

Tutorial Assignment - II (MSA, Phylogeny)

Deadline: 2355hrs, 10th April 2022

SARS-CoV-2 is a zoonotic virus that jumped from animal species to humans. It is known that coronaviruses utilize one of its proteins, called Spike glycoprotein, to infect the host cells. The cellular entry of the coronavirus depends on binding between the viral Spike protein receptor-binding domain (RBD) and the angiotensin converting enzyme 2 (ACE2) target cell receptor. Hence it is important to understand how this protein has evolved for drug target.

To understand how this protein has evolved to modify its host preference, perform multiple sequence alignment of the orthologs of these proteins.

Further, this virus is reported to be closely related to bat and pangolin coronaviruses, and it is hypothesised that perhaps it is a blend of bat and pangolin viruses that emerged by a process, called recombination, in a bat, pangolin or another species. What do you infer from your analysis?

1. Download gene and protein sequences of Spike glycoprotein from six species: SARS-CoV-2, SARS-CoV, MERS-CoV, Bat coronavirus RaTG13, Bat-CoV, and P-CoV. [Hint: Go to NCBI Nucleotide search page, get the genomic sequences, from which search for spike protein and download CDS and protein sequences of spike glycoprotein in FASTA format. Generate a single multi-FASTA file for performing multiple-sequence alignment].
 - (a) Perform multiple sequence alignment and based on percentage identity, identify its closest relative. Compare the results with gene and protein sequences and give the % identity matrix in the two cases.
 - (b) From these results can you infer the possible source of origin of SARS-CoV-2 and that of MERS-CoV.
2. Construct phylogenetic tree using the multiple sequence alignment files obtained in the previous Q. and analyse your results using Phylip. Use one of each parsimony, distance-based, and maximum likelihood methods and compare the three trees obtained (with and without bootstrapping). Submit the trees and summarize your observations.
 - (a) Are the trees obtained by different methods in agreement, topology-wise?
 - (b) Do you observe any difference with or without bootstrapping? What information does bootstrapping provide?
 - (c) Are your inferences in agreement with those in Q.1(b) above.