

Wstęp do bioinformatyki  
Laboratorium 2  
Dopasowanie globalne par sekwencji  
Magdalena Trędak  
236712

1. Schemat blokowy algorytmu dopasowania globalnego

Schematy blokowe algorytmów tworzenia macierzy punktowej oraz optymalnej ścieżki dopasowania ze względu na duży rozmiar i umożliwienia poprawy ich czytelności zamieszczono w repozytorium jako pliki graficzne o nazwach: SchematBlokowyGenerowaniaMacierzyPunktów.jpg, SchematBlokowyGenerowaniaŚcieżkiDopasowania.jpg.

Do wygenerowania schematów użyto programu online znajdującego się na stronie:

<http://www.algorytm.org/narzedzia/edytor-schematow-blokowych.html> (data dostępu 04.04.19)

2. Analiza złożoności obliczeniowej czasowej i pamięciowej

- Oszacowanie złożoności czasowej dla poszczególnych funkcji f

scoringMatrix =  $m * n$

tracBackMatrix =  $m * n$

createInfo =  $m * n + m$

isSequenceCorrect =  $m + n$

$$\sum f = 3 * m * n + 2 * m + n \leq 3 * m * n$$

$O(mn)$  – złożoność czasowa co najwyżej rzędu  $mn$

- Oszacowanie złożoności pamięciowej dla poszczególnych funkcji f

scoringMatrix =  $2 * m * n + m + n$

tracBackMatrix =  $2 * m * n + m + n$

createInfo =  $m * n + m + n + m + n + m * n + 5 * (m + n) + m$

makeFastaContent =  $5 * (m + n)$

readFasta =  $m + n$

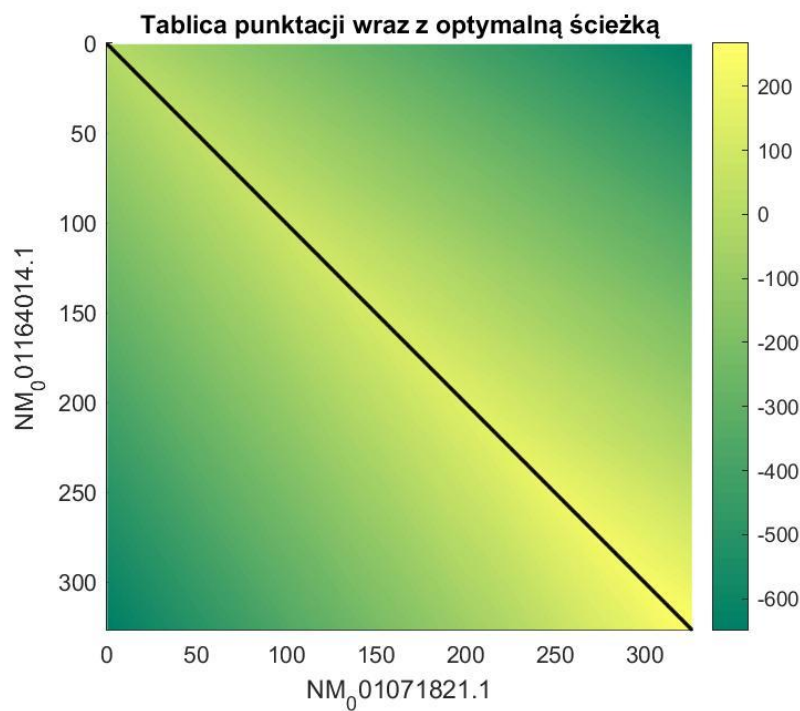
$$\sum f = 6 * m * n + 15 * (m + n) + m \leq 6 * m * n$$

$O(mn)$  – złożoność pamięciowa co najwyżej rzędu  $mn$

### 3. Porównanie przykładowych par sekwencji ewolucyjnie

- Powiązanych

Porównanie cytochromu c konia (*Equus caballus*) - NM\_001164014.1 i szympansa zwyczajnego (*Pan troglodytes*) - NM\_001071821.1 – porównanie nr 1



```
#Sequence1: NM_001071821.1
#Sequence2: NM_001164014.1
#Match: 1
#Mismatch: -1
#Gap: -2
#Mode: similarity
#Score: 267
#Length: 318
#Gaps: 0/318 (0 %)
#Identity: 289/318 (90.8805 %)
ATGGGTGATGTTGAGAAAGGCAAGAAGATTTTATTATGAAGTGTTCCCA(
|||||
ATGGGTGATGTTGAGAAAGGCAAGAAGATTTTGTTCAGAAGTGTTCCCA(
```

- niepowiązanych

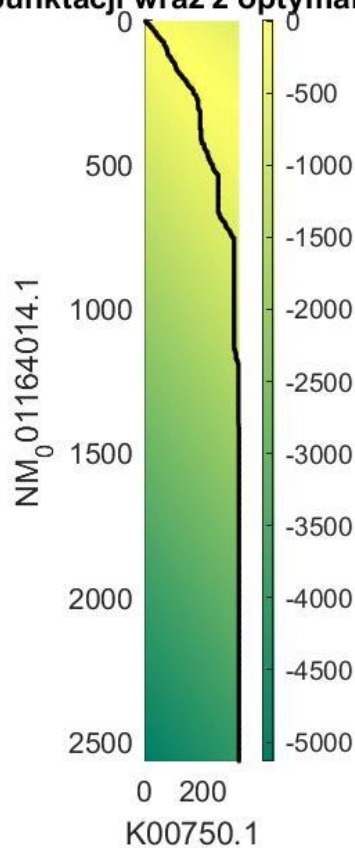
Porównanie cytochromu c szczura wędrownego (*Rattus norvegicus*) – K00750.1 i konia (*Equus caballus*) - NM\_001164014.1 – porównanie nr 2

```

porownanie1 — Notatnik
Plik Edycja Format Widok Pomoc
#Sequence1: K00750.1
#Sequence2: NM_001164014.1
#Match: 1
#Mismatch: -1
#Gap: -2
#Mode: similarity
#Score: -4159
#Length: 2530
#Gaps: 2246/2530 (88.7747 %)
#Identity: 84/2530 (3.3202 %)

```

Tablica punktacji wraz z optymalną ścieżką

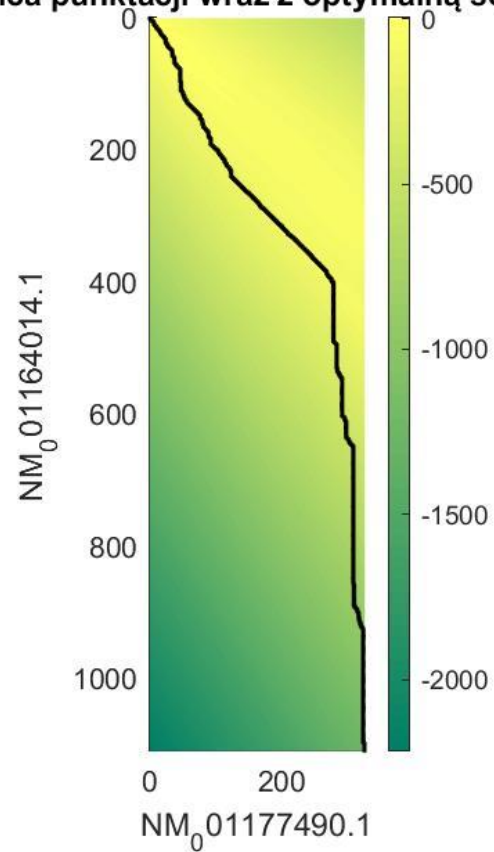


Porównanie cytochromu c pszczoły miodnej (*Apis mellifera*) - NM\_001177490.1 i konia (*Equus caballus*) - NM\_001164014.1 – porównanie nr 3

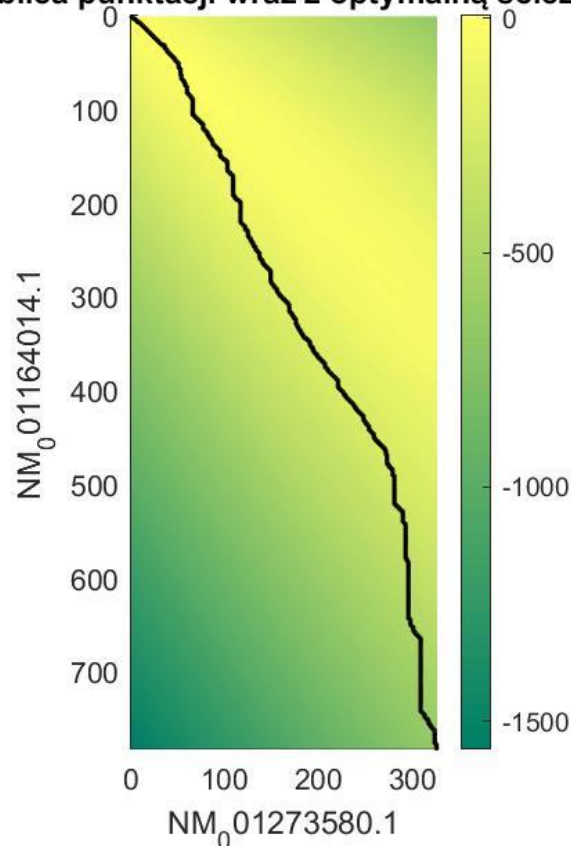
```

porownanie1 — Notatnik
Plik Edycja Format Widok Pomoc
#Sequence1: NM_001177490.1
#Sequence2: NM_001164014.1
#Match: 1
#Mismatch: -1
#Gap: -2
#Mode: similarity
#Score: -1267
#Length: 1097
#Gaps: 796/1097 (72.5615 %)
#Identity: 103/1097 (9.3892 %)

```

[illegible]

Tablica punktacji wraz z optymalną ścieżką



Wnioski:

Porównano ten sam gen dla organizmów powiązanych i niepowiązanych ewolucyjnie. Wszystkie wyniki otrzymano poprzez wykorzystanie algorytmu dopasowania globalnego podobieństwa o parametrach: match = 1, mismatch = -1, gap = -2.

Tabela 1. Porównanie otrzymanych wyników dopasowań globalnych dla różnych par organizmów

Powiązanie ewolucyjne organizmów	Nr porównania	Score [-]	Gap [%]	Identity [%]
Tak	1	267	0	90,88
Nie	2	-4159	88,77	3,32
	3	-1267	72,56	9,39
	4	-643	60,59	12,40

Na podstawie wyników zawartych w Tabeli [1] można zauważyć następujące zależności:

- Dla organizmów powiązanych ewolucyjnie score jest dodatni (największy wśród wyników). W sekwencjach nie ma przerw, co oznacza, że na drodze ewolucji występowały delekcje i insercje pojedynczych nukleotydów lub ich krótkich fragmentów. Podobieństwo sekwencji jest wysokie (ponad 90 %), co potwierdza powiązanie ewolucyjne badanych organizmów.
- Dla organizmów niepowiązanych ewolucyjnie score jest ujemny, wprost proporcjonalny do podobieństwa procentowego (Identity). Procent przerw (gap) jest związany z podobieństwem zależnością odwrotnie proporcjonalną. W analizowanych przypadkach

podobieństwo jest maksymalnie rzędu 10 %, z czego można wnioskować o braku wspólnego drogi ewolucji.