

Bioinformatyka - Raport z zadania 1. - Algorytmy uliniowania sekwencji

Wojciech Celej

1. Implementacja algorytmów

Ponieważ algorytmy Needlemana-Wunscha (global alignment) i Watermana-Smitha (local alignment) różnią się w niewielkim stopniu, stąd obie klasy implementujące algorytmy dziedziczą z klasy

`AlignmentAlgorithm`. Pozwala to uniknąć redundancji kodu.

Wartości zwracane przez funkcje scorujące:

Funkcja scorująca	Match	Mismatch	Gap
1	1	-1	-2
2	2	-1	-1

2. Porównanie genów homologicznych

Porównano ~5000 pierwszych nukleotydów dla 2 genów homologicznych:

- MTHFR metylenetetrahydrofolate reductase [*Homo sapiens* (human)]
- Mthfr metylenetetrahydrofolate reductase [*Rattus norvegicus* (Norway rat)]

Funkcja scorująca	NeedWunsch	WatSMith
1	-565	487
2	4331	4590

3. Porównanie sekwencji białkowych insuliny

- człowieka
- chomika

Funkcja scorująca	NeedWunsch	WatSMith
1	74	74
2	166	166

4. Wnioski

- Złożoność algorytmu postaci $O(mn)$ powoduje, że dla długich sekwencji algorytm staje się nieefektywny obliczeniowo
- Zwiększenie wartości za dopasowanie znaku znacząco zwiększa wartość dopasowania dla długich sekwencji
- Dla podobnych insulin o identycznej długości wynik działania algorytmów NW i WS okazał się być jednakowy