

Sequence Alignment and Phylogenetic Tree Exercise

1. Study Protein: SARS CoV-2 Nsp5 (Apo protein PDB ID: [6yb7](#))

Comparator Proteins: Study Protein homologs from the following viruses:

- SARS-CoV (severe acute respiratory syndrome coronavirus);
- MERS (Middle East respiratory syndrome coronavirus);
- PEDV (Porcine Epidemic Diarrhea Virus);
- TGEV (Porcine Transmissible Gastroenteritis Virus);
- HKU4 (Tylonycteris bat coronavirus HKU4);
- A59 (Mouse hepatitis virus A59);
- HKU1 (Human coronavirus HKU1)
- FIPV (Feline infectious peritonitis virus); and
- NL63 (Human coronavirus NL63).

Hint: Exercises 1.1-1.3 (more involved), Exercises 1.4-1.5 (more involved), and Exercise 1.6 (quick only one person required) can be carried out in parallel if you want to divide the effort and conquer the tasks more quickly.

Exercise Deliverables/Instructions:

1.1 Generate Amino Acid Sequence Alignment of SARS-CoV-2 Study Protein plus homologs using Clustal Omega.

Introductory Documentation: <https://en.wikipedia.org/wiki/Clustal>

Input File: [All_Unique_Nsp5_in_PDB.fasta.txt](#)

Output Files:

All_Unique_Nsp5_in_PDB.clustal_num multiple sequence alignment file produced by the Clustal Omega Website; and

All_Unique_Nsp5_in_PDB.msa.docx MS Word file using Courier non-proportional font flagging

active site residues in the first sequence in the alignment (bold).

Instructions: Go to the homepage of the Clustal Omega website at EMBL-EBI (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

Click STEP 1 Upload All_Unique_Nsp5_in_PDB.fasta.txt as the input sequence file

Click STEP 2 Set Your Parameters to set ORDER to “input” under “More Options”

Click STEP 3 Submit

You will receive the following result after a few minutes:

The word PREFIX below refers to the string clustalo-...-p1m that appears before the “.”.

Click on PREFIX-p1m.clustal_num to download file (Right click/Save As).

Rename as All_Unique_Nsp5_in_PDB.clustal_num

Click on the Alignments tab to view the multiple sequence alignment file produced by the Clustal Omega Website.

See guide below for interpreting the Multiple Sequence Alignment Output.

Click Download Alignment File Tab to generate the .txt file version of the multiple sequence alignment (Right click/Save As). Keep the file extension as .txt

Open with MS Word, use Courier non-proportional font and annotate active site residues on the first line of the sequence alignment with bold: His41, Cys145, Glu166.

Save as All_Unique_Nsp5_in_PDB.msa.docx

Final Steps: Review files All_Unique_Nsp5_in_PDB.clustal_num and All_Unique_Nsp5_in_PDB.msa.docx

1.2 Phylogenetic Tree of SARS CoV-2 Study Protein plus homologs.

Input: PREFIX.ph file produced by the Clustal Omega website

Output: Screen shot of Cladogram Phylogenetic Tree (All_Unique_Nsp5_in_PDB.ph_tree.png); and PREFIX.ph file renamed as All_Unique_Nsp5_in_PDB.ph

Instructions:

- Click on Result Summary Tab

- Click on Phylogenetic Tree Tab and take screen shot of tree (use Cladogram option)
save screen shot as All_Unique_Nsp5_in_PDB.ph_tree.png
- Click on Result Summary Tab
- Click on file name PREFIX.ph to download file for future use with tree drawing software.
- Rename file as All_Unique_Nsp5_in_PDB.ph

Final Steps:

Upload All_Unique_Nsp5_in_PDB.ph_tree.png and
All_Unique_Nsp5_in_PDB.ph to your group results page.

1.3 Amino Acid Sequence Alignment Percent Identity Matrix for SARS CoV-2 Study Protein plus homologs.

Input File: PREFIX.pim file produced by the Clustal Omega website

Output File: MS Word file NxN table, using data from the PREFIX.pim file
(All_Unique_Nsp5_in_PDB.pct_id.docx)

Instructions:

- Click on Result Summary Tab
- Click on Percent Identity Matrix file name Prefix.pim to download the .txt file needed to generate the MS Word file NxN table of sequence identities.
- Use Courier non-proportional font and save as All_Unique_Nsp5_in_PDB.pct_id.docx

Final Steps:

Upload All_Unique_Nsp5_in_PDB.pct_id.docx to your group results page.

1.4 GISAID Amino Acid Sequence Alignment of SARS CoV-2 Study Protein using Clustal Omega.

Introductory Documentation: <https://www.gisaid.org>

Input: [NSP5_GISAID_20200602.fasta.txt](#)

Output: PREFIX.clustal_num multiple sequence alignment file produced by the Clustal Omega Website;
and

MS Word file NSP5_GISAID_20200602.msa.docx

Instructions:

- Go to the homepage of the Clustal Omega website at EMBL-EBI (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).
- Same instructions as for No. 1.1 using NSP5_GISAID_20200602.fasta.txt as the input file.

Final Steps:

Review NSP5_GISAID_20200602.clustal_num and NSP5_GISAID_20200602.msa.docx

1.5 Phylogenetic Tree of GISAID Amino Acid Sequence Alignment of SARS CoV-2 Study Protein.

Input: PREFIX.ph file produced by the Clustal Omega website

Output: Screen shot of Cladogram Phylogenetic Tree, and PREFIX.ph file

Instructions: Same instructions as for No. 1.2.

Final Steps:

Review NSP5_GISAID_20200602.ph_tree.png and NSP5_GISAID_20200602.ph

1.6 Calculate molecular weight and pI (isoelectric point) for SARS-CoV-2 Study Protein using ExPASy website at the Swiss Institute for Bioinformatics.

Introductory Documentation: <https://www.expasy.org/>;

https://web.expasy.org/compute_pi/pi_tool-doc.html

Input: [6yb7_A.fasta.txt](#)

Output: MS Word file with Average molecular weight and pI Filename=6yb7_A.PI&MW.doc

Instructions:

- Go to the eXPASy website page for calculating MW and pI https://web.expasy.org/compute_pi/
- Paste FASTA sequence into input window.
- Select RESOLUTION Average.
- Click here to compute pI/MW.

covid19-bootcamp.rcsb.org

Record results in an MS Word file with pl=# and MW=#, Using filename 6yb7_A.PI&MW.doc