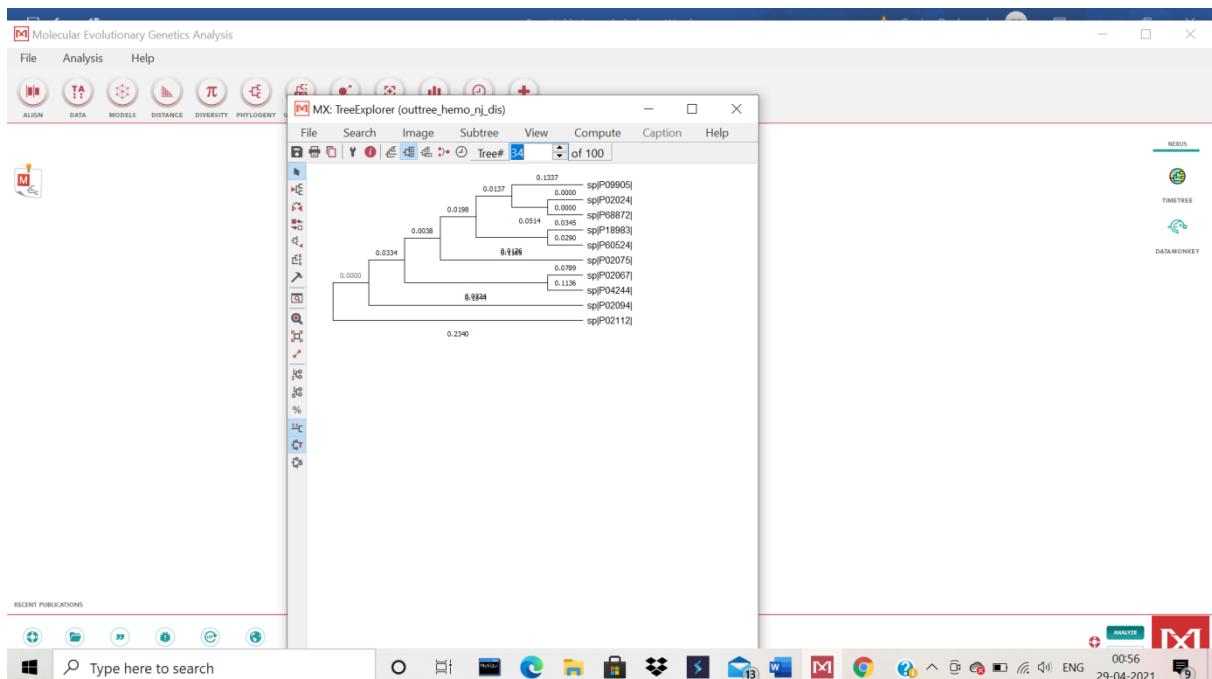
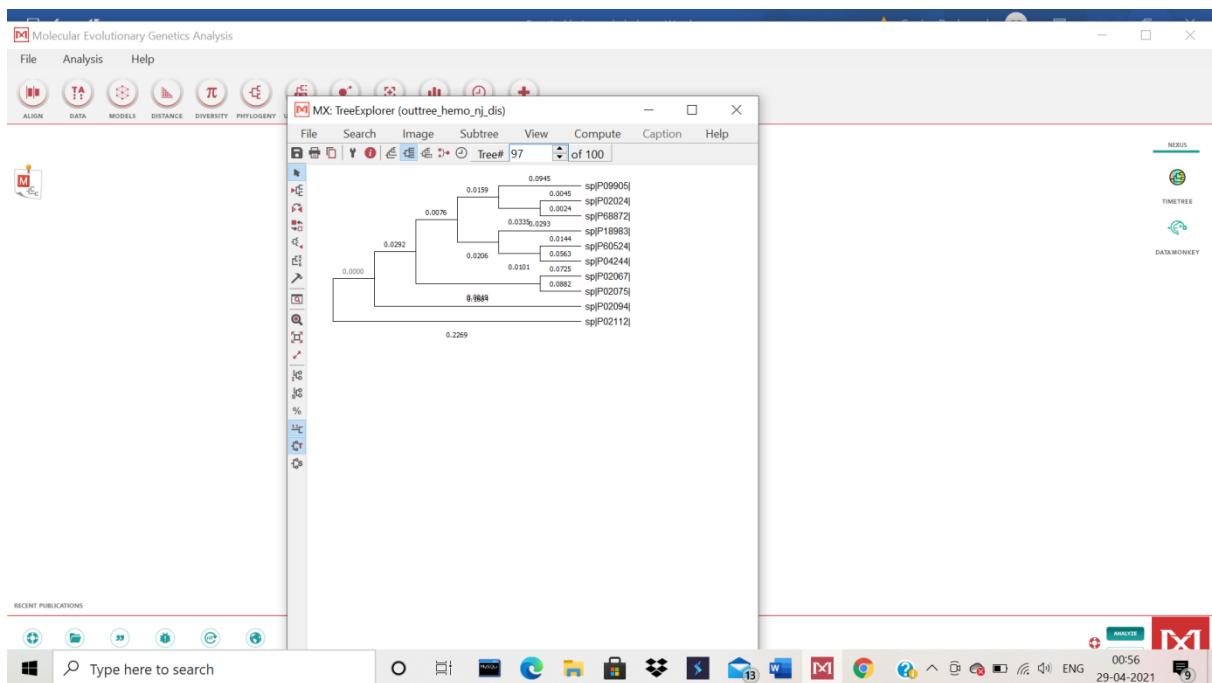


34 of 100



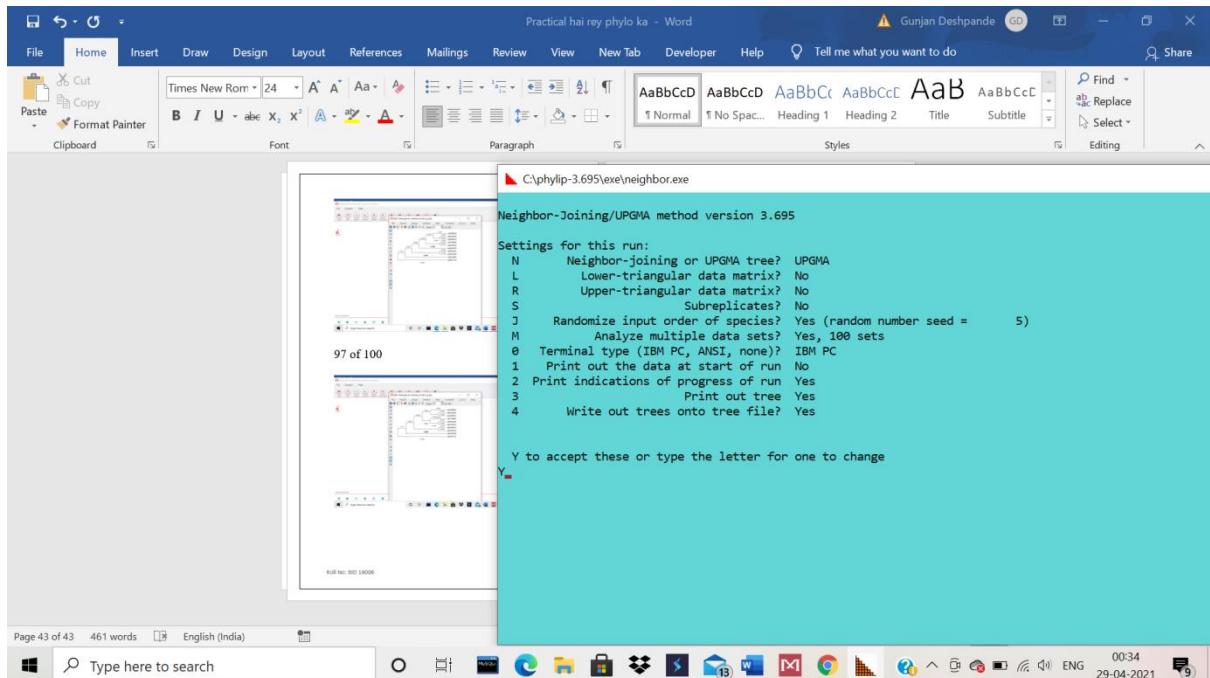
97 of 100



Performing boot strap analysis on distance based phylogenetic method of hemoglobin sequences by **UPGMA** 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
```

Type N to change the parameter to UPGMA method , type M to give number of multiple data sets as 100 , type Y to accept all the parameters and procced:



You will view our tree for all 100 sequences:

C:\phylip-3.695\exe\outtree_hema_upgma_dist - Notepad+

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

outfile_insulin_ni_dist outfile_mulf_d_insulin hemoglobin_seq.php outfile_seq_boot_hemo.outtree_hemo_ni_dis outfile_hemo_ni_dis outfile_hema_upgma_dist.outtree_hema_upgma_dist

```
1 (((((sp|P02067|:0.05569,sp|P02075|:0.05569):0.01874,(((sp|P02024|:0.00695,
2 sp|P68872|:0.00695):0.01433,sp|P18983|:0.02127):0.01455,
3 sp|P60524|:0.03582),((sp|P04244|:0.07106):0.00336):0.02330,
4 sp|P09905|:0.05773):0.02428,sp|P02094|:0.12201):0.05832,
5 sp|P02112|:0.18032);
6 (((((sp|P68872|:0.02135,sp|P02024|:0.02135):0.02930,(sp|P18983|:0.02863,
7 sp|P60524|:0.02863):0.02202):0.01543,(sp|P04244|:0.06192,
8 sp|P02067|:0.06192):0.00416):0.03094,(sp|P02075|:0.07321,
9 sp|P09905|:0.07321):0.02381):0.04179,sp|P02094|:0.13882):0.05522,
10 sp|P02112|:0.19404);
11 (((((sp|P18983|:0.03565,sp|P60524|:0.03565):0.01620,(sp|P68872|:0.00001,
12 sp|P02024|:0.00001):0.05184):0.03548,(sp|P02075|:0.07485,
13 sp|P02067|:0.07485):0.01247,sp|P04244|:0.05759):0.02056,
14 sp|P09905|:0.11815):0.02531,sp|P02094|:0.14346):0.03845,
15 sp|P02112|:0.18191);
16 (((((sp|P68872|:0.00001,sp|P02024|:0.00001):0.03292,(sp|P60524|:0.02570,
17 sp|P18983|:0.02570):0.00722):0.03123,sp|P02067|:0.06415):0.01554,
18 sp|P09905|:0.07969):0.01337,sp|P04244|:0.09306):0.00926,
19 sp|P02075|:0.10232):0.03529,sp|P02094|:0.13761):0.04642,
20 sp|P02112|:0.18403);
21 (((sp|P09905|:0.14793,(sp|P02075|:0.14305,(sp|P02067|:0.10780,
22 (((sp|P60524|:0.04991,sp|P18983|:0.04991):0.01203,(sp|P68872|:0.00351,
23 sp|P02024|:0.00351):0.05844):0.03426,sp|P04244|:0.09620):0.01159):0.03526):0.00488):0.01697,
24 sp|P02094|:0.16491):0.03878,sp|P02112|:0.20368);
25 (((((sp|P04244|:0.07853,((sp|P18983|:0.02815,sp|P60524|:0.02815):0.02322,
26 (sp|P02024|:0.00350,sp|P68872|:0.00350):0.04787):0.02716):0.00114,
27 (sp|P02067|:0.07342,sp|P02075|:0.07342):0.00625):0.03647,
28 sp|P02094|:0.11614):0.00383,sp|P09905|:0.11997):0.08391,
29 sp|P02112|:0.20389);
30 (sp|P02112|:0.18522,((((sp|P68872|:0.00001,sp|P02024|:0.00001):0.06174,
31 (sp|P18983|:0.03269,sp|P60524|:0.03269):0.02906):0.02278,
```

```

C:\phylib-3.695\exe\outtree_hema_upgma_dist - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outline_insulin_nj_dist outline_mult_dis_insulin hemoglobin_seq.phylip outline_seq_boot_hemo outtree_hemo_nj_dis outline_hemo_nj_dis outline_hema_upgma_dist outtree_hema_upgma_dist
426 sp|P02067|:0.08148):0.01882,(sp|P02075|:0.08641,sp|P09905|:0.08641):0.01389):0.06935,
427 sp|P02094|:0.16965):0.06651,sp|P02112|:0.23616);
428 (((((sp|P60524|:0.04861,((sp|P02024|:0.00362,sp|P68872|:0.00362):0.03208,
429 sp|P18983|:0.03569):0.01291):0.02998,sp|P09905|:0.07859):0.01912,
430 sp|P02075|:0.09771):0.00489,(sp|P04244|:0.08104,sp|P02067|:0.08104):0.02156):0.02209,
431 sp|P02094|:0.12468):0.06719,sp|P02112|:0.19187);
432 ((sp|P02094|:0.15889,((((sp|P18983|:0.02540,sp|P60524|:0.02540):0.02485,
433 (sp|P02024|:0.01052,sp|P68872|:0.01052):0.03973):0.04251,
434 sp|P02067|:0.09276):0.01873,sp|P04244|:0.11149):0.01303,
435 sp|P09905|:0.12452):0.00824,sp|P02075|:0.13275):0.02613):0.04640,
436 sp|P02112|:0.20529);
437 ((sp|P02094|:0.13755,((((sp|P02024|:0.00343,sp|P68872|:0.00343):0.03808,
438 (sp|P18983|:0.02108,sp|P60524|:0.02108):0.02043):0.01612,
439 sp|P04244|:0.05762):0.00811,sp|P02067|:0.06574):0.01494,
440 sp|P02075|:0.08068):0.00657,sp|P09905|:0.08725):0.05031):0.03214,
441 sp|P02112|:0.16969);
442 (((sp|P04244|:0.05718,sp|P02067|:0.05718):0.03814,(((sp|P68872|:0.00348,
443 sp|P02024|:0.00348):0.05284,(sp|P18983|:0.03278,sp|P60524|:0.03278):0.02354):0.02912,
444 sp|P09905|:0.08544):0.00988):0.00415,sp|P02075|:0.09947):0.03566,
445 sp|P02094|:0.13512):0.06900,sp|P02112|:0.20412);
446 (((((sp|P60524|:0.03208,sp|P18983|:0.03208):0.00957,(sp|P02024|:0.00001,
447 sp|P68872|:0.00001):0.04165):0.01613,sp|P02067|:0.05778):0.01742,
448 sp|P09905|:0.07520):0.02277,sp|P02094|:0.09797):0.00748,
449 sp|P04244|:0.10544):0.00525,sp|P02075|:0.11069):0.03242,
450 sp|P02112|:0.14311);
451 (((sp|P02067|:0.08693,sp|P04244|:0.08693):0.01378,(((sp|P18983|:0.02845,
452 (sp|P68872|:0.00001,sp|P02024|:0.00001):0.02845):0.01113,
453 sp|P60524|:0.03959):0.03649,sp|P09905|:0.07608):0.01230,
454 sp|P02075|:0.08838):0.01233):0.03308,sp|P02094|:0.13379):0.07220,
455 sp|P02112|:0.20599);
456

```

You will view negative branched trees for all 100 data sets in notepad ++ :

```

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

```

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

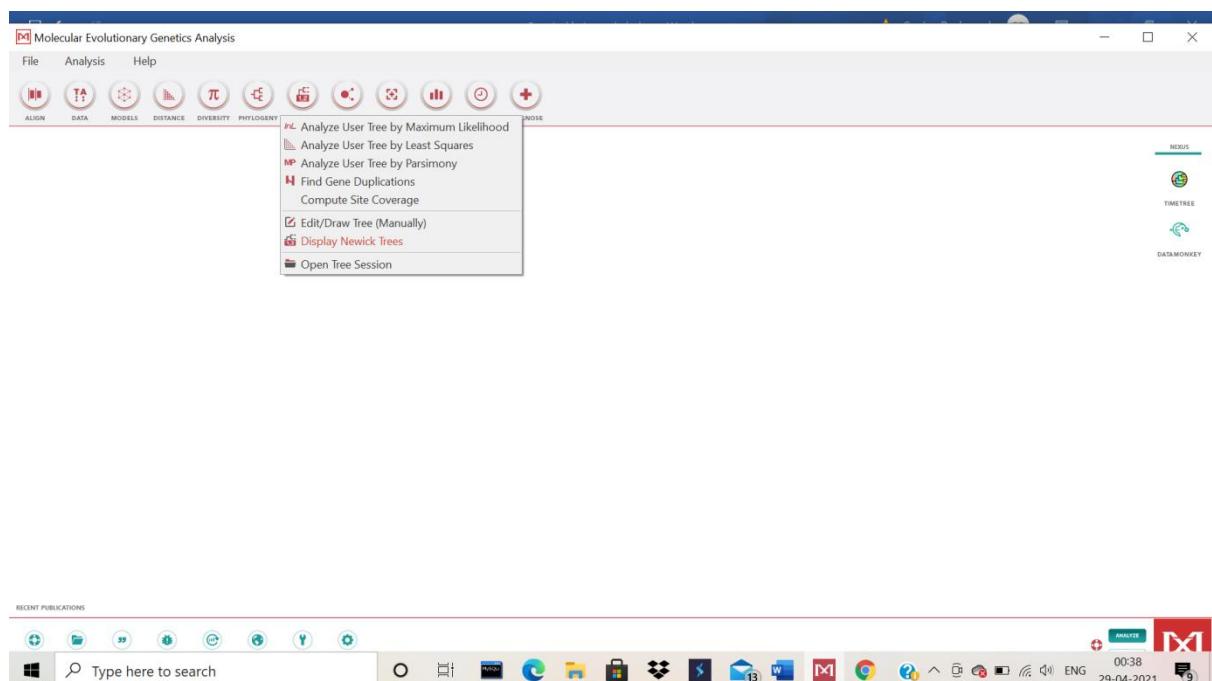
```

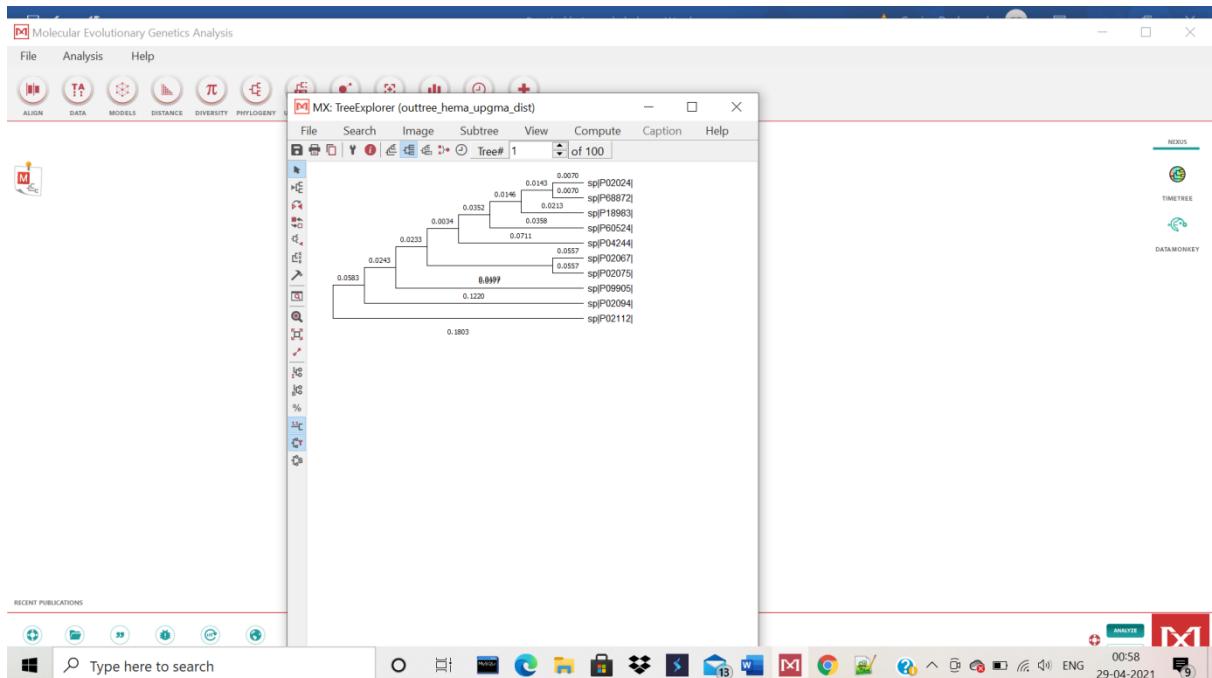
4976 ! !
4977 --9 ! +----sp|P02075|
4978 !
4979 ! +-----sp|P02094|
4980 !
4981 +-----sp|P02112|
4982
4983
4984 From To Length Height
4985 --- --
4986 9 8 0.07220 0.07220
4987 8 7 0.03308 0.10528
4988 7 5 0.01378 0.11906
4989 5 sp|P02067| 0.08693 0.20599
4990 5 sp|P04244| 0.08693 0.20599
4991 7 6 0.01233 0.11761
4992 6 4 0.01230 0.12991
4993 4 3 0.03649 0.16640
4994 3 2 0.01113 0.17753
4995 2 sp|P18983| 0.02845 0.20599
4996 2 1 0.02845 0.20598
4997 1 sp|P68872| 0.00001 0.20599
4998 1 sp|P02024| 0.00001 0.20599
4999 3 sp|P60524| 0.03959 0.20599
5000 4 sp|P09905| 0.07608 0.20599
5001 6 sp|P02075| 0.08838 0.20599
5002 8 sp|P02094| 0.13379 0.20599
5003 9 sp|P02112| 0.20599 0.20599
5004
5005
5006

```

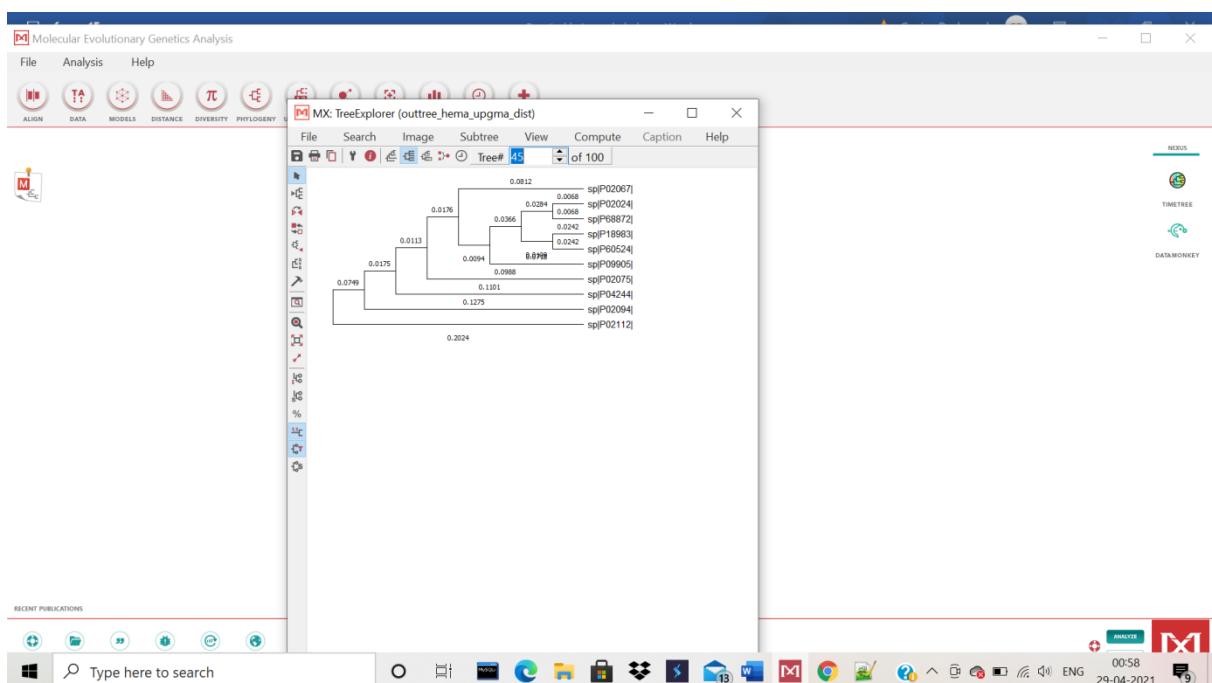
length : 1.46,051 lines : 5,006 Ln : 5,006 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS
Normal text file Type here to search 00:57 29-04-2021

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

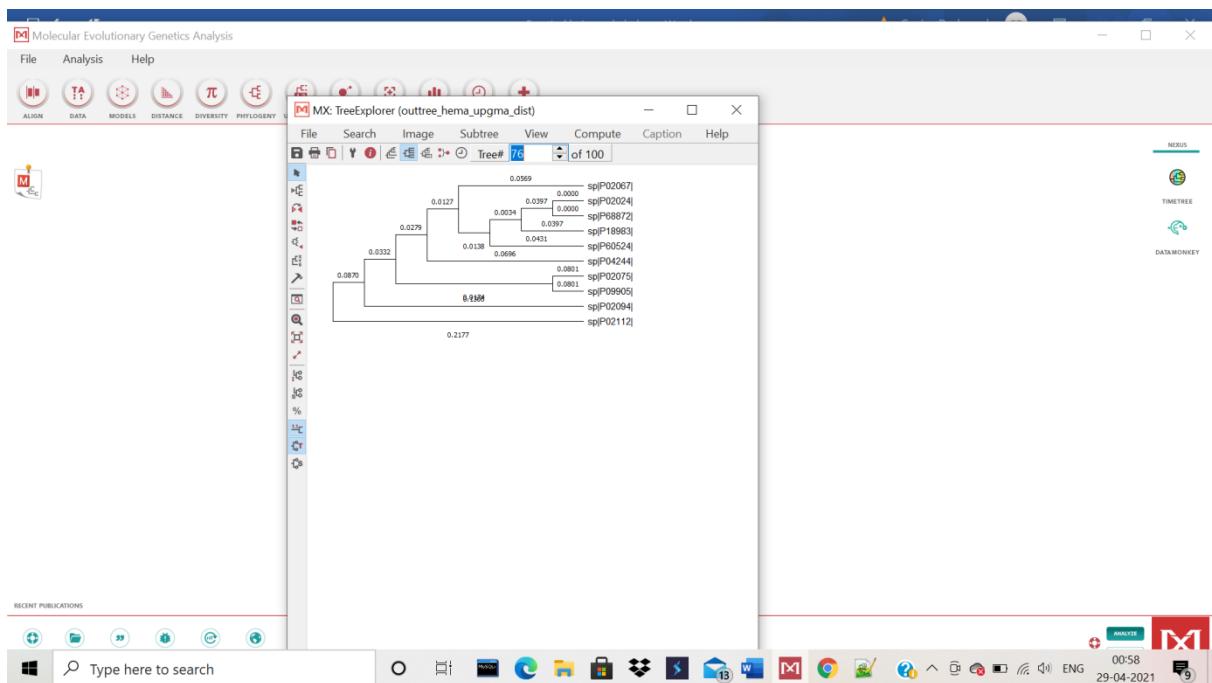




45 of 100



76 of 100



6. Perform boot strap analysis on character based phylogenetic method of hemoglobin sequences
Answer:

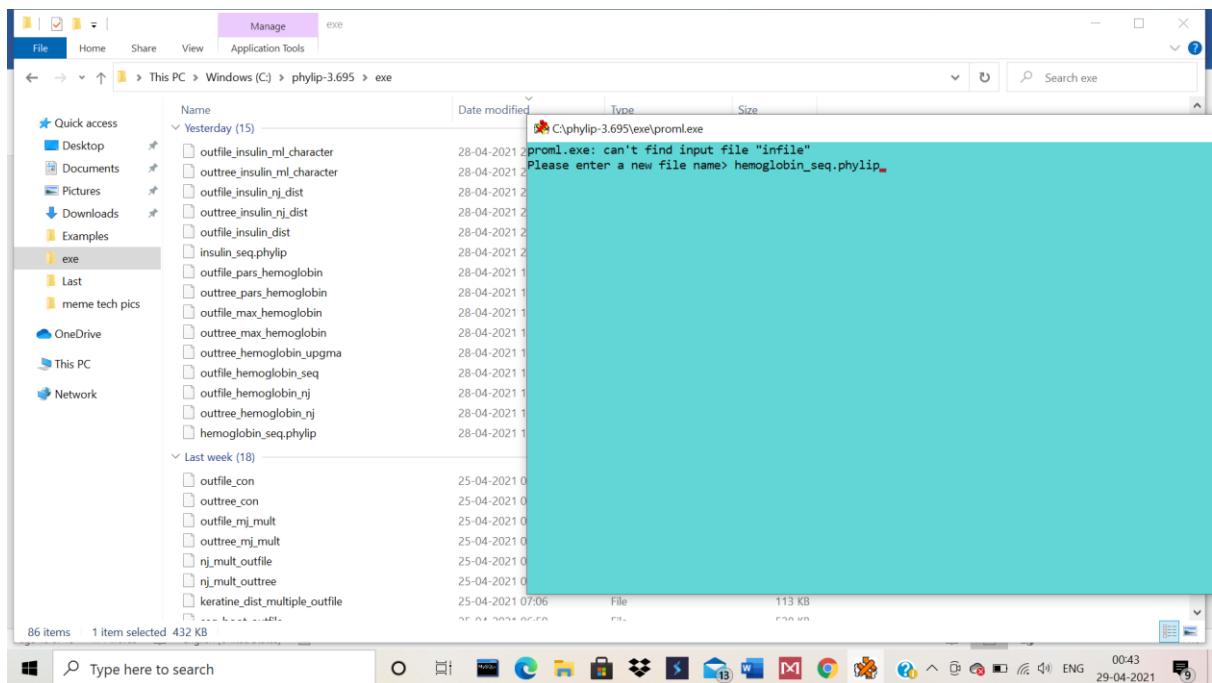
So character based methods includes:-

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

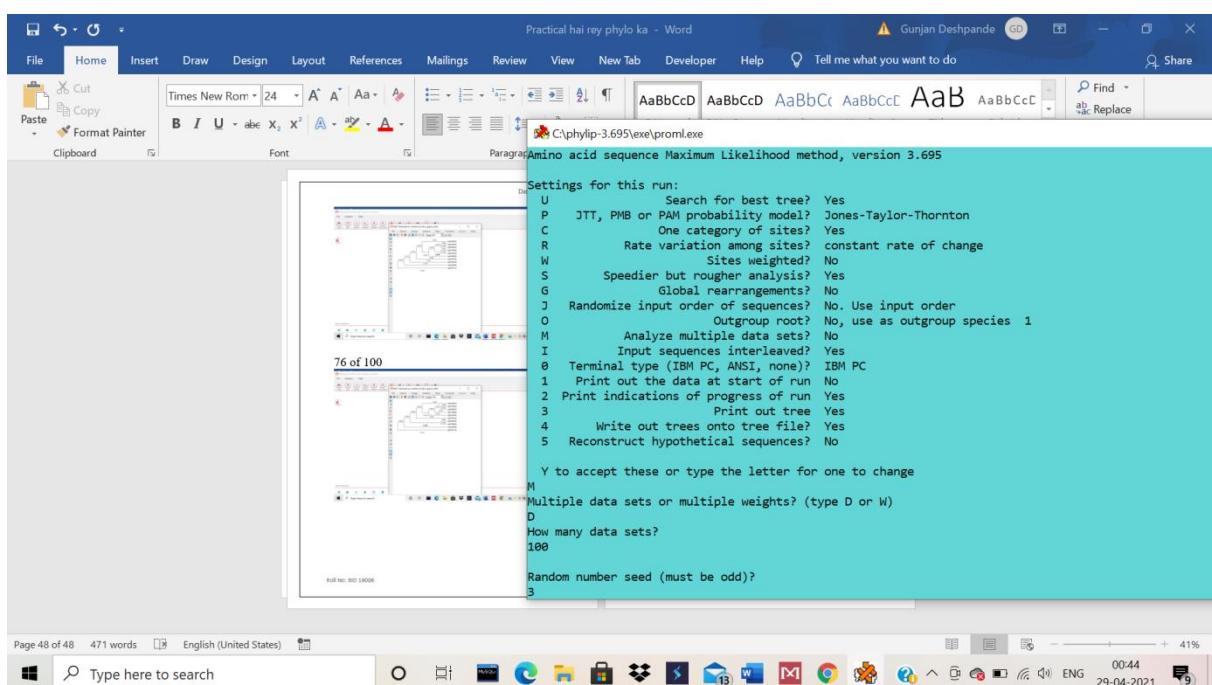
Performing boot strap analysis on character based phylogenetic method of hemoglobin sequences by **Maximum Likelihood method:**

```
>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
```

Go to proml file in exe folder and enter your file name as hemoglobin_seq.phylip:



Type M to analyze multiple data sets, type D to agree to analyze multiple data sets and not data weights , give input of 100 data sets and 3 as random number of seeds; type Y to agree to all changed parameters and procced:



You will view out tree for all 100 data sets:

C:\phylophil-3.695\exe\outtree_hema_char_ml - Notepad+

```

1  ((sp|P02094|:0.11063,(sp|P09905|:0.11358,((sp|P04244|:0.08936,
2  sp|P02067|:0.04982):0.02734,sp|P02075|:0.04129):0.01803):0.01970,
3  ((sp|P60524|:0.04249,sp|P18983|:0.01392):0.01378,
4  (sp|P02024|:0.00010,sp|P68872|:0.01388):0.01403):0.03023):0.07077,
5  sp|P02112|:0.25844);
6  ((sp|P02094|:0.21697,(sp|P09905|:0.10699,sp|P02075|:0.04425):0.00013):0.06375,
7  (sp|P02067|:0.02950,((sp|P02024|:0.04270,sp|P68872|:0.00010):0.04438,
8  ((sp|P18983|:0.04838,sp|P60524|:0.00890):0.04378,
9  sp|P04244|:0.05179):0.02638):0.02078):0.02372,
10 sp|P02112|:0.26023);
11 ((sp|P02094|:0.16104,(((sp|P09905|:0.15794,(sp|P02024|:0.00010,
12 sp|P68872|:0.00010):0.03000):0.02509,((sp|P60524|:0.00437,
13 sp|P04244|:0.10858):0.04208,sp|P18983|:0.02466):0.02399):0.03586,
14 (sp|P02075|:0.07973,sp|P02067|:0.07806):0.03415):0.03570,
15 sp|P02112|:0.25690);
16 ((sp|P04244|:0.11809,sp|P60524|:0.00010):0.01540,
17 (((sp|P68872|:0.00010,sp|P02024|:0.00010):0.02684,
18 ((sp|P09905|:0.09488,sp|P02075|:0.11505):0.02403,
19 (sp|P18983|:0.02015):0.01397,sp|P02112|:0.28789);
20 ((sp|P02094|:0.18445,sp|P02067|:0.07889):0.04129,
21 (((sp|P02024|:0.00010,sp|P68872|:0.00696):0.06143,
22 ((sp|P09905|:0.18890,sp|P18983|:0.00525):0.07023,
23 (sp|P60524|:0.00010,sp|P04244|:0.10924):0.02634):0.01801):0.04671,
24 (sp|P02075|:0.18882):0.00860,sp|P02112|:0.28235);
25 ((sp|P02094|:0.16687,sp|P02075|:0.05476):0.00010,
26 (((sp|P18983|:0.02829,(sp|P04244|:0.09643,sp|P60524|:0.00619):0.02212):0.00594,
27 (sp|P02024|:0.00010,sp|P68872|:0.00697):0.06604):0.04710,
28 sp|P02067|:0.09526):0.01893,sp|P09905|:0.14902):0.02001,
30 sp|P02112|:0.31071);
31 (sp|P02094|:0.10259,(((sp|P09905|:0.12630,(sp|P02024|:0.00010,

```

Normal text file length : 27,164 lines : 483 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS 01:21 29-04-2021

C:\phylophil-3.695\exe\outtree_hema_char_ml - Notepad+

```

453 sp|P09905|:0.08960):0.02640):0.08541,sp|P02112|:0.32950);
454 ((sp|P04244|:0.11774,sp|P60524|:0.01553):0.03226,
455 (sp|P18983|:0.03297,((sp|P02067|:0.09533,(sp|P09905|:0.11378,
456 (sp|P02094|:0.16650,sp|P02075|:0.10147):0.00095):0.02252):0.02949,
457 (sp|P02024|:0.00719,sp|P68872|:0.00010):0.03713):0.01164):0.01236,
458 sp|P02112|:0.28957);
459 ((sp|P04244|:0.10872,((sp|P02024|:0.00010,(sp|P09905|:0.18056,
460 sp|P68872|:0.00010):0.02103):0.06559,sp|P18983|:0.02585):0.02390,
461 sp|P60524|:0.00010):0.05605):0.05023,((sp|P02067|:0.08258,
462 sp|P02075|:0.15439):0.05044,sp|P02094|:0.18630):0.02340,
463 sp|P02112|:0.27936);
464 ((sp|P02094|:0.19916,sp|P02075|:0.07030):0.02017,
465 (((sp|P09905|:0.11753,(sp|P02024|:0.00684,sp|P68872|:0.00010):0.01453):0.03045,
466 (sp|P18983|:0.03257,(sp|P04244|:0.06953,sp|P60524|:0.00010):0.00909):0.02846):0.02147,
467 sp|P02067|:0.06766):0.02147,sp|P02112|:0.25656);
468 (sp|P02094|:0.09244,((sp|P02075|:0.10214,(sp|P04244|:0.08246,
469 sp|P02067|:0.03894):0.05405):0.00460,((sp|P18983|:0.01993,
470 sp|P60524|:0.04486):0.02117,(sp|P68872|:0.00010,
471 sp|P02024|:0.00692):0.06623):0.00710,sp|P09905|:0.12680):0.01898):0.10297,
472 sp|P02112|:0.26811);
473 ((sp|P60524|:0.00010,sp|P04244|:0.11581):0.03114,
474 (((sp|P09905|:0.09439,(sp|P68872|:0.00010,sp|P02024|:0.00010):0.04234):0.02993,
475 sp|P18983|:0.00010):0.01399,((sp|P02094|:0.11999,
476 sp|P02075|:0.11787):0.00698,sp|P02067|:0.05545):0.03701):0.01769,
477 sp|P02112|:0.21329);
478 ((sp|P04244|:0.10505,sp|P60524|:0.00010):0.03826,
479 (sp|P18983|:0.01054,((sp|P02094|:0.17984,((sp|P02067|:0.11916,
480 sp|P02075|:0.06564):0.02227,sp|P09905|:0.09237):0.03665):0.02020,
481 (sp|P02024|:0.00010,sp|P68872|:0.00010):0.03013):0.01670):0.00794,
482 sp|P02112|:0.32914);
483

```

Normal text file length : 27,164 lines : 483 Ln : 483 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS 01:21 29-04-2021

You will view unrooted tree:

```

C:\phylip-3.695\exe\outfile_hema_char_ml - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outfile_hemo_nj_dis outfile_hema_upgma_dist outfile_hema_upgma_dist outfile_hema_char_ml outfile_hema_char_ml

1 Amino acid sequence Maximum Likelihood method, version 3.695
2
3 Data set # 1:
4 Jones-Taylor-Thornton model of amino acid change
5
6
7 +-----sp|P02094|
8 |
9 | +----sp|P09905|
10 | |
11 | +-7 +----sp|P04244|
12 | | | +-8
13 | | | +-2 +-sp|P02067|
14 | | |
15 | | |
16 | 6---5 +-sp|P02075|
17 | |
18 | | | +sp|P60524|
19 | | | +-3
20 | | | +sp|P18983|
21 | | | +-4
22 | | | +sp|P02024|
23 | | | +-1
24 | | | +sp|P68872|
25 |
26 +-----sp|P02112|
27
28 remember: this is an unrooted tree!
29
30 Ln Likelihood = -1097.73246
31

```

Normal text file length : 2,05,409 lines : 5,404 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS
01:22 29-04-2021

You will view unrooted tree for all 100 data sets:

```

C:\phylip-3.695\exe\outfile_hema_char_ml - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outfile_hemo_nj_dis outfile_hema_upgma_dist outfile_hema_upgma_dist outfile_hema_char_ml outfile_hema_char_ml

5374
5375 remember: this is an unrooted tree!
5376
5377 Ln Likelihood = -1166.27673
5378
5379 Between And Length Approx. Confidence Limits
5380 ----- --- -----
5381
5382 1 sp|P02112| 0.32914 ( 0.22264, 0.43563) **
5383 1 6 0.03826 ( 0.00457, 0.07196) **
5384 6 sp|P04244| 0.10505 ( 0.04962, 0.16048) ***
5385 6 sp|P60524| 0.00010 ( zero, infinity)
5386 1 7 0.00794 ( zero, 0.02442)
5387 7 sp|P18983| 0.01054 ( zero, 0.02907) ***
5388 7 2 0.01670 ( zero, 0.03975) ***
5389 2 3 0.02020 ( zero, 0.04797) ***
5390 3 sp|P02094| 0.17984 ( 0.10503, 0.25466) ***
5391 3 5 0.03665 ( 0.00148, 0.07181) ***
5392 5 8 0.02227 ( zero, 0.05398) *
5393 8 sp|P02067| 0.11916 ( 0.05637, 0.18193) ***
5394 8 sp|P02075| 0.06564 ( 0.01826, 0.11303) ***
5395 5 sp|P09905| 0.09237 ( 0.03732, 0.14743) ***
5396 2 4 0.03013 ( zero, 0.06065) **
5397 4 sp|P02024| 0.00010 ( zero, infinity)
5398 4 sp|P68872| 0.00010 ( zero, infinity)

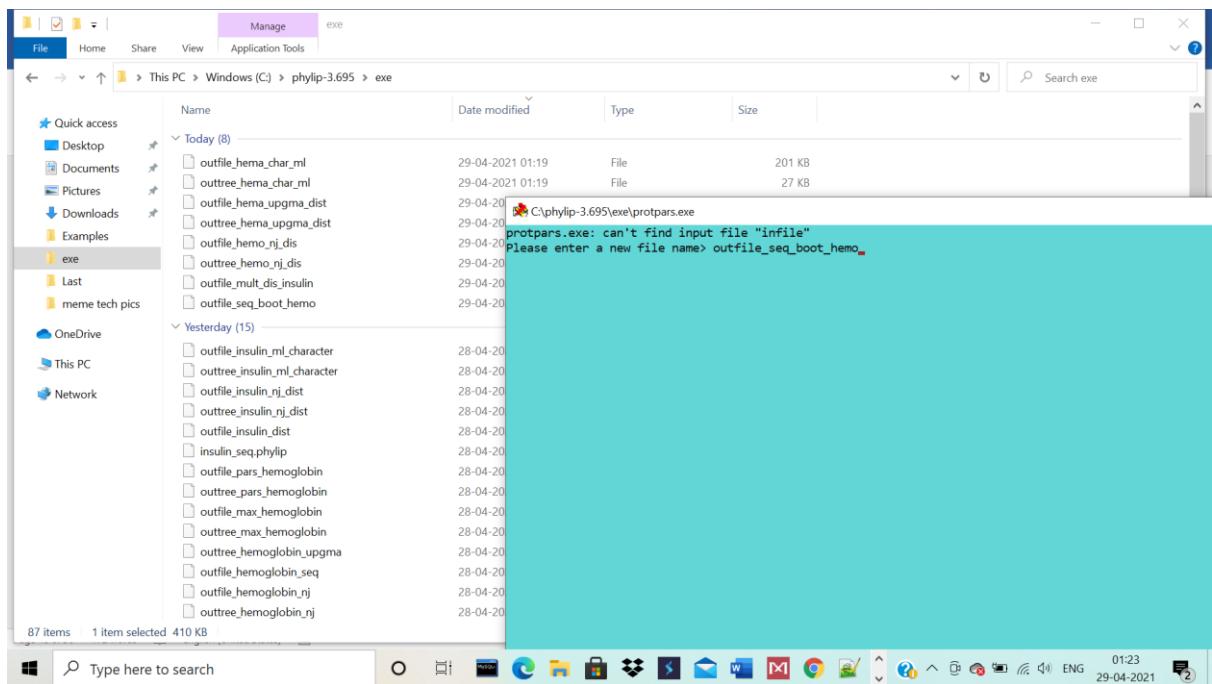
5399 * = significantly positive, P < 0.05
5400 ** = significantly positive, P < 0.01
5401
5402
5403
5404

```

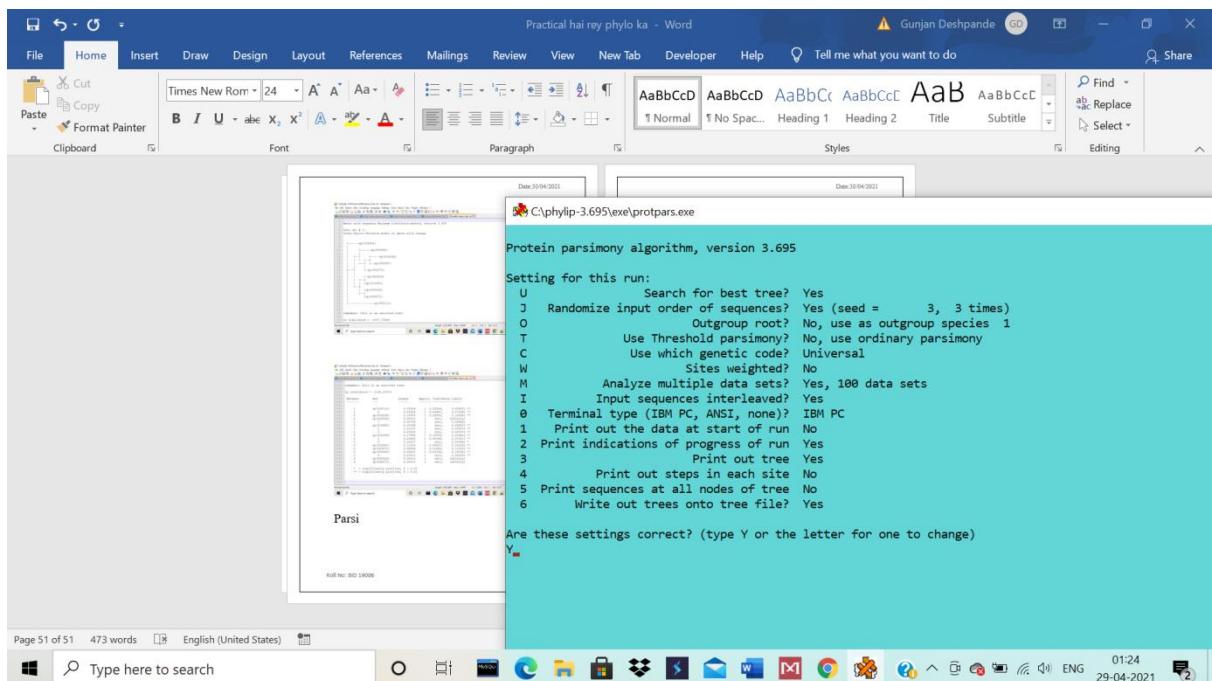
Normal text file length : 2,05,409 lines : 5,404 Ln : 5,404 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS
01:23 29-04-2021

Performing boot strap analysis on character based phylogenetic method of hemoglobin sequences by Parsimony method:

Go to protpars file in exe folder and give input file name as outfile_seq_boot_hemo:



Type Y to agree to all the default parameters and proceed:



You will view out tree for all 100 data sets:

C:\phylip-3.695\exe\outtree_hemo_pari_char - Notepad++

```

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_seq_boot_hemo outfile_hemo_ni outfile_hemo_ni_dis outfile_hema_upgma_dist outfile_hema_upgma_dist.outtree_hema_char_ml outfile_hema_char_ml outfile_hema_pari_char

1 (((sp|P042441,sp|P020671),((sp|P688721,((sp|P189831,sp|P605241),
2 sp|P020241)),(sp|P099051,sp|P020751))),sp|P020941),sp|P021121)[0.5000];
3 (((sp|P042441,sp|P020671),((sp|P189831,sp|P605241),(sp|P688721,
4 sp|P020241)),(sp|P099051,sp|P020751)),sp|P020941),sp|P021121)[0.5000];
5 ((sp|P020671,((sp|P688721,sp|P020241),(sp|P042441,(sp|P605241,
6 (sp|P189831,((sp|P099051,sp|P020751))),sp|P020941))),sp|P021121);
7 (((((sp|P688721,sp|P020241),(sp|P189831,(sp|P605241,sp|P042441)),
8 sp|P099051,sp|P020671),sp|P020751),sp|P020941),sp|P021121);
9 (((sp|P189831,((sp|P042441,sp|P020671),((sp|P688721,sp|P020241),
10 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.2000];
11 (((sp|P189831,sp|P042441,sp|P020671),((sp|P688721,sp|P020241),
12 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.2000];
13 (((sp|P042441,sp|P020671),(sp|P189831,((sp|P688721,sp|P020241),
14 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.2000];
15 (((sp|P042441,(sp|P189831,sp|P020671,((sp|P688721,sp|P020241),
16 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.2000];
17 (((sp|P020671,((sp|P189831,sp|P042441),((sp|P688721,sp|P020241),
18 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.2000];
19 ((sp|P189831,((sp|P042441,sp|P020671),((sp|P688721,sp|P020241),
20 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.3333];
21 (((sp|P605241,sp|P042441),(sp|P189831,((sp|P688721,sp|P020241),
22 (sp|P099051,sp|P020671),(sp|P020941)))),sp|P021121)[0.3333];
23 (((sp|P189831,(sp|P605241,sp|P042441)),((sp|P688721,sp|P020241),
24 (sp|P099051,(sp|P020751),(sp|P020671,sp|P020941)))),sp|P021121)[0.3333];
25 (((sp|P020671,((sp|P688721,sp|P020241),(sp|P189831,(sp|P605241,
26 sp|P042441)),(sp|P099051,sp|P020751))),sp|P020941),sp|P021121);
27 (((sp|P042441,(sp|P605241,(sp|P189831,((sp|P688721,sp|P020241),
28 (sp|P020671,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.3333];
29 (((sp|P042441,(sp|P605241,(sp|P189831,((sp|P688721,sp|P020241),
30 sp|P099051,(sp|P020671,sp|P020751)))),sp|P020941),sp|P021121)[0.3333];
31 (((sp|P042441,(sp|P605241,(sp|P189831,((sp|P688721,sp|P020241),

```

C:\phylip-3.695\exe\outtree_hemo_pari_char - Notepad++

```

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_seq_boot_hemo outfile_hemo_ni outfile_hemo_ni_dis outfile_hema_upgma_dist outfile_hema_upgma_dist.outtree_hema_char_ml outfile_hema_char_ml outfile_hema_pari_char

481 (((sp|P605241,(sp|P189831,(sp|P688721,sp|P020241))),sp|P042441),
482 (sp|P020671,(sp|P099051,sp|P020751))),sp|P020941),sp|P021121);
483 (((sp|P020671,((sp|P688721,sp|P020241),((sp|P189831,sp|P605241),
484 sp|P042441)),(sp|P099051,sp|P020751))),sp|P020941),sp|P021121)[0.2000];
485 (sp|P020671,((sp|P189831,sp|P042441),((sp|P688721,sp|P020241),
486 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.2000];
487 (((sp|P020671,((sp|P688721,sp|P020241),((sp|P189831,sp|P042441),
488 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.2000];
489 (((sp|P020671,((sp|P688721,sp|P020241),(sp|P189831,sp|P042441,
490 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.2000];
491 (((sp|P020671,(sp|P688721,sp|P020241),(sp|P042441,(sp|P189831,
492 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.2000];
493 (((sp|P605241,(sp|P189831,((sp|P688721,sp|P020241),(sp|P042441,
494 sp|P020671,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.1429];
495 (((sp|P605241,(sp|P042441,sp|P020671),((sp|P688721,sp|P020241),
496 (sp|P189831,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.1429];
497 (((sp|P605241,(sp|P042441,sp|P020671),(sp|P189831,((sp|P688721,
498 sp|P020241),(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.1429];
499 (((sp|P042441,sp|P020671),(sp|P605241,(sp|P189831,((sp|P688721,
500 sp|P020241),(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121)[0.1429];
501 (((sp|P099051,((sp|P189831,sp|P605241),(sp|P688721,sp|P020241)),
502 (sp|P042441,sp|P020671),(sp|P020751),sp|P020941),sp|P021121)[0.1429];
503 (((sp|P605241,(sp|P189831,((sp|P688721,sp|P020241),sp|P099051,
504 (sp|P042441,sp|P020671),(sp|P020751)))),sp|P020941),sp|P021121)[0.1429];
505 (((sp|P099051,(((sp|P189831,sp|P605241),(sp|P688721,sp|P020241),
506 (sp|P042441,sp|P020671),(sp|P020751)),sp|P020941),sp|P021121)[0.1429];
507 (((sp|P042441,sp|P020671),(sp|P189831,((sp|P688721,sp|P020241),
508 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121);
509 (((sp|P042441,sp|P020671),(sp|P189831,((sp|P688721,sp|P020241),
510 (sp|P605241,(sp|P099051,sp|P020751)))),sp|P020941),sp|P021121);
511 |

```

You will get the most parsimonious tree out of all in data sets:

```

C:\phylip-3.695\exe\outfile_hemo_pari_char - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_hemo_ni_dis outfile_hemo_ni_dis outfile_hema_upgma_dist outtree_hema_upgma_dist outfile_hema_char_mi outfile_hema_char_mi outtree_hemo_pari_char outfile_hemo_pari_char

1 Protein parsimony algorithm, version 3.695
2
3
4 Data set # 1:
5
6
7 2 trees in all found
8
9
10
11
12
13      +--sp|P04244|
14      +-----5
15      !      +--sp|P02067|
16      !
17      !      +-----sp|P68872|
18      +-3      !
19      ! !      +----7      +--sp|P18983|
20      ! !      !      +--9
21      ! !      !      +--8      +--sp|P60524|
22      ! +----6      !
23      +-2      !      +----sp|P02024|
24      ! !
25      ! !      +-----sp|P09905|
26      ! !      +-----4
27      1 !      +--sp|P02075|
28      !
29      ! +-----sp|P02094|
30      !
31      +-----sp|P02112|


Normal text file length : 1,80,893 lines : 7,744 Ln : 1 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS
Type here to search 01:26 ENG 29-04-2021 2

```

```

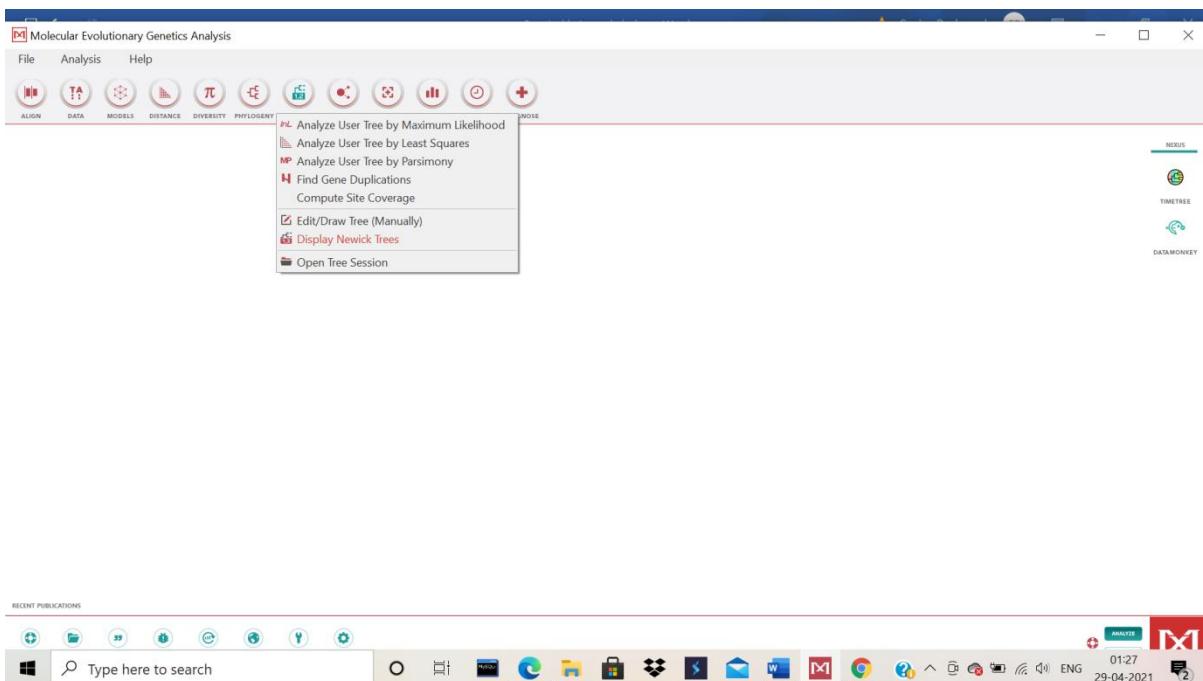
C:\phylip-3.695\exe\outfile_hemo_pari_char - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outtree_hemo_ni_dis outfile_hemo_ni_dis outfile_hema_upgma_dist outtree_hema_upgma_dist outfile_hema_char_mi outfile_hema_char_mi outtree_hemo_pari_char outfile_hemo_pari_char

7714 One most parsimonious tree found:
7715
7716
7717
7718
7719      +--sp|P04244|
7720      +-----5
7721      !      +--sp|P02067|
7722      +-3
7723      ! !      +-----sp|P18983|
7724      ! !
7725      ! +----9      +--sp|P68872|
7726      ! !      !      +----7
7727      ! !      !      +--sp|P02024|
7728      +-2      +--6
7729      ! !      !      +----sp|P60524|
7730      ! !      +----8      ! +--sp|P09905|
7731      ! !
7732      1 !      +--4
7733      ! !      +--sp|P02075|
7734      !
7735      ! +-----sp|P02094|
7736      !
7737      +-----sp|P02112|
7738
7739 remember: this is an unrooted tree!
7740
7741
7742 requires a total of    176.000
7743
7744

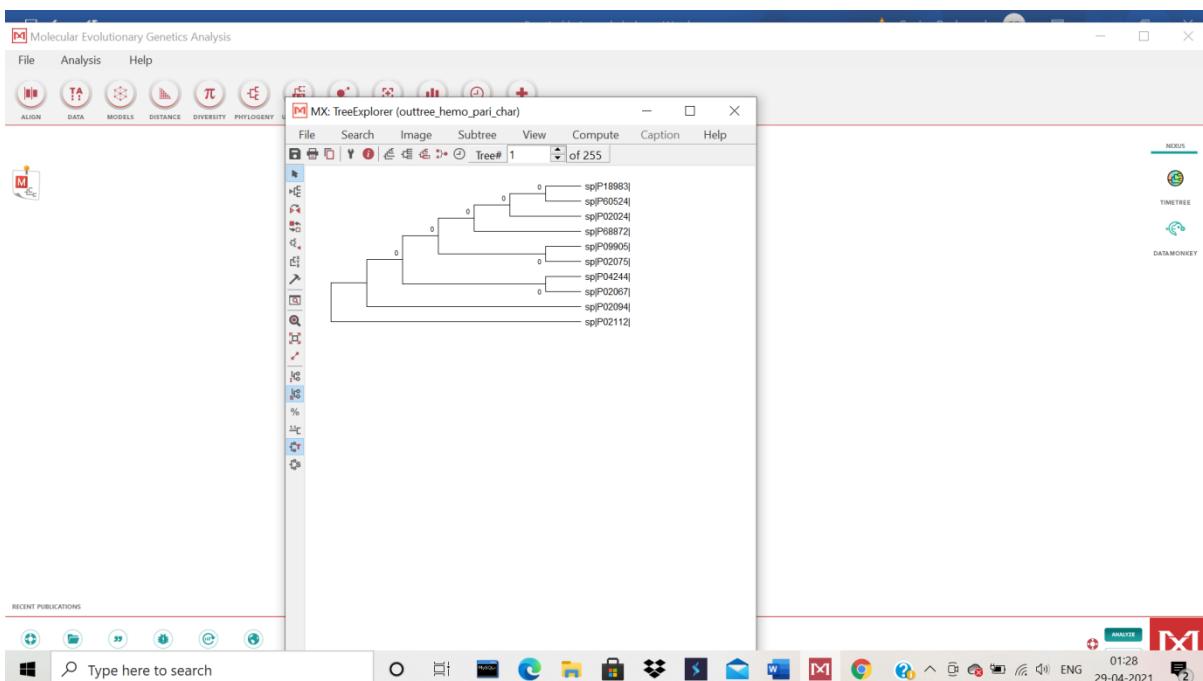

Normal text file length : 1,80,893 lines : 7,744 Ln : 7,744 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS
Type here to search 01:27 ENG 29-04-2021 2

```

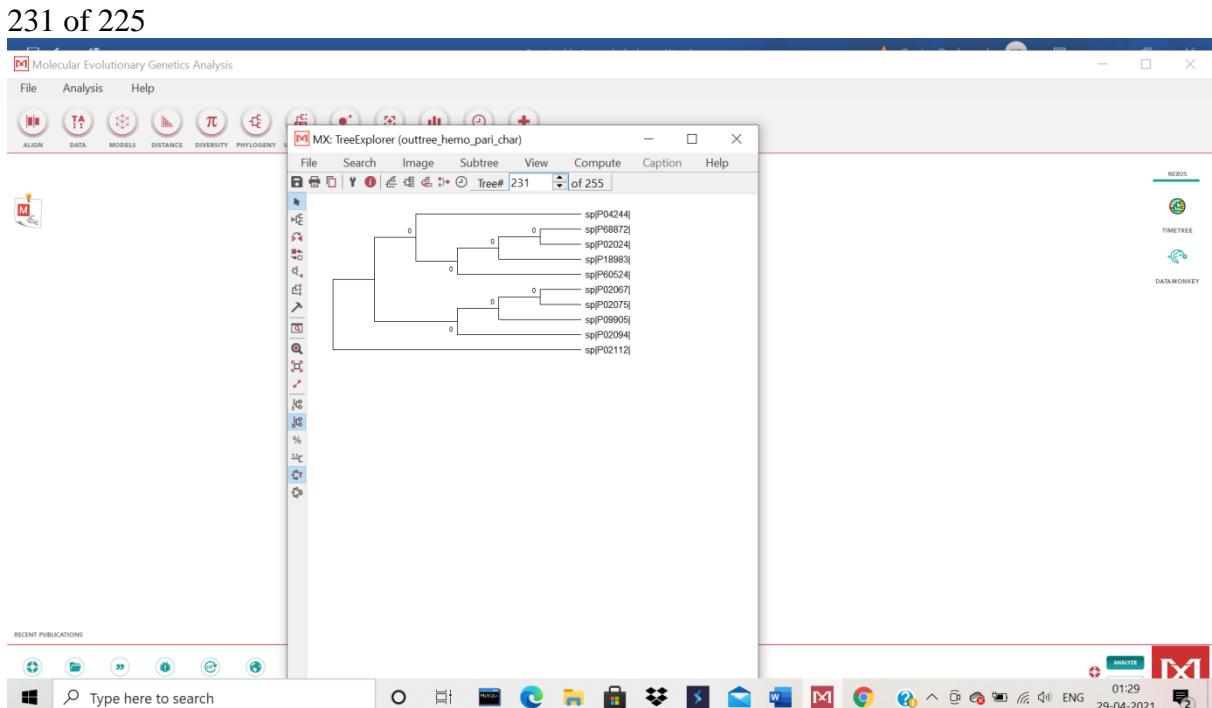
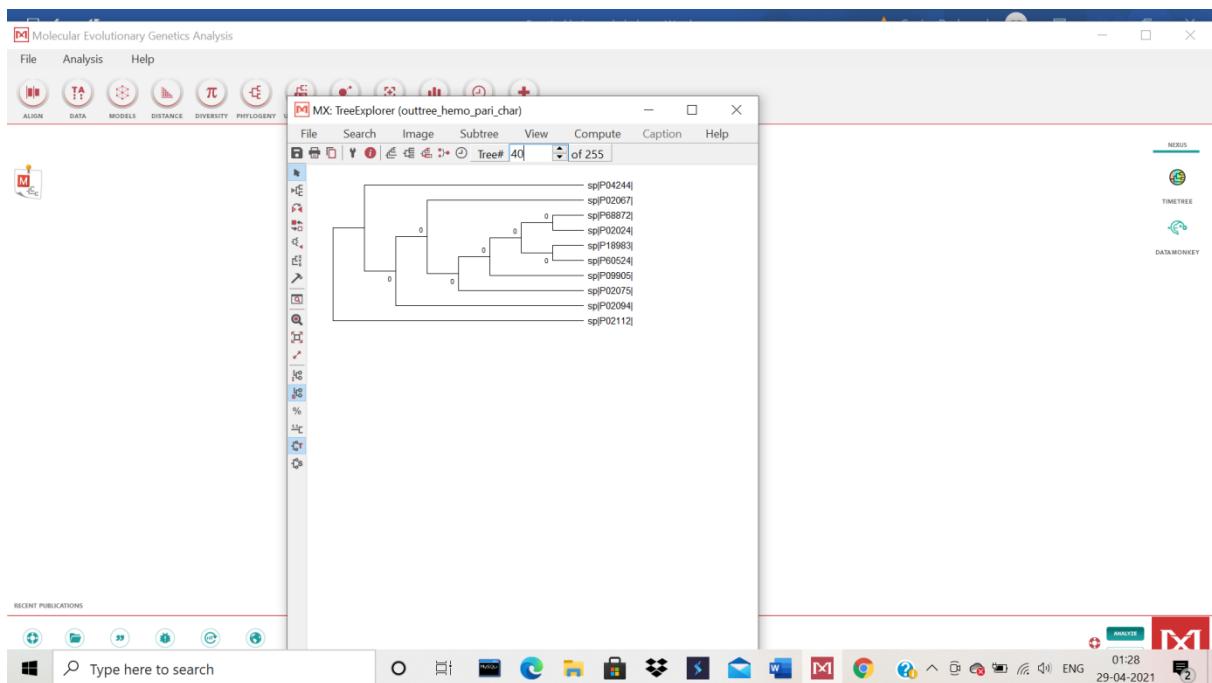
Open MEGA software to view newick phylogenetic tree and go to user tree, click on display newick trees:



1 of 225



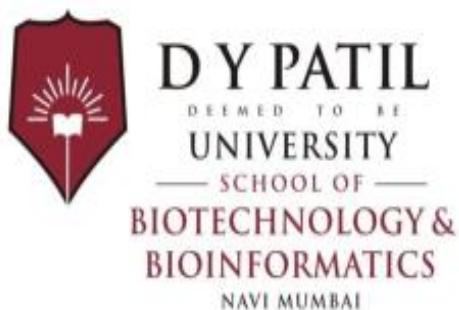
40 of 225



**D. Y. PATIL DEEMED TO BE UNIVERSITY,
NAVI MUMBAI**

**BID-4-P41 JOURNAL
Web Technologies, Databases and
Phyloinformatics
REPORT WRITING**

**B. Tech Bioinformatics
Semester IV
2020 – 2021**



**SCHOOL OF BIOTECHNOLOGY & BIOINFORMATICS
Plot No: 50, Sector 15, CBD Belapur
NAVI MUMBAI 400 614**



Certificate

This is to certify that the virtual lab experiments entered in this section of the journal have been performed by and the reports of the same have been written by **Mr. /**

Ms.Gunjan_Deshpande studying in School of Biotechnology and Bioinformatics, D.Y. Patil Deemed to be University, Navi Mumbai, **B. Tech Bioinformatics**,

Semester IV, Batch 2020-2021 Exam Number/
Enrollment number BID-19006 during the academic year 2020-2021.

Date:

Internal Examiner

External Examiner

Exp No.	Experiment Name	Page No.	Sign
1.	Global alignment of two sequences - Needleman-Wunsch Algorithm		
2.	Smith-Waterman Algorithm - Local Alignment of Sequences		
3.	Pairwise Sequence Alignment using BLAST		
4.	Pairwise sequence alignment using FASTA		
5.	Aligning Multiple Sequences with CLUSTAL W		
6.	Construction of Cladogram		
7.	Phylogenetic Analysis using PHYLIP - Rooted trees		
8.	Phylogenetic Analysis using PHYLIP - Unrooted trees		

Remarks:

Practical In-charge 1:

Virtual Lab Report

Experiment 1: Global alignment of two sequences - Needleman-Wunsch Algorithm

Aim: To perform global sequence alignment between two nucleotide or amino acid sequences and find out structural or functional similarity.

Theory: The most commonly asked question in molecular biology is whether two given sequences are related or not, in order to identify their structure or function. The most simpler way to answer this question is to compare their sequences.

Sequence is a collection of nucleotides or amino acid residues which are connected with each other. Speaking biologically, a typical DNA/RNA sequence consist of nucleotides while a protein sequence consist of amino acids.

Sequencing is the process to determine the nucleotide or amino acid sequence of a DNA fragment or a protein. There are different experimental methods for sequencing, and the obtained sequence is submitted to different databases like NCBI, Genbank etc.

Procedure:

The two sequences can be aligned globally using different algorithms. Needleman-Wunsch algorithm is one of the best algorithm for global alignment, which can be performed using the online tool EMBOSS Needle (European Molecular Biology Open Software Suite).

Steps to align the Sequences:

Step1:

1.To download the data and get access to the tools, go to simulator tab

2. Get access to the tool EMBOSS Needle

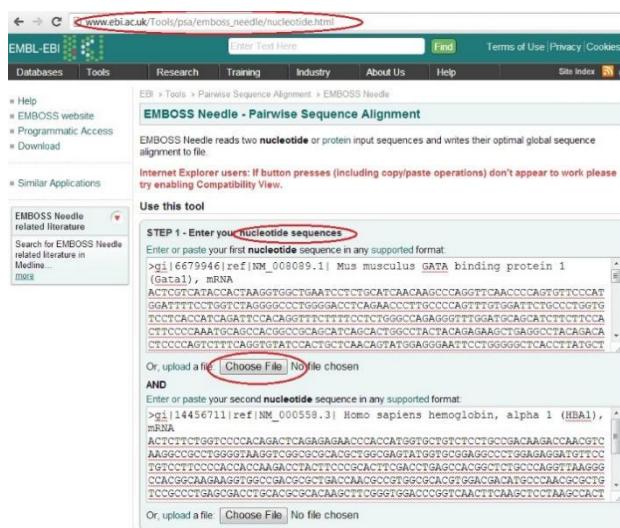


Figure 1: Screen shot to redirect EMBOSS needle and paste the query sequences for sequence alignment

3.Copy and paste the FASTA formatted (Computational representation of the DNA sequence) nucleotide sequence in the step 2 dialog box (as shown in figure 1).

4.One can also choose the file through “Choose File” option and can upload the sequence file.

5.Similarly copy and paste or upload the second sequence for the alignment.

Step 2:

- 1.EMBOSS needle is predefined with the scoring matrices DNAfull for nucleotide sequence, BLOSUM65 for protein sequence (Figure 5)
- 2.The gap open and gap extend penalty can be changed by user defined values. In this example it kept as default values.
- 3.The user can be notified with the results through email, if the checkbox is checked and the mail address is submitted.

STEP 2 - Set your pairwise alignment options

MATRIX	GAP OPEN	GAP EXTEND	OUTPUT FORMAT
DNAfull	10	0.5	pair
END GAP PENALTY	END GAP OPEN	END GAP EXTEND	
false	10	0.5	

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)

Submit

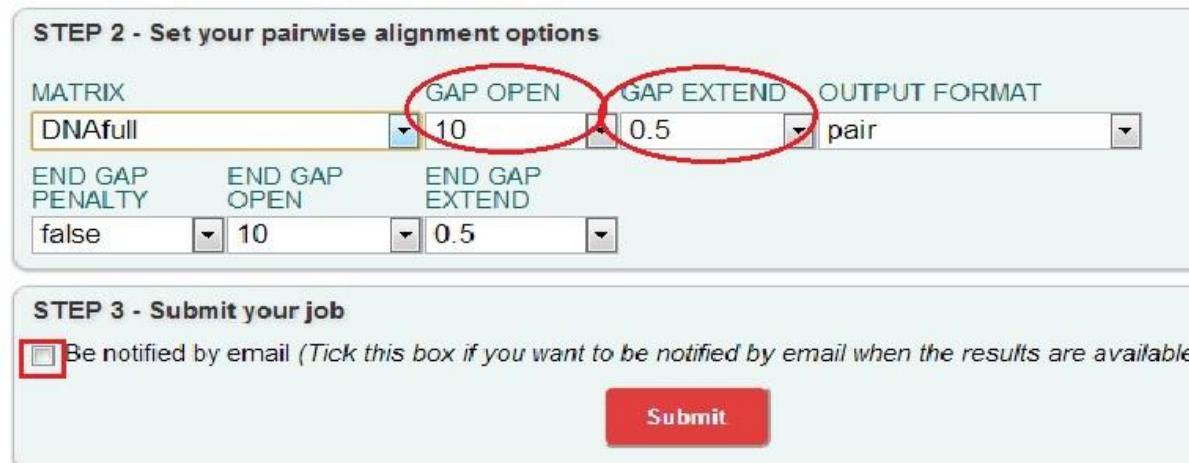


Figure 2: Screenshot to set the parameters for the pairwise sequence alignment using Needleman – Wunsch algorithm

Interpretation:

Once you have clicked on the submit button, the results are displayed within few minutes. The Results page comprises of three tabs namely Alignment, Submission details and submit another job. The Alignment tab shows the alignment of the two sequences, with all the described parameters, used scoring matrices and Gap penalty scored values. The Alignment tab has an option for the user to download the entire alignment file by clicking on the button “View Alignment File”. The submission details tab displays user to show the program used, time and date of when the program has launched and the internal commands used for the program execution. Also we can download the input and output files from this tab.

Experiment 2: Smith-Waterman Algorithm - Local Alignment of Sequences

Aim: To perform local sequence alignment between two nucleotide or amino acid sequences and find out structural or functional similarity.

Theory: The most commonly asked question in molecular biology is whether two given sequences are related or not, in order to identify their structure or function. The most simpler way to answer this question is to compare their sequences.

Sequence is a collection of nucleotides or amino acid residues which are connected with each other. Speaking biologically, a typical DNA/RNA sequence consist of nucleotides while a protein sequence consist of amino acids.

Sequencing is the process to determine the nucleotide or amino acid sequence of a DNA fragment or a protein. There are different experimental methods for sequencing, and the obtained sequence is submitted to different databases like NCBI, Genbank etc.

Procedure:

The two sequences can be aligned pairwise using different algorithms , Smith-Waterman algorithim is one of the best algorithm , which can be performed using the online tool EMBOSS water.

Steps to perform alignment

Step 1:

To download the data , and get access through the tools , go to simulator tab

1. Get access to the tool EMBOSS Water

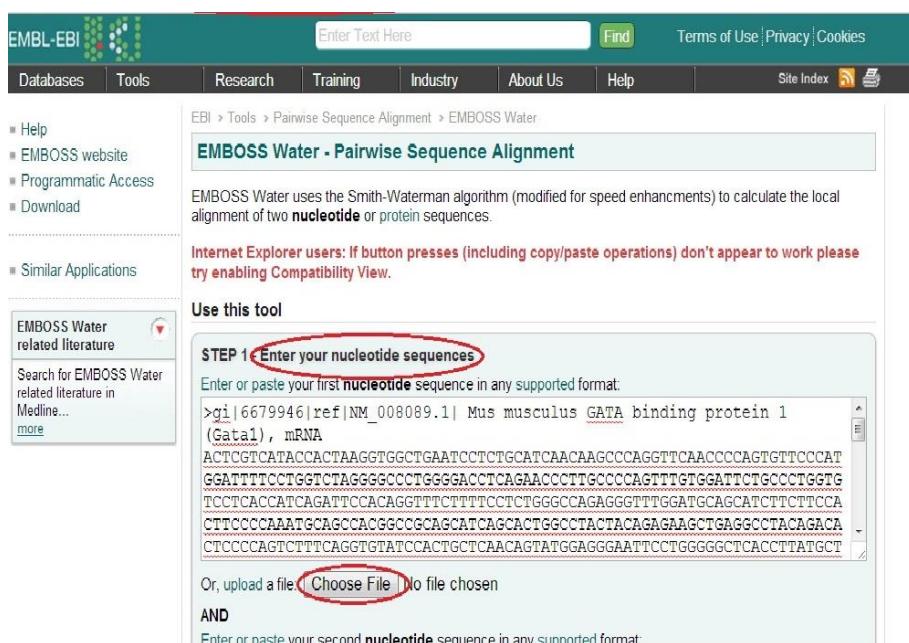


Figure 1: Screen shot to redirect EMBOSS water and paste the query sequences for sequence alignment

2.Copy and paste the FASTA formatted (Computational representation of the DNA sequence) nucleotide sequence in the step1 dialog box (as shown in Figure 1).

3. One can also choose the file through “Choose File” option and can upload the sequence file.

4.Similarly copy and paste or upload the second sequence for the alignment.

Step 2: