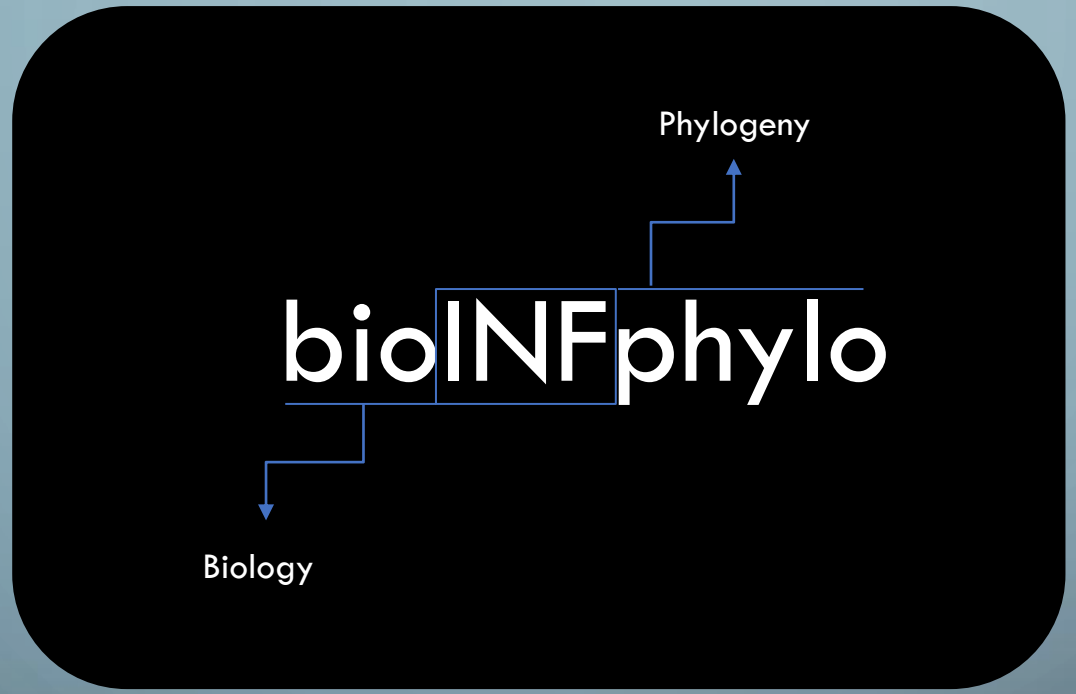


A tree generating app for Bioinformatics

By: Sal Jahromi

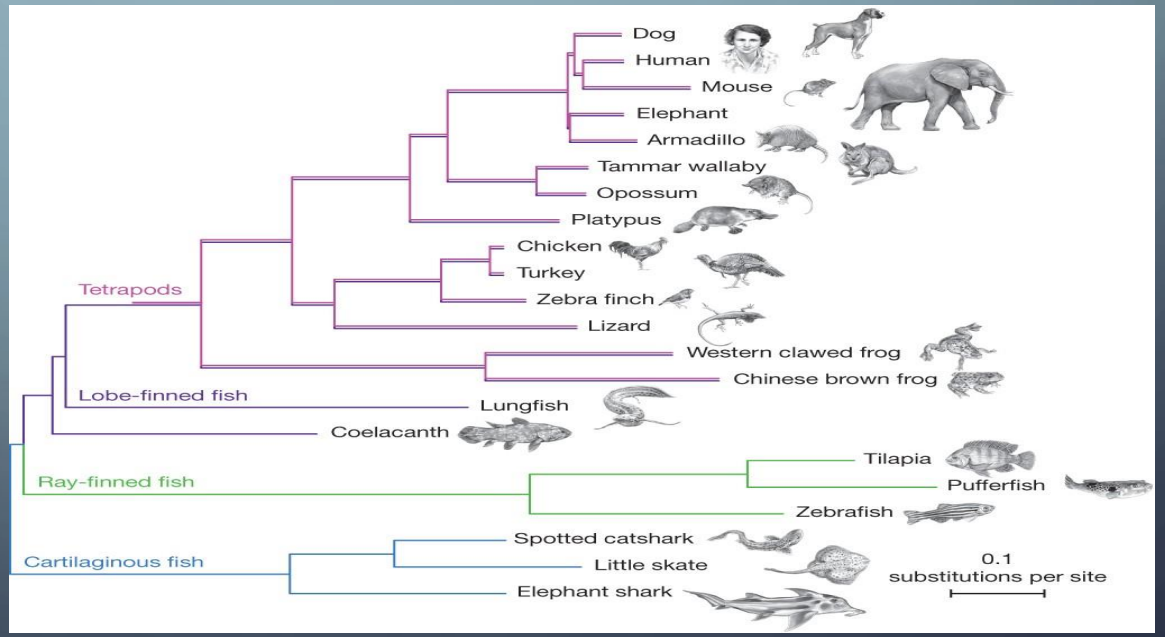
FASTA FILES

This app uses FASTA files to generate phylogenetic trees.

FASTA = A file format used by bioinformaticians and biochemists, which contains nucleotide or amino acid sequences, along with the names of the hosts.

```
>Cat
ACTCCGTCTCGAGCGCAAGCT
>Bird
ACTAAACCTTTCGGTGAGTCC
>Bat
ACTCGGGCCACTACTACTTTCAA
```

Phylogenetic Tree = A tree diagram which represents the evolutionary relationship of organism with each other and their common ancestors

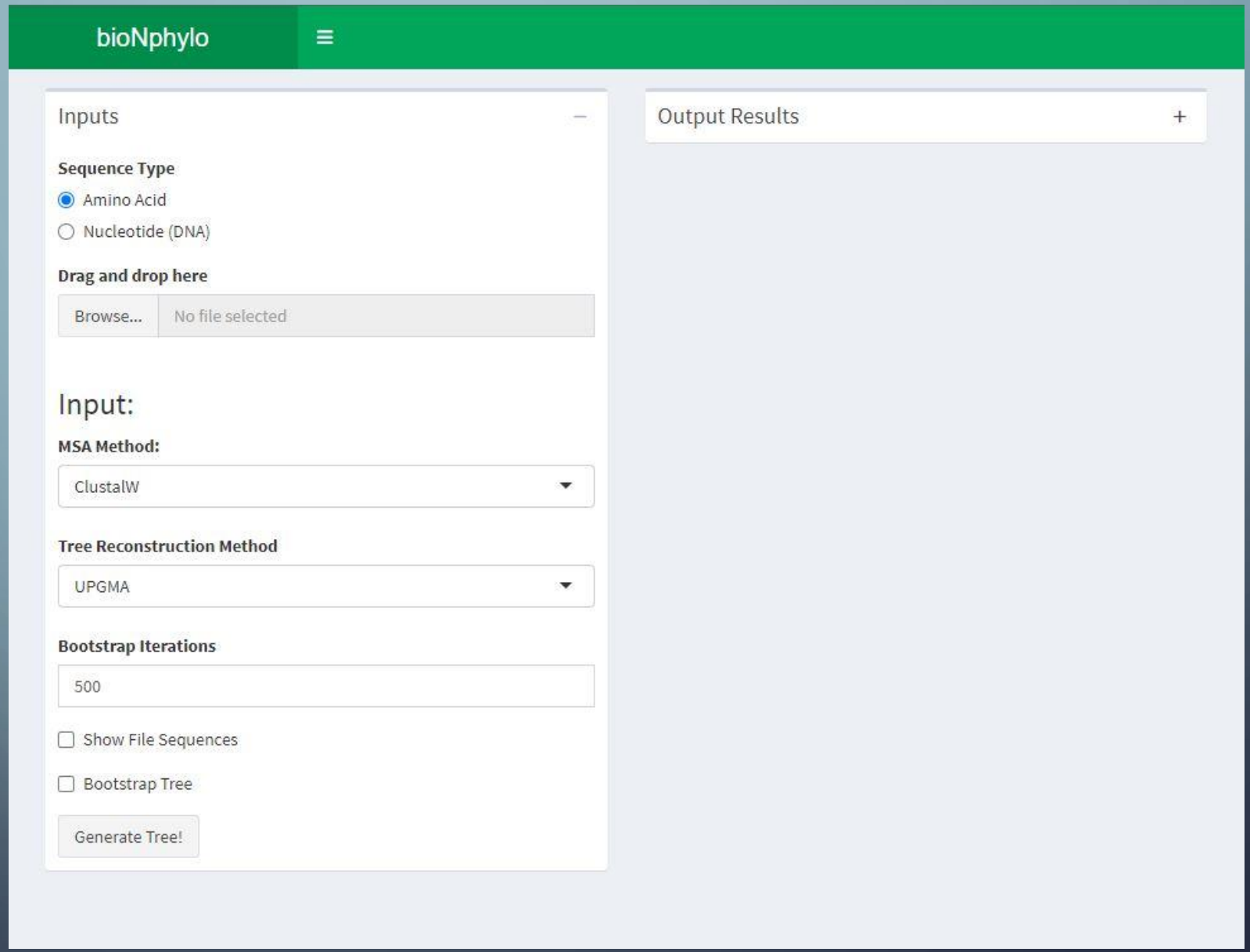


https://plato.stanford.edu/entries/phylogenetic-inference/

APP BUILT IN R-SHINY

- Why R?
  - R is a very powerful tool for data analysis.
    - lots of built-in libraries to read and analyze FASTA files with.
    - Excellent plots (trees, graphs, charts, etc.)
- R Shiny?
  - R Shiny is used to build UI apps based on R
    - Makes for a user-friendly interface

THE APP


bioNphylo
Inputs
Output Results
Sequence Type
Amino Acid
Nucleotide (DNA)
Drag and drop here
Browse...
No file selected
Input:
MSA Method:
ClustalW
Tree Reconstruction Method
UPGMA
Bootstrap Iterations
500
Show File Sequences
Bootstrap Tree
Generate Tree!

Purpose : To depict the following coronavirus types in a phylogenetic tree, to better understand their evolutionary relationships.

- SARS CoV1 E2
- SARS CoV1 AS
- Pangolin coronavirus
- Bat coronavirus
- SARS-CoV-2 iso 1
- SARS CoV2 Wuhan-Hu
- Camel MERS-CoV
- MERS Betacoronavirus England 1
- MERS coronavirus

To accomplish this, the spike protein amino acid sequence of the virus types is collected within a FASTA file.

-Sequences were collected from the NCBI website.

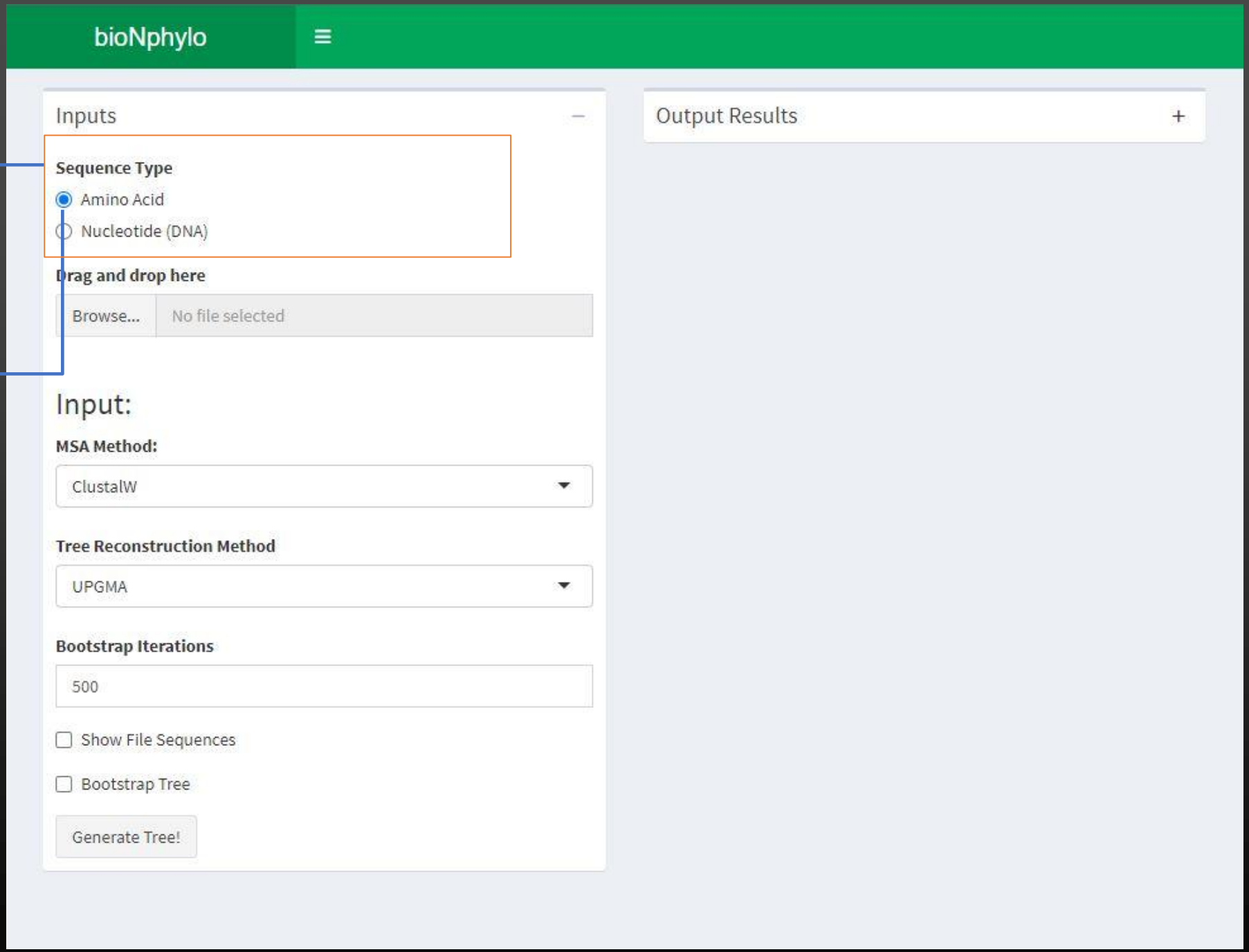
Spike.fa

Amino acid sequence of the spike proteins of the coronavirus types in this study

Default screen on start-up

To select the type of sequence the FASTA file holds

Since “Spike.fa” holds amino acid sequence, we will leave that radio button selected



Can simply drag and drop file or select it through browsing.

bioNphylo

Inputs

Sequence Type

Amino Acid

Nucleotide (DNA)

Drag and drop here

Browse... No file selected

Input:

MSA Method:

ClustalW

Tree Reconstruction Method

UPGMA

Bootstrap Iterations

500

Show File Sequences

Bootstrap Tree

Generate Tree!

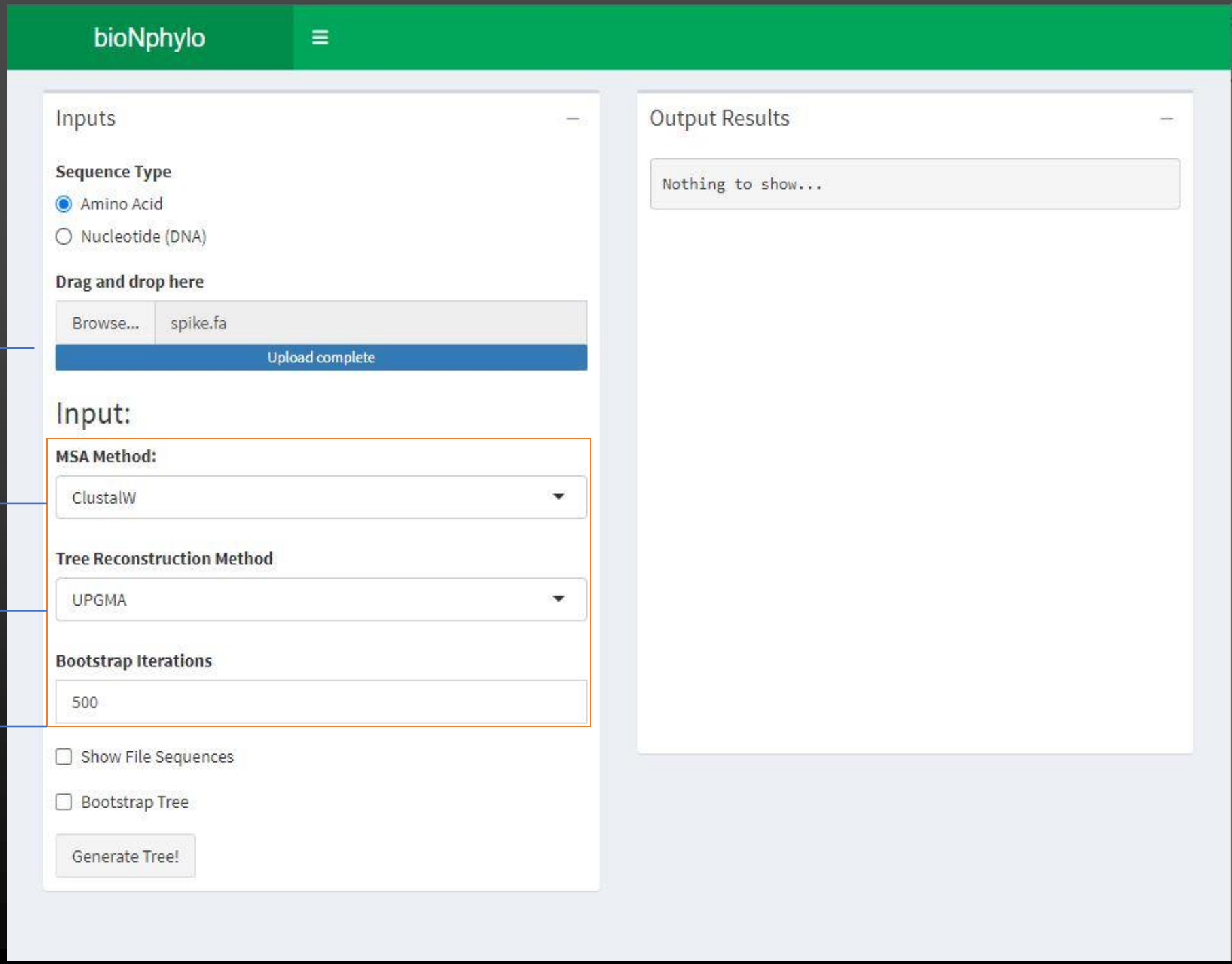
Output Results

Upload complete

Method of multiple sequence alignment

Tree reconstruction method

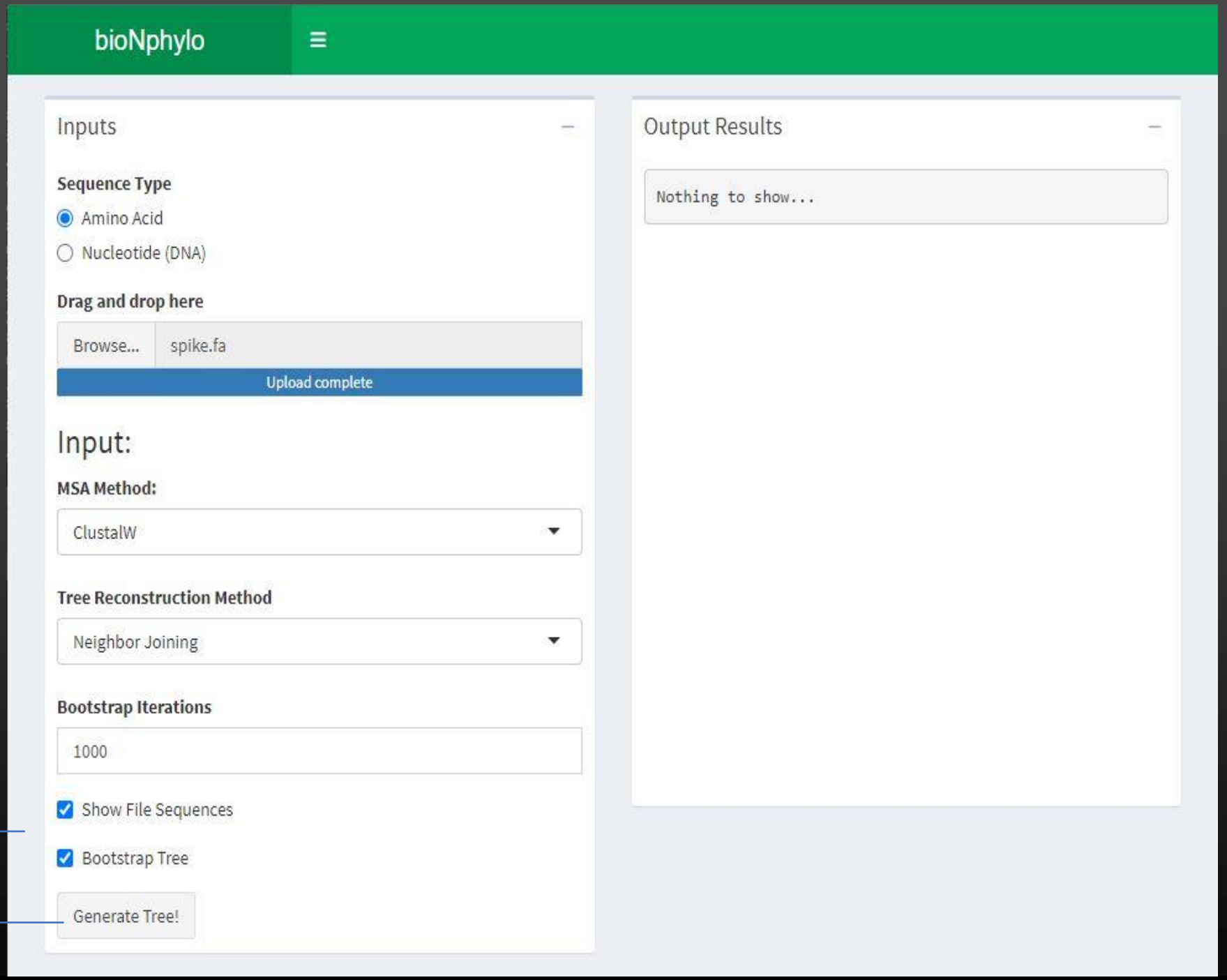
Bootstrap Iterations (1000 is recommended)

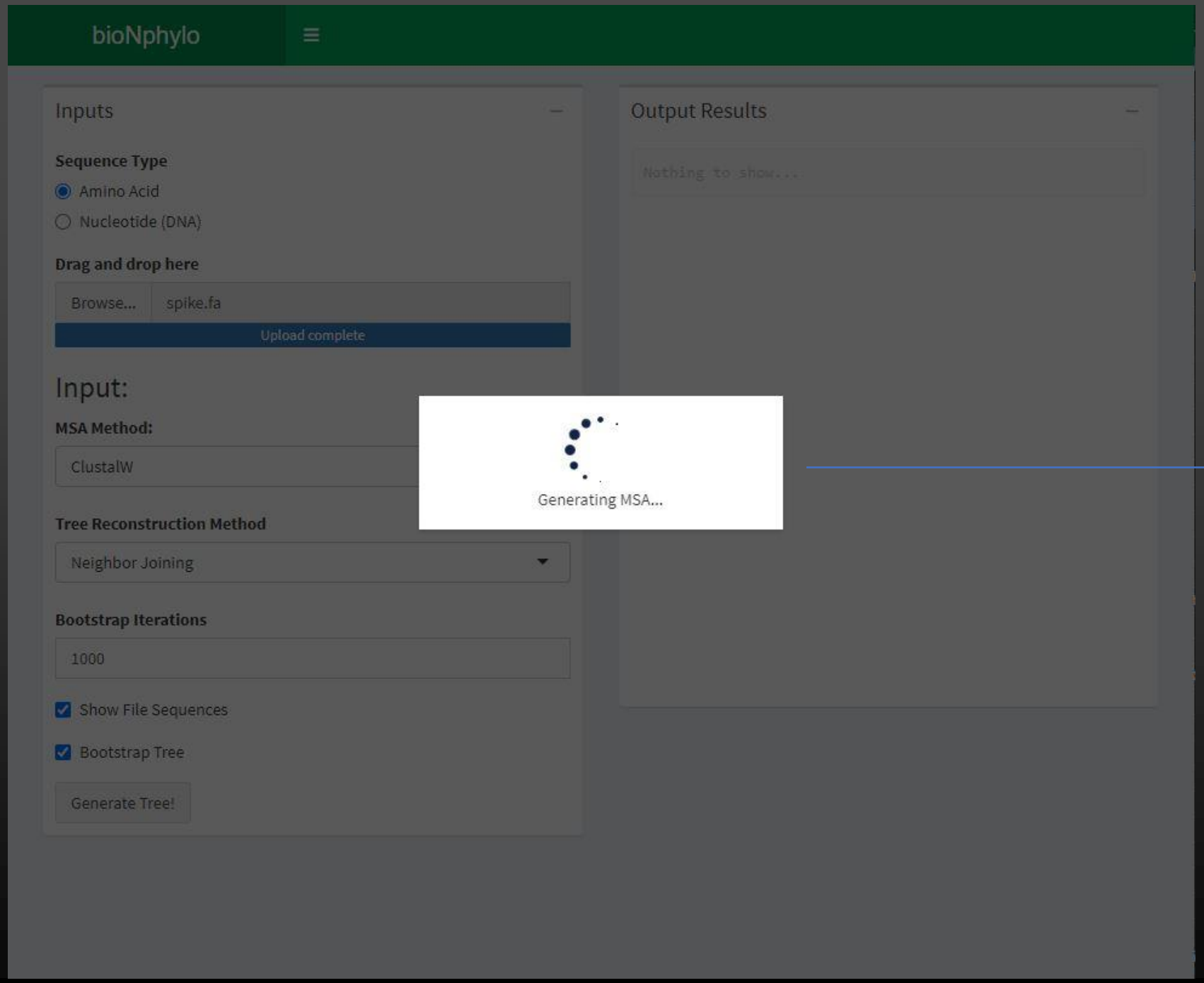


Select desired settings

Select what we want to be shown. If nothing is selected, only a tree is shown

Click to generate the tree!





A loading screen which updates, letting the user know what the program is doing

bioNphylo

Inputs

Sequence Type

- (•) Amino Acid
- ( ) Nucleotide (DNA)

Drag and drop here

Browse... spike.fa

Upload complete

Input:

MSA Method:

ClustalW

Tree Reconstruction Method

Neighbor Joining

Bootstrap Iterations

1000

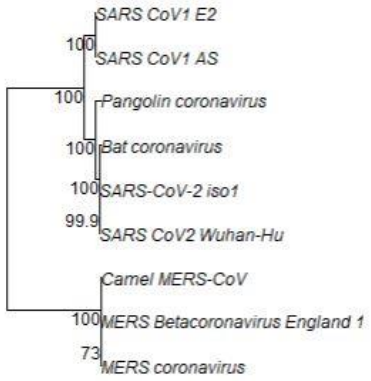
- [x] Show File Sequences
- [x] Bootstrap Tree

Generate Tree!

Output Results

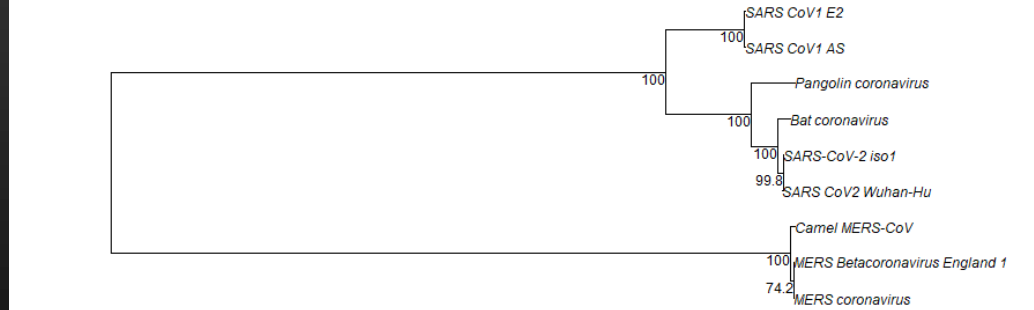
Phylo Result

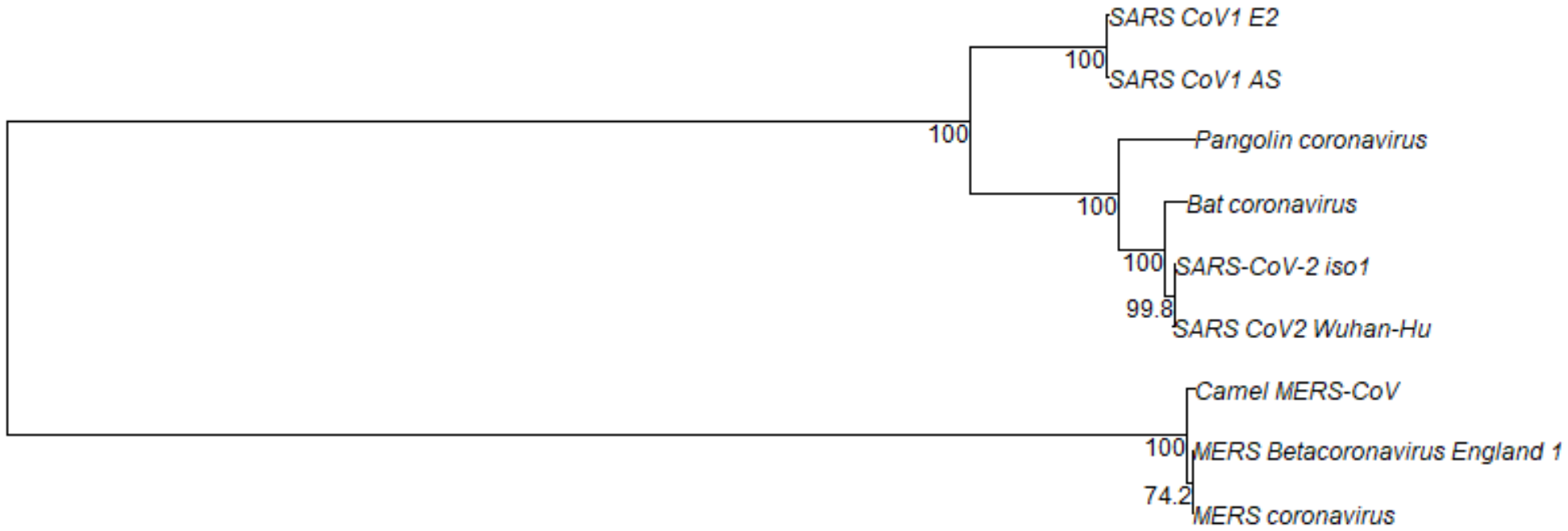
```
AAStringSet object of length 9:
    width seq
[1]  1353 MIHSVFLLMFLLTPTESYVDVGPDSAKSVCIEVDIQQTFFDKTWPRP
[2]  1269 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSSTRGVYYPDKVFRSS
[3]  1265 MLFFFFLHFALVNSQCVNLTGRAAIQPSFTNSSQRGVYYPDTIFRS
[4]  1273 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSS
[5]  1273 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSS
[6]  1255 MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEI
[7]  1255 MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEI
[8]  1353 MIHSVFLLMFLLTPTESYVDVGPDSVKSACIEVDIQQTFFDKTWPRP
[9]  1353 MIHSVFLLMFLLTPTESYVDVGPDSVKSACIEVDIQQTFFDKTWPRP
```



The file sequence is shown here.
There are 9 sequences (9 subjects) and the length of each sequence is under “width”.

Phylogenetic tree given with bootstrap values




SARS CoV1 E2
100
SARS CoV1 AS
100
Pangolin coronavirus
100
Bat coronavirus
100
SARS-CoV-2 iso1
99.8
SARS CoV2 Wuhan-Hu
Camel MERS-CoV
100
MERS Betacoronavirus England 1
74.2
MERS coronavirus