

dr hab. inż. Małgorzata Bogdan
Instytut Matematyki i Informatyki
Politechnika Wroclawska

Wrocław 01.19.2014

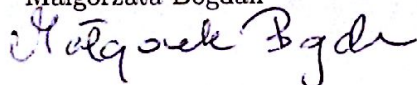
PROMESA RECENZJI ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: Generowanie długich otwartych ramek odczytu w genomach prokariotycznych.

Autor rozprawy: mgr Krystian Bączkowski

Oświadczam, że zapoznałam się z rozprawą doktorską mgr. Krystiana Bączkowskiego i w mojej ocenie spełnia ona wymogi ustawy o stopniach i tytułach naukowych. Zobowiązuję się przygotować pozytywną recenzję w ciągu najbliższych dwóch tygodni.

Małgorzata Bogdan



dr hab. inż. Małgorzata Bogdan
Instytut Matematyki i Informatyki
Politechnika Wroclawska

Wroclaw 03.02.2014

RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: Generowanie długich otwartych ramek odczytu w genomach prokariotycznych.

Autor rozprawy: mgr Krystian Bączkowski

Promotor rozprawy: dr hab. Paweł Mackiewicz

Cel, zakres i charakter rozprawy

Podstawowym celem rozprawy jest analiza możliwości generacji nowych genów w genomach prokariotycznych w wyniku alternatywnego odczytu genów już istniejących. W tym celu wykonano szereg bardzo rozległych i pracochłonnych analiz dotyczących własności ciągów DNA uzyskanych w wyniku alternatywnego odczytu już adnotowanych otwartych ramek odczytu (ORF-ów). Analizie poddano potencjalną długość tak powstałych ORF-ów, jak również ich skład aminokwasowy, różne własności fizykochemiczne i prawdopodobieństwo kodowania białka. Ponadto przeszukano "białkowe" bazy danych pod kątem istnienia sekwencji homologicznych do produktów ORF-ów odczytanych w alternatywnych fazach. Do tych przeszukiwań wykorzystano bardzo wiele istniejących baz danych i programów służących do ich analizy. Zebranie tych wyników i ich czytelna prezentacja była możliwa dzięki utworzeniu relacyjnej bazy, która jest jednym z podstawowych efektów tej pracy doktorskiej. Baza ta może w przyszłości służyć do kolejnych analiz związanych z generowaniem ORF-ów oraz identyfikacji potencjalnie błędnie adnotowanych sekwencji, które faktycznie kodują białko w jednej ze swoich alternatywnych faz odczytu. W pracy omówiono też przykłady genów, będących przypuszczalnie zmodyfikowanymi kopiami innych genów w ich alternatywnych fazach odczytu.

Struktura rozprawy

Rozprawa doktorska składa się z 6 rozdziałów, aneksu i bibliografii. Pierwszy rozdział zawiera wstęp omawiający organizację informacji genetycznej w genomie organizmów prokariotycznych i możliwe mechanizmy generowania genów. W drugim rozdziale wypunktowano cele pracy. Rozdział trzeci zawiera opis wykorzystanych baz danych i narzędzi bioinformatycznych. W rozdziale czwartym zebrano wyniki analiz, których całościowe podsumowanie można znaleźć w rozdziale piątym. Rozdział szósty zawiera wnioski. Aneks zawiera obszerną charakterystykę badanych genomów i spis rycin i tabel.

Dorobek doktoranta

Dorobek doktoranta składa się z trzech publikacji, z których jedna ukazała się w renomowanym wydawnictwie "Lecture Notes in Computer Science". Jest też on współautorem 10 prezentacji konferencyjnych.

Ocena pracy doktorskiej

Rozprawa doktorska mgr. Krystiana Bączkowskiego zawiera bioinformatyczną analizę możliwości generowania genów przez inne geny. Jej celem było potwierdzenie tezy, że nowe geny mogą powstawać w wyniku mutacji alternatywnych faz odczytu innych genów. Zgodnie z tezą mgr. Bączkowskiego, takie kopie genów mogą się "uwolnić" z oryginalnego układu zachodzącego i znaleźć na genomie w znacznej odległości od pierwotnego genu. Do uzasadnienia swoich tez autor wykonał szereg analiz, wykorzystujących bardzo obszerne genomiczne i białkowe zbiory danych i złożone narzędzia bioinformatyczne. W celu organizacji wyników tych analiz mgr Bączkowski stworzył relacyjną bazę danych, która może być w przyszłości rozbudowywana i wykorzystana do kolejnych analiz. Przygotowanie tej rozprawy doktorskiej wymagało od autora opanowania obszernej wiedzy z biologii, genetyki i informatyki.

Wyniki analiz mgr Bączkowskiego potwierdzają tezę, że spośród alternatywnych faz odczytu ORF-ów, największy potencjał do generowania nowych genów posiadają faza czwarta i piąta. W szczególności, zgodnie z oczekiwaniami, w fazie czwartej obserwuje się mniejszą częstość występowania kodonów stopu. W efekcie średnia i maksymalna długość "alternatywnych" ORF-ów w tej fazie zdecydowanie przekraczają średnią i maksymalną długość ORF-ów z innych faz, jak również średnią i maksymalną długość ORF-ów generowanych przez losowe ciągi DNA. Również procent ORF-ów o znaczącym

prawdopodobieństwie kodowania jest istotnie wyższy w tej fazie odczytu niż w innych fazach. Dodatkowo o uprzywilejowanej roli fazy czwartej w procesie tworzenia nowych genów świadczy stosunkowo duże podobieństwo własności ich produktów do rzeczywistych białek i stosunkowo duża ilość białek homologicznych i podobnych domen białkowych. Interesujący jest także fakt, że w znaczącym procencie ORF-ów prawdopodobieństwo kodowania białek w fazie czwartej przekracza p-stwo kodowania w fazie pierwszej, co może świadczyć o błędnej adnotacji pewnej grupy ORF-ów odczytanych w nieprawidłowej fazie.

Pracę mgr Bączkowskiego przeczytałam z dużym zainteresowaniem. W moim odczuciu tezy rozprawy można by istotnie wzmocnić formalnie ilustrując obserwowane efekty za pomocą testów statystycznych. W niektórych przypadkach przydałyby się też dodatkowe porównania z sekwencjami czysto losowymi. Krótką listę pytań/tematów do dyskusji zamieściłam poniżej. Uwagi te nie wpływają zasadniczo na ocenę poprawności zastosowanych przez autora metod i ogólną ocenę rozprawy.

Podsumowując, w mojej ocenie rozprawa doktorska mgr Krysztiana Bączkowskiego spełnia wymogi ustawy o stopniach i tytułach naukowych i wnioskuję o dopuszczenie jej do publicznej obrony.

Z wyrazami szacunku,
Małgorzata Bogdan

Pytania/Tematy do dyskusji

1. Zgodnie z moją wiedzą program Gene-Mark praktycznie nigdy nie zwraca zerowego p-stwa kodowania. Czy w rozprawie został zastosowany pewien próg odcięcia ?
2. Czy można określić o ile procent potencjalnie kodujących ORF-ów w fazie czwartej różni się (jest większy) od procentu potencjalnie kodujących ORF-ów wygenerowanych w sposób losowy ?
3. W moim odczuciu w pracy brakuje wyjaśnienia mechanizmów umożliwiających uwalnianie się ORF-ów z układu zachodzącego. Będę wdzięczna za omówienie tego zjawiska na obronie.