

Laboratory of Bioinformatics and Computational Genomics LB!GO
Faculty of Mathematics and Information Science, Warsaw University of Technology
ul. Koszykowa 75, 00-662 Warsaw, Poland

Warszawa,
20.02.2022

Laboratory of Functional and Structural Genomics LFSG
Centre of New Technologies, University of Warsaw
Banacha 2c Street, 02-097 Warsaw, Poland

mobile: [+48504726203](tel:+48504726203), e-mail: Dariusz.Plewczynski@pw.edu.pl, www: <https://plewczynski-lab.org>

Warsaw, 20/02/2022

Prof. dr hab. Dariusz Plewczyński
Laboratory of Bioinformatics and Computational Genomics,
Faculty of Mathematics and Information Science,
Warsaw University of Technology
Laboratory of Functional and Structural Genomics
Centre of New Technologies
University of Warsaw

Recenzja Rozprawy Doktorskiej
mgr MAŁGORZATY WNĘTRZAK

BADANIE OPTYMALNOŚCI STANDARDOWEGO KODU GENETYCZNEGO

wykonanej w Zakładzie Bioinformatyki i Genomiki
na Wydziale Biotechnologii
Uniwersytetu Wrocławskiego

pod kierunkiem promotorów
profesora doktora habilitowanego Pawła Mackiewicza
oraz
doktora habilitowanego Pawła Błazeja

zgłoszonej do Rady Dyscypliny Naukowej
Nauki Biologiczne
Uniwersytetu Wrocławskiego

Przedstawiona mi do recenzji praca jest wynikiem udanej analizy teoretycznej w paradygmacie obliczeniowej genomiki ewolucyjnej. Autorce udało się twórczo połączyć innowacyjną metodologię teoretyczną z biologicznie istotnym problemem badawczym skupiającym się na głębszym zrozumieniu stabilności standardowego kodu genetycznego (SGC, standard genetic code) oraz jego porównania z kodami sztucznymi. Budowa i analiza modeli obliczeniowych w celu lepszego zrozumienia procesów ewolucji organizmów żywych stanowi przykład stosunkowo niedawno powstałej, wysoko interdyscyplinarnej dziedziny, jaką jest biologia systemowa. Symulacje komputerowe integrują dostępną wiedzę na temat kodowania białek w sekwencji DNA z różnymi rodzajami mutacji obserwowanych w systemach żywych. Takie podejście stanowi rozwiązanie formalizujące i rozszerzające tradycyjną biologię ewolucyjną i umożliwia świeże spojrzenie na podstawowe problemy i wyzwania teorii ewolucji.

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), jak również art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2021 r. poz. 478, 619, 1630), przedmiotem mojej oceny jest oryginalność rozwiązane go problemu naukowego, ogólna wiedza teoretyczna Kandydata w dziedzinie biologii, a także umiejętność samodzielnego prowadzenia pracy naukowej.

Przedmiot mojej oceny, czyli rozprawa doktorska, moim zdaniem jest w pełni zgodny z warunkami określonymi w powyższych Ustawach, a nawet przekracza zwyczajowe wymagania. Autorka prezentuje wyjątkową oryginalność sformułowania i rozwiązania problemu naukowego, szeroką wiedzę teoretyczną z zakresu biologii molekularnej i ewolucyjnej, a także umiejętność prowadzenia wytrwałej i twórczej pracy naukowej. Nie mam żadnych fundamentalnie negatywnych uwag do rozprawy doktorskiej, jakośc przedstawienia metodologii i wyników badań, dorobku publikacyjnego Kandydatki. Całość rozprawy w pełni zasługuje na zaakceptowanie przez Radę Dyscypliny, a nawet jestem przekonany zasługuje na wyróżnienie.

Rozprawa doktorska Pani mgr Małgorzaty Wnętrzak została przygotowana w Zakładzie Bioinformatyki i Genomiki na Wydziale Biotechnologii Uniwersytetu Wrocławskiego we Wrocławiu pod kierownictwem prof. dra hab. Pawła Mackiewicza, oraz dr hab. Pawła Błazeja.

Rozprawa doktorska przedstawia rozważania na temat optymalności standardowego kodu genetycznego (standard genetic code, SGC). Zainteresowanie naukowców tym zagadnieniem zostało zapoczątkowane przez obserwację, że zmiany utraty sensu lub przesunięcia ramki odczytu pozwalają na zachowanie właściwości aminokwasów, takich jak hydrofobowość lub polarność. Doktorantka posługując się algorytmami optymalizacyjnymi bada jak bardzo SGC jest optymalne i jakie zmiany mogą jego optymalność zwiększyć. Praca składa się ze 154 stron, a główna treść rozprawy to 110 stron. Pozostałą część rozprawy to opis osiągnięć doktorantki, dodatkowe tabele i dokumentacja programu napisanego podczas realizacji pracy doktorskiej. W swojej rozprawie doktorskiej doktorantka odwołuje się do blisko 100 pozycji literaturowych. Autorka skupia się na analizie standardowego i szeregu nowych (syntetycznych) kodów genetycznych, przy wykorzystaniu metod optymalizacyjnych i innych narzędzi oraz algorytmów z dziedziny bioinformatyki sekwencyjnej białek i DNA, ciekawie rozszerzonej o analizę cech fizykochemicznych aminokwasów a nawet strukturę (drugorzędową) białek.

Motywacją pracy jest brak pełnej wiedzy na temat odporności kodu genetycznego na skutki mutacji kodonów zmieniających sekwencje białka. Zdaniem recenzenta jest to ważny i często niedoceniany problem badawczy związany z mechanizmami ewolucyjnymi obserwowanymi na Ziemi. Mamy tutaj do czynienia z dwoma poziomami stabilności – jeden związany z sekwencją aminokwasową rozumianą literalnie, a drugi z cechami fizykochemicznymi a w konsekwencji strukturą drugorzędową i trzeciorzędową kodowanych biopolimerów. Autorka rozprawy analizuje oba poziomy stabilności: sekwencji i struktury, oba kodowane w podwójnym łańcuchu DNA stanowiącym główny magazyn informacji biologicznej w komórce – w celu osiągnięcia lepszego zrozumienia skutków mutacji sekwencji genetycznej, popartego przejrzystą prezentacją wyników i użytych algorytmów komputerowych.

Lista prac naukowych opublikowanych przez Doktorantkę w trakcie pracy nad doktoratem obejmuje siedem pozycji powiązanych z tematyką pracy doktorskiej (dwie bez impact factor), dodatkowo Doktorantka wykazała **10** dodatkowych **publikacji**, **pięć wystąpień** konferencyjnych, **siedem posterów** oraz **jeden** samodzielny **grant** badawczy PRELUDIUM.

Poniżej lista publikacji związanych z doktoratem:

- [P1-NOIF] **Małgorzata Wnętrzak**, Paweł Błażej, Paweł Mackiewicz (2020) *Properties of the Standard Genetic Code and Its Alternatives Measured by Codon Usage from Corresponding Genomes*. Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Vol 3: BIOINFORMATICS, pp. 44-51 ***NO Impact Factor***
- [P2] **Małgorzata Wnętrzak**, Paweł Błażej, Paweł Mackiewicz (2019) *Optimization of the standard genetic code in terms of two mutation types: point mutations and frameshifts*. ***Biosystems*** 181, pp. 44-50
- [P3] **Małgorzata Wnętrzak**, Paweł Błażej, Dorota Mackiewicz, Paweł Mackiewicz (2018) *The optimality of the standard genetic code assessed by an eight-objective evolutionary algorithm*. ***BMC Evolutionary Biology*** 18, pp. 192
- [P4] Paweł Błażej, **Małgorzata Wnętrzak**, Dorota Mackiewicz, Paweł Mackiewicz (2018) *Optimization of the standard genetic code according to three codon positions using an evolutionary algorithm*. ***PLoS One*** 13 (8), pp. e0201715
- [P5] Paweł Błażej, **Małgorzata Wnętrzak**, Dorota Mackiewicz, Przemysław Gałat, Paweł Mackiewicz (2019) *Many alternative and theoretical genetic codes are more robust to amino acid replacements than the standard genetic code*. ***Journal of Theoretical Biology*** 464, pp. 21-32
- [P6-NOIF] Paweł Błażej, **Małgorzata Wnętrzak**, Paweł Mackiewicz (2018) *The Importance of Changes Observed in the Alternative Genetic Codes*. Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) - Vol 3: BIOINFORMATICS, pp. 154-159 ***NO Impact Factor***
- [P7] Paweł Błażej, **Małgorzata Wnętrzak**, Paweł Mackiewicz (2016) *The role of crossover operator in evolutionary-based approach to the problem of genetic code optimization*. ***Biosystems*** 150, pp. 61-72

Doktorantka uzyskała grant badawczy pod tytułem „Wielokryterialna optymalizacja kodu genetycznego”, który został przyznany przez Narodowe Centrum Nauki w konkursie PRELUDIUM (nr 2017/27/N/NZ2/00403) w 2018 roku.

W swojej rozprawie doktorskiej Kandydatka skupiła się na analizie stabilności standardowych i sztucznych kodów genetycznych, zastosowaniu dostępnych oraz rozwijaniu nowych metod optymalizacji, analizy wrażliwości, oraz zaprezentowała wyniki, zarówno znane z literatury bądź uzyskane z samodzielnie stworzonych modeli oceny stabilności ewolucyjnej. Autorka przeanalizowała własności kodów genetycznych za pomocą własnej, nowej metody optymalizacyjnej. Rozwinęła algorytmy ewolucyjne maksymalizujące i

minimalizujące skutki fenotypowe mutacji punktowych. Doktorantce udało się zbadać przestrzeń teoretycznie możliwych kodów genetycznych i porównać z rzeczywistym, standardowym kodem genetycznym, co stanowi wyjście poza dotychczas obowiązujący paradygmat porównań przeprowadzanych w odniesieniu do kodów losowych.

Kod rzeczywisty unika zmian polarność aminokwasów w białkach, chociaż nie robi tego w sposób optymalny. Dodatkowo Autorka ustaliła, że standardowy kod genetyczny jest prawie optymalny w kontekście ograniczenia konsekwencji mutacji w trzeciej pozycji w kodonie, natomiast istotnie gorzej ogranicza zmiany powstałe w wyniku mutacji na pozycjach pierwszej i drugiej. Kod rzeczywisty ma cechy pozwalające na minimalizację skutków podstawień, ale nie jest podobny do struktur zoptymalizowanych teoretycznych kodów genetycznych. Co jest ciekawą obserwacją wymagającą dalszych prac.

Dodatkowo Autorka pokazała różnorodny kontekst oceny skutków podstawień: od poziomu zmiany sekwencji białkowych, przez zachowanie właściwości fizykochemicznych aminokwasów (np. ograniczanie zmian w polarność), aż do oceny zachowania tendencji do przyjmowania odpowiednich struktur drugorzędowych. Jednak na żadnym z tych poziomów standardowy kod genetyczny nie jest w pełni optymalny. Na przykład Doktorantka pokazała, że zmiana w kodzie standardowym przypisania nawet jednego kodonu poprawia odporność na skutki podstawień aminokwasów o 15%, a zmiana przypisania trzech kodonów poprawia tę odporność o blisko 50%.

W zakresie projektowania zostały zbadane podobne własności i kryteria oceny skutków podstawień dla teoretycznych kodów genetycznych. Ciekawym wynikiem Autorki jest konkluzja, że większość syntetycznych kodów genetycznych lepiej minimalizuje konsekwencje mutacji punktowych i błędów podczas translacji. Dlaczego miliardy lat ewolucji nie zidentyfikowały alternatywnych metod zapisu genetycznego, które lepiej ograniczają konsekwencje mutacji? Pozostaje zgodzić się z mgr. Małgorzatą Wnętrzak, że minimalizowanie konsekwencji mutacji, ale również błędów podczas translacji odgrywało tylko częściową rolę w procesie ewolucji standardowego kodu genetycznego. Wiele zmian w trakcie rozwoju kodu mogło zostać zaakceptowanych, aby zredukować skutki podstawień aminokwasów i pojawiania się kodonów STOP translacji na skutek mutacji.

Rozdział pierwszy zawiera wstęp wprowadzający podstawowe pojęcia biologiczne używane w rozprawie doktorskiej. Rozdział drugi omawia cele

pracy. Rozdział trzeci prezentuje etapy wykonanych badań. Rozdział czwarty skupia się na opisanu metodologii pracy. Rozdział piąty zawiera wyniki rozprawy. Są to wyniki optymalizacji ze względu na minimalizację zmian w polarności aminokwasów, następnie na inne własności aminokwasów, jak również porównanie z alternatywnymi kodami genetycznymi. Rozdział szósty zawiera dyskusję otrzymanych rezultatów. Rozdział siódmy prezentuje wnioski z prowadzonych badań, ósmy bibliografię przedmiotu. Dodatkowo zamieszczono też Osiągnięcia naukowe autora oraz sekcję zawierającą Dodatki, takie jak Tabele z najlepszymi i najgorszymi kodami syntetycznymi oraz dokumentację programu. Brakuje niestety odniesienia do publicznie dostępnego repozytorium (np. github), które zawierałoby kod źródłowy opracowanego oprogramowania, co jest jedyną moją uwagą dla Doktorantki.

Mocne strony

1. Ważną zaletą pracy jest całościowe podejście zastosowane przez doktorantkę. W badaniach sprawdzała optymalność SGC pod kątem aż ośmiu kryteriów. Dla porównania, prace innych badaczy rozważały do trzech kryteriów.
2. Stosując algorytmy genetyczne doktorantka bardzo pomysłowo zdefiniowała nowe operatory, szczególnie operatora crossover w klasach NUM i GEN. Na pochwałę zasługuje również fakt, że działanie operatorów zostało również wyjaśnione za pomocą schematów.
3. Jednym z najciekawszych wniosków rozprawy doktorskiej jest ten, że alternatywne kody genetyczne z reguły lepiej ogranicza konsekwencje mutacji punktowych i błędów niż SGC. Lekki niedosyt budzi brak przekonującej hipotezy wyjaśniającej to zjawisko.
4. Wyniki pracy doktorskiej doktorantka promowała na wielu konferencjach, w tym tak prestiżowych jak ISMB/ECCB. Zadania badawcze opisane w pracy doktorskiej były również wspierane przez grant Preludium.

Słabe strony

1. Praca doktorska dotyczy tematu optymalności kodu DNA, która jest teraz bardzo istotna w szybko rozwijającym się polu biologii syntetycznej (synthetic biology). Dużym brakiem pracy jest brak odniesienia się do prac takich jak [10.1126/science.1190719](https://doi.org/10.1126/science.1190719) i [10.1073/pnas.1818259116](https://doi.org/10.1073/pnas.1818259116), które mogłyby dać praktyczny kontekst i zastosowanie teoretycznym rozważaniom przedstawionym w pracy

doktorskiej. Czy optymalne kody genetyczne pozwalają projektować bardziej efektywne organizmy? Jeśli tak, to dlaczego? Jakie są pozytywne skutki projektowania organizmów z rekodowanymi genomami?

2. Rozdział 1.1 zawiera szczegółowy opis mechanizmu translacji. Mimo że informacje te nie należą do obszaru wiedzy specjalistycznej, to dziwi brak jakiegokolwiek cytacji.
3. Rycina 1 została w pełni skopiowana ze strony: <https://cdn.kastatic.org/ka-perseus-images/4c267dd9e3999bb3ea41115030843f0aab38d197.png> . W pracy doktorskiej brak informacji, że rycina została w pełni skopiowana, a nie wykonana przez autorkę na podstawie podanej referencji. Nie ma również informacji dotyczących czy licencja obrazka zezwala na jego przedrukowanie w ramach pracy doktorskiej.
4. Do rozwiązania problemu optymalności kodu genetycznego autorka stosuje wyłącznie algorytmy ewolucyjne (jedno- i wielokryterialne), a konkretnie algorytm SPEA2. Na jakiej podstawie doktorantka zdecydowała się wybrać akurat ten algorytm mogąc skorzystać z efektywniejszych SPEA2+ jak i NSGA-III?
5. Dlaczego autorka rozważa tylko 21 z 33 dostępnych alternatywnych kodów genetycznych? Praca nie opisuje jasno kryteriów wyboru poszczególnych kodów.
6. Rozprawa jest oparta na siedmiu pracach, których współautorką jest doktorantka. Tylko w dwóch z tych prac doktorantka jest pierwszym autorem. Dodatkowo, wśród publikacji wymienionych jako dorobek autorki (str 111), dwie z siedmiu publikacji zostały opublikowane w wydawnictwach bez IF. Niestety, żadna z publikacji nie ukazała się w periodyku z IF powyżej 5.
7. Praca zawiera kilkanaście wykresów trójwymiarowych, które są mało czytelne w wersji nieinteraktywnej (wersja interaktywna była częścią oryginalnych publikacji). Rozprawa doktorska powinna albo zawierać odnośniki do interaktywnych wykresów albo wykresy w postaci 2D, np. w postaci wykresu trójkątnego.
8. Praca doktorska jest całkowicie obliczeniowa i dlatego zaskakuje negatywnie brak repozytoriów z kodami niezbędnymi do odtworzenia obliczeń. W czasie obecnego kryzysu reprodukowalności zaleca się publikowanie kodu w otwartych repozytoriach (BitBucket, GitHub). Ważnym elementem jest nie tylko opublikowanie samego kodu, ale również źródłowych danych, zmiennych środowiskowych (takich jak ziarna losowości) i w końcu samych środowisk, które pozwalają na

odtworzenie obliczeń. Praca doktorska nie zawiera odnośników do żadnych repozytoriów. Co więcej, żadna z siedmiu prac powiązanych z tematyką rozprawy doktorskiej takich odnośników również nie zawiera. Jest to poważnym problemem, ponieważ wskutek takiego zaniedbania inni badacze nie są w stanie odtworzyć ciekawych wyników zaprezentowanych w rozprawie. Zaleca się nadrobienie tego zaniedbania poprzez umieszczenie kodów niezbędnych do odtworzenia poszczególnych wyników w repozytorium, wraz z dokumentacją niezbędną do odtworzenia wyników. Taka dokumentacja znajduje się na stronach od xviii do xl, jest jednak a) napisana tylko w języku polskim b) nieprzydatna bez dostępu do programu.

9. Praca zawiera wiele błędów edytorskich. Przykładowo, nazwisko jednego z promotorów doktorantki pojawia się zarówno jako Blazej i jako Błażej (strona 105). Szczególnie problematyczna jest bibliografia, która zawiera bardzo wiele nieścisłości, np. nazwy periodyków czasami są zapisane jako skróty a czasami w pełnej postaci. Np. Knight, R. D., L. F. Landweber, et al. (2001). "How mitochondria redefine the code." *Journal of Molecular Evolution* 53(4-5): 299-313 ale również Sengupta, S., X. Yang, et al. (2007). "The mechanisms of codon reassignments in mitochondrial genetic codes." *J Mol Evol* 64(6): 662-688.
10. W pracy brakuje spisu rycin i spisu tabel. Bardzo utrudnia to lekturę i obniża ogólną ocenę rozprawy doktorskiej.

Ocena końcowa

Rozprawa przedstawiona przez Doktorantkę jest napisana jest niezwykle interesująco. Kandydatka pokazała, że dobrze rozumie i umiejętnie używa różne metody bioinformatyczne i optymalizacyjne, potrafi formułować hipotezy badawcze i je następnie weryfikować. Wyniki badań Doktorantki były poddane szczegółowej ocenie przez recenzentów 7 publikacji z tematu doktoratu oraz aż 10 pozostałych publikacji współautorskich, którzy byli wybrani przez edytorów poszczególnych czasopism. Dorobek naukowy magister Małgorzata Wnętrzak został już oceniony przez wielu ekspertów. Rozprawa doktorska świadczy o dobrym zrozumieniu przez Kandydatkę stawianych zadań badawczych. Cele rozprawy udało się zrealizować. Doktorantka wykazała, że zna i umiejętnie używa odpowiednio dobrane metody obliczeniowe i optymalizacyjne, formułuje hipotezy badawcze i je weryfikuje, tak więc ogólną wiedzę teoretyczną Doktorantki w dziedzinie nauk biologicznych, biotechnologicznych, oraz bioinformatycznych należy ocenić pozytywnie.

Oceniam, że rozprawa ta spełnia z nadmiarem zwyczajowe i ustawowe wymogi, stawiane rozprawom doktorskim, stanowi oryginalne rozwiązanie problemu naukowego, unaocznia ogólną wiedzę teoretyczną kandydata w zakresie matematyki oraz pokazuje umiejętność samodzielnego prowadzenia pracy naukowej. Wnoszę zatem do Rady Dyscypliny Naukowej Nauk Biologicznych Uniwersytetu Wrocławskiego o dopuszczenie Pani magister Małgorzaty Wnętrzak do dalszych etapów przewodu doktorskiego.

Dodatkowo biorąc pod uwagę wysoki poziom merytoryczny rozprawy, jej staranne przygotowanie, przejrzysty sposób prezentacji materiału badawczego oraz unikalny na tym etapie kariery naukowej sukces publikacyjny **wnoszę o wyróżnienie rozprawy stosowną nagrodą.**



Dariusz Plewczynski, PhD, Professor of Exact and Natural Sciences
Principal Investigator and Head

Phone: +48 22 554 36 54 or +48 22 234 7219, mobile: +48 504 726 203

e-mail: d.plewczynski@cent.uw.edu.pl or Dariusz.Plewczynski@pw.edu.pl www: <https://plewczynski-lab.org>

Laboratory of Functional and Structural Genomics LFSG
Centre of New Technologies, University of Warsaw
Banacha 2c Street, 02-097 Warsaw, Poland

Laboratory of Bioinformatics and Computational Genomics LB!GO
Faculty of Mathematics and Information Science, Warsaw University of Technology
ul. Koszykowa 75, 00-662 Warsaw, Poland