

prof. dr hab. Paweł Krajewski
Instytut Genetyki Roślin PAN w Poznaniu

Ocena pracy doktorskiej mgr inż. Anastasii Satyr

„Variation of the receptor-like protein gene family in the *Arabidopsis thaliana* genome revealed by nanopore sequencing and de novo assemblies”

Niniejsza ocena została wykonana na podstawie zlecenia zgodnego z uchwałą Rady Naukowej Instytutu Chemii Bioorganicznej PAN o powołaniu recenzentów w przewodzie doktorskim mgr inż. Anastasii Satyr z dnia 5 grudnia 2025 r.

1. Zakres pracy

Praca doktorska mgr inż. Anastasii Satyr opisuje wyniki badań mających na celu scharakteryzowanie zmienności sekwencji genów należących do rodziny Receptor-Like Protein (RLP) w wybranych akcesjach *Arabidopsis thaliana*. Główna hipoteza badawcza mówiła, że cel ten może być osiągnięty poprzez poznanie, z odpowiednią dokładnością, genomów badanych roślin za pomocą sekwencjonowania. Realizacja badań obejmowała opracowanie protokołu ekstrakcji genomowego DNA, sekwencjonowanie DNA metodą nanoporową, składanie i adnotację genomów, ocenę jakości uzyskanych sekwencji oraz analizę polimorfizmu genów RLP.

2. Opis formalny pracy

Praca składa się z sześciu rozdziałów, o układzie typowym dla rozpraw doktorskich. Wstęp rozpoczyna omówienie mechanizmów reakcji roślin na czynniki chorobotwórcze i przedstawienie głównych rodzin genów odporności, w tym genów RLP, oraz roli wariantów strukturalnych i transpozonów w zmienności genomów. Następnie przedstawione są metody identyfikacji wariantów strukturalnych korzystające z hybrydyzacji lub sekwencjonowania, ze szczególnym uwzględnieniem metody nanoporowej dostarczającej długich odczytów DNA, dla której opisane są także metody ekstrakcji DNA i narzędzia do analizy danych. Wstęp obejmuje też przegląd literatury dotyczącej analizy wariantów strukturalnych w *A. thaliana* i opis kolejnych uzyskanych genomów referencyjnych tej rośliny. Rozdziały 2 i 3 to, odpowiednio, przedstawienie celów prowadzonych badań i wykorzystanych materiałów i metod. Opisany jest sposób uprawy roślin, ekstrakcji DNA i przygotowania bibliotek do sekwencjonowania, po czym przedstawione są narzędzia bioinformatyczne użyte w kolejnych krokach analizy danych. Po przedstawieniu wyników w rozdziale 4 następuje dyskusja, zakończona sformułowaniem dziesięciu wniosków. Spis literatury liczy 192 pozycje, zaś cała praca 119 stron.

3. Ocena sposobu realizacji

Współczesne metody sekwencjonowania DNA pozwalają na szeroko zakrojone badania polimorfizmu genów i zmienności całych genomów pomiędzy formami roślinnymi. Jest to celowe szczególnie dla genów lub rodzin genów determinujących istotne własności roślin, takie jak odporność na choroby. Recenzowana rozprawa opisuje badania wpisujące się w ten nurt.

Informacje przedstawione w części wstępnej pokazują, że szczegółowy dobór tematyki badań odbywał się na podstawie szerokiej analizy interesujących klas genów oraz wiadomości o ich znanych i postulowanych funkcjach. Metodyka badań została dobrana na podstawie przeglądu dostępnych obecnie technik sekwencjonowania, a wybór techniki nanoporowej jest dobrze uzasadniony. Jako materiał badawczy zostały wybrane akcesje *A. thaliana*, dla których złożenia genomów nie były dostępne, pochodzące z różnych regionów geograficznych. Uwagę zwraca wyczerpujące opracowanie rezultatów badań nad zmiennością strukturalną w *Arabidopsis* prowadzonych na przestrzeni lat 2007-2025, które pozwoliło na dobranie materiału i metodyki. Wybór metod i narzędzi przetwarzania danych został poczyniony spośród wielu dostępnych programów właściwych dla analizy długich odczytów DNA; narzędzia te są dobrze opisane, wraz z podaniem zastosowanych wartości ich parametrów.

Przebieg i wyniki badań są opisane w sposób wyczerpujący. Opracowany protokół ekstrakcji DNA okazał się skuteczny, co pokazują przedstawione w Tab. 7 i 9 charakterystyki liczby, długości i jakości odczytów w uzyskanych zbiorach danych oraz dobre pokrycie genomu. Zaobserwowano pewną zmienność charakterystyk pomiędzy akcesjami, lecz wydaje się, że była ona na dopuszczalnym poziomie.

Składanie genomów wykonano najpierw za pomocą pięciu różnych narzędzi dla jednej z badanych akcesji. W wyniku porównania otrzymanych zbiorów kontigów do dalszych obliczeń wybrano narzędzia Canu i Flye, które dostarczyły najdłuższych kontigów, o najkorzystniejszych parametrach N50 i L50. Wyniki składania otrzymane tymi narzędziami dla wszystkich ośmiu akcesji oceniono szczegółowo pod względem liczby i długości kontigów oraz parametrów finalnie otrzymanych (z wykorzystaniem genomu referencyjnego ekotypu Col) skafoldów. Następnie oceniono całe genomy – pod względem zawartości genów (Busco) oraz kompletności regionów telomerowych i centromerowych.

Kolejnym etapem pracy była adnotacja genomów pod względem zawartości genów, wykonana poprzez odnalezienie genów obecnych w genomie referencyjnym oraz adnotację *de novo*. Na tej podstawie szczegółowo przeanalizowano polimorfizm genów RLP pomiędzy badanymi akcesjami, prezentując wszystkie zaobserwowane warianty strukturalne. Warianty przedstawiono i przedyskutowano w odniesieniu do pojedynczych genów oraz skupień genów, na poziomie sekwencji DNA i kodowanych białek. W analizie polimorfizmów i poziomów ekspresji genów posługiwano się dostępnymi publicznie zbiorami krótkich odczytów RNA.

Na podstawie powyższego należy stwierdzić, że zaplanowanie i przygotowanie badań odbyło się w sposób prawidłowy, na podstawie szerokiego przeglądu dostępnej wiedzy. Opracowany ciąg działań badawczych był spójny. Analizy danych zostały wykonane za pomocą właściwych narzędzi

bioinformatycznych. Uwagę zwraca bardzo wnikliwe podejście do oceny jakości uzyskanych danych sekwencyjnych oraz genomów powstałych przez składanie odczytów. Analiza jakości jest wielostronna i pokazała, że wyniki mogły posłużyć do wnioskowania na temat polimorfizmu wybranej rodziny genów. Bardzo wnikliwa jest również analiza wariantów strukturalnych.

Dyskusja wyników dotyczy, w pierwszej części, zagadnień związanych z sekwencjonowaniem nanoporowym i składaniem genomów - w odniesieniu do uzyskanych wyników, w aspekcie ich jakości oraz możliwych innych zastosowań. Druga część dyskusji zawiera uwagi na temat zaobserwowanej zmienności genów RLP, w aspekcie podziału genów ze względu na postulowane funkcje oraz lokalizację jako genów izolowanych lub w obrębie skupień. Elementy dyskusji bądź porównania z wynikami uzyskanymi przez innych autorów są też umieszczone w kilku miejscach w rozdziale 4. Kończące pracę konkluzje są dobrze umotywowane treścią pracy.

Podsumowując ocenę merytoryczną należy stwierdzić, że w przeprowadzonych badaniach zrealizowano postawiony cel polegający na konstrukcji dobrej jakości sekwencji genomowych wybranych akcesji *A. thaliana* i analizy tych sekwencji pod względem zmienności genów RLP wynikającej z obecności wariantów strukturalnych. Badania przeprowadzono za pomocą prawidłowo dobranych metod laboratoryjnych. Zastosowane narzędzia obliczeniowe zostały dobrze dobrane i użyte. Jakkolwiek są one stosowane obecnie w wielu projektach badawczych, to jednak zawsze stawiają wyzwania wynikające z charakterystyki dostępnego materiału biologicznego i wygenerowanych danych, tak więc analizy tego typu są dalekie od rutynowych.

4. Pytania

- Transformacja danych surowych na odczyty nukleotydów jest ważnym etapem przetwarzania danych. Czy na tym etapie testowano inne opcje narzędzia guppy, np. tryb „super dokładny”; jeżeli tak, to jak wypadło porównanie dokładności i kosztów obliczeń?
- W pierwszym kroku analizy, testowano wykorzystanie pięciu assemblerów dla danych z jednej akcesji. Dwa z nich, HASLR i SPAdes, dały wyniki znacznie gorsze od pozostałych. Czy można to wyjaśnić ich innym przeznaczeniem lub jakimiś znanymi własnościami?
- W analizie danych dla ośmiu akcesji zastosowano dwa narzędzia: Canu i Flye. W kilku miejscach następuje porównanie wyników uzyskanych tymi assemblerami, m.in. pod względem dokładności w regionach powtarzalnych (s. 69), ciągłości (dla niektórych akcesji, str. 72), polimorfizmów (s. 80). Czy te różnice można wyjaśnić przez jakieś własności metodyki lub algorytmów zastosowanych w tych programach?

5. Uwagi stylistyczne i terminologiczne

Praca jest napisana zrozumiałym i precyzyjnym językiem, prostym - na ile to możliwe w opisie wyników badań naukowych uzyskanych zaawansowanymi narzędziami. Zauważyłem drobne błędy w streszczeniu w j. polskim (np. „sekwencjonowanie długich odczytów” – lepiej „sekwencjonowanie długich odcinków”, „w regionach powtórzonych” – lepiej „regionach charakteryzujących się

powtarzalnością sekwencji”). Ponadto na str. 45/3 winno być “plant cultivation” zamiast “plant breeding”, zaś na str. 65/15 jest zbędny tekst „in the multiple challenge”.

6. Konkluzja

Recenzowana praca zawiera nowe wyniki naukowe poszerzające wiedzę o polimorfizmie genów RLP w *A. thaliana* i pangenomie tej rośliny. Wyniki zostały uzyskane z wykorzystaniem protokołu, który może być zastosowany w wielu innych sytuacjach. Nastąpiło już wykorzystanie wyników z tej pracy w innych zagadnieniach. Badania wymagały opanowania zarówno technik eksperymentalnych, jak i wielu procedur analizy danych. Doktorantka wykazała się umiejętnością prowadzenia badań w obu tych obszarach i interpretacji ich wyników. Na podkreślenie zasługuje dociekliwość rozważań związanych z oceną jakości danych i rezultatów oraz możliwym znaczeniem znalezionych polimorfizmów.

W związku z tym stwierdzam, że przedstawiona do recenzji rozprawa doktorska spełnia warunki określone w Ustawie z dnia 20 lipca 2018 roku prawo o szkolnictwie wyższym i nauce (Dz.U. z 2018 r. poz. 1668 ze zm.) oraz w Sposobie postępowania w sprawie nadania stopnia doktora w Instytucie Chemii Bioorganicznej PAN w Poznaniu (uchwała Rady Naukowej ICHB PAN nr 40/2025/Internet/RN_140 z dnia 18 marca 2025 r.) i wnioskuję do Rady Naukowej Instytutu Chemii Bioorganicznej PAN o dopuszczenie mgr inż. Anastasii Satyr do dalszych etapów postępowania o nadanie stopnia doktora.

