

Pracownia bioinformatyczna

– arkusz zadań z miejscami do udzielenia odpowiedzi

Zanim zaczniesz rozwiązywać zadania, wpisz swoje imię i nazwisko oraz nr PESEL.

Imię i nazwisko

PESEL

Informacje dotyczące pracowni bioinformatycznej

Zanim zaczniesz rozwiązywać zadania, zastąp „nnnnnnnnnn” w nazwie tego pliku swoim numerem PESEL. Po zakończeniu egzaminu niezwłocznie prześlij ten plik opiekunowi.

Zadanie 1. (21 pkt)

Punktem wyjścia do zadań na pracowni bioinformatyczno-filogenetycznej stanowi struktura kompleksu białkowego o numerze dostępu **6MOJ** w bazie danych Protein Data Bank. Możesz pobrać plik .pdb i analizować pobraną strukturę lokalnie w takich programach jak np. UCSF Chimera lub PyMol. Możesz także wykorzystać funkcję „3D View” bezpośrednio dostępną w bazie danych PDB.

Rozwiązując zadania z tej pracowni, możesz korzystać zarówno z programów zainstalowanych na komputerze, jak również z narzędzi analitycznych i obliczeniowych dostępnych przez Internet, takich jak np.: NCBI, UniProt, czy Expasy. Możesz także korzystać z literatury źródłowej, do której prowadzą odnośniki w bazach danych.

Zadanie 1.1. (3 pkt)

Określ liczbę reszt aminokwasowych wchodzących w skład fragmentu białka S pochodzącego z SARS-CoV-2 zawartego w strukturze z numerem dostępu 6MOJ.

Zadanie 1.2. (4 pkt)

Określ całkowitą liczbę reszt aminokwasowych wchodzących w skład białka S pochodzącego z SARS-CoV-2. Do określenia tej liczby wykorzystaj referencyjną sekwencję genomu SARS-CoV-2 zdeponowaną w bazie danych GenBank z numerem dostępu NC_045512.2.

Zadanie 1.3. (5 pkt)

Określ, które stwierdzenia dotyczące struktury zdeponowanej pod numerem dostępu 6M0J są prawdziwe, a które fałszywe.

Stwierdzenie	Prawda czy fałsz?
1. Do uzyskania homogennego preparatu obu białek – białka S i receptora ACE2 – użyto chromatografii powinowactwa z wykorzystaniem motywu His-tag.	prawda / fałsz
2. Receptor dla SARS-CoV-2, białko ACE2, jest hydrolazą.	prawda / fałsz
3. Jeden z atomów asparaginy w pozycji 501 w białku S bierze udział w tworzeniu wiązania kowalencyjnego z jednym z atomów receptora ACE2.	prawda / fałsz
4. Struktura białka S jest stabilizowana przez obecność jonu Zn^{2+} , który oddziałuje z histydyną w pozycji 374 i 378, a także z kwasem glutaminowym w pozycji 402.	prawda / fałsz
5. Dominującą strukturą drugorzędową białka ACE2 jest α -helisa.	prawda / fałsz

Zadanie 1.4. (3 pkt.)

Podaj numer dostępu z bazy danych GenBank dla kompletnej sekwencji genomu wirusa występującego u *Rhinolophus acuminatus* o największym stopniu podobieństwa do referencyjnego genomu SARS-CoV-2: NC_045512.2.

Użyj programu blastn, a do pól „Organism” i „Entrez Query” wpisz, odpowiednio „viruses (taxid:10239)” i „Rhinolophus acuminatus[FEATURES]”.

Zadanie 1.5. (3 pkt.)

Wykonaj w programie blastp przyrównanie sekwencji aminokwasowych białek S obu wirusów korzystając z domyślnych ustawień, ale zaznacz „Align two or more sequences”.

Podaj identyczność sekwencji aminokwasowych obu białek (ang. *identities*) jako wartość wyrażoną w procentach.

Zadanie 1.6. (3 pkt)

Wykonaj w programie blastn przyrównanie sekwencji nukleotydowych kodujących białka S obu wirusów. Wykonaj przyrównanie przy domyślnych ustawieniach, ale zaznacz „Align two or more sequences” i w części „Program Selection” zaznacz „Somewhat similar sequences (blastn)”, aby można było wykonać przyrównanie dwóch sekwencji nukleotydowych.

Określ, czy identyczność pełnych sekwencji nukleotydowych kodujących białka S obu wirusów jest mniejsza czy większa w porównaniu do identyczności sekwencji aminokwasowych określonej w zadaniu 1.5.

A. mniejsza / B. większa

Zadanie 2. (9 pkt)

Przygotuj sekwencje nukleotydowe następujących genomów:

- SARS-CoV-2 wyizolowany z *Homo sapiens* (NC_045512.2),
- SARS Tor2 wyizolowany z *Homo sapiens* (NC_004718.3),
- RacCS203 wyizolowany z *Rhinolophus acuminatus* (MW251308.1),
- RaTG13 wyizolowany z *Rhinolophus acuminatus* (MN996532.2),
- Neoromicia/PML-PHE1/RSA/2011 wyizolowany z *Neoromicia capensis* (KC869678.4),
- MP789 wyizolowany z *Manis javanica* (MT121216.1),
- SW1 wyizolowany z *Delphinapterus leucas* (NC_010646.1),
- HKU14 wyizolowany z *Oryctolagus cuniculus* (NC_017083.1).

Następnie przygotuj plik w formacie FASTA z sekwencjami nukleotydowymi kodującymi białko S z wyżej wymienionych 8 genomów wirusów.

Wykorzystaj wersję on-line programu ClustalO (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) do przygotowania przyrównania sekwencji kodujących białko S. Sekwencje w formacie FASTA możesz wkleić do pola „STEP 1 – Enter your input sequences” lub załadować, korzystając z przycisku znajdującego się bezpośrednio pod tym polem. Pamiętaj, aby wybrać typ sekwencji – RNA.

Pozostałych ustawień nie zmieniaj i kliknij w części „STEP 3 – Submit your job” przycisk „Submit”.

Uwaga: obliczenia mogą potrwać nawet kilka minut!

Gdy wyświetli się przyrównanie, kliknij na zakładkę „Phylogenetic Tree” i odpowiedz na następujące pytania dotyczące relacji filogenetycznych 8 wirusów odtworzonej z wykorzystaniem sekwencji aminokwasowej białka S.

Uwaga: drzewo jest obliczane za pomocą metody odległościowej Neighbour-joining, która jest w ogólnych założeniach podobna do metody UPGMA, ale zwraca drzewo formalnie niezakorzone i nie łamie gałęzi dokładnie w połowie ich długości.

Zadanie 2.1. (9 pkt)

Określ, które stwierdzenia dotyczące interpretacji uzyskanego drzewa filogenetycznego są prawdziwe, a które fałszywe.

Stwierdzenie	Prawda czy fałsz?
1. SARS-CoV-2 i RaTG13 stanowią taksony siostrzane.	prawda / fałsz
2. SARS Tor2 i SARS-CoV-2 tworzą razem grupę monofiletyczną.	prawda / fałsz
3. SARS-CoV-2 jest bliżej spokrewniony z Neoromicia/PML-PHE1/RSA/2011 z <i>Manis javanica</i> (łuskowca) niż SARS Tor2.	prawda / fałsz