

# Science 2 - TA -2

Naimeesh Narayan Tiwari

2020101074

## ▼ Question 1

DNA:

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

1: MERS-COV	100.00	45.82	46.15	46.00	46.74	46.62
2: batcov	45.82	100.00	70.16	68.11	68.42	68.24
3: SARS-COV	46.15	70.16	100.00	73.56	73.38	73.33
4: P-COV	46.00	68.11	73.56	100.00	83.36	82.87
5: SARS-COV2	46.74	68.42	73.38	83.36	100.00	93.12
6: RaTG13	46.62	68.24	73.33	82.87	93.12	100.00

Protein:

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

1: MERS-COV	100.00	31.69	32.28	31.99	32.30	32.32
2: batcov	31.69	100.00	75.85	71.61	72.18	72.28
3: SARS-COV	32.28	75.85	100.00	77.38	77.22	77.82
4: P-COV	31.99	71.61	77.38	100.00	92.36	93.05
5: SARS-COV2	32.30	72.18	77.22	92.36	100.00	97.64
6: RaTG13	32.32	72.28	77.82	93.05	97.64	100.00

Steps:

- get single multi-FASTA file for the 6 species

- perform multiple sequence alignment > Clustal Omega with “ClustalW with character count” as output format
- get result summary
- get percentage identity matrix

#### Closest relatives

- MERS-CoV : SARS-COV2
- batcov: SARS-COV
- SARS-COV: P-COV
- P-COV: SARS-COV2, RaTG13
- SARS-COV2: RaTG13
- RaTG13: SARS-COV2

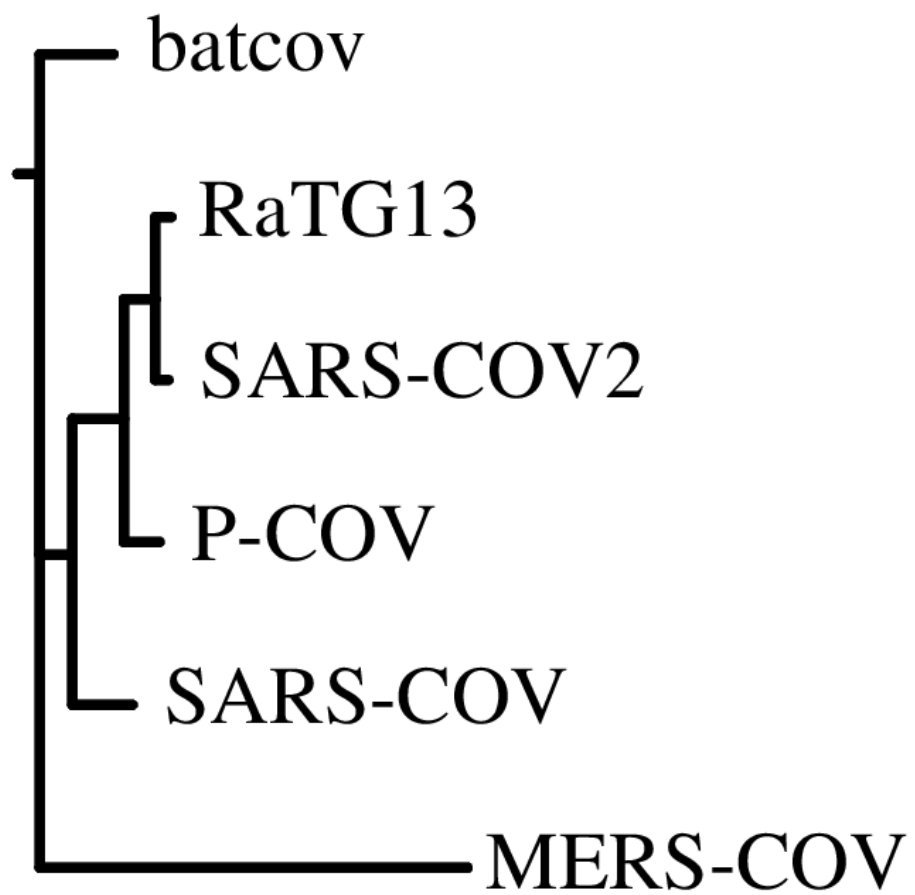
#### Q1>b>

- SARS-Cov2: RaTG13 bat coronavirus
- MERS-COV : cant say as all have nearly same percentage identity

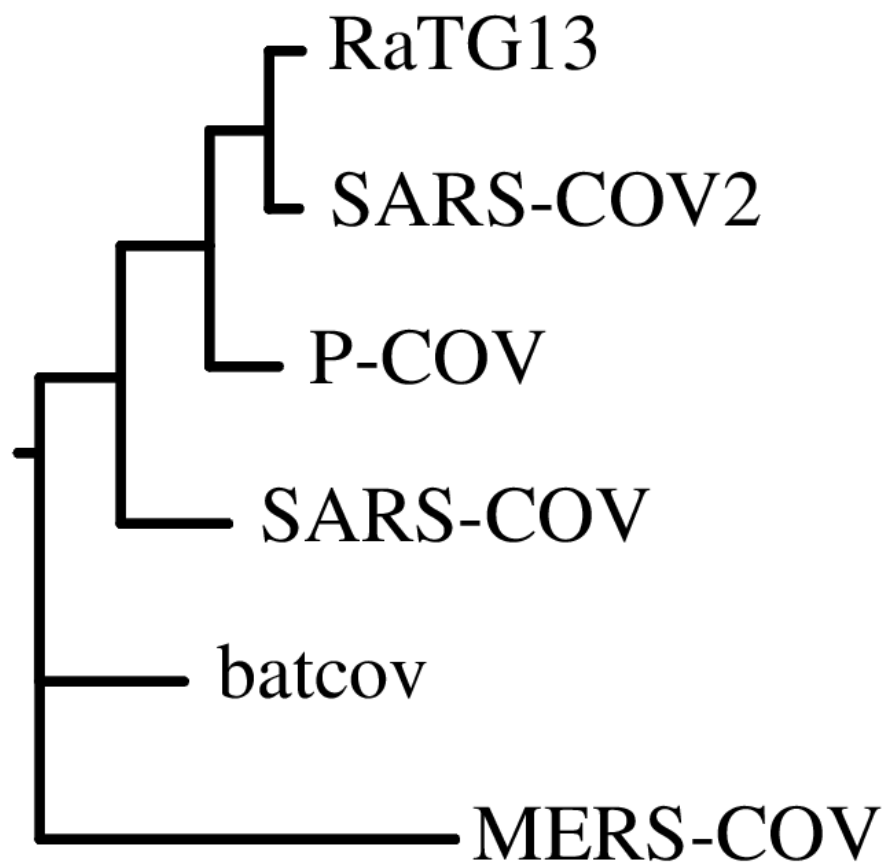
#### ▼ Question 2

##### ▼ Part a>

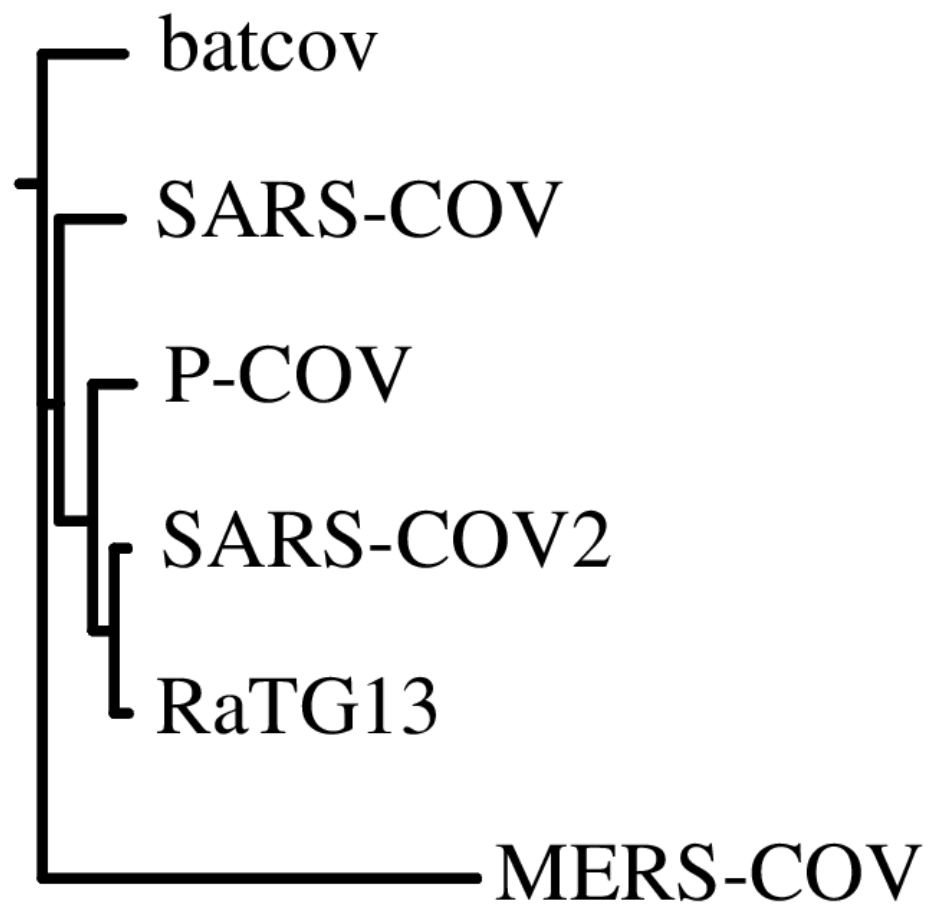
##### ▼ dnaml



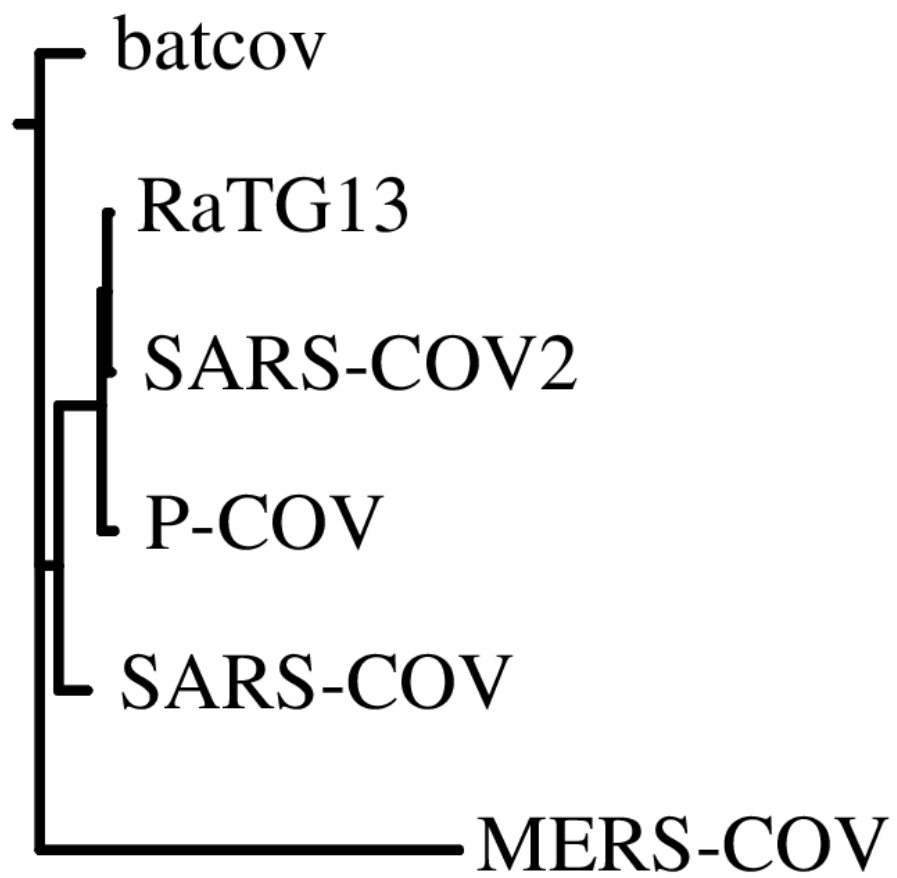
▼ dnapars



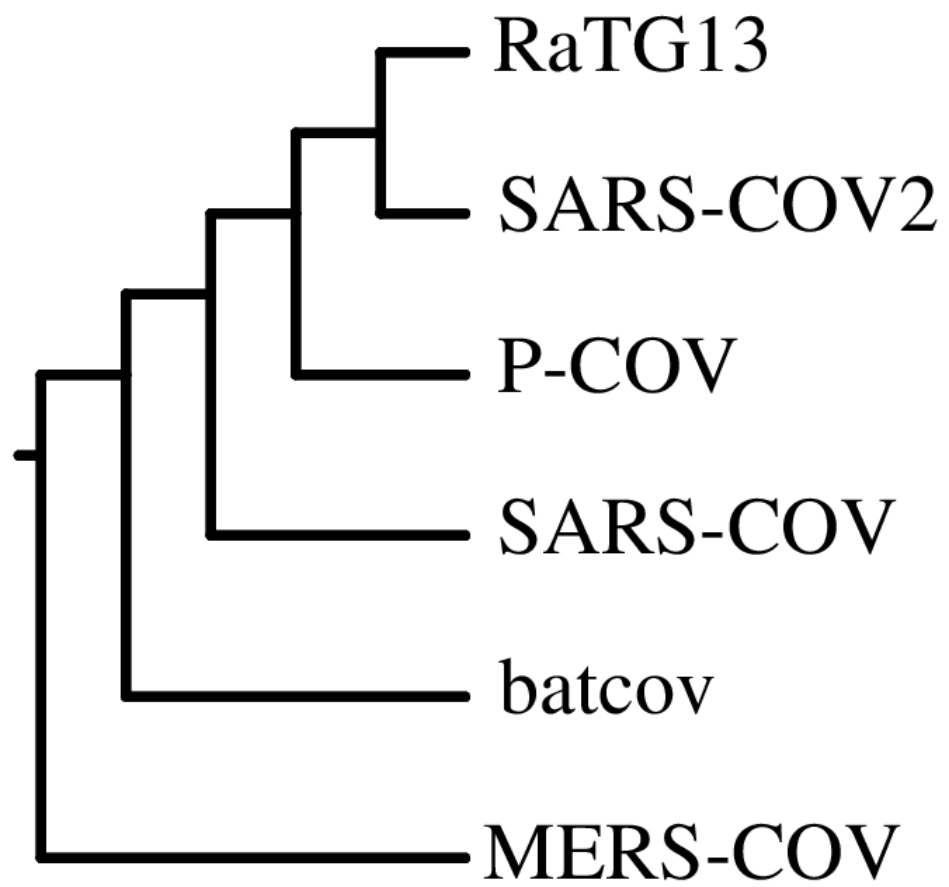
▼ dnadist



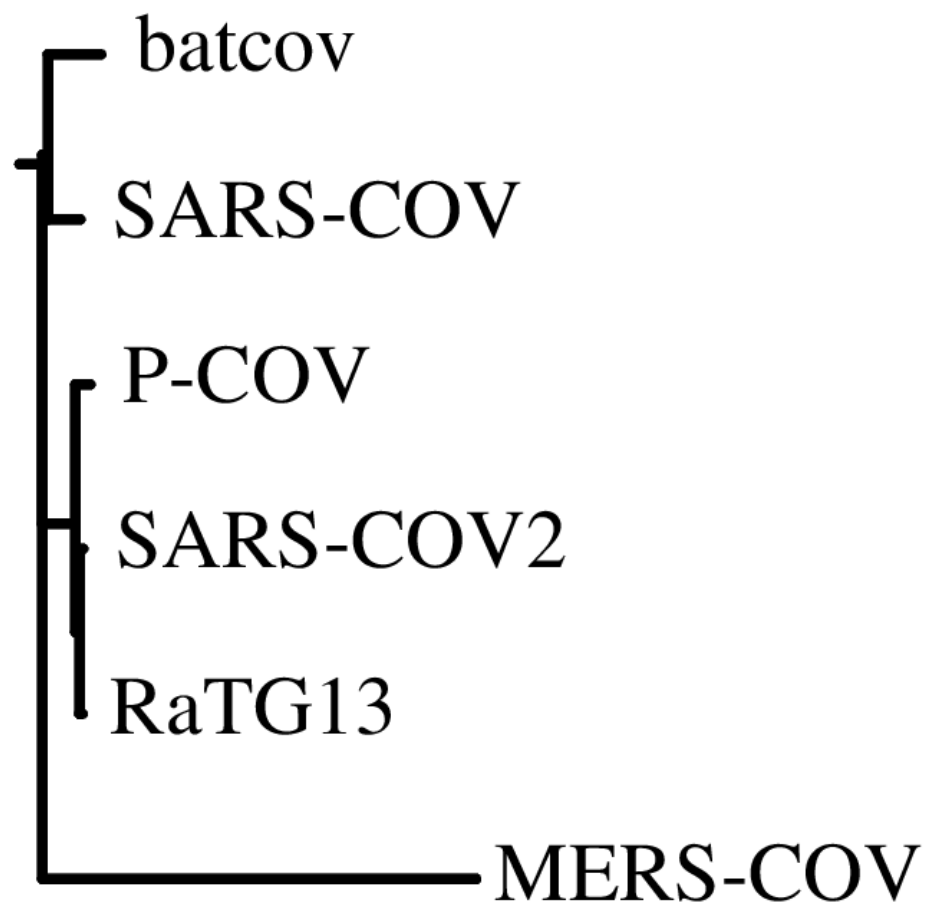
▼ proml



▼ propars



▼ prodist



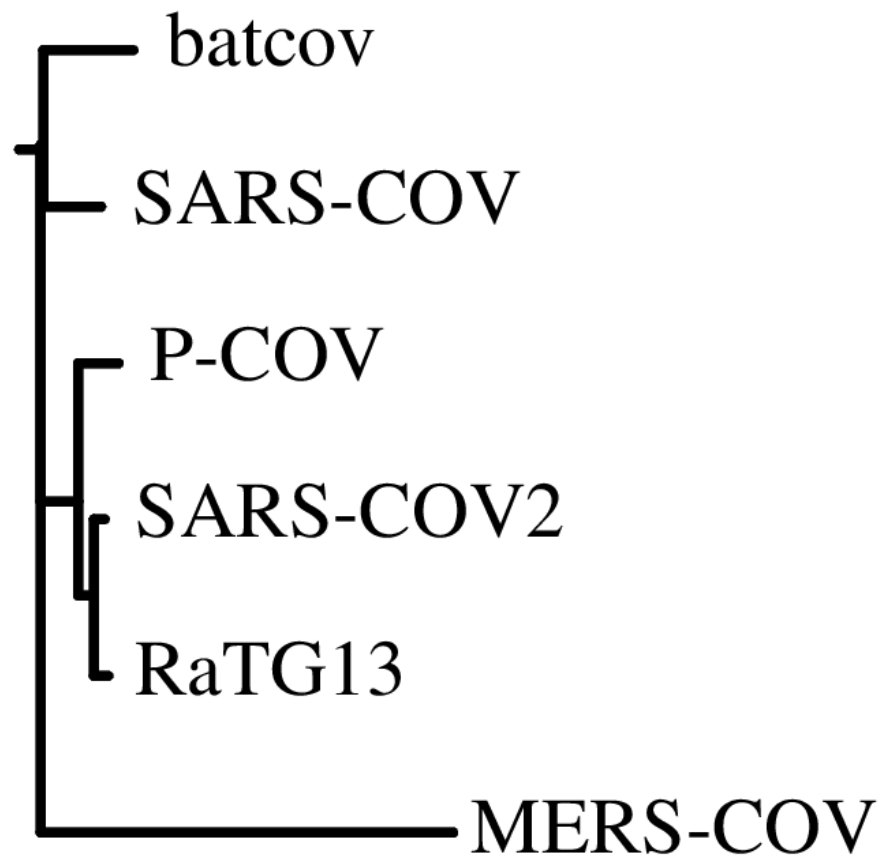
Yes, topology-wise the trees seem to be in agreement in different methods. For example, RaTG13 and SARS-COV2 are closely related in all the methods whereas MERS-CoV is a distant relative in all the different methods.

▼ Part b>

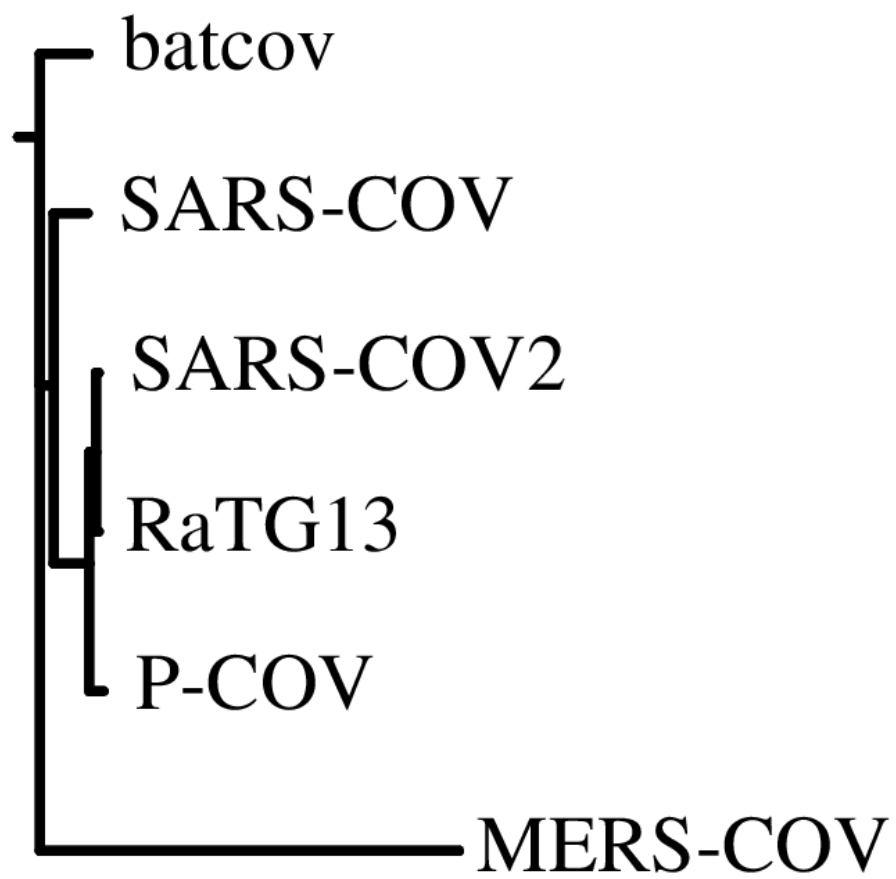
Following are the trees obtained using Consensus:

▼ dnadist

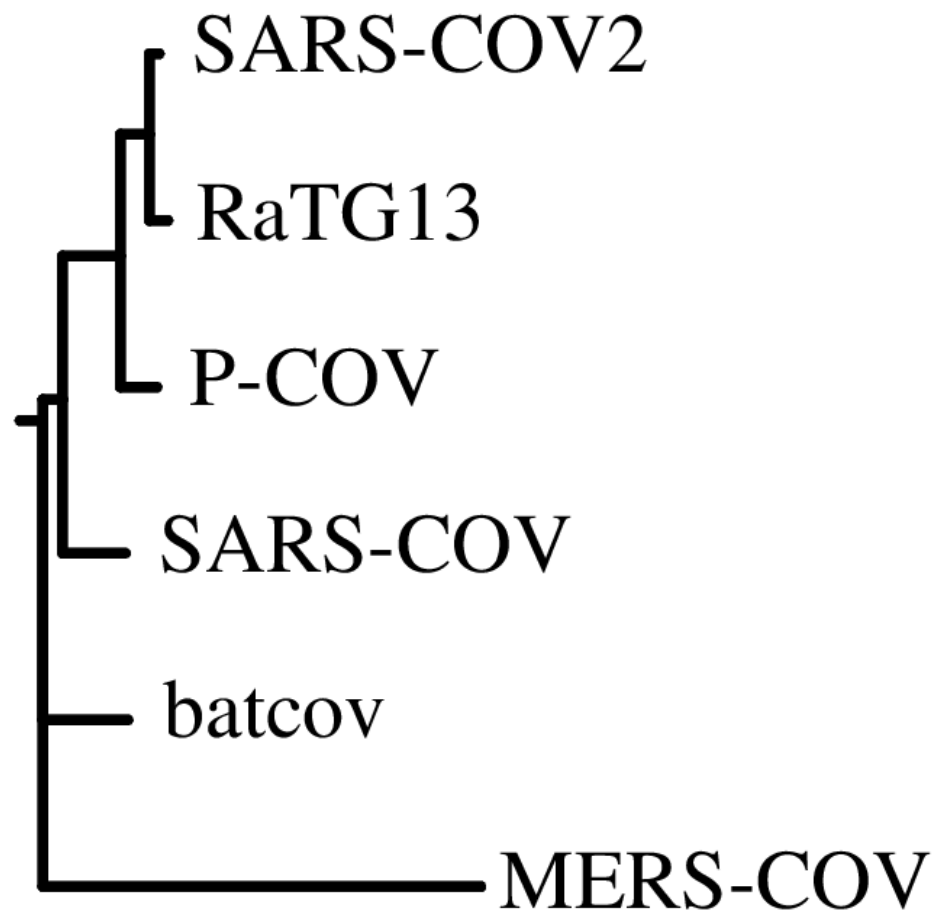




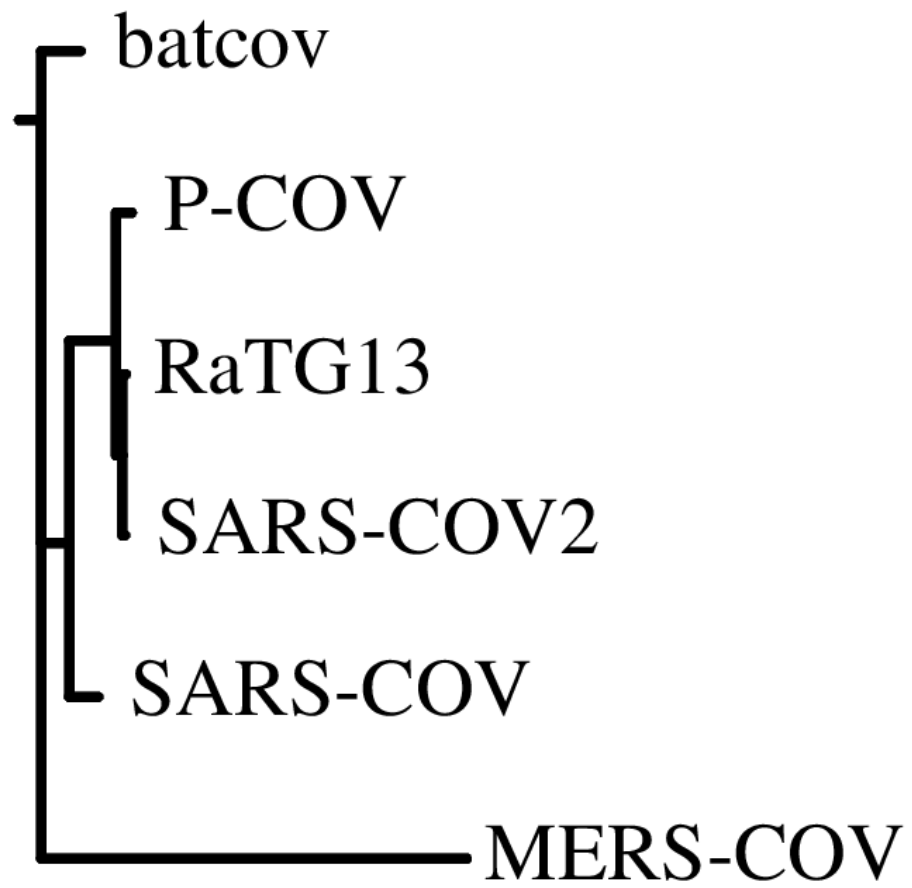
▼ protdist



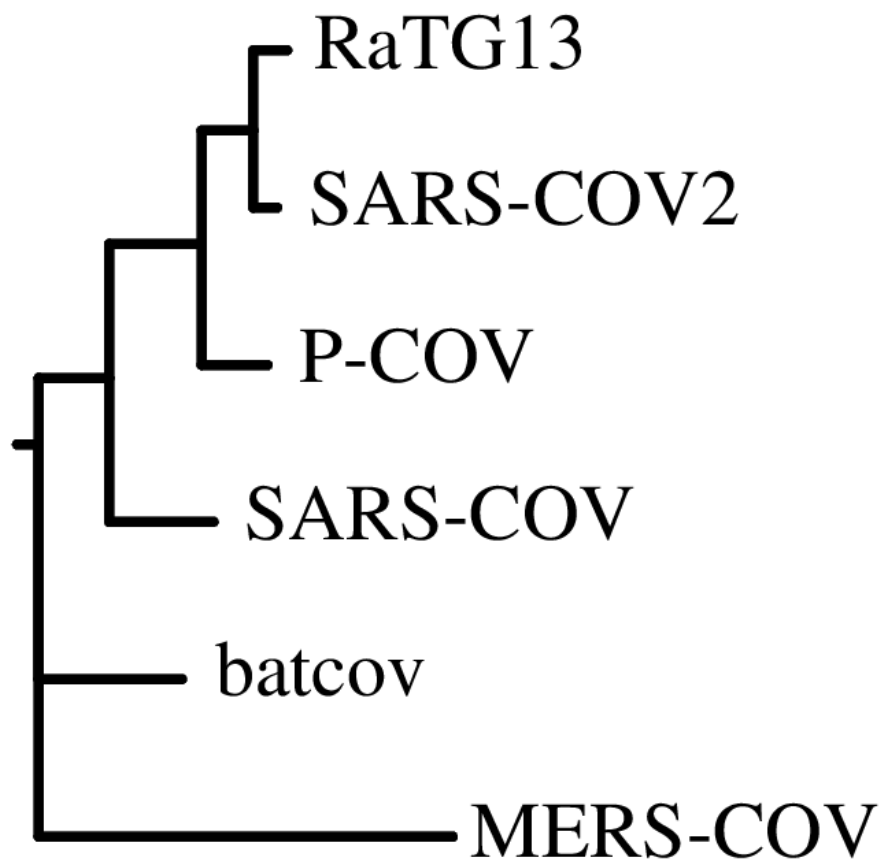
▼ dnaml



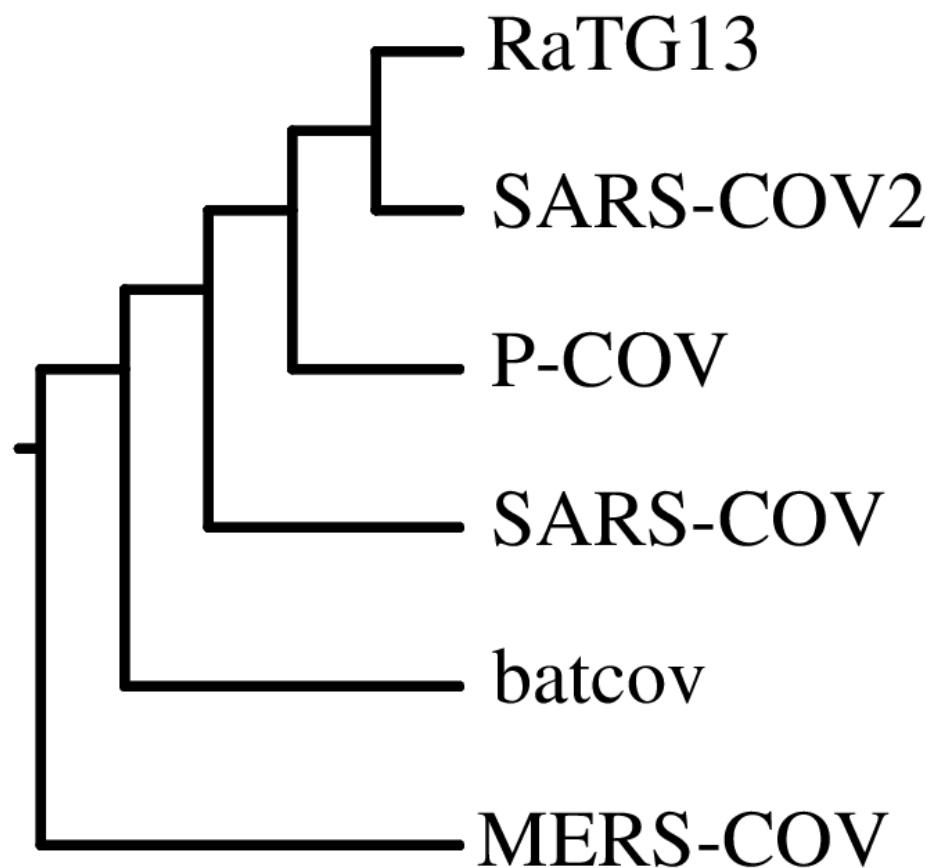
▼ proml



▼ dnapars



▼ protpars



The general method of bootstrapping involves running SEQBOOT to make multiple bootstrapped data sets out of your one data set, then running one of the tree-making programs with the Multiple data sets option to analyze them all, then running CONSENSE to make a majority rule consensus tree from the resulting tree file.

Though the figures obtained look quite similar with and without bootstrap, we can still notice minor differences.

Here,

▼ Part c>

Yes they do match

