

# WIARYGODNOŚĆ MODELI HORYZONTALNEGO TRANSFERU GENÓW

## Streszczenie

Ewolucja rodzin genów jest skomplikowanym procesem i nie zawsze odpowiada ewolucji samych gatunków. Różnice te mogą być wyjaśnione za pomocą modelu drzew uzgadniających, w którym rozbieżności między drzewem ewolucji gatunków a drzewem ewolucji genów tłumaczy się występowaniem duplikacji i strat genów. W badaniach związanych z bakteriami poza duplikacjami i stratami należy również uwzględnić zachodzenie zjawiska horyzontalnego transferu genów (HGT). Rozszerzenie modelu drzew uzgadniających o HGT sprawia jednak, że algorytmy stają się złożone obliczeniowo co znacząco wpływa na czas ich działania. Do tej pory opracowano kilka algorytmów rekonstrukcji HGT, które dzięki uproszczonym modelom, pozwalają osiągnąć wielomianową złożoność czasową. Pomimo dobrej znajomości obliczeniowych własności tych algorytmów, niewiele wiadomo na temat ich stosowalności do danych rzeczywistych. W szczególności badanie wiarygodności osiągniętych rozwiązań jest kwestią otwartą. W niniejszym projekcie planujemy następujące cele naukowe:

1. Opracowanie metod do określania wiarygodności scenariuszy z HGT.
2. Analiza własności algorytmów transferowych ze szczególnym uwzględnieniem ich stosowalności do danych empirycznych i symulowanych.
3. Zastosowanie opracowanych metod do różnych problemów z HGT, w szczególności do problemu rekonstrukcji relacji gen-gatunek w kontekście eksperymentów metagenomicznych.
4. Opracowanie efektywnych algorytmów w formie ogólnodostępnych narzędzi.

Punktem startowym naszych badań jest opracowany przez nas algorytm umożliwiający określanie wiarygodności wpisania drzewa genów w drzewo gatunków, który zostanie zaprezentowany w czerwcu na konferencji AICoB 2016 i opublikowany w LNCS. Stworzone przez nas narzędzie pozwala na ocenę wiarygodności poszczególnych zdarzeń ewolucyjnych – duplikacji genów i specjacji. W pierwszym kroku zamierzamy rozszerzyć dotychczasowe rozwiązanie o model uwzględniający HGT. Następnie, z wykorzystaniem danych rzeczywistych oraz symulowanych, zamierzamy zbadać własności algorytmów rekonstrukcji HGT opierając się na opracowanych przez nas algorytmach i modelach. Na ich podstawie chcemy stworzyć hierarchię algorytmów rekonstrukcji HGT, a następnie zbadać wiarygodność uzyskiwanych za ich pomocą wyników. Dane do eksperymentów symulować będziemy za pomocą opracowanego przez nas programu, który również rozszerzymy o model z HGT. Metodę oceny wiarygodności chcemy wykorzystać również w kontekście rekonstrukcji relacji gen-gatunek w eksperymentach metagenomicznych. Zbadamy również jak na poprawę wiarygodności znajdujących relacji gen-gatunek wpływa integracja informacji dla wielu rodzin genów.

Mikroorganizmy stanowią niezwykle istotny element ziemskiego ekosystemu. Reakcje przeprowadzane przez bakterie mają wpływ na skład chemiczny i odczyn środowiska, a produkty tych reakcji są często wykorzystywane przez organizmy niezdolne do ich samodzielnej syntezy. W organizmach zwierząt i ludzi bakterie wspomagają procesy trawienia, rozkładu toksyn, wzmacniają odporność, przeprowadzają syntezę witamin i pełnią wiele innych niezbędnych funkcji. Bakterie mają również duże znaczenie przemysłowe. Wykorzystuje się je do produkcji antybiotyków, nawozów, wielu artykułów spożywczych wymagających fermentacji, a także do oczyszczania wód i ścieków z zanieczyszczeń przemysłowych. Badania nad HGT pozwalające lepiej zrozumieć ich ewolucję mają więc duże znaczenie dla rozwoju medycyny, przemysłu a także ochrony środowiska. Jedną z niezwykle istotnych kwestii jest nabywanie przez bakterie antybiotykooporności, która może być nabywana przez kolejne szczepy właśnie na drodze HGT. Nasze algorytmy mogą przyczynić się do ułatwienia analizy bioinformatycznej danych z eksperymentów metagenomicznych i zwiększenia wiarygodności otrzymywanych wyników.