

```

C:\phylib-3.695\exe\outfile_pars_hemoglobin - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_hemoglobin_upgma outfile_max_hemoglobin outfile_max_hemoglobin outfile_pars_hemoglobin outfile_pars_hemoglobin
1 Protein parsimony algorithm, version 3.695
2
3
4
5
6 2 trees in all found
7
8
9
10
11 +-----5
12 +-----sp|P04244|
13 ! +-----sp|P02067|
14 !
15 !
16 +---3 +----9
17 ! ! !
18 ! ! +----8
19 ! ! ! !
20 ! ! +----7
21 +-2 +----6 +--sp|P02024|
22 !
23 ! !
24 ! !
25 1 !
26 !
27 ! +-----sp|P02094|
28 !
29 +-----sp|P02112|
30
31 remember: this is an unrooted tree!

```

Normal text file length : 1.467 lines : 64 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS  
19:29 28-04-2021

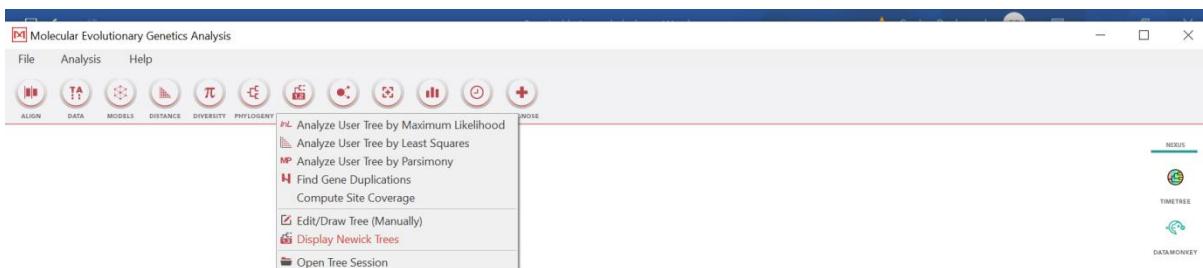
```

C:\phylib-3.695\exe\outfile_pars_hemoglobin - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_hemoglobin_upgma outfile_max_hemoglobin outfile_max_hemoglobin outfile_pars_hemoglobin outfile_pars_hemoglobin
34 requires a total of 194.000
35
36
37
38
39 +-----9
40 +-----sp|P18983|
41 !
42 +----8
43 ! !
44 ! +----7
45 +----6 +--sp|P02024|
46 !
47 !
48 +---3 +----5
49 ! !
50 !
51 +-2 !
52 ! !
53 ! !
54 1 !
55 ! +-----sp|P02094|
56 !
57 +-----sp|P02112|
58
59 remember: this is an unrooted tree!
60
61
62 requires a total of 194.000
63
64

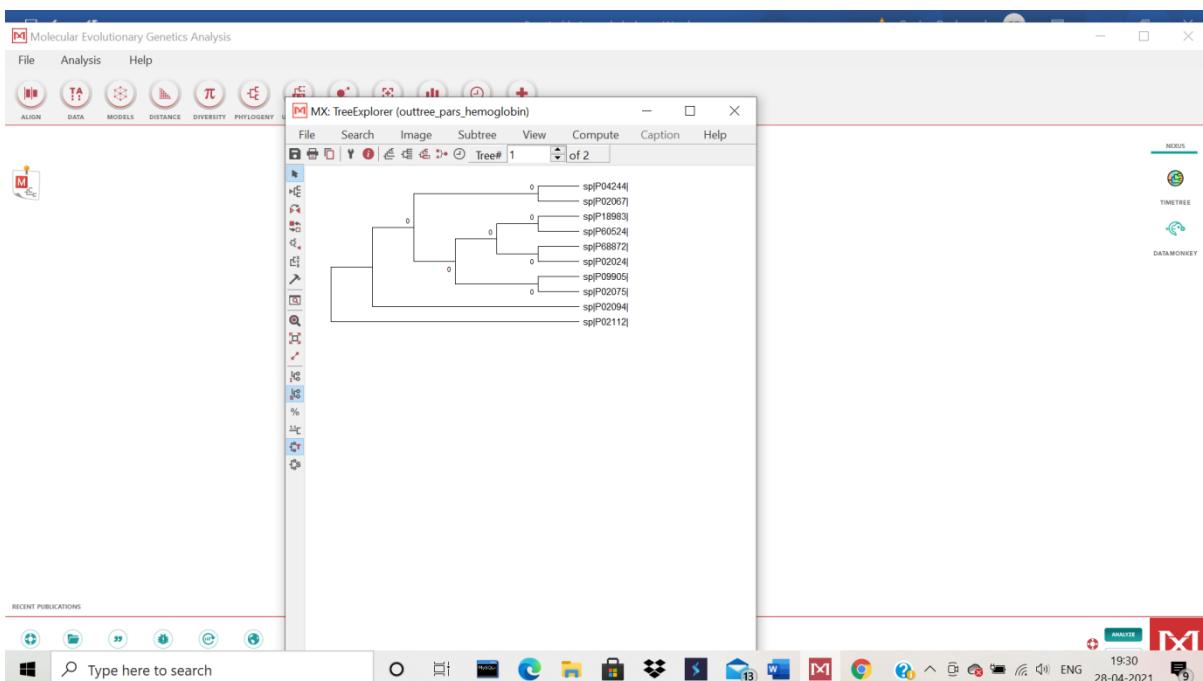
```

Normal text file length : 1.467 lines : 64 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS  
19:29 28-04-2021

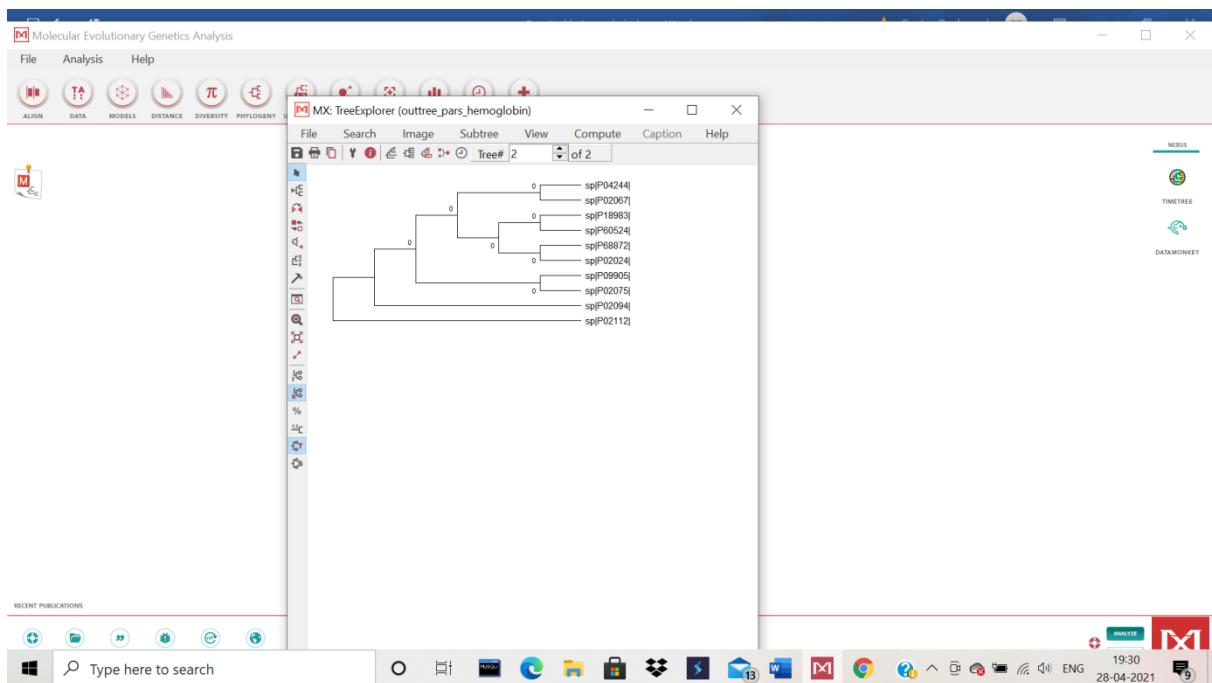
Open MEGA software to view newick phylogenetic tree and go to user tree, click on display newick trees:



1 of 2



2 of 2



3. Perform distance base phylogenetic analysis of nucleotide sequences (any protein or transcription factor) from 5 different organisms. Any method.

**Answer:**

So for performing nucleotide analysis we have two methods:-

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

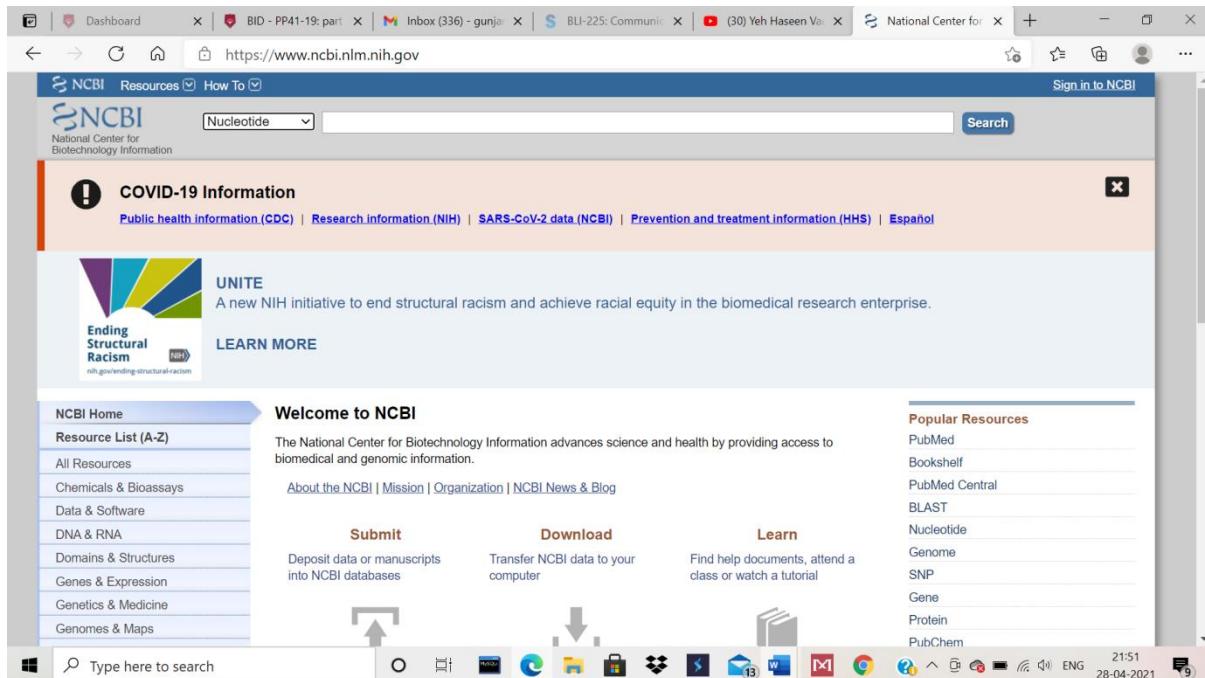
- UPGMA
- Neighbour Joining

2.Character based methods:-

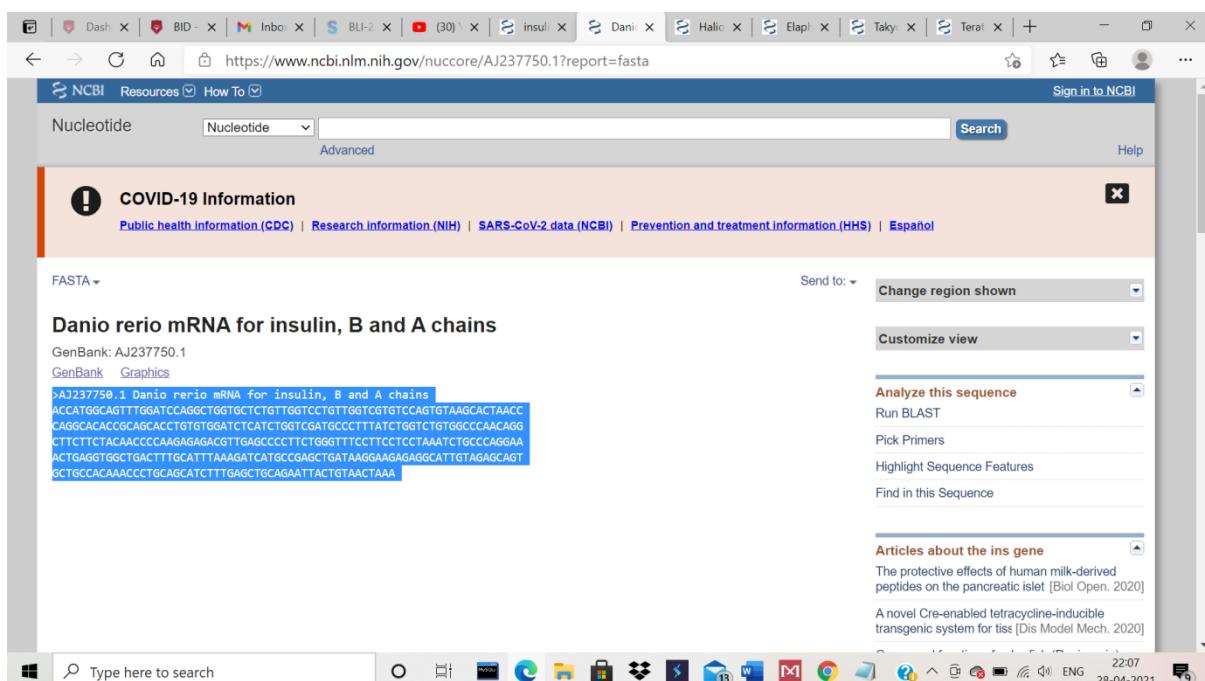
- Maximum Likelihood: Where proml is the program used for protein/nucleotides in ML.
- Parsimony: Which has protpars as nucleotides/protein programs used in Parsimony.

```
>AJ237750.1 = Danio rerio mRNA for insulin
>FJ536845.1 = Haliothis corrugata insulin mRNA
>LC020055.1 = Elaphe climacophora INS mRNA for insulin
>LC020053.1 = Takydromus tachydromoides INS mRNA for insulin
>LC020046.1 = Teratoscincus roborowskii INS mRNA for insulin
```

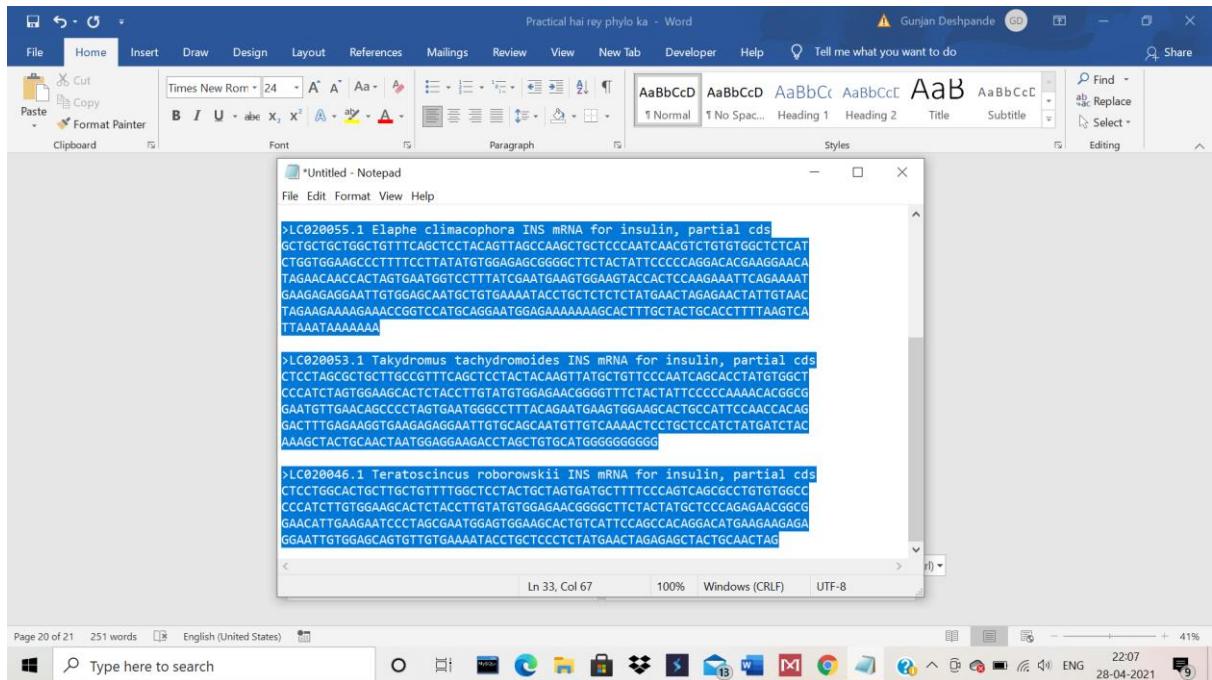
Go to NCBI homepage and select nucleotide as database and type insulin, select 5 different organisms to analysis their sequences download each and every fasta format :



Copy paste every fasta format :



Copy pasting every fasta format in notepad:

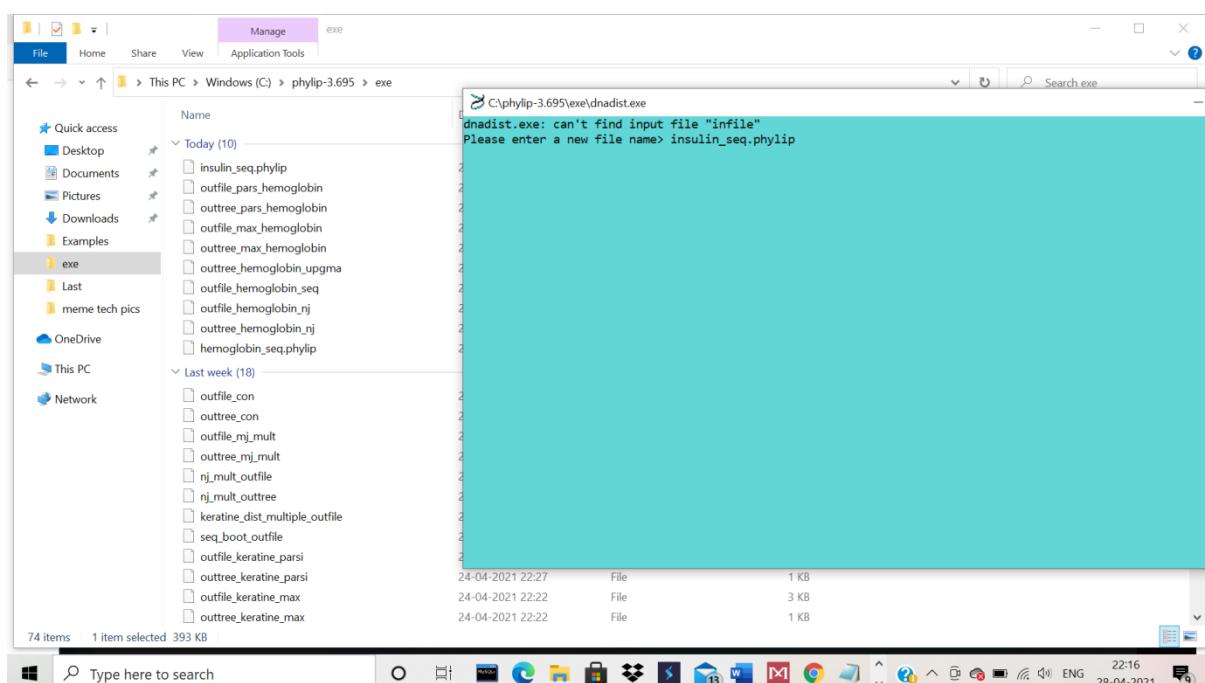


In clustal omega change the set of sequence to DNA and paste your sequences in the box. Change the out put file format to PHYLIP and click on submit:

Download Alignment File:

The screenshot shows the Clustal Omega results page. At the top, there's a navigation bar with links for Dashboard, BID - PP41-19, Your job "DNA", Results < Clustal, BUI-225: Com..., (30) Yeh Hasee..., and insulin - Nucleic... The main title 'Clustal Omega' is displayed prominently. Below it are links for Input form, Web services, Help & Documentation, Bioinformatics Tools FAQ, Feedback, and Share. The page content includes a 'Tools > Multiple Sequence Alignment > Clustal Omega' breadcrumb, a 'Results for job clustalo-E20210428-173858-0814-18820773-p1m' section, and tabs for Alignments (selected), Result Summary, Guide Tree, and Submission Details. A 'Download Alignment File' button is present. The main area displays a sequence alignment with several rows of DNA/RNA sequences. At the bottom, there's a search bar with 'Type here to search' and a taskbar with various icons.

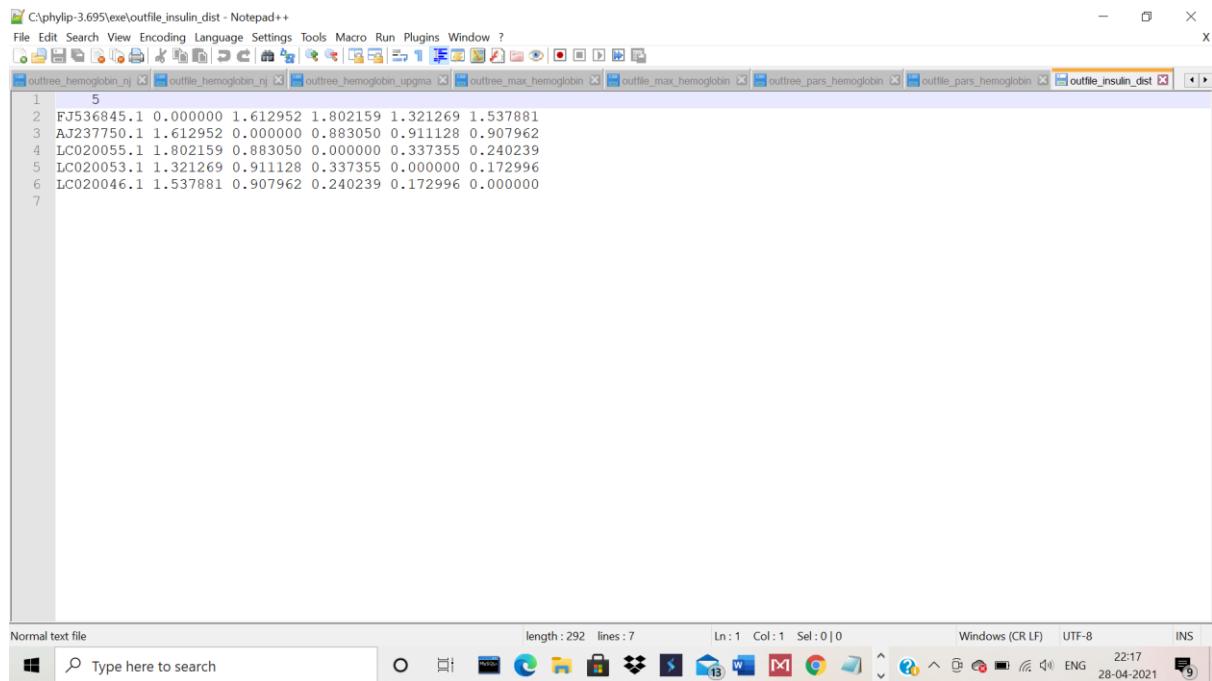
Go to dnadist in exe folder and enter a file name .i.e. insulin\_seq.phylip.



You will view Distance Matrix:

>AJ237750.1 = *Danio rerio* mRNA for insulin  
>FJ536845.1 = *Haliotis corrugata* insulin mRNA  
>LC020055.1 = *Elaphe climacophora* INS mRNA for insulin

>LC020053.1 = Takydromus tachydromoides INS mRNA for insulin  
 >LC020046.1 = Teratoscincus roborowskii INS mRNA for insulin



```

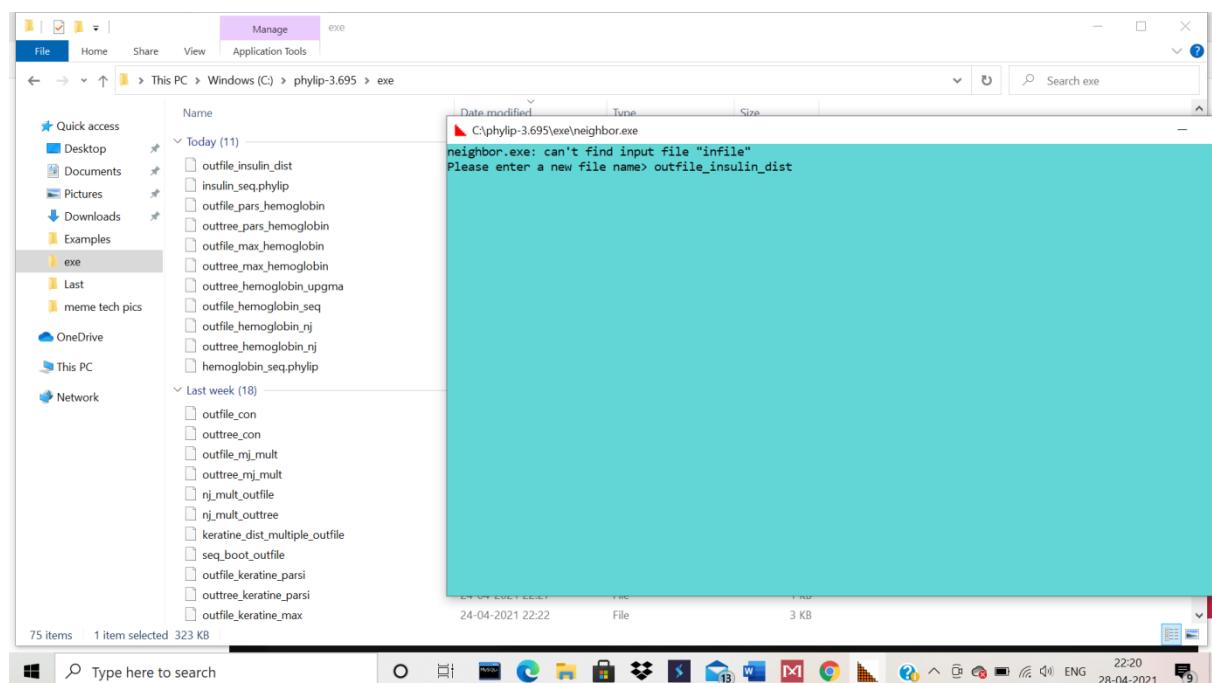
C:\phylib-3.695\exe\outfile_insulin_dist - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_hemoglobin_nj outfile_hemoglobin_nj outfile_hemoglobin_upgma outfile_max_hemoglobin outfile_max_hemoglobin outfile_pars_hemoglobin outfile_pars_hemoglobin outfile_insulin_dist
1 5
2 FJ536845.1 0.000000 1.612952 1.802159 1.321269 1.537881
3 AJ237750.1 1.612952 0.000000 0.883050 0.911128 0.907962
4 LC020055.1 1.802159 0.883050 0.000000 0.337355 0.240239
5 LC020053.1 1.321269 0.911128 0.337355 0.000000 0.172996
6 LC020046.1 1.537881 0.907962 0.240239 0.172996 0.000000
7

```

Normal text file length : 292 lines : 7 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS  
 Type here to search 22:17 ENG 28-04-2021 9

Performing distance base phylogenetic analysis of nucleotide sequences from 5 different organisms by **neighbor joining method**:

Go to neighbor in exe folder and enter your file name .i.e. outfile\_insulin\_dist:



Out tree obtained :

```

C:\phylib-3.695\exe\outtree_insulin_nj_dist - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_pars_hemoglobin outfile_pars_hemoglobin outfile_insulin_dist outtree_insulin_nj_dist
1 (AJ237750.1:0.47995, ((LC020055.1:0.19113, LC020046.1:0.04911):0.09075,
2 LC020053.1:0.04431):0.26542, FJ536845.1:1.13300);
3

```

Normal text file length : 121 lines : 3 Ln : 3 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 22:21 28-04-2021 ENG INS

You will get an unrooted tree for neighbor method of nucleotide sequences from 5 different organisms:

```

C:\phylib-3.695\exe\outfile_insulin_nj_dist - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_pars_hemoglobin outfile_pars_hemoglobin outfile_insulin_dist outtree_insulin_nj_dist outfile_insulin_nj_dist
1
2 5 Populations
3
4 Neighbor-Joining/UPGMA method version 3.695
5
6
7 Neighbor-joining method
8
9 Negative branch lengths allowed
10
11
12 +-----AJ237750.1
13 !
14 !       +---LC020055.1
15 !         +-2
16 !         3   +LC020046.1
17 !           !
18 !         +-LC020053.1
19 !
20 +-----FJ536845.1
21
22
23 remember: this is an unrooted tree!
24
25 Between      And      Length
26 -----  ---  -----
27 1          AJ237750.1  0.47995
28 1          3          0.26542
29 3          2          0.09075
30 2          LC020055.1  0.19113
31 2          LC020046.1  0.04911

```

Normal text file length : 735 lines : 36 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 22:22 28-04-2021 ENG INS

```

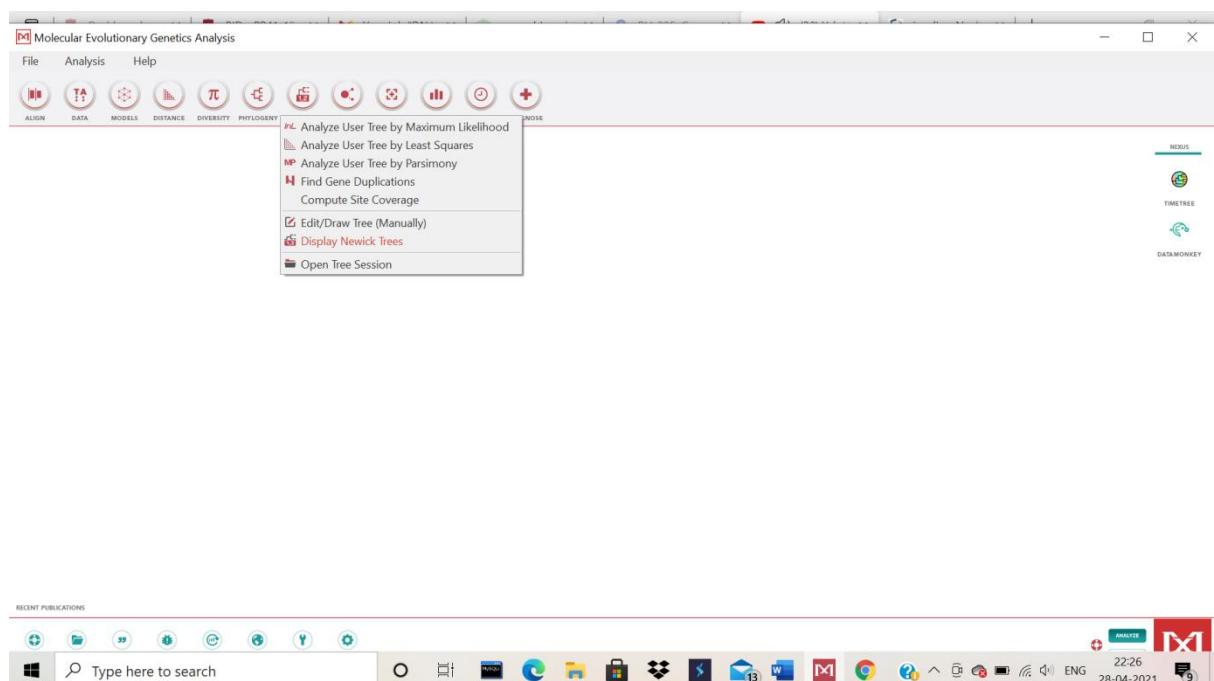
C:\phylib-3.695\exe\outfile_insulin_nj_dist - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_pars_hemoglobin outfile_pars_hemoglobin outfile_insulin_dist outfile_insulin_nj_dist outfile_insulin_nj_dist

6 Neighbor-joining method
7
8 Negative branch lengths allowed
9
10
11
12 +-----AJ237750.1
13 !
14 ! +----LC020055.1
15 !
16 ! +--2
17 ! !
18 ! +-LC020046.1
19 !
20 +-----FJ536845.1
21
22
23 remember: this is an unrooted tree!
24
25 Between And Length
26 -----
27 1 AJ237750.1 0.47995
28 1 3 0.26542
29 3 2 0.09075
30 2 LC020055.1 0.19113
31 2 LC020046.1 0.04911
32 3 LC020053.1 0.04431
33 1 FJ536845.1 1.13300
34
35
36

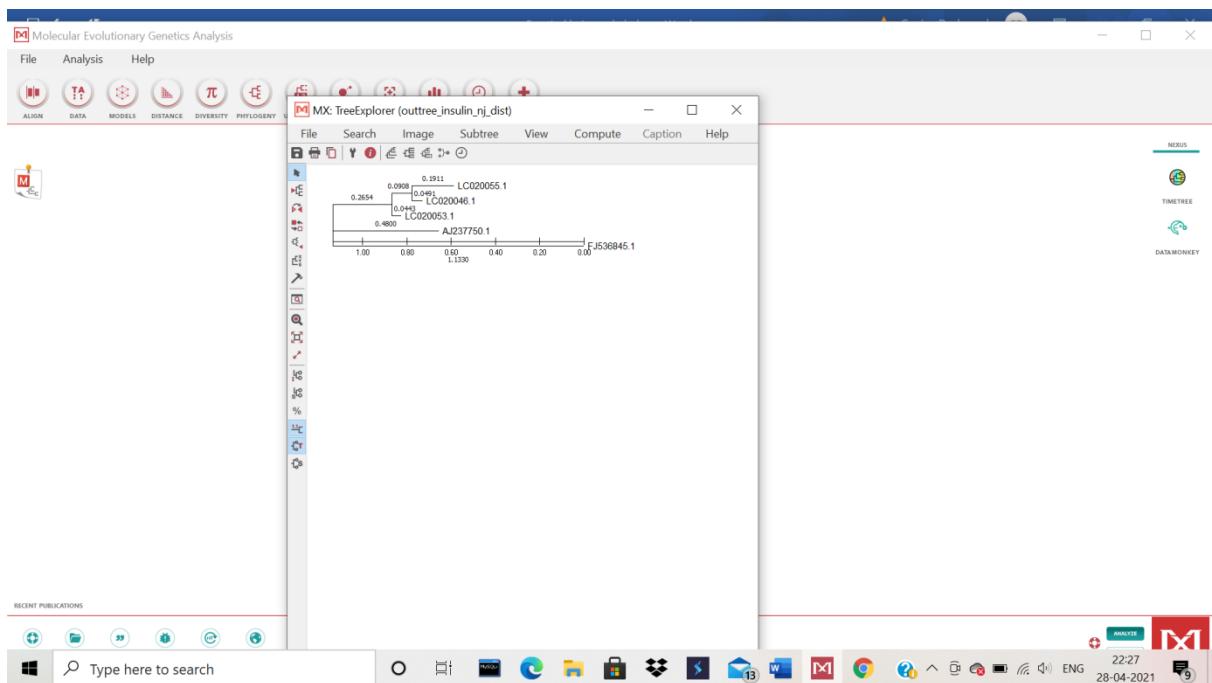
```

Normal text file length : 735 lines : 36 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 22:22 ENG 28-04-2021 9

Open MEGA software to view newick phylogenetic tree and go to user tree, click on display newick trees:



Distance based phylogenetic newick tree for neighbor joining method:



4. Perform character base phylogenetic analysis of nucleotide sequences (any protein or transcription factor) from 5 different organisms

Answer:

So character based methods includes:-

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

>AJ237750.1 = Danio rerio mRNA for insulin

>FJ536845.1 = Haliotis corrugata insulin mRNA

>LC020055.1 = Elaphe climacophora INS mRNA for insulin

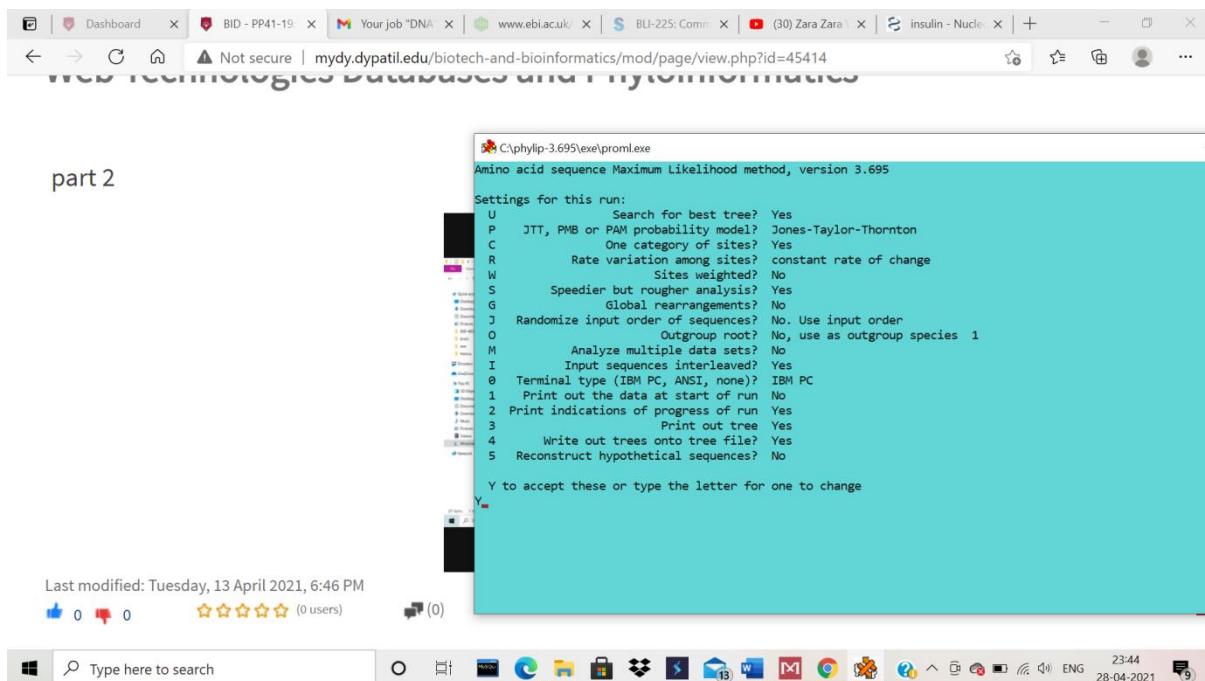
>LC020053.1 = Takydromus tachydromoides INS mRNA for insulin

>LC020046.1 = Teratoscincus roborowskii INS mRNA for insulin

Performing character based phylogenetic analysis of insulin sequences of 5 different organisms

**Maximum Likelihood:**

Go to proml in exe folder and type Y to accept all the default parameters:



You will fill an unrooted tree in outfile for the maximum likelihood method:

```

C:\phylip-3.695\exe\outfile_insulin_parsi_character - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outfile_pars_hemoglobin outfile_insulin_dist outtree_insulin_nj_dist outfile_insulin_nj_dist outfile_insulin_parsi_character

1 Amino acid sequence Maximum Likelihood method, version 3.695
2 Jones-Taylor-Thornton model of amino acid change
3
4
5
6
7 +-----sp|P02094|
8 |
9 | +----sp|P02075|
10 | |
11 | | +----sp|P02067|
12 1--2 |
13 | | |
14 | | | +sp|P18983|
15 | | | +-8
16 | | | +sp|P60524|
17 | | | +-7
18 | | | +-----sp|P04244|
19 | | | +sp|P68872|
20 | | +-5 +-6
21 | | +sp|P02024|
22 |
23 | +-----sp|P09905|
24 |
25 +-----sp|P02112|
26
27
28 remember: this is an unrooted tree!
29
30 Ln Likelihood = -1246.88212
31

```

Normal text file length: 2,137 lines: 57 ln: 1 col: 1 sel: 0 | 0 Windows (CR LF) UTF-8 INS

C:\phylip-3.695\exe\outfile\_insulin\_parsi\_character - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

outfile\_pars\_hemoglobin outfile\_insulin\_dist.outtree\_insulin\_nj\_dist.outfile\_insulin\_nj\_dist.outfile\_insulin\_parsi\_character

```

27 remember: this is an unrooted tree!
28
29 Ln Likelihood = -1246.88212
30
31 Between And Length Approx. Confidence Limits
32 ----- --- -----
33
34
35 1 sp|P02112| 0.29537 ( 0.19208, 0.39864) **
36 1 sp|P02094| 0.16327 ( 0.08750, 0.23904) **
37 1 2 0.04845 ( zero, 0.09724) **
38 2 sp|P02075| 0.09514 ( 0.04026, 0.15001) **
39 2 4 0.01391 ( zero, 0.03821) *
40 4 sp|P02067| 0.10275 ( 0.04479, 0.16071) **
41 4 5 0.02611 ( zero, 0.05900) **
42 5 3 0.04136 ( 0.00468, 0.07804) **
43 3 8 0.01650 ( zero, 0.04052) **
44 8 sp|P18983| 0.02814 ( zero, 0.05655) **
45 8 7 0.03129 ( 0.00044, 0.06213) **
46 7 sp|P60524| 0.00482 ( zero, 0.01943)
47 7 sp|P04244| 0.13104 ( 0.06834, 0.19375) **
48 3 6 0.05676 ( 0.01543, 0.09809) **
49 6 sp|P68872| 0.00696 ( zero, 0.02071) *
50 6 sp|P02024| 0.00010 ( zero, 0.01694)
51 5 sp|P09905| 0.12395 ( 0.06114, 0.18676) **

* = significantly positive, P < 0.05
** = significantly positive, P < 0.01

```

Normal text file length : 2.137 lines : 57 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS

Type here to search

Out tree for maximum likelihood method of insulin sequences of 5 different organisms:

C:\phylip-3.695\exe\outtree\_insulin\_parsi.character - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

outfile\_pars\_hemoglobin outfile\_insulin\_dist.outtree\_insulin\_nj\_dist.outfile\_insulin\_nj\_dist.outfile\_insulin\_parsi\_character.outtree\_insulin\_parsi\_character

```

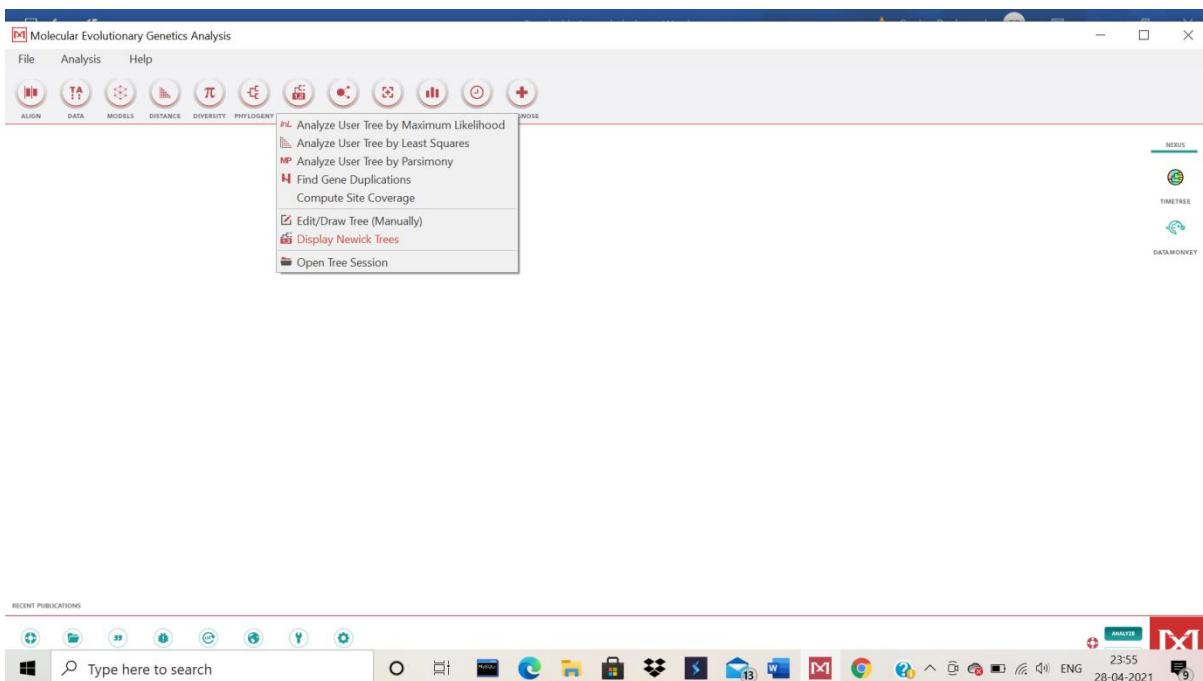
1 ((sp|P02094|:0.16327, (sp|P02075|:0.09514, (sp|P02067|:0.10275,
2 (((sp|P18983|:0.02814, (sp|P60524|:0.00482, sp|P04244|:0.13104):0.03129):0.01650,
3 (sp|P68872|:0.00696, sp|P02024|:0.00010):0.05676):0.04136,
4 sp|P09905|:0.12395):0.02611):0.01391):0.04845,
5 sp|P02112|:0.29537;
6

```

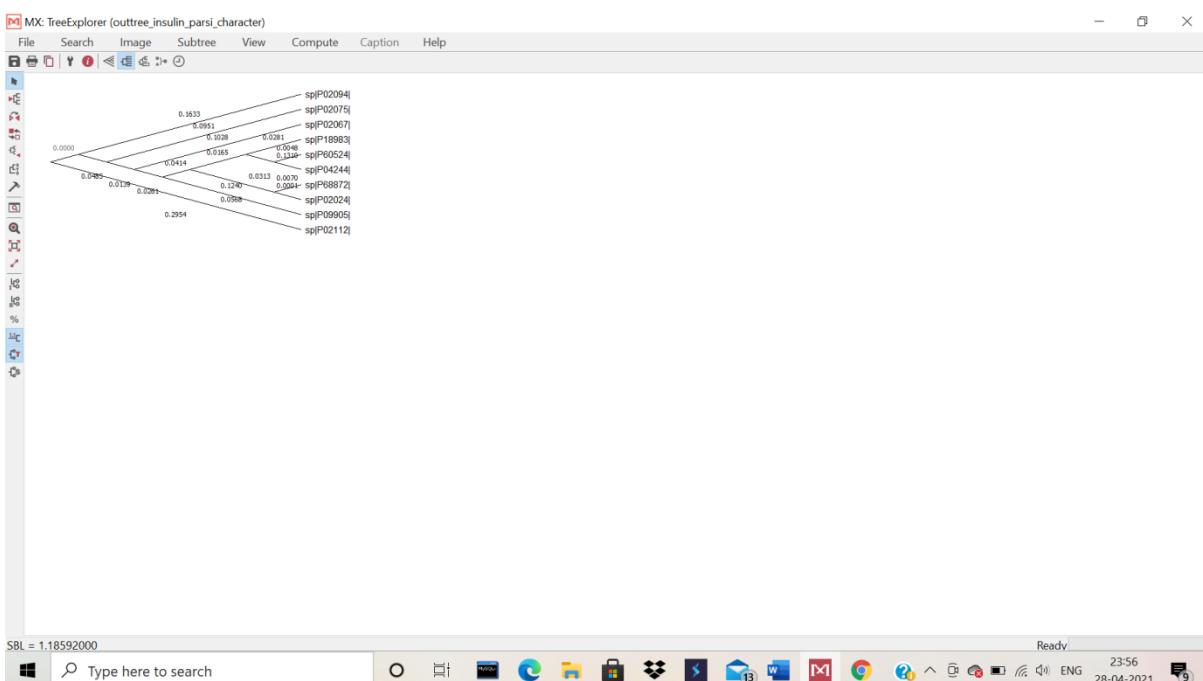
Normal text file length : 272 lines : 6 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS

Type here to search

Open MEGA software to view newick phylogenetic tree and go to user tree, click on display newick trees:



You will view character base phylogenetic newick tree for maximum likelihood method:



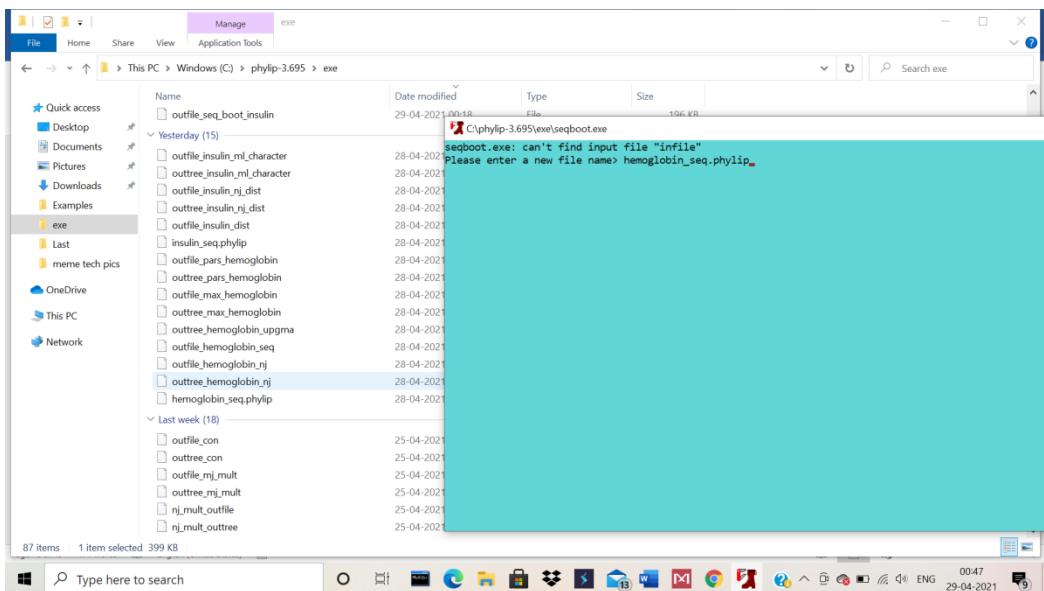
5. Perform boot strap analysis on distance based phylogenetic method of hemoglobin sequences.

Answer:

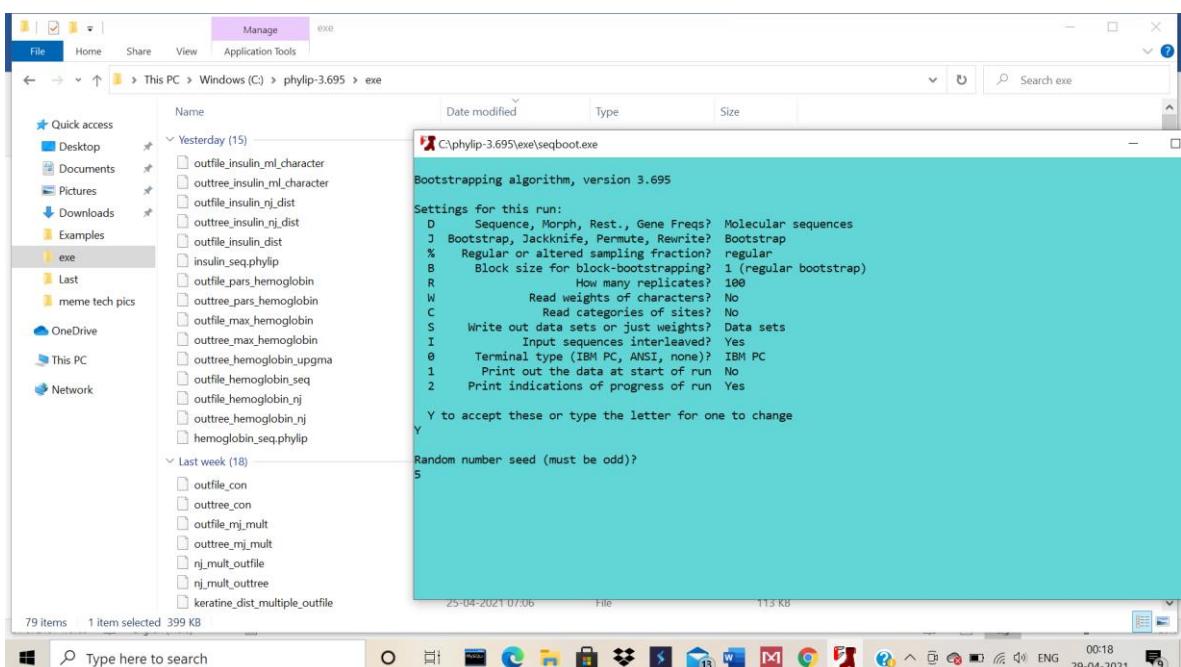
**Bootstrapping:-** It is basically a type of alignment which is already done , but we are actually reshuffling the positions of the sequences .i.e. certain positions are repeated, certain positions are deleted in short the shuffling of alignment. Bootstrap basically is used for building up the confidence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

```
>sp|P02024|= Gorilla Hemoglobin  
>sp|P68872|= Pan paniscus Hemoglobin  
>sp|P02112|= CHICK Hemoglobin  
>sp|P02067|= PIG Hemoglobin  
>sp|P02075|= SHEEP Hemoglobin  
>sp|P09905|= Physeter macrocephalus Hemoglobin  
>sp|P04244|= Panthera pardus orientalis Hemoglobin  
>sp|P02094|= Mesocricetus auratus Hemoglobin  
>sp|P60524|= Canis lupus Hemoglobin  
>sp|P18983|= Ailuropoda melanoleuca
```

Open seq boot in exe folder and type the file name as hemoglobin\_seq.phylip :



Type 5 for random number and Y to accept all the parameters and press enter:



Distance Matrix for hemoglobin sequences:

C:\phylip-3.695\exe\outfile\_seq\_boot\_hemo - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

outtree\_insulin.upgma.dist outfile\_insulin\_upgma\_dist hemoglobin\_seq.phylo outfile\_seq\_boot\_hemo

1 10 147

2 sp|P02141| MMHTTAEKK KKIITGKNNV AAECGEAARL I IIIVVVWQRF FFFFANNLLLS SPPTTAAILP

3 sp|P02094| MMHTTDEEKK KKVVVTGKNAA DDAVGEGEARL VVVVVVVWQRF FFFEDDDLLLS SAASSAAVMP

4 sp|P02075| --MTTAEKKKK KKVVVTGKVN DDEVGEGEARL VVVVVVVWQRF FFFEDDDLLLS NAADDAAVMP

5 sp|P02067| MMHSSAAEKKKK KKVLVLGKNNV DDEVGEGEARL VVVVVVVWQRF FFFEDDDLLLS NAADDAAVMP

6 sp|P09905| --HTTGEKKKK KKLLTAKNNV EEEIIGEEARL VVVVVVVWQRF FFFEDDDLLLS TAADDAAVMP

7 sp|P04244| MMFSSAAEKKKK KKVVVSGKVN DDEVGEGEARL VVVVVVVWQRF FFFQDQLLSS SAADDAAIMA

8 sp|P02041| MMHTTPEEKKKK KKVVVTGKNNV DDEVGEGEARL VVVVVVVWQRF FFFEDDDLLLS TPPDDAAVMP

9 sp|P68721| MMHTTPEEKKKK KKVVVTGKNNV DDEVGEGEARL VVVVVVVWQRF FFFEDDDLLLS TPPDDAAVMP

10 sp|P60524| --HTTAEKKKK KKVVVSGKVN DDEVGEGEARL IIIIVVVWQRF FFFEDDDLLLS TPPDDAAVMA

11 sp|P18983| MMHTTGEKKKK KKVVTSKVN DDEVGEGEARL VVVVVVVWQRF FFFDQLLSS TPPDDAAVMP

12

13 RAAHHGKKKK TFGGANDTFSEELHCKKHVDPPFFRRGGGDIIIIIVVL LAAHHSSKKD

14 KAAHHGKKKKK HFAAGHDAFSSEELHCKKHVDPPFFKKGGG GNNIIIIIVVL LSHDDGKGKD

15 KAAHHGKKKKK DFSSGGHDTFAAEELHCKHV DPPFFRRGGG GNNVVVVVVL LARHHCGNN

16 KAAHHGKKKKQ FSSSGHDTFAAEELHCOQHV DPPFFRRGGG GNNVVVVVVL LARRGGHHD

17 KKKHHHGKKKK AFGGGHDTFAAEELHCKHV DPPFFRRGGG GNNVVVVVVL LARHHHGKKE

18 KAAHHGKKKKK NFSSGGNDTFAAEELHCKHV DPPFFRRGGG GNNVCCVVL LARHHHGHH

19 KAAHHGKKKK GFSSGGHDTFAAEELHCKHV DPPFFKKGGG GNNVCCVVL LAAHHHGKKE

20 KAAHHGKKKK NFSSGGNDTFAAEELHCKHV DPPFFRRGGG GNNVCCVVL LAAHHHGKKE

21 KAAHHGKKKK NFSSGGNDTFAAEELHCKHV DPPFFKKGGG GNNVCCVVL LAAHHHGKKE

22 KAAHHHGKKKK NFSSGGNDTFAAEELHCKHV DPPFFRKGGG GNNVCCVVL LAAHHHGKR

23

24 FFQQQAAAALV LVVVVVAHLR KYYYYHHH

25 FFQQQSSAAK VGVVVANLHL KYYYYHHH

26 FFQQQAADDK VGVVVANLHL KYYYYHHH

27 FFQQQAAAALVG VGVVVANLHL KYYYYHHH

28 FFQQQTTAAK VGVVVANLHL KYYYYHHH

29 FFQQQAAAALK VGVVVASLHL RYYYYHHH

30 FFQQQAAAALK VGVVVANLHL KYYYYHHH

31 FFQQQAAAALK VGVVVANLHL KYYYYHHH

Normal text file length : 1,99,900 lines : 3,401 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 00:52 29-04-2021 ENG IN

For all data sets:

C:\phylip-3.695\ex\outfile\_seq\_boot\_hemo - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

outtree\_outfile\_seq\_boot\_hemo\_dist outfile\_insulin\_upgma\_dist hemoglobin\_seq.phyip outfile outfile\_seq\_boot\_hemo

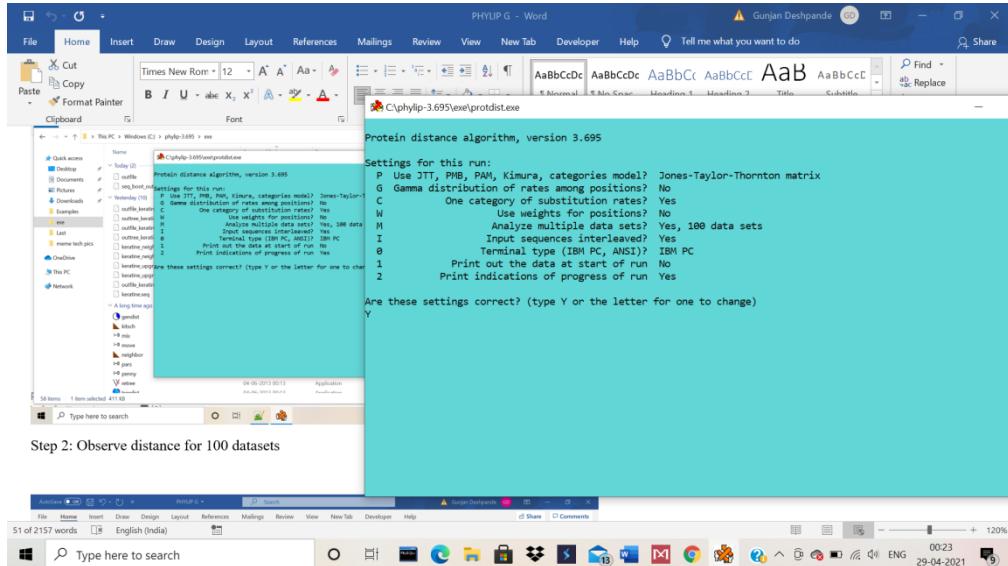
3371 sp|P020671 MMVHLLSSS EEAIVLLLIG GGKKNVVVDE EVALRRLVWV QQQFFESFFG GDDDDLSNAP  
3372 sp|P099051 ---VHLTTT ESGLITLTTAA KAKNNVVVEE EIALRRLVWV QQQFFEHFFG GDDDDLSSTAP  
3373 sp|P04244 MMMSFLSSS ENLVWSSSGG GKGNVVVDE EVALRRLVWV QQQFFQSFFG GDDDDLSAAA  
3374 sp|P020241 MMMVHLLTTT ESAVVTTTA AGKKNVVVDE EVALRRLVWV QQQFFESFFG GDDDDLSTPP  
3375 sp|P688721 MMVHLLTTT ESAVVTTTA AGKKNVVVDE EVALRRLVWV QQQFFESFFG GDDDDLSTPP  
3376 sp|P605241 ---VHLTTT ESLVWSSSGG GKGNVVVDE EVALRRLVWV QQQFFDSFFG GDDDDLSTPA  
3377 sp|P189831 MMMVHLLTTT EA AVVTTTG GS KKNVVVDE EVALRRLVWV QQQFFDSFFG GDDDDLSTPP  
3378  
3379 M M M VVHHKL TFTFGVNVILLI KKFFFSLLLL LHCCDKHVP PENNFLLII IVVVAAAAAA  
3380 QQQQQVHHKKI HFFFAILHHLLKK KFFFFSLLLL LHCCDKHVP PENNFLLIMI IVVVSSHHHH  
3381 KKKVVVHKKL DFFSMHHHLL KFFFFFALLL LHCCDKHVP PENNFLLVV VVVVAARRR  
3382 KKKVVVHKKL QFFSLLHHLL KFFFFFALLL LHCCDQHVP PENNFLLVV VVVVAARRR  
3383 KKKVVVHKKL AFGGLLHLL KFFFFFALLL LHCCDKHVP PENNFLLVV VVVVAARRR  
3384 KKKVVVHKKL NFFSLLNLL KFFFFFALLL LHCCDKHVP PENNFLLVV VVVVAHHHH  
3385 KKKVVVHKKL GFFSLLHHLL KFFFFFALLL LHCCDKHVP PENNFLLVV VVVVAHHHH  
3386 KKKVVVHKKL GFPSLLHHLL KFFFFFALLL LHCCDKHVP PENNFLLVV VVVVAHHHH  
3387 KKKVVVHKKL NFFSLLNLL KFFFFFALLL LHCCDKHVP PENNFLLVV VVVVAHHHH  
3388 KKKVVVHKKL NFFSLLNLL KFFFFFALLL LHCCDKHVP PENNFLLVV VVVVAHHHH  
3389  
3390 AFSSKTPPEC QQQQQWQKLL LLAARYH  
3391 HLGGKTPPSA QQQQQFHKVW VLAAYHYH  
3392 RHGGNTPPVL QQQQQFQKVW VLAAYHYH  
3393 RLGGHNPNV QQQQQFQKVW VLAAYHYH  
3394 RFGGKTPPEL QQQQQYQKVW VLAAYHYH  
3395 HFGGHNPQPV QQQQQFQKVW VLAAYHYH  
3396 HFGGKTPPPV QQQQQYQKVW VLAAYHYH  
3397 HFGGKTPPPV QQQQQYQKVW VLAAYHYH  
3398 HFGGKTPPOV QQQQQYQKVW VLAAYHYH  
3399 HFGGKTPQPV QQQQQYQKVW VLAAYHYH  
3400  
3401

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

programs will

## ➤ Neighbour Joining

So performing boot strap analysis on distance based phylogenetic method of hemoglobin sequences:  
 Open protdist in exe folder and type M to give number of multiple data sets as 100 and type Y to accept all the changed parameters:



Step 2: Observe distance for 100 datasets

Out file for all 100 data sets of hemoglobin sequences:

```
C:\phylip-3.695\exe\outfile_mult_dis_insulin - Notepad+
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outfile_mult_dis_insulin.outfile_insulin_ng.outfile_insulin_nj.outfile_insulin_upgma.dist.outfile_insulin_upgma_dist.hemoglobin_seq.phylip.outtree.outfile_seq_boot_hemo

2071 0.222409 0.222409 0.117799 0.182621
2072 sp|P02024| 0.266848 0.209663 0.228941 0.119643 0.131260 0.222409
2073 0.000000 0.000010 0.094188 0.072416
2074 sp|P68872| 0.266848 0.209663 0.228941 0.119643 0.131260 0.222409
2075 0.000010 0.000000 0.094188 0.072416
2076 sp|P60524| 0.241468 0.196620 0.217004 0.111599 0.161493 0.117799
2077 0.094188 0.094188 0.000000 0.064153
2078 sp|P18983| 0.241929 0.155042 0.173882 0.111347 0.126210 0.182621
2079 0.072416 0.072416 0.064153 0.000000
2080 10
2081 sp|P02112| 0.000000 0.419166 0.442415 0.490970 0.393485 0.478613
2082 0.374204 0.374204 0.387091 0.347597
2083 sp|P02094| 0.419166 0.000000 0.272031 0.318952 0.299644 0.318220
2084 0.223921 0.223921 0.261806 0.222066
2085 sp|P02075| 0.442415 0.272031 0.000000 0.178598 0.165807 0.257250
2086 0.166860 0.166860 0.218991 0.165265
2087 sp|P02067| 0.490970 0.318952 0.178598 0.000000 0.234608 0.173852
2088 0.191205 0.191205 0.205122 0.193425
2089 sp|P09905| 0.393485 0.299644 0.165807 0.234608 0.000000 0.305207
2090 0.122771 0.122771 0.199570 0.163501
2091 sp|P04244| 0.478613 0.318220 0.257250 0.173852 0.305207 0.000000
2092 0.194165 0.194165 0.107876 0.164129
2093 sp|P02024| 0.374204 0.223921 0.166860 0.191205 0.122771 0.194165
2094 0.000000 0.000010 0.089443 0.056909
2095 sp|P68872| 0.374204 0.223921 0.166860 0.191205 0.122771 0.194165
2096 0.000010 0.000000 0.089443 0.056909
2097 sp|P60524| 0.387091 0.261806 0.218991 0.205122 0.199570 0.107876
2098 0.089443 0.089443 0.000000 0.056641
2099 sp|P18983| 0.347597 0.222066 0.165265 0.193425 0.163501 0.164129
2100 0.056909 0.056909 0.058641 0.000000
```

C:\phyllip-3.695\exef\outfile\_mult\_dis\_insulin - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

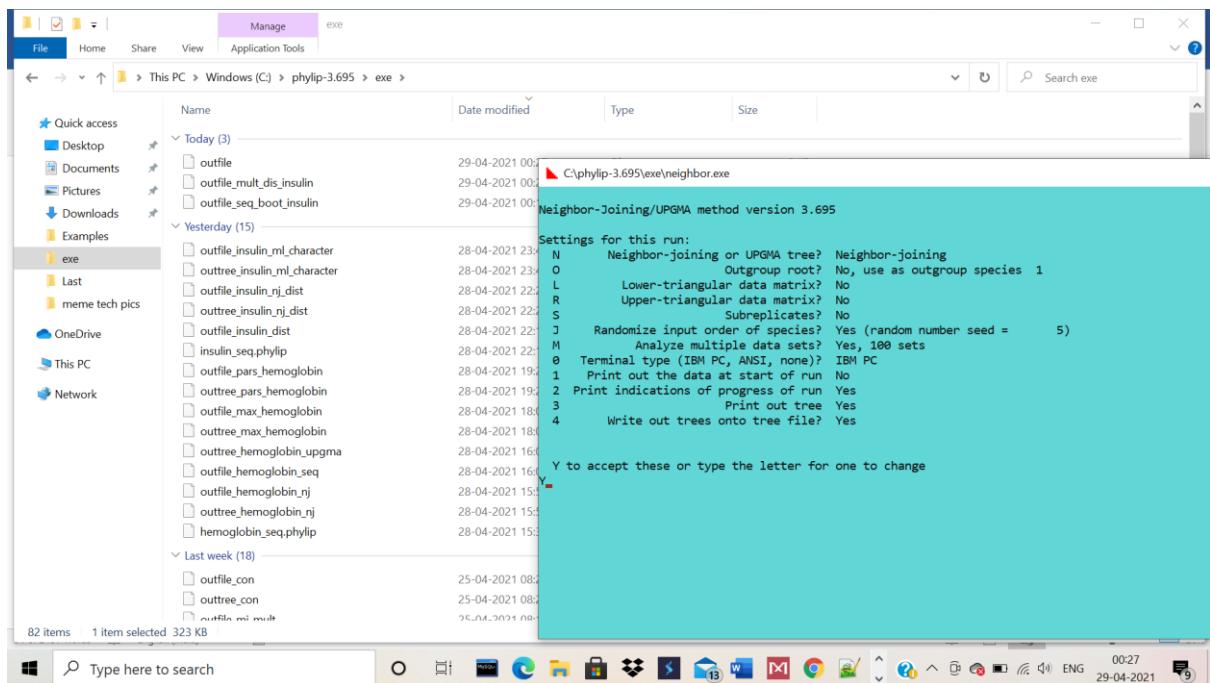
outfile\_mult\_dis\_insulin outfile\_insulin\_nj\_dists outfile\_insulin\_nj\_dists outfilefree\_insulin\_upgma\_dist outfile\_insulin\_upgma\_dist hemoglobin\_seq.phylip outfiletree outfile\_seq.boot.hemo

2041	sp P02094	0.347927	0.000000	0.271689	0.283234	0.294472	0.283620
2042		0.276739	0.285244	0.231698	0.235272		
2043	sp P02075	0.451017	0.271689	0.000000	0.166852	0.218012	0.225545
2044		0.205391	0.197012	0.213395	0.166324		
2045	sp P02067	0.453769	0.283234	0.166852	0.000000	0.228248	0.114359
2046		0.167045	0.158553	0.159340	0.148424		
2047	sp P09905	0.414110	0.294472	0.218012	0.228248	0.000000	0.274380
2048		0.176001	0.167625	0.182112	0.157770		
2049	sp P04244	0.437996	0.283620	0.225545	0.114359	0.274380	0.000000
2050		0.242155	0.233535	0.112464	0.182166		
2051	sp P02024	0.431661	0.276739	0.205391	0.167045	0.176001	0.242155
2052		0.000000	0.006956	0.121358	0.096256		
2053	sp P68872	0.421516	0.285244	0.197012	0.158553	0.167625	0.233535
2054		0.006956	0.000000	0.129111	0.103790		
2055	sp P60524	0.339732	0.231698	0.213395	0.159340	0.182112	0.112464
2056		0.121358	0.129111	0.000000	0.065552		
2057	sp P18983	0.376428	0.235272	0.166324	0.148424	0.157770	0.182166
2058		0.096256	0.103790	0.065552	0.000000		
2059		10					
2060	sp P02112	0.000000	0.277480	0.412055	0.262136	0.300800	0.306415
2061		0.266848	0.266848	0.241468	0.241929		
2062	sp P02094	0.277480	0.000000	0.253637	0.186971	0.217635	0.276244
2063		0.209663	0.209663	0.196620	0.155042		
2064	sp P02075	0.412055	0.253637	0.000000	0.196316	0.252735	0.219637
2065		0.228941	0.228941	0.217004	0.173882		
2066	sp P02067	0.262136	0.186971	0.196316	0.000000	0.201747	0.167156
2067		0.119643	0.119643	0.111599	0.111347		
2068	sp P09905	0.300800	0.217635	0.252735	0.201747	0.000000	0.287584
2069		0.131260	0.131260	0.161493	0.126210		
2070	sp P04244	0.306415	0.276244	0.219637	0.167156	0.287584	0.000000
2071		0.222409	0.222409	0.117799	0.182621		

Performing boot strap analysis on distance based phylogenetic method of hemoglobin sequences by **Neighbour joining** method:

>sp|P02024| = Gorilla Hemoglobin  
>sp|P68872| = Pan paniscus Hemoglobin  
>sp|P02112| = CHICK Hemoglobin  
>sp|P02067| = PIG Hemoglobin  
>sp|P02075| = SHEEP Hemoglobin  
>sp|P09905| = Physeter macrocephalus Hemoglobin  
>sp|P04244| = Panthera pardus orientalis Hemoglobin  
>sp|P02094| = Mesocricetus auratus Hemoglobin  
>sp|P60524| = Canis lupus Hemoglobin  
>sp|P18983| = Ailuropoda melanoleuca

Open neighbor in exe folder and type M to set number of data sets as 100 , random number seeds as 5 , than type Y to procced and accept all the shown parameters:



## Out tree for the neighbor joining method in notepad++ :

```
C:\phylib-3.695\exe\outtree_hemo_nj_dis - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_insulin_nj_dis.outfile_insulin_nj_dis.outtree_insulin_upgma_dis.outfile_insulin_upgma.phylib.outtree.outfile_seq.boot.hemo.outtree_hemo_nj_dis
1 ((sp|P09905|:0.11840,((sp|P02024|:0.00634,sp|P68872|:0.00755):0.01089,
2 (sp|P18983|:0.01221,sp|P60524|:0.04446):0.01467):0.01868,
3 ((sp|P02067|:0.06690,sp|P02075|:0.04448):0.01276,sp|P02441|:0.07773):0.02364):0.01446):0.03555,
4 sp|P02094|:0.11946,sp|P02112|:0.23666);
5 (sp|P02067|:0.03898,(((sp|P68872|:0.01722,sp|P02024|:0.02548):0.02963,
6 (sp|P18983|:0.03049,sp|P60524|:0.02676):0.02169):0.01539,
7 sp|P02441|:0.07006):0.01685,((sp|P02075|:0.07431,sp|P09905|:0.07211):0.01928,
8 sp|P02094|:0.17915):0.02734):0.01216,sp|P02112|:0.27087);
9 (((sp|P02075|:0.07621,sp|P02067|:0.07350):0.01909,((sp|P18983|:0.02877,
10 (sp|P02441|:0.09498,sp|P60524|:0.01816):0.03714):0.02377,
11 (sp|P09905|:0.13389,(sp|P68872|:0.00001,sp|P02024|:0.00001):0.04046):0.01656):0.01139):0.02835,
12 sp|P02094|:0.15557,sp|P02112|:0.23287);
13 (sp|P02094|:0.13936,(sp|P02067|:0.06844,(((sp|P68872|:0.00001,
14 (sp|P02024|:0.00001):0.02573,((sp|P02441|:0.10639,sp|P60524|:0.01495):0.02356,
15 sp|P18983|:0.01829):0.01389):0.01595,(sp|P02075|:0.10919,
16 sp|P09905|:0.09734):0.01412):0.01032):0.04284,sp|P02112|:0.23181);
17 ((sp|P02067|:0.09685,sp|P02094|:0.17783):0.01227,(((sp|P09905|:0.15454,
18 sp|P18983|:0.03967):0.03419,(sp|P68872|:0.00226,sp|P02024|:0.00475):0.06093):0.01289,
19 (sp|P60524|:0.01353,sp|P04244|:0.09571):0.04163):0.01509,
20 sp|P02075|:0.17362):0.00874,sp|P02112|:0.27037);
21 (sp|P02094|:0.11470,(sp|P09905|:0.12860,(sp|P02075|:0.05339,
22 (((sp|P02441|:0.07891,sp|P60524|:0.02443):0.01466,sp|P18983|:0.03627):0.01904,
23 (sp|P02024|:0.00472,sp|P68872|:0.00228):0.04925):0.01537,
24 sp|P02067|:0.08918):0.01281):0.01546):0.02080,sp|P02112|:0.28142);
25 (sp|P02094|:0.10196,(((sp|P68872|:0.00001,sp|P02024|:0.00001):0.03241,
26 sp|P09905|:0.10563):0.03721,(sp|P18983|:0.04036,(sp|P60524|:0.03123,
27 sp|P02441|:0.10155):0.02177):0.02936):0.01863,(sp|P02075|:0.11860,
28 sp|P02067|:0.05841):0.01619):0.05900,sp|P02112|:0.20558);
29 (((sp|P02075|:0.09854,sp|P09905|:0.10355):0.01615,((sp|P60524|:0.02202,
30 sp|P02441|:0.10177):0.01731,sp|P02067|:0.00486):0.02057,
31 (sp|P18983|:0.02468,(sp|P68872|:0.00215,sp|P02024|:0.00479):0.07165):0.01090):0.02030):0.03323,
```

## Out tree for the neighbor joining method in notepad++ for all 100 data sets :

```

C:\phylib-3.695\exe\outtree_hemo_nj_dis - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_insulin_nj_dis outfile_insulin_nj_dis outtree_insulin_upgma_dist outfile_insulin_upgma_dist hemoglobin_seq.phylip outtree outfile_seq_boot_hemo outtree_hemo_nj_dis
356 sp|P02024|:0.00352):0.01740):0.01646):0.02835):0.00443,sp|P02067|:0.08400):0.07077,
357 sp|P02094|:0.18307,sp|P02112|:0.29836);
358 (((sp|P02075|:0.08481,sp|P09905|:0.08800):0.02688,((sp|P04244|:0.06824,
359 (sp|P18983|:0.03441,(sp|P68872|:0.00001,sp|P02024|:0.00001):0.03514,
360 sp|P60524|:0.00620):0.00364):0.02657):0.01922,sp|P02067|:0.08592):0.00877):0.07599,
361 sp|P02094|:0.17030,sp|P02112|:0.30318);
362 (sp|P02094|:0.11928,((sp|P02024|:0.00453,sp|P68872|:0.00270):0.03536,
363 ((sp|P60524|:0.04387,(sp|P04244|:0.10198,sp|P02067|:0.06009):0.03531):0.02004,
364 (sp|P18983|:0.03034):0.00344):0.00983,(sp|P09905|:0.08885,
365 sp|P02075|:0.10324):0.02381):0.03816,sp|P02112|:0.25339);
366 (((sp|P18983|:0.02529,sp|P60524|:0.02550):0.02766,((sp|P02024|:0.01448,
367 sp|P68872|:0.00656):0.03981,sp|P09905|:0.14269):0.01035):0.02594,
368 (sp|P04244|:0.13957,(sp|P02075|:0.14688,sp|P02067|:0.00875):0.00617):0.01061):0.03781,
369 sp|P02094|:0.16351,sp|P02112|:0.25328);
370 (((sp|P02075|:0.08818,sp|P02067|:0.07253):0.00187,((sp|P02024|:0.00450,
371 sp|P68872|:0.00235):0.03348,sp|P09905|:0.09452):0.01586,
372 (sp|P18983|:0.02931,(sp|P60524|:0.01440,sp|P04244|:0.05630):0.01005):0.02062):0.00761):0.02917,
373 sp|P02094|:0.16837,sp|P02112|:0.22690);
374 (((sp|P04244|:0.07043,sp|P02067|:0.04393):0.03664,sp|P02075|:0.01238):0.01793,
375 (((sp|P68872|:0.00209,sp|P02024|:0.00486):0.06714,sp|P09905|:0.10119):0.01117,
376 (sp|P18983|:0.02799,sp|P60524|:0.03756):0.01143):0.00730):0.07301,
377 sp|P02094|:0.10120,sp|P02112|:0.24673);
378 (((sp|P02067|:0.05240,(sp|P02075|:0.12227,sp|P04244|:0.09737):0.01952):0.01084,
379 sp|P60524|:0.03779):0.00792,(sp|P18983|:0.02075,((sp|P02024|:0.00001,
380 sp|P68872|:0.00001):0.03996,sp|P09905|:0.09130):0.01293):0.00795):0.02727,
381 sp|P02094|:0.10194,sp|P02112|:0.17554);
382 (((sp|P09905|:0.08051,sp|P02075|:0.08530):0.02452,(((sp|P02067|:0.08595,
383 sp|P04244|:0.08790):0.03323,sp|P60524|:0.03634):0.01970,
384 sp|P18983|:0.02076):0.00903,(sp|P68872|:0.00001,sp|P02024|:0.00001):0.03060):0.01539):0.03489,
385 sp|P02094|:0.13945,sp|P02112|:0.27971);
386

```

You will get unrooted and negative branched out file tree for neighbour joining method:

```

C:\phylib-3.695\exe\outfile_hemo_nj_dis - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_insulin_nj_dis outfile_insulin_nj_dis outtree_insulin_upgma_dist outfile_insulin_upgma_dist hemoglobin_seq.phylip outtree outfile_seq_boot_hemo outfile_hemo_nj_dis
1
2 10 Populations
3
4 Neighbor-Joining/UPGMA method version 3.695
5
6 Data set # 1:
7
8 Neighbor-joining method
9
10 Negative branch lengths allowed
11
12
13 +-----sp|P09905|
14 !
15 ! +sp|P02024|
16 ! +7
17 +-4 ! +sp|P68872|
18 ! ! +-8
19 ! ! ! +sp|P18983|
20 ! ! ! +6
21 ! +-5 +-sp|P60524|
22 !
23 ! ! +---sp|P02067|
24 ! ! +2
25 ! +-3 +-sp|P02075|
26 !
27 ! +----sp|P04244|
28 !
29 1-----sp|P02094|
30 !
31 +-----sp|P02112|

```

For all 100 data sets you will get unrooted and negative branched out file tree for neighbour joining method:

```

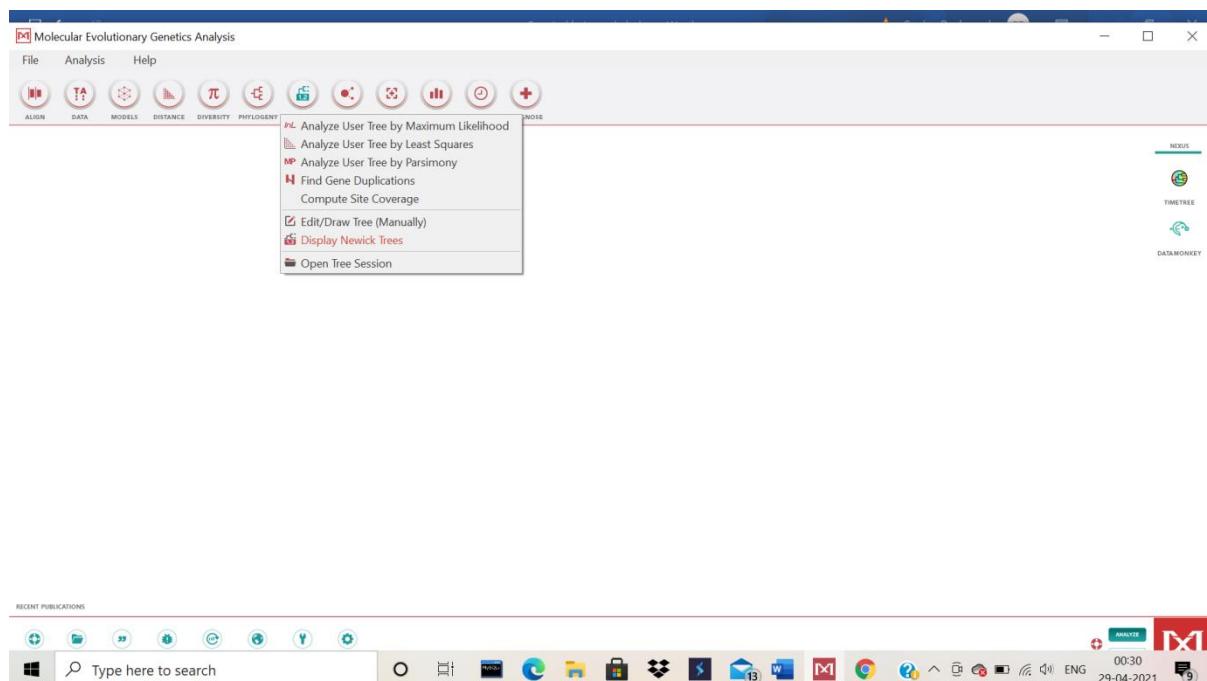
C:\phylip-3.695\exe\outfile_hemo_nj_dis - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outfile_insulin_nj_dis outfile_upgma_dist outfile_insulin_upgma_dist hemoglobin_seq.phylip outfile_tree outfile_seq_boot_hemo outfile_hemo_nj_dis outfile_hemo_nj_dis
5076 ! +sp|P02024|
5077 !
5078 2-----sp|P02094|
5079 !
5080 +-----sp|P02112|
5081
5082
5083 remember: this is an unrooted tree!
5084
5085 Between And Length
5086 -----
5087 2 6 0.03489
5088 6 4 0.02452
5089 4 sp|P09905| 0.08051
5090 4 sp|P02075| 0.08530
5091 6 8 0.01539
5092 8 7 0.00903
5093 7 5 0.01970
5094 5 1 0.03323
5095 1 sp|P02067| 0.08595
5096 1 sp|P04244| 0.08790
5097 5 sp|P60524| 0.03634
5098 7 sp|P18983| 0.02076
5099 8 3 0.03060
5100 3 sp|P68872| 0.00001
5101 3 sp|P02024| 0.00001
5102 2 sp|P02094| 0.13945
5103 2 sp|P02112| 0.27971
5104
5105
5106

```

Normal text file length : 1,22,526 lines : 5,106 Ln : 5,106 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS

Type here to search 00:55 29-04-2021 ENG

Open MEGA software to view newick phylogenetic tree and go to user tree, click on display newick trees:



You will view 100 distance based phylogenetic newick tree by bootstrap method for hemoglobin sequences for all 100 data sets:

1 of 100