

Warszawa, 6/12/2020

Prof. dr hab. Dariusz Plewczyński  
Laboratorium Genomiki Funkcjonalnej i Strukturalnej  
Centrum Nowych Technologii  
Uniwersytet Warszawski  
ul. Banacha 2c, 02-097 Warszawa, Polska

## RECENZJA

rozprawy doktorskiej magistra Marcina Tomasz Gradowskiego  
ANALIZA BIOINFORMATYCZNA PEŁNEGO ZBIORU LUDZKICH KINAZ ORAZ KINAZ  
ORGANIZMÓW ZWIĄZANYCH Z CZŁOWIEKIEM

wykonanej w Katedrze Biochemii i Mikrobiologii  
Instytutu Biologii  
Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie

pod kierunkiem promotora  
dr hab. Krzysztofa Piotra Pawłowskiego

Przedstawiona mi do recenzji praca jest owocem udanej analizy komputerowej rozległego zbioru kinaz w organizmie ludzkim i organizmach związanych z człowiekiem. W pracy udało się twórczo połączyć wysoko-wydajne narzędzia bioinformatyczne z istotnym biologicznie problemem badawczym skupiającym się na poznaniu funkcji biologicznej pełnego repertuaru kinaz. Już we wstępie chciałbym podkreślić nie tak często spotykaną staranność w przygotowaniu rozprawy, jej szaty graficznej, szczegółowości opisu zagadnień, narzędzi i wyników, świetnych i przekonujących ilustracji, wizualizacji struktur trójwymiarowych biopolimerów, tabel i uliniowień sekwencji analizowanych białek.

Przedmiotem mojej oceny, w myśl wymagań Ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki z dnia 14 marca 2003 r. (Dz.U. 2017 poz. 1789, z późn. zm.) oraz Rozporządzenia Ministra Nauki i Szkolnictwa Wyższego z dnia 19 stycznia 2018 r. w sprawie szczegółowego trybu i warunków przeprowadzania czynności w przewodzie doktorskim, w postępowaniu habilitacyjnym oraz w postępowaniu o nadanie tytułu profesora (Dz.U. 2018 poz. 261), jest oryginalność rozwiązane problemu naukowego, ogólna wiedza teoretyczna Kandydata w dziedzinie biologii, a także umiejętność samodzielnego prowadzenia pracy naukowej.

Rozprawa doktorska Pana mgr Marcina Gradowskiego została przygotowana w Laboratorium dr hab. Krzysztofa Pawłowskiego w Katedrze Biochemii i Mikrobiologii w Instytucie Biologii Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie. Autor skupia się na systematycznym przeglądzie kinaz ludzkich a następnie zestawów kinaz różnych organizmów zamieszkujących organizm ludzki, ze szczególnym uwzględnieniem patogenu Legionella. Następnie autor proponuje nową metodologię bioinformatyczną, która pozwala na kompleksowe wyszukiwanie nowych rodzin kinaz i innych enzymów. Następnie Autor przedstawił szczegółowo 6 nowych rodzin kinaz, które nie były wcześniej opisane. Na przykładzie kluczowej dla cyklu życiowego Legionelli wewnątrz makrofaga kinazy SidJ Autor sugeruje ścieżkę do odkrycia nowych sposobów walki z patogenami ludzkimi.

Autor w jednostronicowym streszczeniu pracy przedstawia główne tezy rozprawy. Podkreśla znaczenie opracowanego pipeline bioinformatycznego. Opisane są również wyniki zastosowania bioinformatycznego narzędzia do zbioru kinaz ludzkich oraz patogenów, tj. bakterii oportunistycznej Legionelli. Następnie Autor przedstawia listę skrótów użytych w pracy oraz przedstawia bardzo obszerny spis ilustracji.

W rozdziale pierwszym autor przedstawia w skrócie podstawowe tło prowadzonych badań, zaś w rozdziale drugim skupia się na rozległym przeglądzie literatury, wprowadzeniu podstawowych pojęć wykorzystywanych w rozprawie. Zaczynając o opisania białek efektorowych, różnych typów kinaz przechodzi do ich szczegółowej klasyfikacji, a następnie odnosząc się do wiedzy na temat Legionelli jako patogenu organizmu ludzkiego.

W rozdziale trzecim autor przedstawia główne cele i założenia rozprawy. Odnosi się do wielkości aktualnie poznanego Kinomu człowieka (529 białek i pseudokinaz), następnie wspominając o nowych rodzinach kinaz bakteryjnych. Wspomniany jest kluczowe źródło danych sekwencyjnych, tj. sekwencjonowanie następnej generacji mikrobiomu ludzkiego (6,217k sekwencji z prawie 3k próbek Human Microbiome Project).

W rozdziale czwartym opisuje główne materiały, narzędzia, algorytmy i metody wykorzystywane w trakcie badań. Opis narzędzi bioinformatycznych jest bardzo przejrzysty, zaczyna się od języka python jako wyboru środowiska programistycznego. Następnie Autor opisuje biblioteki użyte przez siebie, takie jak biopython, pandas, csv i glob. W kolejnym podrozdziale Autor skupia się na opisanie źródła z którego korzystał przygotowując dane wejściowe do dalszych analiz. Są to suplement artykułu Bursteina oraz baza Protein NCBI. Następnie Autor przechodzi do opisanie metodyki i narzędzi bioinformatycznych. Jest to przede wszystkim algorytm porównywania sekwencji białek (dopasowanie, lub uliniowanie) przy użyciu macierzy substytucji aminokwasów. Autor omawia oba podstawowe typy używane w bioinformatyce: BLOSUM oraz PAM. Następnie krótko omówione są algorytmy lokalnego i globalnego uliniowania za pomocą programowania dynamicznego.

Następnie przedstawiony jest algorytm BLAST, filtrowanie krótkich sekwencji, fragmentaryzacja pozostałych celem identyfikacji domen, oraz metoda klastrowania sekwencji białkowych – CD-HIT. W pracy oparto się na algorytmie RPS-BLAST i CD-search do wyszukiwania domen w sekwencjach, oraz FFAS/Phyre2 do wykrywania odległych homologów. Autor szczegółowo opisuje metodę FFAS, pakiet HH-suite, oraz serwer Phyre2 pozwalający na szybkie wyszukanie dalekiej homologii białek z dodatkowym krokiem modelowania struktury trójwymiarowej białek. Następnie przedstawiona jest wizualizacja klastrów sekwencji, dopasowanie wielu sekwencji za pomocą algorytmu MAFFT. Struktura drugorzędowa została przewidziana za pomocą PSIPRED. Informację genomiczną uzyskano z bazy Integrated Microbial Genomes, przeprowadzono analizę współwystępowania genów. Następnie opisano metody Filogenetyczne, oraz ciekawy algorytm DeepCoil wyszukujący za pomocą uczenia głębokiego sploty helis w sekwencji białek. Do treningu użyto biblioteki keras

i sieci zbudowanej z dwóch warstw konwolucyjnych. Do wizualizacji użyto programu chimera pochodzącego z UCSF.

Na stronie 73 znajduje się szczegółowy protokół analizy danych obejmujący wszystkie kroki pracy. Podzielono go na etapy wstępne, etapy przewidywania dalekiej homologii oraz dodatkowe analizy i klasyfikacji.

W rozdziale piątym przedstawiono wyniki i dyskusję. Opisana została pseudokinaza SidJ, następnie prawdopodobna kinaza efektorowa LLO\_2159, białkowe Lani\_1194, Lcin\_0519, efektorową Lsha\_0263 z insercją w strukturze PKL, oraz złożona kinaza białkowa typu HipA zbudowana z produktów trzech genów (lpg2378, lpg2379, lpg2380).

Autor swoją pracę doktorską zbudował wokół 41 genomów Legionella analizowanych złożoną metodologią bioinformatyczną (struktura drugorzędowa, u liniowanie sekwencji białek, daleka homologia, przewidywanie i analiza struktury trójwymiarowej). Ciekawe jest, że mimo braku istotności statystycznej miar podobieństwa, mimo wszystko dokładniejsza analiza ortogonalna (fizyko-chemiczna) pozwoliła pozytywnie zweryfikować takie relacje odległej homologii. Za pomocą CLANS zbudowano wizualizację nowych rodzin kinaz w szerszym kontekście całej rodziny PKL, oraz kinaz białkowych Legionella, człowieka i ameb. Przeanalizowano horyzontalny transfer genów między tymi gatunkami. Na kolejnych stronach została pokazana dokładna analiza rodziny pseudokinaz SidJ wraz z ich drzewem filogenetycznym, a następnie kolejnych nowych zaproponowanych rodzin. Zostały one szczegółowo opisane, czytelnie i z dużą erudycją biomolekularną i ewolucyjną.

Na koniec zamieszczono wnioski z przeprowadzonych analiz oraz przedstawiono listę publikacji w których wykorzystano wyniki rozprawy. Autor wyraził nadzieję opracowania w przyszłości nowego typu terapii przez opracowanie inhibitorów kinaz / pseudo-kinaz, które spowolnią lub nawet zatrzymają infekcję. Dodatkowo Autor zamieścił suplement i bibliografię.

Podsumowując – autor wykazał się znajomością bioinformatyki, zarówno od strony istotnych problemów biologicznych, ale również wielu najczęściej używanych narzędzi. W przedstawionej pracy przekonująco zaprezentował swoje umiejętności i zdolności sformułowania pytania biologicznego, jego algorytmizacji a następnie implementacji metody komputerowej, świetną wiedzę o narzędziach bioinformatycznych oraz przekonującą analizą danych. Wyniki opracowane w trakcie doktoratu zostały opublikowane w trzech pracach w wysoko punktowanych czasopismach o zasięgu międzynarodowym, w tym jednym jako pierwszy autor.

Poniżej postaram się krótko wypunktować kluczowe wg. recenzenta osiągnięcia badawcze doktoranta, w tym zaproponowane analizy i opracowane algorytmy, oraz analizy danych doświadczalnych bazujące na porównywaniu sekwencji i struktur trójwymiarowych białek.

- zebranie rozległego zestawu kinaz ludzkich, regulujących wiele istotnych procesów komórkowych;
- zebranie komplementarnego zestawu kinaz patogenów, tj. bakterii kolonizujących populację;
- szczegółowe przeanalizowanie mechanizmu patogenności organizmu *Legionella* powodującego zapalenie płuc u ludzi;
- wskazanie zestawu 114 kinaz *Legionelli* stanowiącej pełny poznany zestaw białek z tej rodziny zidentyfikowanych w wielu badaniach;
- zaproponowanie zestawu 15 nowych rodzin kinaz, które wcześniej nie były przypisane do tej klasy funkcjonalnej;
- przeanalizowanie najciekawszych przypadków z ich interpretacją biologiczną;
- opracowanie nowej procedury bioinformatycznej do wyszukiwania nowych rodzin białek, w tym przypadku kinaz;
- potwierdzenie skuteczności opracowanej metody bioinformatycznej w wykrywaniu nowych rodzin białek oraz porównanie jej do istniejących metod;
- zasugerowanie metod szacujących istotność opracowanych wyników.

## Ocena końcowa

W podsumowaniu mojej oceny rozprawy doktorskiej Pana magistra Marcina Gradowskiego stwierdzam, że zaprezentowaną pracę oceniam bardzo wysoko. Biorąc pod uwagę przejrzystość opisu problemu, czytelność wyników, liczne walory naukowe przedstawionej rozprawy doktorskiej, a dodatkowo wyjątkowo udane połączenie starannie opisanych algorytmów bioinformatycznych, oraz istotnych biologicznie pytań badawczych, oceniam rozprawę doktorską mgr Marcina Gradowskiego jako istotny wkład Autora do biologii w zakresie bioinformatyki. Z całym przekonaniem oceniam, że rozprawa spełnia zwyczajowe i ustawowe wymogi, stawiane rozprawom doktorskim, stanowi oryginalne rozwiązanie problemu naukowego w dziedzinie biologii, unaoacza ogólną wiedzę biologiczną kandydata, oraz znajomość metod w bioinformatyce, oraz pokazuje umiejętność samodzielnego prowadzenia pracy naukowej. Wnoszę zatem do Rady Dyscypliny Nauk Biologicznych Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie o dopuszczenie Pana magistra Marcina Gradowskiego do dalszych etapów przewodu doktorskiego.

Dodatkowo biorąc pod uwagę wysoki poziom merytoryczny rozprawy, jej staranne bioinformatyczne przygotowanie, przekonujący i bogato zilustrowany sposób prezentacji tematyki badawczej oraz istotność biologiczną wyników przeprowadzonych analiz wnoszę o wyróżnienie rozprawy stosowną nagrodą.



Profesor dr hab. Dariusz Plewczyński  
Laboratorium Genomiki Funkcjonalnej i Strukturalnej  
Centrum Nowych Technologii  
Uniwersytet Warszawski  
ul. Banacha 2c, 02-097 Warszawa, Polska