

Science II Tutorial 2 Assignment

Roll No: 2021101064

Name: Aman Raj

1)

(A) The following results were obtained by using the multiple sequence alignment. The similarity matrix derived comes out to be:

For DNA:

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

1: MERS	100.00	45.98	46.26	46.08	46.54	46.66
2: BATCOV	45.98	100.00	70.16	68.11	68.24	68.42
3: COV	46.26	70.16	100.00	73.56	73.32	73.35
4: PCOV	46.08	68.11	73.56	100.00	82.83	83.39
5: BATRAG13	46.54	68.24	73.32	82.83	100.00	93.12
6: COV2	46.66	68.42	73.35	83.39	93.12	100.00

For Protein:

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

1: MERS-CoV	100.00	31.77	32.28	31.91	32.32	32.30
2: BatCOV	31.77	100.00	75.85	71.61	72.28	72.26
3: SARS-CoV	32.28	75.85	100.00	77.38	77.76	77.30
4: PCOV	31.91	71.61	77.38	100.00	92.98	92.43
5: RATG13	32.32	72.28	77.76	92.98	100.00	97.71
6: SARS	32.30	72.26	77.30	92.43	97.71	100.00

As we can see from the Identity Matrix, the closest relative to SARS CoV 2 is Bat RatG13 since the percentage match is around 93.13 with SARS in DNA and 97.71 in

the Protein.

(B)

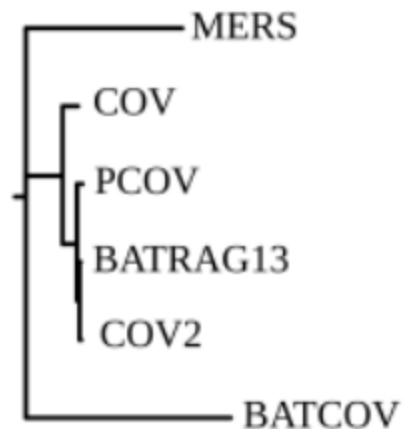
We can describe this from (A) that the source of SARS CoV 2 could be most likely be the Bat RatG13 due to their similarity.

For MERS, we cannot infer the from the data.

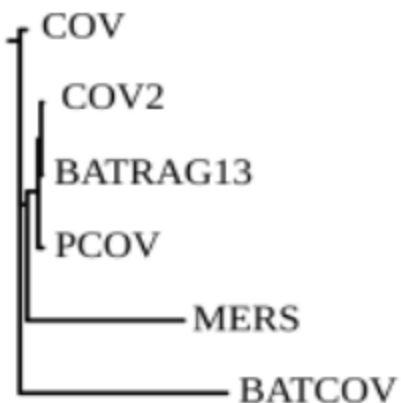
2)

The trees generated are as follows:

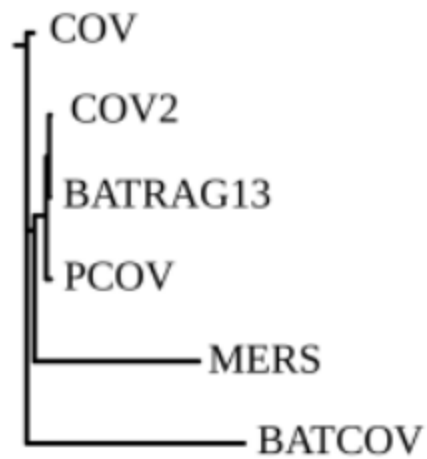
Distance Without Bootstrap : -



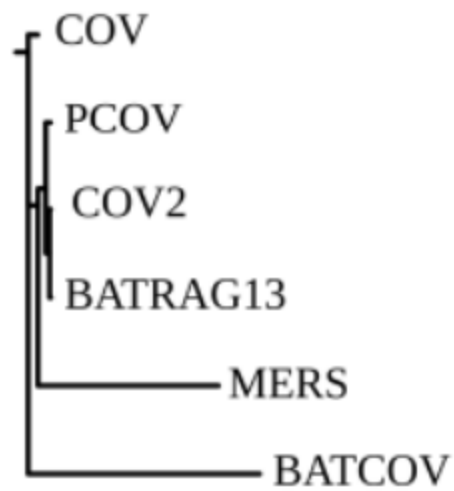
Distance With Bootstrap : -



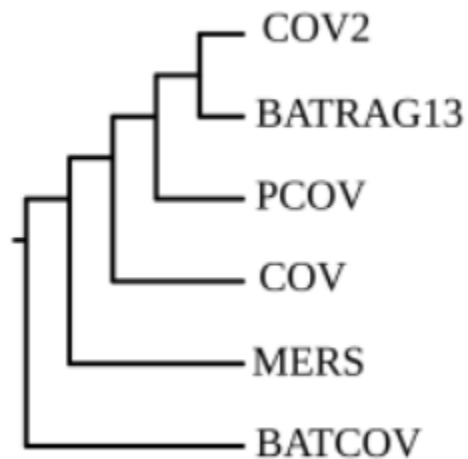
Likelihood Without Bootstrap : -



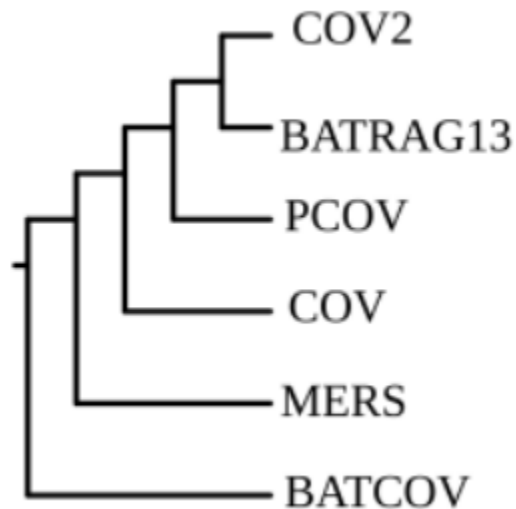
Likelihood With Bootstrap : -



Parsing Without Bootstrap : -



Parsing With Bootstrap : -



- a) No, the trees that are obtained by the methods are not topology-wise.
- b) There are essentially no differences between the bootstrapped and non bootstrapped data here. Bootstrapping is a statistical method that is used for resampling a large set to make a large number of stimulated samples.
- c) Yes, the inferences derived from here are in really striking agreement to Q1(b). This is because here, pairing is similar for all the cases and for the one which we can't

determine, it is paired with the others.

Job Title

Sequence1

RID

4JJCUTM0016

Search expires on 04-27 23:56 pm

Download All

Program

BLASTN

Citation

Database

nt

See details

Query ID

lcl|Query_73073

Description

Sequence1

Molecule type

dna

Query Length

4706

Other reports

Distance tree of results

MSA viewer

Filter Results

Organism

only top 20 will appear

☐ exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity

E value

Query Coverage

to

to

to

Filter

Reset

Descriptions

Graphic Summary

Alignments

Taxonomy

Alignment view

Pairwise

☐ CDS feature

Restore defaults

Download

1000 sequences selected

Download

GenBank

Graphics

Next

Previous

Descriptions

Arabidopsis thaliana genome assembly, chromosome: 2

Sequence ID: LR782543.1

Length: 20021703

Number of Matches: 1

Range 1: 8458301 to 8463006

GenBank

Graphics

Next Match

Previous Match

Score	Expect	Identities	Gaps	Strand
9048 bits(4706)	0.0	4706/4706(100%)	0/4706(0%)	Plus/Plus
Query 1	CCACACaaaaaaagtcgcagaaaaatataagggaaacaaaaagcgaagacgacaaaaaa	60		
Sbjct 8458301	CCACACAAAAAAGTCGAGAAAAATATATGAGGAAACAAAAAGCGAAGACGACAAAAAA	8458360		
Query 61	aaaaaaCTCTGAttttttttGTTATCTCTCTATCTGAGAGGCACACATTTGCTTC	120		
Sbjct 8458361	AAAAAACTCTGATTTTTTGTATCTCTCTATCTGAGAGGCACACATTTGCTTC	8458420		
Query 121	GTCTTCTTCAATTTATTTATTGGTTTCTCCACTTATCTCCGATCTCAATTCTCCCAT	180		
Sbjct 8458421	GTCTTCTTCAATTTATTTATTGGTTTCTCCACTTATCTCCGATCTCAATTCTCCCAT	8458480		
Query 181	TTCTTCTTCTCAAGTTCAAAATCTTGAGAAATTAGCTCTACCGAATTCGTCTCCGAT	240		
Sbjct 8458481	TTCTTCTTCTCAAGTTCAAAATCTTGAGAAATTAGCTCTACCGAATTCGTCTCCGAT	8458540		
Query 241	AACTAGTGGATGATGATTCACCTAAATCCTTCCTGTCTCAAGGTAATTCGAGAAAT	300		
Sbjct 8458541	AACTAGTGGATGATGATTCACCTAAATCCTTCCTGTCTCAAGGTAATTCGAGAAAT	8458600		

[Edit Search](#)
[Save Search](#)
[Search Summary](#)

[How to read this report?](#)
[BLAST Help Videos](#)
[Back to Traditional Results Page](#)

Job Title

Seq2

RID

4JJRWJ0R013

Search expires on 04-27 23:55 pm

Download All

Program

BLASTN

Citation

Database

nt

See details

Query ID

lcl|Query_45425

Description

Seq2

Molecule type

dna

Query Length

261

Other reports

Distance tree of results

MSA viewer

Filter Results

Organism

only top 20 will appear

exclude

Type common name, binomial, taxid or group name

Add organism

Percent Identity

E value

Query Coverage

to

to

to

Filter

Reset

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Select columns

Show

1000

select all

243 sequences selected

GenBank

Graphics

Distance tree of results

MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Tomato spotted wilt virus intergenic region isolate TSWV-IT-CE, genomic RNA	Tomato spotted...	502	502	100%	1e-137	100.00%	261	AJ309924.1
Tomato spotted wilt orthotospovirus isolate T1012 segment M, complete sequence	Tomato spotted...	481	481	100%	3e-131	99.23%	4762	Q840011.1
Tomato spotted wilt orthotospovirus isolate 104DOT22S segment M, complete sequence	Tomato spotted...	475	475	100%	2e-129	98.85%	4762	QO534350.1
Tomato spotted wilt orthotospovirus isolate 102SEC22S segment M, complete sequence	Tomato spotted...	475	475	100%	2e-129	98.85%	4762	QO534347.1
Tomato spotted wilt orthotospovirus isolate 71SET22S segment M, complete sequence	Tomato spotted...	475	475	100%	2e-129	98.85%	4762	QO534344.1
Tomato spotted wilt orthotospovirus isolate 106DOT22S segment M, complete sequence	Tomato spotted...	475	475	100%	2e-129	98.85%	4762	QO534341.1
Tomato spotted wilt orthotospovirus isolate 106DOT22S segment M, complete sequence	Tomato spotted...	475	475	100%	2e-129	98.85%	4762	QO507128.1
Tomato spotted wilt orthotospovirus isolate 107DOT22S segment M, complete sequence	Tomato spotted...	475	475	100%	2e-129	98.85%	4762	QO507125.1
Tomato spotted wilt orthotospovirus isolate 105DOT22S segment M, complete sequence	Tomato spotted...	475	475	100%	2e-129	98.85%	4762	QO507122.1
Tomato spotted wilt orthotospovirus isolate 86DOW22S segment M, complete sequence	Tomato spotted...	475	475	100%	2e-129	98.85%	4762	QO507119.1
Tomato spotted wilt orthotospovirus isolate 85DOP22S segment M, complete sequence	Tomato spotted...	475	475	100%	2e-129	98.85%	4762	QO507116.1
Tomato spotted wilt virus isolate intergenic region, isolate TSWV-IT-TQ	Tomato spotted...	475	475	100%	2e-129	98.85%	260	AJ3428574.1
Tomato spotted wilt virus RNA segment M, complete sequence	Tomato spotted...	435	435	100%	3e-117	96.98%	4768	AB190818.1
Tomato spotted wilt orthotospovirus isolate P349 nonstructural protein (NSm) gene, complete cds	Tomato spotted...	433	866	100%	1e-116	97.32%	1564	Q8867574.1

Sequence 1 is closely related to “Arabidopsis thaliana genome” from mustard family.

Sequence 2 is closely related to “Tomato spotted wilt virus intergenic region, isolate TSWV-IT-CE, genomic RNA” from Arachis hypogea plant.