HW9-TrucHuynh

Truc Huynh

4/16/2021

## Objectives

* Query and read list of sequence data.
* Produce and analyze multiple sequence alignments.
* Construct, plot, and analyze Phylogenetic trees.

### Question 1:

Retrieve/download the following DNA sequences from GenBank: SARS coronavirus MA15 ExoN1 (FJ882953) SARS coronavirus Frankfurt (AB257344) SARS coronavirus ZJ0301 (DQ182595) SARS coronavirus strain CV7 (DQ898174) SARS-CoV-2 (MT334547)

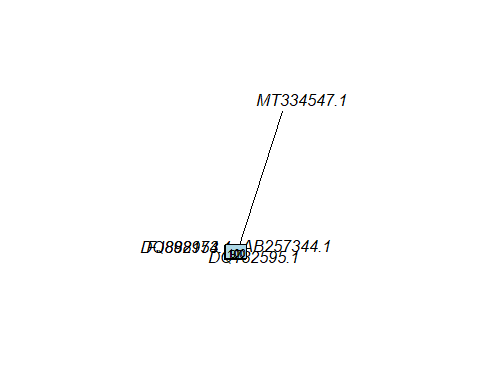
# Download all and paste them in 1 FASTA file

### Question 2:

Build and plot an unrooted phylogenetic tree from the sequence data using the neighbor-joining algorithm. For better display results you may set the type of the phylogenetic tree using type=phylogram

# Read in Multiple Alignment Sequences  
objAln<-read.alignment(file="SARCOV.phy", format = 'phylip')  
unrootedNJtree(objAln,'DNA')

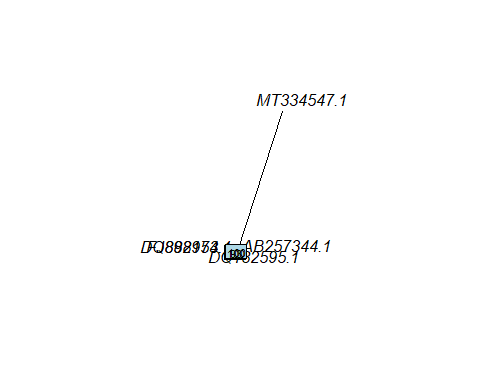
## Running bootstraps: 100 / 100  
## Calculating bootstrap values... done.



##   
## Phylogenetic tree with 5 tips and 3 internal nodes.  
##   
## Tip labels:  
## FJ882953.1, DQ182595.1, DQ898174.1, AB257344.1, MT334547.1  
## Node labels:  
## 100, 97, 100  
##   
## Unrooted; includes branch lengths.

objAlnClean<-cleanAlignment(objAln,30,30)  
unrootedTree<-unrootedNJtree(objAlnClean,'DNA')

## Running bootstraps: 100 / 100  
## Calculating bootstrap values... done.



unrootedTree

##   
## Phylogenetic tree with 5 tips and 3 internal nodes.  
##   
## Tip labels:  
## FJ882953.1, DQ182595.1, DQ898174.1, AB257344.1, MT334547.1  
## Node labels:  
## 100, 93, 100  
##   
## Unrooted; includes branch lengths.

write.tree(unrootedTree,'unrootedTree.tre')

### Question 3:

Which are the most closely related sequences based on the tree?

* After using the write tree function to see the closely related, the result is:
* (DQ898174.1:3.462104461e-05,DQ182595.1:0.0002701546253,((AB257344.1:0.007896816713,MT334547.1:0.2350000195)100:0.002098656081,FJ882953.1:0.001319169067)93:0.0001046645142)100;
* Therefore, (AB257344.1:0.007896816713,MT334547.1:0.2350000195)100:0.002098656081 AB257344.1 and MT334547.1 are the most closely related sequences

### Question 4:

* Original tree Node labels: 100, 91, 100
* After apply the clean Alignment, the Node Label is: 100, 92, 100
* The bootstrap value in all case is very high and most case they are close to 100%. Therefore we can say we are very confident about the group formed in this tree.

### Question 5:

* Build a rooted phylogenetic tree of the SARS virus sequences. Note: the Bovine viral diarrhea virus (GenBank accession JX297515) is related to Corona viruses and so can be used as an outgroup in this case.

objAln<-read.alignment(file="SARCOV2.phy", format = 'phylip')  
  
dist.alignment(objAln)

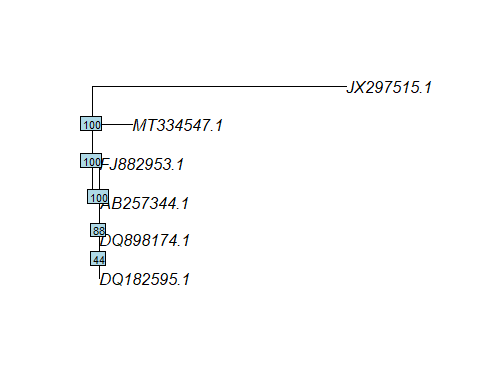
## FJ882953.1 DQ898174.1 DQ182595.1 AB257344.1 MT334547.1  
## DQ898174.1 0.03808347   
## DQ182595.1 0.04106646 0.01740601   
## AB257344.1 0.10654812 0.09978207 0.10095898   
## MT334547.1 0.45021147 0.44894247 0.44932062 0.45358804   
## JX297515.1 0.73404030 0.73386303 0.73392213 0.73404030 0.75305490

objAln$nam[6]

## [1] "JX297515.1"

SARCOVrootedTree<-rootedNJtree(objAln,objAln$nam[6],'DNA')

## Running bootstraps: 100 / 100  
## Calculating bootstrap values... done.



write.tree(SARCOVrootedTree,"SARCOVrootedTree.tre")  
  
SARCOVrootedTree

##   
## Phylogenetic tree with 6 tips and 5 internal nodes.  
##   
## Tip labels:  
## FJ882953.1, DQ898174.1, DQ182595.1, AB257344.1, MT334547.1, JX297515.1  
## Node labels:  
## 100, 44, 88, 100, 100  
##   
## Rooted; includes branch lengths.

### Question 6:

Which are the most closely related SARS virus sequences, based on the tree constructed in Q5: - After using the write tree function to see the closely related, the result is: - (((((DQ182595.1:0.0005925764163,DQ898174.1:1.484407215e-05)48:4.092907134e-05,AB257344.1:0.0009142187806)79:0.0005764947768,FJ882953.1:0.001409215861)100:0.028459734,MT334547.1:0.1530354779)100:0,JX297515.1:0.9622086774)100; - Therefore, AB257344.1,FJ882953.1 and MT334547.1 they have 100% bootstrap values, so that they are the most closely related sequences

### Question 7:

What extra information does the tree in Q5 tell you, compared to the unrooted tree in Q6. The rooted tree have more information than the unrooted tree as we know the direction that evolutionary time ran. Another way to say is that DQ182595.1 and DQ898174.1 share a common ancestor. Then their common ancestor share a common ancestor with AB257344.1 Because it is a rooted tree, we know that time ran from left to right along the branches of the tree The lengths of branches in this tree are proportional to the amount of evolutionary change