Practicals – 10

-BS19B032

-R. Vasantha Kumar

1) As per the given hints in the question and the detailed steps in the presentation given, I obtained phylogenetic trees for the given datasets.

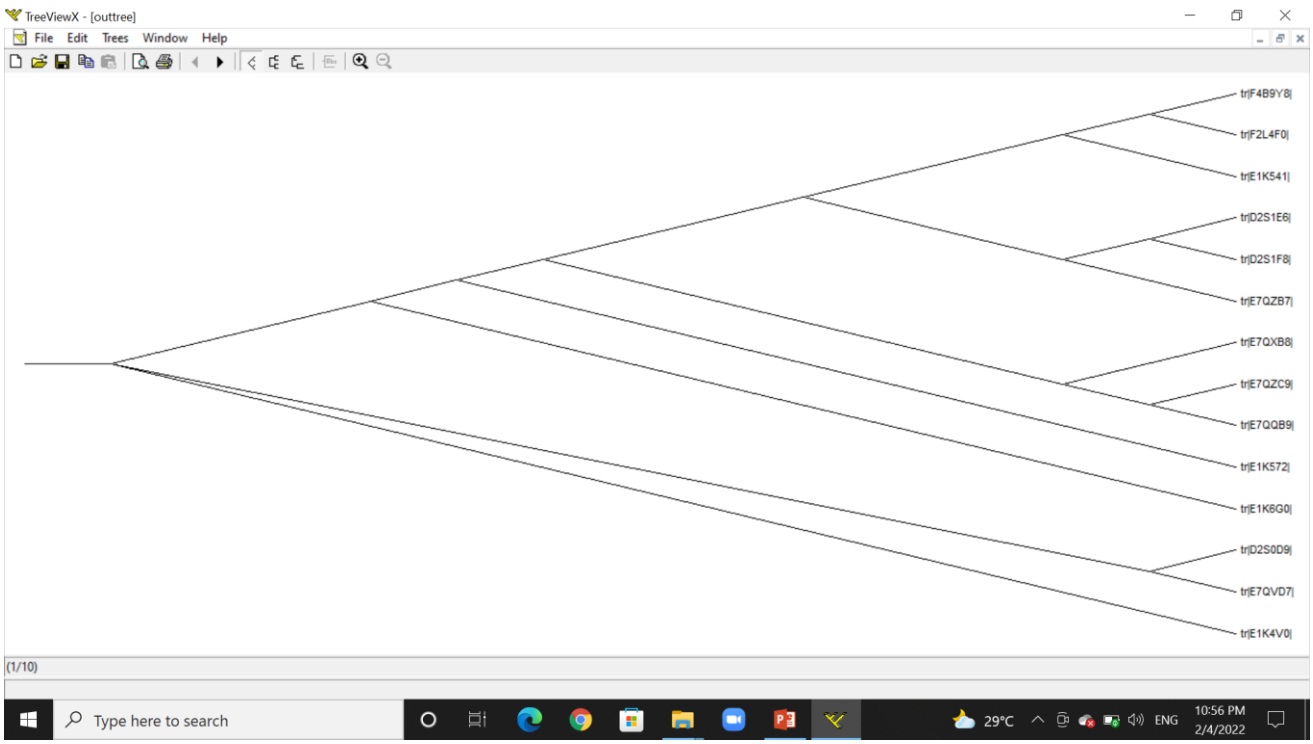
For each dataset, I obtained four trees:

- Normal Phylogenetic tree
- Consensus Phylogenetic tree
- NJ Phylogenetic tree
- Consensus NJ Phylogenetic tree

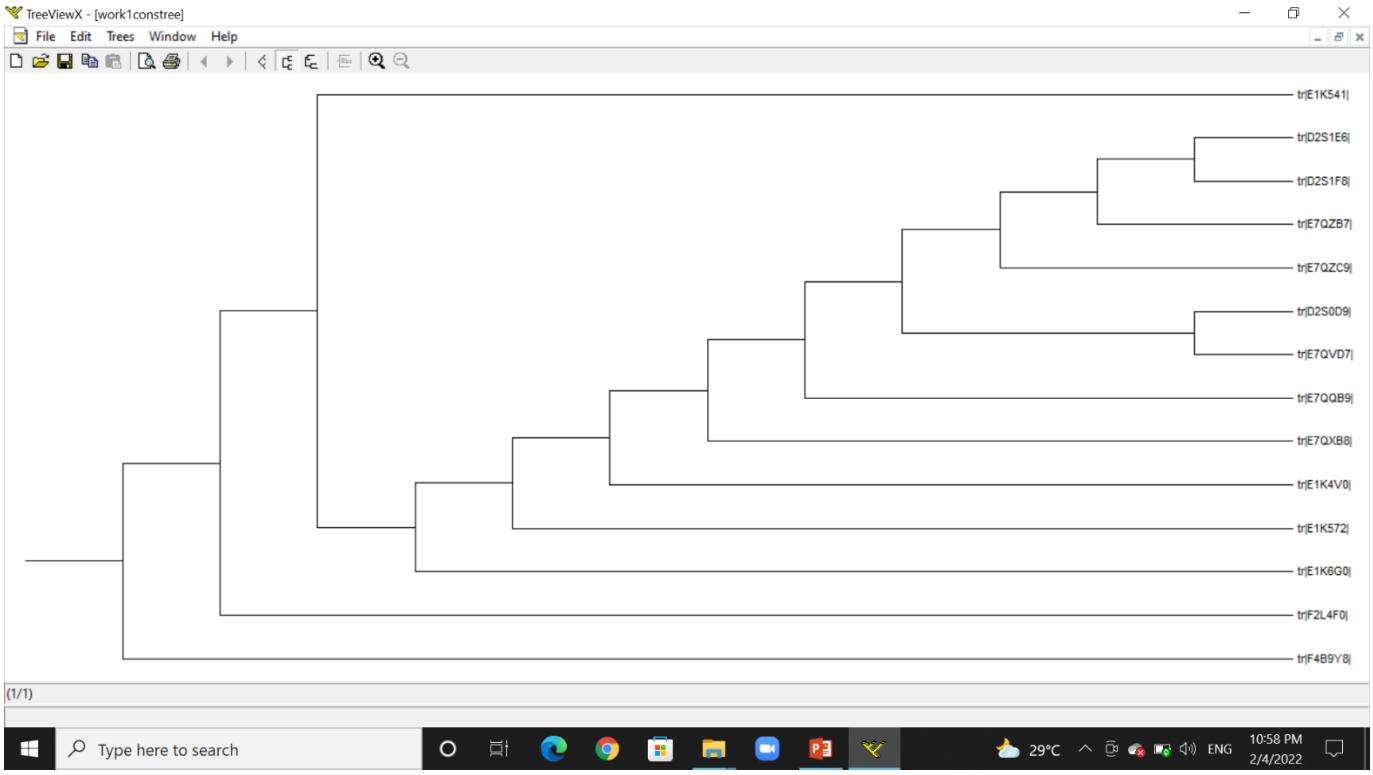
For viewing the trees obtained, I downloaded Treeview X program. Files are attached in submission.

For set1: tim.dat –

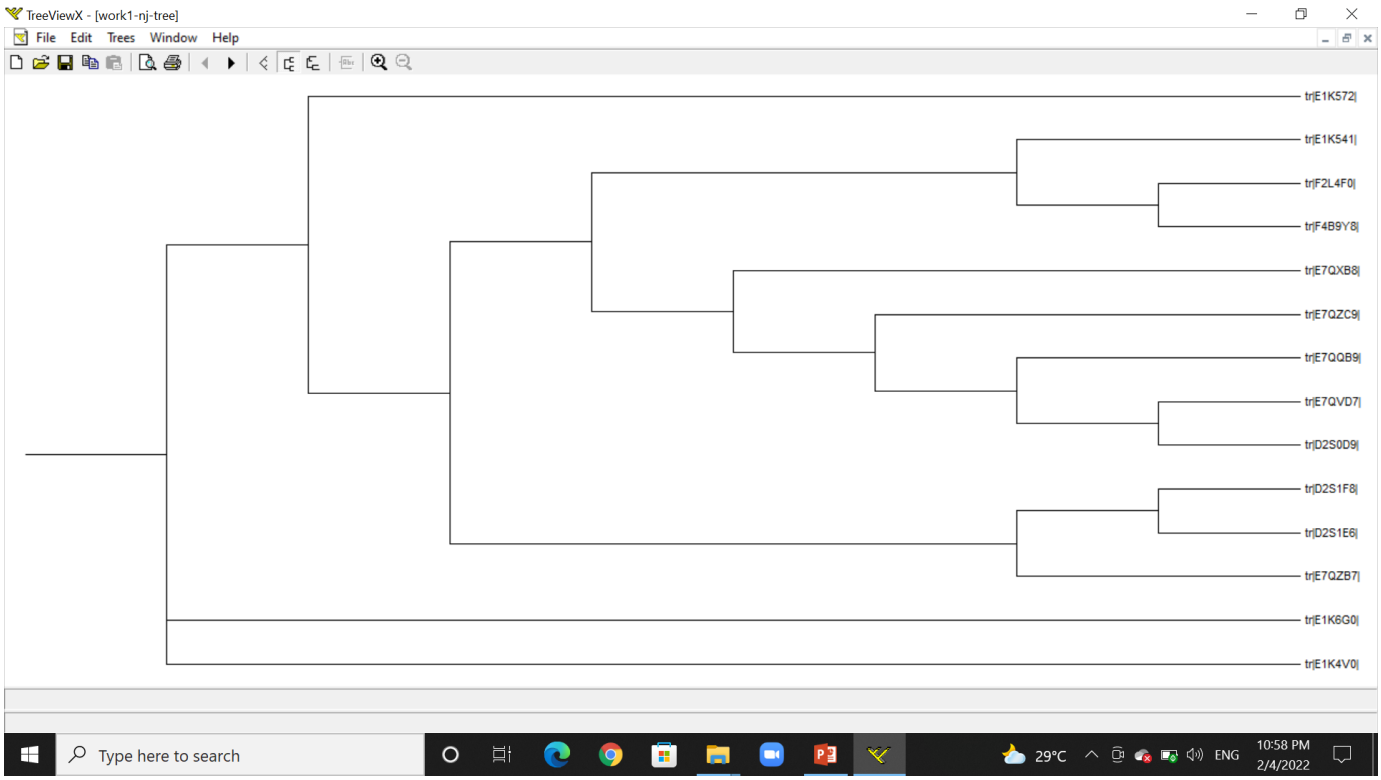
Normal Phylogenetic tree: outtree



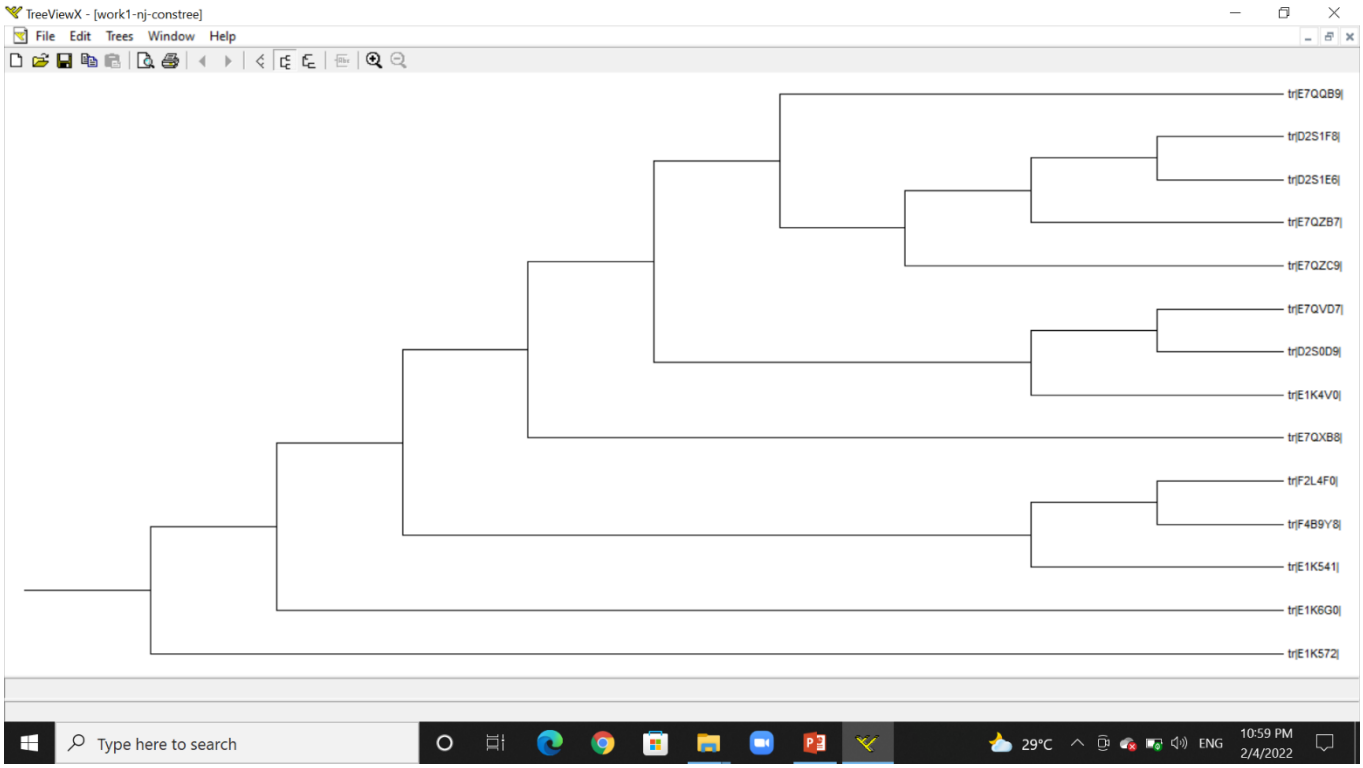
Consensus Phylogenetic tree: work1constree



NJ Phylogenetic tree: work1-nj-tree

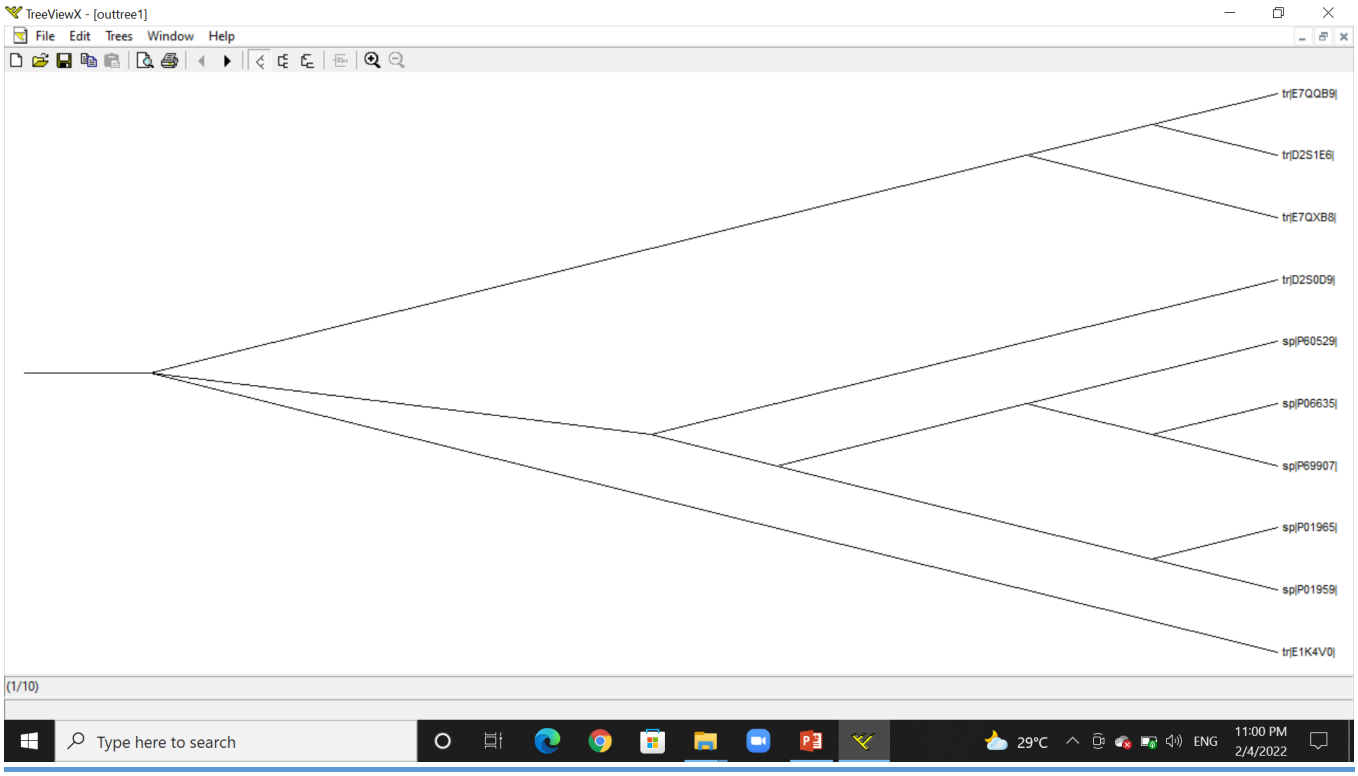


Consensus NJ Phylogenetic tree: work1-nj-constree

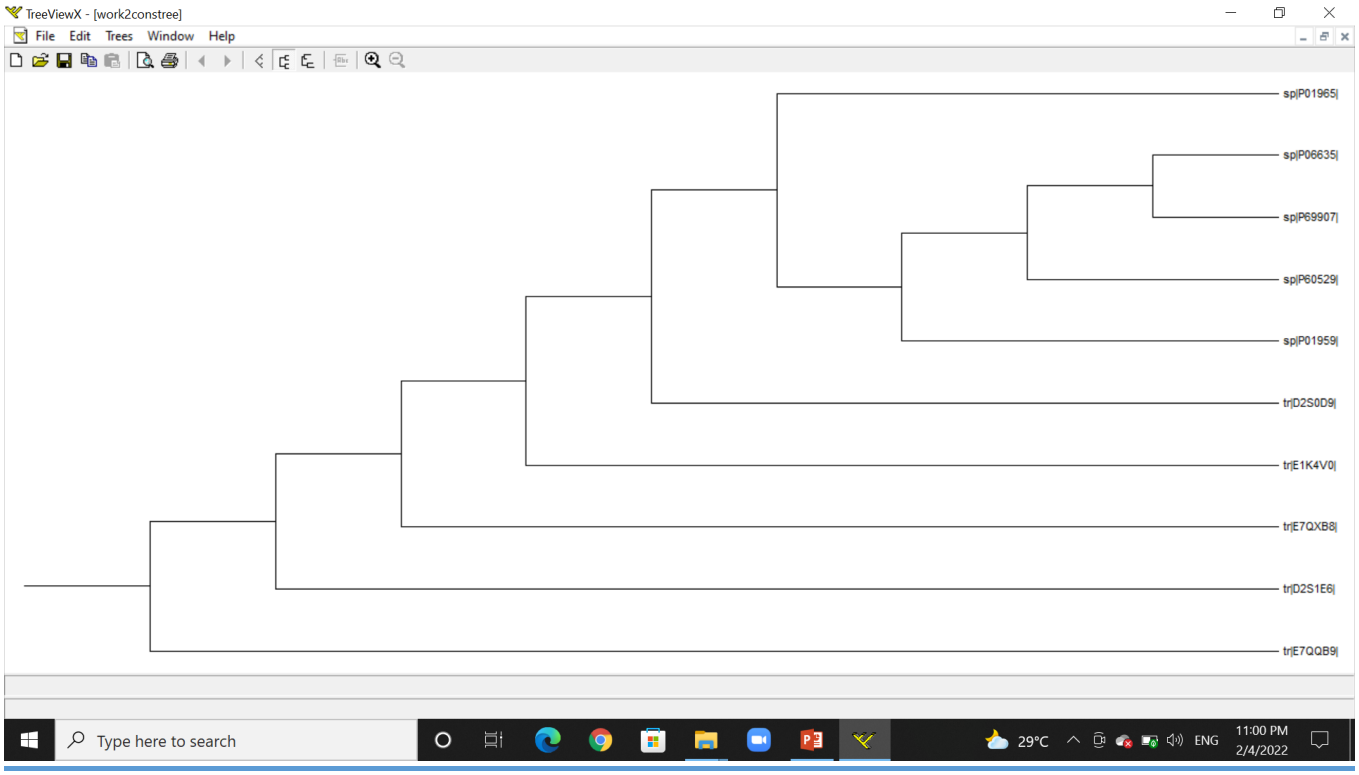


For set2: tim-hemo.dat –

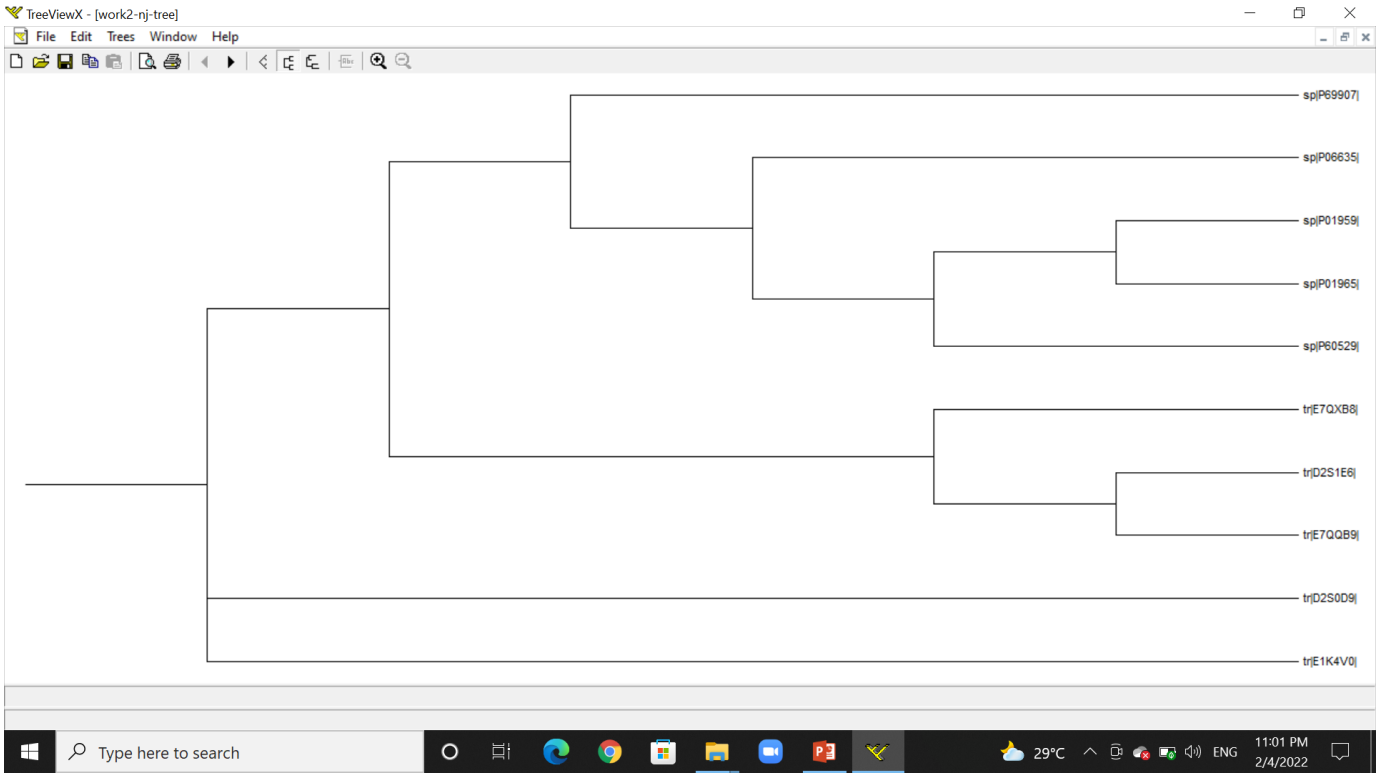
Normal Phylogenetic tree: outtree1



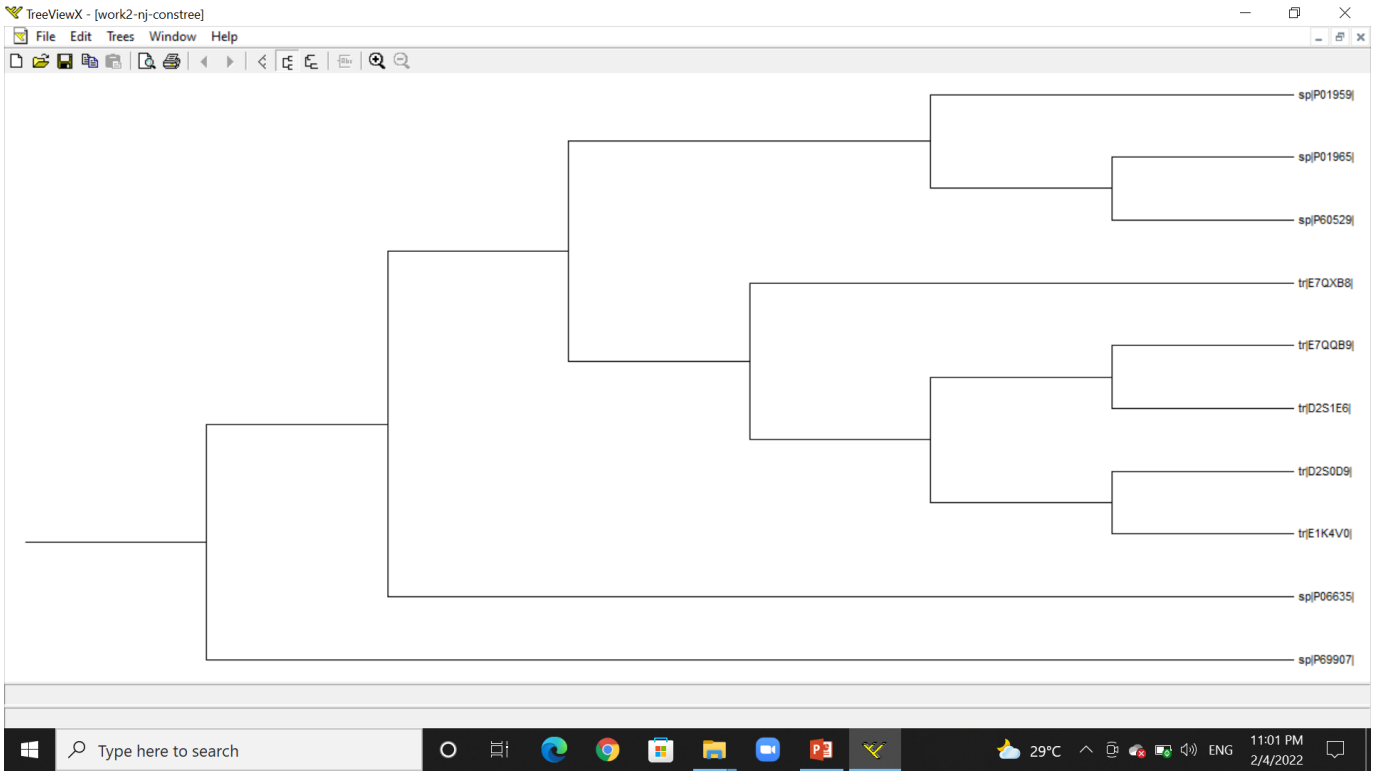
Consensus Phylogenetic tree: work2constree



NJ Phylogenetic tree: work2-nj-tree



Consensus NJ Phylogenetic tree: work2-nj-constree



2) I wrote a code to create the weight matrix for the given sequences using the given formula. I attached the code in submission.(weight_matrix.py)

$$w[i][j] = \ln(((n_{i,j} + 0.05)/(N+1))/0.05)$$

$n_{i,j}$ = frequency of amino acid i, in position j.

0.05 = 1/20, since, total amino acids = 20

N = total number of sequences

I attached the weight matrix as an excel file in submission. (w_mat.xlsx)