



Profesor dr hab. Stanisław Cebrat
Zakład Genomiki
Wydział Biotechnologii
Uniwersytet Wrocławski

Wrocław, dnia 19 grudnia 2011.



RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: Algorithms for modeling of the evolution of complex stochastic genetic systems. (*Algorytmy modelowania ewolucji stochastycznych systemów genetycznych o dużej złożoności*).

Autor rozprawy: mgr Tomasz Wojdyła

Promotor rozprawy: Prof. dr hab. inż. Marek Kimmel

Praca została napisana w języku angielskim. Ma ona charakter rozprawy interdyscyplinarnej, matematyk może nazwałby jej zakres „zastosowaniem matematyki”. Dla osób spoza dziedziny genetyki populacji mogłoby się wydawać, że niewiele w tym zakresie pozostało do zrobienia po przedstawieniu bardzo precyzyjnych opisów matematycznych mechanizmów ewolucyjnych przez Fishera, Wrighta, Hardy’ego, Haldane’a i Weinberga w pierwszej połowie dwudziestego wieku. To z połączenia ich teorii i praw Mendla z darwinowską teorią selekcji naturalnej powstał tak zwany neodarwinizm zwany również ewolucjonizmem syntetycznym. W tej postaci jest on ciągle propagowany w popularnej biologii, niestety tej programowo nauczanej – też, jako jedyny słuszny pogląd na mechanizmy ewolucji. Biologowie często nie zauważają, że znakomita większość tych praw i opisów dotyczy populacji mendlowskiej, bo taką można było opisywać ówczesnymi narzędziami i taka była wtedy nasza wiedza genetyczna. Obecnie powinniśmy sobie zdawać sprawę z tego, że średniopoloowe opisy ewolucji populacji nie są w stanie dać prawidłowego obrazu rzeczywistości, jeżeli choćby rozmieszczenie osobników w przestrzeni i genów na liniowych chromosomach z bardzo nielosowymi miejscami rekombinacji całkowicie zmieniają nasze poglądy na sposób dziedziczenia. Najlepszym przykładem stereotypowego myślenia jest przyjmowanie z wielkim zdziwieniem i niezrozumieniem wyniku badania dietności

Biurowo Dziekana

Wpłynęło dnia 02.01.2012.

Nr 282 / zat.

RAU

w populacji człowieka – dzietność maleje wraz z odległością genetyczną między małżonkami, a sukces ewolucyjny jest niemonotoniczną funkcją tej odległości. A przecież wydawałoby się, że heterozja powinna promować pary z większą odległością genetyczną.

Dlatego uważam, że praca, w której pojawiają się hasła koalescencja, dryf (dryft) genetyczny i demografia jest bardzo obiecująca już na wstępie. We wstępnych częściach rozprawy Autor podaje podstawowe definicje i informacje z genetyki populacji. Należy podkreślić, że robi to bardzo zwięźle, ale i bardzo klarownie i precyzyjnie.

W rozprawie można wydzielić trzy części. W pierwszej części Autor zajmuje się modelem Morana z dryftem, mutacjami i rekombinacjami. Stawia interesujące pytanie: czy jest możliwe wykazanie w niesymulacyjnym modelu, że populacja z rekombinującymi genomami może osiągnąć stan, w którym jest nieodróżnialna od populacji poddanej w czasie ewolucji jedynie dryftowi genetycznemu i mutacjom. Jego odpowiedź na to pytanie jest pozytywna. Zasadnicze elementy tej części rozprawy zostały opublikowane w 2010 roku w *J. Mathematical Biology* w pracy, w której doktorant był drugim autorem. Możliwość opisanego rozkładu stopnia pokrewieństwa osobników w populacji stwarza niebanalne możliwości w analizie wyników genotypowania populacji człowieka i dlatego jest istotnym przyczynkiem dla genetyki populacji.

W drugiej części doktorant zajmuje się w pewnym sensie podobnymi problemami - algorytmami umożliwiającymi wyliczenie dystrybucji odległości do najbliższego wspólnego przodka w próbkowaniu populacji różnych wielkości. Przedstawia wyniki zastosowania tej metody do analizy populacji człowieka. Ta część została opublikowana w czasopiśmie *Theoretical Population Biology* w pracy, której pierwszym autorem jest doktorant. Analizy statystyczne populacji mające na celu odtworzenie ich filogenezy, czy genealogii są obecnie bardzo często prowadzone i każde nowe podejście do rozwiązania problemów w tej dziedzinie biologii musi być docenione.

W trzeciej części Autor przedstawia narzędzia matematyczne opisujące sieci demograficzne. W modelach tych populacje mogą dzielić się, łączyć i możliwa jest również migracja osobników między populacjami, a genomy osobników podlegają mutacjom. Również w tej części Autor stosuje swoje narzędzia do analizy struktur

demograficznych rzeczywistych populacji człowieka oraz do analizy międzygatunkowej człowiek/szympan. Nie jest dla mnie jasne, czy we wszystkich analizach Autor posługiwał się tymi samymi markerami mikrosatelitarnymi z chromosomu Y. Z pewnością dla opisu populacji Bałtów i Słowian. W związku z tym mam pytanie do doktoranta: czy wyniki uzyskane z badań etnicznych przy użyciu markerów chromosomu Y powinny być podobne do wyników uzyskanych z zastosowaniem markerów mitochondrialnych, albo wyników uzyskanych z użyciem markerów autosomalnych?

Wartość modelowania matematycznego lub komputerowego zależy nie tylko od poprawności przeprowadzenia formalnych dowodów i odpowiednich algorytmów, ale również od poprawności przyjętych założeń. To właśnie założenia modeli matematycznych są najczęściej krytykowane przez biologów. Dlatego moje pytania do Doktoranta będą dotyczyć założeń.

1. We wszystkich częściach rozprawy Autor wprowadza mutacje. Rozumiem, że w analizowanych modelach mamy do czynienia z allelami wielokrotnymi, to znaczy, że w danym locus może pojawić się w puli genetycznej więcej niż dwa różne allele. W jaki sposób mutacje zamieniają te allele?

2. W opisie jednego z wykresów Autor podaje, że przyjął efektywną wielkość populacji człowieka $\sim 1 \times 10^4$, a efektywną wielkość populacji szympana trzy razy większą. Dlaczego?

3. Autor informuje, że przyjął częstość mutacji 10^{-4} na generację. Nie jest jasne, czy jest to częstość mutacji pojedynczej sekwencji mikrosatelitarnej czy też częstość liczona na genom i generację. Jeżeli te częstości dotyczą mutacji sekwencji mikrosatelitarnych chromosomu Y, to wydaje mi się, że powinny one być wyższe nawet o jeden rząd. Czy istotnym dla wyników analizy tymi metodami jest fakt, że poszczególne markery mutują z różnymi częstościami?

Powyższe pytania nie stanowią krytyki tej rozprawy ani nie kwestionują poprawności stosowanych metod, chciałbym jedynie wskazać Autorowi gdzie może spotkać się z ostrą krytyką biologów, jeżeli zależy mu na współpracy z nimi, a mam nadzieję, że tak i że zechce nadal pracować w tej dziedzinie.

Pewne nieścisłości, które pojawiają się w tekście wynikają z przyjmowania innych definicji terminów genetycznych, co niestety w genetyce i wśród genetyków jest dość

częste, lub patrzenia na problem genetyczny jedynie z perspektywy rozwiązywanego problemu. Stąd pojawiły się takie dziwne sformułowania:

1. „W celu modelowania rekombinacji, na ogół wyróżnialiśmy więcej niż jeden locus”. Jeżeli „na ogół”, to znaczy, że można było modelować rekombinację z jednym locus. Czy możliwe jest przeprowadzenie rekombinacji z genomem złożonym z jednego locus, jeżeli zachodzi ona tylko pomiędzy loci?

2. We wstępnej części pracy Autor opisuje pokrótce model Wrighta-Fishera, z którego to opisu można by wywnioskować, że w modelu tym osobnikowi haploidalnemu odpowiada jeden gen, albo dwa geny – osobnikowi diploidalnemu. Dalej Autor pisze: „W podstawowym modelu Wrighta-Fishera przyjmujemy, że osobniki w populacji nie są poddawane mutacjom, ani innym siłom ewolucyjnym. W szczególności **przyjmujemy, że nie ma rekombinacji między genami ...**” Może gdyby Autor nie powoływał się na pracę Fishera z 1930 roku, to byłoby mi łatwiej zrozumieć Jego intencje. Praca Fishera dotyczy bowiem tak zwanej populacji mendelowskiej, nieskończenie dużej, na którą nie działają siły selekcyjne ani mutacje, **a geny dziedziczą się niezależnie**, jak tego wymagało drugie prawo Mendla, a to znaczy, że prawdopodobieństwo rekombinacji między sąsiadującymi allelami wynosi 0,5. Tak naprawdę Autorowi chodziło chyba o allele, a nie o osobniki.

Z terminologii jeszcze: „rare matrix” czy „sparse matrix”?

W moim przekonaniu Pan mgr Tomasz Wojdyła przedstawił do oceny bardzo dobrą pracę. Wykazał w niej, że precyzyjnie potrafi postawić problem do rozwiązania, jest to problem istotny z naukowego punktu widzenia, a jego rozwiązanie daje interesujący pogląd na opisywane zjawisko. Wykazał również, że jego spojrzenie na dziedzinę genetyki populacji nie jest wrywkowe, obejmuje wiele istotnych zagadnień tej dziedziny biologii. Wiele z przedstawionych narzędzi matematycznych i wyników doczeka się wkrótce weryfikacji eksperymentalnej, ponieważ właśnie czekamy na rezultaty kilku wielkich programów genotypowania w populacji człowieka, a nawet sekwencjonowania dużej liczby genomów ludzkich. Dodając do tej oceny fakt, że część wyników pracy doktoranta została już opublikowana w dobrych czasopismach wnioskuję wyróżnienie pracy.

Stwierdzam, że rozprawa spełnia warunki Ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym, określone w art. 13 ust. 1 i ust. 2.
Wniosuję o dopuszczenie doktoranta do dalszych etapów przewodu doktorskiego.

Stanisław Cebrat

A handwritten signature in dark ink, appearing to read 'Stanisław Cebrat', written in a cursive style.