SCIENCE 2

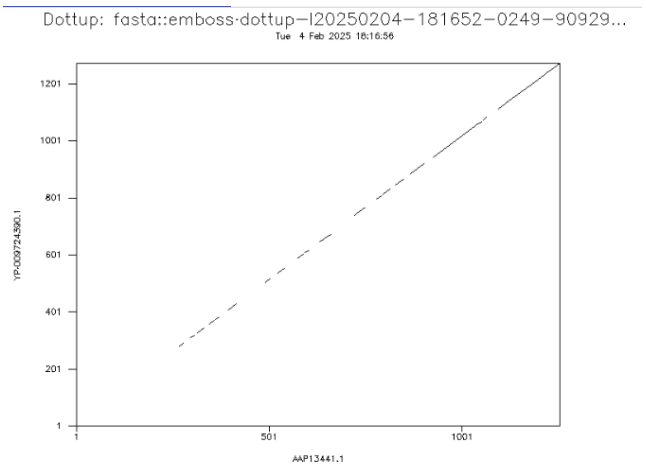
TUTORIAL-1

Name: Varun Gupta
Roll No: 2023101108

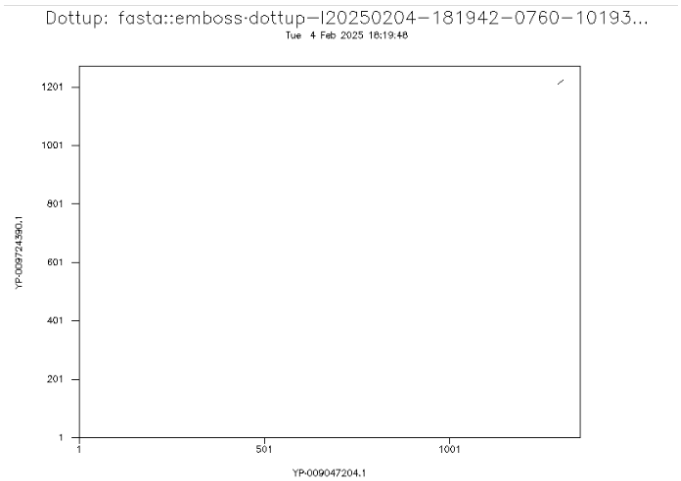
Question 1

Dottup

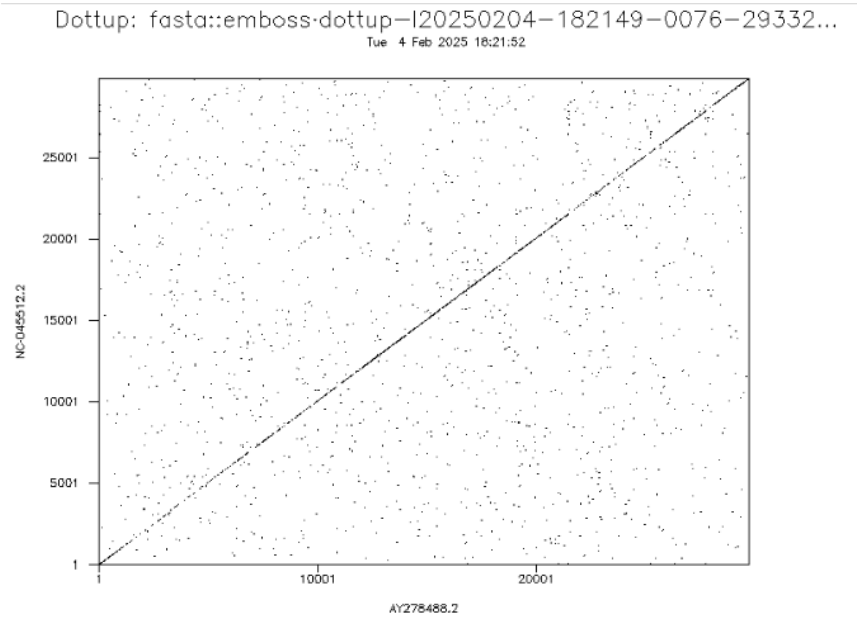
SARS-CoV2 vs SARS-CoV (Protein)



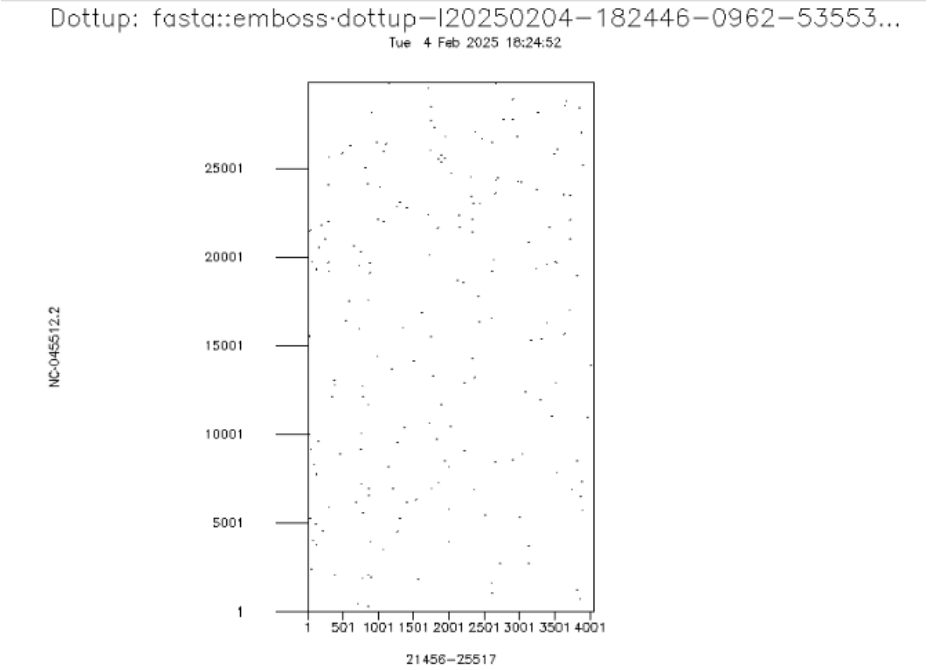
SARS-Cov2 vs MERS-CoV (Protein)



SARS-CoV2 vs SARS-CoV (DNA)

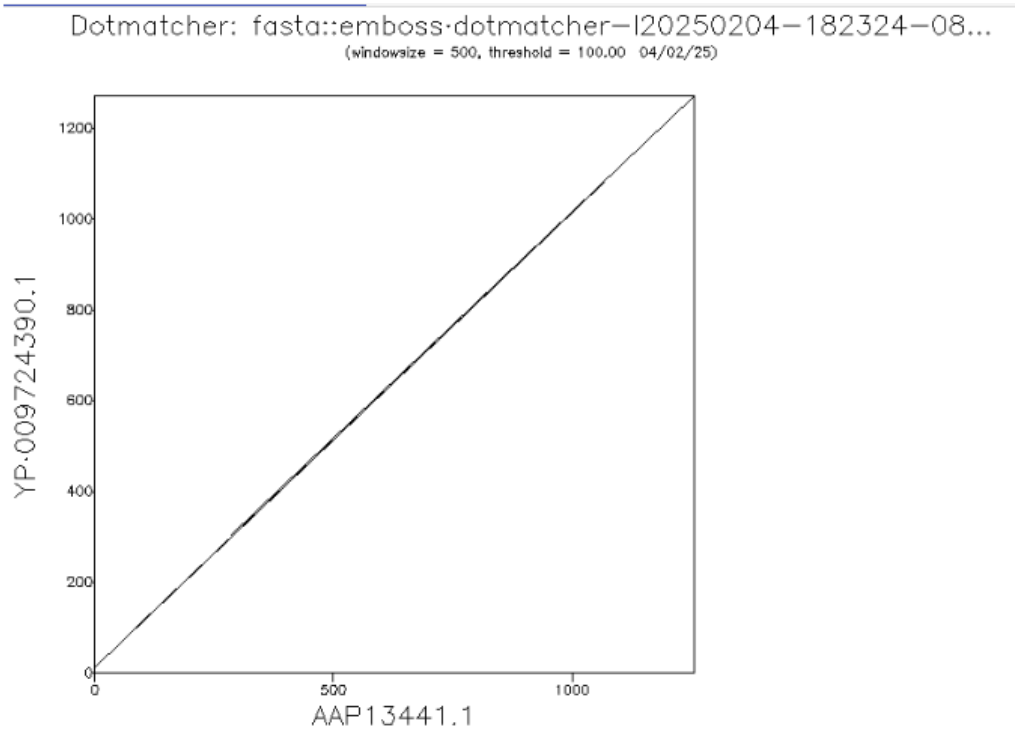


SARS-Cov2 vs MERS-CoV (DNA)

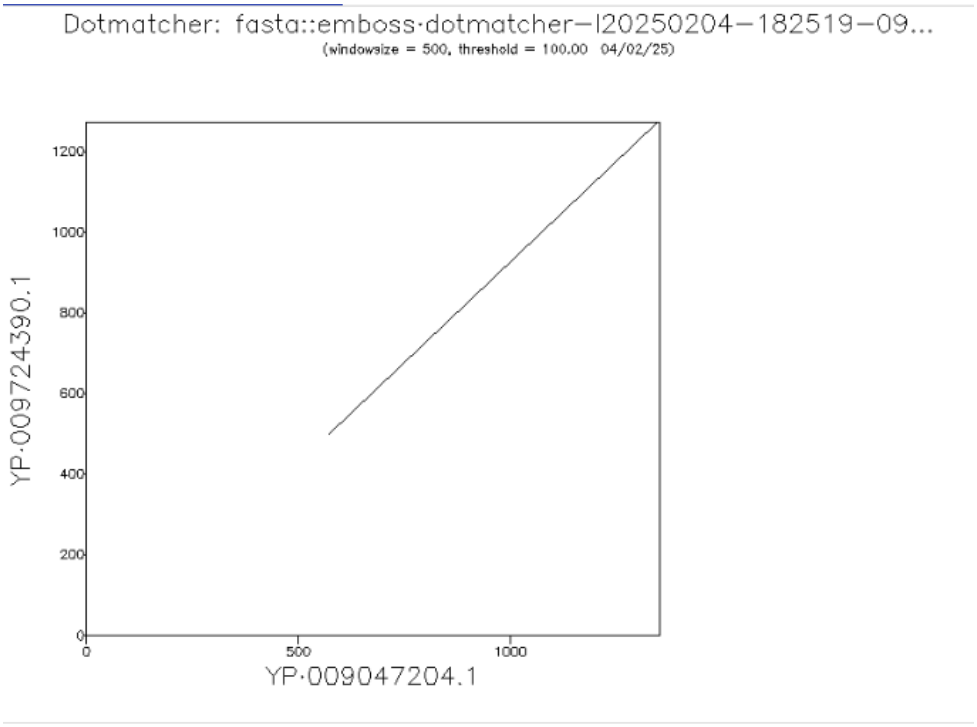


Dotmatcher

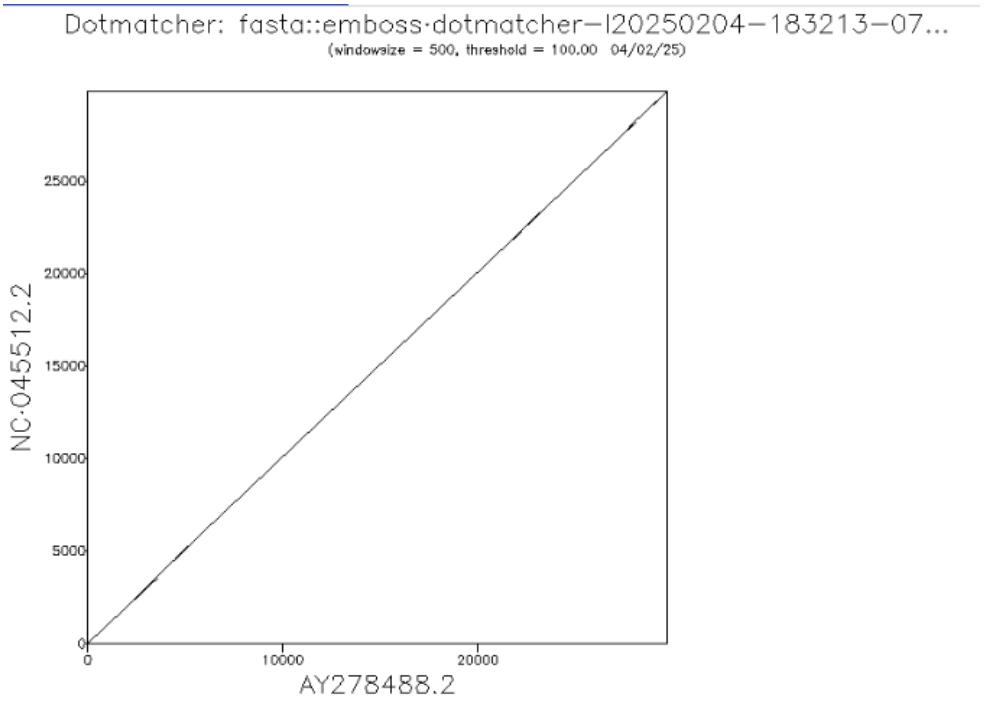
SARS-CoV2 vs SARS-CoV (Protein)



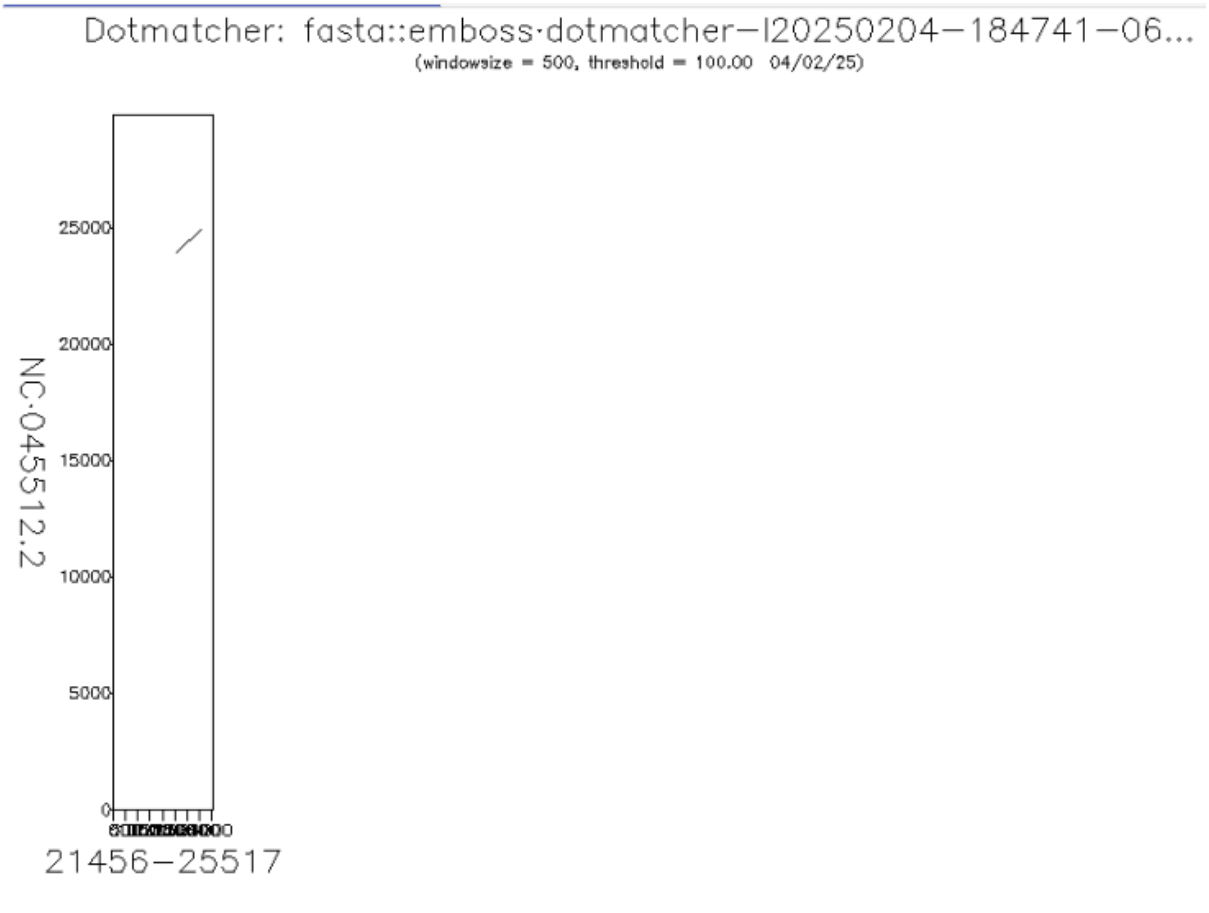
SARS-Cov2 vs MERS-CoV (Protein)



SARS-CoV2 vs SARS-CoV (DNA)



SARS-Cov2 vs MERS-CoV (DNA)



1. SARS-CoV-2 is more similar to SARS-CoV (2003) than to MERS-CoV (2012). The dot plots reveal a strong, continuous diagonal pattern between SARS-CoV-2 and SARS-CoV, indicating

high sequence similarity. In contrast, the plots comparing SARS-CoV-2 and MERS-CoV show fewer matching points and a lack of a continuous diagonal, suggesting a more distant relationship. Dottup results confirm an almost continuous alignment with small gaps between SARS-CoV-2 and SARS-CoV, while no such diagonal is observed for SARS-CoV-2 and MERS-CoV. Similarly, Dotmatcher results show a continuous diagonal alignment for SARS-CoV-2 and SARS-CoV, but only scattered matches toward the latter part of the sequence with MERS-CoV.

2. It is easier to identify similarities using protein sequences than DNA sequences. Protein sequence dot plots are clearer and more continuous, as seen in the 7th plot (Dotmatcher between protein sequences of SARS-CoV-2 and SARS-CoV), which shows a continuous diagonal line compared to the slightly broken diagonal in the 3rd plot (DNA sequences). DNA sequence plots are more fragmented due to silent mutations (nucleotide changes that do not alter amino acids) and codon redundancy, where multiple codons code for the same amino acid, causing variations in DNA without affecting the resulting protein.

3. In Dottup, the K-tuple represents the word size, which is set to 10. In Dotmatcher, the window size is 500, and the threshold value is 100.

Question 2

Part A

(i) At the DNA level for SARS-CoV-2 and SARS-CoV:

- Percentage identity: 73.3%
- Percentage similarity: 73.3%

At the protein level for two proteins:

- Percentage identity: 76.4%
- Percentage similarity: 87.0%

The higher percentage similarity at the protein level is due to the use of the BLOSUM62 matrix, which accounts for not only exact matches but also biochemical similarities between amino acids. In contrast, nucleotide comparisons are binary—bases either match or they don't—resulting in identical percentage values for identity and similarity. This highlights that proteins provide a more detailed measure of sequence similarity, making them more effective for comparing sequences than nucleotides.

(ii) Sequence identity strictly counts the exact matches between the bases of two sequences without accounting for gaps or penalties. For sequences of different lengths, the analysis is based on the shorter sequence.

In contrast, sequence similarity is a broader metric that measures the overall resemblance between two sequences. It considers both matches and mismatches and evaluates how well the sequences align. Unlike identity, similarity uses a more flexible and nuanced approach, making it less rigid in its assessment.

(iii) Global alignment aligns the entire length of both sequences from start to end, making it ideal for sequences of similar lengths. In contrast, local alignment focuses on specific regions of similarity, making it suitable for identifying conserved segments within longer sequences.

Global alignment is performed using Emboss Needle, while local alignment is done using Emboss Water. As shown, the similarity and identity scores from both Needle and Water aligners are very close (73.3% and 72.8%), indicating that global and local alignments yield comparable results in this case.

(iv) EMBOSS WATER (Pairwise Alignment)

SARS-CoV2, SARS-CoV (Protein)

```
########################################
# Program: water
# Rundate: Tue  4 Feb 2025 18:31:47
# Commandline: water
#    -auto
#    -stdout
#    -asequence emboss_water-I20250204-183144-0276-3181322-p1m.asequence
#    -bsequence emboss_water-I20250204-183144-0276-3181322-p1m.bsequence
#    -datafile EBLOSUM62
#    -gapopen 10.0
#    -gapextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: YP_009724390.1
# 2: AAP13441.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1277
# Identity:     975/1277 (76.4%)
# Similarity:  1111/1277 (87.0%)
# Gaps:          26/1277 ( 2.0%)
# Score: 5230.0
#
#
#=======================================
```

SARS-CoV2 vs SARS-CoV (DNA)

```
########################################
# Program: water
# Rundate: Tue  4 Feb 2025 18:40:11
# Commandline: water
#    -auto
#    -stdout
#    -asequence emboss_water-I20250204-183901-0647-27915578-p1m.asequence
#    -bsequence emboss_water-I20250204-183901-0647-27915578-p1m.bsequence
#    -datafile EDNAFULL
#    -gapopen 10.0
#    -gapextend 0.5
#    -aformat3 pair
#    -snucleotide1
#    -snucleotide2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: NC_045512.2
# 2: AY278488.2
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 30425
# Identity:   24132/30425 (79.3%)
# Similarity: 24132/30425 (79.3%)
# Gaps:        1257/30425 ( 4.1%)
# Score: 95783.5
#
#
#=======================================
```

EMBOSS NEEDLE (Pairwise Alignment)

SARS-CoV2, SARS-CoV (Protein)

```
########################################
# Program: needle
# Rundate: Tue  4 Feb 2025 18:39:39
# Commandline: needle
#    -auto
#    -stdout
#    -asequence emboss_needle-I20250204-183936-0008-88772737-p1m.asequence
#    -bsequence emboss_needle-I20250204-183936-0008-88772737-p1m.bsequence
#    -datafile EBLOSUM62
#    -gapopen 10.0
#    -gapextend 0.5
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: YP_009724390.1
# 2: AAP13441.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1277
# Identity:     975/1277 (76.4%)
# Similarity:  1111/1277 (87.0%)
# Gaps:          26/1277 ( 2.0%)
# Score: 5230.0
#
#
#=======================================
```

SARS-CoV2 vs SARS-CoV (DNA)

```
########################################
# Program: needle
# Rundate: Tue  4 Feb 2025 18:45:28
# Commandline: needle
#    -auto
#    -stdout
#    -asequence emboss_needle-I20250204-184434-0664-53518741-p1m.asequence
#    -bsequence emboss_needle-I20250204-184434-0664-53518741-p1m.bsequence
#    -datafile EDNAFULL
#    -gapopen 10.0
#    -gapextend 0.5
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -snucleotide1
#    -snucleotide2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: NC_045512.2
# 2: AY278488.2
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 30481
# Identity:   24125/30481 (79.1%)
# Similarity: 24125/30481 (79.1%)
# Gaps:        1334/30481 ( 4.4%)
# Score: 95783.5
#
#
#=======================================
```

For protein alignments:

- **Matrix**: BLOSUM62
- **Gap penalty**: 10
- **Extend penalty**: 0.5

For DNA alignments:

- **Matrix**: DNAfull
- **Gap penalty**: 10
- **Extend penalty**: 0.5

Part B

EMBOSS WATER (Pairwise Alignment)

SARS-Cov2 vs MERS-CoV (Protein)

```
########################################
# Program: water
# Rundate: Tue  4 Feb 2025 18:46:05
# Commandline: water
#    -auto
#    -stdout
#    -asequence emboss_water-I20250204-184558-0828-55458133-p1m.asequence
#    -bsequence emboss_water-I20250204-184558-0828-55458133-p1m.bsequence
#    -datafile EBLOSUM62
#    -gapopen 10.0
#    -gapextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: YP_009724390.1
# 2: YP_009047204.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1440
# Identity:     433/1440 (30.1%)
# Similarity:   655/1440 (45.5%)
# Gaps:         276/1440 (19.2%)
# Score: 1568.5
#
#
#=======================================
```

SARS-Cov2 vs MERS-CoV (DNA)

```
########################################
# Program: water
# Rundate: Tue  4 Feb 2025 18:50:58
# Commandline: water
#    -auto
#    -stdout
#    -asequence emboss_water-I20250204-185049-0200-91566408-p1m.asequence
#    -bsequence emboss_water-I20250204-185049-0200-91566408-p1m.bsequence
#    -datafile EDNAFULL
#    -gapopen 10.0
#    -gapextend 0.5
#    -aformat3 pair
#    -snucleotide1
#    -snucleotide2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: NC_045512.2
# 2: 21456-25517
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 6807
# Identity:    2779/6807 (40.8%)
# Similarity:  2779/6807 (40.8%)
# Gaps:        3258/6807 (47.9%)
# Score: 4835.0
#
#
#=======================================
```

EMBOSS NEEDLE (Pairwise Alignment)

SARS-Cov2 vs MERS-CoV (Protein)

```
########################################
# Program: needle
# Rundate: Tue  4 Feb 2025 18:50:54
# Commandline: needle
#     -auto
#     -stdout
#     -asequence emboss_needle-I20250204-185049-0136-56216348-p1m.asequence
#     -bsequence emboss_needle-I20250204-185049-0136-56216348-p1m.bsequence
#     -datafile EBLOSUM62
#     -gapopen 10.0
#     -gapextend 0.5
#     -endopen 10.0
#     -endextend 0.5
#     -aformat3 pair
#     -sprotein1
#     -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: YP_009724390.1
# 2: YP_009047204.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1454
# Identity:      434/1454 (29.8%)
# Similarity:    658/1454 (45.3%)
# Gaps:          282/1454 (19.4%)
# Score: 1564.0
#
#
#=======================================
```

SARS-Cov2 vs MERS-CoV (DNA)

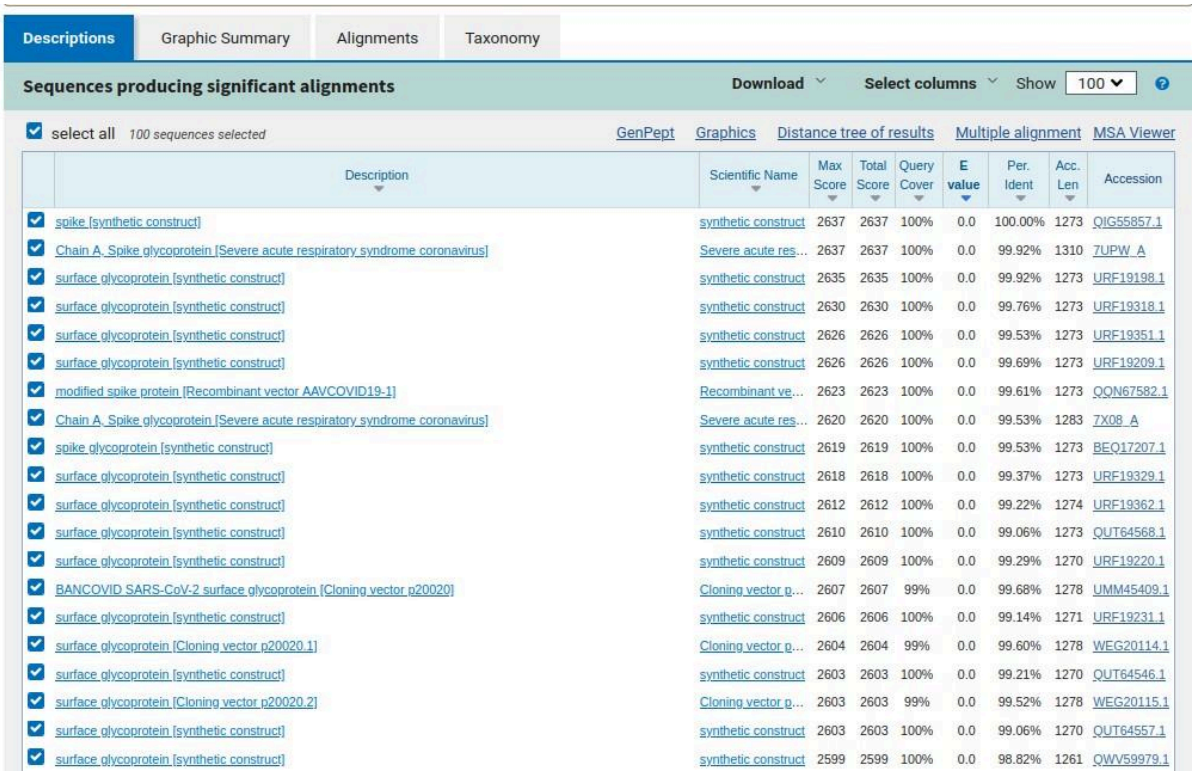
```
########################################
# Program: needle
# Rundate: Tue  4 Feb 2025 20:02:42
# Commandline: needle
#    -auto
#    -asequence /var/lib/emboss-explorer/output/135364/.asequence
#    -bsequence /var/lib/emboss-explorer/output/135364/.bsequence
#    -noendweight
#    -brief
#    -outfile outfile
#    -aformat3 srspair
# Align_format: srspair
# Report_file: outfile
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: 21456-25517
# 2: 21563-25384
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 4867
# Identity:    2330/4867 (47.9%)
# Similarity:  2330/4867 (47.9%)
# Gaps:        1850/4867 (38.0%)
# Score: 4414.5
#
#
#=======================================
```

(i) Proteins with a percentage similarity exceeding 30% are considered homologous. According to the Emboss Water results, the similarity percentage between SARS-CoV-2 and MERS-CoV is 51%. Thus, it can be concluded that SARS-CoV-2 and MERS-CoV are homologous.

(ii) The conclusions are based on protein sequence alignments performed using Emboss Water.

Question 3

BLAST Protein Search of SARS-Cov2



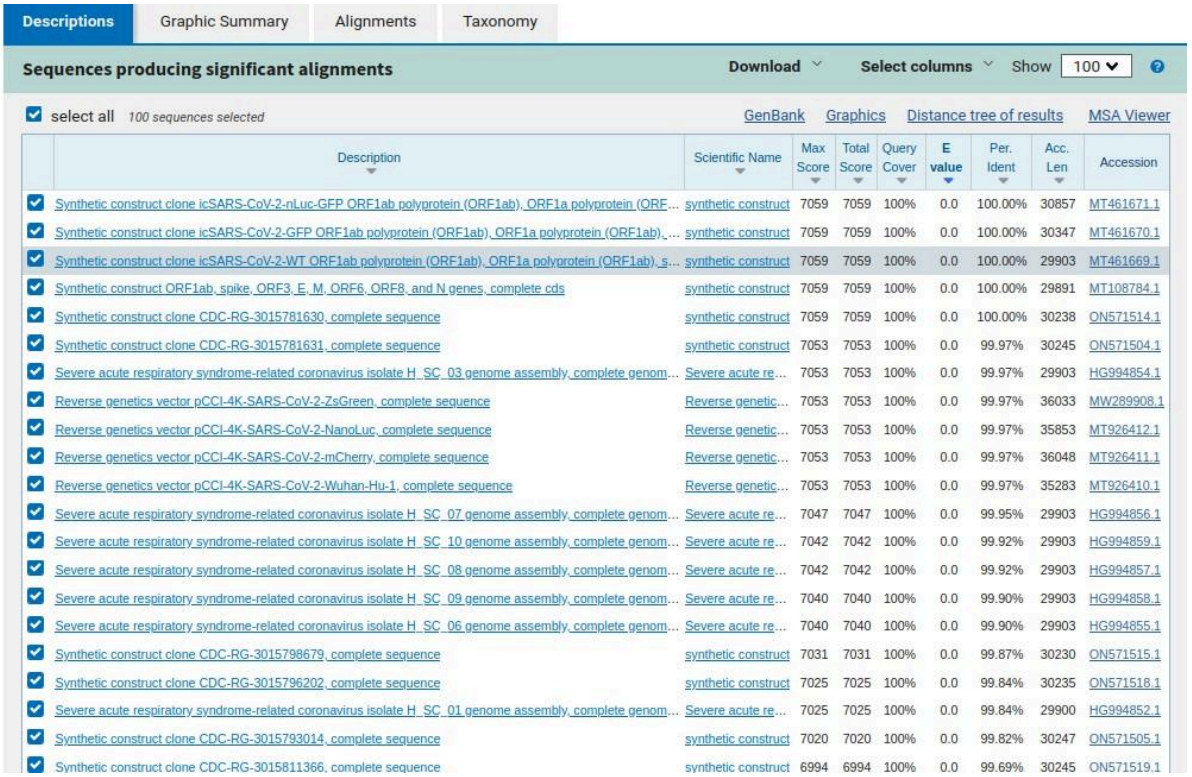
Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download | Select columns | Show 100

select all *100 sequences selected* — GenPept | Graphics | Distance tree of results | Multiple alignment | MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
spike [synthetic construct]	synthetic construct	2637	2637	100%	0.0	100.00%	1273	QIG55857.1
Chain A, Spike glycoprotein [Severe acute respiratory syndrome coronavirus]	Severe acute res...	2637	2637	100%	0.0	99.92%	1310	7UPW_A
surface glycoprotein [synthetic construct]	synthetic construct	2635	2635	100%	0.0	99.92%	1273	URF19198.1
surface glycoprotein [synthetic construct]	synthetic construct	2630	2630	100%	0.0	99.76%	1273	URF19318.1
surface glycoprotein [synthetic construct]	synthetic construct	2626	2626	100%	0.0	99.53%	1273	URF19351.1
surface glycoprotein [synthetic construct]	synthetic construct	2626	2626	100%	0.0	99.69%	1273	URF19209.1
modified spike protein [Recombinant vector AAVCOVID19-1]	Recombinant ve...	2623	2623	100%	0.0	99.61%	1273	QQN67582.1
Chain A, Spike glycoprotein [Severe acute respiratory syndrome coronavirus]	Severe acute res...	2620	2620	100%	0.0	99.53%	1283	7X08_A
spike glycoprotein [synthetic construct]	synthetic construct	2619	2619	100%	0.0	99.53%	1273	BEQ17207.1
surface glycoprotein [synthetic construct]	synthetic construct	2618	2618	100%	0.0	99.37%	1273	URF19329.1
surface glycoprotein [synthetic construct]	synthetic construct	2612	2612	100%	0.0	99.22%	1274	URF19362.1
surface glycoprotein [synthetic construct]	synthetic construct	2610	2610	100%	0.0	99.06%	1273	QUT64568.1
surface glycoprotein [synthetic construct]	synthetic construct	2609	2609	100%	0.0	99.29%	1270	URF19220.1
BANCOVID SARS-CoV-2 surface glycoprotein [Cloning vector p20020]	Cloning vector p...	2607	2607	99%	0.0	99.68%	1278	UMM45409.1
surface glycoprotein [synthetic construct]	synthetic construct	2606	2606	100%	0.0	99.14%	1271	URF19231.1
surface glycoprotein [Cloning vector p20020.1]	Cloning vector p...	2604	2604	99%	0.0	99.60%	1278	WEG20114.1
surface glycoprotein [synthetic construct]	synthetic construct	2603	2603	100%	0.0	99.21%	1270	QUT64546.1
surface glycoprotein [Cloning vector p20020.2]	Cloning vector p...	2603	2603	99%	0.0	99.52%	1278	WEG20115.1
surface glycoprotein [synthetic construct]	synthetic construct	2603	2603	100%	0.0	99.06%	1270	QUT64557.1
surface glycoprotein [synthetic construct]	synthetic construct	2599	2599	100%	0.0	98.82%	1261	QWV59979.1

BLAST Nucleotide Search of SARS-Cov2



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download | Select columns | Show 100

select all *100 sequences selected* — GenBank | Graphics | Distance tree of results | MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Synthetic construct clone icSARS-CoV-2-nLuc-GFP ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF...	synthetic construct	7059	7059	100%	0.0	100.00%	30857	MT461671.1
Synthetic construct clone icSARS-CoV-2-GFP ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), ...	synthetic construct	7059	7059	100%	0.0	100.00%	30347	MT461670.1
Synthetic construct clone icSARS-CoV-2-WT ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), s...	synthetic construct	7059	7059	100%	0.0	100.00%	29903	MT461669.1
Synthetic construct ORF1ab, spike, ORF3, E, M, ORF6, ORF8, and N genes, complete cds	synthetic construct	7059	7059	100%	0.0	100.00%	29891	MT108784.1
Synthetic construct clone CDC-RG-3015781630, complete sequence	synthetic construct	7059	7059	100%	0.0	100.00%	30238	ON571514.1
Synthetic construct clone CDC-RG-3015781631, complete sequence	synthetic construct	7053	7053	100%	0.0	99.97%	30245	ON571504.1
Severe acute respiratory syndrome-related coronavirus isolate H_SC_03 genome assembly, complete genom...	Severe acute re...	7053	7053	100%	0.0	99.97%	29903	HG994854.1
Reverse genetics vector pCCI-4K-SARS-CoV-2-ZsGreen, complete sequence	Reverse genetic...	7053	7053	100%	0.0	99.97%	36033	MW289908.1
Reverse genetics vector pCCI-4K-SARS-CoV-2-NanoLuc, complete sequence	Reverse genetic...	7053	7053	100%	0.0	99.97%	35853	MT926412.1
Reverse genetics vector pCCI-4K-SARS-CoV-2-mCherry, complete sequence	Reverse genetic...	7053	7053	100%	0.0	99.97%	36048	MT926411.1
Reverse genetics vector pCCI-4K-SARS-CoV-2-Wuhan-Hu-1, complete sequence	Reverse genetic...	7053	7053	100%	0.0	99.97%	35283	MT926410.1
Severe acute respiratory syndrome-related coronavirus isolate H_SC_07 genome assembly, complete genom...	Severe acute re...	7047	7047	100%	0.0	99.95%	29903	HG994856.1
Severe acute respiratory syndrome-related coronavirus isolate H_SC_10 genome assembly, complete genom...	Severe acute re...	7042	7042	100%	0.0	99.92%	29903	HG994859.1
Severe acute respiratory syndrome-related coronavirus isolate H_SC_08 genome assembly, complete genom...	Severe acute re...	7042	7042	100%	0.0	99.92%	29903	HG994857.1
Severe acute respiratory syndrome-related coronavirus isolate H_SC_09 genome assembly, complete genom...	Severe acute re...	7040	7040	100%	0.0	99.90%	29903	HG994858.1
Severe acute respiratory syndrome-related coronavirus isolate H_SC_06 genome assembly, complete genom...	Severe acute re...	7040	7040	100%	0.0	99.90%	29903	HG994855.1
Synthetic construct clone CDC-RG-3015798679, complete sequence	synthetic construct	7031	7031	100%	0.0	99.87%	30230	ON571515.1
Synthetic construct clone CDC-RG-3015796202, complete sequence	synthetic construct	7025	7025	100%	0.0	99.84%	30235	ON571518.1
Severe acute respiratory syndrome-related coronavirus isolate H_SC_01 genome assembly, complete genom...	Severe acute re...	7025	7025	100%	0.0	99.84%	29900	HG994852.1
Synthetic construct clone CDC-RG-3015793014, complete sequence	synthetic construct	7020	7020	100%	0.0	99.82%	30247	ON571505.1
Synthetic construct clone CDC-RG-3015811366, complete sequence	synthetic construct	6994	6994	100%	0.0	99.69%	30245	ON571519.1

(i) Based on the BLASTp (protein) search, the closest homolog is the Spike glycoprotein (Chain A) from Severe Acute Respiratory Syndrome Coronavirus. Similarly, the BLASTn (nucleotide) search identifies the closest homolog as the genome assembly of the Severe Acute Respiratory Syndrome-related coronavirus isolate H_SC_03.

(ii) Protein Alignment:

- Percentage Identity: 99.92%
- Length: 1310
- Score: 2637
- E-value: 0.0

Nucleotide Alignment:

- Percentage Identity: 99.97%
- Length: 30,245
- Score: 7053
- E-value: 0.0

Both alignments show high identity with highly significant E-values (0.0).

(iii) The **Spike glycoprotein** of **SARS-CoV** stands out as the top hit in the BLAST search results, indicating it as the closest homolog to the query sequence. The percentage identity and similarity values obtained from this search align closely with those generated by the **'water'** alignment tool, further supporting the similarity between the sequences. The images below illustrate these alignment results.

Nucleotide

```
########################################
# Program: water
# Rundate: Tue  4 Feb 2025 16:56:54
# Commandline: water
#    -auto
#    -stdout
#    -asequence /var/lib/emboss-explorer/output/493580/.asequence
#    -bsequence /var/lib/emboss-explorer/output/493580/.bsequence
#    -datafile EDNAFULL
#    -gapopen 10.0
#    -gapextend 0.5
#    -aformat3 pair
#    -snucleotide1
#    -snucleotide2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: 21563-25384
# 2: HG994854.1
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 3822
# Identity:      3821/3822 (100.0%)
# Similarity:    3821/3822 (100.0%)
# Gaps:             2/3822 ( 1.0%)
# Score: 19101.0
#
#
#=======================================
```

Protein

```
########################################
# Program: water
# Rundate: Tue  4 Feb 2025 16:56:54
# Commandline: water
#    -auto
#    -stdout
#    -asequence emboss_water-120240405-175911-0682-16077024-pim.asequence
#    -bsequence emboss_water-120240405-175911-0682-16077024-pim.bsequence
#    -datafile EBLOSUM62
#    -gapopen 10.0
#    -gapextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: YP_009724390.1
# 2: 7UPWA
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1273
# Identity:     1272/1273 (99.9%)
# Similarity:   1272/1273 (99.9%)
# Gaps:            1/1273 ( 0.9%)
# Score: 6715.0
#
#
```

The consistency between the results highlights the effectiveness and precision of the **BLAST** search algorithm in sequence alignment. Unlike **Water**, where sequences are directly input for comparison, **BLAST** performs a search across the entire database. The agreement in the results reinforces the reliability and accuracy of the **BLAST** search method in identifying homologous sequences.

(iv)

When searching for **SARS-CoV-2** in **BLAST**, a hit for **bat SARS coronavirus** is found. Below are the alignment metrics for the **spike glycoprotein** of **SARS-CoV-2** and the **spike glycoprotein** of **bat SARS coronavirus**:

- **Percentage Similarity**: NA
- **Percentage Identity**: 98.43%
- **Length of Alignment**: 1269
- **Score**: 2593
- **E-value**: 0.0

Question 4

Database Statistics:

- **UniProtKB (Protein Database)**:
  - Entries: 248,805,733
  - Total amino acids: 87,574,368,369
  - Reference
- **GenBank (Nucleotide Database)**:
  - Total bases: 2,570,711,588,044
  - Total sequences: 249,060,436
  - Reference

Dynamic Programming (DP) Search:

- **Search Complexity**: O(m * n), where **m** is the query sequence length and **n** is the target sequence length.
- **Processing speed**: 10 million matrix cells per second, implying a time of 10^-7 seconds per cell.

(i) Search Time Estimates:

- **UniProt (Protein Database)**:
  - **Bases**: 1000
  - **Amino acids**: ~333 (since each amino acid is represented by 3 bases)
  - **Total cells**: ~2.92×10^{13}
  - **Calculation Time**: ~2.92×10^{6} seconds
- **GenBank (Nucleotide Database)**:
  - **Total cells**: 2,570,711,588,044,000
  - **Calculation Time**: ~257,071,158.8044 seconds

(ii) Space Complexity for Chromosomes:

- **Human Chromosome 1**:
  - **Size**: 249 Mbp
  - **Query Sequence Length**: 1000
  - **Space Complexity**: ~2.49×10^{11} bits
- **Mouse Chromosome 1**:
  - **Size**: 195 Mbp
  - **Query Sequence Length**: 1000
  - **Space Complexity**: ~1.95×10^{11} bits