UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytut Biologii Molekularnej i Biotechnologii

Poznań, 02.10.2025 r.

dr hab. Andrzej Zieleziński, prof. UAM
Pracownia Biologii Obliczeniowej

RECENZJA
rozprawy doktorskiej mgr Marii Nuc
pt. „Metodologia statystyczna badania zmian jakościowych transkryptomu roślin w aspekcie interakcji genotypu i środowiska”

1. Wartość naukowa rozprawy

Największą wartością naukową recenzowanej pracy jest opracowanie przez Autorkę oryginalnego podejścia do badania zmian jakościowych transkryptomu roślin w kontekście interakcji genotyp–środowisko. Metoda łączy bioinformatyczną analizę globalnego podobieństwa sekwencji transkryptomów (odległość Mash) z narzędziami statystycznymi, takimi jak redukcja wymiarowości (PCoA) oraz analiza wariancji (ANOVA/MANOVA), co pozwala formalnie uchwycić różnice jakościowe między transkryptomami. Cenny jest również formalny dowód, że odległość Mash spełnia aksjomaty metryki. W literaturze wspominano, że Mash jest metryką, jednak dotychczas brakowało potwierdzenia matematycznego. Argumentacja zawarta w rozprawie stanowi solidną podstawę, by traktować tę miarę jako użyteczny dystans filogenetyczny. W mojej ocenie, zarówno opracowana metoda, jak i jej formalne uzasadnienie stanowią istotny wkład w rozwój narzędzi do analizy porównawczej transkryptomów i otwierają perspektywę dalszych badań nad poszczególnymi genami i aspektami ilościowymi.

2. Ocena merytoryczna rozprawy

Pozytywnie oceniam również stronę merytoryczną pracy. Wstęp dobrze wprowadza w tematykę badań, definiuje kluczowe pojęcia i opisuje czynniki wpływające na zmienność transkryptomów. Być może warto byłoby jednoznacznie wyjaśnić, co Autorka rozumie przez „zmiany jakościowe transkryptomu”. Z dalszej części tekstu wynika, że chodzi o różnice w składzie i funkcji transkryptów, ale warto byłoby doprecyzować to już na początku, z zaznaczeniem, że zależą one także od czynników ilościowych, takich jak próg detekcji czy głębokość sekwencjonowania. Takie uściślenie ułatwiłoby zdefiniowanie celów pracy, zwłaszcza tych związanych z biologicznym znaczeniem metody, które obecnie mogłyby być sformułowane precyzyjniej.

Zastosowane metody zostały dobrane właściwie, a ich opis świadczy o wysokim poziomie warsztatu matematyczno-statystycznego Autorki. Wyniki przedstawiono przejrzyście, a wnioski są poprawne. Choć biologiczne interpretacje są ograniczone, Autorka jest tego świadoma i podejmuje próbę ich rozwinięcia poprzez analizę ontologii genów. Dyskusja jest rzeczowa i obiektywna, z uwzględnieniem zarówno mocnych stron, jak i ograniczeń metody, co potwierdza dobre zrozumienie konsekwencji przyjętego podejścia.

3. Uwagi merytoryczne do rozważenia i dyskusji

Doktorantka prezentuje wysoki poziom naukowy, a poniższe uwagi mają jedynie charakter dyskusyjny i wynikają z mojego szczególnego zainteresowania tematem.

1. Sekwencjonowanie NovaSeq przeprowadzono przy łagodnych kryteriach filtrowania (Phred ≥20) i krótkiej długości odczytów (101 nt), co nie wykorzystuje w pełni możliwości tej technologii (2 × 150 nt). Z kolei, do składania transkryptomów użyto programu SOAPdenovo2, przeznaczonego do składania genomów, zamiast wariantu SOAPdenovo-Trans czy popularnych narzędzi takich jak

ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań
tel. +48 61 829 58 42
andrzej.zielezinski@amu.edu.pl
www.combio.pl

Trinity lub rnaSPAdes. Te czynniki mogły wpłynąć na uzyskanie stosunkowo krótkich kontigów (średnio 500 nt), które nie odzwierciedlają pełnych transkryptów. Ograniczenie to nie powinno wpłynąć na wnioski dotyczące zmian jakościowych całego transkryptomu, ale może mieć znaczenie przy analizach wykorzystujących ontologię genów. Powstaje zatem pytanie, czym kierowała się Autorka przy wyborze tej strategii?

2. Czy Doktorantka próbowała ocenić wiarygodność i powtarzalność wyników opracowanej metody (Mash + PCoA + ANOVA/MANOVA), np. poprzez sprawdzenie, czy odległości Mash między replikami są istotnie mniejsze niż między różnymi próbami eksperymentalnymi. Taka analiza mogłaby również pomóc w doborze parametrów Mash, np. wielkości szkicu, oraz w ocenie powtarzalności wniosków statystycznych przy zmianie parametrów.
3. Czy Autorka rozważała sposoby poprawy jakości wyników opartych na surowych odczytach, np. poprzez filtrowanie *k*-merów w szkicach Mash według minimalnej liczby wystąpień? Mogłoby to przybliżyć jakość wyników do tej uzyskanej dla kontigów, bez konieczności ich składania.
4. Analiza Gene Ontology jest interesująca, lecz jej wartość interpretacyjna może być ograniczona przez niepełną lub zbyt ogólną adnotację genomów roślinnych. Warto podać, jaki procent kontigów udało się przypisać do terminów GO, oraz rozważyć uzupełnienie analizy o domeny białkowe, które często dają bardziej szczegółowe informacje i być może pozwoliłyby lepiej powiązać obserwowane zmiany z mechanizmami odpowiedzi roślin.
5. Twórczym pomysłem jest analiza kowariancji, pozwalająca skorygować porównania ilościowe o różnice w sekwencji. Problemem może być jednak stosowanie tej poprawki na poziomie transkryptomu – w takim modelu geny o niskiej ekspresji lub poniżej progu detekcji w jednym z wariantów mogą być błędnie eliminowane. Jeśli celem jest kontrola różnic sekwencyjnych, być może bardziej odpowiednie byłoby użycie genomu obejmującego pełny zestaw genów?

4. Komentarze techniczne

Poniższe komentarze mają charakter techniczny i redakcyjny. Są one mało istotne z punktu widzenia oceny rozprawy i Doktorantka może je pominąć podczas publicznej obrony.

- Na str. 50 oraz w Tabeli S2 podano procent zmienności wyjaśnianej przez osie składowych głównych PCoA. Wartości są niewielkie i nie przekraczają jedności. Czy są one w skali 0–1, czy rzeczywiście w procentach? Warto to wyraźnie oznaczyć, aby uniknąć niejasności.
- Czy analizowanie i dyskutowanie konkretnych wartości współrzędnych PCoA oraz prezentowanie ich w tabelach głównych jest rzeczywiście informatywne? Zakres osi PCoA zależy od danych, a dodanie lub usunięcie pojedynczego transkryptomu może całkowicie zmienić zakres wartości, co utrudnia biologiczną interpretację.
- We wzorze na współczynnik determinacji brakuje potęg drugich, zarówno przy obliczaniu odchyleń od średniej, jak i w symbolu współczynnika (R^2). Wygląda to raczej na przeoczenie edytorskie niż błąd obliczeń.
- Warto byłoby udostępnić implementację opracowanej metody, np. w serwisie GitHub. Dzięki temu jej zastosowanie w pracy biologa byłoby łatwiejsze i bardziej praktyczne.

5. Ocena końcowa

Wysoko oceniam recenzowaną rozprawę. Stanowi ona oryginalne rozwiązanie problemu naukowego wprowadzając nowatorskie podejście do analizy jakościowych zmian transkryptomu w kontekście interakcji genotyp–środowisko. Rozprawa wykazuje zarówno wiedzę teoretyczną Kandydatki w dyscyplinie naukowej rolnictwo i ogrodnictwo, jak i wysokie umiejętności w samodzielnym prowadzeniu badań naukowych. Rozprawa spełnia zatem warunki określone w art. 187 ustawy "Prawo o szkolnictwie wyższym i nauce" (Dz.U. z 2023 poz. 742 ze zm.) i dlatego wnioskuję do Rady Naukowej Instytutu Genetyki Roślin Polskiej Akademii Nauk o dopuszczenie mgr Marii Nuc do dalszych etapów przewodu doktorskiego.

PODPIS ZAUFANY
ANDRZEJ
ZIELEZIŃSKI
20.10.2025 10:06:01 GMT+0200
Dokument podpisany elektronicznie
podpisem zaufanym

Andrzej Zieleziński