

Question 1

- Following are the results for the MSA with Clustal Flag:
 - Protein_msa_sequence:
 - https://www.ebi.ac.uk/Tools/services/rest/clustalo/result/clustalo-I20200415-064910-0241-46268859-p1m/aln-clustal_num
 - Gene MSA sequence:
 - https://www.ebi.ac.uk/Tools/services/rest/clustalo/result/clustalo-I20200415-065439-0099-28478834-p2m/aln-clustal_num
- Following are the results for the MSA with Phylip flag with inputs as the CLUSTAL files:
 - Protein Phylip Sequence:
 - https://www.ebi.ac.uk/Tools/services/rest/clustalo/result/clustalo-I20200415-085230-0050-12856455-p1m/aln-clustal_num
 - Gene Phylip Sequence:
 - <https://www.ebi.ac.uk/Tools/services/rest/clustalo/result/clustalo-I20200415-073710-0003-82176554-p1m/aln-phylip>
- Percent Identity matrix is as follows:

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
1: ABS86623.1      100.00   16.38   16.67   21.57   21.57   21.57
2: AZF86130.1      16.38   100.00   25.66   25.63   25.72   25.71
3: YP_009047204.1  16.67   25.66  100.00   31.31   31.63   31.53
4: QI054048.1      21.57   25.63   31.31  100.00   92.91   92.43
5: QHR63300.2       21.57   25.72   31.63   92.91  100.00   97.64
6: YP_009724390.1  21.57   25.71   31.53   92.43   97.64  100.00
```

-
- These ID's signify the different Proteins
- All the relevant files have been included alongside the folder in the submission folder

Question 1 : B

- According to the Percentage Identity Matrix, SARS CoV 2 is the most identical to Bat coronavirus.
 - The possible source of SARS-CoV-2 are bats.
- MERS CoV is not identical to other viruses in the list, so its source is difficult to determine.

Question 2 :

- a) Yes, the trees do seem to be in agreement. They are identical topology wise. This applies to the trees generated without bootstrapping. Since the trees are

unrooted, anyone of the viruses can be selected as a root and the trees on being redrawn will have the same structure. Hence ,the agreement.

- b) There are very minor details observed with and without bootstrapping.
 - i) By bootstrapping, we can get some extra details like a percentage on the tree, and better graphs.
 - ii) With a larger dataset, we got a better generalised structure of the tree.
 - c) Yes, the observations made are in line with the observations from the trees obtained in this question. Both SARS COV2 and bat coronavirus have the closest common ancestor, unlike others. So source=bat. MERS is the farthest, so cannot find the source.
- All the relevant outfile for single and multi-data sets have been included alongside the folder in the submission folder.