David Bereszczyński 236807

**Wstęp do bioinformatyki**

**ZADANIE 2: Dopasowanie globalne par sekwencji**

**Odnośnik do repozytorium:**

<https://github.com/236807/WBI>

1. **Cel ćwiczenia**

Celem ćwiczenia było napisanie programu znajdującego optymalne dopasowanie pary sekwencji DNA, RNA lub białek, oraz wykonanie z jego pomocą przykładowej analizy.

1. **Złożoność czasowa**

Dla każdego z elementów wyznaczana jest wartość maksymalna z trzech elementów . W mojej implementacji wymaga to trzech porównań dla (i-1)(j-1) elementów.

Podana metoda jest rzędu Θ(ij).

W metodzie wyznaczającej optymalną drogę używam w najgorszym przypadku 2 porównań dla każdego elementu który jest na optymalnej ścieżce. Metoda jest rzędu Θ(i+j).

Więc złożoność czasowa algorytmu jest rzędu Θ(ij).

1. **Złożoność pamięciowa**

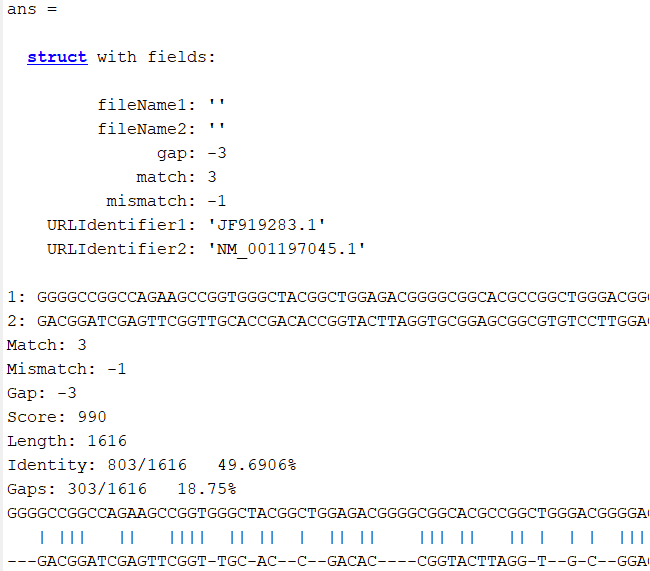
Macierz którą implementuje z 2 sekwencji które są rozmiaru i i j ma rozmiar ich iloczynu a więc złożoność pamięciowa jest rzędu Θ(ij).

1. **Porównanie par sekwencji ewolucyjnie powiązanych**

Do pokazania sekwencji powiązanych ewolucyjnie wykorzystano sekwencję DNA kodujące cytochrom c mRNA u Canis Lupus (wilka) i Canis Lupus familiaris (psa).

Wynik otrzymano komendą:

NeedWunInterface('URLIdentifier1','JF919283.1','URLIdentifier2','NM\_001197045.1','match',3,'gap',-3,'mismatch',-1)



Dopasowanie tych gatunków w podanym cytochromie c mRNA jest na poziomie 50%. Ilość przerw w sekwencji jest na niskim poziomie około 19% a więc potwierdza to powiązanie ewolucyjnie cytochormu c mRNA u wilka i psa.

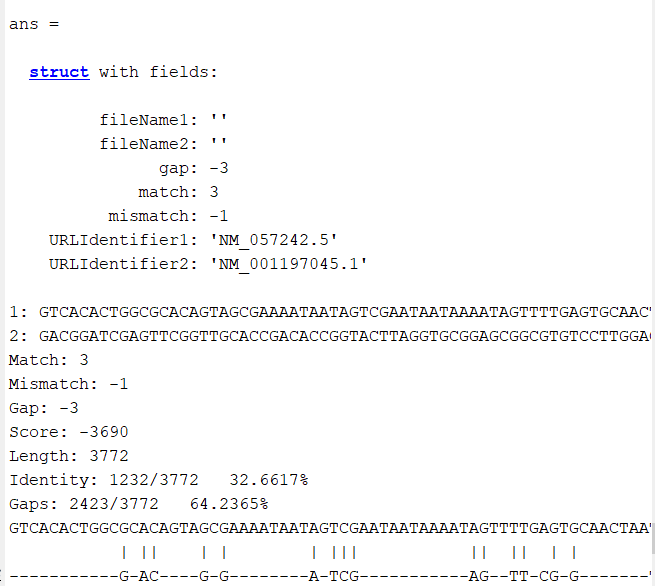
1. **Porównanie par sekwencji ewolucyjnie niepowiązanych**

**Kineaza u muszki owocowej**

Do pokazania sekwencji niepowiązanych ewolucyjnie wykorzystano sekwencję DNA kodujące cytochrom c mRNA u Canis familiaris (psa) i kineazy u drosophila melanogaster (muszki owocowej).

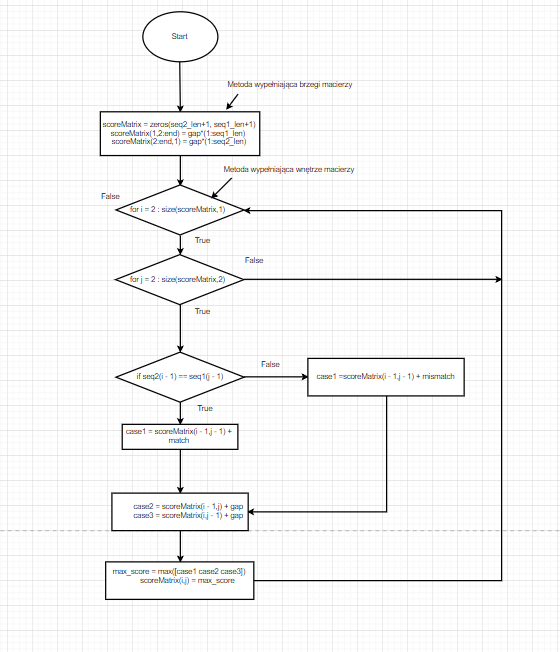
Wynik otrzymano komendą:

NeedWunInterface('URLIdentifier1','NM\_057242.5','URLIdentifier2','NM\_001197045.1','match',3,'gap',-3,'mismatch',-1)



Dopasowanie tych gatunków niewielkie bo wspólne sekwencje to tylko 32% a ilość przerw występujących jest dominująca względem nukleotydów (stanowi 64%). Pokazuje to brak powiązań ewolucyjnych porównywanych sekwencji.

1. **Schemat blokowy algorytmu**



Schemat blokowy algorytmu optymalnego dopasowania