**Final Project**

Each one of you are given a sequence from SAR-CoV-2. Perform the following tasks and submit the results. You need to submit snapshots of all the outputs. Save the results in a folder with your Roll no. and upload.

1. Perform database search of your sequence using BLAST program. [Hint: make use of available options on the search page to obtain hits from other organisms].
2. Which BLAST program did you consider for the search?
3. Which database was it searched against and why?
4. Based on the search what can you say about your sequence?
5. Report the organism name, sequence coverage, %age identity and e-value of the top non-SARS-CoV-2 sequence.
6. Find 5 hits to your sequence with identity in the range ~ 95% - 25% - list the accession id, organism name, query coverage, percentage identity, and e-value for the selected 5 sequences.
7. Give the dotplots of the query with the 1st and 5th sequences from the list in (iv).
8. Give the pairwise alignment by Needleman-Wunsch algorithm – is it similar to the alignment given by BLAST?
9. Perform multiple sequence alignment of your query sequence with the 5 ‘hits’ selected from the database search in step-1 using CLUSTALΩ. Submit the alignment and score.
10. Construct phylogenetic tree using the MSA obtained in step-2. You may choose a method of your choice from the PHYLIP suite of programs.
11. Which method was used to construct the tree? Give reasons for choosing this method.
12. Give bootstrap values.
13. Submit the tree and inferences drawn from this tree about the relatedness of the species considered?
14. Give the name and accession id of your protein. Give a brief description of your protein and its function (based on the latest literature).