



Poznań, dnia 11.02.2026

Dr hab. Marek Żywicki

Zakład Biologii Obliczeniowej

Wydział Biologii

Uniwersytet im. Adama Mickiewicza w Poznaniu

**Recenzja pracy doktorskiej mgr Anastasii Satyr pt. „Variation of the receptor-like protein gene family in the *Arabidopsis thaliana* genome revealed by nanopore sequencing and *de novo* assemblies”**

Przedstawiona do recenzji praca doktorska mgr Anastasii Satyr wykonana została pod kierunkiem dr hab. Agnieszki Żmieńko w Zakładzie Genomiki Roślin Instytutu Chemii Bioorganicznej PAN w Poznaniu. Praca skupia się na badaniach wzorców zmienności strukturalnej (SV) w rodzinie genów RLP w ośmiu zróżnicowanych liniach *Arabidopsis thaliana*, z wykorzystaniem sekwencjonowania długich odczytów oraz składaniem sekwencji genomowych *de novo*.

Głównymi celami pracy doktorskiej były:

- opracowanie zoptymalizowanego protokołu izolacji wysokocząsteczkowego DNA (HMW DNA) z liści *Arabidopsis*;
- zsekwencjonowanie oraz złożenie *de novo* wysokiej jakości sekwencji genomowych dla ośmiu wybranych linii *Arabidopsis*;
- identyfikacja, adnotacja oraz porównanie regionów obejmujących geny RLP, w celu analizy ich zmienności strukturalnej oraz identyfikacji nowych wariantów genów RLP.

W wyniku realizacji pracy wszystkie powyższe cele zostały osiągnięte.

Wstęp pracy jest obszerny, logiczny i wzorowo udokumentowany. Stanowi klarowne i rzetelne wprowadzenie do problematyki naturalnej zmienności genetycznej roślin, ze szczególnym uwzględnieniem wariantów strukturalnych (SV) oraz ich roli w funkcjonowaniu systemu odpornościowego. Autorka charakteryzuje organizację odporności roślin typu PTI oraz ETI — oraz szczegółowo charakteryzuje rolę receptorów PRR, a zwłaszcza rodzin RLP i RLK, w rozpoznawaniu sygnałów pochodzących od patogenów i aktywowaniu odpowiedzi wewnątrzkomórkowych. Istotną część rozdziału poświęca również znaczeniu wariantów

strukturalnych jako kluczowego źródła zróżnicowania genomów roślinnych. Uzasadnienie wyboru *Arabidopsis thaliana* jako organizmu modelowego zastosowanego w pracy zostało przeprowadzone kompetentnie i spójnie, z odwołaniem do bogatej zmienności naturalnej i szerokiego zakresu dostępnych danych genomowych. Autorka wyczerpująco opisuje dotychczasowy stan wiedzy w tym zakresie oraz klarownie uzasadnia konieczność zastosowania w swojej pracy sekwencjonowania długimi odczytami. Wstęp stanowi silny fundament merytoryczny rozprawy, wyraźnie nakreślając kontekst biologiczny i metodologiczny pracy wraz z uzasadnieniem wyboru narzędzi i strategii badawczych. Pod względem struktury, argumentacji i kompletności prezentuje wysoki poziom naukowy i dobrze wprowadza czytelnika w zagadnienia analizowane w dalszych częściach pracy.

Następnie w sposób klarowny, konkretny i adekwatny do zakresu rozprawy Autorka przedstawia cele badawcze, po czym w kolejnym rozdziale zamieszcza szczegółowy i kompletny opis procedur eksperymentalnych oraz metod bioinformatycznych zastosowanych w trakcie realizacji pracy. Autorka zwięźle przedstawia wszystkie etapy pracy — od przygotowania materiału biologicznego, poprzez sekwencjonowanie, aż po zaawansowane analizy bioinformatyczne. Pomimo zwięzłości, opisy są kompletne a ich szczegółowość pozwala na jednoznaczną ocenę rzetelności uzyskanych wyników.

Główną część pracy stanowi rozdział 4, w którym przedstawione zostały uzyskane wyniki. Rozdział zaczyna się od opisu zoptymalizowanego protokołu izolacji wysokocząsteczkowego DNA, który został zoptymalizowany dla celów pracy (szczegółowo przedstawionego w Załączniku 1). Umożliwił on uzyskanie długich fragmentów genomowego DNA niezbędnych dla sekwencjonowania Nanopore. Wysoka jakość uzyskanego materiału znalazła swoje odbicie w przedstawionych wynikach sekwencjonowania Nanopore — dla każdej linii uzyskano średnio 14,5 Gb, z pokryciem zazwyczaj przekraczającym 90×. Odczyty charakteryzowały się znaczną długością (N50 w zakresie 12,6 do 19,8 z maksymalną długością odczytów powyżej 500 kb), stanowiąc solidną podstawę do wysokiej jakości złożeń genomów.

Następnie Autorka opisuje proces analizy bioinformatycznej uzyskanych odczytów. Jednym z elementów przeprowadzonych analiz była walidacja dostępnych assemblerów w celu wyboru najbardziej efektywnego dla uzyskanych danych. Spośród pięciu przetestowanych narzędzi wybrano *canu* i *flye*, generujące najpełniejsze i najstabilniejsze złożenia dla testowego genomu linii Mitterberg 2-185. Wybrane narzędzia posłużyły do opracowania dla każdego z ośmiu genomów dwóch niezależnych zestawów kontigów oraz reprezentacji chromosomowych, wykorzystując genom referencyjny Col-CC.v1 jako wzorzec dla konstrukcji scaffoldów. Dla każdego złożenia wykonana została kompleksowa analiza jakości złożenia obejmująca ocenę ciągłości (NG50, LG50), kompletności (BUSCO, reprezentacja telomerów i centromerów) oraz poprawności (QV). Zgodnie z uzyskanymi metrykami, uzyskane złożenia wykazywały wysoką jakość w obszarach euchromatycznych, jednak odtwarzalność rejonów powtórzonych była niska, pomimo zastosowania technologii długich odczytów. Zaobserwowano także komplementarne mocne strony obu assemblerów: *canu* lepiej odtwarza telomery, natomiast *flye* wykazuje wyższą efektywność dla centromerów.

W wyniku adnotacji uzyskanych genomów za pomocą narzędzia *Liftoff* zidentyfikowano ponad 27 tysięcy genów w każdym złożeniu. Dodatkowo w każdym genomie zidentyfikowano *de novo* ok. 300 nowych genów. Adnotacja transpozonów wykazała, że około 10% zsekwencjonowanych genomów stanowią elementy ruchome, w większości silnie metylowane (5mC), co wskazuje na ich epigenetyczne wyciszenie.

Najobszerniejsza część rozdziału dotyczy analizy zmienności genów RLP. Autorka identyfikuje wszystkie potencjalne geny RLP w każdym genomie i porównuje ich sekwencje, organizację domen oraz ewolucję między analizowanymi liniami. Wyniki wskazują na zróżnicowaną podatność tej rodziny na zmienność strukturalną: geny niezgrupowane są na ogół zachowawcze, natomiast geny występujące w parach i klastrach podlegają licznym zmianom strukturalnym, w tym delecjom, duplikacjom i insercjom transpozonowym. Przykłady obejmują m.in. PAV genu RLP3 w parze RLP2-RLP3, duplikację tandemową RLP30, insercje TE w parach RLP32-RLP33 oraz złożone zdarzenia w klastrach RLP13-RLP16, RLP39-RLP42 i RLP47-RLP50. Poprzez szczegółową i dogłębną analizę zachowawczości poszczególnych klastrów Autorka wykazuje, że geny RLP są znacznie bardziej dynamiczne ewolucyjnie, niż dotąd zakładano na podstawie danych uzyskanych technologiami krótkich odczytów.

Praca została napisana poprawnym językiem, z właściwym stosowaniem odniesień. Rozdziały dyskusja oraz wnioski trafnie podsumowują uzyskane wyniki oraz umieszczają je w kontekście obecnego stanu wiedzy. Niemniej, proszę o ustosunkowanie się podczas obrony do następujących pytań:

1. Czy wydajność opracowanego protokołu izolacji HMW DNA była porównywana z innymi protokołami?
2. Przedstawiony na Rycinie 9 rozkład długości odczytów uzyskanych dla linii Mitterberg 2-185 odstaje od pozostałych zestawów danych. Z jakich względów właśnie ta linia (zestaw danych) została wybrana jako referencyjna do przeprowadzenia walidacji assemblerów i czy odmienny rozkład długości odczytów mógł mieć wpływ na wyniki analizy i wybór assemblera?
3. Czy analizowane linie wykazywały podobieństwa w zakresie częstości występowania wariacji w obrębie genów RLP lub w rodzajach wariacji obserwowanych w obrębie klastrów genów RLP (np. czy są linie bardziej podatne na transpozycje lub PAV)?
4. W analizie wariacji wykorzystywane były uzyskane w pracy dane z sekwencjonowania długich odczytów, jak również dane uzyskane za pomocą krótkich odczytów. Czy pomiędzy poszczególnymi klastrami występowały typy wariacji, których częstość występowania w poszczególnych liniach powiązana byłaby z zastosowaną techniką sekwencjonowania?

Przedstawione powyżej pytania w żaden sposób nie umniejszają wysokiej jakości przedstawionej pracy doktorskiej. Zaprezentowane w pracy wyniki badań stanowią spójną całość, w kompleksowy sposób charakteryzując zmienność genów RLP u *Arabidopsis thaliana*. Zostały one opatrzone trafnymi wnioskami i wyczerpująco przedyskutowane w odniesieniu do istniejącej wiedzy. Przedstawiona praca doktorska świadczy o wysokiej dojrzałości naukowej Autorki oraz stanowi cenny wkład w badania nad mechanizmami zmienności



genetycznej roślin. Opisane w recenzowanej pracy wysokiej jakości złożenia genomów ośmiu linii *Arabidopsis thaliana* stanowią cenny zasób dla środowiska naukowego oraz dla dalszych analiz. Ponadto, ujęte w pracy aspekty techniczne (protokół izolacji HMW DNA, wnioski z walidacji assemblerów) stanowią cenną wiedzę dla naukowców podejmujących podobne tematy badawcze.

**Podsumowując, 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 Anastasii Satyr do dalszych etapów postępowania o nadanie stopnia doktora.**

Poznań, 11.02.2026

Dr hab. Marek Żywicki