

Wojciech Miśta, 236453, Piątek 11:15

GitHub: <https://github.com/Cloud11PL/Bioinformatyka/tree/zadanie1>

Z uwagi na fakt, że MatLab nie jest w stanie przetworzyć dużej ilości danych, w szczególności większości sekwencji nukleotydów z bazy danych NCBI, poniższe wyniki dotyczą głównie pojedynczych, wybranych białek o ich niewielkiej ilości.

Wyznaczenie pokrewieństwa różnych gatunków

Między sobą porównano dwa gatunki zwierząt: nosorożca oraz słonia azjatyckiego.

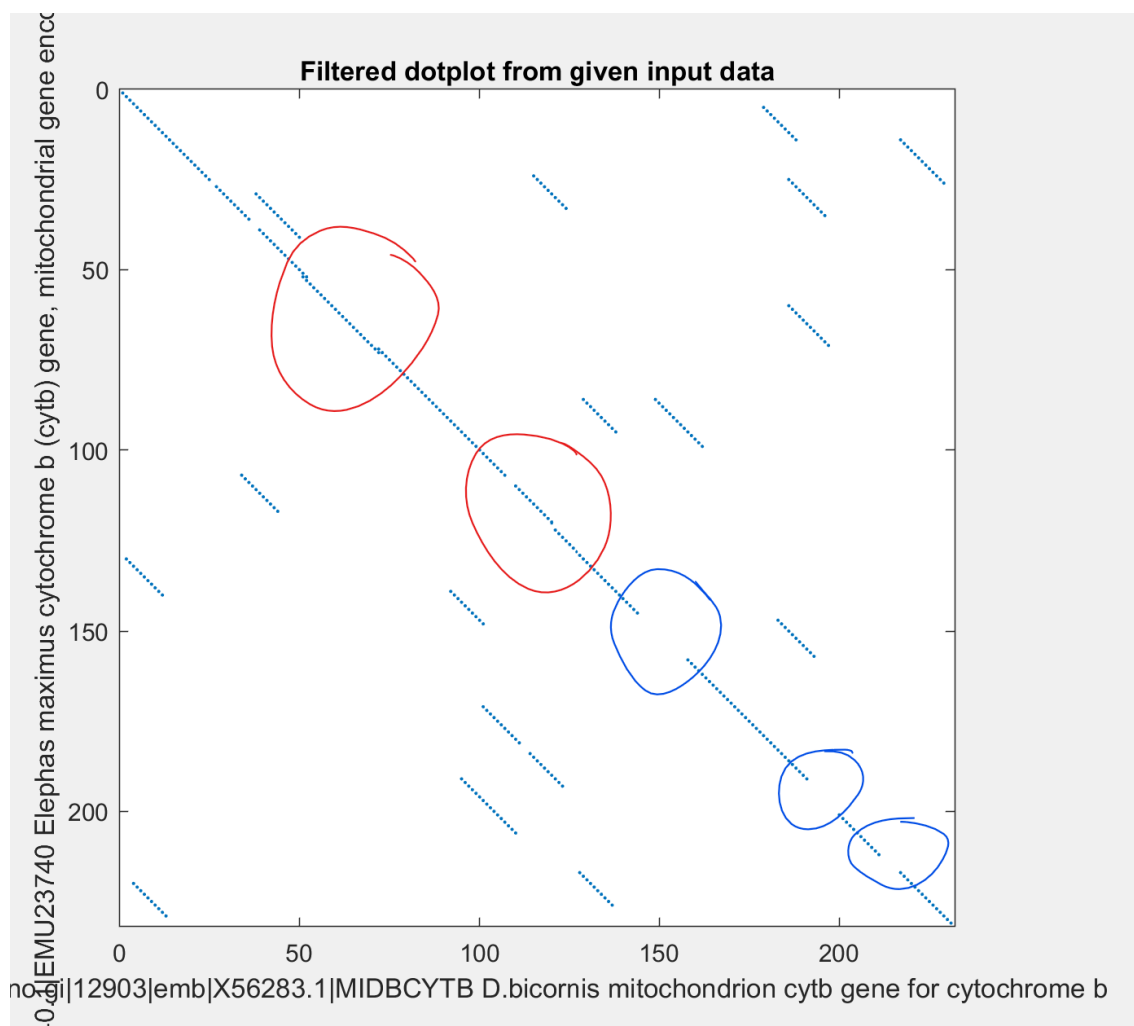


Figure 1 Filtrowany wykres dot-plot sekwencji cytochromu B nosorożca oraz słonia azjatyckiego

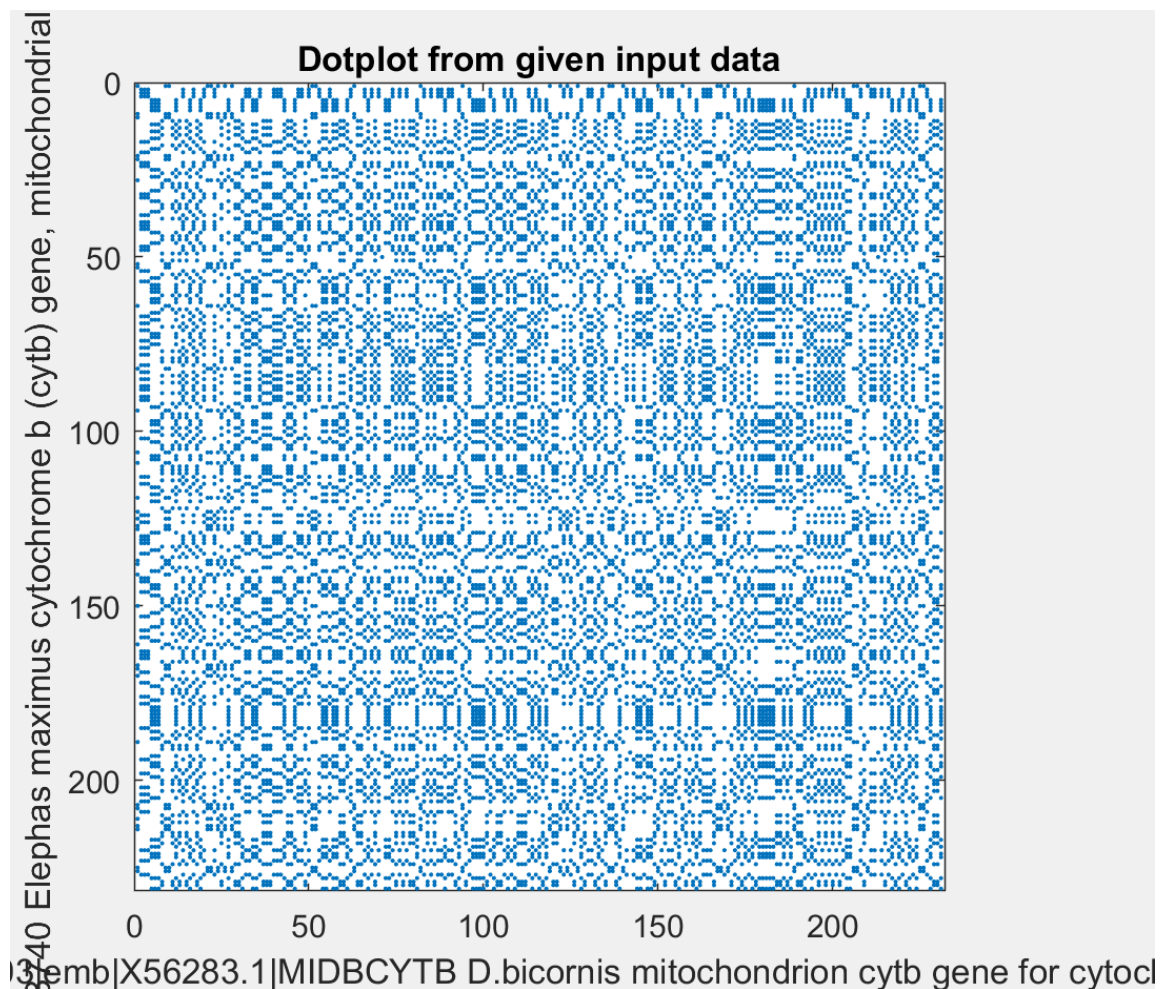


Figure 2 Wykres dot-plot sekwencji cytochromu B nosorożca oraz słonia azjatyckiego

Za pomocą wykresu kropkowego oceniono stopień podobieństwa pomiędzy gatunkami oraz zaobserwowano mutacje w sekwencji. Na filtrowanym wykresie (Window: 10, Threshold: 2) zauważalne są charakterystyczne przekątne linie świadczące o pewnej zgodności sekwencji. Linie te wskazują na średni stopień podobieństwa względem tego samego białka.

Kolorem niebieskim na *Figure 1* zaznaczono substytucje. Kolorem czerwonym zaznaczono możliwe kolejno insercje, delecję oraz ponowna insercję. Na pierwszym zaznaczeniu od góry o kolorze czerwonym zauważamy nachodzące się linie co wskazuje na duplikację.

Wysokie powiązanie

Zastosowano dużą wartość okna oraz mały stopień odcięcia w celu wyznaczenia pokrewieństwa między gatunkami – Window: 19, Threshold: 5. Porównano sekwencję białka insuliny gatunków Homo Sapiens oraz Gorilla Gorilla Gorilla.

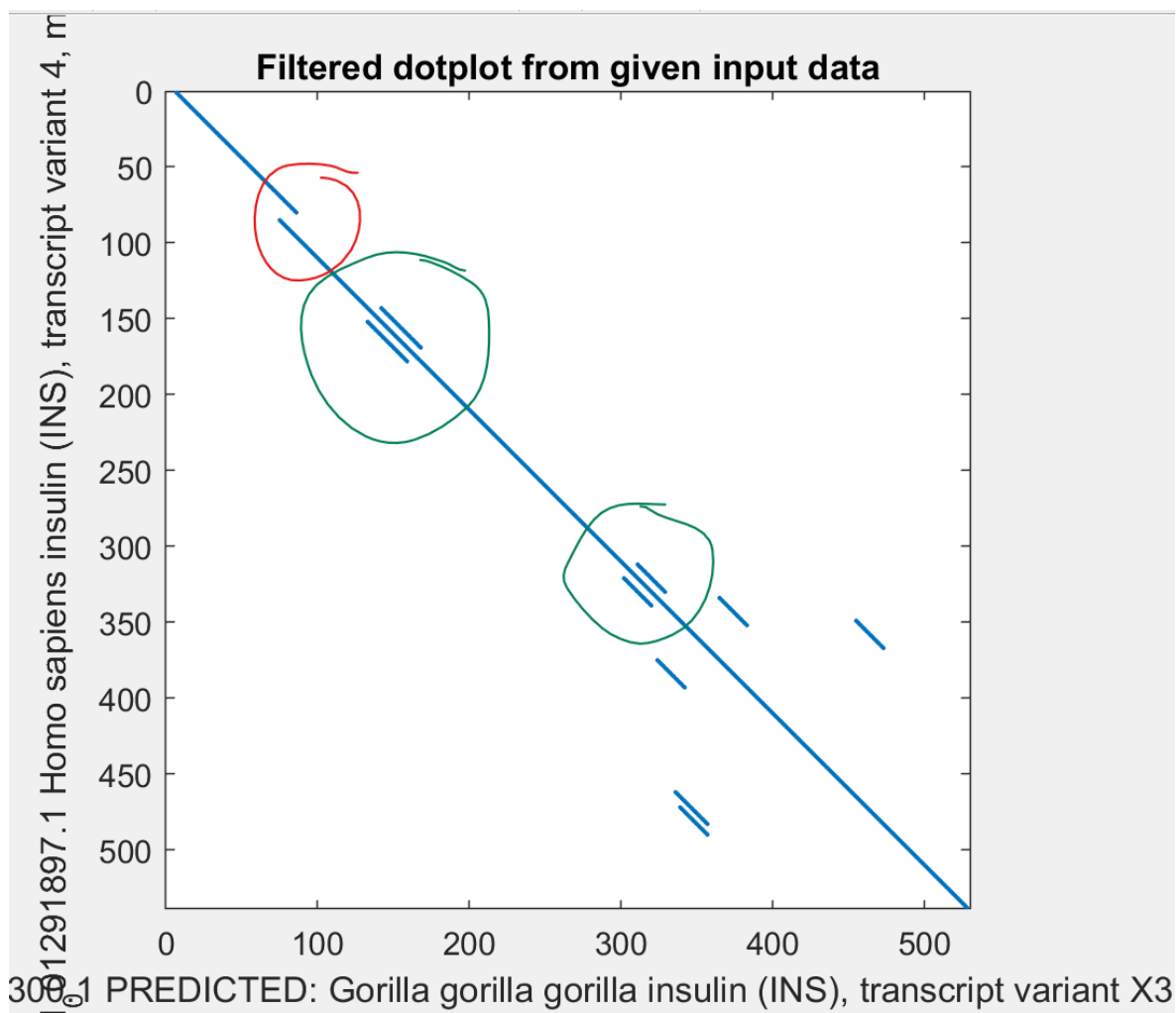


Figure 3 Filtrowany wykres dot-plot sekwencji insuliny człowieka oraz goryla

Kolorem czerwonym zaznaczono mutację, która może być insercją bądź delecją. Widoczny jest brak substytucji na linii przekątnej. Wynik ten wskazuje na bliskie pokrewieństwo gatunków co ma potwierdzenie ewolucyjne.

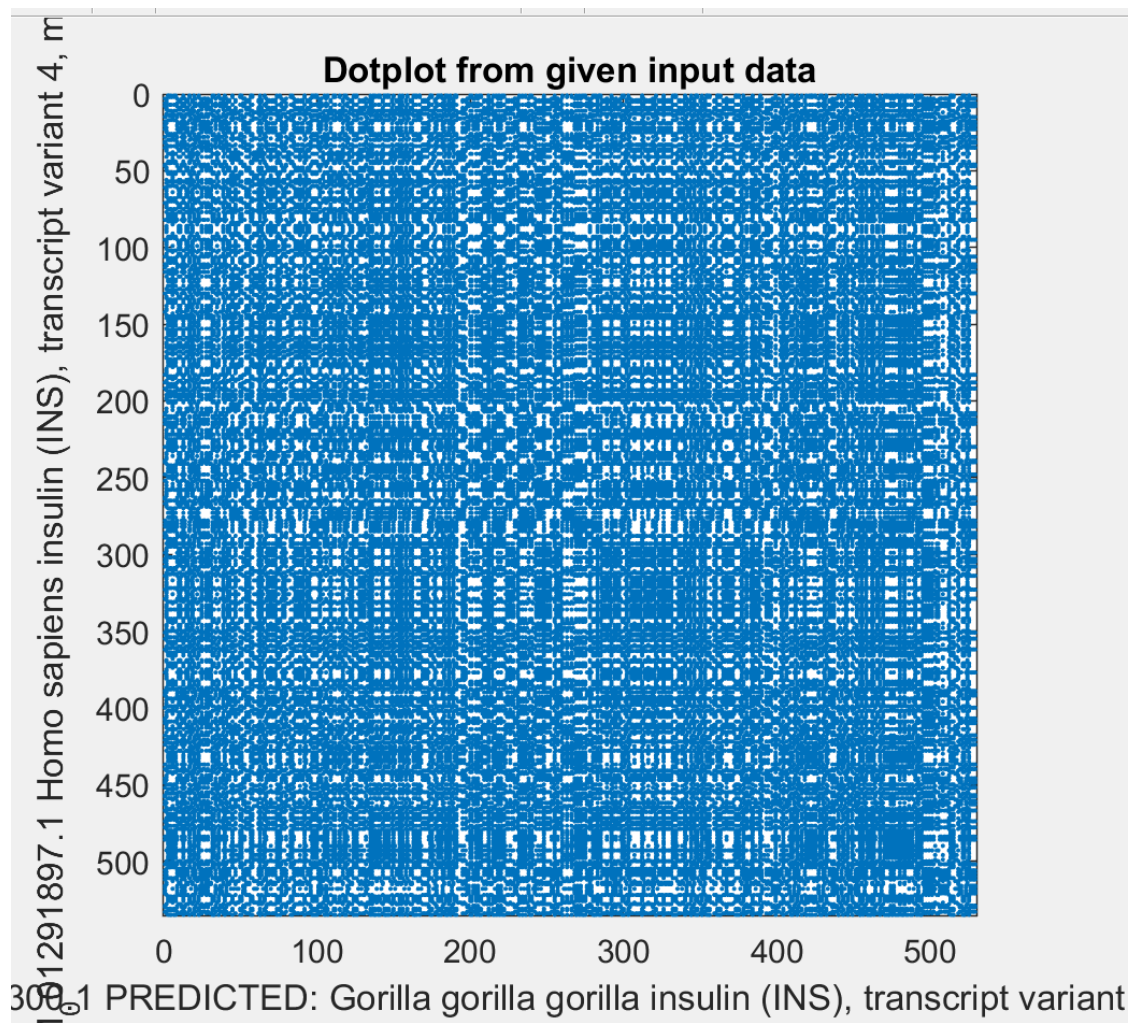


Figure 4 Niefiltrowany wykres dot-plot sekwencji insuliny człowieka oraz goryla

Niefiltrowany wykres również ukazuje wysokie podobieństwo sekwencji białka obu gatunków.

Niskie powiązanie

W celu znalezienia niskiego powiązania porównano sekwencję białka insuliny kota domowego i goryla. (Window: 15, Threshold: 2)

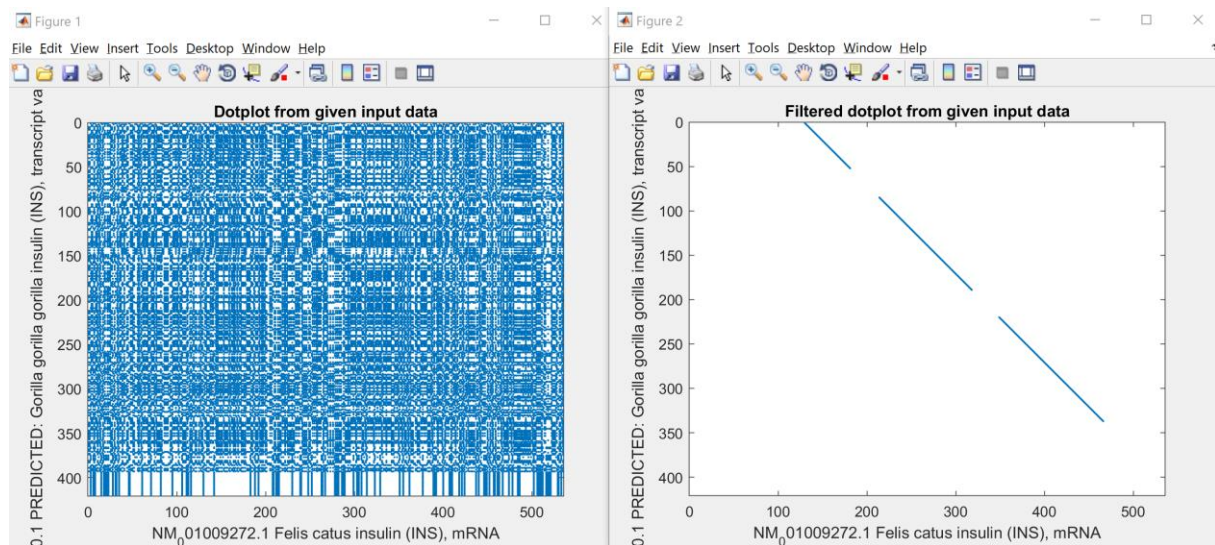


Figure 5 Wykresy dot-plot sekwencji insuliny kota domowego oraz goryla

Przez liczne i znaczne delecje możemy stwierdzić niskie powiązanie obu gatunków.

Do wyznaczenia **czasowej złożoności obliczeniowej** danego programu pominięto elementy wspomagające interakcję użytkownika z programem i skupiono się tylko na kodzie, który bezpośrednio odpowiada za analizę danych. W kodzie znajdują się odpowiednie komentarze, które pokazują liczbę operacji podstawowych potrzebnych do realizacji zadania.

Przykładowa analiza czasowej złożoności obliczeniowej:

```
function [window] = InputMatrix(matrix,sizer,error)
    [rows,columns] = size(matrix);
    window = zeros(size(matrix));
    for w = 1:rows
        for k = 1:columns
            if matrix(w,k) == 1
                foundZeros = 0;
                for i = 1:sizer-1
                    if ~(w+i < rows && k+i < columns && matrix(w+i,k+i) == 1)
                        foundZeros = foundZeros+1;
                    end
                end
                if foundZeros <= error
                    window(w,k) = 1;
                    for i = 1:sizer-1
                        window(w+i,k+i) = 1;
                    end
                end
            end
        end
    end
end
```

%Set values of columns and rows to
%Preallocation +1
%Checking condition +1
%Checking condition +1
%Checking condition +1
%Setting value +1
%Checking condition +1
%Checking 3 conditions +3
%Setting value +1

%Checking condition +1
%Setting matrix values +1
%Checking condition +1
%Setting matrix values +1

- 1 przypisanie zmiennej,
- 1 prealokacja
- 1 sprawdzenie warunku (*x)
 - 1 sprawdzenie warunku (*k)
 - 1 sprawdzenie warunku
 - 1 przypisanie
 - 1 sprawdzenie warunku (*i)
 - 3 sprawdzenia warunków
 - 1 przypisanie zmiennej
 - 1 sprawdzenie warunku
 - 1 Przypisanie zmiennej
 - 1 Sprawdzenie warunku (*i)
 - 1 Przypisanie zmiennej

Tylko dla metody sprawdzającej okno macierzy, łączna liczba operacji wynosi:

$$T(w, k, i) = 1 + 1 + x * k * (1 + i * 4 + 1 + 1 + i * 1) = 2 + w * k * (3 + 5i)$$

Przykładowa analiza złożoności przestrzennej:

- Zmienne: 8 (wraz z argumentami wejściowymi)
- Macierz: 1 w*k (argument wejściowy)

Złożoność: $8 + w * k$