

Tytuł: Budowanie modelu: rozwój zasobów genomowych dla *Ferula communis*, tradycyjnej rośliny leczniczej z rodziny baldaszkowatych (Apiaceae)

Autor: Mergi Daba Dinka

Streszczenie

Rośliny okrytozalążkowe reprezentują najbardziej zróżnicowaną grupę roślin lądowych. Za jedną z przyczyn ich sukcesu ewolucyjnego uważa się bardzo specyficzną budowę genomów charakteryzującą się dużą plastycznością. Jednak z powodu braku danych dla większości grup okrytonasiennych nasze zrozumienie architektury genomu jest na dzień dzisiejszy ograniczone. W tej pracy podjąłem badania, których celem jest wypełnienie luki w naszej wiedzy o genomach roślin okrytozalążkowych, składając i analizując genom *Ferula communis*, gatunku z ekonomicznie ważnej rodziny baldaszkowatych (Apiaceae).

Genom jądrowy *F. communis* został złożony z wykorzystaniem sekwencji uzyskanych dzięki technologii sekwencjonowania nowej generacji, obejmującej zarówno technologię Illumina, jak i Oxford Nanopore Technology (ONT). W procesie składania genomu użyto szerokiego spektrum narzędzi bioinformatycznych, w tym programów wykorzystujących tylko krótkie fragmenty, tylko długie oraz oba typy odczytów. Najlepszy jakościowo genom został złożony przez program Flye wykorzystujący tylko długie odczyty z technologii ONT. Wielkość uzyskanego genomu wynosiła 2,8 Gb, na który składało się 59 178 kontigów. Jakość genomu oszacowana na podstawie metryki N50 wynosiła 0,17 Mb, zaś kompletności genomu bazująca na genach BUSCO została określona na 93,2%. Elementy mobilne stanowiły 86,67% genomu, przy czym retrotranspozony *Gypsy* i *Copia* były szczególnie liczne, co podkreśla ich kluczową rolę w ewolucji genomu *F. communis*. Dzięki funkcjonalnej adnotacji genomu zidentyfikowałem 68 318 genów kodujących białka i 1 262 genów tRNA. W porównaniu z genomami marchwi, selera i kolendry należących do rodziny baldaszkowatych, *F. communis* wykazuje większą liczbę unikalnych genów oraz charakteryzuje się większą ekspansją rodzin genów. Wśród tych rozszerzonych rodzin zidentyfikowałem geny związane z mechanizmami obronnymi, odpowiedziami na stres i wernalizacją. Co ciekawe, *F. communis* dzieli więcej genów z odlegle spokrewnioną kolendrą niż z marchwią i selerem. Przykładem są geny związane z biosyntezą terpenoidów. Prawdopodobną przyczyną tego zaskakującego wyniku jest redukcja genomu u marchwi, prowadząca do redukcji liczby genów. Natomiast u selera, jego pierwotne środowisko wodne mogło wymagać specyficznego zestawu genów, który różni się od tego wymaganego w środowiskach lądowych.

Genom plastydowy *F. communis* został złożony przy użyciu dwóch programów: GetOrganelle i Novoplasty. Złożony genom obejmuje 166 696 pz i koduje 132 geny, w tym 87 kodujących białka, 37 tRNA i 8 genów rRNA. Genom *F. communis* odzwierciedla standardową strukturę czteroczęściową obserwowaną w innych genomach plastydowych okrytonasiennych, która obejmuje regiony IR, LSC i SSC. W genomie znajduje się 217 sekwencji mikrosatelitarnych (SSR), szczególnie w regionie LSC. Kilka spośród przestrzeni międzygenowych charakteryzowało się dużą różnorodnością nukleotydową co umożliwia potencjalne wykorzystanie ich jako znaczników filogenetycznych. Wśród wszystkich genów kodujących białka w genomie plastydowym *F. communis*, tylko gen *ccsA*, który odpowiada za przyłączanie hemu do cytochromu c, podlega silnej selekcji pozytywnej.

Genom mitochondrialny został złożony przy użyciu programu GetOrganelle i przeanalizowany pod kątem zawartości genów i wariantów strukturalnych. Analiza sekwencji mitochondrialnych dała wynik w postaci 16 kontigów o łącznej długości 250 278 pz co wskazuje, że genom *F. communis* nie występuje w postaci jednej kolistej cząsteczki DNA. W obrębie tych kontigów udało się scharakteryzować 37 genów kodujących białka, 3 geny rRNA i 20 genów tRNA. Spośród adnotowanych genów, osiem genów kodujących białka i trzy tRNA zawierały introny o różnych długościach. W genomie zidentyfikowano 183 sekwencji mikrosatelitarnych, z których 82% to powtórzenia mono- i dimerowe. Ponadto, określono 385 miejsc edycji RNA w 25 genach kodujących białka. 61,14% edytowanych miejsc prowadzi do zmiany aminokwasu z hydrofilowego na hydrofobowy, zaś 31,09% obejmuje zmiany między aminokwasami hydrofobowymi.

Prezentowane tutaj wyniki stanowią podstawę i dostarczają cenne informacje dla przyszłych badań nad genomiką, ewolucją i właściwościami leczniczymi tego ważnego gatunku z rodziny baldaszkowatych.