David Bereszczyński 236807

**Wstęp do bioinformatyki**

**ZADANIE 1: Dopasowanie par sekwencji – algorytm kropkowy**

**Odnośnik do repozytorium:**

https://github.com/236807/WBI

1. **Cel ćwiczenia**

Celem zadania było przygotowanie programu do wizualnego porównywania sekwencji DNA za mpomocą za pomocą algorytmu kropkowego oraz analiza otrzymanych wyników.

1. **Opis oraz funkcjonalność programu :**

Program posiada możliwość wczytywania sekwencji w formacie FASTA z dysku, zdalnie z bazy danych NCBI oraz wpisując ręcznie. Następnie odbywa się parsowanie tego pliku i zapis w strukturę tablicy. Program obsługuje pliki w formacie FASTA zawierające 1 lub więcej sekwencji i w zależności od potrzeby można się do tego odnieść. Program pozwala na zapis otrzymanych wykresów.

Program jest wywoływany z poziomu linii komend:

Przykładowe wywołanie:

dotPlot('fileAdres1','DQ334813.1','fileAdres2','LT994967.1','window',23,'error',13,'pngPlotName','plot')

dotPlot('fastaFileName1','sequence1','fastaFileName2','sequence2','window',14,'error',6,'pngPlotName','plot')

Podczas wywoływania program można podać mu następujące argumenty:

‘fastaFileName1’,’Nazwa pliku bez dopisku .txt’

‘fastaFileName2’,’Nazwa pliku bez dopisku .txt’

‘fileAdres1’,’Identyfikator sekwencji FASTA z bazy NCBI’

‘fileAdres2’,’Identyfikator sekwencji FASTA z bazy NCBI’

‘window’,’Szerokość okna’

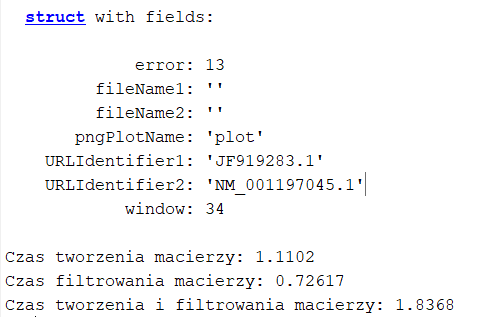
‘error’,’Ilość dopuszczalnych błędów w oknie’

‘pngPlotName’,’Nazwa pliku w którym chcemy zapisać plik .png, w wypadku gdy nie chcemy zapisywać pole należy zostawić puste’

1. **Analiza złożoności obliczeniowej czasowej i pamięciowej**
   1. **Czasowa**

Podczas działania programu wykonywano pomiar czasu trwania wykonania funkcji odpowiadającej za wygenerowanie macierzy kropkowej oraz osobny pomiar dla funkcji filtrującej otrzymaną macierz. W zależności od wykorzystanego pliku FASTA program przeprowadza jeden lub kilka pomiarów w zależności od ilości sekwencji. Analiza czasowa programu jest poprzedzona informacją o id sekwencji na podstawie których macierz została utworzona.

Przykład testu:



* 1. **Pamięciowa**

Analiza pamięciowa została przeprowadzona osobno dla funkcji tworzącej macierz kropkową oraz dla funkcji filtrującej. W każdym z wypadków została wykonana dla sytuacji z najmniejsza możliwą ilością operacji oraz największą możliwą ilością operacji.

**Funkcja tworząca macierz kropkową:**

Najmniejsza możliwa ilość kroków:

Największa możliwa ilość kroków:

**Funkcja filtrująca macierz kropkową:**

Najmniejsza możliwa ilość kroków:

Pomijając stałe dla dużych wartości:

Największa możliwa ilość kroków:

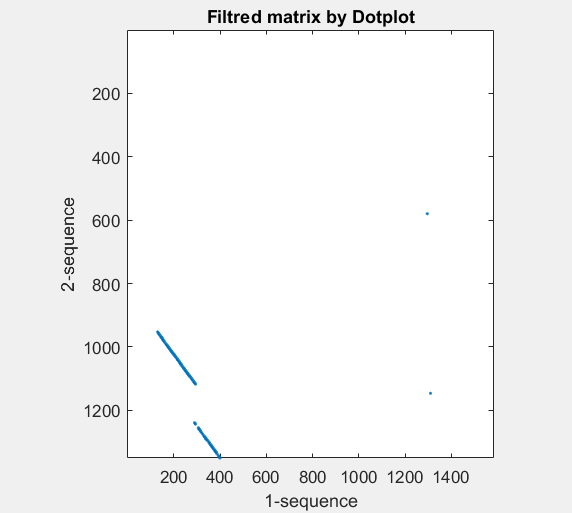
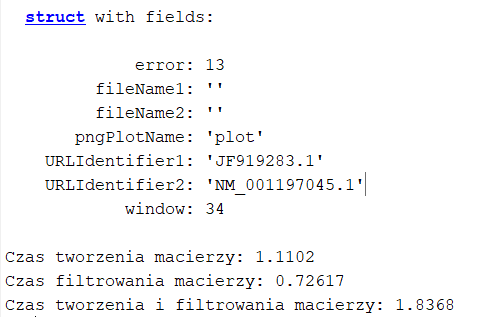
Pomijając stałe dla dużych wartości:

1. **Porównanie oraz interpretacja macierzy kropkowej dla przykładowych par sekwencji ewolucyjnie**

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

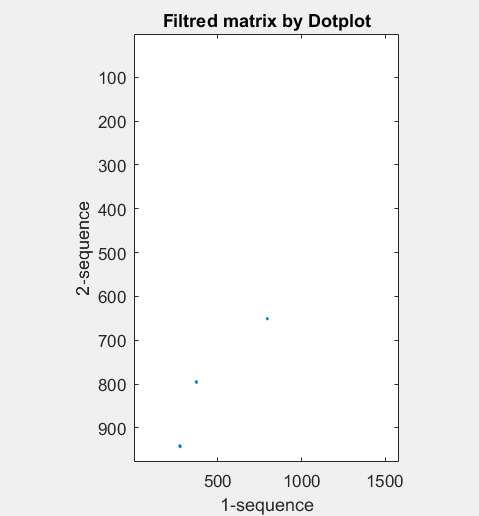
Wynik otrzymano komendą:

dotPlot('fileAdres1','JF919283.1','fileAdres2','NM\_001197045.1','window',34,'error',13,'pngPlotName','plot')

Pokrewieństwo u tych gatunków można zauważyć od około 950 nukleotydu w cytochromie c. Pomiędzy około 1150 do 1250 wystąpiła przerwa (przesunięcie) w podobieństwie co świadczy o wystąpieniu insercji sekwencji po lewej stronie bądź delecji sekwencji po prawej.

Do pokazania sekwencji niepowiązanych genetycznie wykorzystano sekwencję DNA kodujące cytochrom c mRNA u Canis Lupus familiaris (psa) i genu APOB u Canis Lupus (wilka).



W powyższym przykładzie nie wykazano podobieństwa w sekwencjach tylko pojedyncze szumy.