

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

1. Schemat blokowy algorytmu dopasowania lokalnego

Schematy blokowe algorytmów tworzenia macierzy punktowej oraz optymalnej ścieżki dopasowania ze względu na duże rozmiary 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 24.04.19)

2. Analiza złożoności obliczeniowej czasowej i pamięciowej

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
```

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

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

- Oszacowanie złożoności pamięciowej dla poszczególnych funkcji f

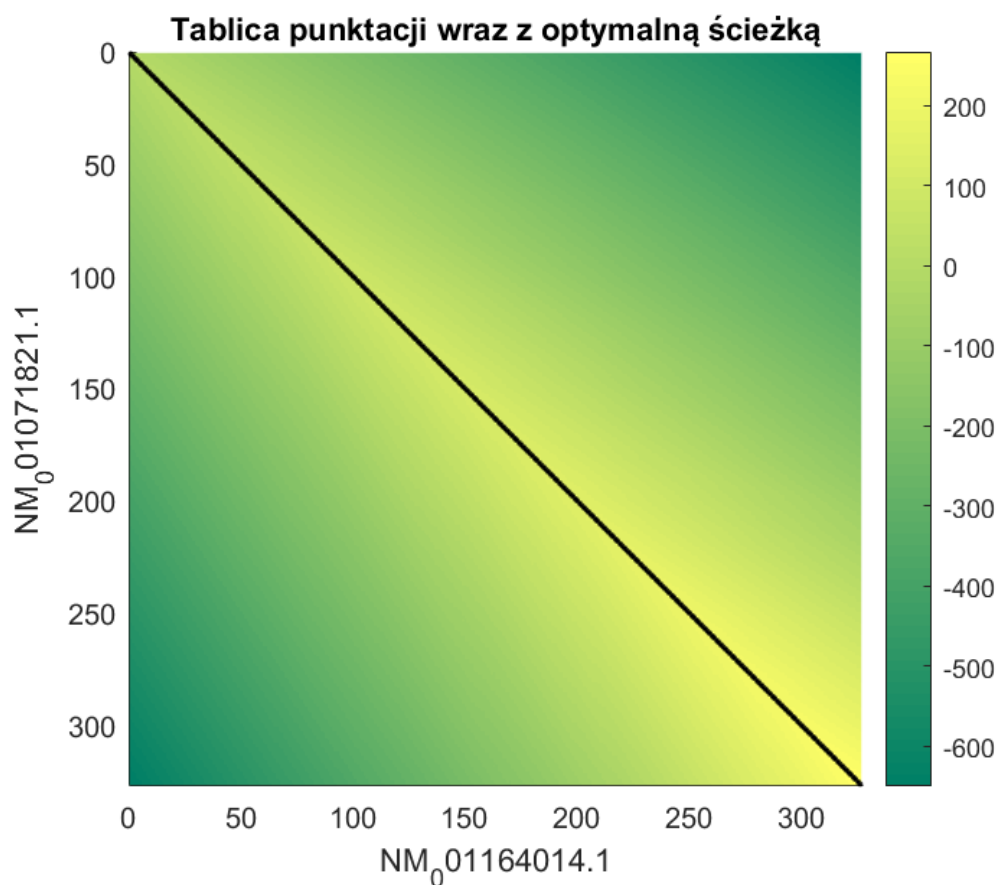
```
scoringMatrix = 2*m*n + m + n  
tracBackMatrix = 4 + (m*n)*6  
createInfo = m*n + m + n + m + n + m*n + 5*(m+n) + m  
makeFastaContent = 5 * (m+n)  
readFasta = m + n
```

$$\Sigma f = 19*m*n + 15*(m+n) + m + 4 \leq 6*m*n$$

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

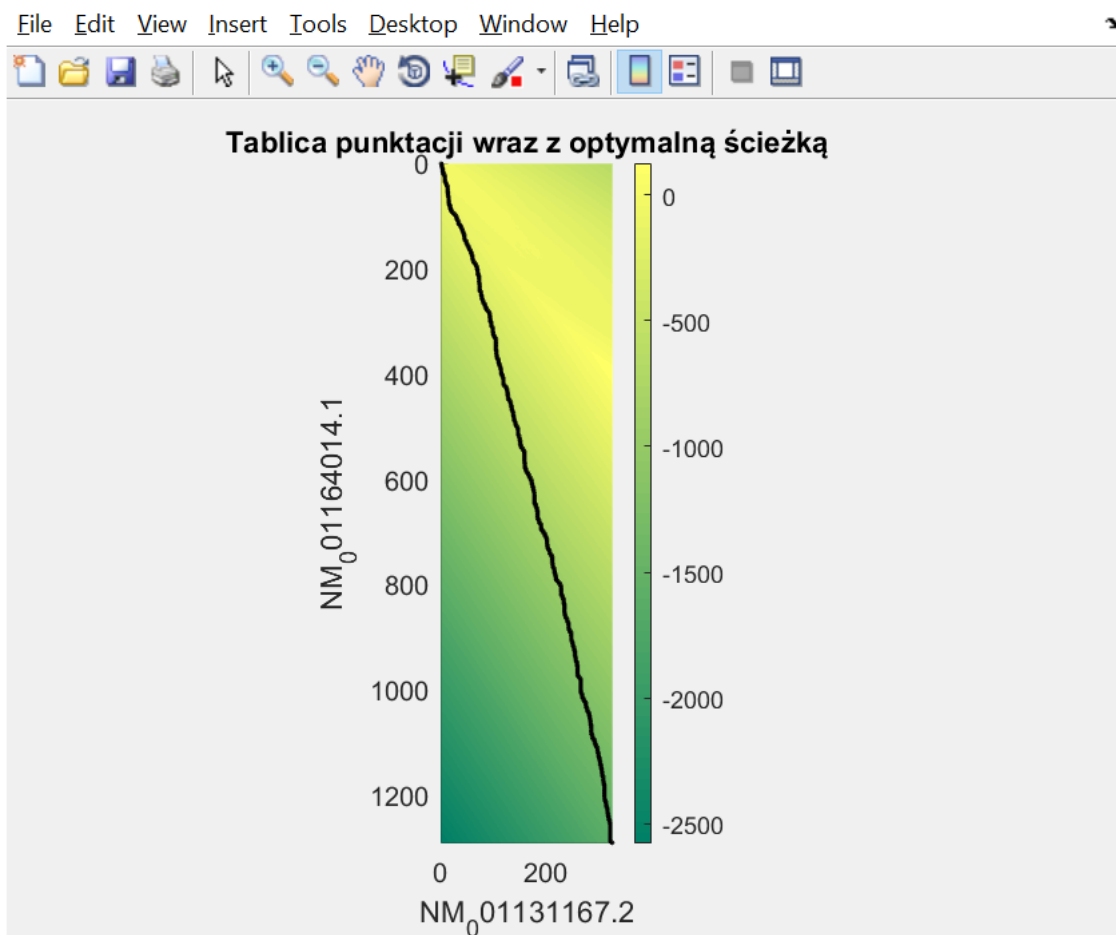
3. Porównanie przykładowych par sekwencji

Porównanie cytochromu c konia (*Equus caballus*) - NM_001164014.1 i szympansa zwyczajnego (*Pan troglodytes*) - NM_001071821.1 – porównanie nr 1

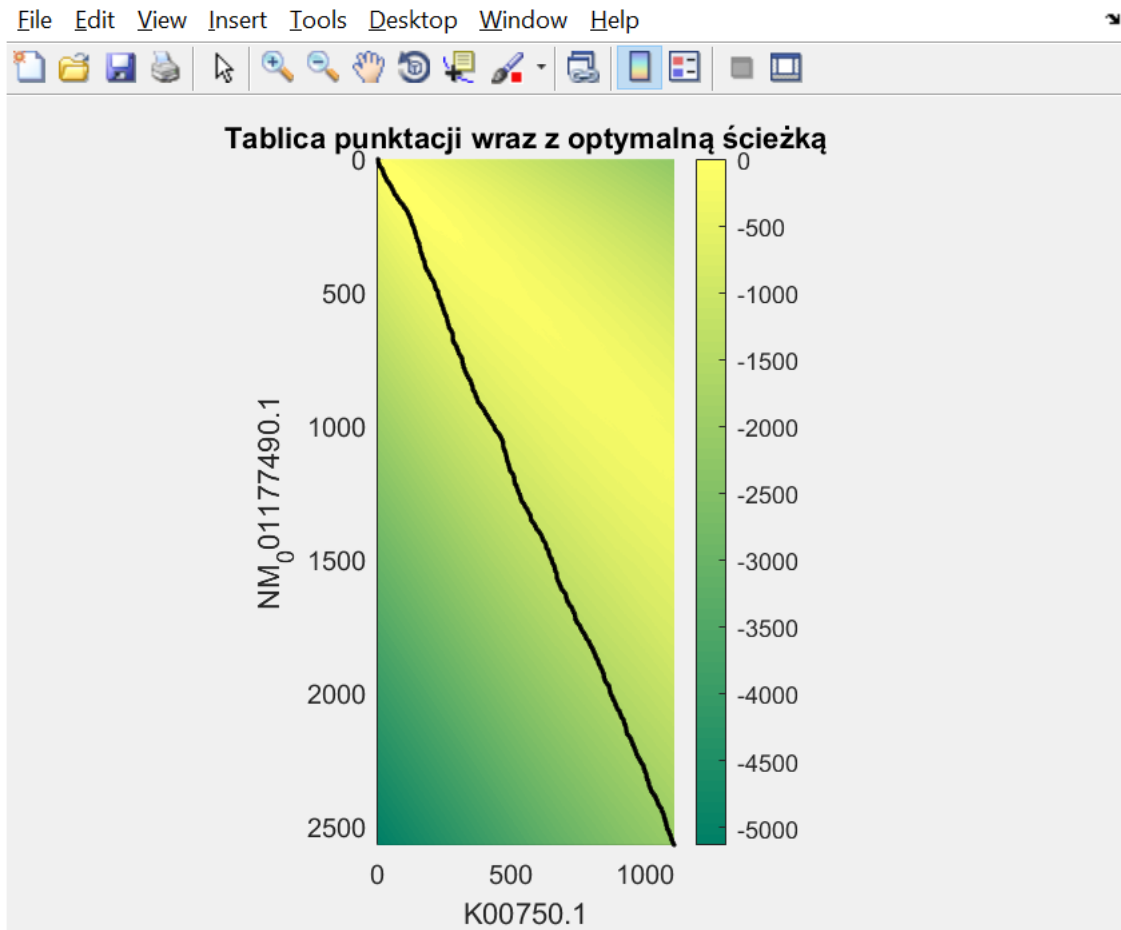


```
porownanie1 — Notatnik
Plik Edycja Format Widok Pomoc
#Sequence1: NM_001164014.1
#Sequence2: NM_001071821.1
#Match: 1
#Mismatch: -1
#Gap: -2
#Mode: similarity
#Score: 267
#Length: 318
#Gaps: 0/318 (0 %)
#Identity: 289/318 (90.8805 %)
ATGGGTGATGTTGAGAAAGGCAAGAAGATTTTGTTCAGAAGTGTGCCAGTGCCATACCGTGAAAAAGGGAGGCAAGCACAAGACTGGGCCAAACCTCCATGGTCTATTTGGGC
|||||
ATGGGTGATGTTGAGAAAGGCAAGAAGATTTTATTATGAAGTGTTCAGTGCCATACCGTTGAAAAAGGGAGGCAAGCACAAGACTGGGCCAAATCTCCATGGTCTCTTCGGGC
```

Porównanie cytochromu c orangutana (*Pongo abelii*) – NM_001131167 i konia (*Equus caballus*) - NM_001164014.1 – porównanie nr 2

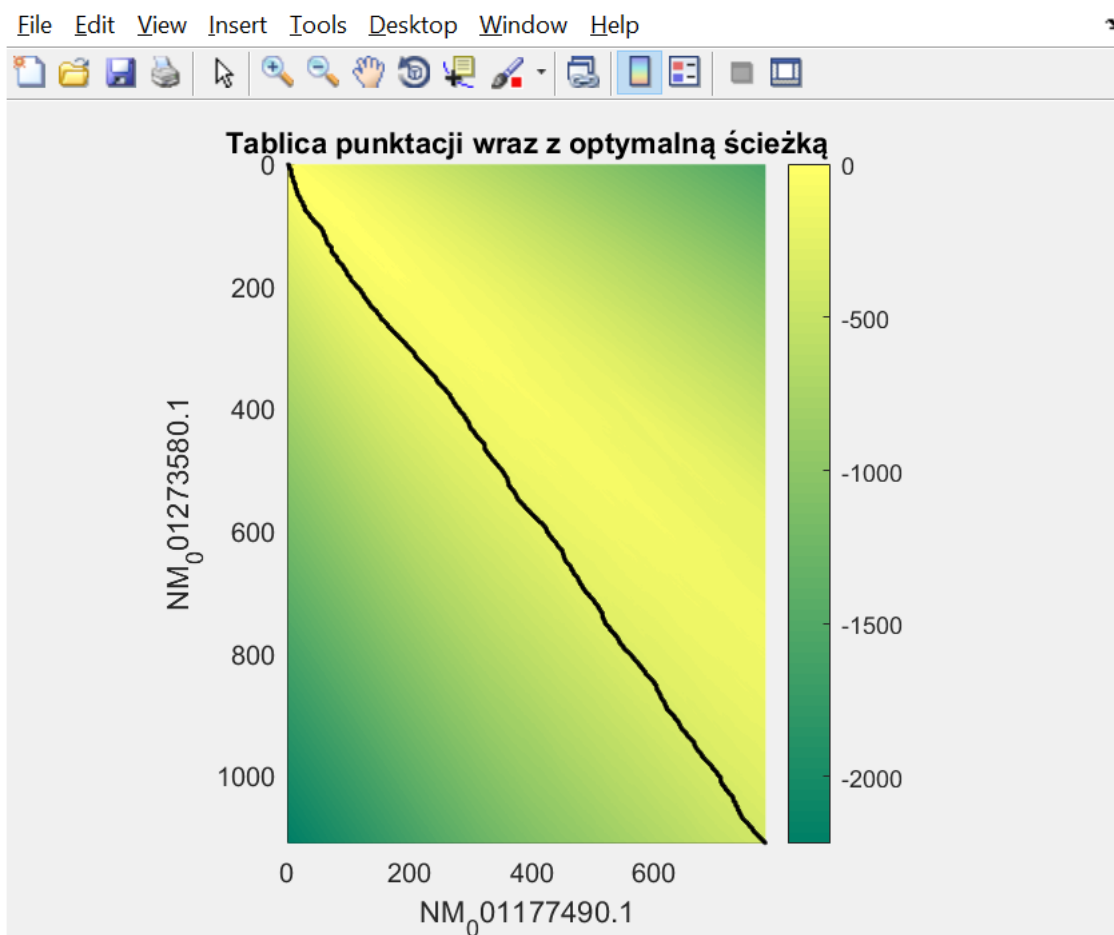
[illegible]

Porównanie cytochromu c szczura wędrownego (*Rattus norvegicus*) – K00750.1 i pszczoły miodnej (*Apis mellifera*) - NM_001177490.1 – porównanie nr 3



```
% GlobalAlignmentPlot', 'C:\Nauka\b\wstep do Bioinformatyki\Lab\Lista 1\Flikl',...  
Command Window  
>> Tester  
#Sequence1: K00750.1  
#Sequence2: NM_001177490.1  
#Match: 1  
#Mismatch: -1  
#Gap -2  
#Mode: similarity  
#Score: -2075  
#Gaps: 1458/2528 (57.6741 %)  
#Identity: 280/2528 (11.0759 %)  
GCTCCTCCCCAACACGCAGGCCGGAGGGTGACGTCAATCCACGTCCACGCCTTACGTCCAAGGGCGCTGTCGAAGTTGTCGGGCAAACGAGGCCCTCTAGAGGAAGGGCGCCCTCTCG-  
      | ||          ||| ||||   ||    |         ||     ||       |              |  
-----A-TCA----G-TGGTTAC-TCAT---CG--A----T-CG---AA-G-C-ATC--T-A---TC-CACGGTC-AGG---CTATA-AA---CG---TCT-
```

Porównanie cytochromu c muszki owocowej (*Drosophila melanogaster*) variant B - NM_001273580.1 i pszczoły miodnej (*Apis mellifera*) - NM_001177490.1 – porównanie nr 4

[illegible]

Wnioski:

Porównano ten sam gen dla organizmów powiązanych i niepowiązanych ewolucyjnie. Wszystkie wyniki otrzymano poprzez wykorzystanie algorytmu dopasowania lokalnego podobieństwa o parametrach: gap = -2 oraz parametrach zawartych w macierzy substytucji dołączonej jako plik sMatrix.txt

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	-1617	76,02	6,94
Nie	3	-2075	57,67	11,07
Tak	4	-402	30,77	18,17

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

Dla wszystkich organizmów score jest wprost proporcjonalny do podobieństwa procentowego (Identity).

- Dla organizmów silnie powiązanych ewolucyjnie score jest dodatni (największy wśród wyników). W sekwencjach nie ma przerw, co oznacza, że na drodze ewolucji występowały delecje i insercje pojedynczych nukleotydów lub ich krótkich fragmentów. Podobieństwo sekwencji jest wysokie (ponad 90 %), co potwierdza powiązanie ewolucyjne badanych organizmów.
- Dla organizmów niepowiązanych ewolucyjnie. 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.