

Zadanie 0.

Dokończyć zadania z poprzedniego zestawu.

Zadanie 1.

Pobrać pierwsze 10^5 nukleotydów sekwencji ludzkiego chromosomu 1. Zapisać do pliku tekstowego. Odpowiedzieć na następujące pytania dotyczące tej sekwencji:

- (a) obliczyć ilość i częstość występowania A, C, G i T
- (b) znaleźć pozycję i ilość kodonów STOP
- (c) pociąć sekwencję na odcinki zawierające STOP na końcu; obliczyć ich długości i zapisać na dysku
- (d) znaleźć ilość i pozycje na których znajduje się AGATA; czy znalezione pozycje zachodzą na siebie?
- (e) znaleźć najdłuższą sekwencję powtarzających się nukleotydów (np. AAAAAA...)

Zadanie 2*

Jak w zadaniu 1, ale dla całego chromosomu 1 oraz pozostałych. Porównać wyniki.

UWAGA: W wielu miejscach brakuje danych: należy te obszary pominąć. Efektywne rozwiązanie problemu może wymagać ładowania fragmentów sekwencji w „paczkach”, po kilkaset - kilka tysięcy nukleotydów.

Zadanie 3

Dokonać transkrypcji oraz translacji ciągów z Zad. 1 stosując standardowy kod genetyczny.

WSKAZÓWKA: Można zacząć od Zad. 4 i zastosować wynik do Zad. 3

Zadanie 4

Napisać funkcje dokonujące transkrypcji i translacji ciągu nukleotydów. Zastosować standardowy kod genetyczny. Sprawdzić wyniki za pomocą WolframAlpha (np. <http://www.wolframalpha.com/input/?i=AGATAC>)