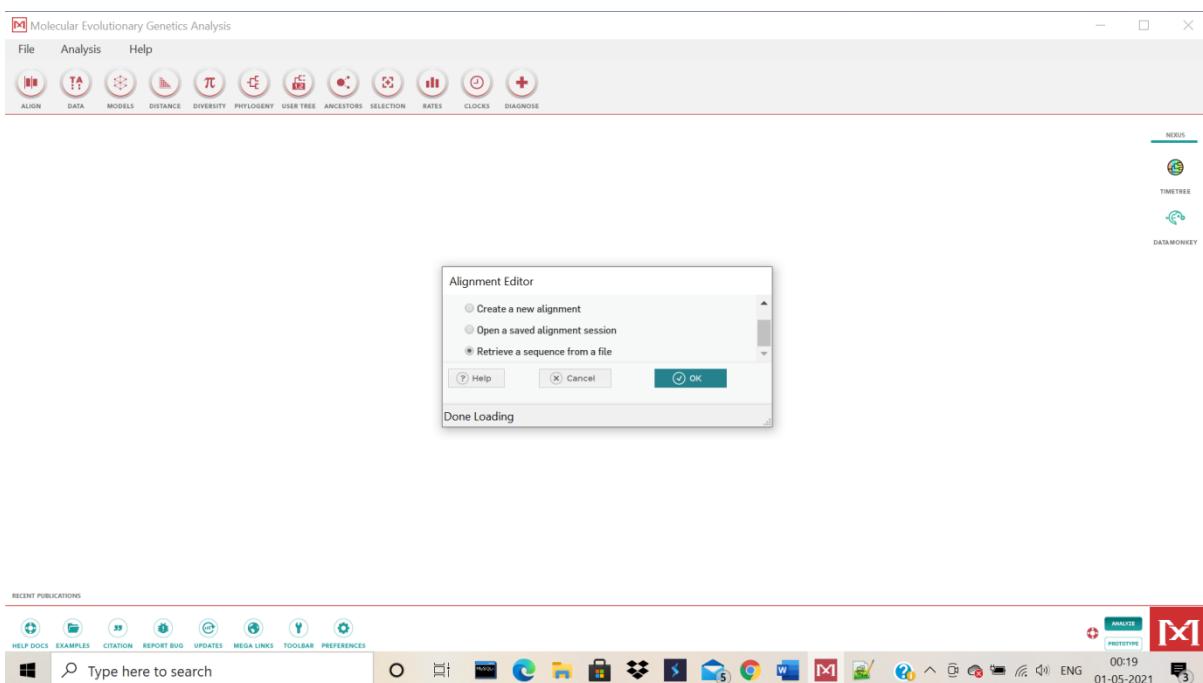


In the alignment editor you need to select the option of retrieve a sequence from a file:



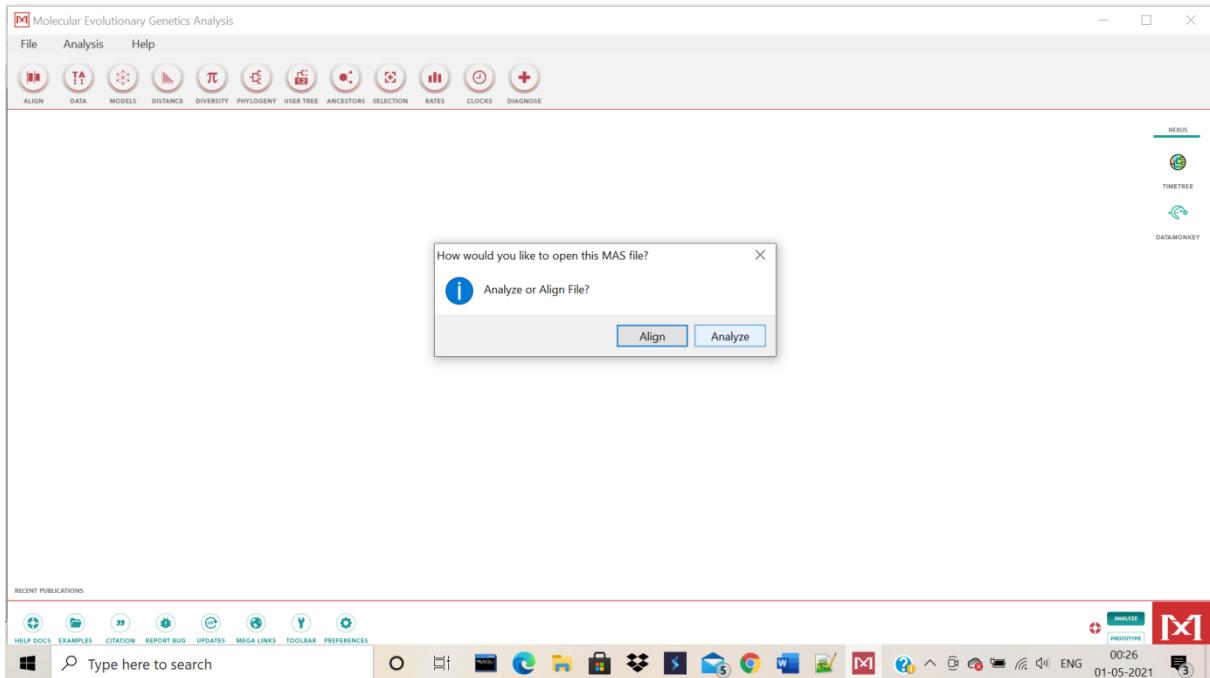
After retrieving a sequence we get this output:

Go to align using the muscle algorithm to align your sequence algorithm:

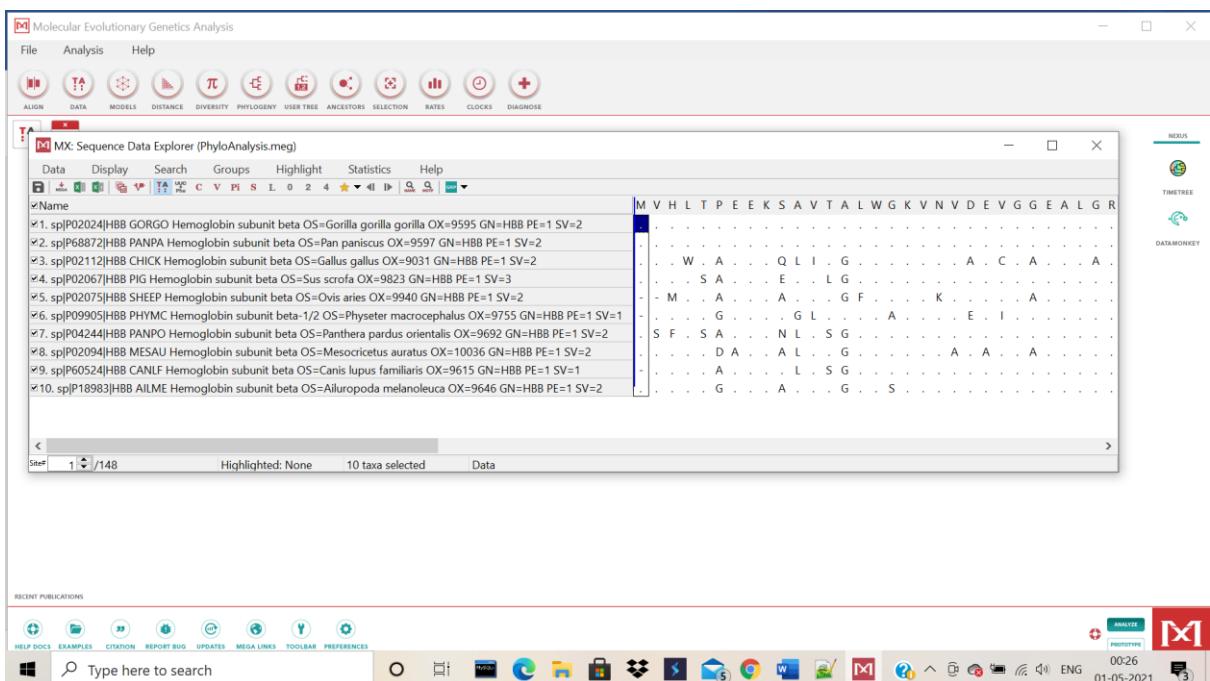
Select UPGMA as your cluster method, keep min diag length as 24 and click on ok:

You will get your new aligned sequence:

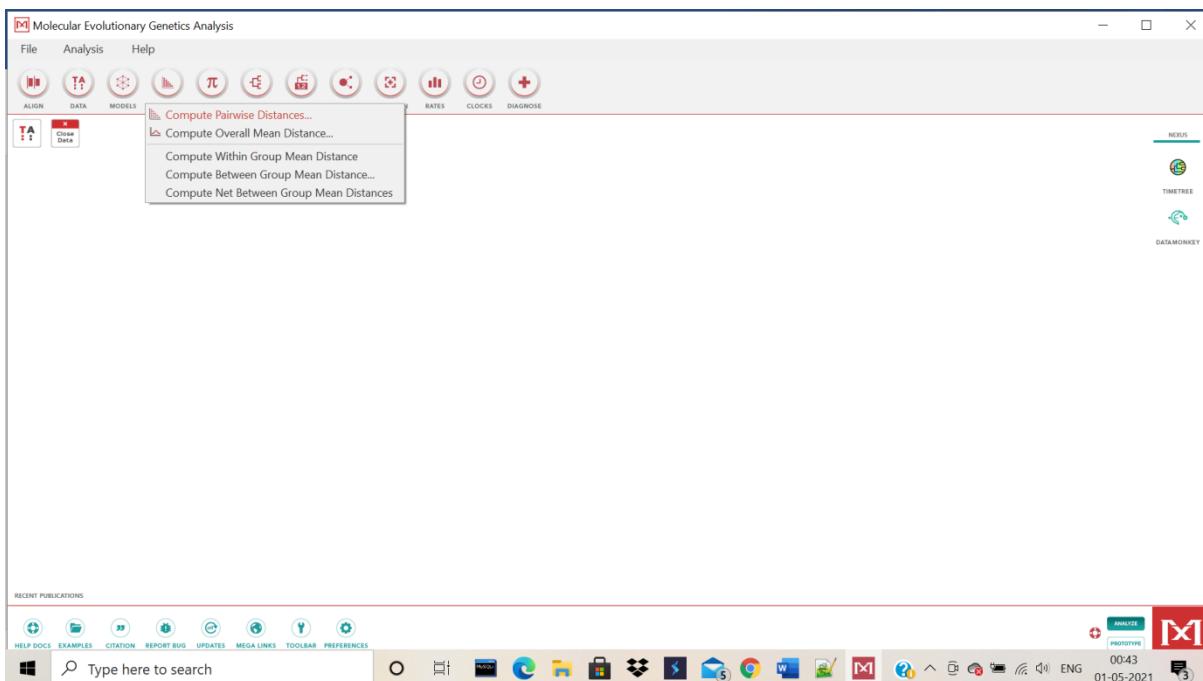
Click on analyze:



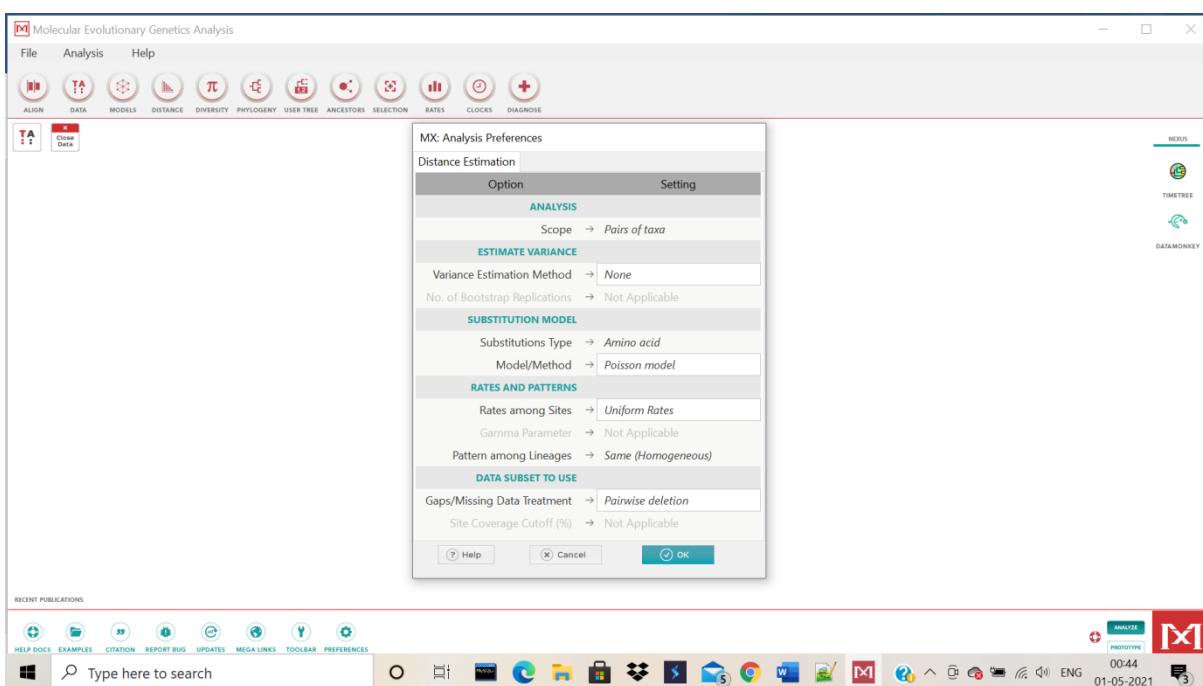
You will find your new analyzed sequence for 10 hemoglobin sequences of different organisms:



Go to distance and select compute pairwise distances:



Select poisson model, select pairwise deletion as data subset to be used and press ok:



Distance based Matrix for hemoglobin sequences of 10 different organisms:

1. sp P02024 HBB GORGO Hemoglobin subunit beta OS=Gorilla gorilla gorilla OX=9595 GN=HBB PE=1 SV=2							
2. sp P68872 HBB PANPA Hemoglobin subunit beta OS=Pan panicus OX=9597 GN=HBB PE=1 SV=2	0.0050						
3. sp P02112 HBB CHICK Hemoglobin subunit beta OS= Gallus gallus OX=9031 GN=HBB PE=1 SV=2	0.3750	0.3654					
4. sp P02067 HBB PIG Hemoglobin subunit beta OS=Sus scrofa OX=9823 GN=HBB PE=1 SV=3	0.17015	0.16212	0.40547				
5. sp P02075 HBB SHEEP Hemoglobin subunit beta OS=Ovis aries OX=9940 GN=HBB PE=1 SV=2	0.20605	0.19761	0.42206	0.18094			
6. sp P09905 HBB PHYMC Hemoglobin subunit beta-1/2 OS=Physeter macrocephalus OX=9755 GN=HBB PE=1 SV=1	0.16310	0.19529	0.40990	0.23002	0.22143		
7. sp P04244 HBB PANPO Hemoglobin subunit beta OS=Panthera pardus orientalis OX=9692 GN=HBB PE=1 SV=2	0.21875	0.21131	0.41572	0.19841	0.24054	0.30148	
8. sp P20204 HBB MESAU Hemoglobin subunit beta OS=Mesocricetus auratus OX=10036 GN=HBB PE=1 SV=2	0.27193	0.29050	0.41572	0.26304	0.26535	0.32017	0.39647
9. sp P60524 HBB CANLF Hemoglobin subunit beta OS=Canis lupus familiaris OX=9615 GN=HBB PE=1 SV=1	0.10800	0.19841	0.33932	0.17959	0.23002	0.30294	0.13158
10. sp P18983 HBB AILME Hemoglobin subunit beta OS=Aluropoda melanoleuca OX=9646 GN=HBB PE=1 SV=2	0.09259	0.10000	0.36546	0.17625	0.19824	0.17142	0.17625

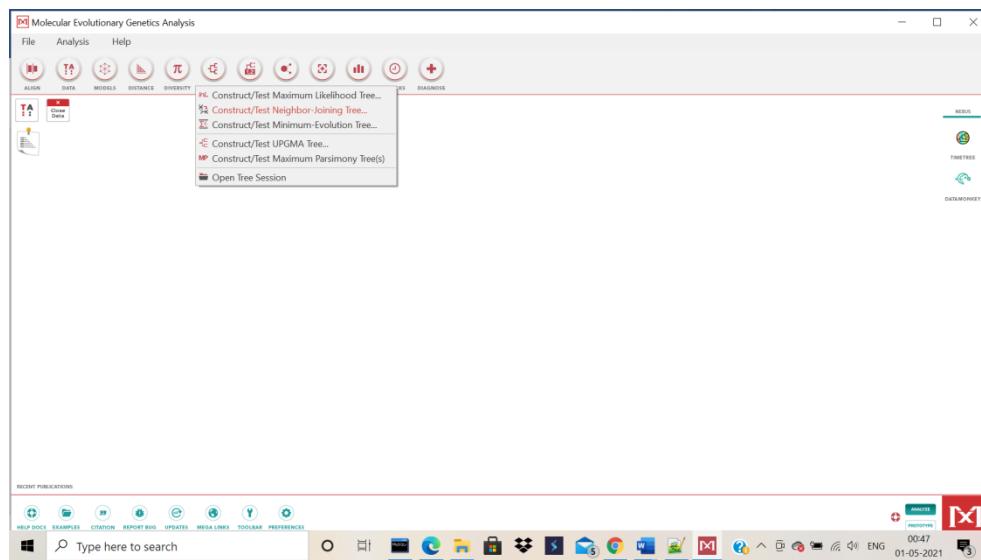
### Caption:

**Table. Estimates of Evolutionary Divergence between Sequences**

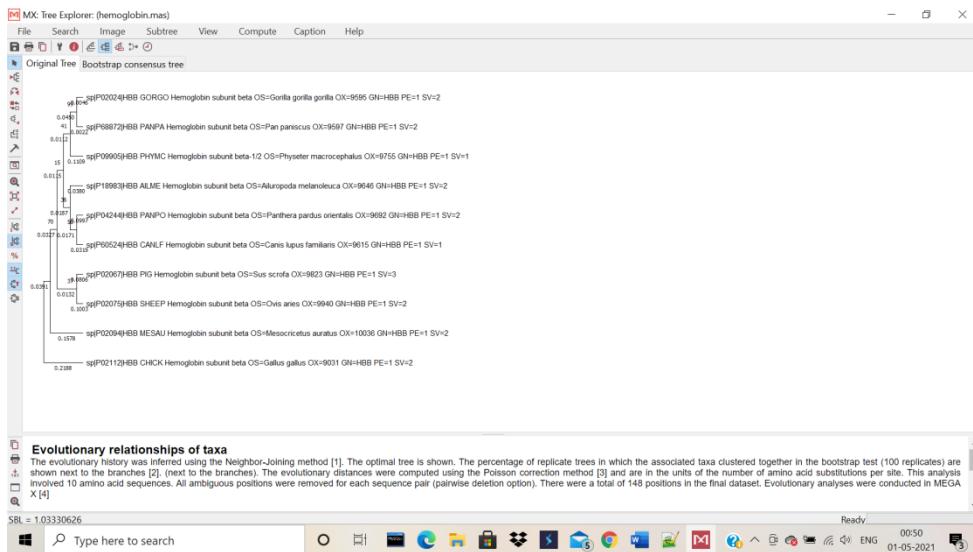
The number of amino acid substitutions per site from between sequences are shown. Analyses were conducted using the Poisson correction model [1]. This analysis involved 10 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 148 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [2].

Performing distance based phylogenetic analysis of hemoglobin sequences of 10 different organisms by **Neighbor-joining** method:

Go to phylogeny and click on construct/test neighbor joining tree:



Original tree:

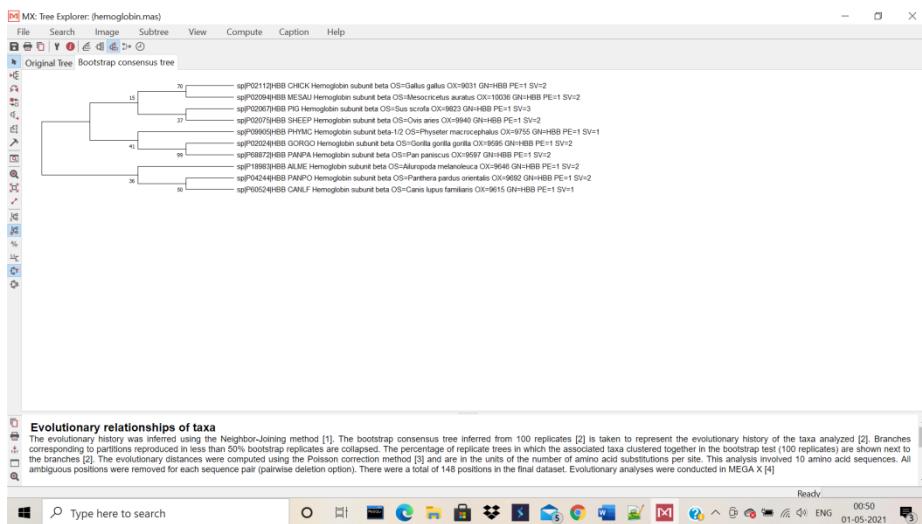


Caption:

### Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. 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]. (next to the branches). The evolutionary distances were computed using the Poisson correction method [3] and are in the units of the number of amino acid substitutions per site. This analysis involved 10 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 148 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [4]

Bootstrap consensus tree:



Caption:

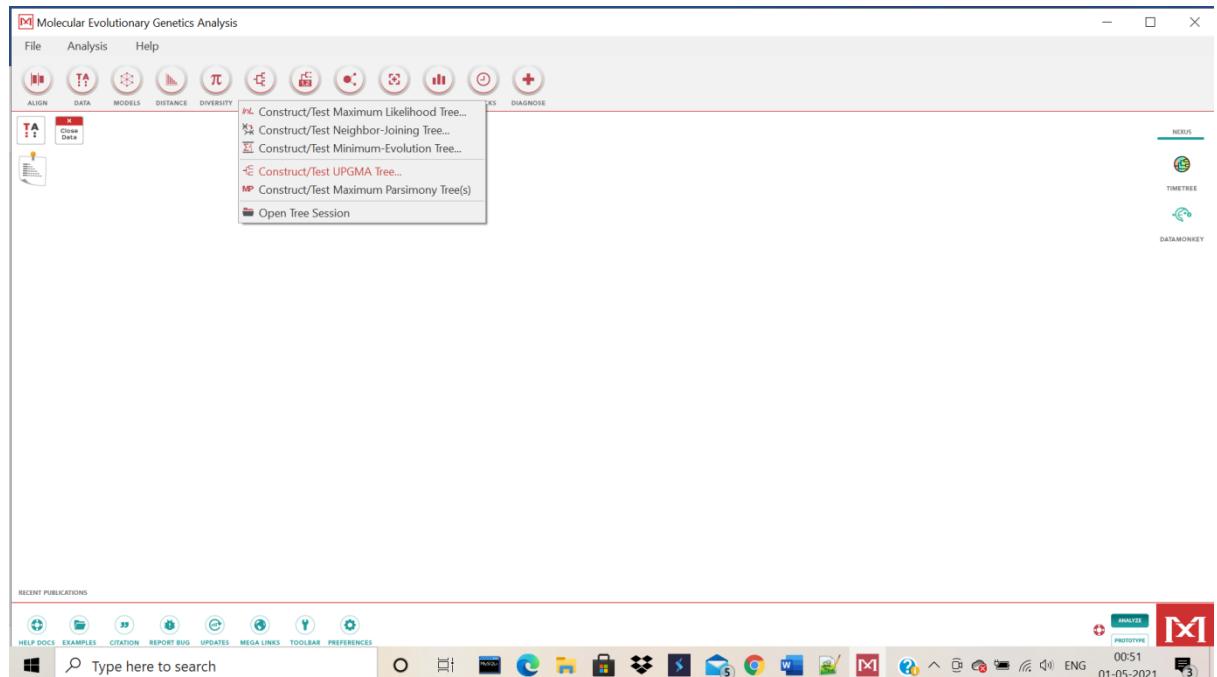
### Evolutionary relationships of taxa

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

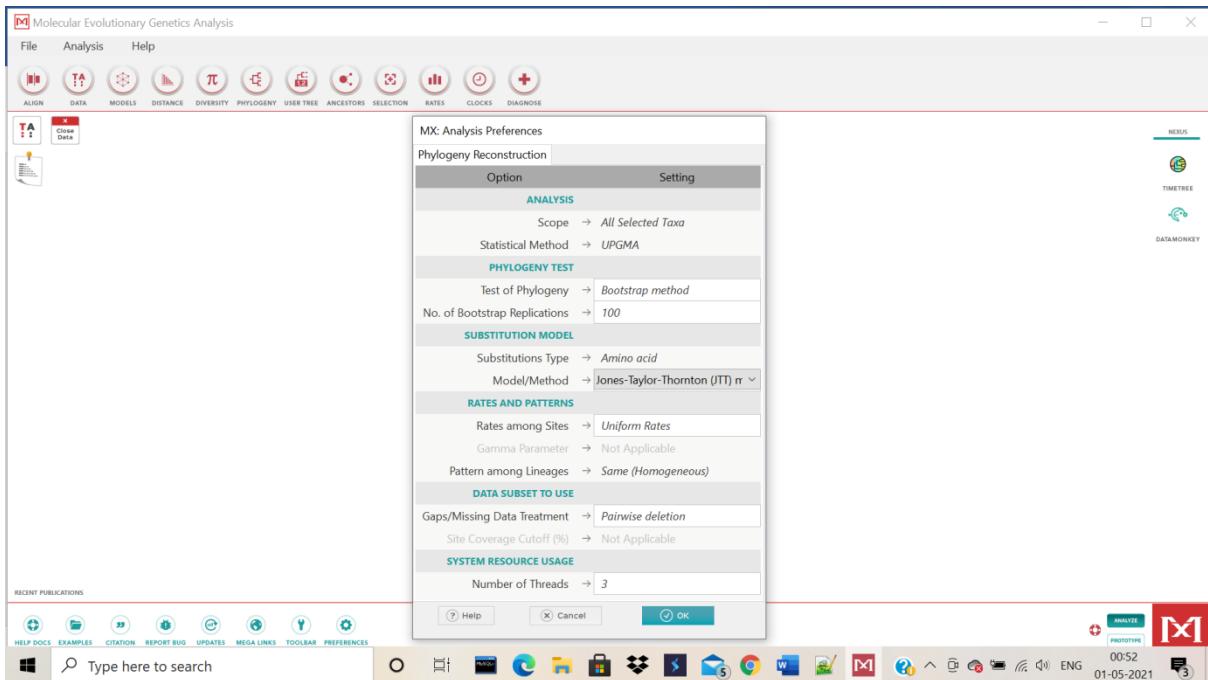
replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches [2]. The evolutionary distances were computed using the Poisson correction method [3] and are in the units of the number of amino acid substitutions per site. This analysis involved 10 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 148 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [4]

Performing distance based phylogenetic analysis of hemoglobin sequences of 10 different organisms by **UPGMA** method:

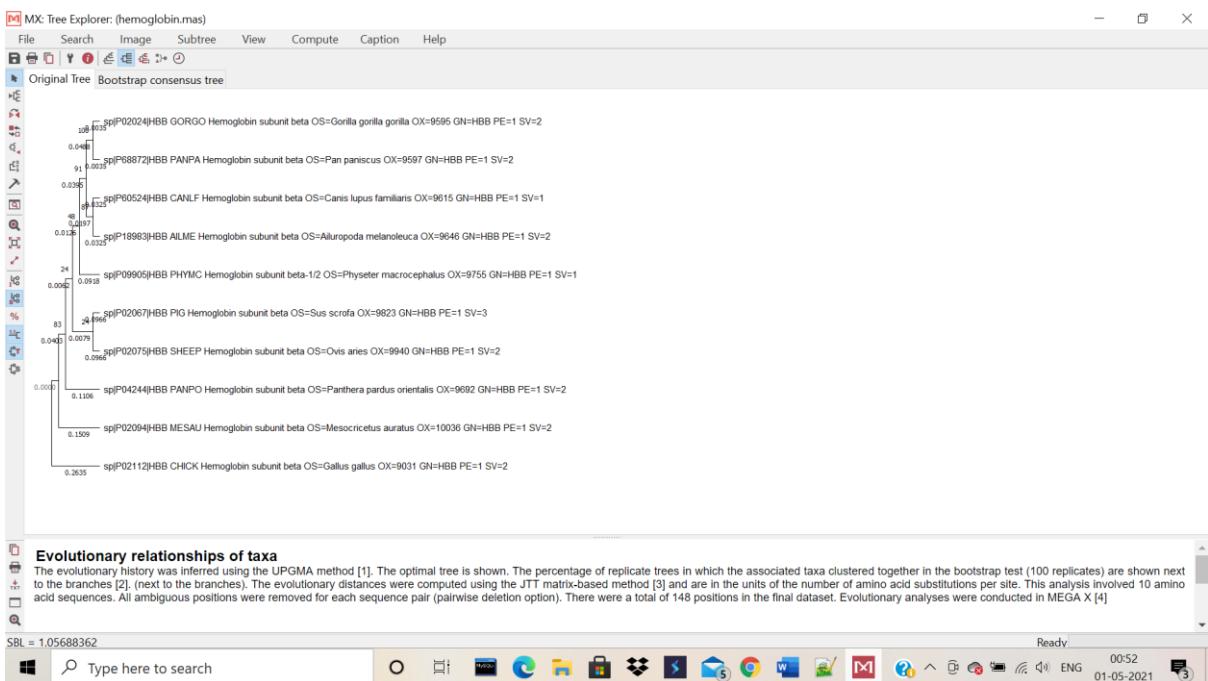
Go to phylogeny and click on construct/test upgma tree:



Select bootstrap method as test of phylogeny , type 100 as number of bootstrap replications, take 3 as number of threads:



## Original Tree:

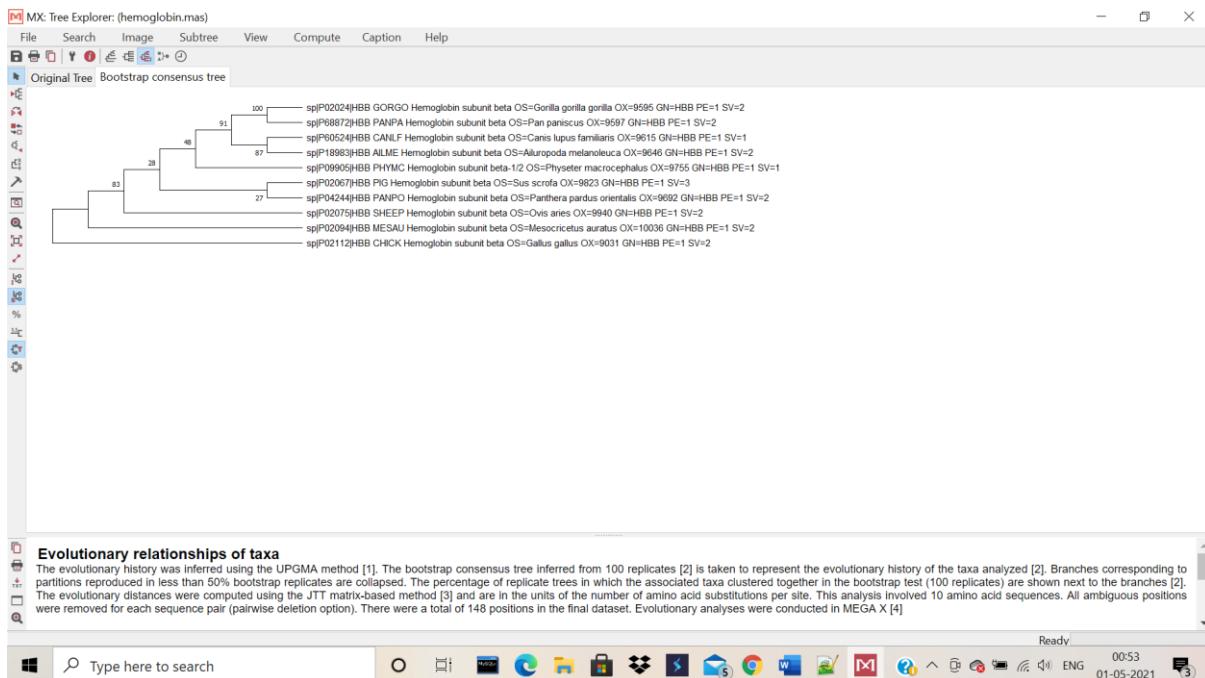


## Caption:

### Evolutionary relationships of taxa

The evolutionary history was inferred using the UPGMA method [1]. The optimal tree is shown. 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]. (next to the branches). The evolutionary distances were computed using the JTT matrix-based method [3] and are in the units of the number of amino acid substitutions per site. This analysis involved 10 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 148 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [4]

## Bootstrap consensus tree:



Caption:

### Evolutionary relationships of taxa

The evolutionary history was inferred using the UPGMA 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 JTT matrix-based method [3] and are in the units of the number of amino acid substitutions per site. This analysis involved 10 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 148 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [4]

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

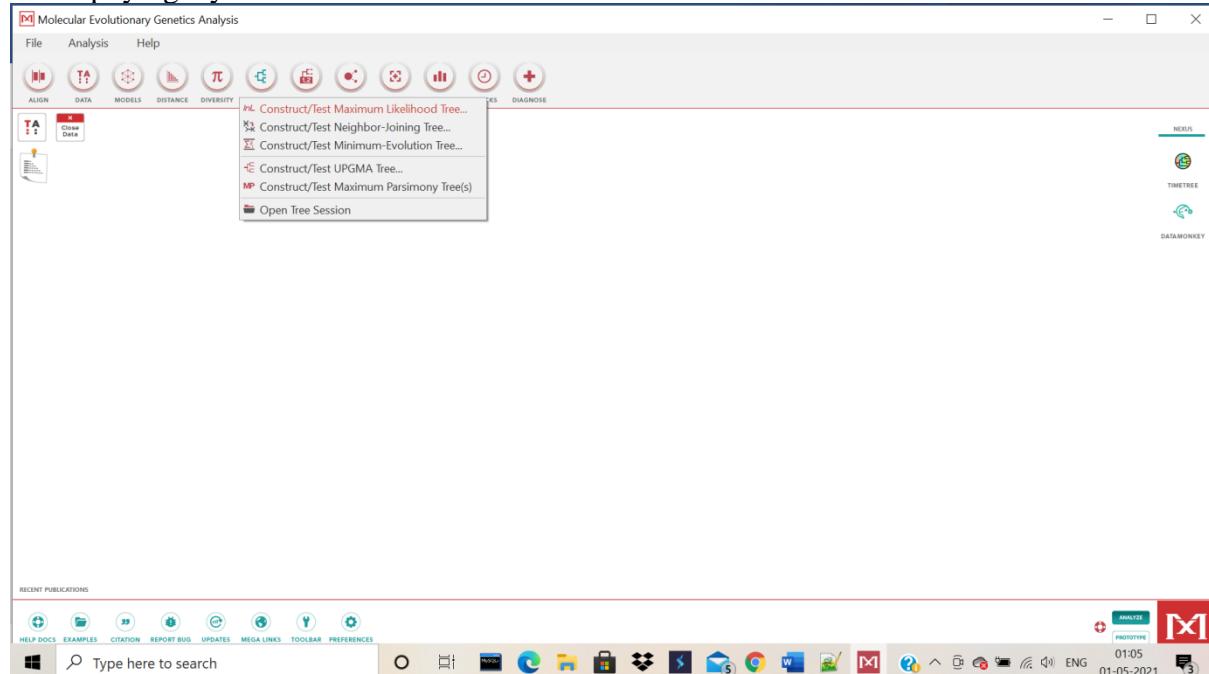
Answer:

Hemoglobin sequences of 10 different organisms:

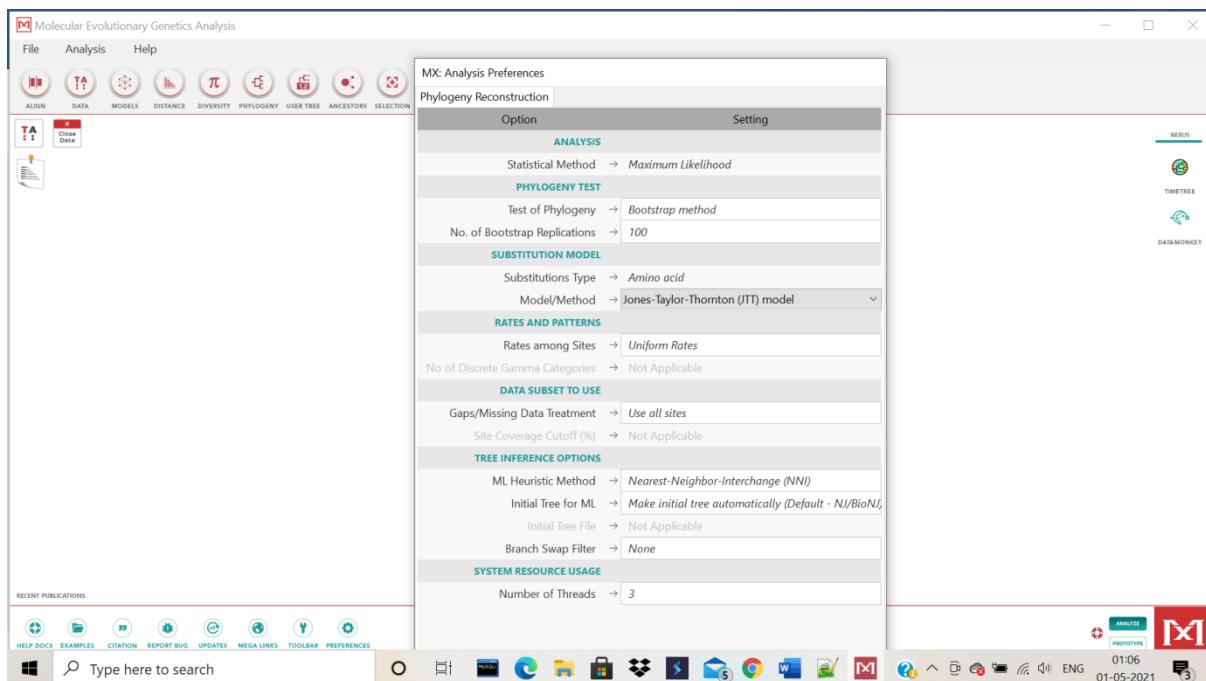
```
>sp|P02024|= Gorilla Hemoglobin
>sp|P68872|= Pan paniscus Hemoglobin
>sp|P02112|= CHICK Hemoglobin
>sp|P02067|= PIG Hemoglobin
>sp|P02075|= SHEEP Hemoglobin
>sp|P09905|= Physeter macrocephalus Hemoglobin
```

>sp|P04244| = Panthera pardus orientalis Hemoglobin  
>sp|P02094| = Mesocricetus auratus Hemoglobin  
>sp|P60524| = Canis lupus Hemoglobin  
>sp|P18983| = Ailuropoda melanoleuca

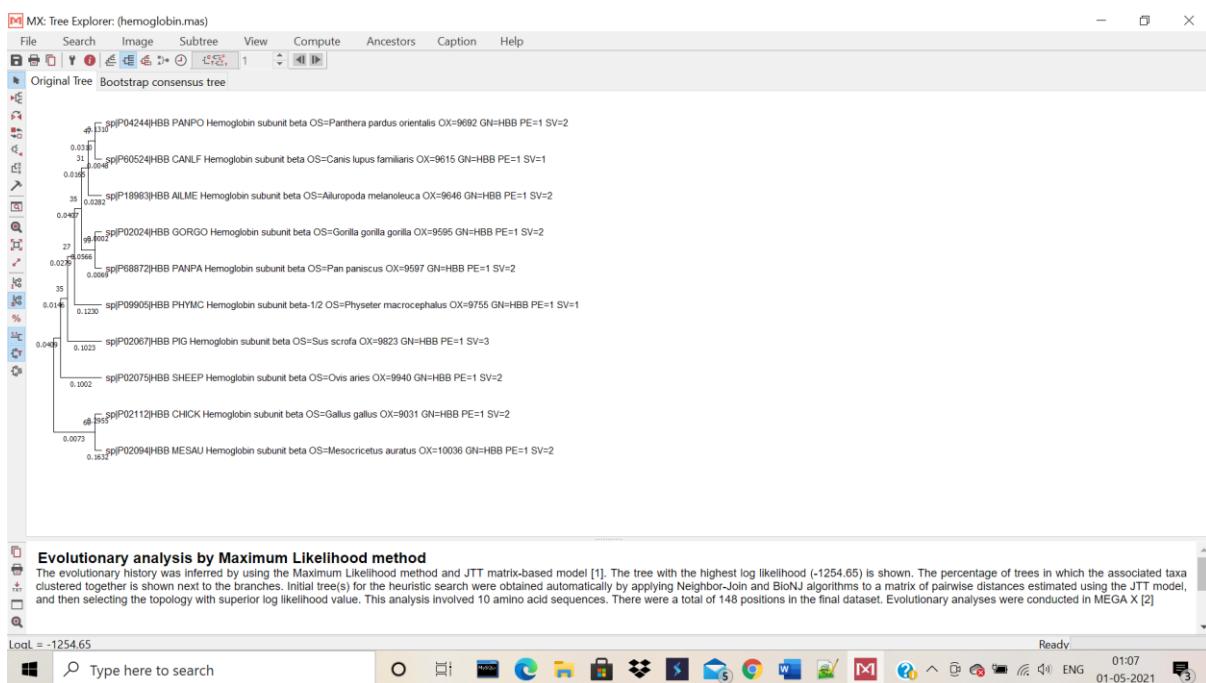
Go to phylogeny and select construct/test maximum likelihood tree:



Take 100 number of bootstrap replications, take jones-taylor-thomton (JTT) model, number of threads as 3:



## Original Tree:

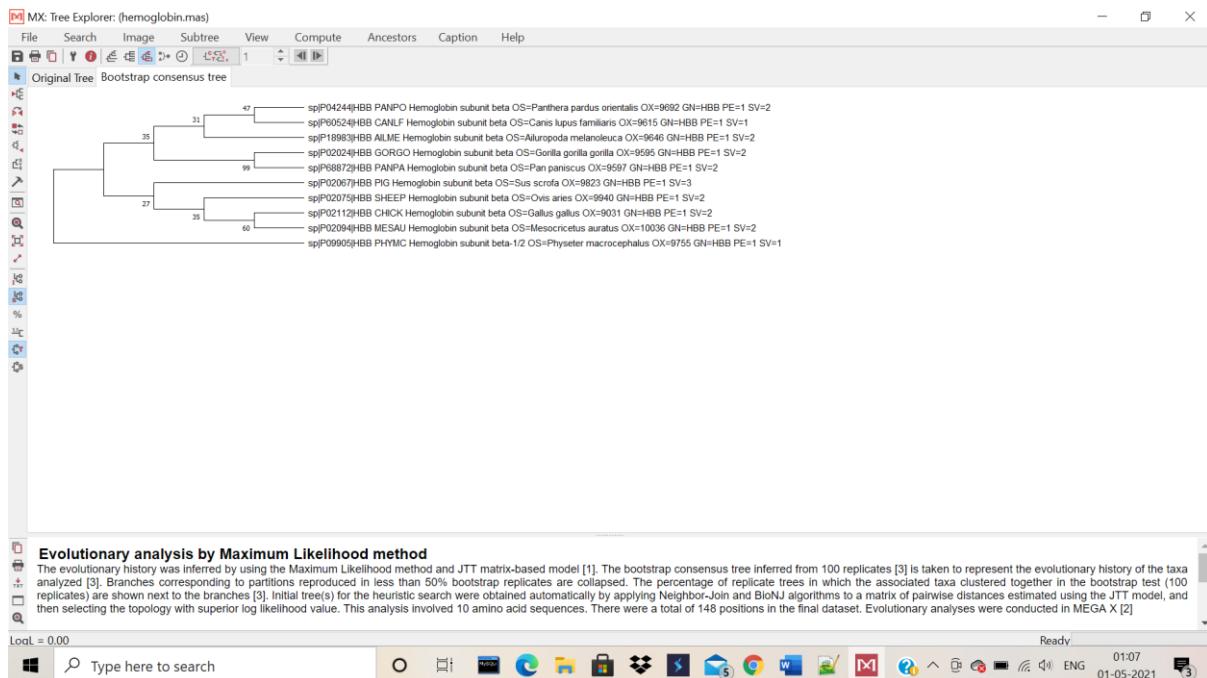


## Caption:

### Evolutionary analysis by Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model [1]. The tree with the highest log likelihood (-1254.65) 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 JTT model, and then selecting the topology with superior log likelihood value. This analysis involved 10 amino acid sequences. There were a total of 148 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [2]

## Bootstrap consensus tree:



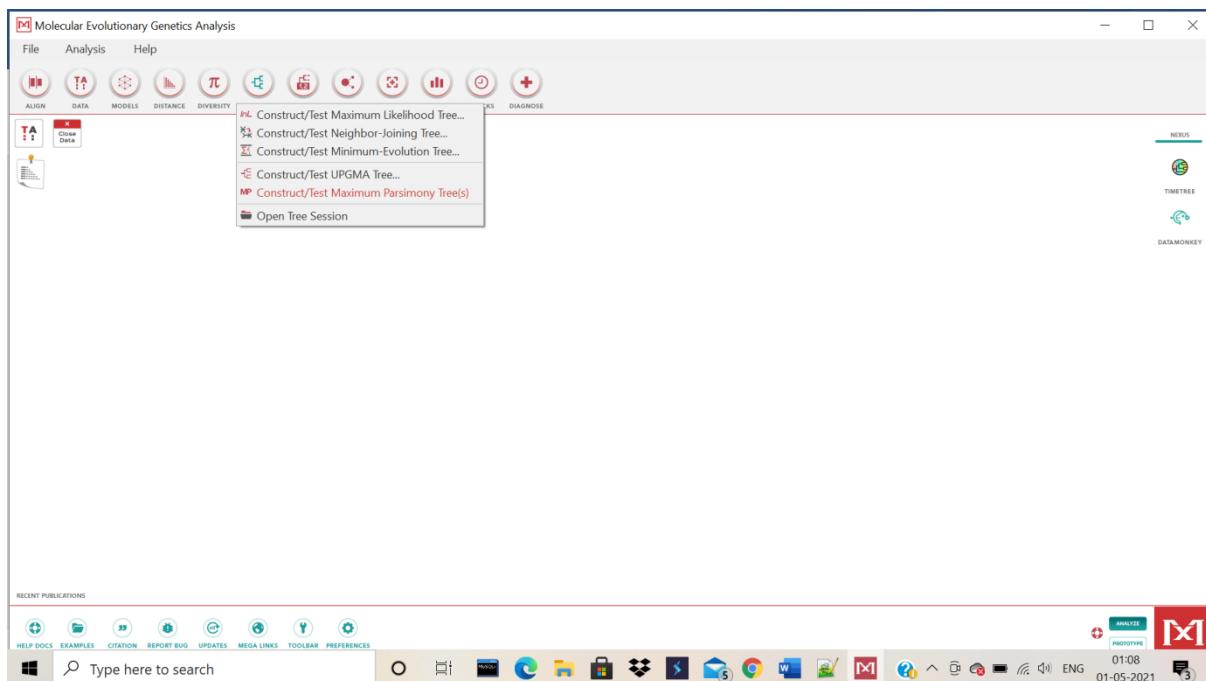
Caption:

### Evolutionary analysis by Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based 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 JTT model, and then selecting the topology with superior log likelihood value. This analysis involved 10 amino acid sequences. There were a total of 148 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [2]

Performing distance based phylogenetic analysis of hemoglobin sequences of 10 different organisms by **Parsimony** method:

Go to phylogeny and click on construct/test maximum parsimony tree:

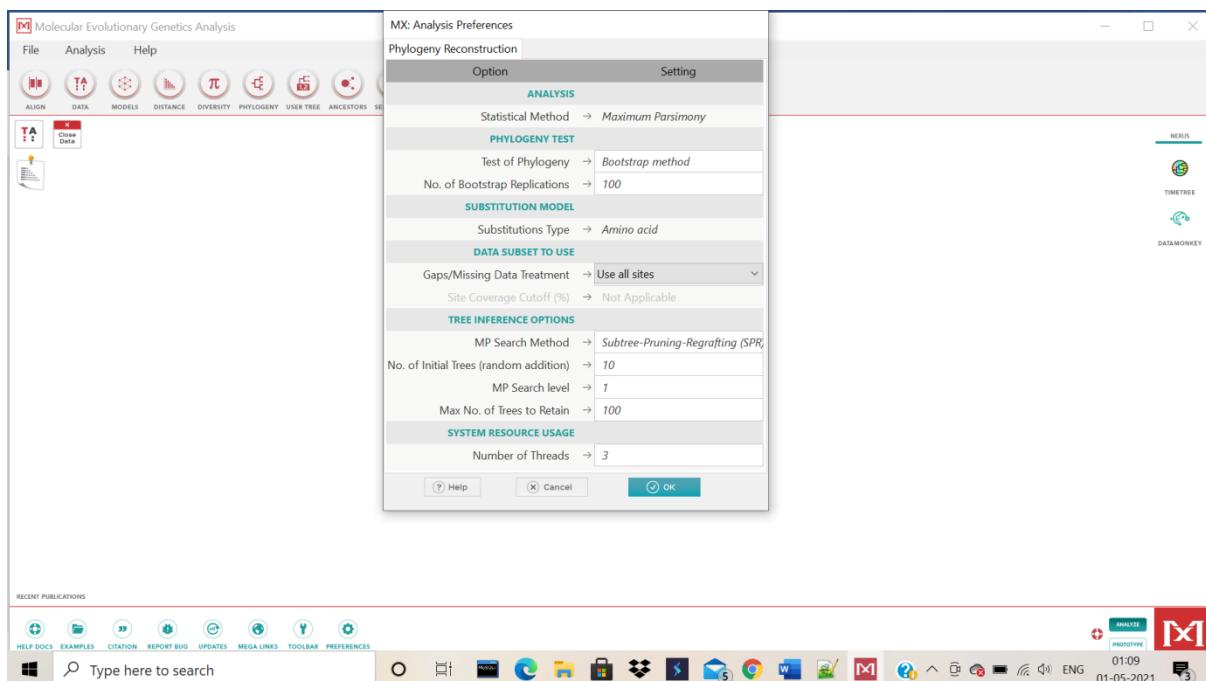


Select the following options in your box given:

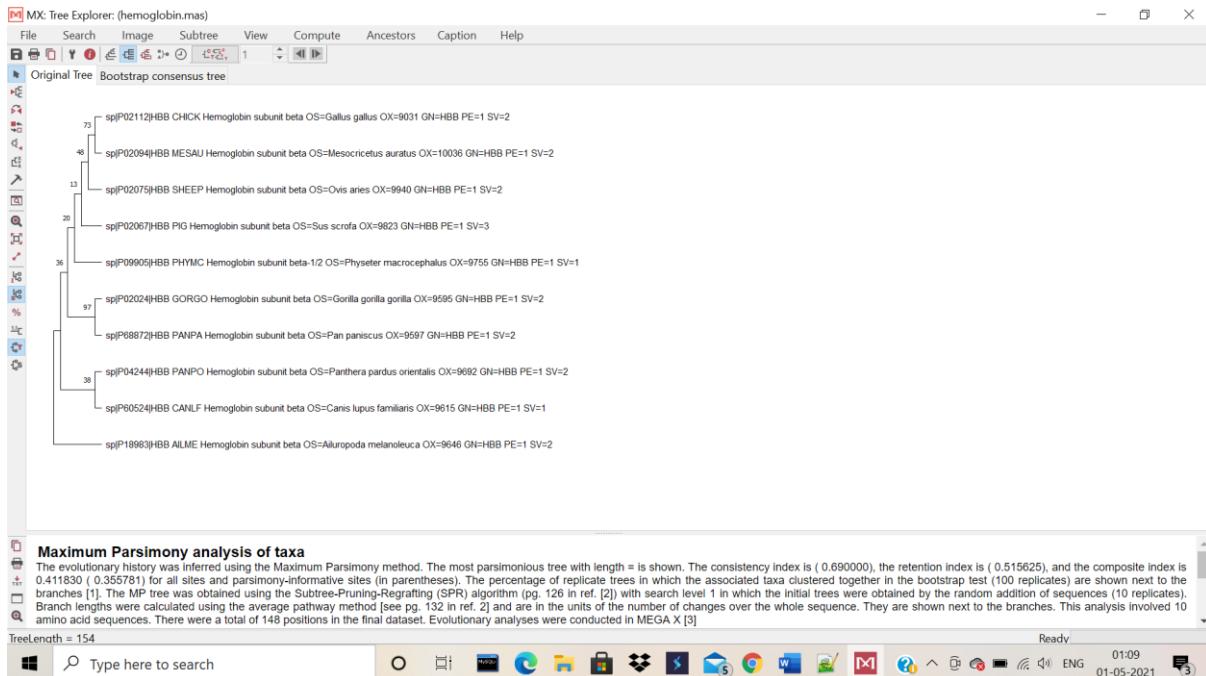
Test of phylogeny: Bootstrap method

No. of bootstrap replications: 100

No. of threads: 3



Original tree:



Caption:

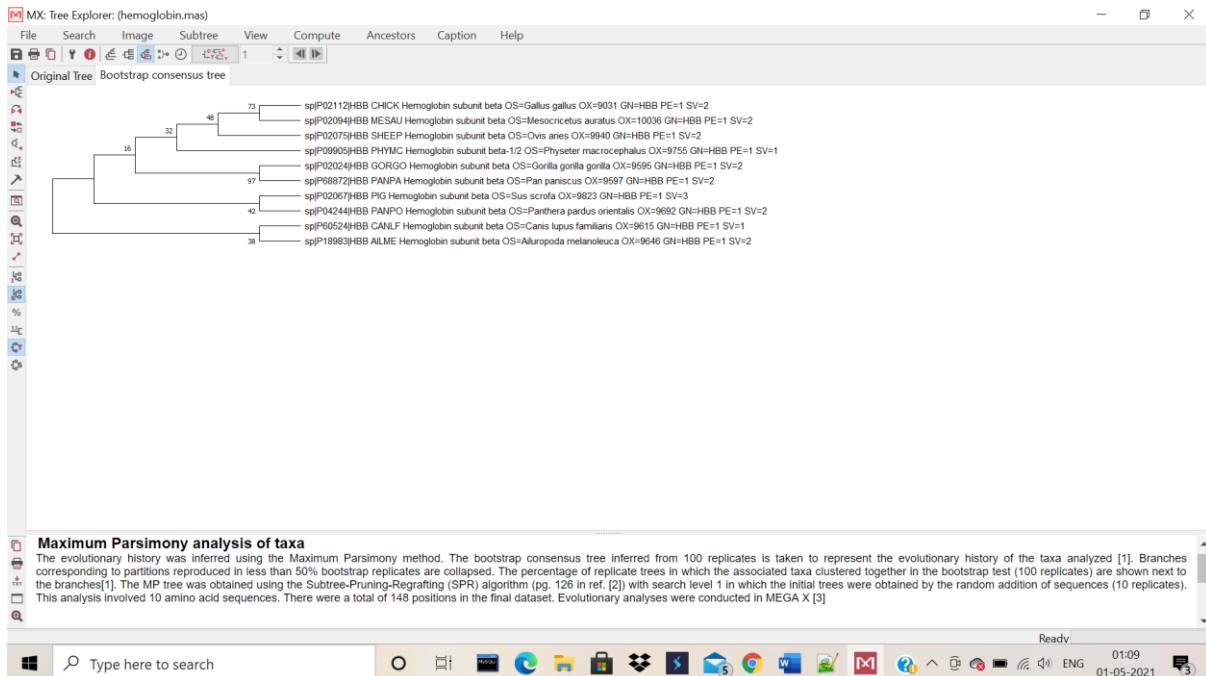
### Maximum Parsimony analysis of taxa

The evolutionary history was inferred using the Maximum Parsimony method. The most parsimonious tree with length = is shown. The consistency index is ( 0.690000), the retention index is ( 0.515625), and the composite index is 0.411830 ( 0.355781) 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 10 amino acid sequences. There were a total of 148 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [3]

Bootstrap

consensus

tree:



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 10 amino acid sequences. There were a total of 148 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [3]

3. Perform distance 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

Copy paste all the nucleotide sequences in notepad ++ :

```
C:\Users\Prasad\Downloads\ginsulink.fas - Notepad+
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
outfile_parse_seqboot outfile_parse_seqboot ginsulink.fas name index.php myfirstprogram.php StudentForm.html folder ginsulink.fas
1 >M57671.1 Octodon degus insulin mRNA, complete cds
2 GCATCTGAGGCACTTCAACAGGTTCTGGACCTCCGGCATGGCCCGTGGATGCATCTCCCTCACCGT
3 GCTGGCCCTGCTGGCCCTTGAGGACCAACTCTGTGAGGCTTATTCCAGGCCAGCACCTGTGGCTCC
4 AACCTAGTGGAGGCACTGTACATGACATGTGGACGGAGTGGCTTCTATAGACCACGACCAGCAGAGC
5 TGGAGGACCTTCAGGTGGAGCAGGCAACTGGGCTTGAGGAGCAGGGCCCTGCAGCCTTCGGCCCTGGA
6 GATGATTCTGAGAACGGCCGATTTGGATCAGTGGCTATAACATTTGACACATTAAACAGCTGCA
7 AACTACTGCAATGTCCCTAGACACCTGCTTGGCCCTGGCTGCTGCTCTGCCCTGGCAACCAAATAAAC
8 CCCTTGATGAG
9
10 >M61153.1 Oryctolagus cuniculus insulin mRNA, partial cds
11 CGGGCTGGGAGGAGCTGGGGCAGGGGAGCTGGGGCAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
12 CCTGGGGCTTATCTGGCCCTGAGAACGGGGCATCTGGAGCAGTTGACCATCTGCTGGCT
13 CTACCAAGCTGGAGAACTACTGCAACTAGGGGTGCCCCCACCCACCCCTGGCCGCCCCCACGCC
14 CGCCCTGCCCAACCAATAAACCCCTCCAGCGCCCCC
15
16
17 >U03610.1 Oryctolagus cuniculus New Zealand White insulin mRNA, complete cds
18 TCATGGCTCTGACCATGGCCCTGGCCGCTCTGGCCCTGCTGGCCCTGCTGGCCCTCTGAGA
19 CTGGATCTGCCCAAGGCTCTGCAACACAGCACCTGTGGGCTCTCACCTGGTGGAGGCGCTGTACCTGG
20 TGTCGGGGAGCCGGGTTTTATACACCCAAAGTCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
21 CGCGGAGCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
22 CGCGGCATCTGGAGCAGTTGACCAAGCATCTGCTGCTCTACAGCTGGAGAAGACTGCAACTAGG
23 GGTGCCCCCACCCACCCCTGGCCGCCCCCACGCCCTGGCCCAATAAACCCCTC
24 CACCGGCCCCGGG
25
26
27 >JF909299.1 Homo sapiens insulin (INS) mRNA, partial cds
28 CTGGGACCTGACCCAGCCGAGCCTTGTGAACCAACACTGTGGGCTCACACCTGGTGGAGCTCTC
```

Go to distance and click on compute pairwise distance :

The screenshot shows the MEGA X software interface with the 'Distance Estimation' analysis preferences dialog box open. The dialog box is divided into several sections: ANALYSIS, ESTIMATE VARIANCE, SUBSTITUTION MODEL, RATES AND PATTERNS, and DATA SUBSET TO USE. Under ANALYSIS, the scope is set to 'Pairs of taxa'. In the ESTIMATE VARIANCE section, the variance estimation method is 'None' and the number of bootstrap replications is 'Not Applicable'. The SUBSTITUTION MODEL section shows 'Nucleotide' as the substitution type, 'Not Applicable' for the genetic code table, and 'Maximum Composite Likelihood' as the model/method. Under RATES AND PATTERNS, 'Uniform Rates' is selected for rates among sites, and 'Same (Homogeneous)' is selected for patterns among lineages. In the DATA SUBSET TO USE section, 'Pairwise deletion' is chosen for gaps/missing data treatment, and 'Not Applicable' is selected for site coverage cutoff. For selecting codon positions, options for 1st, 2nd, and 3rd positions are checked, while 'Noncoding Sites' is unchecked. At the bottom of the dialog box are 'Help', 'Cancel', and 'OK' buttons.

You will get a distance matrix for your nucleotide sequences:

MX: Pairwise Distances (ginsulinkg.mas)

File Display Average Caption Help

	1	2	3	4	5
1. M57671.1 Octodon degus insulin mRNA complete cds					
2. M61153.1 Oryctolagus cuniculus insulin mRNA partial cds	0.3693				
3. U03610.1 Oryctolagus cuniculus New Zealand White insulin mRNA complete cds	0.3684	0.0203			
4. JF909299.1 Homo sapiens insulin (INS) mRNA partial cds	0.3471	0.1669	0.1825		
5. AJ237750.1 Danio rerio mRNA for insulin B and A chains	0.6409	0.7615	0.6956	0.6614	

[1,1] (M57671.1 Octodon degus insulin mRNA complete cds-M57671.1 Octodon degus insulin mRNA complete cds) / Nucleotide: Maximum Composite Likelihood

Search 17:22 12-05-2021 ENG 9

MX: Alignment Explorer (ginsulinkg.meg)

Data Edit Search Alignment Web Sequencer Display Help

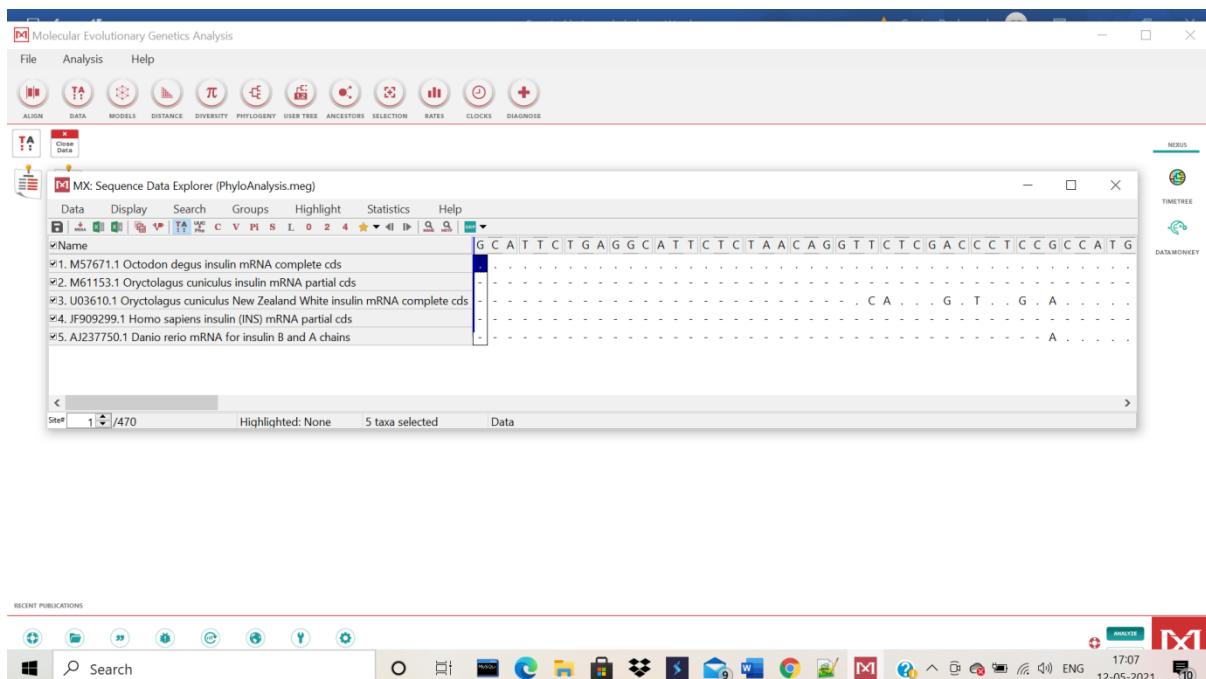
DNA Sequences Translated Protein Sequences

Species/Abby

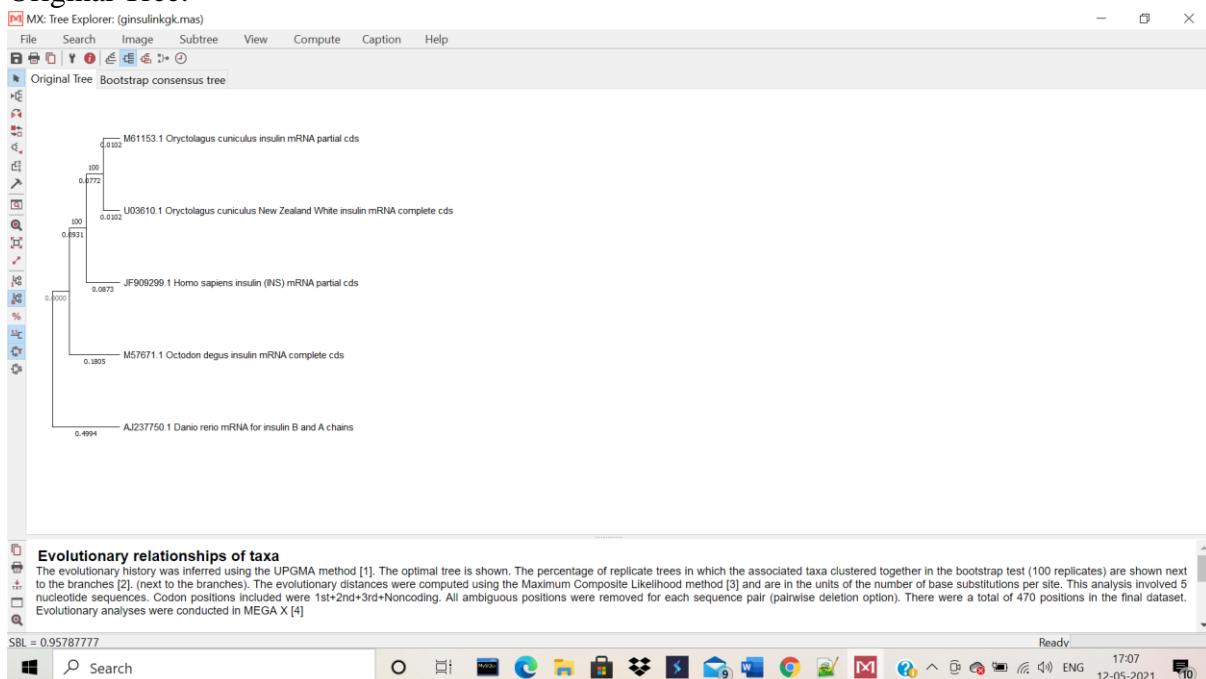
1. M57671.1 Octodon degus insulin mRNA complete cds  
2. M61153.1 Oryctolagus cuniculus insulin mRNA partial cds  
3. U03610.1 Oryctolagus cuniculus New Zealand White insulin mRNA complete cds  
4. JF909299.1 Homo sapiens insulin (INS) mRNA partial cds  
5. AJ237750.1 Danio rerio mRNA for insulin B and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

Site # 1 Selected genetic code: Standard

Search 17:07 12-05-2021 ENG 10



UPGMA:  
Original Tree:

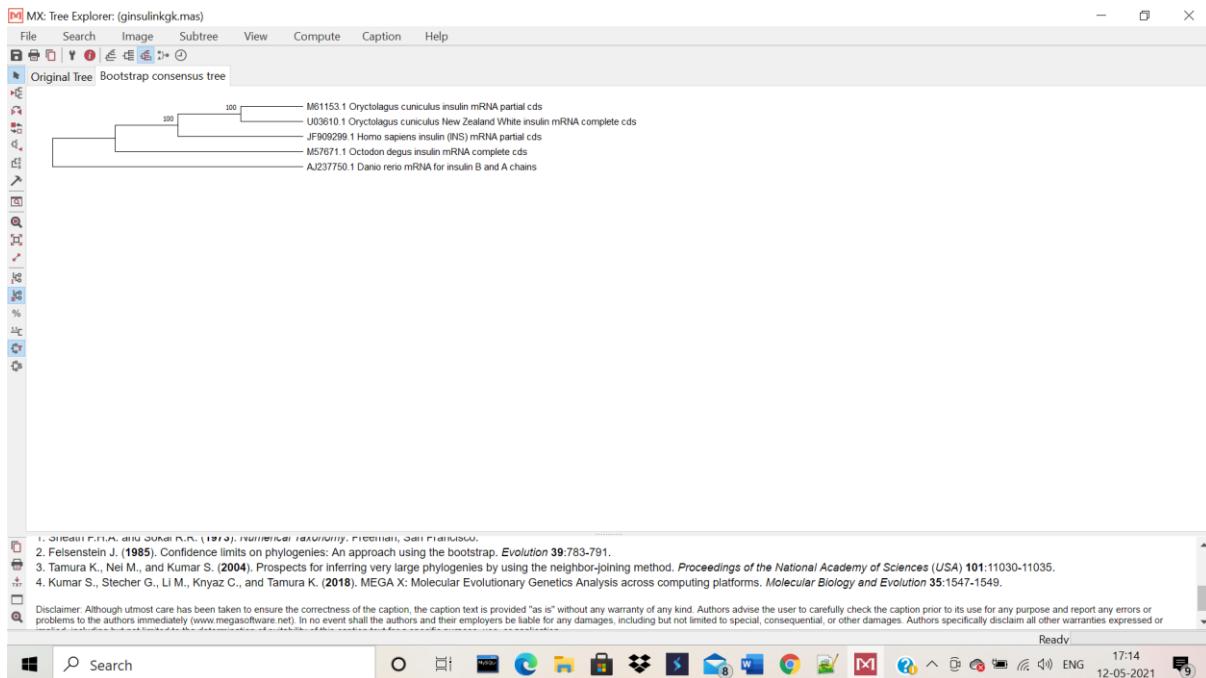


Caption:

### **Evolutionary relationships of taxa**

The evolutionary history was inferred using the UPGMA method [1]. The optimal tree is shown. 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]. (next to the branches). 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].

## Bootstrap Tree:

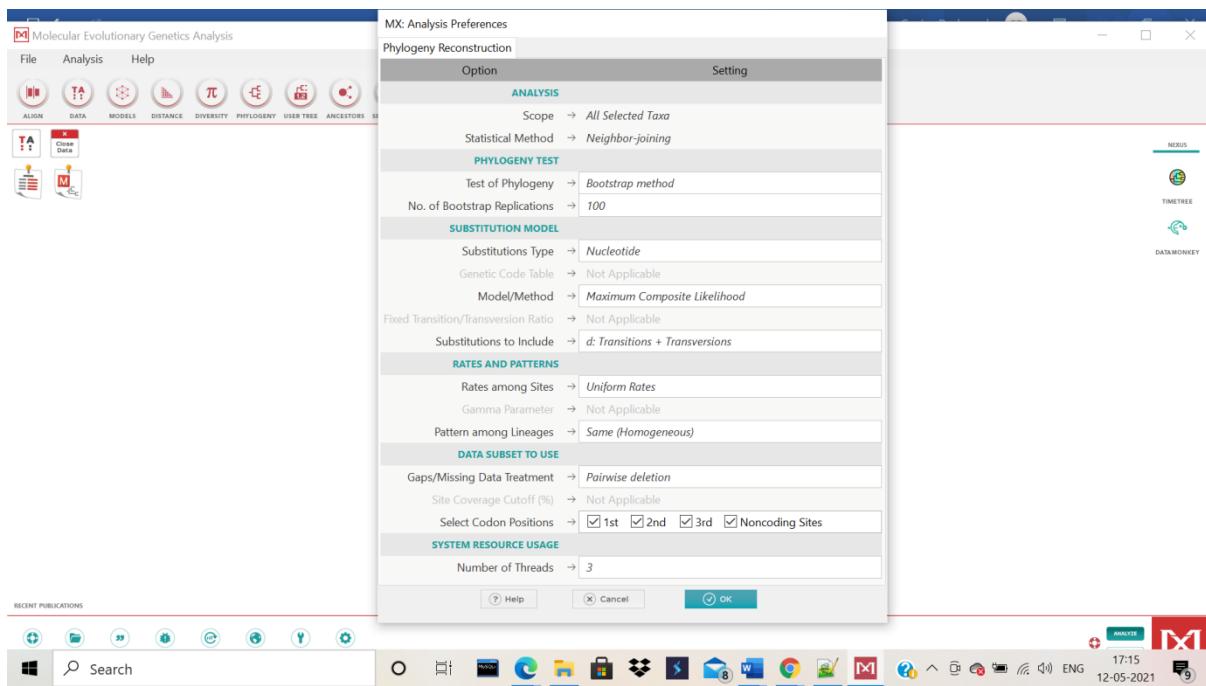


Caption:

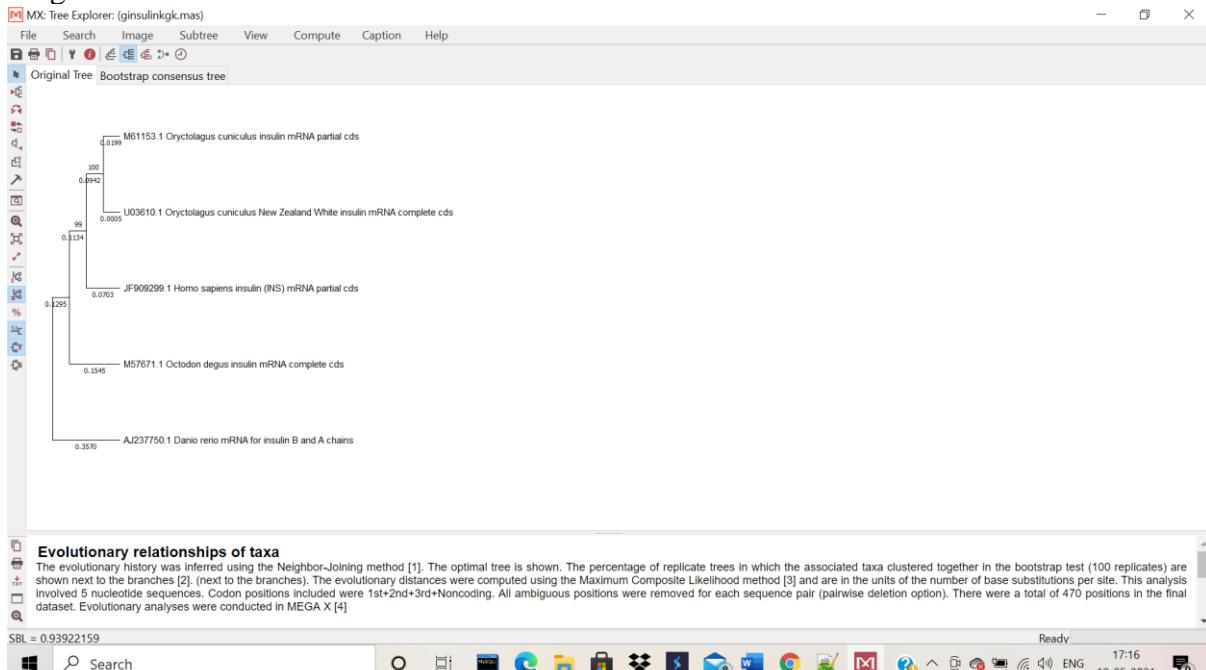
### **Evolutionary relationships of taxa**

The evolutionary history was inferred using the UPGMA 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]

Neighbor Joining:



### Original Tree:



Caption:

### Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. 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]. (next to the branches). 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].

### Bootstrap Tree: