экз. Зак. 7363 Цена договорная

Учредители: Федеральное государственное бюджетное учреждение  
 «Национальный исследовательский центр «Курчатовский институт»

Издатель: НИЦ «Курчатовский институт», 123182, Россия, г. Москва, пл. Академика Курчатова, д. 1  
 Исполнитель по контракту № 35-3-24-44-5 от 22.04.2024 г. ООО «Объединённая редакция»,

109028, г. Москва, Подкопаевский пер., д. 5, каб. 6

Отпечатано в типографии «Book Jet» (ИП Коняхин А.В.),  
 390005, г. Рязань, ул. Пушкина, 18, тел. (4912) 466-151

16+