

## Поиск бифункциональных 3-дегидрохинатадегидратазы/шикиматдегидрогеназ *Cornus sericea* L. на основе *de novo* секвенирования транскриптома

Е.Р. Никонорова\*, Т.А. Кроль, Д.Н. Балеев

ФГБНУ «Всероссийский научно-исследовательский институт лекарственных и ароматических растений»,  
Москва, Россия

\* [gatiatulinaer@gmail.com](mailto:gatiatulinaer@gmail.com)

**Введение:** Гидролизруемые танины (ГТ) – класс фенольных соединений растений, основной функцией которых является защита растений от абиотических и биотических стрессоров, таких как патогены, вредители и УФ-излучение. Однако механизм их синтеза и накопления в различных видах растений изучен слабо. Растения *Cornus sericea* L. накапливают большое количество ГТ, но в связи со сложностью их химической структуры и отсутствием данных секвенирования генома и транскриптома *C. sericea*, точный механизм синтеза ГТ не известен.

**Цель:** Целью данной работы было провести секвенирование, сборку *de novo* и анализ транскриптома *C. sericea*, в особенности бифункциональной 3-дегидрохинатадегидратазы/шикиматдегидрогеназы (ДХД/ШДГ), одного из ключевых ферментов синтеза ГТ.

**Материалы и методы:** Образцы молодых и зрелых листьев, незрелых плодов, бутонов и цветков *C. sericea* были собраны в июле 2024 г. в Ботаническом саду ФГБНУ ВИЛАР (Москва). Тотальную РНК извлекали с помощью набора RNeasy Mini Kit и обрабатывали турбо-ДНКазой. Концентрацию определяли с помощью Qubit RNA HS Assay Kit. Подготовку библиотеки для секвенирования нового поколения (NGS) проводили с использованием набора TruSeq Stranded mRNA (Illumina). Секвенирование проводили на платформе NovaSeq 6000, 2 × 101 п.н. (Illumina Technologies, Сан-Диего, США). Для подготовки библиотеки для нанопорового секвенирования (Oxford Nanopore Technologies, ONT) РНК была преобразована в кДНК с использованием набора для синтеза кДНК Mint с 20-25 циклами амплификации. Библиотеку готовили с использованием стандартного протокола для геномной ДНК с набором SQK-LSK114. Обработка ONT прочтений выполняли с помощью Guppy 6.5.7 и NTproc (<https://github.com/shelkmike/NTproc/>). Хвосты PolyA ONT прочтений удаляли с помощью cutadapt v.5.0 с параметром “-a A{100}”. Демультиплексирование коротких ридов и обрезку адаптеров – с помощью Illumina bcl2fastq 2.20 и Skewer 0.2.2. Качество FASTQ анализировали с помощью FastQC для Illumina и NanoPlot для ONT. Сборку *de novo* выполняли с помощью trans2express (<https://github.com/albidgy/trans2express>), tnaQUAST и BUSCO (Benchmarking Unique Single Copy Orthologs) использовали для анализа качества сборки. Аннотацию проводили с помощью blastx с e-value 1e-10 и баз данных Uniprot и RefSeq. Анализы путей KEGG выполняли на сервере KAAS (<https://www.genome.jp/tools/kaas/>), GO – при помощи eggNOG-mapper. Картирование коротких и длинных прочтений на сборку транскриптома выполняли с помощью minimap2. ProtParam и InterProScan использовали для проверки потенциальной стабильности и функциональности изоформ ДХД/ШДГ.

**Результаты:** Окончательная сборка транскриптома состояла из 29124 транскриптов и 48139967 п.н. с самым большим транскриптом длиной 27663 п.н. По данным BUSCO, сборка транскриптома охватила 88.4% ортологичных генов из набора данных эвдикот (eudicots\_odb10), 90.4% генов растений (viridiplantae\_odb10) и 90.6% эукариотических генов (eukaryota\_odb10), что характеризует ее как достаточно полную. При анализе GO данных только часть транскриптов (19 098 – 65,6%) была аннотирована, где 4555 транскриптов были функционально классифицированы как относящиеся к «биологическому процессу», 854 – к «клеточному компоненту», 2756 – «молекулярной функции». По данным анализа KEGG были аннотированы 8742 транскриптов. С помощью KEGG Mapper были реконструированы карты путей, из которых 3741 относились к метаболизму, 1876 – заболеваниям человека (что возможно, так как присвоение идентификаторов основано на ортологии, и многие из них имеются и у растений, и у животных), 1024 – к обработке генетической информации, 734 – к клеточным процессам, 706 – системам организма и 474 – к обработке информации из окружающей среды. После функциональной аннотации были определены 5 потенциальных изоформ ДХД/ШДГ, с длиной 500–532 аминокислот и предсказанной молекулярной массой от 54,3 до 57,8 кДа, причем для одной из изоформ был выявлен дополнительный минорный вариант за счет удержания интрона.

**Выводы:** Впервые проведены секвенирование, сборка *de novo* и функциональная аннотация транскриптома *C. sericea*. Идентифицированы 5 потенциальных изоформ ДХД/ШДГ.

**Финансирование:** Работа выполнена согласно государственному заданию по теме FGUU-2024-0001.