

Phyloinformatics

Assignment-8

Q-1 Construct a phylogenetic tree using PHYLIP software based on any ten keratin protein sequences.

- Design tree using **NJ and ML** methods (default parameters)
- Perform Bootstrapping of the results with default parameters and generate consensus tree accordingly.
- Compare your results.

Report contains 1. Input sequences 2. parameters 3. NJ and ML Tress 4. Bootstrap parameters and consensus tree of NJ and ML

Answer:

So for performing protein analysis we have two methods:-

1.Distance based methods :- It is made using prodist method and tree is designed using neighbour programs which has two methods of tree building:-

- UPGMA
- Neighbour Joining

2.Character based methods:-

- Maximum Likelihood: Where proml is the program used for proteins in ML.
- Parsimony: Which has protpars as protein programs used in Parsimony.

Step 1: Download PHYLIP software and MEGA software from chrome.

PHYLIP:

PHYLIP (the PHYLogeny Inference Package) is a package of programs for inferring phylogenies (evolutionary trees). It is available free over the Internet, and written to work on as many different kinds of computer systems as possible. The source code is distributed (in C), and executables are also distributed. In particular, already-compiled executables are available for Windows (95/98/NT/2000/me/xp/Vista), Mac OS X, and Linux systems. Older executables are also available for Mac OS 8 or 9 systems. Complete documentation is available on documentation files that come with the package.

Methods that are available in the package include **parsimony, distance matrix, and likelihood methods, including bootstrapping and consensus trees**. Data types that can be handled include molecular sequences, gene frequencies, restriction sites and fragments, distance matrices, and discrete characters.

The programs are controlled through a menu, which asks the users which options they want to set, and allows them to start the computation. The data are read into the program from a text file, which the user can prepare using any word processor or text editor (but it is

important that this text file *not* be in the special format of that word processor -- it should instead be in "flat ASCII" or "Text Only" format). Some sequence analysis programs such as the **ClustalW** alignment program can write data files in the PHYLIP format. Most of the programs look for the data in a file called "infile" -- if they do not find this file they then ask the user to type in the file name of the data file.

Output is written onto special files with names like "**outfile**" and "**outtree**". **Trees written onto "outtree" are in the Newick format**, an informal standard agreed to in 1986 by authors of a number of major phylogeny packages.

At this stage we do not have a mouse-windows interface for PHYLIP.

PHYLIP is probably the most widely-distributed phylogeny package. It is the sixth most frequently cited phylogeny package, after MrBayes, PAUP*, RAxML, Phyml, and MEGA. PHYLIP is also the oldest widely-distributed package. It has been in distribution since October, 1980, and has over 30,000 registered users. It is still being updated.

MEGA:

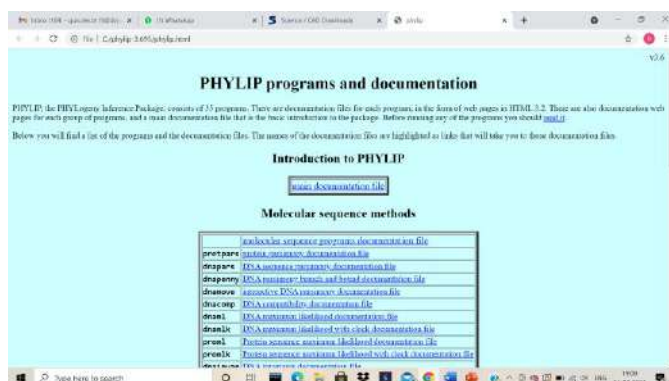
MEGA stands for Molecular Evolutionary Genetic Analysis and the X stands for version 10.

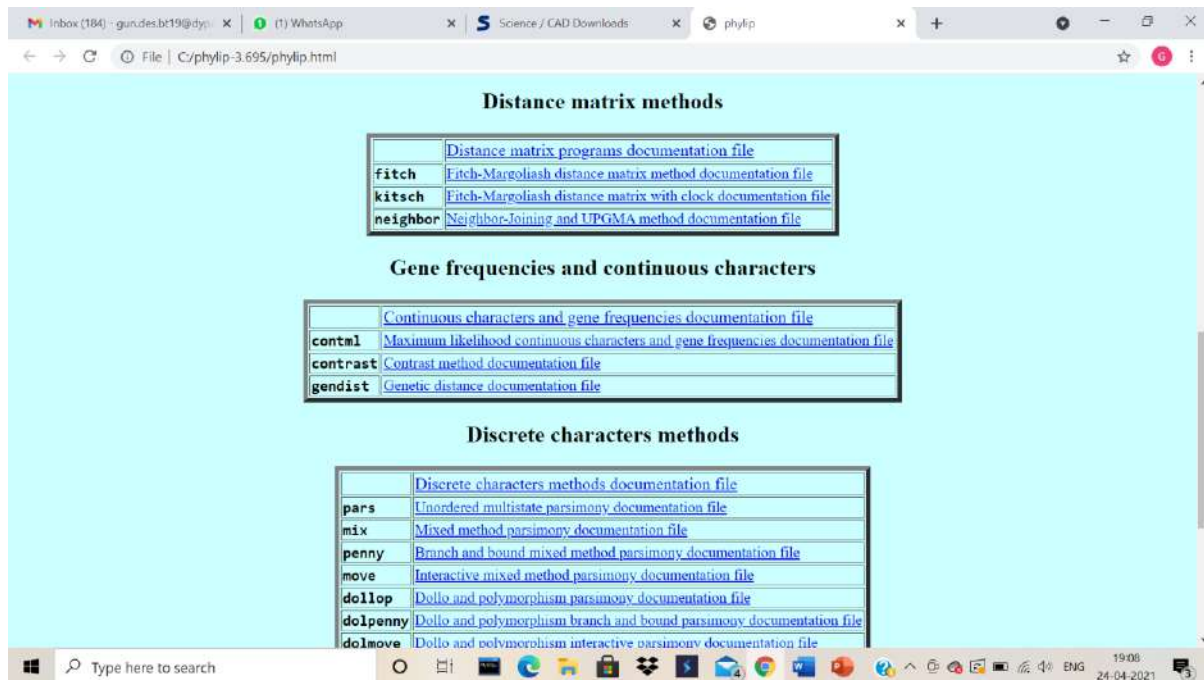
The objective of the MEGA software has been to provide tools for exploring, discovering, and analyzing DNA and protein sequences from an evolutionary perspective. The first version was developed for the limited computational resources that were available on the average personal computer in early 1990s. MEGA1 made many methods of evolutionary analysis easily accessible to the scientific community for research and education. MEGA2 was designed to harness the exponentially greater computing power and a graphical interface of the late 1990's, fulfilling the fast-growing need for more extensive biological sequence analysis and exploration software. It expanded the scope of its predecessor from single gene to genome wide analyses.

Step 2: Extract all the PHYLIP files to the C drive for convenient working.

Step 3: Go to the home page of PHYLIP.

The choice of molecular marker - **Protein analysis.**





To construct a phylogenetic tree using **PHYLIP** software based on any **ten keratin protein sequences**:

Step 1: Go to the home page of uniprot

UniProtKB:

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added. This includes widely accepted biological ontologies, classifications and cross-references, and clear indications of the quality of annotation in the form of evidence attribution of experimental and computational data.

The UniProt Knowledgebase consists of two sections: a section containing manually-annotated records with information extracted from literature and curator-evaluated computational analysis, and a section with computationally analyzed records that await full manual annotation. For the sake of continuity and name recognition, the two sections are referred to as "UniProtKB/Swiss-Prot" (reviewed, manually annotated) and "UniProtKB/TrEMBL" (unreviewed, automatically annotated), respectively.

Step 2: In the search bar type **keratin protein**

The screenshot shows the UniProtKB search results for 'keratin protein'. The page title is 'UniProtKB 2021_02 results'. A search bar at the top contains 'keratin protein'. Below the search bar, there are navigation links: BLAST, Align, Retrieve/ID mapping, Peptide search, and SPARQL. A 'Basket' button is on the right. A large blue box explains that UniProtKB consists of two sections: Reviewed (Swiss-Prot) - Manually annotated and Unreviewed (TrEMBL) - Computationally analyzed. Below this, a 'Filter by' section shows 'Reviewed (511)' and 'Unreviewed (18,961)'. A 'Popular organisms' section lists 'Human (310)'. A table of results is displayed with columns: Entry, Entry name, Protein names, Gene names, Organism, and Length. The table shows three entries: Q9BYR3 (KRA44_HUMAN), P26371 (KRA59_HUMAN), and Q07627 (KRA11_HUMAN). A 'Do you mean keratin' suggestion box is also visible.

Entry	Entry name	Protein names	Gene names	Organism	Length
Q9BYR3	KRA44_HUMAN	Keratin-associated protein 4-4	KRTAP4-4 KAP4.13, KAP4.4, KRTAP4-13, KRTAP4.13, KRTAP4.4	Homo sapiens (Human)	166
P26371	KRA59_HUMAN	Keratin-associated protein 5-9	KRTAP5-9 KAP5.9, KRN1, KRTAP5.9, UHSK1	Homo sapiens (Human)	169
Q07627	KRA11_HUMAN	Keratin-associated protein 11-1	KRTAP11-1 KAP11.1, KAP11.2, KAP11.3, KAP11.4, KAP11.5, KAP11.6, KAP11.7, KAP11.8, KAP11.9, KAP11.10, KAP11.11, KAP11.12, KAP11.13, KAP11.14, KAP11.15, KAP11.16, KAP11.17, KAP11.18, KAP11.19, KAP11.20, KAP11.21, KAP11.22, KAP11.23, KAP11.24, KAP11.25, KAP11.26, KAP11.27, KAP11.28, KAP11.29, KAP11.30, KAP11.31, KAP11.32, KAP11.33, KAP11.34, KAP11.35, KAP11.36, KAP11.37, KAP11.38, KAP11.39, KAP11.40, KAP11.41, KAP11.42, KAP11.43, KAP11.44, KAP11.45, KAP11.46, KAP11.47, KAP11.48, KAP11.49, KAP11.50, KAP11.51, KAP11.52, KAP11.53, KAP11.54, KAP11.55, KAP11.56, KAP11.57, KAP11.58, KAP11.59, KAP11.60, KAP11.61, KAP11.62, KAP11.63, KAP11.64, KAP11.65, KAP11.66, KAP11.67, KAP11.68, KAP11.69, KAP11.70, KAP11.71, KAP11.72, KAP11.73, KAP11.74, KAP11.75, KAP11.76, KAP11.77, KAP11.78, KAP11.79, KAP11.80, KAP11.81, KAP11.82, KAP11.83, KAP11.84, KAP11.85, KAP11.86, KAP11.87, KAP11.88, KAP11.89, KAP11.90, KAP11.91, KAP11.92, KAP11.93, KAP11.94, KAP11.95, KAP11.96, KAP11.97, KAP11.98, KAP11.99, KAP11.100	Homo sapiens (Human)	177

Step 3: Select any 10 keratin protein sequences.

Step 4: Download all the fasta formats.

The screenshot shows the UniProtKB entry page for P60014 (KR10A_HUMAN). The page title is 'UniProtKB - P60014 (KR10A_HUMAN)'. The entry is for 'Keratin-associated protein 10-10' (KRTAP10-10) from 'Homo sapiens (Human)'. The status is 'Reviewed - Annotation score: 5.00 - Experimental evidence at protein level'. The function is described as 'In the hair cortex, hair keratin intermediate filaments are embedded in an interfibrillar matrix, consisting of hair keratin-associated proteins (KRTAP), which are essential for the formation of a rigid and resistant hair shaft through their extensive disulfide bond cross-linking with abundant cysteine residues of hair keratins. The matrix proteins include the high-sulfur and high-glycine-tyrosine keratins.' The GO - Biological process is 'keratinization' (Source: Reactome). The page also includes a 'Function' section with a description and a 'GO - Biological process' section with a list of terms.

Function

In the hair cortex, hair keratin intermediate filaments are embedded in an interfibrillar matrix, consisting of hair keratin-associated proteins (KRTAP), which are essential for the formation of a rigid and resistant hair shaft through their extensive disulfide bond cross-linking with abundant cysteine residues of hair keratins. The matrix proteins include the high-sulfur and high-glycine-tyrosine keratins.

GO - Biological process

- keratinization (Source: Reactome)

Complete GO annotation on QuickGO ...

Enzyme and pathway databases

Sequence status: Complete.

Length: 251
Mass (Da): 25,571
Last modified: November 14, 2003 - V1
Checksum: A128CDB845B74962

BLAST GO

Sequence data in FASTA format:

```

10      20      30      40      50
MAASTNSIGS SACTDSNRWV DCEGCEPC CCAAPSLTL VCTPVSCVSS
60      70      80      90     100
PCQTACEPS ACQSGYTSSC TTPCVQSSC QPCCTSSC QACQVPCV
110     120     130     140     150
VPVCCVPCN KPVCVPTCS ESQPSQSS SQPTCTSS PCQACQVPCV
160     170     180     190     200
CSKVCVPCV CSASTSCQ QSCQACCT ASQPSQSS SLLQVPCV
210     220     230     240     250
TCCVPCVPCV ASASTSCQ CRTASVLL CRPVCSAPAC YSLCSQSS
C

```

Natural variant

Feature key	Position(s)	Description	Actions	Graphical view	Length
Natural variant (VAR_03063)	20	V → D. Corresponds to variant dbSNP:rs2838602	Ensembl		1

Step 5: Copy paste all fasta formats in notepad.

FASTA format sequence:

```

>sp|P60014|K18A1_HUMAN Keratin-associated protein 18-10 OS=Homo sapiens OX=9606 GN=K18-10 PE=1 SV=1
MAASTNSIGS SACTDSNRWV DCEGCEPC CCAAPSLTL VCTPVSCVSS PCQTACEPS ACQSGYTSSC
TTPCVQSSC QPCCTSSC QACQVPCV VPVCCVPCN KPVCVPTCS ESQPSQSS SQPTCTSS PCQACQVPCV
CSKVCVPCV CSASTSCQ QSCQACCT ASQPSQSS SLLQVPCV TCCVPCVPCV ASASTSCQ CRTASVLL
CRPVCSAPAC YSLCSQSS

```

Notepad content:

```

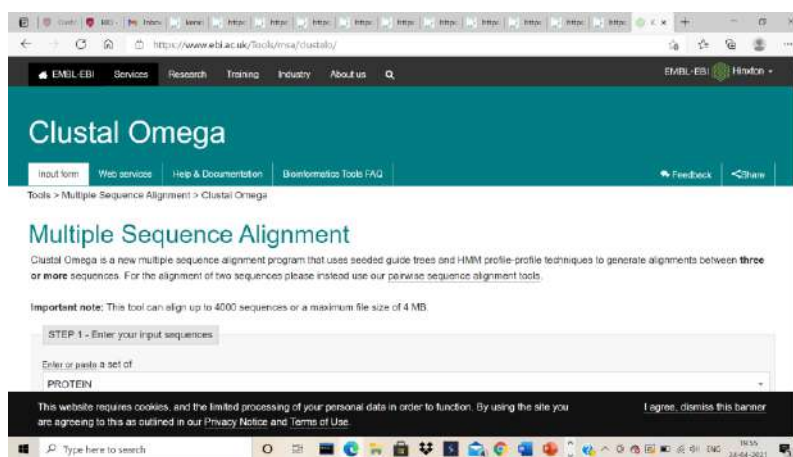
>sp|P60014|K18A1_HUMAN Keratin-associated protein 18-10 OS=Homo sapiens OX=9606 GN=K18-10 PE=1 SV=1
MAASTNSIGS SACTDSNRWV DCEGCEPC CCAAPSLTL VCTPVSCVSS PCQTACEPS ACQSGYTSSC
TTPCVQSSC QPCCTSSC QACQVPCV VPVCCVPCN KPVCVPTCS ESQPSQSS SQPTCTSS PCQACQVPCV
CSKVCVPCV CSASTSCQ QSCQACCT ASQPSQSS SLLQVPCV TCCVPCVPCV ASASTSCQ CRTASVLL
CRPVCSAPAC YSLCSQSS

```

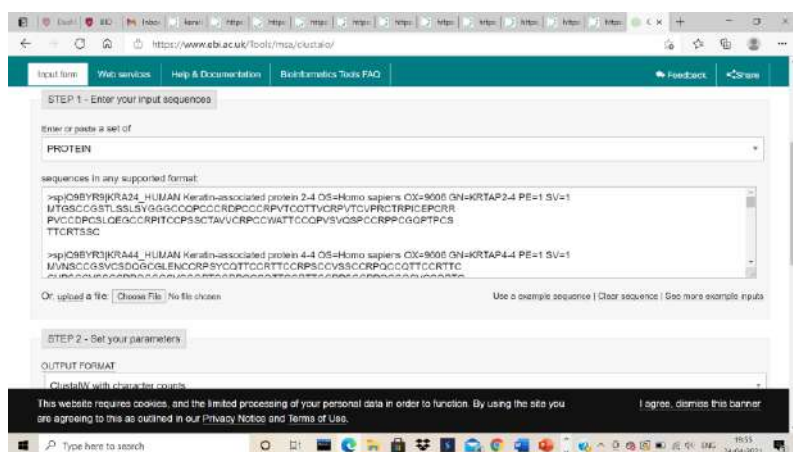

Step 6: Go to Clustal omega home page .

ClustalΩ (Omega): The current standard version.

ClustalΩ (alternatively written as Clustal O and Clustal Omega) is a fast and scalable program written in C and C++ used for multiple sequence alignment. It uses seeded guide trees and a new HMM engine that focuses on two profiles to generate these alignments. The program requires three or more sequences in order to calculate the multiple sequence alignment, for two sequences use pairwise sequence alignment tools (EMBOSS, LALIGN). Clustal Omega is consistency-based and is widely viewed as one of the fastest online implementations of all multiple sequence alignment tools and still ranks high in accuracy, among both consistency-based and matrix-based algorithms.



Step 7: Paste all 10 sequences in the box.



Step 8: Change the output format to PHYLIP.

Or, upload a file: No file chosen

[Use a example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

STEP 2 - Set your parameters

OUTPUT FORMAT

- ClustalW with character counts
- ClustalW
- Pearson/FASTA
- MSF
- NEXUS
- PHYLIP**
- SELEX
- STOCKHOLM
- VIENNA

EMAIL:

TITLE:

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#). [I agree, dismiss this banner](#)

STEP 3 - Submit your job

☒ Be notified by email (Tick this box if you want to be notified by email when the results are available)

EMAIL:

TITLE:

If available, the title will be included in the subject of the notification email and can be used as a way to identify your analysis

If you use this service, please consider citing the following publication: **The EMBL-EBI search and sequence analysis tools APIs in 2019**

Please read the provided [Help & Documentation](#) and [FAQs](#) before seeking help from our support staff. If you have any feedback or encountered any issues please let us know via [EMBL-EBI Support](#). If you plan to use these services during a course please contact us. Read our [Privacy Notice](#) if you are concerned with your privacy and how we handle personal information.

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#). [I agree, dismiss this banner](#)

Result:

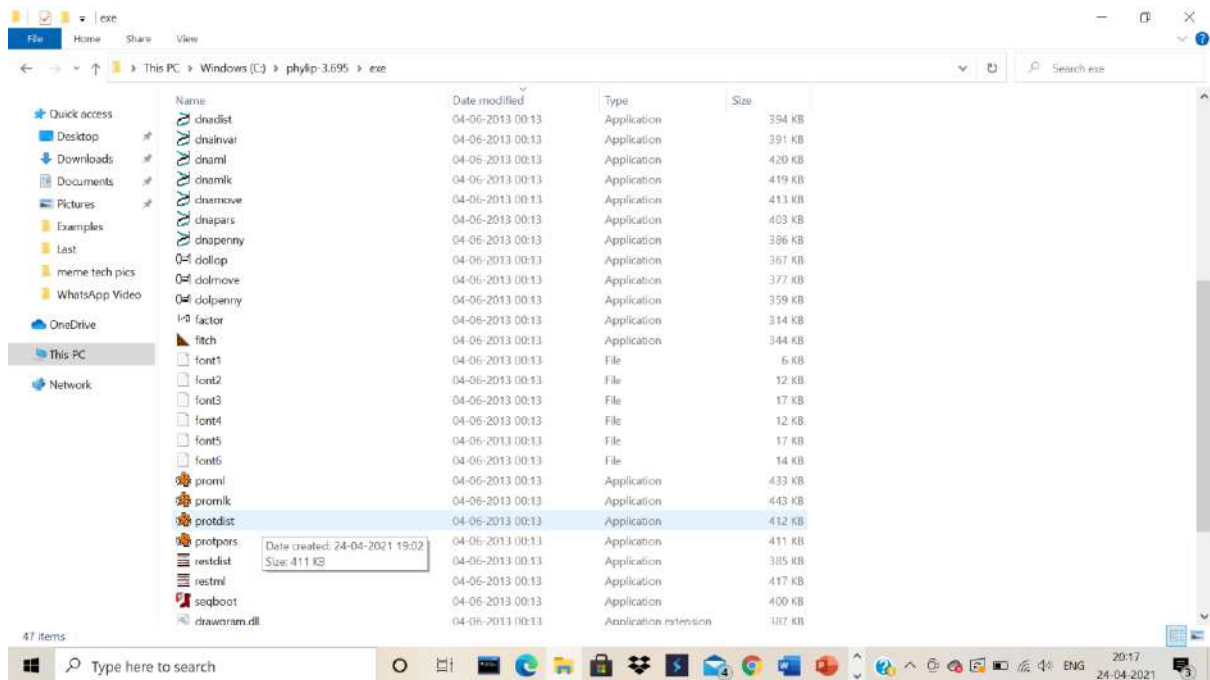
Step 9: Go to Download alignment file.

Step 10: Save this sequences in notepad and than save as in exe folder of PHYLIP file.

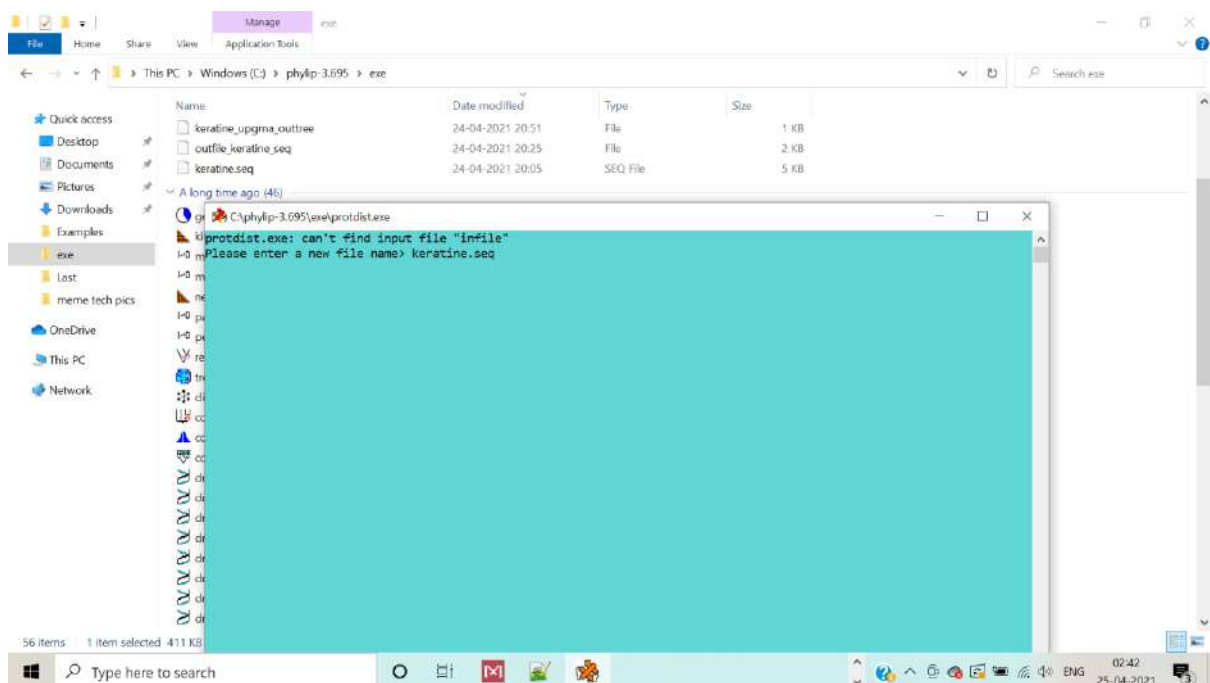
✓ Performing protein analysis using Distance based methods

Obtaining a distance matrix:

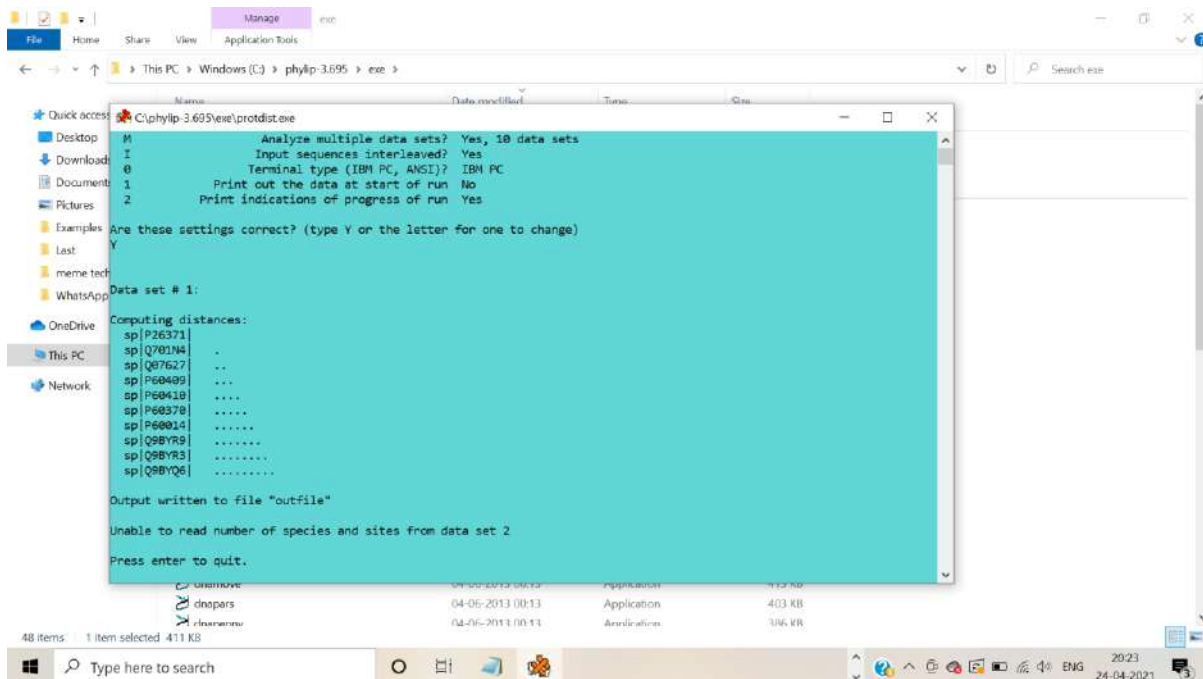
Step 1: Go to protdist



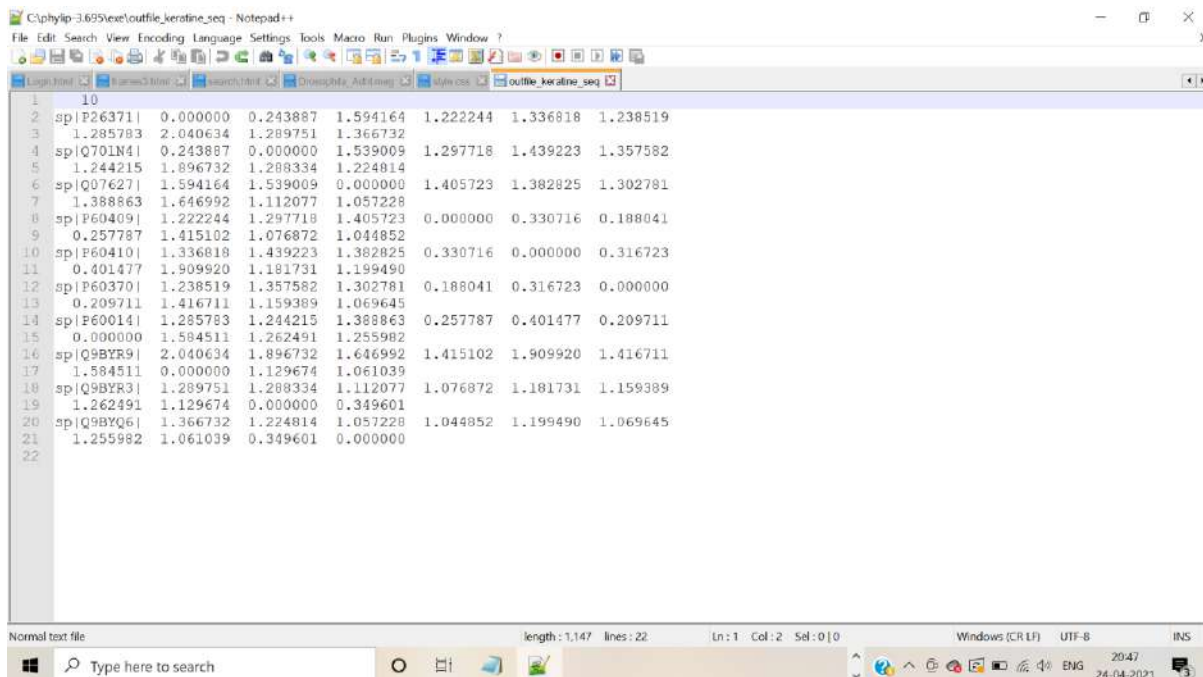
Step 2: Enter your new file name.



Step 3: Press Y and agree all the conditions and enter to exit.



Step 4: Open the output folder and you will get your **Distance matrix**.



✓ **To perform tree building using neighbour- joining method:**

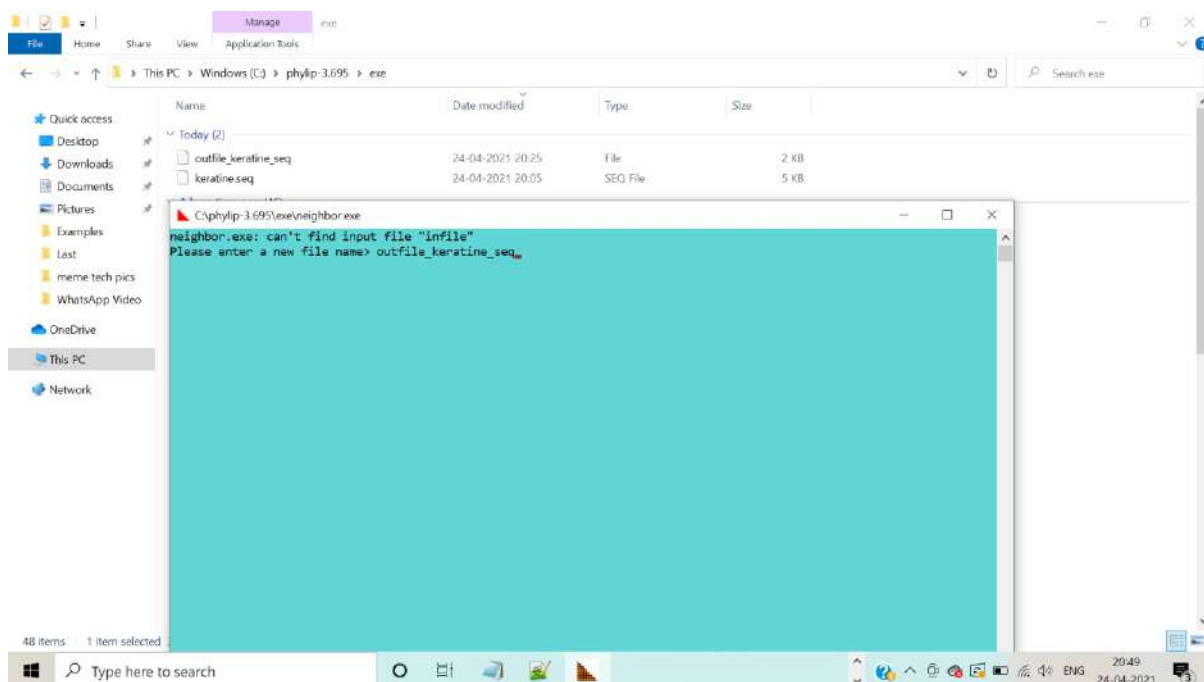
Neighbour- Joining Method:-

This method (Saitou and Nei 1987) is a simplified version of the minimum evolution (ME) method (Rzhetsky and Nei 1992). The ME method uses distance measures that correct for multiple hits at the same sites; it chooses a topology showing the smallest value of the sum of all branches (S) as an estimate of the correct tree. However, construction of an ME tree is time-consuming because, in principle, the S values for all topologies must be evaluated. Because the number of possible topologies (unrooted trees) rapidly increases with the number of taxa, it becomes very difficult to examine all topologies.

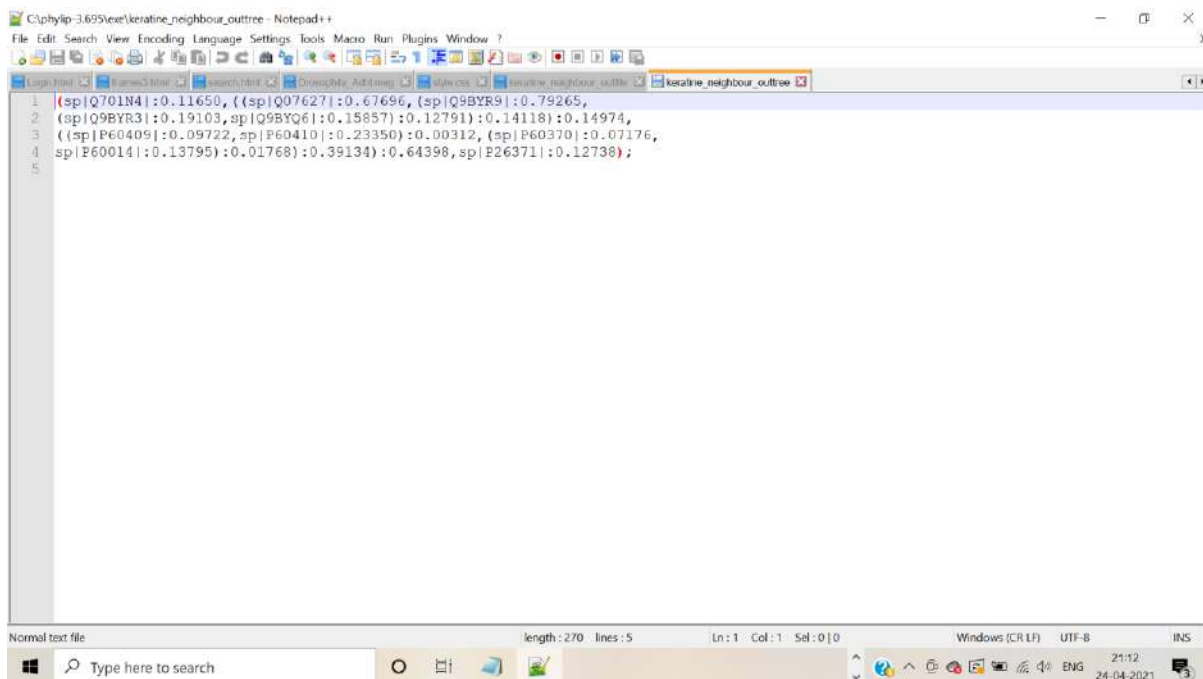
In the case of the NJ method, the S value is not computed for all or many topologies, but the examination of different topologies is embedded in the algorithm, so that only one final tree is produced.

The NJ method produces an unrooted tree because it does not require the assumption of a constant rate of evolution. Finding the root requires an outgroup taxon. In the absence of outgroup taxa, the root is sometimes given at the midpoint of the longest distance connecting two taxa in the tree, which is referred to as mid-point rooting.

Step 1: Enter the file name:



Step 2: Obtain the out tree.

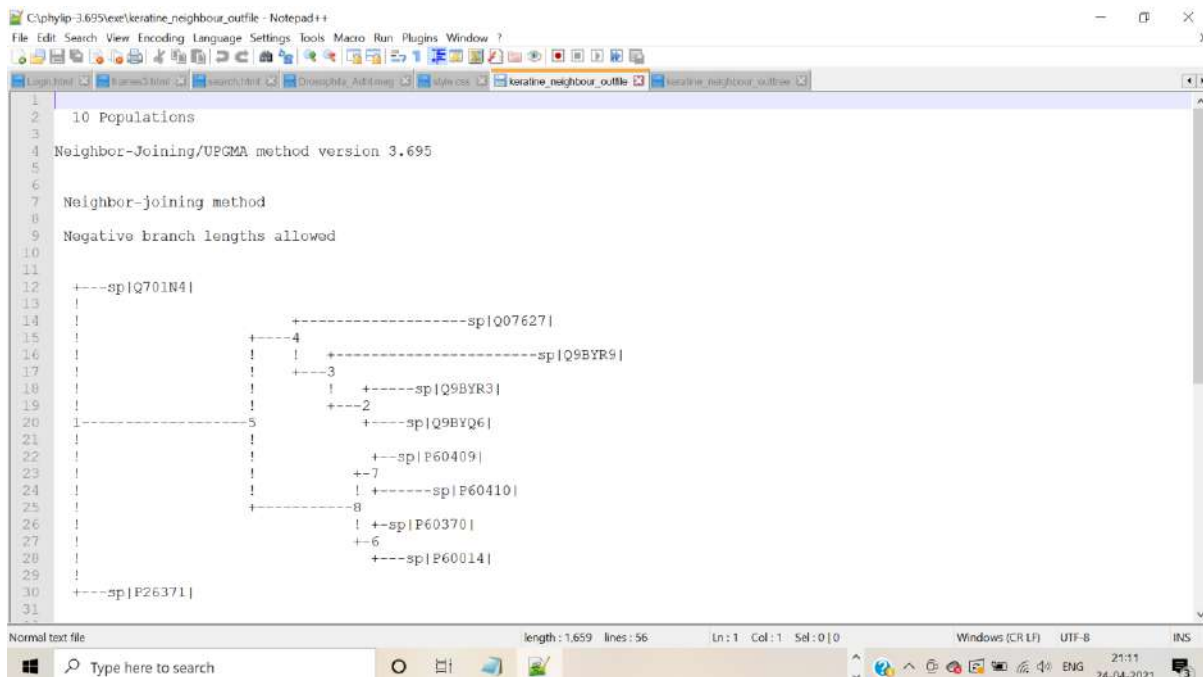


```

1  ((sp|Q701N4|:0.11650, ((sp|Q07627|:0.67696, (sp|Q9BYR9|:0.79265,
2  (sp|Q9BYR3|:0.19103, sp|Q9BYQ6|:0.15857):0.12791):0.14118):0.14974,
3  ((sp|P60409|:0.09722, sp|P60410|:0.23350):0.00312, (sp|P60370|:0.07176,
4  sp|P60014|:0.13795):0.01768):0.39134):0.64398, sp|P26371|:0.12738));
5

```

Step 3: Obtain a Un-rooted tree for Neighbour-joining method



```

1  10 Populations
2
3  Neighbor-Joining/UPGMA method version 3.695
4
5  Neighbor-joining method
6
7  Negative branch lengths allowed
8
9
10
11
12  +---sp|Q701N4|
13  !
14  !
15  !
16  !
17  !
18  !
19  !
20  !
21  !
22  !
23  !
24  !
25  !
26  !
27  !
28  !
29  !
30  !
31  +---sp|P26371|

```

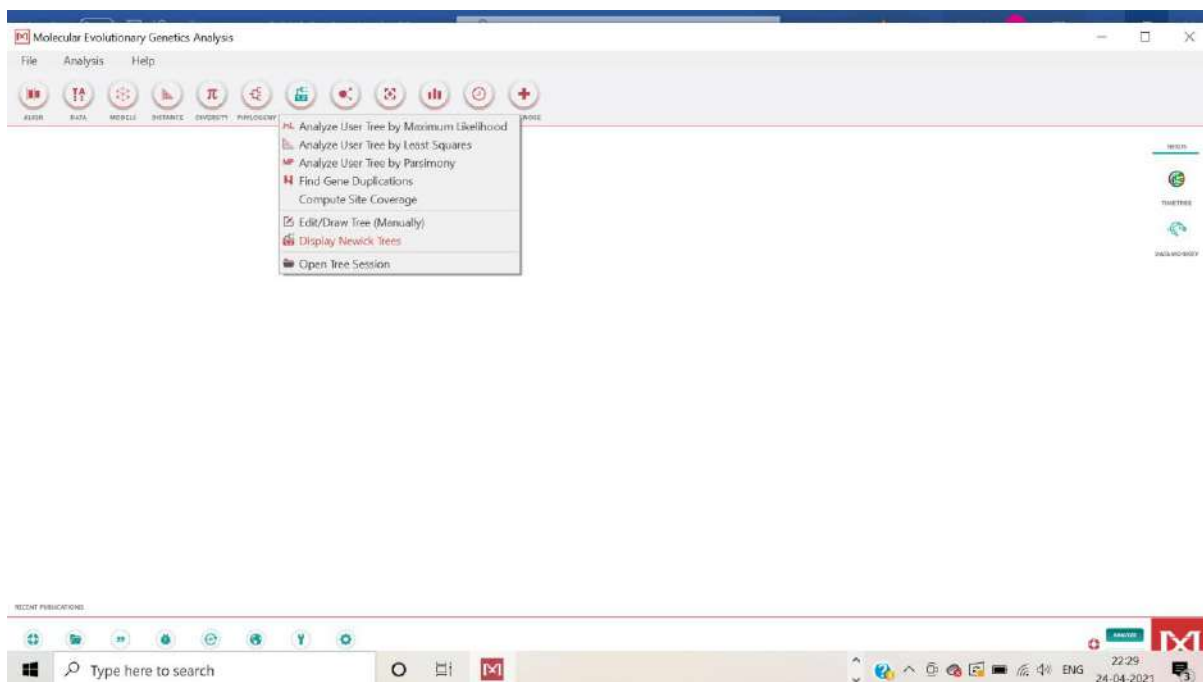
```

26 | +-sp|P60370|
27 |
28 | +-6
29 |
30 | +---sp|P60014|
31 |
32 |
33 | remember: this is an unrooted tree!
34 |
35 | Between      And      Length
36 | -----
37 | 1            sp|Q701N4|  0.11650
38 | 1            5          0.64398
39 | 5            4          0.14974
40 | 4            sp|Q07627| 0.67696
41 | 4            3          0.14118
42 | 3            sp|Q9BYR9|  0.79265
43 | 3            2          0.12791
44 | 2            sp|Q9BYR3|  0.19103
45 | 2            sp|Q9BYQ6|  0.15857
46 | 5            8          0.39134
47 | 8            7          0.00312
48 | 7            sp|P60409|  0.09722
49 | 7            sp|P60410|  0.23350
50 | 8            6          0.01768
51 | 6            sp|P60370|  0.07176
52 | 6            sp|P60014|  0.13795
53 | 1            sp|P26371|  0.12738
54 |
55 |
56 |

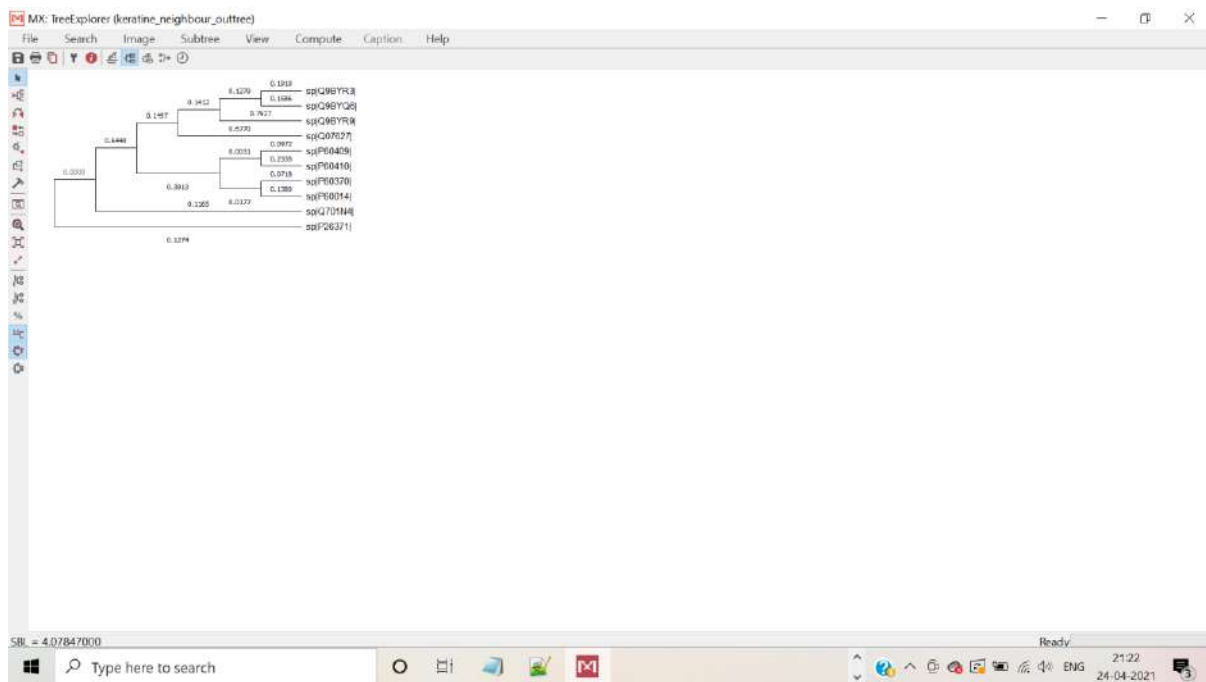
```

Normal text file length: 1,659 lines: 56 Ln: 1 Col: 1 Sel: 0 | 0 Windows (CRLF) UTF-8 INS

Step 4: View the tree in MEGA software, go to user tree and click on Display newick tree.



Step 5: Phylogenetic tree obtained for Neighbour-joining method.

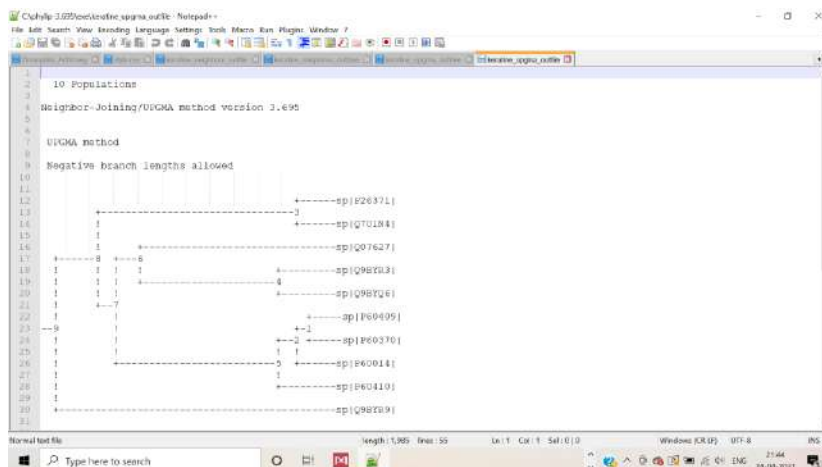


✓ Performing Distance Based Tree Building

UPGMA:-

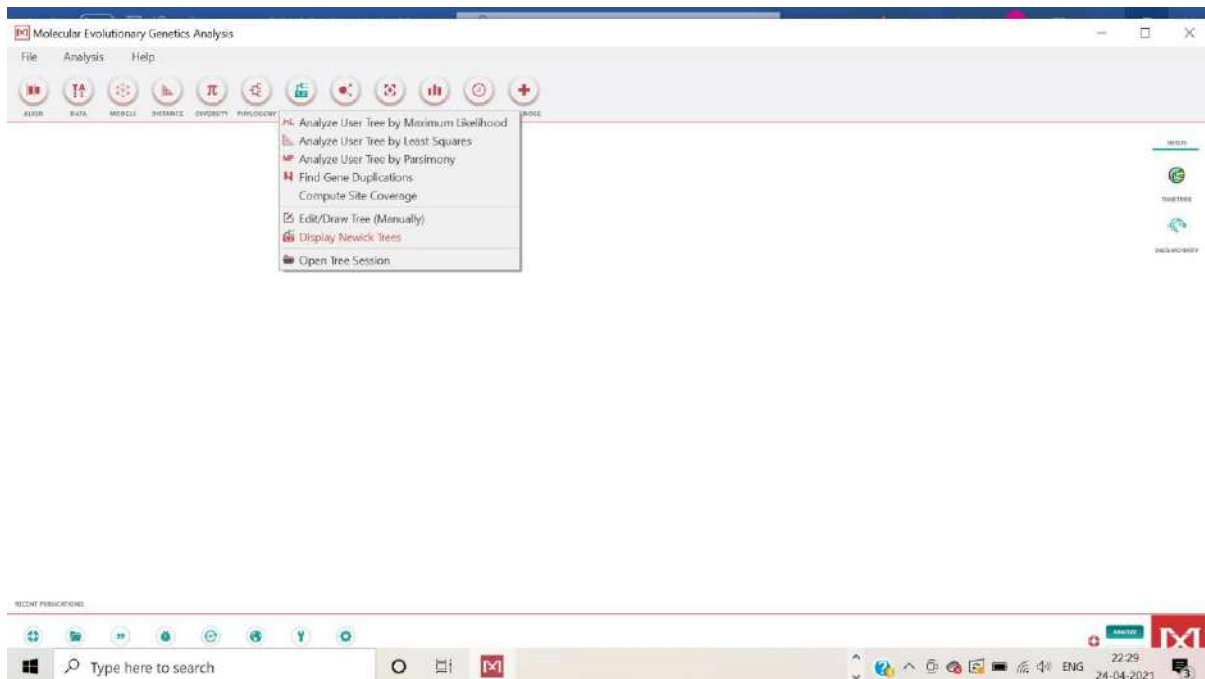
This method assumes that the rate of nucleotide or amino acid substitution is the same for all evolutionary lineages. An interesting aspect of this method is that it produces a tree that mimics a species tree, with the branch lengths for two OTUs being the same after their separation. Because of the assumption of a constant rate of evolution, this method produces a rooted tree, though it is possible to remove the root for certain purposes.

Step 1: Open the saved keratine_upgma_outfile in notepad ++ and you will see Negative branched tree in UPGMA method.

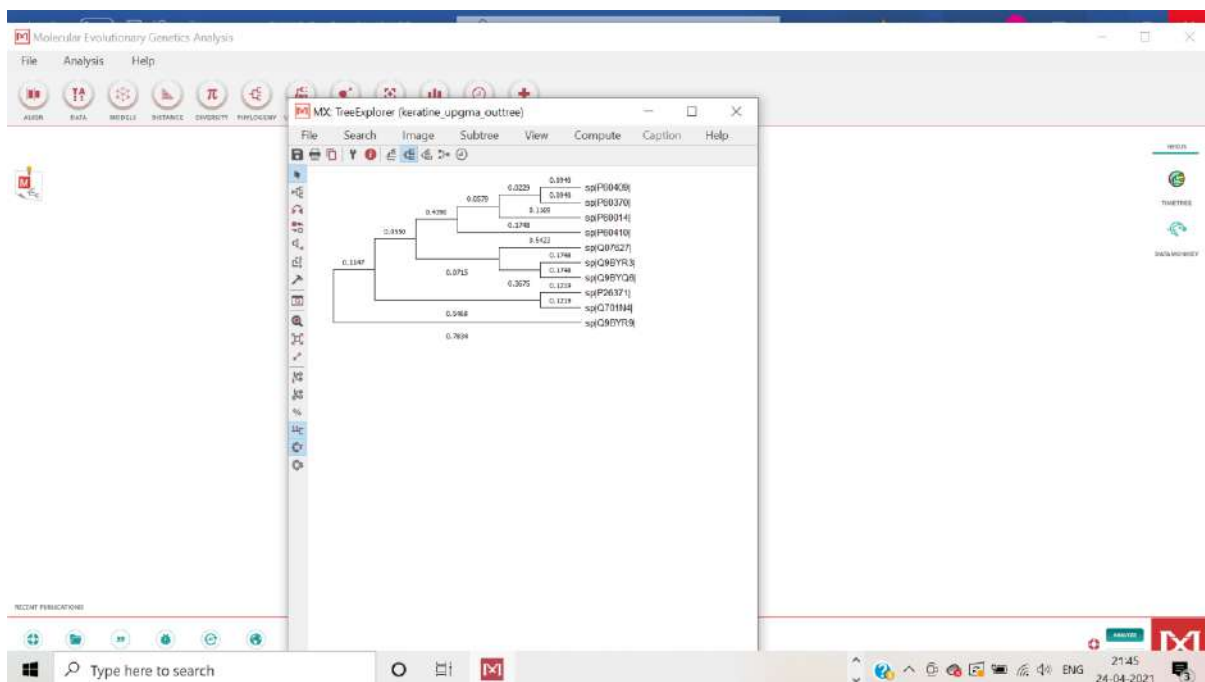


From	To	Length	Height
9	8	0.11066	0.11066
8	3	0.54680	0.66146
3	sp P26371	0.12196	0.78341
3	sp Q70184	0.12196	0.78341
8	7	0.05597	0.16963
7	6	0.07145	0.24108
6	sp Q07627	0.54233	0.78341
6	4	0.28753	0.60961
4	sp Q98703	0.17480	0.78341
4	sp Q98706	0.17480	0.78341
7	5	0.43096	0.60959
5	2	0.05796	0.66833
2	1	0.02285	0.68939
1	sp P60409	0.09402	0.78341
1	sp P60370	0.09402	0.78341
2	sp P60014	0.11697	0.78341
5	sp P60410	0.17482	0.78341
9	sp Q98709	0.78341	0.78341

Step 2: Go to MEGA to obtain a phylogenetic tree. Click on user tree and select Display Newick tree.



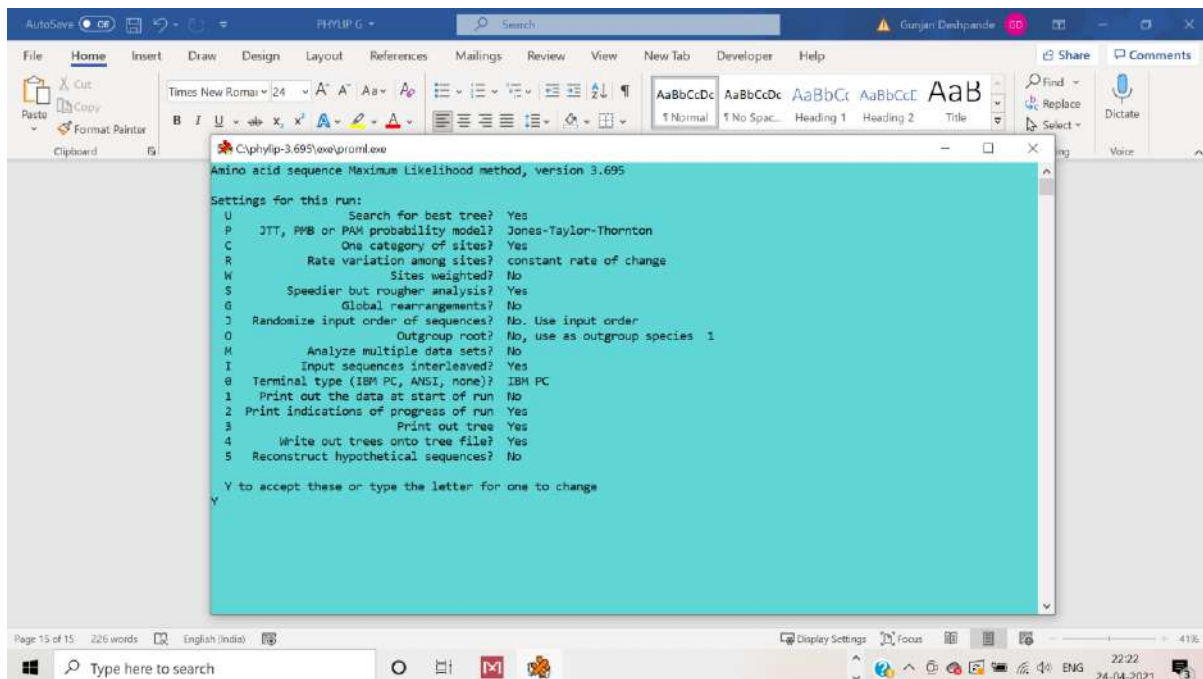
Step 3: UPGMA tree is obtained.



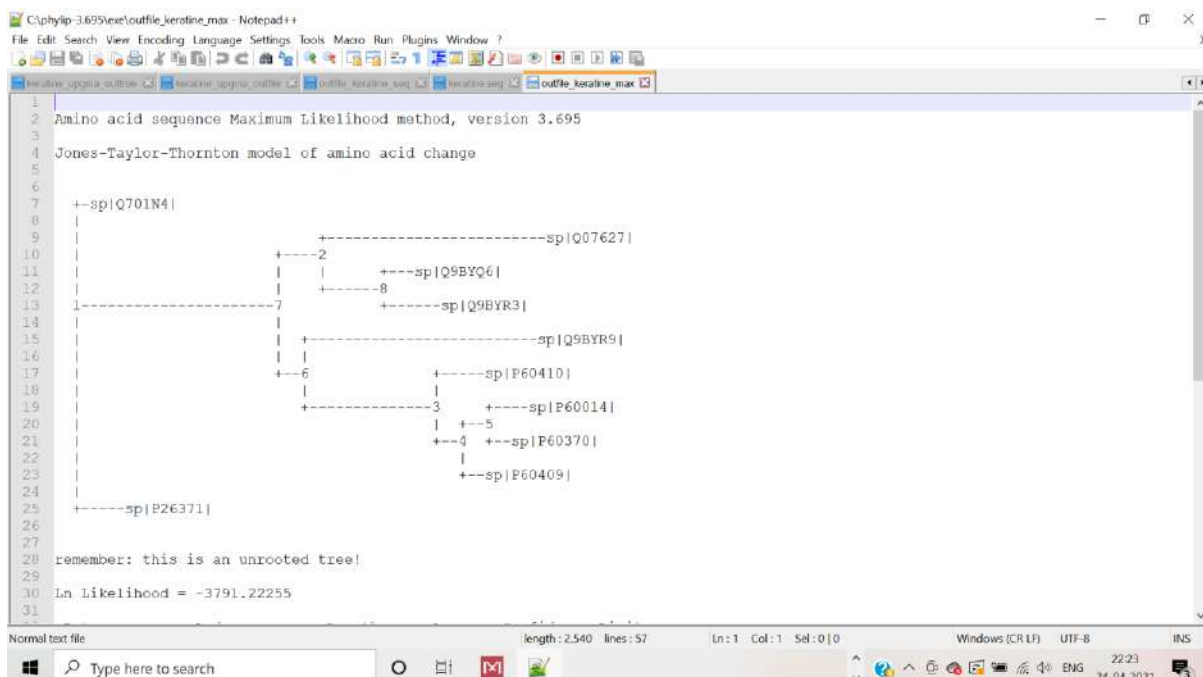
✓ Performing protein analysis using Character based methods:-

Maximum Likelihood: Where proml is the program used for proteins in ML.

Step 1: Open proml and type Y to agree and proceed with the conditions.



Step 2: Jones – Taylor – Thornton model is obtained for the sequence using maximum likelihood method.



```

C:\phylip-3.695\exe\outtree_keratin_max - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
keratin_upper.outtree keratin_upper.outtree outtree_keratin_max outtree_keratin_max
27
28 remember: this is an unrooted tree!
29
30 Ln Likelihood = -3791.22255
31
32 Between:      And      Length      Approx. Confidence Limits
33 -----
34
35 1      sp|P26371|      0.17559      ( 0.08909, 0.26208) **
36 1      sp|Q701N4|      0.05900      ( zero, 0.12361) **
37 1      7      0.75668      ( 0.52311, 0.99025) **
38 7      2      0.15273      ( 0.01834, 0.28712) **
39 2      sp|Q07627|      0.85750      ( 0.61803, 1.09697) **
40 2      8      0.23121      ( 0.09545, 0.36698) **
41 8      sp|Q9BYQ6|      0.13245      ( 0.04740, 0.21749) **
42 8      sp|Q9BYR3|      0.23216      ( 0.13385, 0.33046) **
43 7      6      0.00010      ( zero, 0.08292)
44 6      sp|Q9BYR9|      0.88086      ( 0.61812, 1.14359) **
45 6      3      0.49672      ( 0.33959, 0.65386) **
46 3      sp|P60410|      0.19624      ( 0.12669, 0.26579) **
47 3      4      0.05709      ( 0.00639, 0.10777) **
48 4      5      0.02832      ( zero, 0.06178) *
49 5      sp|P60014|      0.14953      ( 0.09563, 0.20345) **
50 5      sp|P60370|      0.06998      ( 0.03269, 0.10728) **
51 4      sp|P60409|      0.10349      ( 0.05814, 0.14884) **
52
53 * = significantly positive, P < 0.05
54 ** = significantly positive, P < 0.01
55
56
57
Normal text file      length: 2540 lines: 57      Ln: 1 Col: 1 Sel: 0 | 0      Windows (CRLF) UTF-8 INS
Type here to search

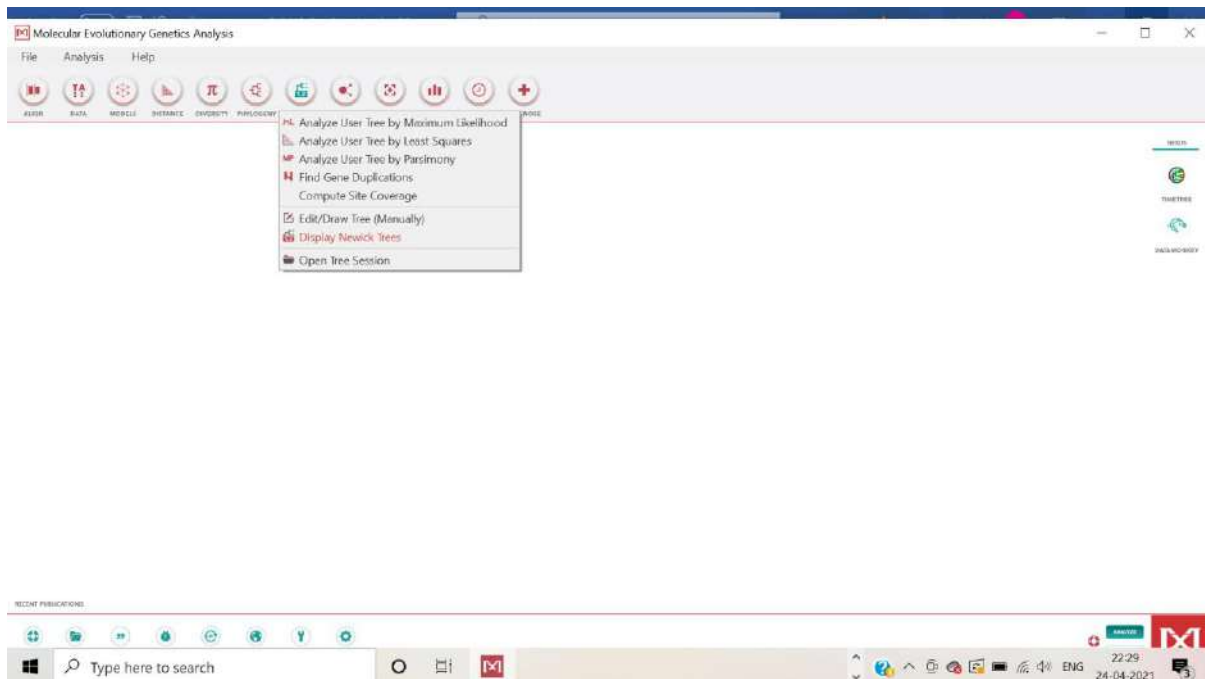
```

```

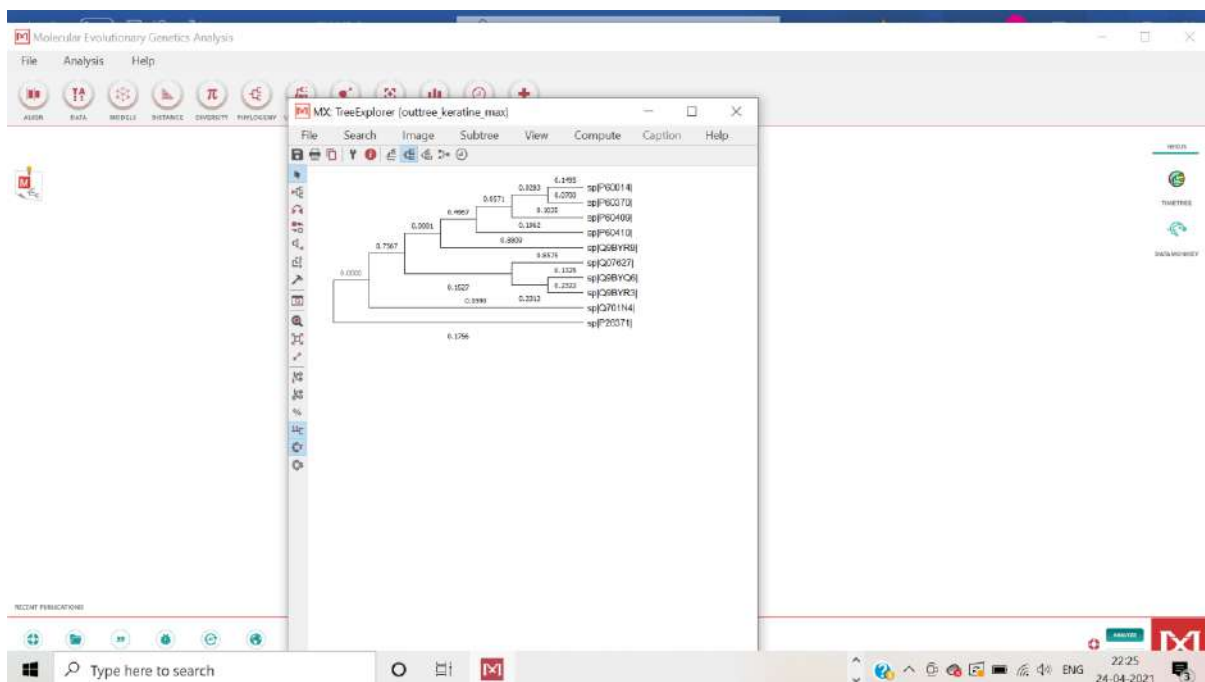
C:\phylip-3.695\exe\outtree_keratin_max - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
keratin_upper.outtree keratin_upper.outtree outtree_keratin_max outtree_keratin_max
1 (sp|Q701N4|:0.05900, ((sp|Q07627|:0.85750, (sp|Q9BYQ6|:0.13245,
2 sp|Q9BYR3|:0.23216):0.23121):0.15273, (sp|Q9BYR9|:0.88086,
3 (sp|P60410|:0.19624, ((sp|P60014|:0.14953, sp|P60370|:0.06998):0.02832,
4 sp|P60409|:0.10349):0.05709):0.49672):0.00010):0.75668,
5 sp|P26371|:0.17559);
6
Normal text file      length: 272 lines: 6      Ln: 1 Col: 1 Sel: 0 | 0      Windows (CRLF) UTF-8 INS
Type here to search

```


Step 3: Go to MEGA to obtain a phylogenetic tree. Click on user tree and select Display Newick tree.



Step 4: Phylogenetic tree is obtained using maximum likelihood method.



- ✓ Performing protein analysis using Character based methods:-

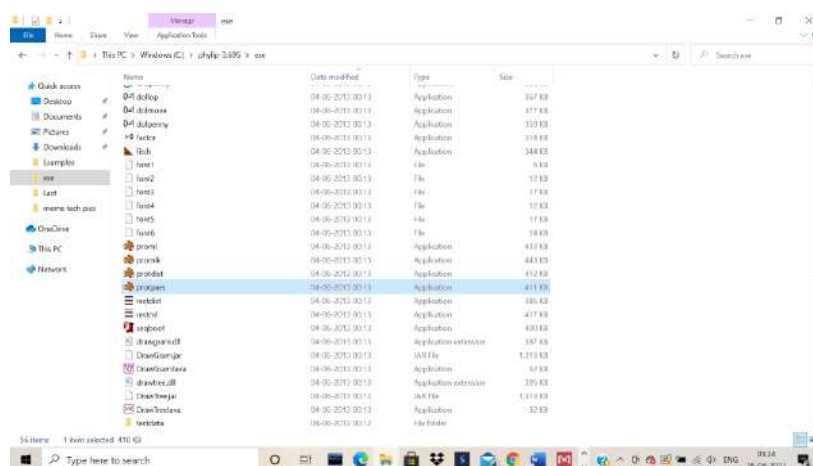
Parsimony: Which has protpars as protein programs used in Parsimony.

This will test the tree which you provide, and report on how accurate the tree is in relation to the data file you have open. The best tree with this method will be the one with the least evolutionary change required.

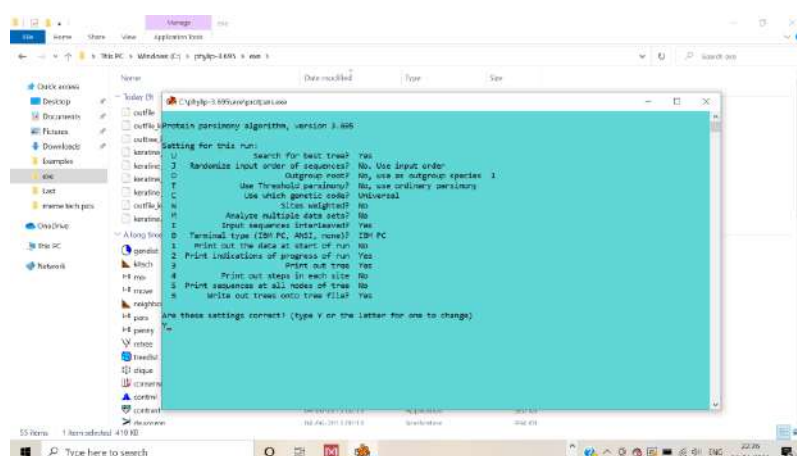
Reconstruction of the evolutionary history of genes and species is currently one of the most important subjects in molecular evolution. If reliable phylogenies are produced, they will shed light on the sequence of evolutionary events that generated the present day diversity of genes and species and help us to understand the mechanisms of evolution as well as the history of organisms.

Phylogenetic relationships of genes or organisms usually are presented in a treelike form with a root, which is called a *rooted tree*. It also is possible to draw a tree without a root, which is called an *unrooted tree*. The branching pattern of a tree is called a *topology*.

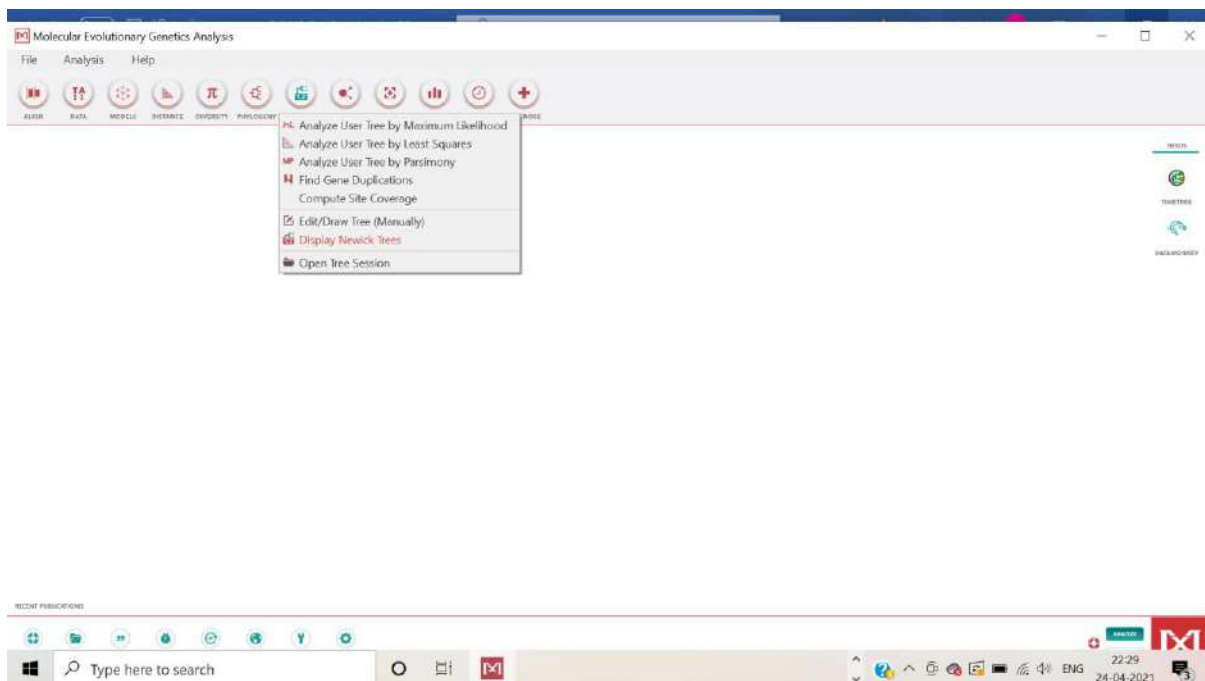
Step 1 : Open the protpars document.



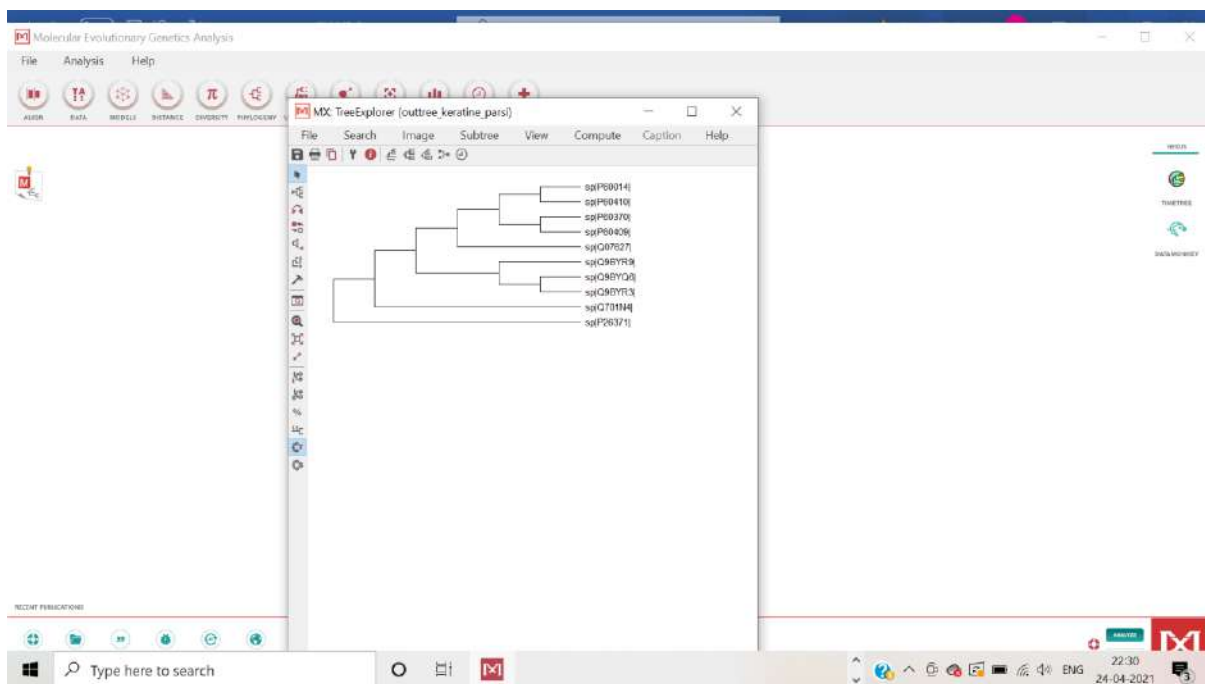
Step 2: Press Y and accept all the parameters.



Step 5: Go to MEGA and click on user tree and select Display Newick Tree.



Step 6: You will get the phylogenetic tree for parsimony method.

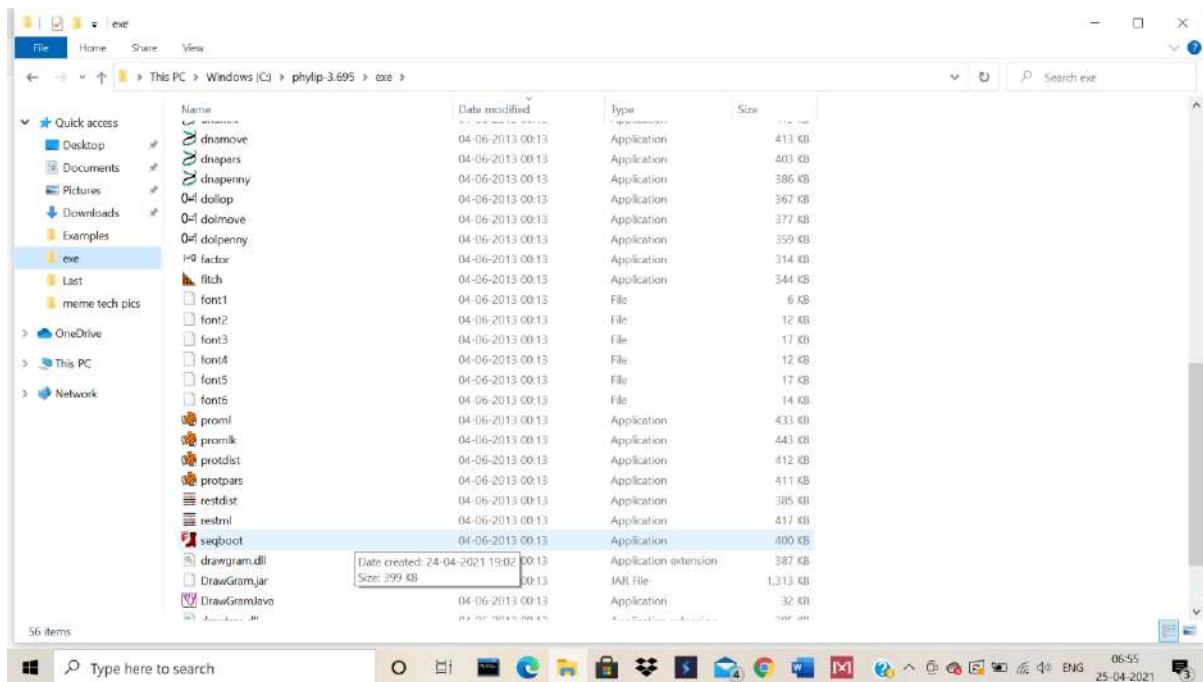


✓ Performing Bootstrapping from the results of NJ and ML methods:

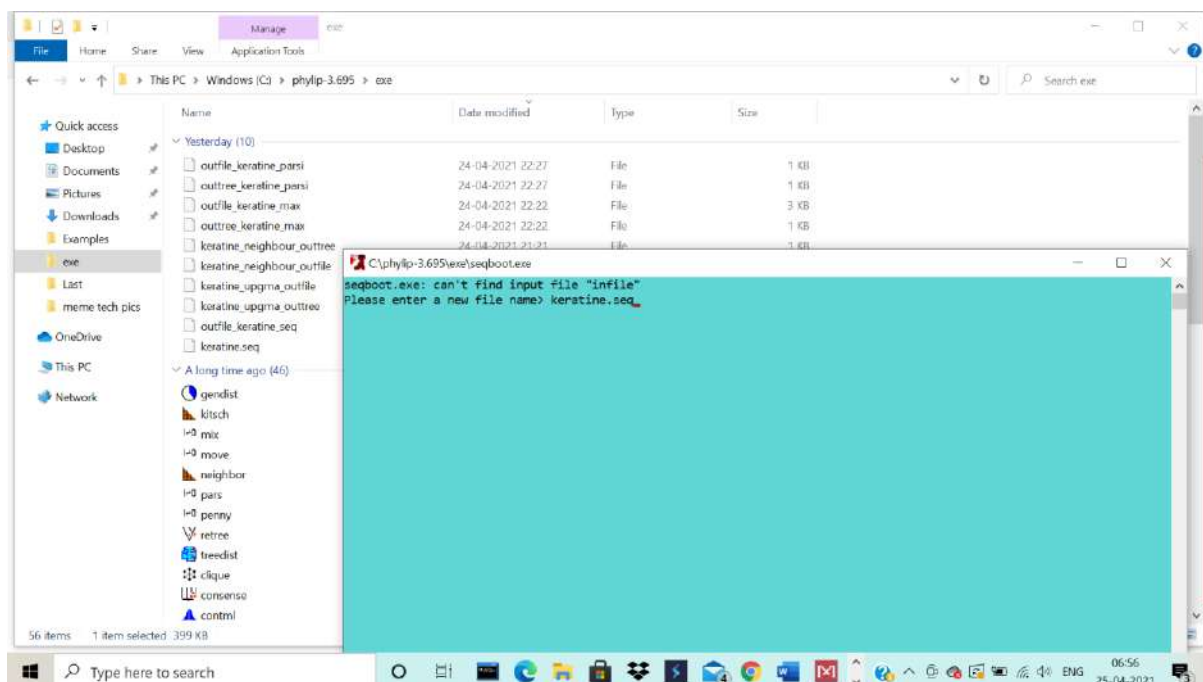
Bootstrapping:- It is basically a type of alignment which is already done , but we are actually reshuffling the positions of the sequences .i.e. certain positions are repeated, certain positions are deleted in short the shuffling of alignment. Bootstrap basically is used for building up the confidence values of the tree.

NJ:

Step 1: Select seqboot from exe folder.



Step 2: Enter the saved file name = keratine.seq




```

C:\phylip-3.695\seqboot.exe
C
  Read categories of sites? No
  Write out data sets or just weights? Data sets
  Input sequences interleaved? Yes
  Terminal type (IBM PC, ANSI, none)? IBM PC
  Print out the data at start of run No
  Print indications of progress of run Yes

Y to accept these or type the letter for one to change

Random number seed (must be odd)?
5

completed replicate number 10
completed replicate number 20
completed replicate number 30
completed replicate number 40
completed replicate number 50
completed replicate number 60
completed replicate number 70
completed replicate number 80
completed replicate number 90
completed replicate number 100

Output written to file "outfile"

Done.

Press enter to quit.

```

Step 3: Once opened the file in notepad ++ you will get 100 different alignment sets for the sequences.

```

C:\phylip-3.695\seq_boot.outfile - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
1 10 411
2 sp|P26371| ----- MGSSG GGGGGGGGS CGG----- RRGCGSCCA
3 sp|Q701N4| ----- MGSSR RRRGGGGGS CGGSGGGCG CGGCGSCCV
4 sp|Q07627| ----- MAA ASSTMSDDL LLLSYGLLPS CD----- SSWQQ DDDDCPSCC
5 sp|P60409| ----- MAA ASSTMSDDL LLLSYGLLPS CD----- SSWQQ DDDDCPSCC
6 sp|P60410| MDACRTTIAA ASSTMSDDL VVGVHVSP -T----- SSWQQ DDDDCPSCC
7 sp|P60370| ----- MAA ACCTMS-- -- -- -- SSWRR DDDDCPSCC
8 sp|P60014| ----- MAA ASSTMS-- -- -- -- SSWRR VDDDCPSCC
9 sp|Q9BYR9| ----- MGG GSSCCLLS-- -- -- -- SYYG GGGCCGCC
10 sp|Q9BYR3| ----- MNM NSSCCCHH-- -- -- -- GCGG EE--- NCCR
11 sp|Q9BYQ6| ----- MNM NSSCCCHH-- -- -- -- GCGG DDLQCQCCR
12
13 PPVCCPVVP CCSCSSSSG GK----- CC GSGGGGSG GCGCCGCC
14 PPVCCPVVP CCSCSSSSG GSCCGKKG CGGSGGKCG GSGGGGSG GCGCCGCC
15 ----- MATS CCG----- --FSCSTSG TCCCCCQQP
16 PPFC-CPASL CCTPPSSSV VSSSSCRIT CESEF--CQ SGCSCTTPS CCCCCCQQL
17 PPRS-CSSPS CCAPPPPPPL LALLCPSSE CESEF--CQ SGCSCTTPS CCCCCCQQP
18 PPFC--TATL CCTPPSSSV VSSSSCQAA CESEF--CQ SGCSCTTPS C--CCQQP
19 PPCC-CPATL CCTPPSSSV VSSSSCQAA CESEF--CQ SGYSCITTP CYYYYCQQP
20 DDPCCPVVP CCRPPPTTV VF----- -- -- --
21 PPSY-CTTT CCRPPCCCV VS----- -- -- --
22 PPSC-CTTT CCRPPCCCV VS----- -- -- --
23
24 PCCSCSSSSG GSSCCQCY YCCSSCSCS CCCCPRRS SCCCCQGS CCKPPCCCC
25 PCCSCSSSSG GSSCCQSCS CCCCCSSG CCCC----- SCCCCQGS CCKPPCCCC
26 SCCCCTTSC QPRCCCTSS SCCCCCTSS FFCCF----- GGG CDSSSCCCC
27 AGCCSSSSG QACCVPCV VVCCCPVV YCCCGDSS SCCCCQGS CQSAACCCC
28 AGCCSSSSG QACCVPCV VVCCCPVV YCCCGDSS SCCCCQGS CQSAACCCC
29 AGCCSSSSG QACCVPCV VVCCCPVV YCCCGDSS SCCCCQGS CQSAACCCC
30 DCCSCSSSSG QACCVPCV VVCCCPVV YCCCGDSS SCCCCQGS CQSAACCCC
31 -----RT RPICBP-RV VVCCCPCCS SS----- -- -- --

```

```

C:\phylip-3.695\exe\seq_boot_outfile - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
seq_boot_outfile
3125 sp1P604091 -----AAA ATMSVSSSS LLYYGSRLG GSSS-----
3126 sp1P604101 MADDTRTAAA ATMSVSSSS VVHHVSRSS SS-C-----
3127 sp1P603701 -----AAA ATMSVSSSS -----A-----
3128 sp1P600141 -----AAA ATMSISSSS -----A-----
3129 sp1Q9BYR91 -----TTT GCCCGSLSS -----
3130 sp1Q9BYR31 -----VVV NCCCGSCSS -----
3131 sp1Q9BYQ61 -----VVV NCCCGSCSS -----
3132 -----CGG SSCCVYOCK VCCCVVAAS SSSS-----RRGSS
3133 CSSCCCGG RRCCYVCOCK VCCCVVAAS STTSCCGS SKKCCSSGG SSSRGGGSS
3135 -----ACCCCTTFFG G-----
3136 -SSSVVCCP SSCCPC-CA APPPCSVVT TVVSSSSPC CRKCCPP--SSP--
3137 -SSSVVCCP SSCCPC-CA SSCCPCPCA AAAAALLVA APPCCPP--SSP--
3138 -SSSVVCCP SSCCPC-G APPPCTTVT TVVSSSSPC CQCCCP--SSP--
3139 -SSSVVCCP SSCCPC-CA APPPSTTVT TVVSSSSPC CQCCCP--SSA--
3140 -SSSGGCGG CCCCPCCCR VTTCCTTVR RVVVT-----
3141 -GGGLLL--NCCCSY-CQ TCCCTTCR RSSC-----
3142 -GGRRRCQG TCCSCC-CE TCCCTTYR RSSC-----
3143 -----
3144 SSCGGGSGK GCGCGGGGS SCCSCKPCS SGGSSSSSC QCCSCKYCS SSQSSCCG
3145 SCGGGSGK GCGCGGGGS SSSSCKPCS SGGSSSSSC QSSSCKKCC CQSSSSCCG
3146 -----FSS CSTTGTTCG GSSCCPCT TSQPPPC ETTSCGSC CQSSSFFF
3147 -QSSSGGTS SCTTSCCGG QSSSCLACC SPPQQQACC VPPVCCVCC CCKPVVYYY
3148 -QSSSGGTD SCTTSCCGG QSSSCFACC SPPQQQACC VPPVCCVCC CCKPVVYYY
3149 -QSSSGGTS SCTTSCCGG -----CPACC SPPQQQACC VPPVCCVCC CCLP-----
3150 -QSSSGGTS SCTTSCCGG QSSSCFACC SPPQQQACC VPPVCCVCC CCLP-----
3151 -----CCRPPTCC P-----CCRVCC CCRPCCSSS
3152 -----CCRPPTCC P-----CCRVCC CCRPCCSSS
3153 -----CCRPPTCC P-----CCRVCC CCRPCCSSS
3154 -----CCRPPTCC P-----CCRVCC CCRPCCSSS
3155 CCKKPPRGG GSSSCCGG QSSKPPPPC SCGSS-----
Normal text file length: 5,39,700 lines: 7,801 Ln: 1 Col: 1 Sel: 0|0 Windows (CRLF) UTF-8 INS
Type here to search

```

```

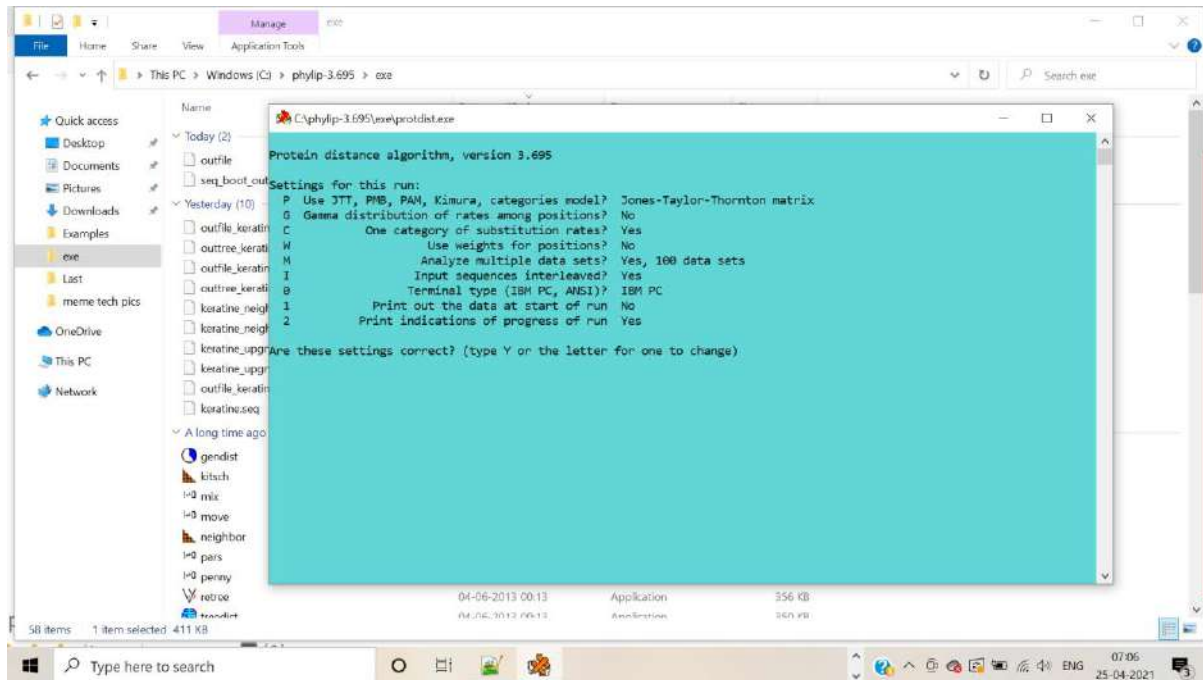
C:\phylip-3.695\exe\seq_boot_outfile - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
seq_boot_outfile
6574 -----GGGDDDD CCGCTCCRP CCGCCCCCR TCCRV-----
6575 -----
6576 GSCGGGGS QCCSCCPCP CCGSSSGCC CCGSSSQSC CCGSSQCC CCKPPPPSS
6577 GSCGGGGS QSSSSCCPCP CCGSSSGCC CCGSSSQSC CCGSSQCC CCKPPPPSS
6578 -FSTTTCG SSSSCCSCS CCGTSSSCC CCGPPEESC CCGCCTTF FFFGFFFFF
6579 CCGCTTTCG QSSSSCCAAA CCGSSSPCC CCGQAVVVC CCGCCKPPY YYYVPPPPC
6580 CCGCTTTCG QSSSSCCAAA CCGSSSPCC CCGQAVVVC CCGCCKPPY YYYVPPPPC
6581 CCGCTTTCG -----CCAAA CCGSSSPCC CCGQAVVVC CCGCCKPPY YYYVPPPPC
6582 GYCTTTTYQ QSSSSCCDD CCGSSSPCC CCGQAVVVC CCGCCKPPY YYYVPPPPC
6583 -----RRCTT TTRFPFP CCGCCKPPY YYYVPPPPC
6584 -----SSCC CCGPQQTC CCGCCKPPY YYYVPPPPC
6585 -----SSCC CCGPQQTC CCGCCKPPY YYYVPPPPC
6586 -----SSCC CCGPQQTC CCGCCKPPY YYYVPPPPC
6587 SSSGSSCGG SCGCCPCG-----
6588 -----
6589 FFS-----GG TCCDSCEC QPCTTTCCT GGGI--GGG-----EESS GV-----
6590 CCGSSSQGQ SCQQACPC VPPCPPPSI SSSCCSSSC QVCEPPPPS GISSSSSCT
6591 -----CPPCCS II-----CSSGGGAAS
6592 CCGSSSQGQ SCQQACPC VPPCPPPSI SSSCCSSSC QVCEPPPPS GISSSSSCT
6593 -----PPCPPPFP TT-----CSSSEEDDS
6594 -----PPCPPPFP TT-----CSSSEEDDS
6595 -----PPCPPPFP TT-----CSSSEEDDS
6596 -----PPCPPPFP TT-----CSSSEEDDS
6597 -----PPCPPPFP TT-----CSSSEEDDS
6598 -CCSSCKKP PQS-----RC VPPQI-----
6599 -CCSSCKKP PQS-----NC VPPQI-----
6600 -STIRCCRP PVVETLPPC VVVC-----CC-----
6601 TCCQSCCKP PTTSPQAC VPPKCCVET TSSGGGSCP ACCTSSSQQ QGCVVPPCC
6602 SCCQSCQPS STTFPPQAC VPPK-----
6603 SCCQSCQPS PTTSPQAC VPPK-----
6604 SCCQSCQPS PTTSPQAC VPPK-----
Normal text file length: 5,39,700 lines: 7,801 Ln: 1 Col: 1 Sel: 0|0 Windows (CRLF) UTF-8 INS
Type here to search

```

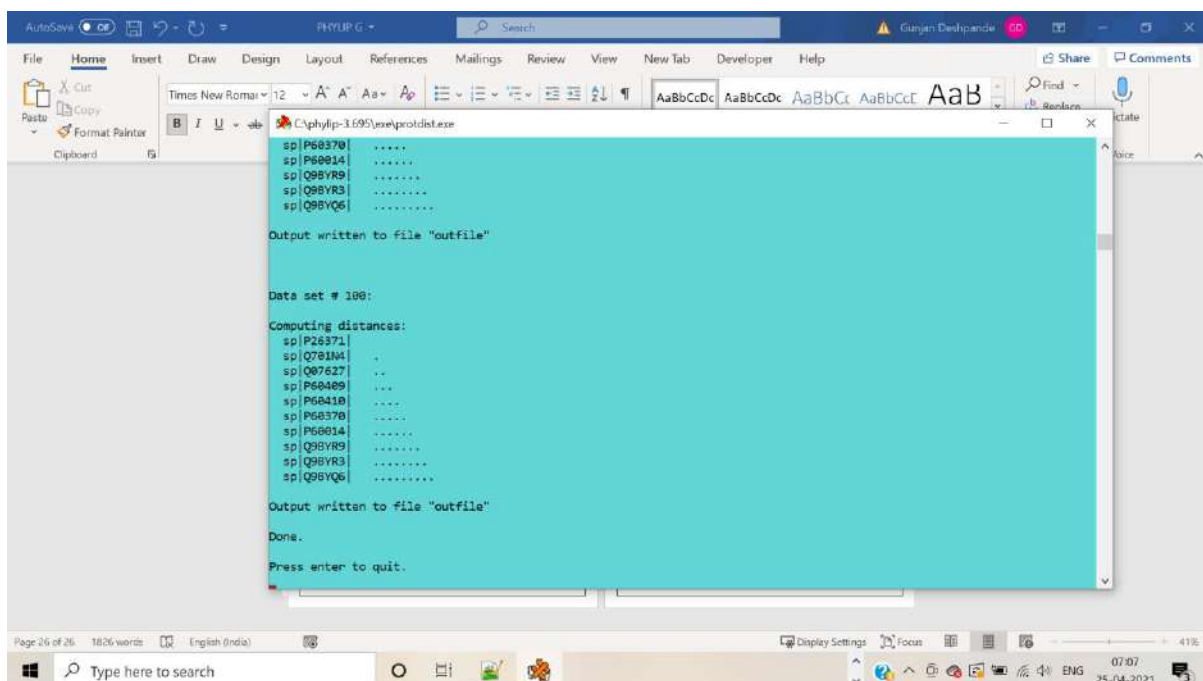
✓ Performing Bootstrapping by distance method :

Creating distance for 100 datasets:

Step 1: Select the prodist file and Modify/Change the parameters.



Step 2: Observe distance for 100 datasets

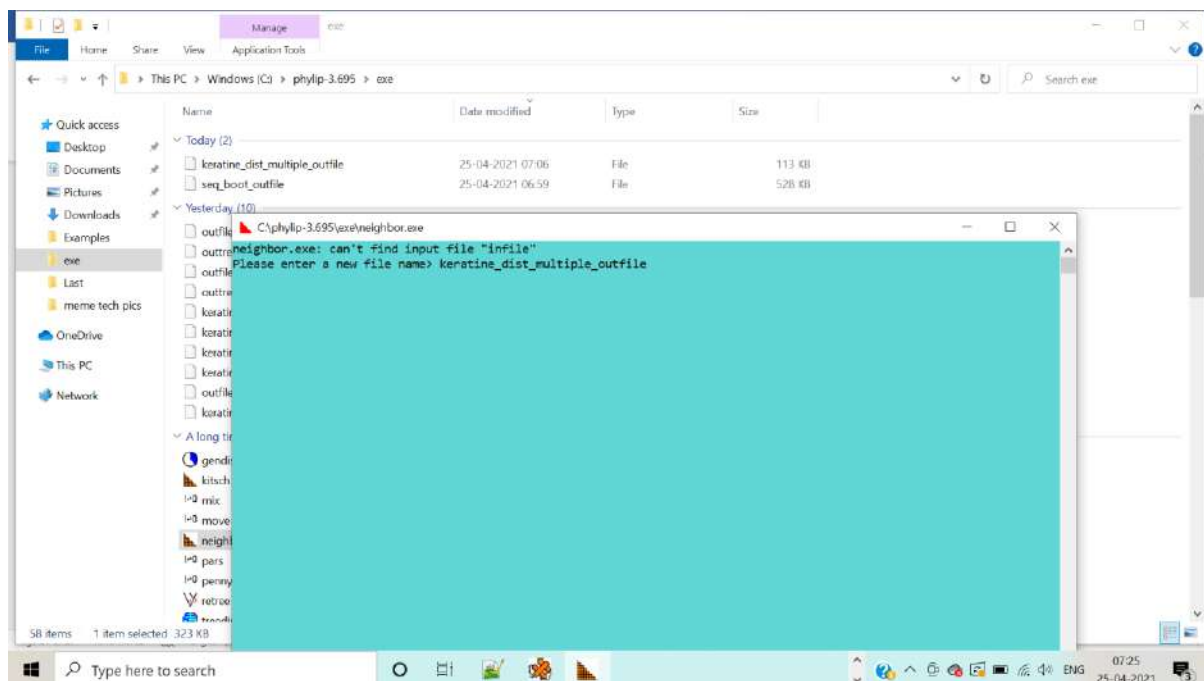


Step 3: Obtaining 100 distance matrices

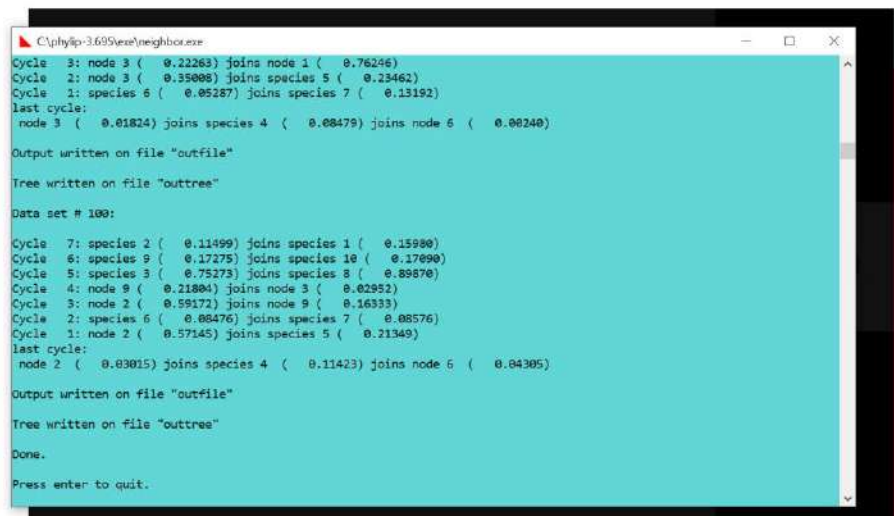
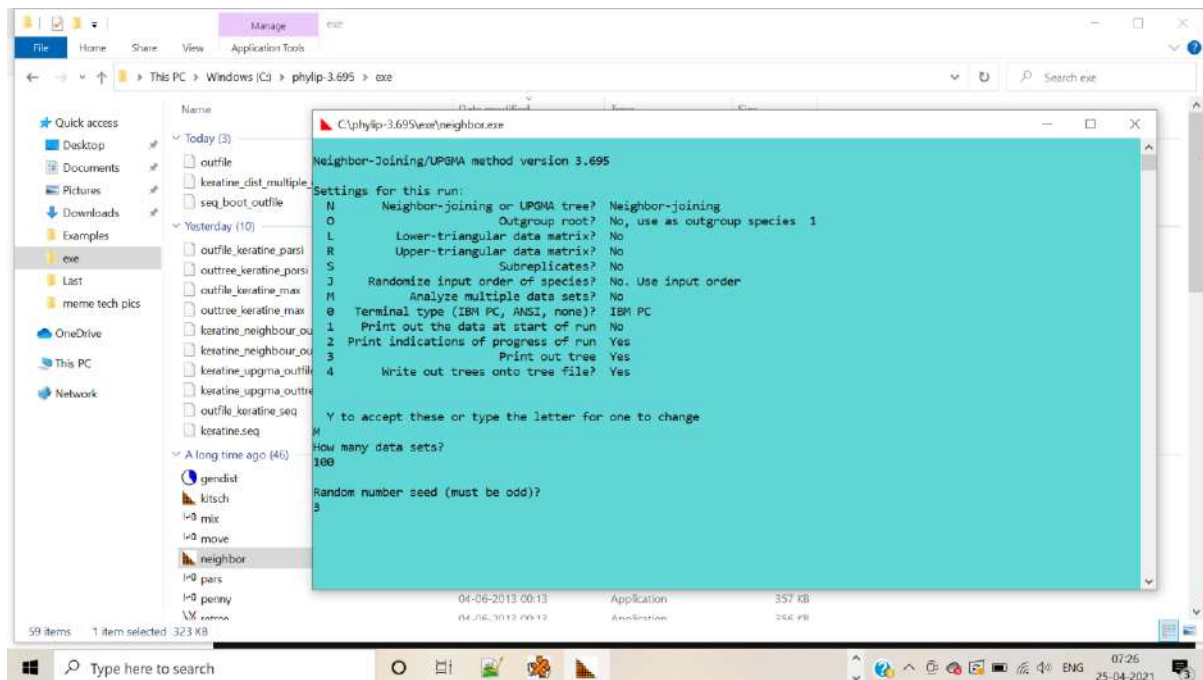
```

1 10
2 sp|P26371| 0.000000 0.235283 1.671817 1.254859 1.543415 1.138627
3 1.248512 2.318984 1.397221 1.287708
4 sp|Q701N4| 0.235283 0.000000 1.580261 1.263742 1.698652 1.141459
5 1.212450 1.843947 1.389025 1.162573
6 sp|Q07627| 1.671817 1.580261 0.000000 1.322628 1.289484 1.284513
7 1.294597 1.827726 1.207906 1.217074
8 sp|P60409| 1.254859 1.263742 1.322628 0.000000 0.327438 0.138960
9 0.227468 1.501303 0.965589 0.950938
10 sp|P60410| 1.543415 1.698652 1.289484 0.327438 0.000000 0.312342
11 0.416778 2.074238 1.241231 1.215168
12 sp|P60370| 1.138627 1.141459 1.284513 0.138960 0.312342 0.000000
13 0.137545 1.409623 0.997383 0.958282
14 sp|P60014| 1.248512 1.212450 1.294597 0.227468 0.416778 0.137545
15 0.000000 1.490143 1.080655 1.121647
16 sp|Q9BYR9| 2.318984 1.843947 1.827726 1.501303 2.074238 1.409623
17 1.490143 0.000000 1.381587 1.224728
18 sp|Q9BYR3| 1.397221 1.389025 1.207906 0.965589 1.241231 0.997383
19 1.080655 1.381587 0.000000 0.317735
20 sp|Q9BYQ6| 1.287708 1.162573 1.217074 0.950938 1.215168 0.958282
21 1.121647 1.224728 0.317735 0.000000
22 10
23 sp|P26371| 0.000000 0.350629 1.351127 1.588979 1.290319 1.538891
24 1.506696 2.426165 1.284847 1.499296
25 sp|Q701N4| 0.350629 0.000000 1.339785 1.364145 1.182077 1.353477
26 1.367485 1.968204 1.421094 1.364574
27 sp|Q07627| 1.351127 1.339785 0.000000 1.436904 1.498026 1.407429
28 1.736736 1.909386 1.184741 1.088273
29 sp|P60409| 1.588979 1.364145 1.436904 0.000000 0.268688 0.144776
30 0.301070 1.524622 1.166851 1.038074
31 sp|P60410| 1.290319 1.182077 1.498026 0.268688 0.000000 0.247672
  
```

Step 4 : Obtaining Phylogenetic tree by Neighbour method.



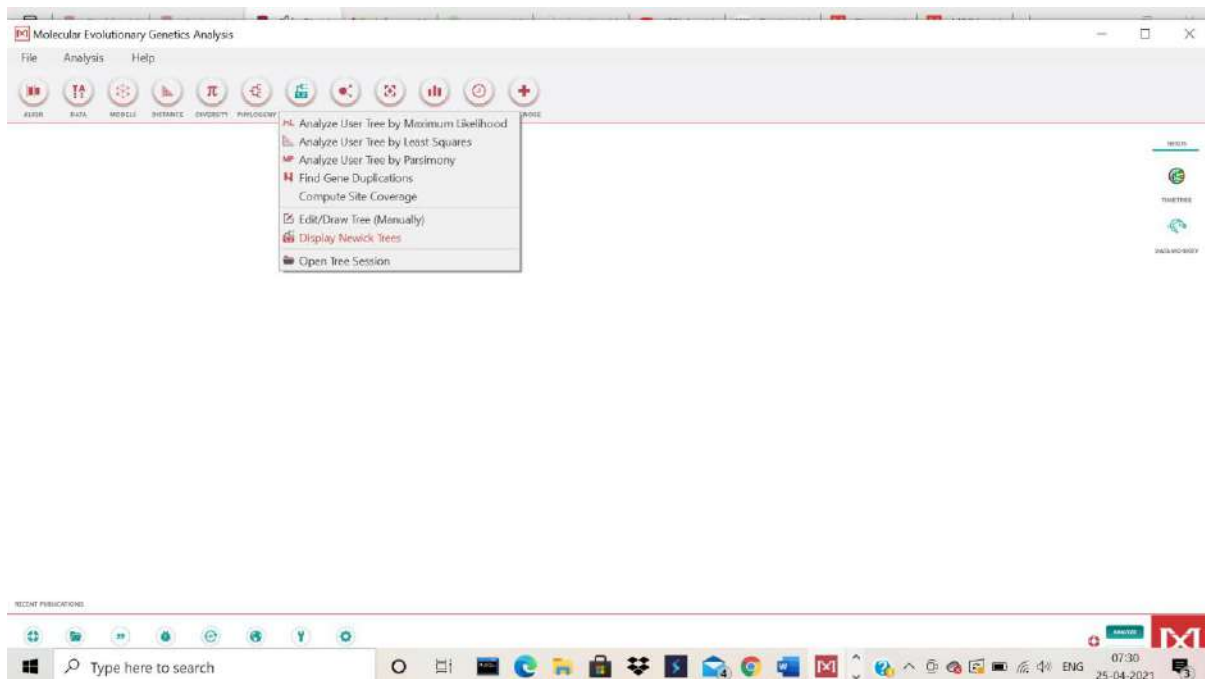
Setting the parameters:



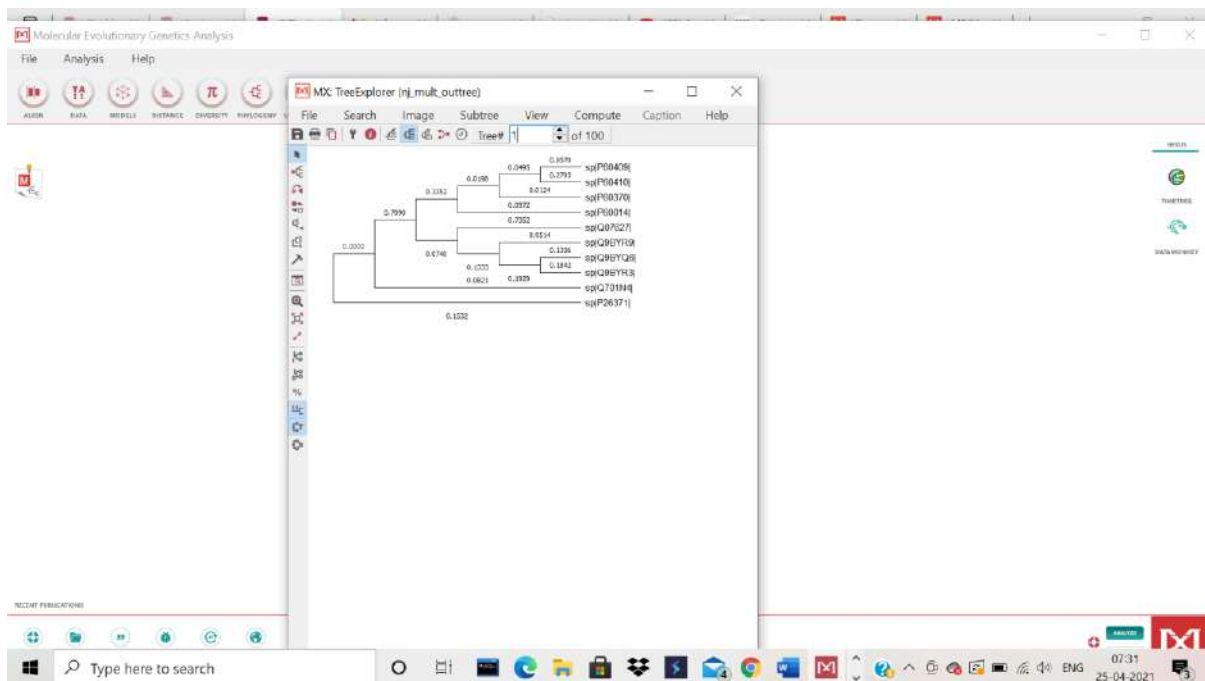
3 April 2021, 6:46 PM

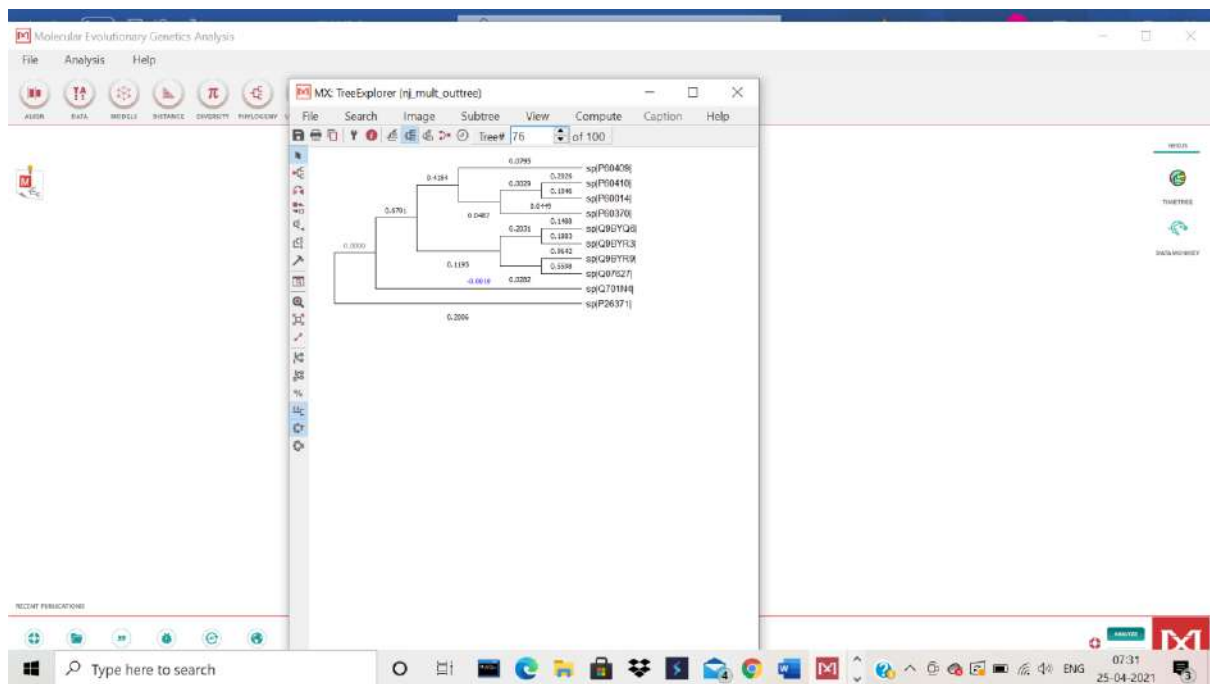


Step 5: To compare all 100 trees go to MEGA:



Step 6: Go to user tree and click on display newick tree and see all the 1 of 100 trees.

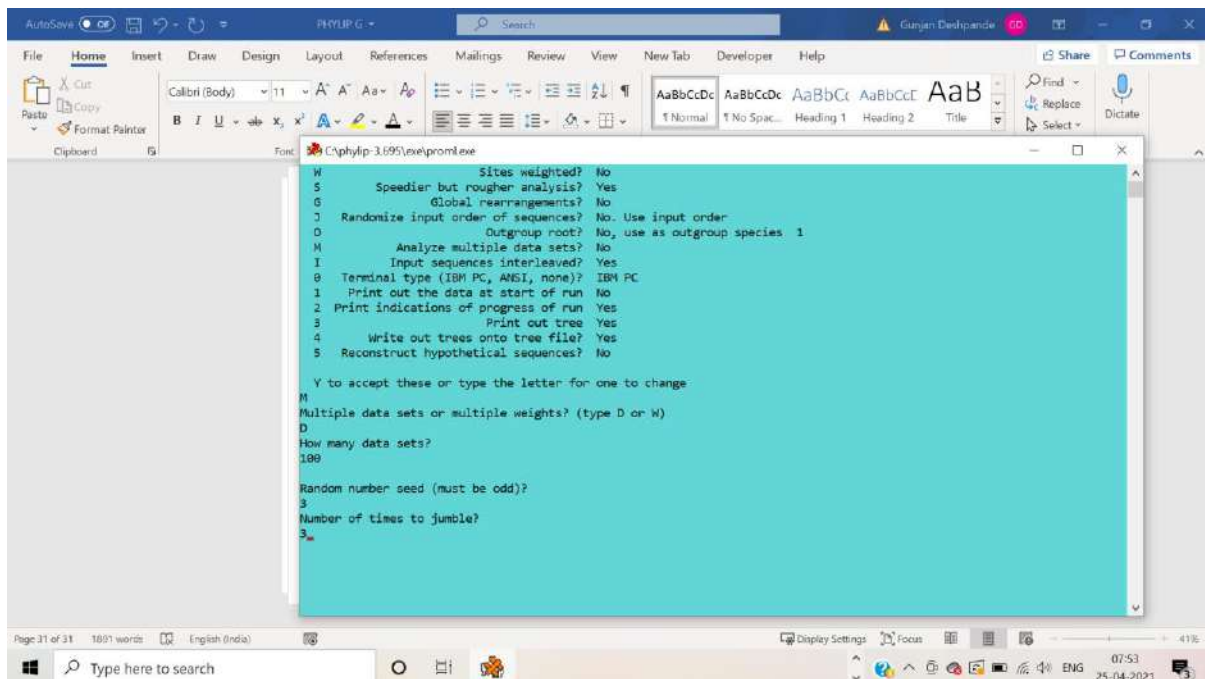




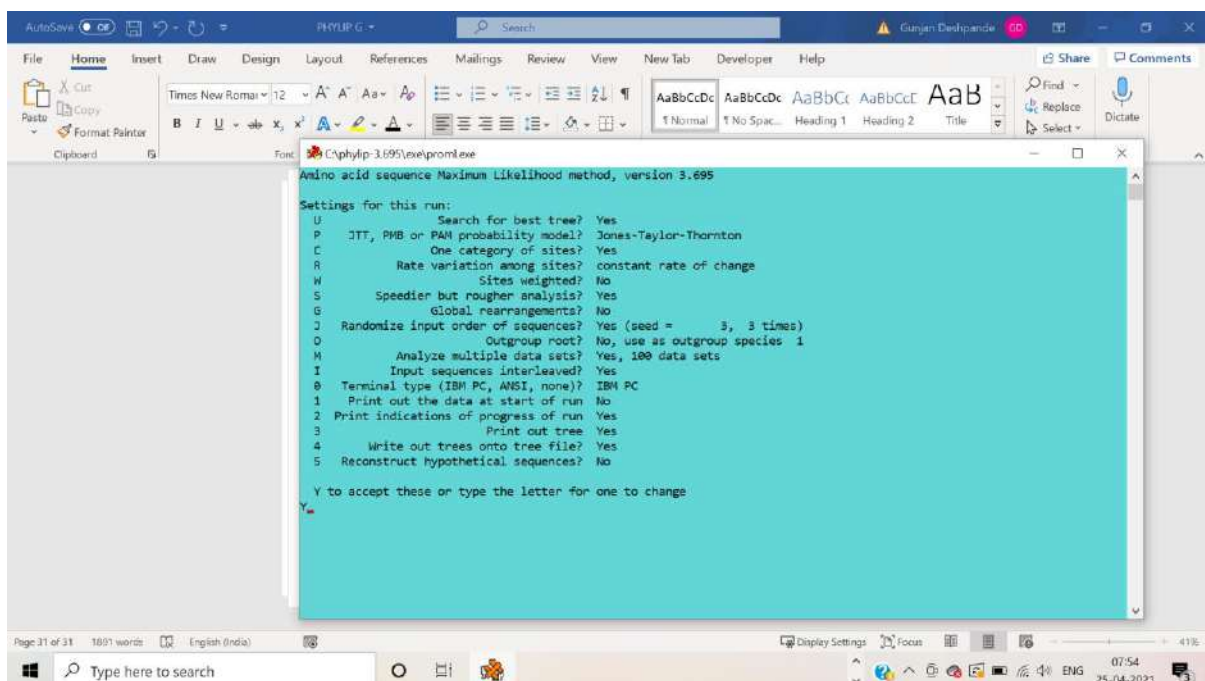
✓ Performing Bootstrapping for Character based methods:-

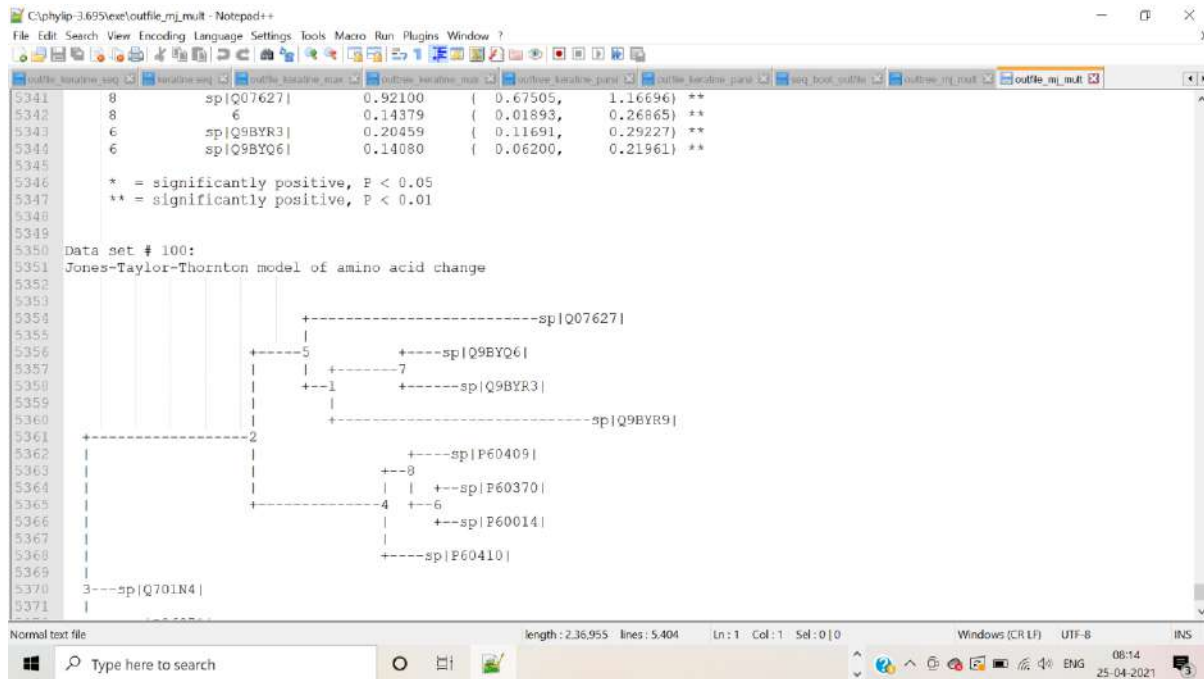
Maximum Likelihood: Where proml is the program used for proteins in ML.

Step 1: Select parameters:

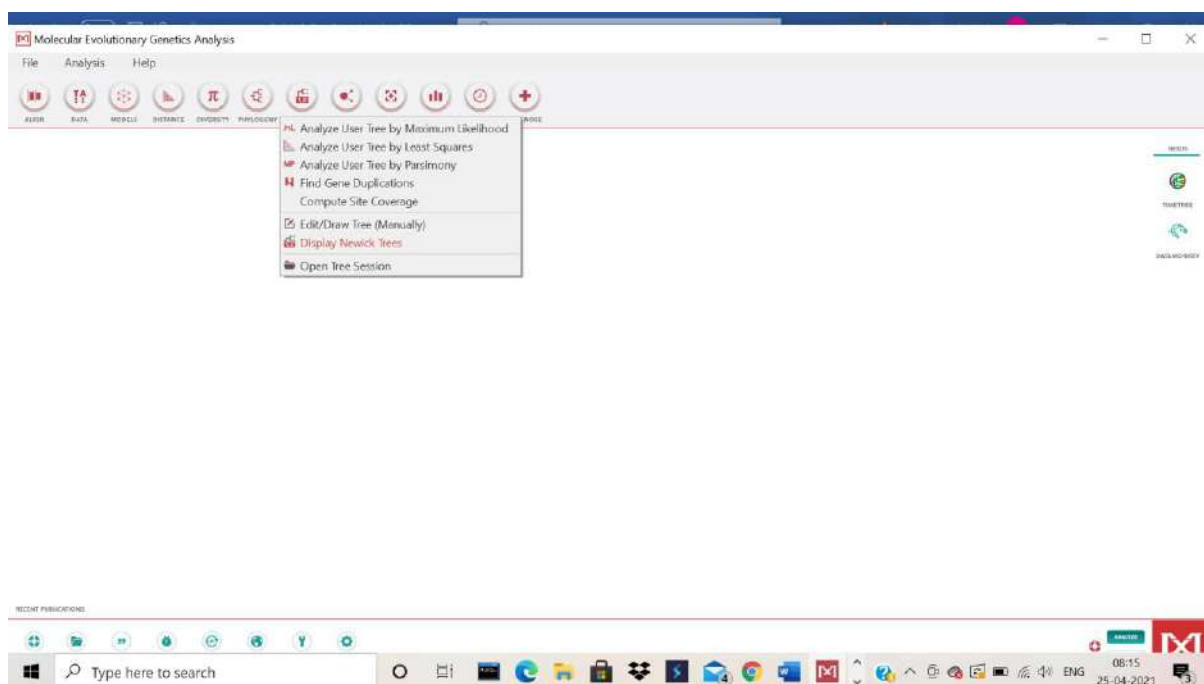


Step 2: Accept all the respective modified parameters by typing Y.

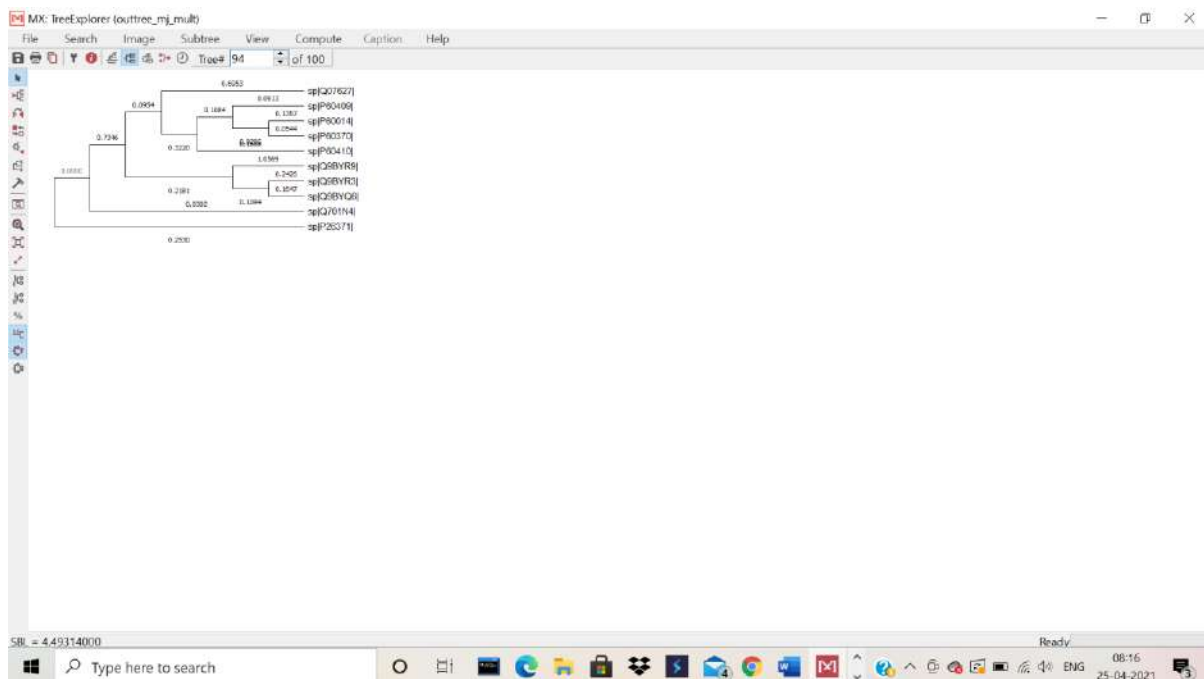




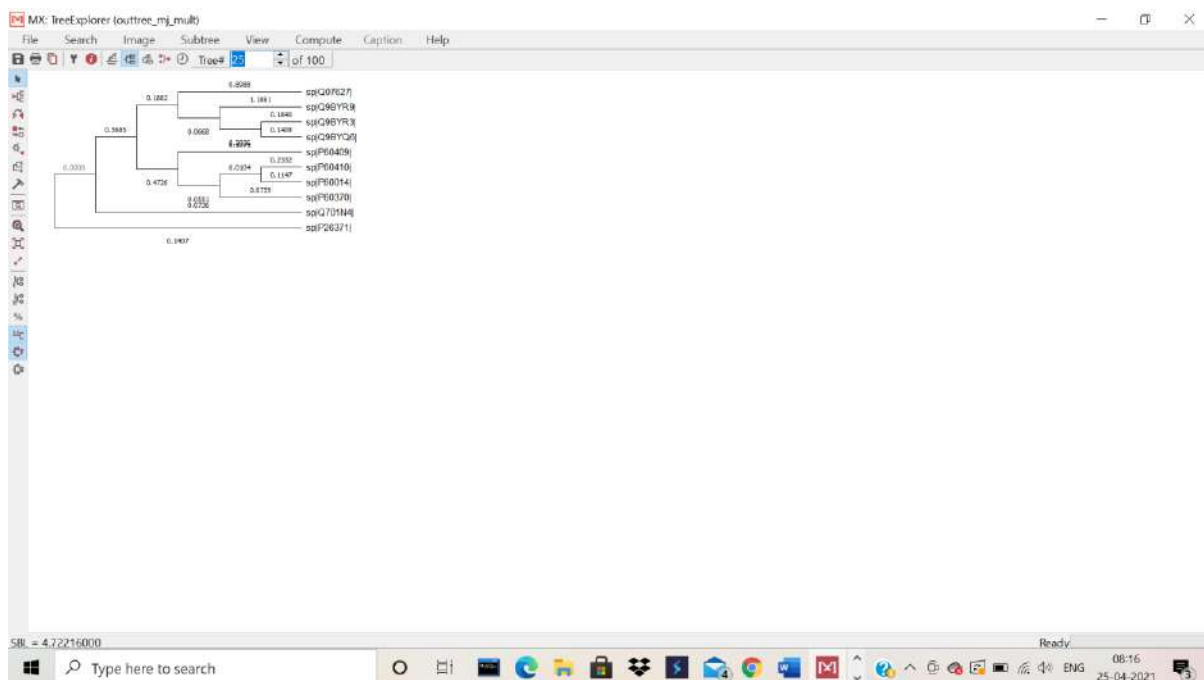
Step 5: Go to MEGA and select user tree and click on display network tree.



94 of 100

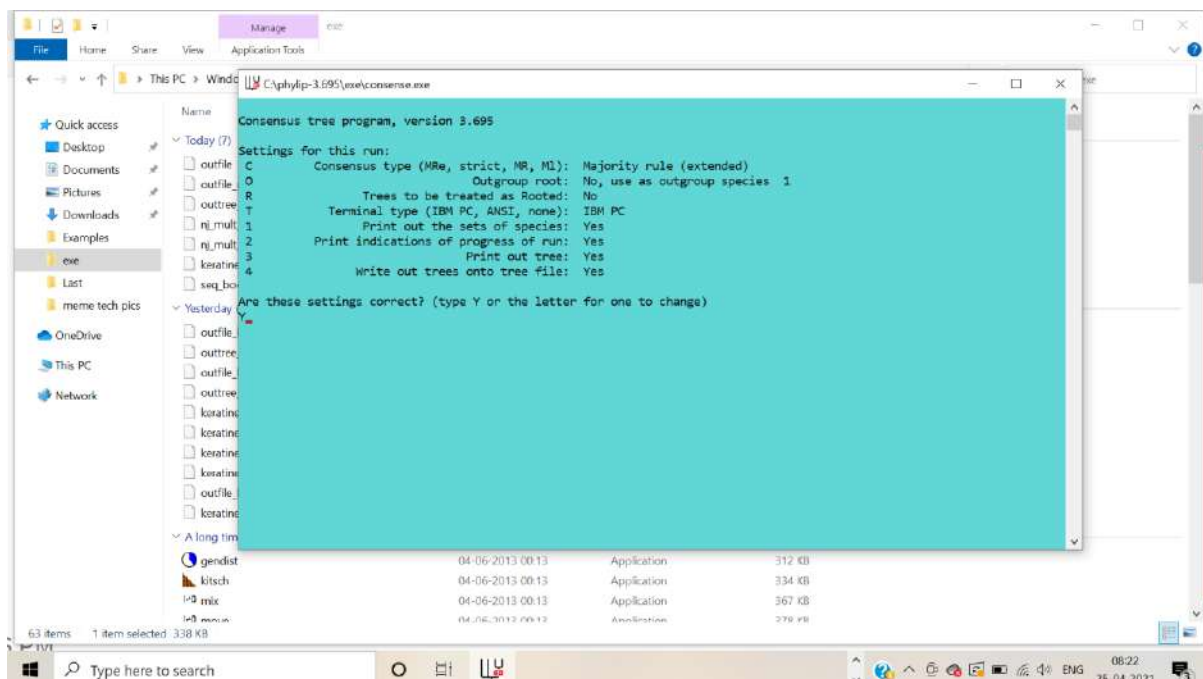
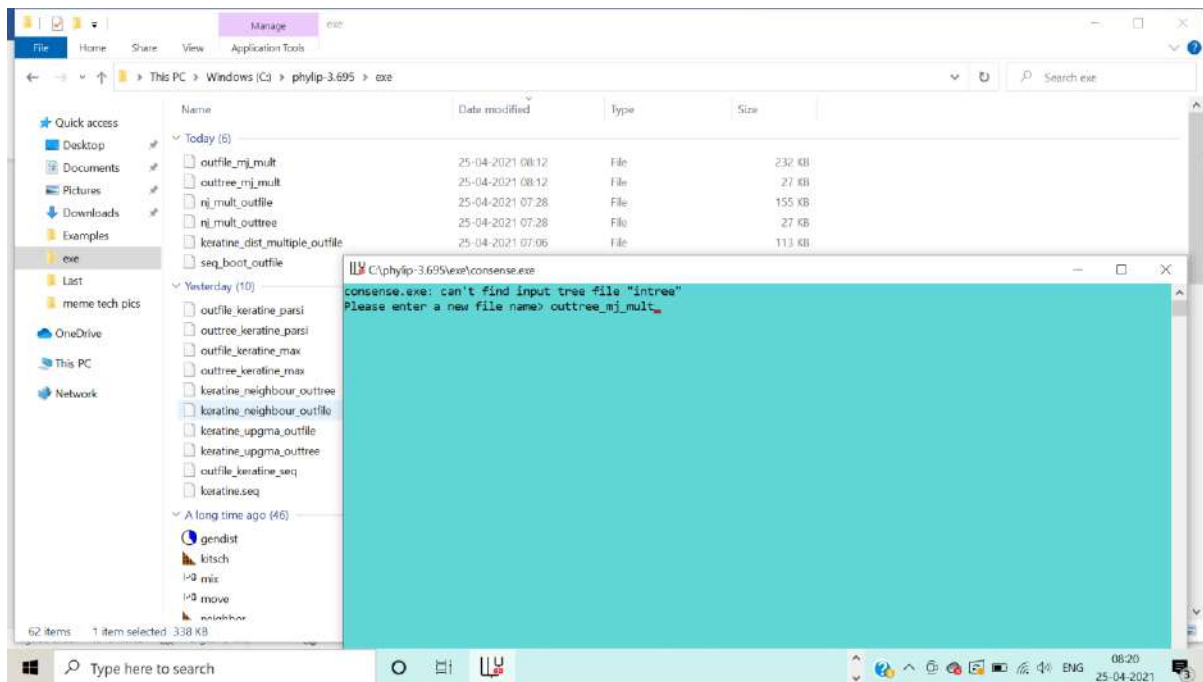


25 of 100

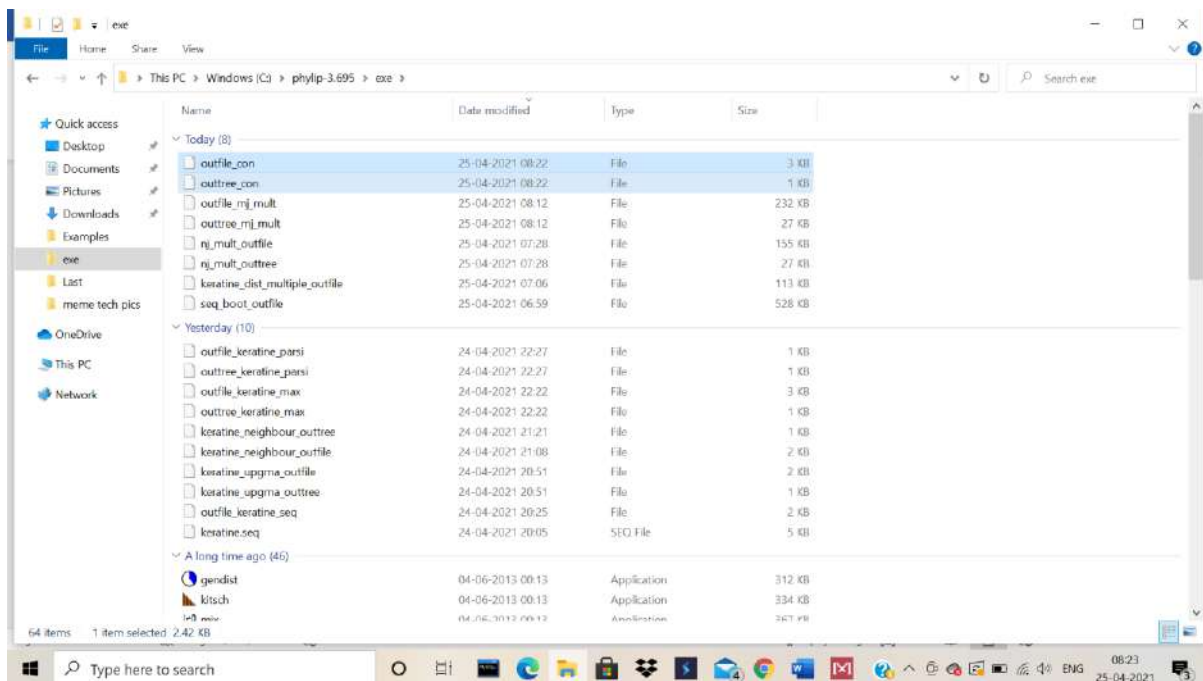


✓ Performing consensus tree of NJ and ML:

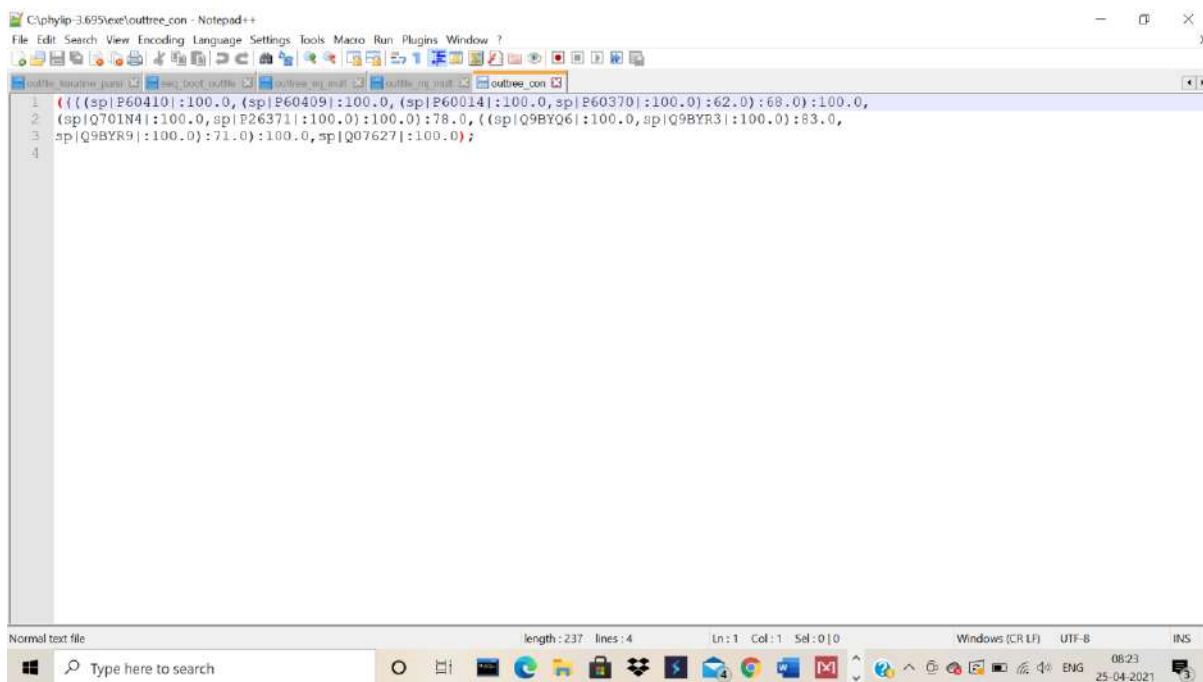
Step 1: Select a file which has atleast 100 trees.



Select 2: Outfile and outtree appears as soon as you press Y and agree their terms.

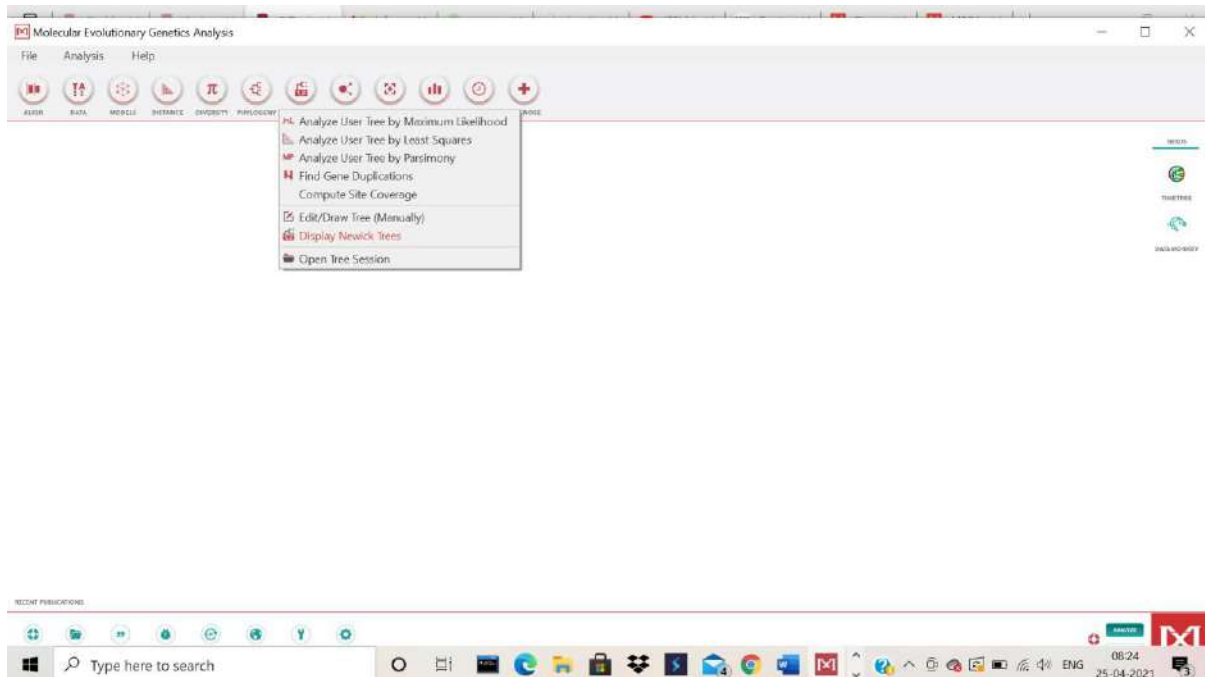


Step 3: This is the outtree for all the sequence evaluated.



This outtree is consenses of those 100 trees.

Step 4: Open MEGA go to user tree and click on display newrick tree.



Step 5: We successfully got consensus tree.

