

# 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 $\alpha$ -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

# Pracownia bioinformatyczna

## – zasady oceniania rozwiązań zadań

### Zadanie 1.1. (3 pkt)

3 pkt – za podanie liczby reszt aminokwasowych wchodzących w skład fragmentu białka S w łańcuchu E struktury o numerze dostępu 6M0J (229 reszt) lub liczbę pomniejszoną o dołączony do tej sekwencji aminokwasowej znacznik histydynowy (223 reszty).

0 pkt – za błędną odpowiedź lub brak odpowiedzi.

Find similar proteins by:  (by identity cutoff) | [Structure](#)

Entity ID: 2					
Molecule	Chains	Sequence Length	Organism	Details	Image
Spike protein S1	E	229	<a href="#">Severe acute respiratory syndrome coronavirus 2</a>	Mutation(s): 0 Gene Names: <a href="#">S</a> , <a href="#">2</a>	

Find proteins for [PODTC2](#) (*Severe acute respiratory syndrome coronavirus 2*)    Explore [PODTC2](#)    Go to UniProtKB: [PODTC2](#)

### Zadanie 1.2. (4 pkt)

4 pkt – za podanie liczby reszt aminokwasowych wchodzących w skład białka S ustalonej na podstawie informacji zawartej w referencyjną sekwencję genomu SARS-CoV2 zdeponowaną w bazie danych GenBank z numerem dostępu NC\_045512.2 (1273 reszty aminokwasowe).

0 pkt – za błędną odpowiedź lub brak odpowiedzi.

```

21361 aatccaattc agtggcttc ctattcttta ttgacatga gtaaaattcc cctiaaatta
21421 aggggtactg ctggtatgct tttaaaagaa ggtcaaatca atgatatgat tttatctctt
21481 cttagtaaaq gtgactat aattagagaa aacaacagag ttgttatctc tagtgatggt
21541 cttgtaaca actaaacgaa caatgtttgt tttcttgg ttattgccac tagtctctag
21601 tcagtggtt aatcttacia ccagaactca attaccctcc gcatacacta atctttccac
21661 acggtggtt tattaccctg acaaaagttt cagatcctca gttttacatt caactcagga
21721 cttgttcta cttttctttt ccaatgttac ttggttccat gctatacatg tctctgggac
21781 caatggtaac aagaggtttg aataacctgt cctaccattt aatgatggtg tttatttgc
21841 ttccactgag aagtctaa caataagagg ctggattttt ggtactactt tagattgcaa
21901 gaccagctcc ctacttatg ttaataacgc tactaatgtt gttataaag tctgtaatt
21961 caattttgt aatgatccat ttttgggtgt ttattaccac aaaaacaaca aaagtggat
22021 ggaagatgag ttcaagattt attctagtc gaataattgc acttttgaat atgtcttcca
22081 gcttttctt atgacactgt aaggaaaca gggaatttc aaaaactcta ggaatttgt
22141 gtttaagaat atgatggtt attttaaatt atattctaaq cacacgcta ttaattagt
22201 gctgatctc cctcaggtt tttcggcttt agaaccattg gtatagttgc caataggtat
22261 taacatcact agtttcaaa ctttacttgc ttacataga agttatttga ctctcgttga
22321 ttcttctca ggttggacag ctggtctctc agcttattat gttgggtatc ttaacactag
22381 gaattttcta ttaaaatata atgaaatgga aaccattaca gatgctgtag actgtaact
22441 tgacctctc tcagaacaaa agttagcttt gaaatcttc actgtagaaa aaggaactca
22501 tcaaaacttc aactttagag tccaaccaac agaactattt gttagatttc ctaattatc
22561 aacttctgc ccttttgggt aagtttttaa cgccaccaga ttgcatctg tttatgctg
22621 gaacaggaag agaatacagca actgtgttgc tgattattc gtccatatata attcgcactc
22681 atttccact ttaagtgtt atggagtctc tctactaaa ttaaatgac tctgcttacc
22741 taatgctat gcagattcat ttgtaattag aggtgatga gtcagacaaa tgcctcagg
22801 gcaactgga aagattgctg attataatta taattacca gatgatttca caggctcgt
22861 tatacctgg aatttcaaca actttgattc taaggttgtt gtaattata attaccgtta
22921 tagattgtt aggaagtcta atctcaaac ttttgagaga gatatttcaa ctgaacttca
22981 tcaggccgtt agcacactt gtaattggtt tgaaggtttt aattgttact ttcctttaca
23041 atcataggt ttccaaccca ctaattggtt tggttacca ccaatacagag tagtagtact
23101 ttctttgaa ctctacatg caccagcaac tgtttgga cctaaaagt ctactaattt
23161 gtttaaaaa aatgtgtca atttcaactt caatgttga acagggcacag gtgtcttacc
23221 tgagtctaac aaaaagtctc tgcctttcca acaatttggc agagacattg ctgacactac
23281 tgatgctgc cgtgatcac agacacttga gattcttgc ataccacat gttcttttgg
23341 tgggtcagt gttataaac caggaacaaa tacttctaac caggttctg tctttatca
23401 ggatgtaac tgcaganaag tcccttggc tattcatga gatcaacta ctctacttgc
23461 gcggttttat tctacaggtt ctaattgttt tcaaacactg gcaggtctgt taataggggc
23521 ttaacatgac aacaactcat atgagatga cataccattt ggtgcagga tatggcctag
23581 ttatcagact cagactaatt ctctcgggc ggcactgat gtagctagc aattccatcat
23641 tgcctacact atgtactctg gtgcagaaa ttcagttgct tactctaata atctatttgc
  
```

```

21563..25384
/gene="S"
/locus_tag="GU280_gp02"
/gene_synonym="spike glycoprotein"
/notes="structural protein; spike protein"
/codon_start=1
/product="surface glycoprotein"
/protein_id="YP_009724390.1"
/db_xref="GeneID: 43740568"
/translation="MFVFLVLLPLVSSQCNLTTTRQLPPATNSFRGVYDPKQVFR
SSVLSHSDQLFLPFSNVTFHAIHVSQNGTKRFDPVLPFDNGVYFSTEKSNIR
GWIFGTLDSKTSLLIWNNAIIVKVCDFQNDPFLGVYHKNKSMSEFRVY
SSANNCTFEYVQPLMDLEGQGNFKNREFVKNIDYFKIYKHTPINLVRDLQ
GFSALPLVDLPTGINITRFTLLALHRSYLPDSSSAGTAGAAAYVGLQPRFL
LKYNGITDAVDCALDPLSEKTKLXFTVEKGIYQTSNFRVQPTESIVRFPNITN
LCPFGVFNATRFASVYAMNKRISNVCADYSVLSVNSASFVFKYGVSPKTLNDLFC
TNVYADSVIRGDEVQIAPGGTQKIDADYNYKLPDPTGCVIAMSNNLSDKVVGN
YLRLFRKSNLKPFRDISEIYQAGSTPONGVEGFCNYPQLQSYGFQPTNGVYQPY
RVVFLFELHAPATVCGPKSNLTKNKNVNFNGLTGTGLTESNKKLFPQFGG
RDIADTDAVRDPTLEILDITPCFSGVSVITPNTSNQVAVLVDVNCTEVPVAI
HADQLPTWRVYSTGSNVFTRAGCLIGAHEVNNSEYCDIPGAGICASYQVTKNSPR
RARSVASQSIAYTMSLGAENSVAVNSIAIPTNFTISVITTEILPVSMTKTSVDCTM
YICGDSSTCSNLLQYGSFCTDLNRLATGIAVEQDNKTEVEFAQVKQIYKTPKIDFG
GNFSQILPDPSPKSRSFIEDLNFKNVLTADAGFIKQYGDCLDIIARDLCAQKFN
GLTVLPLLDHEIAQYTSALLAGTITSGWTFGAGAAQLQIPFAMQMAIRFNGIGVTON
VLVENKLIANQFNIAIGKIQDLSSTASALGKLDVNNQNAQALNTLVKLSNFGA
ISSVLDIISRLDKVEAEVQIDRLITGRQLSLOTYVTQQLIRAAETIRASANLAATKMS
ECLVGSKRVDGCKGYHLMSFPOSAHPHQVFLHWYVADQEKFTTAPALCIDGKAH
FRFEEGLDKYFNHTSPVDLGDIGSINASVNIQKEDIRLNEVAKNLSLIDQLG
KVEYIKWFWIWLGFAGLIAIIVMTIMCLTSCSCSCLKGCSSCKCFDEDDSE
PVLKGVKLYHT"
  
```

### Zadanie 1.3. (5 pkt)

5 pkt – za poprawną ocenę wszystkich pięciu stwierdzeń.

1–4 pkt – odpowiednio za liczbę poprawnie ocenionych stwierdzeń.

0 pkt – za błędną ocenę wszystkich stwierdzeń lub brak odpowiedzi.

**Uwaga:**

W stwierdzeniu 1. należało stwierdzić obecność sekwencji aminokwasowej HHHHHH (znacznik histydynowy) w obu łańcuchach (A i E) w strukturze o numerze dostępu 6MOJ.

Reference Sequence 6MOJ\_1

PDB ENTITY 6MOJ_1	P L L N Y F E P L F T W L K D Q N K N S F V G W S T D W S P Y A D H H H H H H
UNIPROT ALIGN Q9BYF1	P L L N Y F E P L F T W L K D Q N K N S F V G W S T D W S P Y A D
UNMODELED A	
ARTIFACT	

Stwierdzenie 4. jest prawdziwe, ale w odniesieniu do białka ACE2, które jest w kompleksie z fragmentem białka S. Dlatego stwierdzenie 4., mówiące o tym, że jon  $Zn^{2+}$  stabilizuje białko S, należało ocenić jako fałszywe.

Angiotensin-converting enzyme 2  
6MOJ | Model 1 | Instance ASM\_1 | A | GLU 357 [auth 375]

ZINC ION: ZN 901 | C [auth A]

Measurements

Structural Motif Search

Components 6MOJ

- Carbohydrate 2 reprs
- Water Ball & Stick
- Ion Ball & Stick
- [Focus] Target Ball & Stick
- [Focus] Surroundings (5 Å)

Unit Cell P 41 21 2

Density

Assembly Symmetry

Export Animation

**Zadanie 1.4. (3 pkt)**

3 pkt – za podanie właściwego numeru dostępu (MW251308.1).

0 pkt – za błędną odpowiedź lub brak odpowiedzi.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Bat coronavirus RacCS264, partial genome	Bat coronavirus RacCS264	27475	42052	92%	0.0	94.70%	29820	MW251311.1
<input checked="" type="checkbox"/> Bat coronavirus RacCS203, complete genome	Bat coronavirus RacCS203	27475	42217	94%	0.0	94.70%	29832	MW251308.1
<input checked="" type="checkbox"/> Bat coronavirus RacCS271, partial genome	Bat coronavirus RacCS271	27470	42076	93%	0.0	94.69%	29820	MW251312.1
<input checked="" type="checkbox"/> Bat coronavirus RacCS253, partial genome	Bat coronavirus RacCS253	27470	42082	92%	0.0	94.70%	29820	MW251310.1
<input checked="" type="checkbox"/> Bat coronavirus RacCS224, partial genome	Bat coronavirus RacCS224	27470	42107	92%	0.0	94.70%	29820	MW251309.1

**Uwaga:**

Należało wybrać numer dostępu sekwencji, która w nazwie (kolumna „Description”) ma określenie „complete genome”, ponieważ tylko ta sekwencja zawiera kompletną sekwencję.

**Zadanie 1.5. (3 pkt)**

3 pkt – za podanie identyczność sekwencji wynoszącą 72,75% lub zawierającą się w przedziale [72%, 73%].

0 pkt – za błędną odpowiedź lub brak odpowiedzi.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> unnamed protein product		1834	1834	100%	0.0	72.75%	1227	Query_546525

**Zadanie 1.6. (3 pkt)**

3 pkt – za wybór odpowiedzi A (mniejsza).

0 pkt – za błędną odpowiedź lub brak odpowiedzi.

**Uwaga:**

Zastosowanie opcji „Somewhat similar sequences (blastn)” pozwala określić identyczność, która wynosi 79,24%. Jednak „Query Cover” (stopień pokrycia sekwencji) wynosi 79%.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> None provided		2503	2606	79%	0.0	79.24%	3684	Query_39677

Biorąc pod uwagę, że zadanie dotyczy sekwencji kodujących białka S (a nie ich fragmentów), należy wziąć pod uwagę pozostałe części sekwencji (21%), które przez program nie zostały zidentyfikowane jako podobne do siebie. Wartość identyczności wynikająca z porównania całej sekwencji będzie więc mniejsza niż 72,75%, która wynika z porównania sekwencji aminokwasowych białka S (zadanie 1.5.). Ponadto warto zwrócić uwagę na to, że ze względu na degenerację kodu genetycznego należy spodziewać się większego zróżnicowania sekwencji nukleotydowych niż aminokwasowych kodujących ortologiczne białka.

**Zadanie 2.1. (9 pkt)**

9 pkt – za poprawną ocenę wszystkich trzech stwierdzeń.

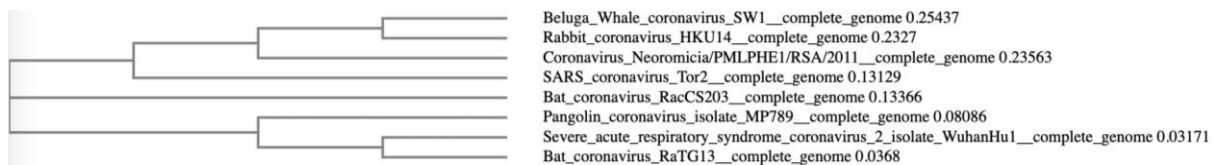
6 pkt – za poprawną ocenę dwóch stwierdzeń.

3 pkt – za poprawną ocenę jednego stwierdzenia.

0 pkt – za błędną ocenę wszystkich stwierdzeń lub brak odpowiedzi.

**Uwaga:**

Każdy uczestnik uzyskał 3 pkt za to zadanie, ponieważ w stwierdzeniu 3. błędnie przypisano wirusa Neoromicia/PML-PHE1/RSA/2011 do gatunku *Manis javanica*, przez co niemożliwa była prawidłowa jego ocena.



# Pracownia bioinformatyczna

## – arkusz zadań z miejscami do udzielenia odpowiedzi

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

Imię i nazwisko

KGOB

PESEL

2222222222

*Informacje dotyczące pracowni bioinformatycznej*

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

### Zadanie 1. (21 pkt)

Punktem wyjścia do zadań na pracowni bioinformatyczno-filogenetycznej stanowi struktura kompleksu białkowego o numerze dostępu **6M0J** 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 6M0J.**

229

*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.**

1273



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.	<input checked="" type="checkbox"/> prawda / <input type="checkbox"/> fałsz
2. Receptor dla SARS-CoV-2, białko ACE2, jest hydrolazą.	<input checked="" type="checkbox"/> prawda / <input type="checkbox"/> fałsz
3. Jeden z atomów asparaginy w pozycji 501 w białku S bierze udział w tworzeniu wiązania kowalencyjnego z jednym z atomów receptora ACE2.	<input type="checkbox"/> prawda / <input checked="" type="checkbox"/> 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.	<input type="checkbox"/> prawda / <input checked="" type="checkbox"/> fałsz
5. Dominującą strukturą drugorzędową białka ACE2 jest $\alpha$ -helisa.	<input checked="" type="checkbox"/> prawda / <input type="checkbox"/> 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]”.

MW251308.1

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.

72,75%

**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.	<input checked="" type="checkbox"/> prawda / <input type="checkbox"/> fałsz
2. SARS Tor2 i SARS-CoV-2 tworzą razem grupę monofiletyczną.	<input type="checkbox"/> prawda / <input checked="" type="checkbox"/> fałsz
3. SARS-CoV-2 jest bliżej spokrewniony z Neoromicia/PML-PHE1/RSA/2011 z <i>Manis javanica</i> (łuskowca) niż SARS Tor2.	<input checked="" type="checkbox"/> prawda / <input checked="" type="checkbox"/> fałsz