



UNIwersytet  
Warszawski

CeNT CENTRUM  
NOWYCH  
TECHNOLOGII

Warszawa, 17.02.2025

Dr hab. Mateusz Baca  
Centrum Nowych Technologii  
Uniwersytet Warszawski  
Ul. S. Banacha 2c  
02-097 Warszawa

## Recenzja rozprawy doktorskiej Pani mgr Katarzyny Burdy pt.: „Uwarunkowania obciążenia genetycznego w dzikich populacjach gupików *Poecilia reticulata*”

### Formalna charakterystyka rozprawy doktorskiej

Rozprawa Pani Katarzyny Burdy została przygotowana w Pracowni Biologii Ewolucyjnej Wydziału Biologii Uniwersytetu Adama Mickiewicza w Poznaniu pod kierunkiem Pana prof. dr. hab. Jacka Radwana oraz Pana dr. hab. Mateusza Konczala. Rozprawa składa się z dwóch głównych rozdziałów, które mają formę wieloautorskich artykułów naukowych napisanych w języku angielskim. Pierwszy z nich został opublikowany w czasopiśmie „Molecular Ecology Reseources”, drugi jest niepublikowanym maszynopisem. Badania opisane w obydwu rozdziałach były finansowane z projektu NCN SONATA 2018/31/D/NZ8/00091 kierowanego przez Pana dr. hab. Mateusza Konczala. Pierwszy rozdział dotyczy oszacowania tempa mutacji zachodzących z pokolenia na pokolenie w genomach gupików (*Poecilia reticulata*). Drugi rozdział dotyczy obciążenia genetycznego w naturalnych populacjach gupików z Trynidadu i Tobago, a w szczególności zależności pomiędzy obciążeniem genetycznym a efektywną wielkością populacji oraz wpływu historii populacji na obciążenie genetyczne. W skład rozprawy wchodzi też streszczenia w języku polskim i angielskim, wstęp, w którym autorka w jasny i zwięzły sposób wprowadza czytelnika w tematykę badań, rozdział, w którym przedstawione są cele pracy, oraz informacja, w którym z głównych rozdziałów zostały one zrealizowane. Po dwóch głównych rozdziałach autorka zamieściła rozdział „Final conclusions”, w którym podsumowuje zwięzłe wnioski płynące z przedstawionych wyników badań. Na końcu rozprawy przedstawione są deklaracje autorstwa precyzujące wkład autorki w powstanie prezentowanych artykułów.

Z deklaracji tych wynika, że w przypadku pierwszego artykułu Pani Katarzyna Burda uczestniczyła w powstaniu koncepcji i planu pracy, wyizolowała DNA z badanych prób oraz wykonała wszystkie analizy bioinformatyczne poza maskowaniem regionów powtarzalnych, uczestniczyła w przygotowaniu pierwszej wersji maszynopisu oraz w dalszych rewizjach. Zwracam tu uwagę na pewną nieścisłość, w sekcji author contribution opublikowanego artykułu jako osoba odpowiedzialna za

koncepcję badań wskazany jest jedynie Pan dr hab. Mateusz Konczal. W przypadku drugiego artykułu Pani Burda uczestniczyła w powstaniu koncepcji pracy, pobieraniu prób do analiz oraz wykonała ekstrakcję DNA. Autorka rozprawy wykonała znakomitą większość analiz bioinformatycznych. Analizy statystyczne wykonała wspólnie z Panem prof. Jackiem Radwanem. Pierwszą wersję maszynopisu przygotowała wspólnie z Panem dr. hab. Mateuszem Konczalem, uczestniczyła też w rewizjach manuskryptu. W mojej ocenie, pomimo wskazanej nieściśłości, deklaracje te potwierdzają wiodący wkład Pani Katarzyny Burdy w powstanie prezentowanych artykułów.

### **Merytoryczna ocena rozprawy doktorskiej**

W pierwszym artykule wchodzącym w skład rozprawy doktorskiej zatytułowanym „**Validation of machine learning approach for direct mutation rate estimation**” Pani Katarzyna Burda oszacowała tempo mutacji z pokolenia na pokolenie w genomie gupika (*Poecilia reticulata*). W tym celu zsekwencjonowała do wysokiego pokrycia genomu 24 osobników, dwóch par rodzicielskich i po dziesięciuro potomstwa każdej z nich. Następnie zastosowała dwie metody przywoływania wariantów oraz szereg filtrów, aby uzyskać wstępny zestaw mutacji *de novo* obecnych jedynie u potomstwa. Po czym wykorzystała dwa alternatywne podejścia do dalszej weryfikacji kandydackich mutacji. Pierwsze polegało na zastosowaniu szeregu predefiniowanych filtrów mających na celu odrzucenie mutacji będących wynikiem błędów sekwencjonowania czy mutacji w komórkach somatycznych. Jako alternatywę autorka zastosowała filtrowanie w oparciu o uczenie maszynowe. W tym celu przygotowała zestaw treningowy zawierający zweryfikowane oraz błędnie zidentyfikowane mutacje, a następnie pozwoliła, żeby algorytm sam przeprowadził filtrowanie pozostałych kandydackich mutacji. Ostatecznie, tempo mutacji oszacowane z zastosowaniem tych dwóch sposobów filtrowania było zbliżone. W dyskusji autorka najpierw szeroko omawia i analizuje znaczenie poszczególnych filtrów i porównuje, które z nich miały największy udział w określaniu mutacji jako błędnych w obydwu zastosowanych podejściach filtrowania. Zastanawia mnie w tym kontekście filtr AD (allelic depth in parents), który wymaga, aby u rodziców nie było ani jednego odczytu wspierającego zweryfikowaną mutację obecną u potomstwa. Mając na uwadze, że pojedyncze zmiany nukleotydów mogą pojawiać się w wyniku błędów sekwencjonowania należy rozważyć, czy tak skonstruowany filtr nie jest zbyt konserwatywny?

Niedługo po publikacji omawianej tu pracy w czasopiśmie „Genome Research” ukazał się artykuł, w którym stosując podobną metodykę, uzyskano prawie dwukrotnie wyższe szacunki tempa mutacji na pokolenie u gupików (Genome Res 2023 Aug;33(8):1317–1324. doi: [10.1101/gr.277936.123](https://doi.org/10.1101/gr.277936.123)). Czy różnice te mogły wynikać z subtelnych różnic w zastosowanej metodyce filtrowania i walidacji mutacji *de novo*, czy też wskazują może na naturalną wariację w tempie mutacji pomiędzy populacjami/rodzinami?

W dalszej części dyskusji autorka omawia uzyskane tempo mutacji na tle szacunków uzyskanych dla innych organizmów. W szczególności szuka możliwego wyjaśnienia niskiego tempa mutacji u ryb kostnoszkieletowych (teleosts) w porównaniu do innych grup kręgowców.

W mojej opinii omawiany artykuł prezentuje wysoki poziom naukowy. Bezpośrednie szacowanie tempa mutacji na podstawie analizy genomów rodziców i potomstwa jest relatywnie rzadkie i takie szacunki dostępne są dla nielicznych organizmów. Wynika to z ciągle wysokich kosztów sekwencjonowania, ograniczonej dostępności wysokiej jakości genomów referencyjnych, a także z trudności w identyfikacji mutacji *de novo*. Zaprezentowany eksperyment został dobrze zaplanowany,

autorka zastosowała najnowsze standardy analityczne do filtrowania kandydackich mutacji i szacowania tempa mutacji. Wymagało to bez wątpienia dobrego przygotowania bioinformatycznego i merytorycznego.

Efekt, jakim jest bezpośrednio oszacowanie tempa mutacji w genomach gupików, jest ważnym wkładem w rozwój badań nad ich ewolucją, szczególnie, że jest to gatunek modelowy. Jest to parametr bardzo istotny, choćby w rekonstrukcji historii ewolucyjnej i demograficznej. Pozwala umieścić zaobserwowane procesy na bezwzględnej skali czasu i powiązać je z przeszłymi zmianami klimatu czy innymi czynnikami bio- czy abiotycznymi. W samym artykule rekonstrukcja zmian efektywnej wielkości populacji ( $N_e$ ) w czasie wskazana jest jako jedno z możliwych zastosowań oszacowanego tempa mutacji. Autorka rekonstruuje te zmiany dla czwórki osobników rodzicielskich metodą PSMC i odnotowuje ciągły spadek  $N_e$  na przestrzeni ostatnich kilkudziesięciu tysięcy lat. W dyskusji zabrakło mi jednak jakiegokolwiek odniesienia do tej rekonstrukcji. Czy zaobserwowany spadek współgra z dostępnymi informacjami o historii populacji Trynidadzkich gupików? Czy zrekonstruowany przez metodę PSMC spadek  $N_e$  może być efektem innych procesów zachodzących w populacji?

Biorąc pod uwagę, że podobne szacunki tempa mutacji są ciągle rzadkie, omawiana artykuł ma też szersze znaczenie w badaniach ewolucyjnych jako element metaanaliz identyfikujących wzorce i ich ewolucyjne przyczyny, jak choćby omawiana w pracy obserwacja niższego tempa mutacji wśród ryb kostnoszkieletowych w porównaniu do innych grup kręgowców.

W drugim artykule zatytułowanym „Genetic load is affected by demographic histories in Trinidadian guppies (*Poecilia reticulata*), but does not explain invasiveness after a recent artificial translocation” przedstawione są wyniki analiz obciążenia genetycznego (*ang. genetic load*) w naturalnych populacjach gupików. Do analiz zsekwencjonowano genomy 190 osobników z 14 populacji z ośmiu rzek na Trynidadzie i 2 na Tobago. Eksperyment został zaplanowany w sposób umożliwiający przetestowanie szeregu ogólnych hipotez dotyczących zależności pomiędzy historią populacji i wielkością obciążenia genetycznego, między innymi: Czy istnieje zależność pomiędzy efektywną wielkością populacji a obciążeniem genetycznym? Jaki wpływ na obciążenie genetyczne ma gwałtowne zmniejszenie liczebności populacji (*ang. bottleneck*) związane z translokacją oraz następująca po nim ekspansja populacji? Czy obciążenie genetyczne różni się w populacjach o różnym reżimie drapieżnictwa? Zastosowano dwa wykorzystywane wcześniej wskaźniki obciążenia genetycznego: stosunek liczby alleli pochodnych (*ang. derived allele*) w silnie ewolucyjnie konserwowanych pozycjach genomu do liczby tychże w pozycjach bardziej zmiennych oraz stosunek alleli pochodnych, które powodują mutacje typu utraty funkcji (*ang. loss of function*) do wszystkich alleli pochodnych w genomie. Strukturę populacji zrekonstruowano standardowymi narzędziami (PCA, Admixture) i uwzględniono ją we wnioskowaniu.

Wykazano związek pomiędzy wielkością populacji a obciążeniem genetycznym, gdzie małe izolowane populacje wykazywały większe obciążenie, jednak nie zaobserwowano ogólnej zależności pomiędzy zmiennością neutralną populacji (tu rozumianą jako efektywna wielkość populacji) a obciążeniem genetycznym.

Wbrew teoretycznym oczekiwaniom nie stwierdzono efektu oczyszczania (*ang. purging*) ze szkodliwych wariantów w populacji, która w wyniku translokacji doznała gwałtownego spadku liczebności. Autorzy artykułu wskazują raczej na osłabienie selekcji oczyszczającej w tej populacji. Sugerują też, że w populacji tej nastąpiła prawdopodobnie akumulacja zmian o umiarkowanym efekcie

(*ang. moderate effect*), ale też, że zaobserwowano allele pochodne w silnie konserwowanych pozycjach genomu, w których zmiany mają prawdopodobnie poważne konsekwencje. Zastanawiam się, czy istnieje możliwość, aby bardziej precyzyjnie scharakteryzować obciążenie w tej populacji? Oprogramowanie SNPeff dzieli warianty na kategorie w zależności od ich domniemanego wpływu. Czy w tym przypadku dałoby się takie rozróżnienie uzyskać?

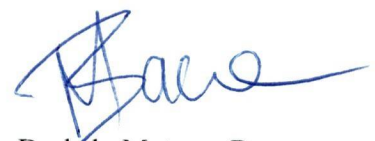
Kolejną ciekawą obserwacją poczynioną przez autorów był brak akumulacji obciążenia genetycznego wraz z ekspansją populacji. Wzrost obciążenia podczas ekspansji był postulowany w pracach teoretycznych i obserwowany w niektórych badaniach. Autorzy zwracają jednak uwagę, że w tym przypadku mógł on jednak być niwelowany przez mieszanie z lokalną populacją, czego ślady stwierdzono w badanej populacji.

Omawiany artykuł oceniam bardzo wysoko. Przeprowadzone analizy stoją na najwyższym poziomie. Tematyka obciążenia genetycznego, jego faktycznego wpływu na dostosowanie populacji i tego, jak kształtowane jest ono przez procesy demograficzne, to obecnie intensywnie badane i dyskutowane zagadnienie. Jest to możliwe w związku z coraz większą liczbą projektów sekwencjonowania genomów w skali całej populacji i ma wyraz w licznych publikacjach w prestiżowych czasopiśmie. W mojej opinii omawiany artykuł dostarcza nowych, ważkich wyników i będzie istotnym głosem w tej dyskusji.

Podsumowując, w mojej opinii rozprawa doktorska Pani Katarzyny Burdy prezentuje wysoką jakość merytoryczną i metodyczną. Nawet jeżeli przyjmiemy założenie, że ogólna koncepcja prac wchodzących w skład rozprawy pochodziła głównie od promotora pomocniczego, kierownika projektu, z którego badania sfinansowano, wkład doktorantki był znaczący. Uważam, że Pani Katarzyna Burda zademonstrowała bardzo dobre przygotowanie merytoryczne w zakresie genetyki populacji, co przełożyło się na wnikliwą dyskusję nad uzyskanymi wynikami. Realizacja projektu, ze względu na wielkość zestawów danych (190 genomów!) i wieloetapową ich analizę, wymagała solidnego warsztatu bioinformatycznego. Umiejętność manipulowania dużymi zbiorami danych genomowych jest obecnie bardzo dużym atutem.

#### **Wniosek końcowy**

W mojej opinii oceniana rozprawa Pani Katarzyny Burdy spełnia wymogi stawiane rozprawom doktorskim określone w art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2022 r. poz. 574 ze zm.). Rozprawa prezentuje ogólną wiedzę teoretyczną kandydatki, co przejawia się w zastosowanej metodyce prac i dyskusji nad wynikami. Skala zrealizowanych projektów sekwencjonowania, ich analizy i interpretacji wskazuje na umiejętność samodzielnego prowadzenia badań naukowych. Bez wątplenia uzyskane wyniki w postaci tempa mutacji czy też nowych informacji na temat powiązania historii populacji z obciążeniem genetycznym stanowią oryginalne rozwiązanie problemu naukowego. W związku z powyższym wnioskuję do Rady Naukowej Dyscypliny Nauki Biologiczne Uniwersytetu im. Adama Mickiewicza w Poznaniu o dopuszczenie Pani Katarzyny Burdy do dalszych etapów przewodu doktorskiego.



Dr. hab. Mateusz Baca