

Imię, Nazwisko i numer indeksu:

Kolokwium ze Wstępu do Biologii Obliczeniowej

12. czerwca 2018. roku

Czas rozwiązywania: 90 minut

Mamy 5 pytań otwartych, proszę o odpowiedzi zwięzłe, ale wyczerpujące wszystkie aspekty wspomniane w pytaniu. Można pisać na kartce z pytaniami i na dodatkowych kartkach. Proszę nie zapomnieć o podpisaniu kartek z odpowiedziami. Odpowiedzi typu tak/nie zawsze trzeba uzasadnić.

Powodzenia!

Zadania do rozwiązania:

1. **Macierze substytucji aminokwasów.** Podaj zasady konstrukcji macierzy PAM i BLOSSUM. Czemu odpowiadają liczby w nazwach macierzy i dlaczego potrzebujemy takich macierzy z różnymi liczbami? Czemu odpowiadają wartości dodatnie i ujemne w macierzach substytucji? Dlaczego nazywamy te macierze addytywnymi?
2. **Uliniowienia.** Podaj sformułowanie problemu znajdowania optymalnego uliniowienia dwóch sekwencji i jaka jest relacja tego problemu z problemem odległości edycyjnej? Wypisz macierz programowania dynamicznego łącznie ze "strzałkami" pozwalającymi na rekonstrukcję uliniowienia, która powstałaby podczas uruchomienia algorytmu Smitha-Watermana dla napisów CCGTTGA i GTG (kara za przerwy -2, substytucje: -1, nagroda za dopasowanie: 1). Jaki jest wynik (uliniowienie)? Czy wynik byłby inny, gdyby kara za przerwy była mniejsza niż za substytucje?
3. **Ukryte modele Markowa.** Podaj elementy składowe ukrytego modelu Markowa. Podaj schemat działania algorytmu Baum-Welch'a. Czy wynik algorytmu B-W jest zawsze optymalny? Opisz relację modeli Markowa i multiuliniowień (na przykładzie modelu HMMer). Dlaczego jest to dobry model do reprezentacji profili rodzin białek?
4. **Algorytm BLAST** Jaki problem rozwiązuje algorytm BLAST, które jego parametry są kluczowe dla odróżnienia go od uliniowienia lokalnego? Jakie heurystyki wykorzystuje algorytm BLAST? Co opisuje wartość E-value i jak jest wyliczana? Czy wynik działania BLAST dla zadanej sekwencji zapytania na publicznie dostępnej bazie danych może się różnić zależnie od dnia, kiedy wykonamy zapytanie?
5. **Relacja homologii i uzgadnianie drzew** Czym różnią się paralogi od ortologów - jeśli chodzi o definicję i oczekiwane własności sekwencji? Czym jest rozwiązanie problemu uzgadniania drzew? Jaka jest relacja zdarzeń ewolucyjnych (duplikacji, specjacji, strat) z paralogami i ortologami? Jaka jest rola mapowania LCA w znalezieniu optymalnego (w sensie duplikacji i strat) scenariusza uzgadniającego?