

Caption:

Evolutionary relationships of taxa

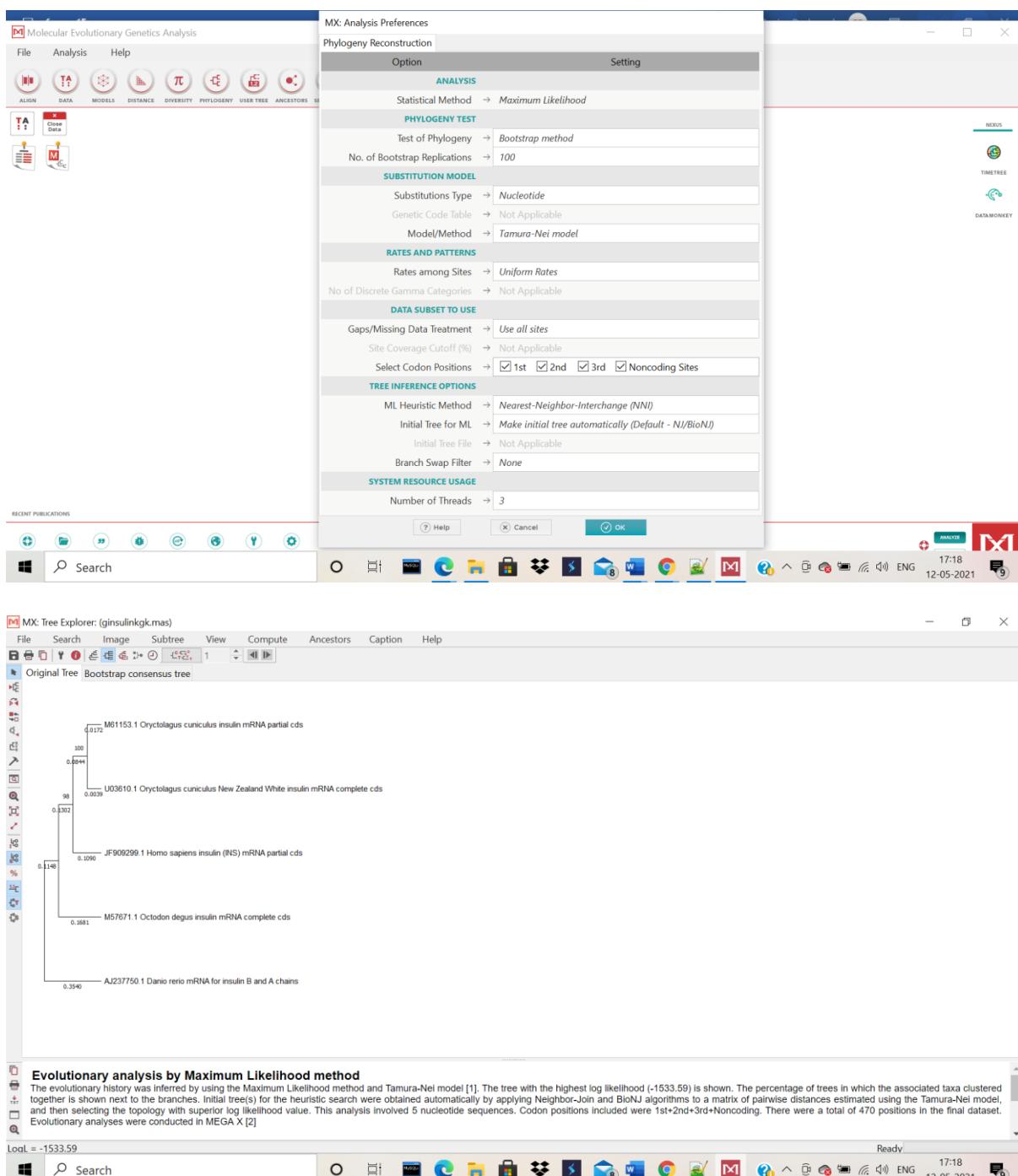
The evolutionary history was inferred using the Neighbor-Joining method [1]. The bootstrap consensus tree inferred from 100 replicates [2] is taken to represent the evolutionary history of the taxa analyzed [2]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches [2]. The evolutionary distances were computed using the Maximum Composite Likelihood method [3] and are in the units of the number of base substitutions per site. This analysis involved 5 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 470 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [4]

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

Answer:

>M57671.1 Octodon degus insulin mRNA, complete cds
 >M61153.1 Oryctolagus cuniculus insulin mRNA, partial cds
 >U03610.1 Oryctolagus cuniculus New Zealand White insulin mRNA, complete cds
 >JF909299.1 Homo sapiens insulin (INS) mRNA, partial cds
 >AJ237750.1 Danio rerio mRNA for insulin, B and A chains

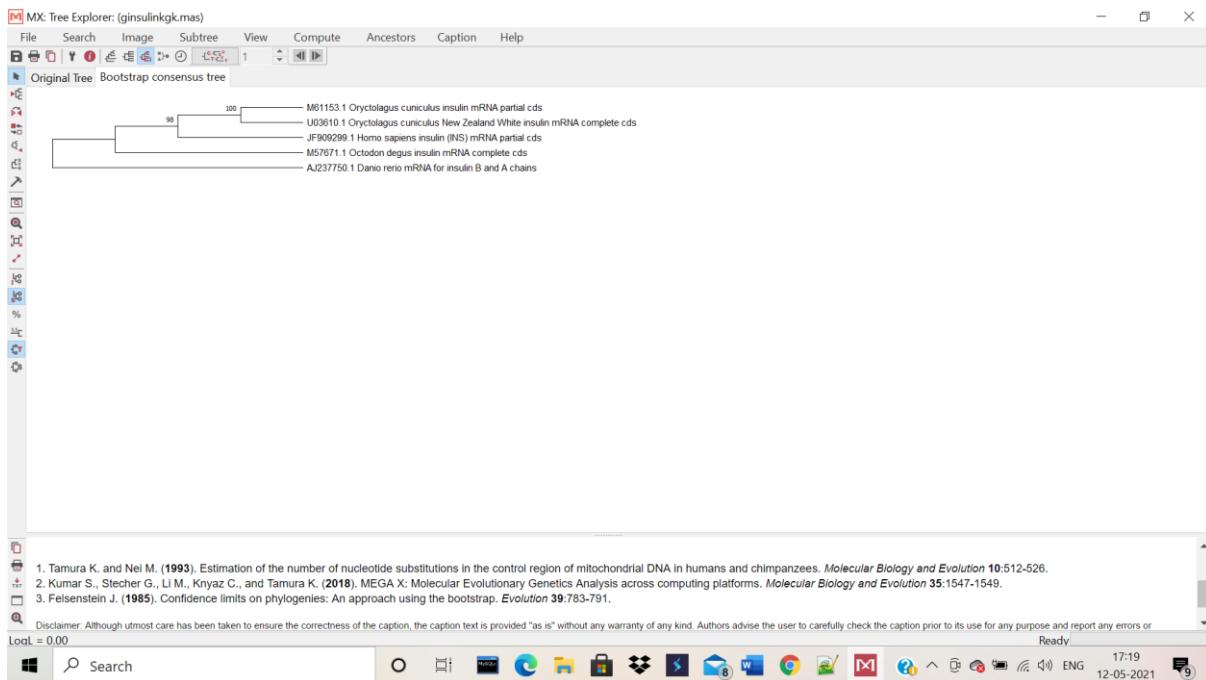
Maximum Likelihood:



Caption:

Evolutionary analysis by Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-1533.59) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. This analysis involved 5 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 470 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [2]

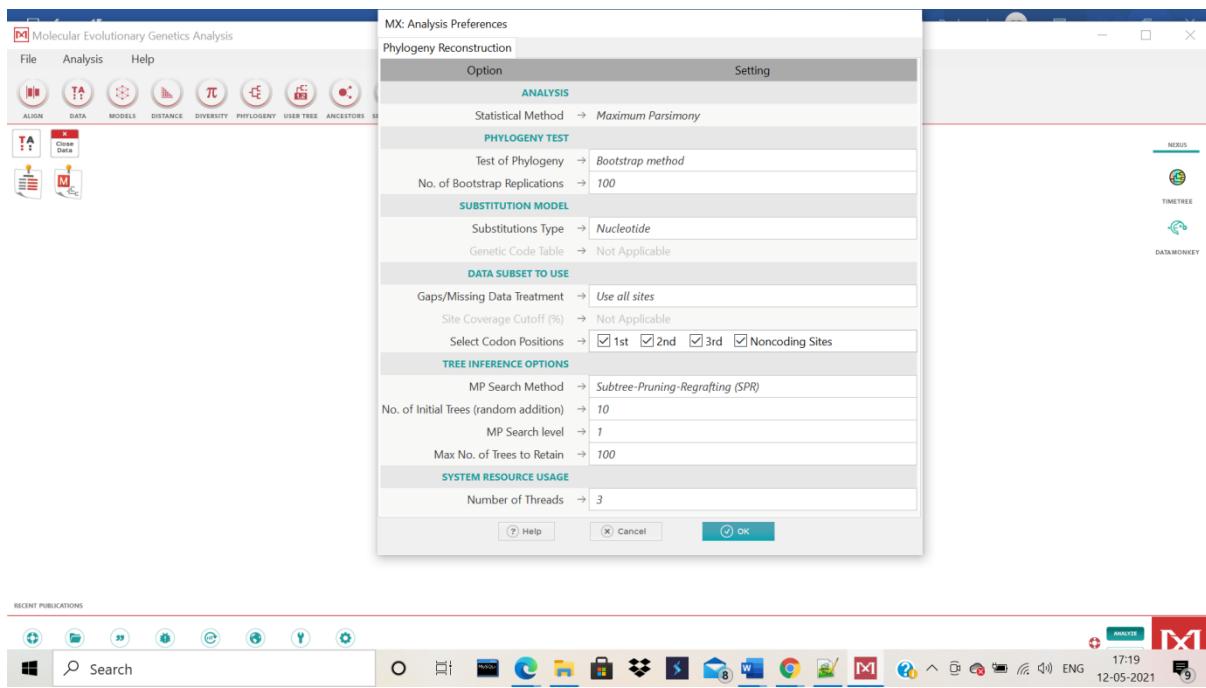


Caption:

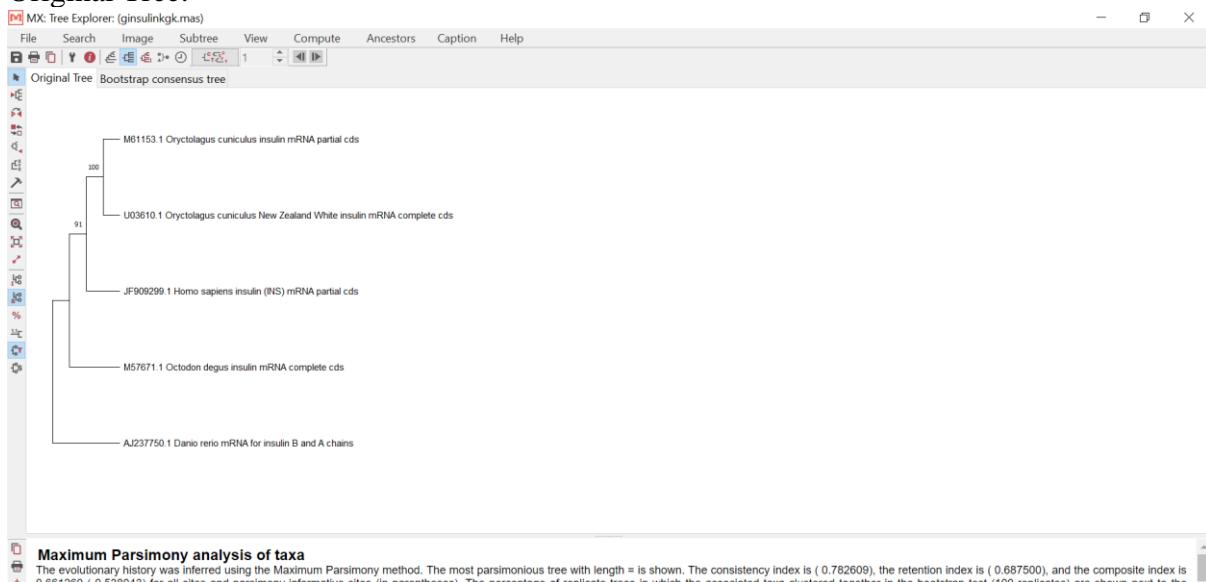
Evolutionary analysis by Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The bootstrap consensus tree inferred from 100 replicates [3] is taken to represent the evolutionary history of the taxa analyzed [3]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches [3]. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. This analysis involved 5 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 470 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [2]

Parsimony



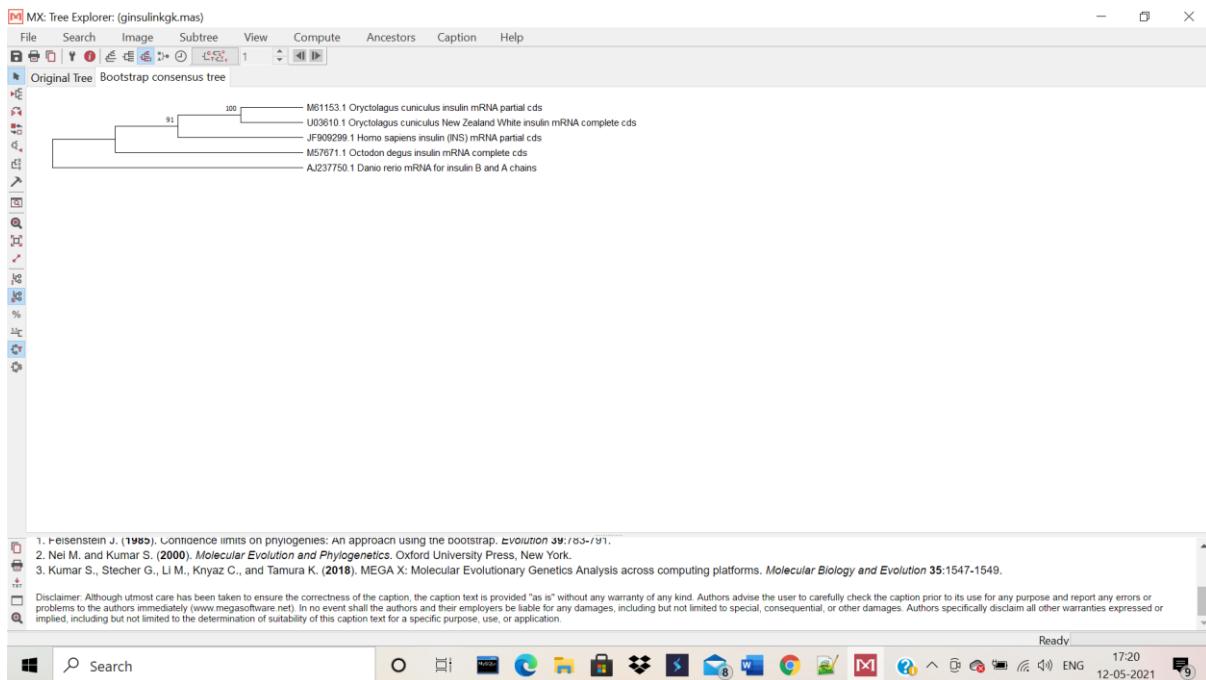
Original Tree:



Caption:

Maximum Parsimony analysis of taxa

The evolutionary history was inferred using the Maximum Parsimony method. The most parsimonious tree with length = is shown. The consistency index is (0.782609), the retention index is (0.687500), and the composite index is 0.661260 (0.538043) for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches [1]. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (pg. 126 in ref. [2]) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). Branch lengths were calculated using the average pathway method [see pg. 132 in ref. 2] and are in the units of the number of changes over the whole sequence. They are shown next to the branches. This analysis involved 5 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 470 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [3]



Caption:

Maximum Parsimony analysis of taxa

The evolutionary history was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 100 replicates is taken to represent the evolutionary history of the taxa analyzed [1]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches[1]. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (pg. 126 in ref. [2]) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). This analysis involved 5 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 470 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [3]

Exercise 2:

Phylogenetics Analysis In PHYLIP software:

1. Perform distance based phylogenetic analysis of hemoglobin sequences of 10 different organisms (protein sequences-UPGMA and NJ method)

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.

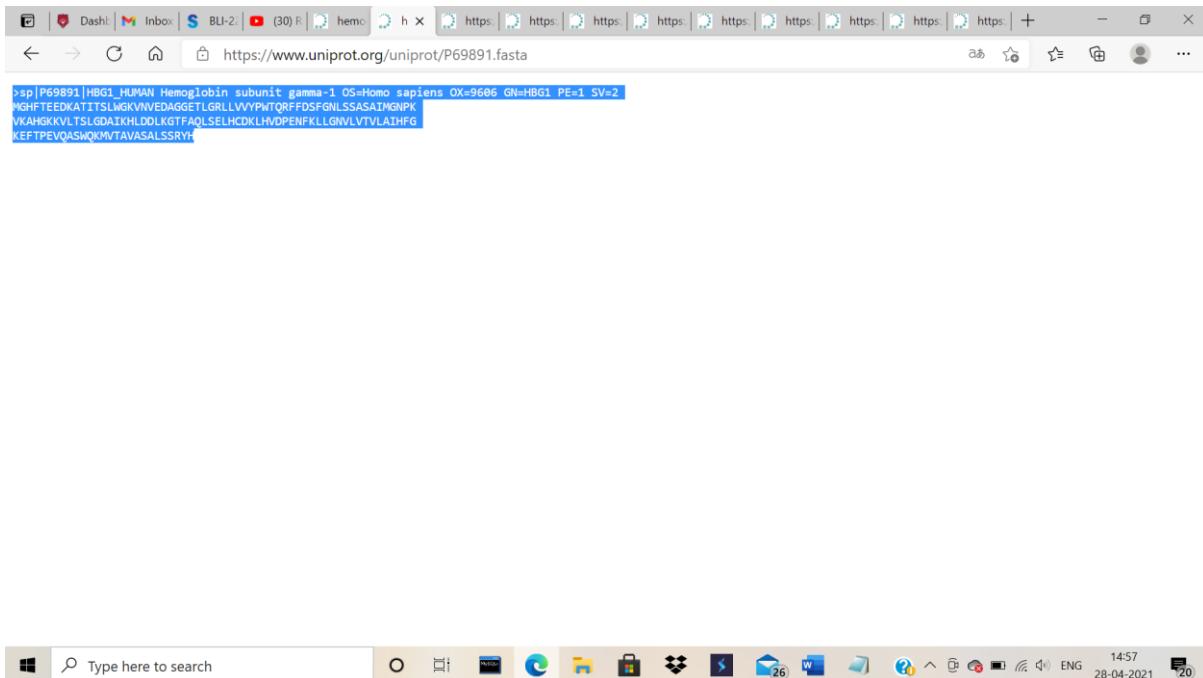
Select your 10 hemoglobin protein sequence for different organisms from uniprot:

```
>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
```

The screenshot shows the UniProtKB 2021_02 results page. The search bar at the top contains "hemoglobin protein". Below the search bar, there are links for BLAST, Align, Retrieve/ID mapping, Peptide search, and SPARQL. The main content area is titled "UniProtKB 2021_02 results". It displays two sections: "Reviewed (Swiss-Prot) - Manually annotated" and "Unreviewed (TrEMBL) - Computationally analyzed". The "Reviewed" section has 1,135 entries, and the "Unreviewed" section has 25,340 entries. A table below shows the first few entries for the reviewed section, including columns for Entry, Entry name, Protein names, Gene names, Organism, and Length. The table rows include P9WN23 (TRHBO_MYCTU), P69891 (HBG1_HUMAN), and P6Q9Q2 (HRG2_HUMAN). The UniProt logo is visible in the top left corner of the page.

Entry	Entry name	Protein names	Gene names	Organism	Length
P9WN23	TRHBO_MYCTU	Group 2 truncated hemoglobin GlbO	glbO Rv2470, MTV008.26	Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv)	128
P69891	HBG1_HUMAN	Hemoglobin subunit gamma-1	HBG1 PRO2979	Homo sapiens (Human)	147
P6Q9Q2	HRG2_HUMAN	Hemoglobin subunit	HRG2	Homo canis (Human)	147

Go to sequences and download the fasta format of each organism:



Copy paste each fasta sequence in notepad for all 10 different organisms:

```
*Untitled - Notepad
File Edit Format View Help
>sp HUMAN Hemoglobin
MGHTFEEDKATITSLNKGKVIVEDAGGETLGRLLVVYPTQRFDFSGNLSSASA1MGNPK
VIKAHGKKVLTSLGDAIKHLDLKGTFQALSELHCDKLHVDPENFKLLGNVLTVLA1HFG
KEFTPEVEQASWQMVMTAVASALSSRYH

>sp RAT Hemoglobin
MWHLTDAEKAAVANGLWKVNPDDVGEALGRLLVVYPWTQRFDFSGNLSSASA1MGNPK
VIKAHGKKVLTSLGDAIKHLDLKGTFQALSELHCDKLHVDPENFKLLGNVLTVLA1HFG
KEFTPECAQAAFQKVAGVASALAHRYH

>sp BOVIN Hemoglobin
MLTAEEKAAVTAFWKGKVKDEVGEALGRLLVVYPWTQRFDFSGNLSTADAAMNNPKVK
AIGHKKVLDLSFSNGMKHLDLKGTFQALSELHCDKLHVDPENFKLLGNVLVVVLA1HFG
FTPVLQADFQKVAGVASALAHRYH

>sp ECOLX Escherichia coli
MRIRIYSLRYSAVARGFIAVSEFARKCVHKSVRRLCFPVLLIPVLFSGSLAGTVNNELG
YQLFRDFAENKGMPRGATNIAYNKQGEFVGLDKAAMPDFSAVDSEIGVATLNPQVI
ASVKHNGYGTINVSFGDENRYNIIVDRNNAAPSDFHAPRDLKLVTVEAPTAQGAVAGA
YLDKERYPVFVRLGSGTQYIKDSNGLTKMGGAYSWLTGGTVGSLSSQNGEMISTSSL
VFDFYKLNGAMPYIEADGSPLFAFDTVQNWKWLVLGVLTAGNGAGGRGNNAV1PLDFI
GQKFNEENDNAPVTFTSEGGALEWSFNSSSTGAGALTQGTTTYAHHQCGNDLNAGKNLIF
QGQNGQINLKDSVSGAGSLTFRDNTYVTTNSWTITGAGIVDNGSVNVWQVNGVKGDN
LHKIGEGLTLVQGTG1NEGGGLKVGDGKVKVNQQADNGQVQAFSSVNAISGRPTVLTDE
RQVNPDTDSVSWGYRGRTLDVNGNSLTFHQLKAADYGAVLANNVDRATZTLDYALRADKVA
LNGWESGGKGTAGNLVYKNNPYTNTDFILQKSTYGYFPTDQSMMATWEFVGHSGQDAQ
KLVADRFNTAGYLFHGQLKGNNLVDRLPPEGTVGALVMDGAADISGTTFQENGRLTQGH
PVJTHAYNTQSADKLAQSDHSLVLTQTSFSEQDWENRSEFTFDRLSLKNTDFGLGRNATL
NTTQADNSVTLGDSRVFIDKNDGGTATFLLEEGTSVATKADAKSVFNGTVNLDNQSVL
NNDIFNGGIQANNSTVVISDSAVLGNSTLTSTALNNGANALASQFVSDGPVNISD
ATLSLNSRPDEVSHTLLPVYDAGSNLKGDDARLNVPYSMILSGNININVQDGTVTLGGE
GELSPDLTLQNMQLYSLFNGYRNWIWSGSLNAPDATVSMTDQWSMMNGNSTAGNMKLRTI
VGFNGTGSPFTTLTONLDAVOSAFVMRTDLNKADKLVINKSATGHDNSIWVFLKKPSN
KDTLDIPLVSAPEATADNLFRASTRVVGFSDVTPILSRKEDGKKEWLGDGYQVARNDQ
```

Type ‘DNA’ in the drop box of “enter or paste a set of”. Change the output to “PHYLIP”. Copy paste each fasta sequence in clustal omega for all 10 different organisms.

Dashboard | Inbox (337) - gunja | BLU-225: Communic | (30) Tu Jaane | hemoglobin protein | Clustal Omega < M | +

EMBL-EBI Services Research Training Industry About us Search

EMBL-EBI Hinxton

Clustal Omega

Input form Web services Help & Documentation Bioinformatics Tools FAQ

Feedback Share

Tools > Multiple Sequence Alignment > Clustal Omega

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.

STEP 1 - Enter your input sequences

Enter or paste a set of

PROTEIN

sequences in any supported format:
HDFNPNVNVQAAFQKVAGVNALAHKYH

Type here to search

Download Alignment File:

10 147
sp |P02112|MVHVTAAEKKOLITGLWGVNVNVAECGEAELARLLVPPWTORFFASGNLS
sp |P02094|MVHLTDAEKALVTLGWLWGVNVNADEVGEAELGRLLVVPPWTORFEHFQDLS
sp |P02875|,-MLTAEEKAATVGFWGKVNVDEVEGAELGRLLVVPPWTORFEHFQDLS
sp |P02867|MVHLTAAEKKLWAVGLWGVNVDEVEGAELGRLLVVPPWTORFEHFQDLS
sp |P09985|,VHLTGEKSGLTALWAKNVNVVEEIGEALGRLLVVPPWTORFEHFQDLS
sp |P04244|MSFLSAEKKNLVSGLWGVNVDEVEGAELGRLLVVPPWTORFQSFQDLS
sp |P02824|MVHLTPZEKAATVGLWGVNVDEVEGAELGRLLVVPPWTORFESFQDLS
sp |P68872|MVHLTPZEKAATVGLWGVNVDEVEGAELGRLLVVPPWTORFESFQDLS
sp |P60524|VHLTAAEKKLWAVGLWGVNVDEVEGAELGRLLVVPPWTORFDSDQDLS
sp |P18983|MVHLTGEKAATVGLWGVNVDEVGGEALGRLLVVPPWTQRFFDSFGDLS

SPTAILGNPWPRAHGKVLTSFGDVKNLNLIKNTFSQSELHCDI LHD
SASAIVMNPQVKAHGKVKTHFSADGLKLHDLNLKGAFSSLSSELHCDI LHD
NADAVMNPQVKAHGKVKLDSFSNGWQHDLKLGTFAQLSELHCDI LHD
NADAVMNPQVKAHGKVKLQFSQDGLKLHDLNLKGTFAKSELHCDQJLH
DAVAMNPQVKKHQOKVLASFEGGLHLNDNLKGTFATLSELHCDI LHD
SADAJMSNAVKVKAHGKVKLNSFSQDGLKNIQDGLKGAFAKSELHCDI LHD
TPDAVMNPQVKAHGKVKLAFSGDGLAHLDNLKGTFATLSELHCDI LHD
TPDAVMNPQVKAHGKVKLAFSGDGLAHLDNLKGTFATLSELHCDI LHD
TPDAVMNPQVKAHGKVKLNSFSQDGLKLNLDNLKGTFAKSELHCDI LHD
TPDAVMNPQVKAHGKVKLNSFSQDGLKLNLDNLKGTFAKSELHCDI LHD

PENFRLQLDILITVLALAHPSKDFTPCEQAIIWOKLVRVVAHALARKYH
PENFKLQLGMIIIZTVSHOLGIKDFTPSAQSFAHVIVVAGVANALAHKYH
PENFRLQLGVNVLVLVVAHARHGKFETPVLOADFQKVVAGVANALAHKYH
PENFRLQLGVNVLVLVVAHARLGHFDNNVQOAFQKVVAGVANALAHKYH
PENFRLQLGVNVLVLVVAHMFKEFTPELOTAQXQVAGVANALAHKYH
PENFRLQLGVNLVCVLVAHMFQHENPQOQAFQKVVAGVANALAHRYH
PENFKLQLGVNLVCVLVAHMFKEFTPQOQAYQKVVAGVANALAHYH
PENFRLQLGVNLVCVLVAHMFKEFTPQOQAYQKVVAGVANALAHYH
PENFKLQLGVNLVCVLVAHMFKEFTPQOQAYQKVVAGVANALAHYH

15:31
28-04-2021 ENG

Distance Matrix for hemoglobin protein sequences for 10 different organisms:

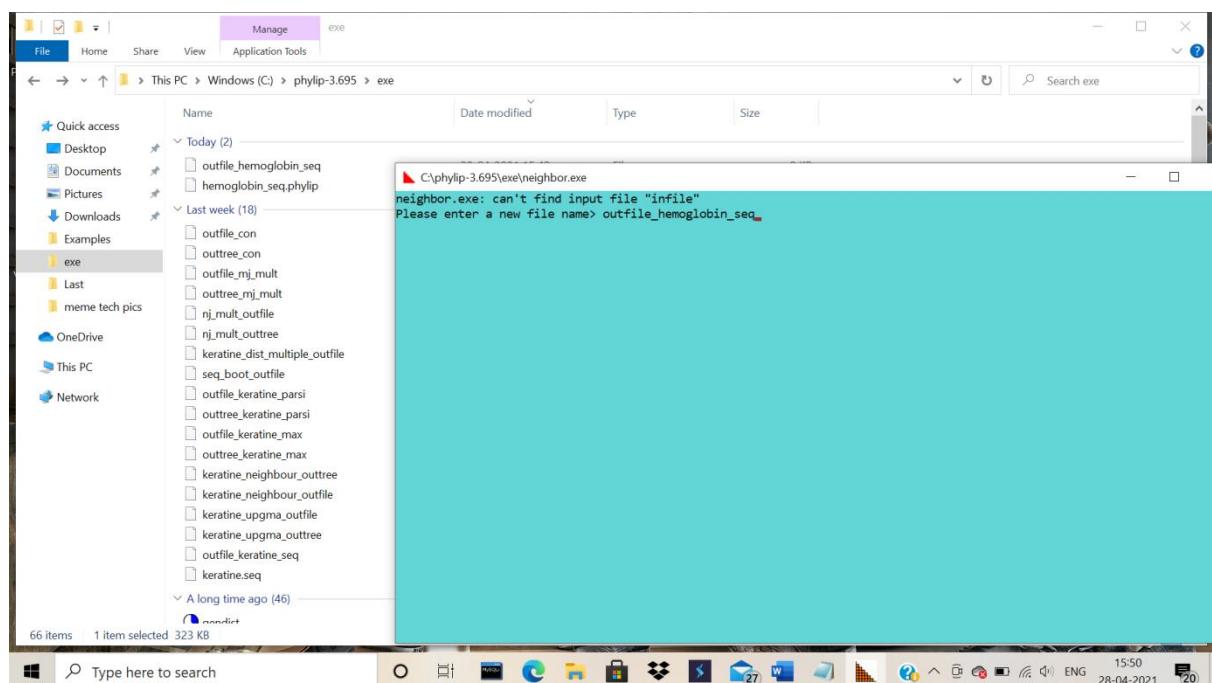
```

C:\phylip-3.695\exe\outfile_hemoglobin_seq - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outfile_keratine_max outfile_keratine_max outfile_keratine_parsi outfile_keratine_parsi seq_boot_outfile outfile_mj_mult outfile_mj_mult outfile_con outfile_hemoglobin_seq
1 10
2 sp|P02112| 0.000000 0.436579 0.447520 0.426213 0.438591 0.443980
3 0.404947 0.395195 0.353874 0.382129
4 sp|P02094| 0.436579 0.000000 0.296793 0.279049 0.333256 0.354713
5 0.288472 0.297027 0.291129 0.272213
6 sp|P02075| 0.447520 0.296793 0.000000 0.193049 0.226737 0.251436
7 0.219113 0.210489 0.231091 0.198394
8 sp|P02067| 0.426213 0.279049 0.193049 0.000000 0.250192 0.195463
9 0.185761 0.177389 0.188678 0.186290
10 sp|P09905| 0.438591 0.333256 0.226737 0.250192 0.000000 0.322109
11 0.174979 0.166660 0.215833 0.178760
12 sp|P04244| 0.443980 0.354713 0.251436 0.195463 0.322109 0.000000
13 0.230725 0.222278 0.137162 0.188564
14 sp|P02024| 0.404947 0.288472 0.219113 0.185761 0.174979 0.230725
15 0.000000 0.007012 0.104683 0.096720
16 sp|P68872| 0.395195 0.297027 0.210489 0.177389 0.166660 0.222278
17 0.007012 0.000000 0.112300 0.104279
18 sp|P60524| 0.353874 0.291129 0.231091 0.188678 0.215833 0.137162
19 0.104683 0.112300 0.000000 0.065058
20 sp|P18983| 0.382129 0.272213 0.198394 0.186290 0.178760 0.188564
21 0.096720 0.104279 0.065058 0.000000
22

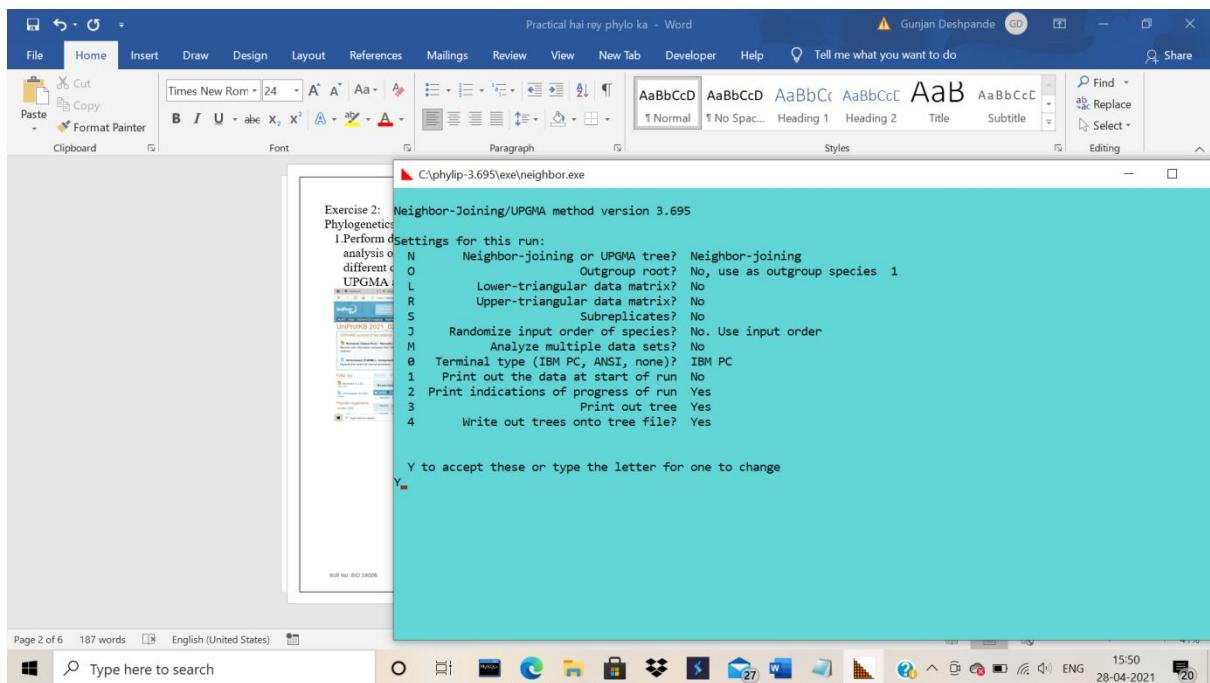
```

Normal text file length : 1.147 lines : 22 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS
Type here to search 15:43 28-04-2021 20

Performing distance based phylogenetic analysis of hemoglobin sequences of 10 different organisms using **Neighbor Joining Method**:



Type 'Y' to accept all the default parameters shown:



Out tree for neighbour- joining method of the sequences:

The screenshot shows a Notepad++ window with the following text:

```

C:\phylip-3.695\exe\outtree_hemoglobin.nj - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_keratine.max outtree_keratine.pars1 outfile_keratine_pars1 seq_boot_outfile outtree_mj_mult outfile_mj_mult outtree.con outfile_hemoglobin_seq outtree_hemoglobin_nj
1 (sp|P02094|:0.16311,((sp|P02075|:0.10247,sp|P02067|:0.09058):0.01184,
2 (sp|P09905|:0.11783,(((sp|P04244|:0.10623,sp|P60524|:0.03093):0.01879,
3 sp|P18983|:0.03944):0.01151,(sp|P02024|:0.00474,sp|P68872|:0.00227):0.05558):0.01252):0.01287):0.03574,sp|P02112|:0.27347);
4

```

The Notepad++ status bar indicates: length : 268 lines : 4 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS

Un rooted tree by neighbor joining method in the outfile when viewed in notepad ++ :

```

C:\phylip-3.695\exe\outfile_hemoglobin_nj - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_keratine_parsi outfile_keratine_parsi seq_boot_outtree outtree_mj_mult outfile_mj_mult outtree_con outfile_hemoglobin_seq outfile_hemoglobin_nj outfile_hemoglobin_nj

1 10 Populations
2
3 Neighbor-Joining/UPGMA method version 3.695
4
5
6 Neighbor-joining method
7
8 Negative branch lengths allowed
9
10
11 +-----sp|P02094|
12 !
13 ! +----sp|P02075|
14 ! +-7
15 ! ! +---sp|P02067|
16 ! !
17 !
18 2-8 +----sp|P09905|
19 ! !
20 ! ! ! +----sp|P04244|
21 ! +-6 +-3
22 ! ! +4 +-sp|P60524|
23 ! !
24 ! +5 +-sp|P18983|
25 !
26 ! ! +sp|P02024|
27 ! +-1
28 ! +sp|P68872|
29 !
30 +-----sp|P02112|
31

```

Normal text file length : 1,276 lines : 56 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS
15:53 ENG 28-04-2021 20

```

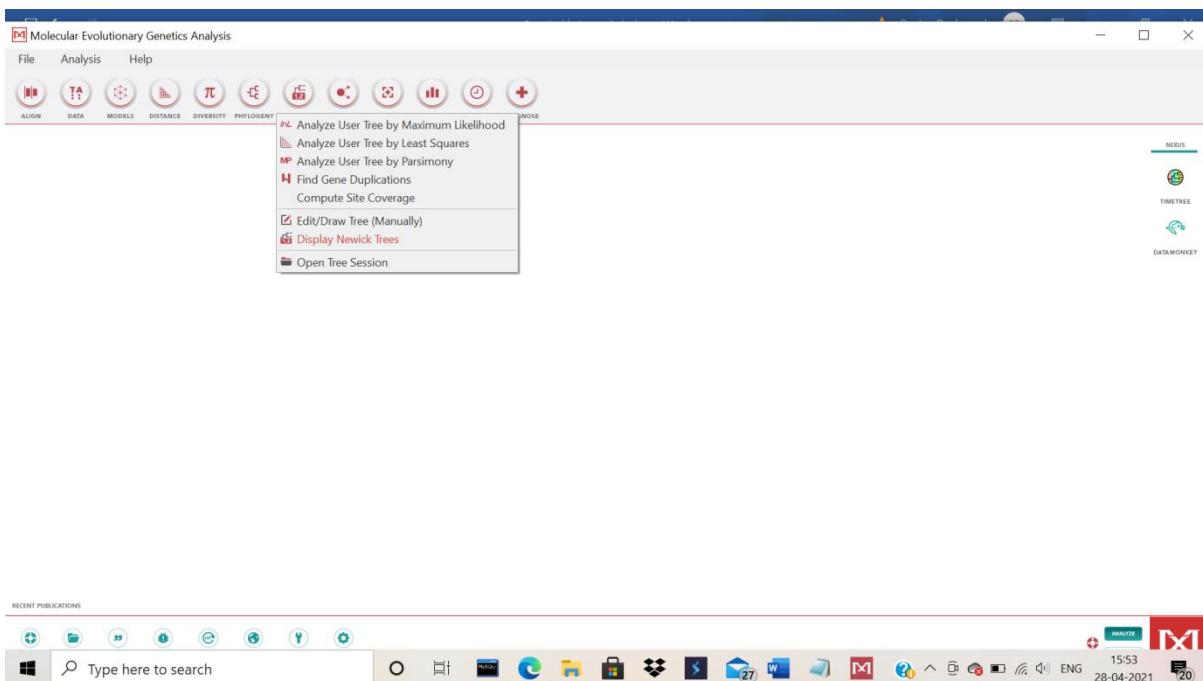
C:\phylip-3.695\exe\outfile_hemoglobin_nj - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_keratine_parsi outfile_keratine_parsi seq_boot_outtree outtree_mj_mult outfile_mj_mult outtree_con outfile_hemoglobin_seq outfile_hemoglobin_nj outfile_hemoglobin_nj

26 ! ! +sp|P02024|
27 !
28 ! +sp|P68872|
29 !
30 +-----sp|P02112|
31
32
33 remember: this is an unrooted tree!
34
35 Between And Length
36 -----
37 2 sp|P02094| 0.16311
38 2 8 0.03574
39 8 7 0.01184
40 7 sp|P02075| 0.10247
41 7 sp|P02067| 0.09058
42 8 6 0.01287
43 6 sp|P09905| 0.11783
44 6 5 0.01252
45 5 4 0.01151
46 4 3 0.01879
47 3 sp|P04244| 0.10623
48 3 sp|P60524| 0.03093
49 4 sp|P18983| 0.03944
50 5 1 0.05558
51 1 sp|P02024| 0.00474
52 1 sp|P68872| 0.00227
53 2 sp|P02112| 0.27347
54
55
56

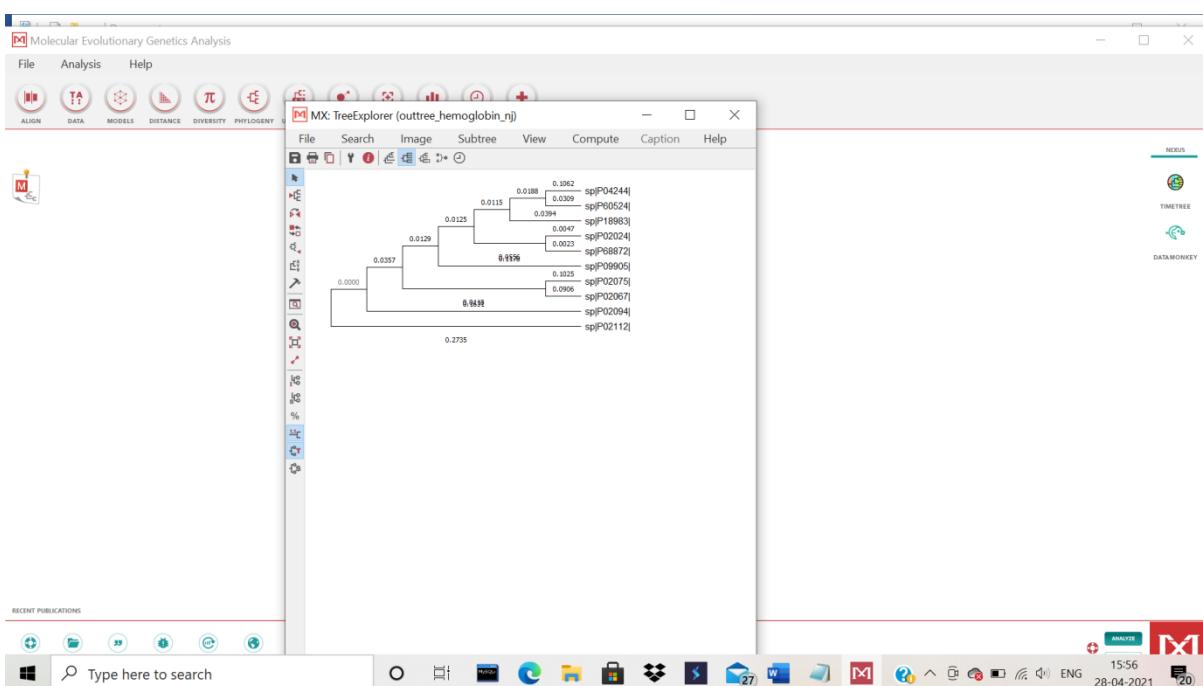
```

Normal text file length : 1,276 lines : 56 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS
15:53 ENG 28-04-2021 20

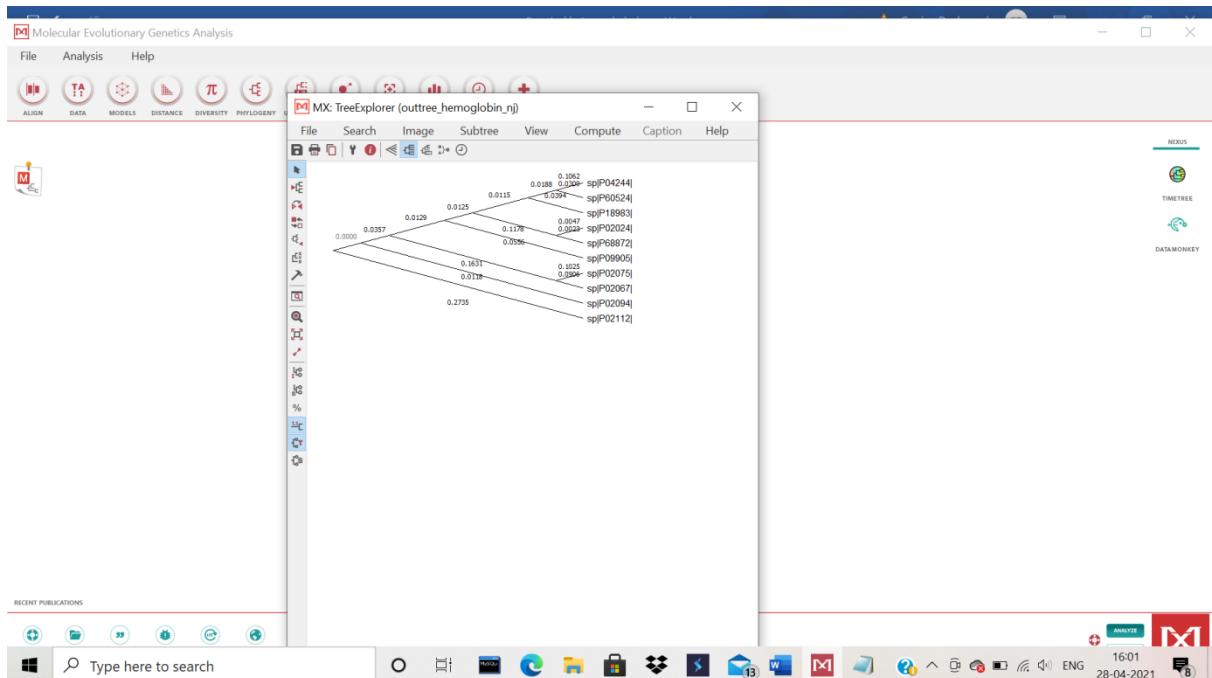
Open MEGA software to view newick phylogenetic tree and go to user tree, click on display newick trees:



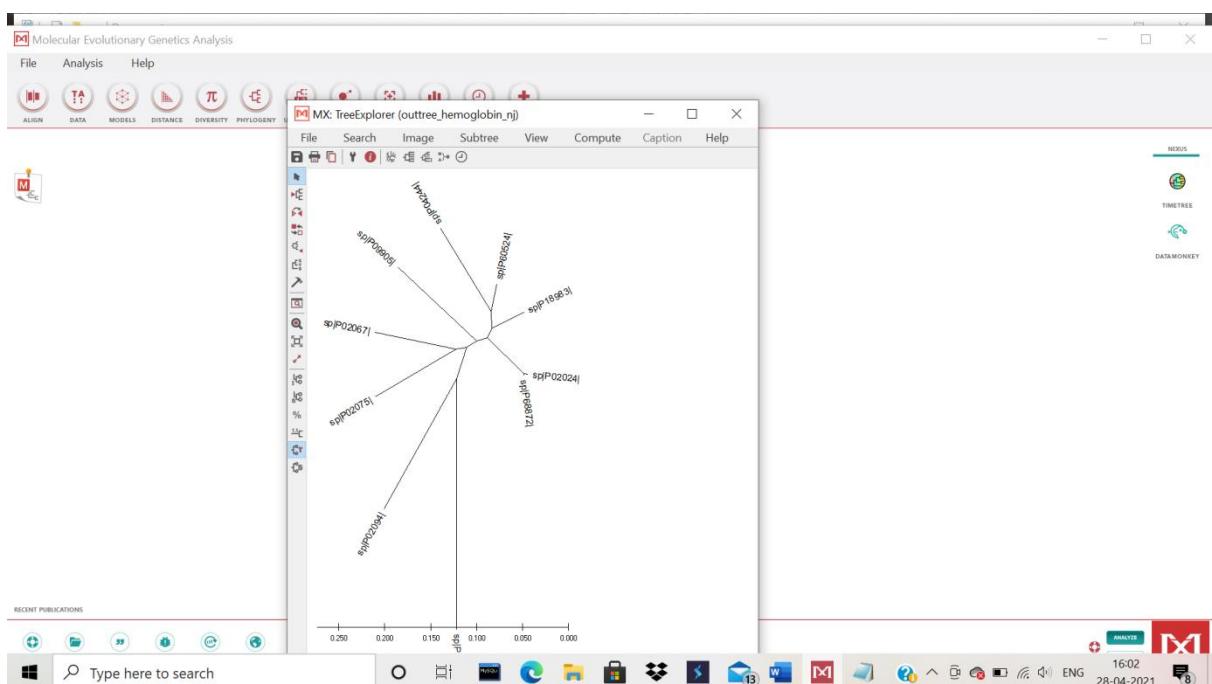
Newick tree in MEGA software for neighbor joining method:



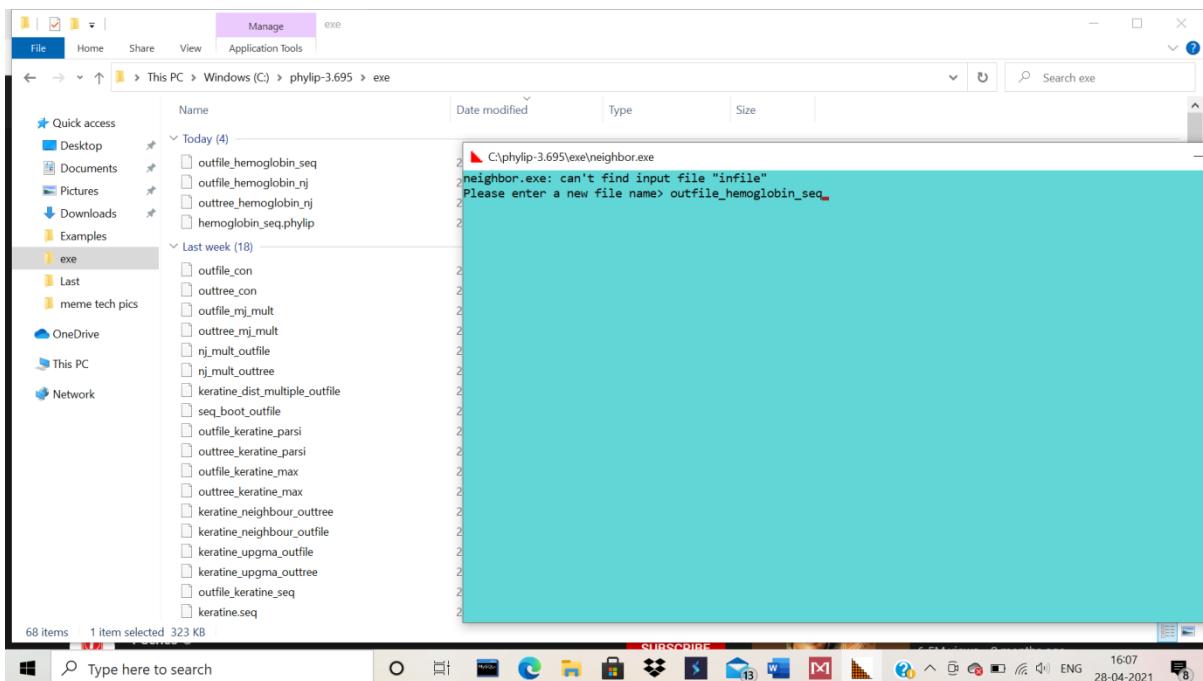
You can view the tree in different formats –
Straight :



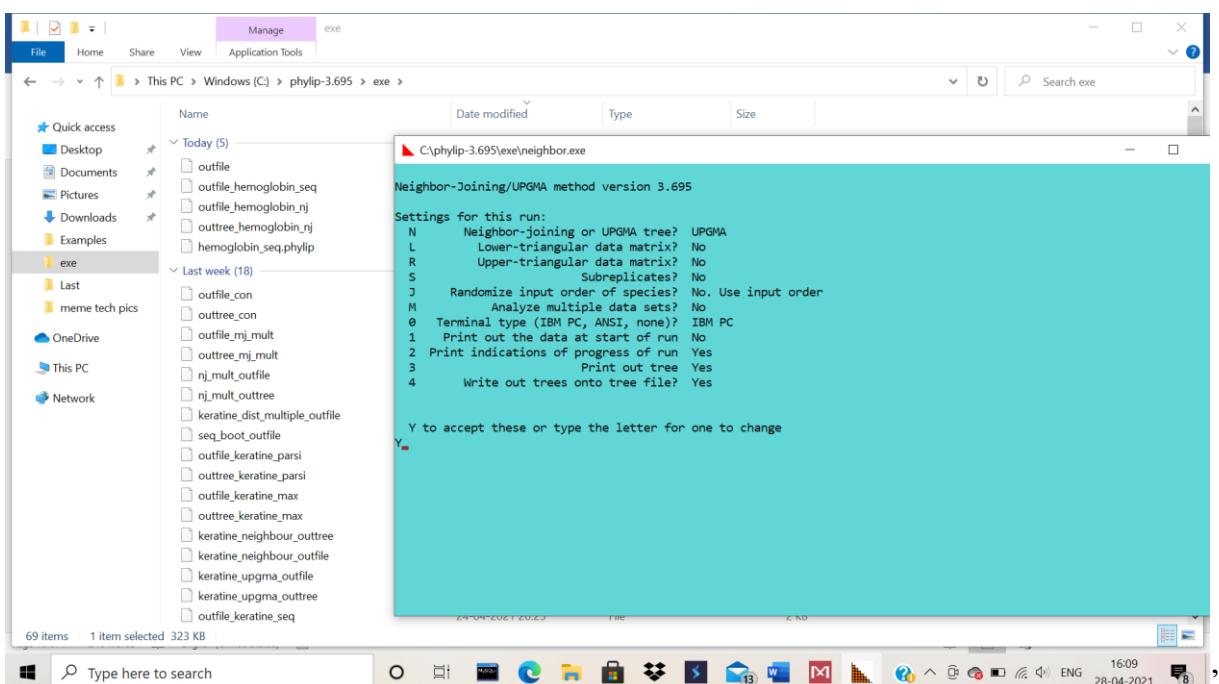
Radiation:



Performing distance based phylogenetic analysis of hemoglobin sequences of 10 different organisms using **UPGMA Method**:



Type 'Y' to accept all the default parameters shown:



The following is the out tree for UPGMA method in notepad ++ :

```

C:\phylip-3.695\exe\outfile_hemoglobin_upgma - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_mj_mult outfile_mj_mult outfile_con outfile_hemoglobin_seq outfile_hemoglobin_nj outfile_hemoglobin_nj outfile_hemoglobin_nj outfile_hemoglobin_upgma
1 (sp|P02112|:0.20717, (sp|P02094|:0.15079, ((sp|P02075|:0.09652,
2 sp|P02067|:0.09652):0.00718, (sp|P09905|:0.09203, ((sp|P02024|:0.00351,
3 sp|P68872|:0.00351):0.04874, (sp|P60524|:0.03253, sp|P18983|:0.03253):0.01972):0.03978):0.01168):0.00685,
4 sp|P04244|:0.11055):0.04024):0.05638);
5

```

Normal text file length : 280 lines : 5 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS
Type here to search 16:10 ENG 28-04-2021 8

Out file for hemoglobin sequences for 10 different organisms in notepad ++:

```

C:\phylip-3.695\exe\outfile_hemoglobin_upgma - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_mj_mult outfile_mj_mult outfile_con outfile_hemoglobin_seq outfile_hemoglobin_nj outfile_hemoglobin_nj outfile_hemoglobin_nj outfile_hemoglobin_upgma outfile_hemoglobin_upgma
1
2 10 Populations
3
4 Neighbor-Joining/UPGMA method version 3.695
5
6 UPGMA method
7
8 Negative branch lengths allowed
9
10
11 +-----sp|P02112|
12 !
13 !
14 ! +----sp|P02094|
15 --9 !
16 ! ! +---sp|P02075|
17 ! ! +-5
18 ! ! ! +---sp|P02067|
19 +-8 +-6
20 ! ! ! +---sp|P09905|
21 ! ! !
22 ! ! +-4 +sp|P02024|
23 ! ! ! +-1
24 +-7 ! ! +sp|P68872|
25 ! +-3
26 ! ! +sp|P60524|
27 ! +-2
28 ! +sp|P18983|
29 !
30 +----sp|P04244|
31

```

Normal text file length : 1,509 lines : 55 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS
Type here to search 16:10 ENG 28-04-2021 8

C:\phylib-3.695\exe\outfile_hemoglobin_upgma - Notepad+

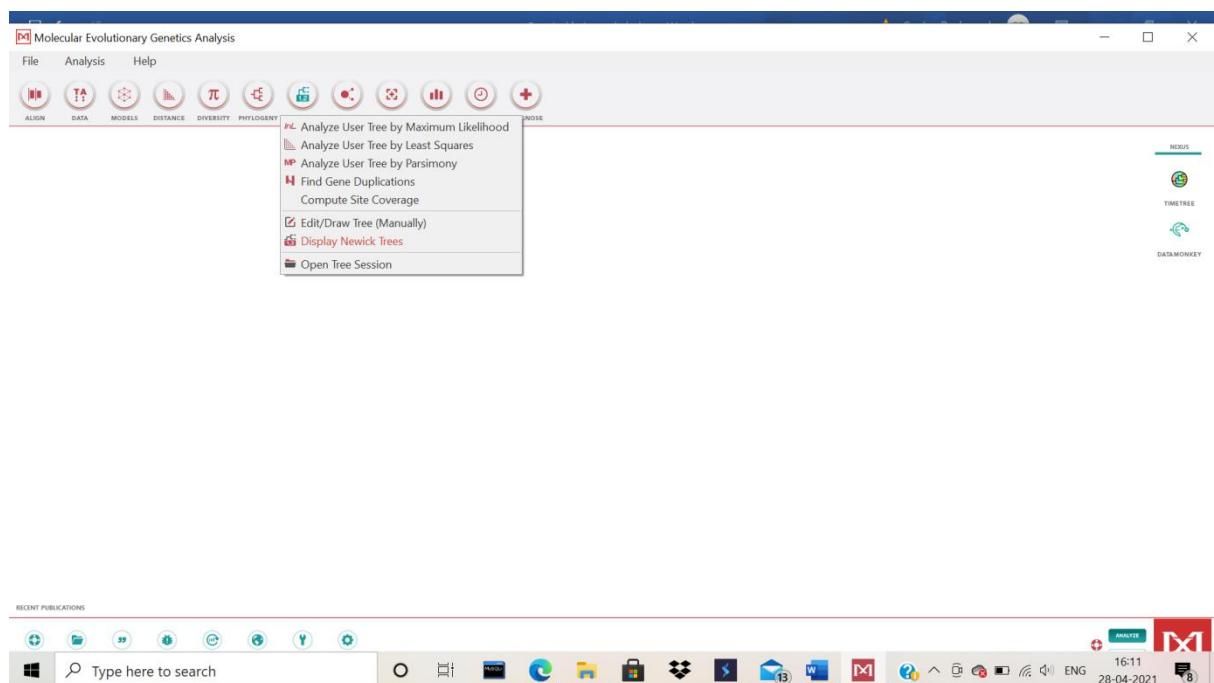
```

25      ! +--3
26      !   +-sp|P60524|
27      !     +-2
28      !       +-sp|P18983|
29      !
30     +----sp|P04244|
31
32
33 From    To      Length      Height
34 -----
35   9      sp|P02112|  0.20717    0.20717
36   9      8          0.05638    0.05638
37   8      sp|P02094|  0.15079    0.20717
38   8      7          0.04024    0.09662
39   7      6          0.00685    0.10346
40   6      5          0.00718    0.11064
41   5      sp|P02075|  0.09652    0.20717
42   5      sp|P02067|  0.09652    0.20717
43   6      4          0.01168    0.11514
44   4      sp|P09905|  0.09203    0.20717
45   4      3          0.03978    0.15492
46   3      1          0.04874    0.20366
47   1      sp|P02024|  0.00351    0.20717
48   1      sp|P68872|  0.00351    0.20717
49   3      2          0.01972    0.17464
50   2      sp|P60524|  0.03253    0.20717
51   2      sp|P18983|  0.03253    0.20717
52   7      sp|P04244|  0.11055    0.20717
53
54
55

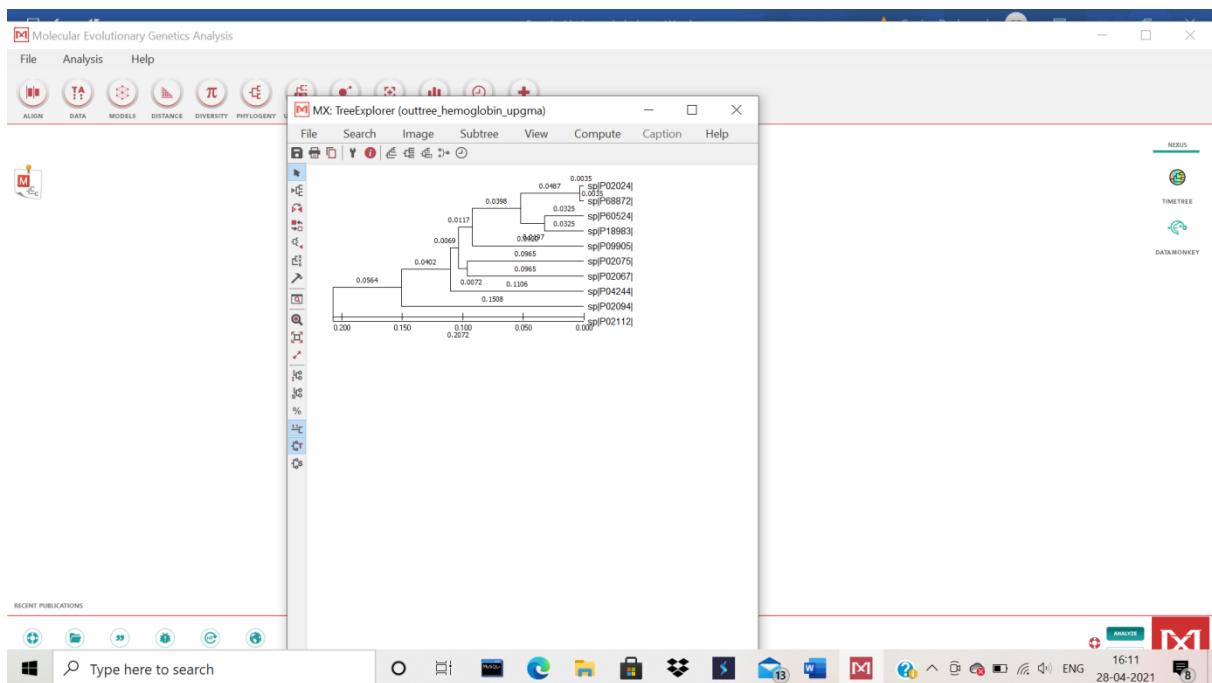
```

length : 1.509 lines : 55 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS

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



Phylogenetic newick tree for the 10 different organisms sequences is shown:



2. Perform character based phylogenetic analysis of hemoglobin sequences of 10 different organisms. Maximum likelihood and parsimony (protein sequences)

Answer:

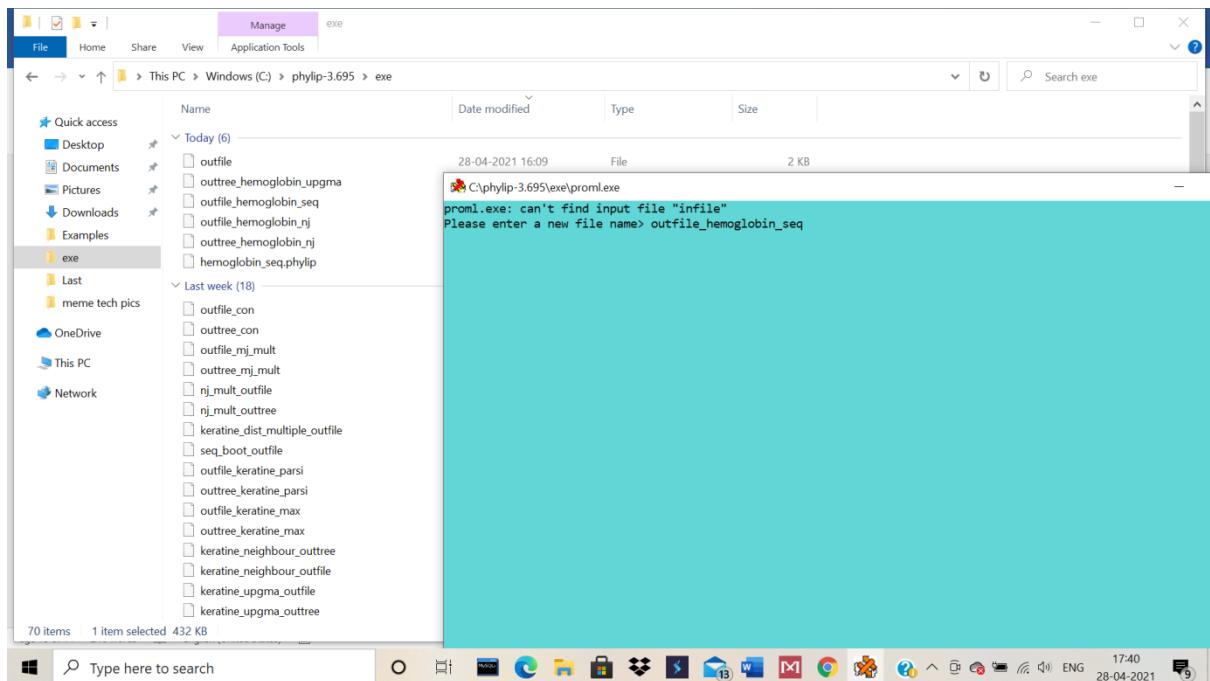
So Character based methods includes:-

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

Performing character based phylogenetic analysis of hemoglobin sequences of 10 different organisms

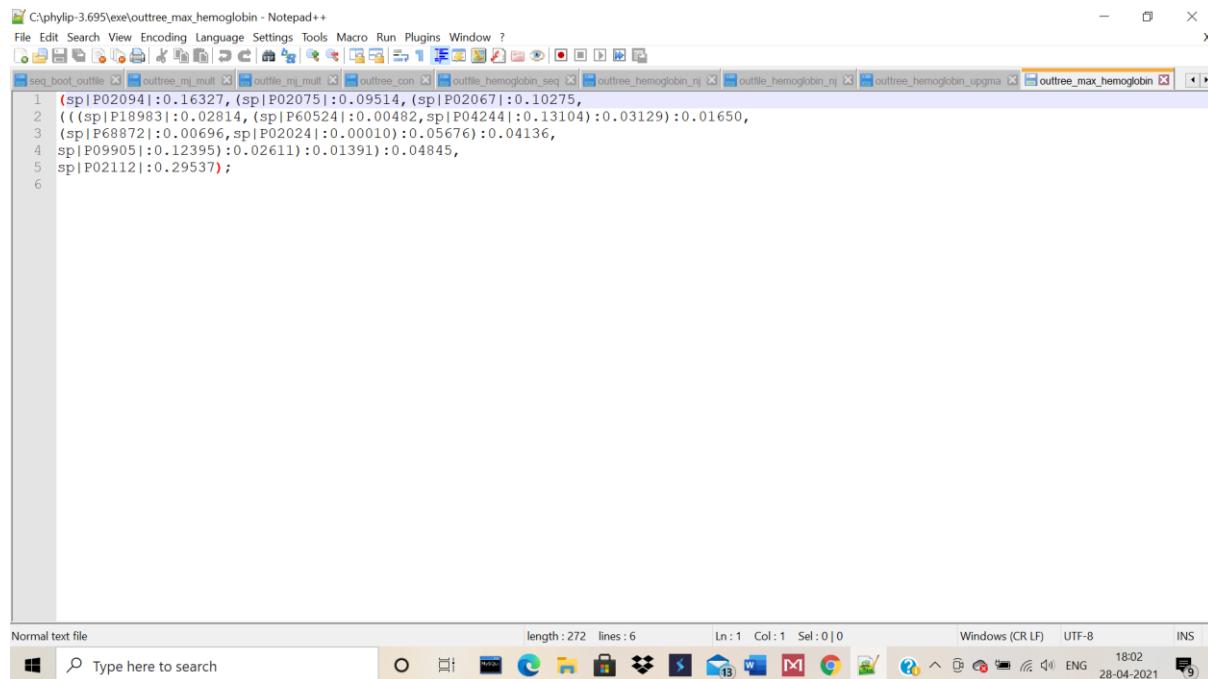
Maximum Likelihood:

Go to proml in the exe folder and enter the infile name ‘outfile_hemoglobin_seq’ and type Y to accept all the parameters and press enter:



```
>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
```

You can few the matrix .i.e. out tree for the hemoglobin sequences in ML method in notepad ++ :



```
C:\phylip-3.695\exe\outtree_max_hemoglobin - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
seq_boot_outfile.outree_mj_mull.outfile_mj_must.outtree.com.outfile_hemoglobin.seq.outtree_hemoglobin.nj.outfile_hemoglobin.nj.outtree_hemoglobin.upgma.outtree_max_hemoglobin
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
```

Out file for Maximum likelihood method in notepad ++:

```

C:\phylip-3.695\exe\outfile_max_hemoglobin - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outfile_hemoglobin_nj outfile_hemoglobin.upgma outfile_max_hemoglobin outfile_max_hemoglobin

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 | | | +sp|P18983|
14 | | | +-8
15 | +-4 | | +sp|P60524|
16 | | | +-7
17 | | +-3 +-----sp|P04244|
18 | | |
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
Type here to search 18:02 28-04-2021

ML method will always give you an unrooted tree:

```

C:\phylip-3.695\exe\outfile_max_hemoglobin - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outfile_hemoglobin_nj outfile_hemoglobin.upgma outfile_max_hemoglobin outfile_max_hemoglobin

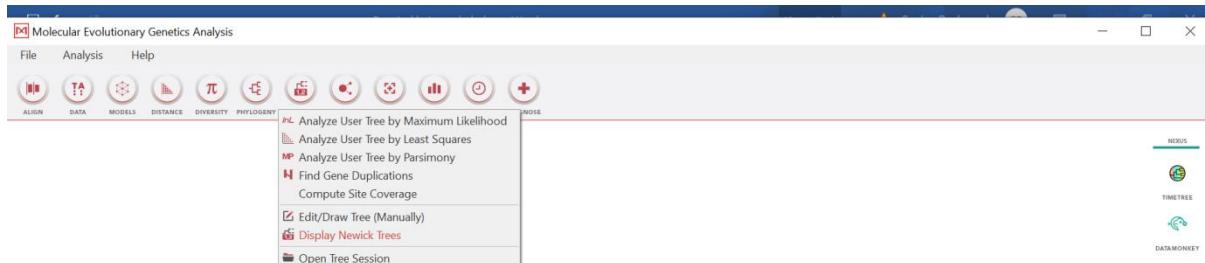
27
28 remember: this is an unrooted tree!
29
30 ln Likelihood = -1246.88212
31
32 Between And Length Approx. Confidence Limits
33 ----- --- -----
34
35 1 sp|P02112| 0.29537 ( 0.19208, 0.39864) **
36 1 sp|P02094| 0.16327 ( 0.08750, 0.23904) **
37 1 2 0.04845 ( zero, 0.09724) **
38 2 sp|P02075| 0.09514 ( 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

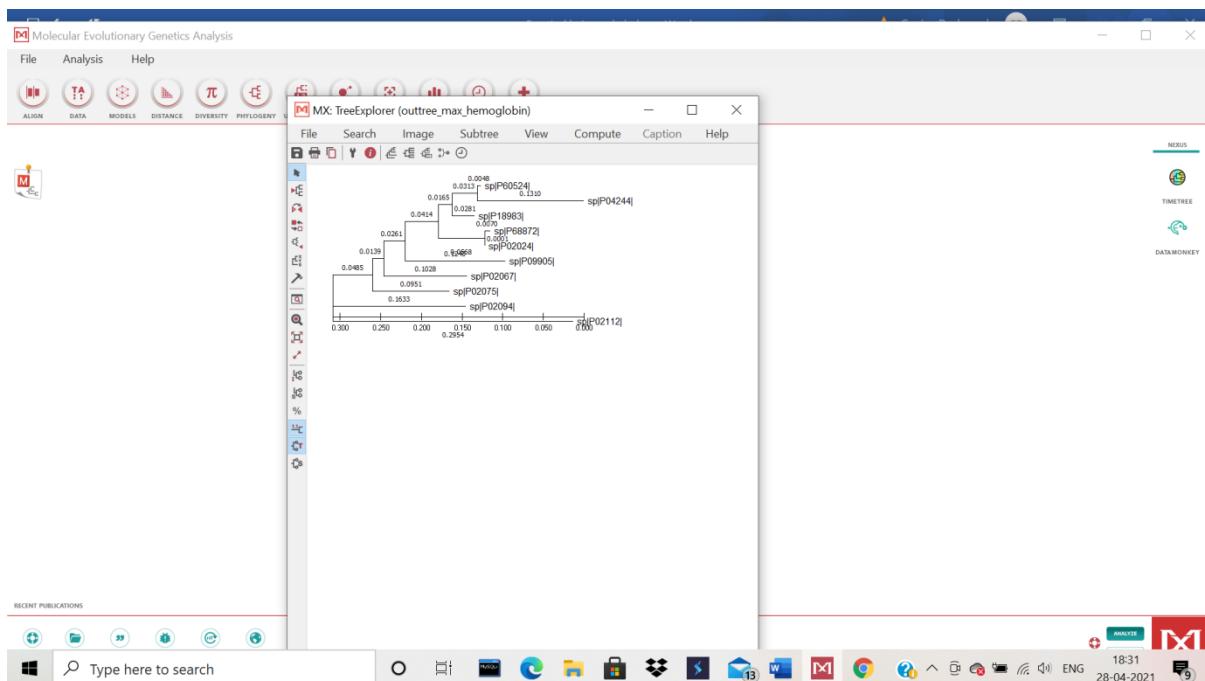
```

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 18:02 28-04-2021

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

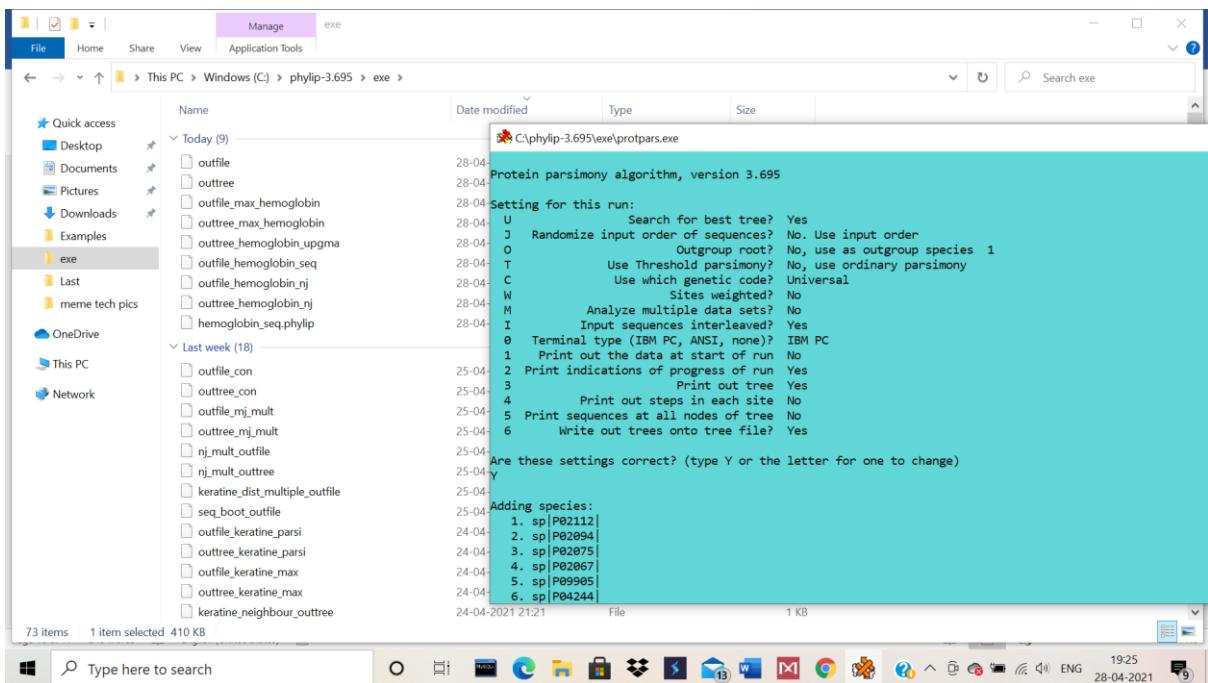


You can see phylogenetic newick tree for ML method of hemoglobin sequences of 10 different organisms:



Performing character based phylogenetic analysis of hemoglobin sequences of 10 different organisms
Parsimony:

Enter the file name and type Y to accept the changed parameters:



Out tree for parsimony method of hemoglobin sequences of 10 different organisms:

```

C:\phylip-3.695\exe\outtree_pars_hemoglobin - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_hemoglobin.upgns  outtree_max_hemoglobin  outfile_max_hemoglobin  outtree_pars_hemoglobin
1 (((sp|P042441, sp|P020671), (((sp|P189831, sp|P605241), (sp|P688721, sp|P020241)), (sp|P099051, sp|P020751))), sp|P020941, sp|P02112) [0.5000];
2 (((((sp|P189831, sp|P605241), (sp|P688721, sp|P020241)), (sp|P042441, sp|P020671)), (sp|P099051, sp|P020751)), sp|P020941), sp|P02112) [0.5000];
3
4
5

```

Normal text file length : 280 lines : 5 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS
Type here to search 19:28 28-04-2021

Unrooted tree for parsimony method in notepad ++ :