SCHOOL OF COMPUTING
Faculty of Engineering

Semester II 2024/2025

Subject   : Bioinformatics I (SECB2103)
Section   : 01 – Dr Haslina Hashim
Topic   : Lab 07 – Phylogenetic
Name   : Welson Woong Lu Bin (A23CS0196)

1) Determine whether human and chimpanzee mitochondrial DNA sequences have equal evolutionary rates between lineages. To do this, use Tajima's relative rate test as implemented in MEGA

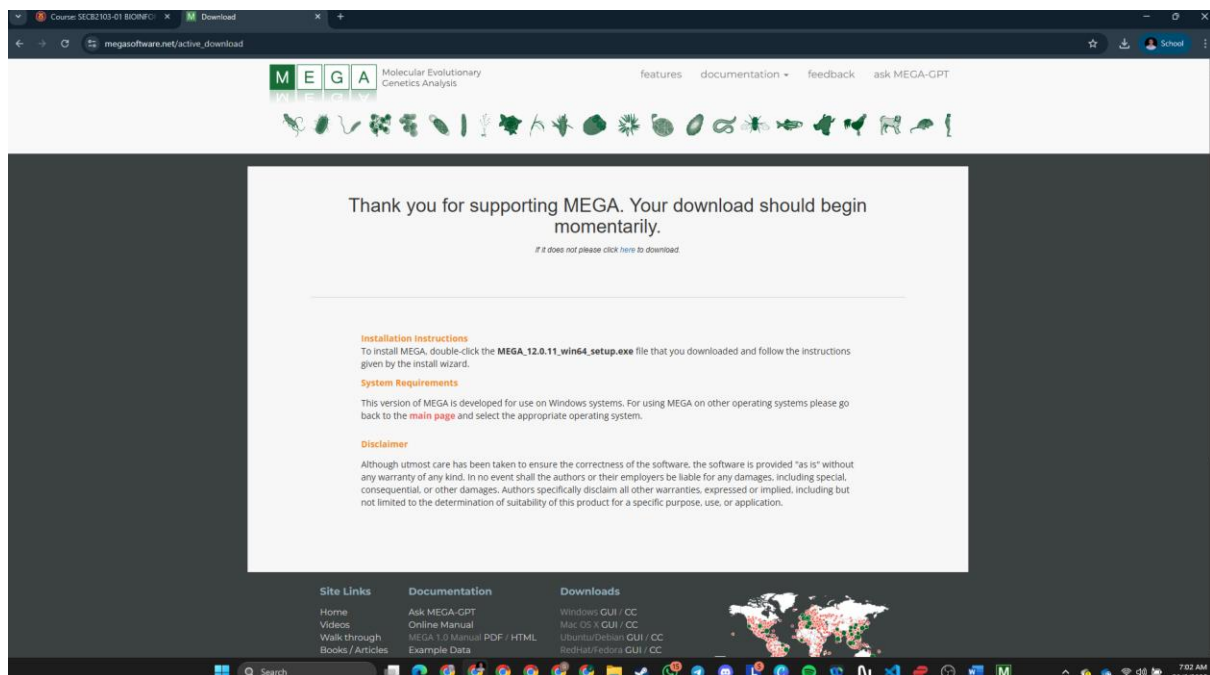   i) Obtain MEGA software Perform pairwise alignment at the NCBI BLAST website.

Figure 1.1: Download of MEGA

ii) Obtain mitochondrial DNA sequences from human, chimpanzee, bonobo, orangutan, gorilla, and gibbon
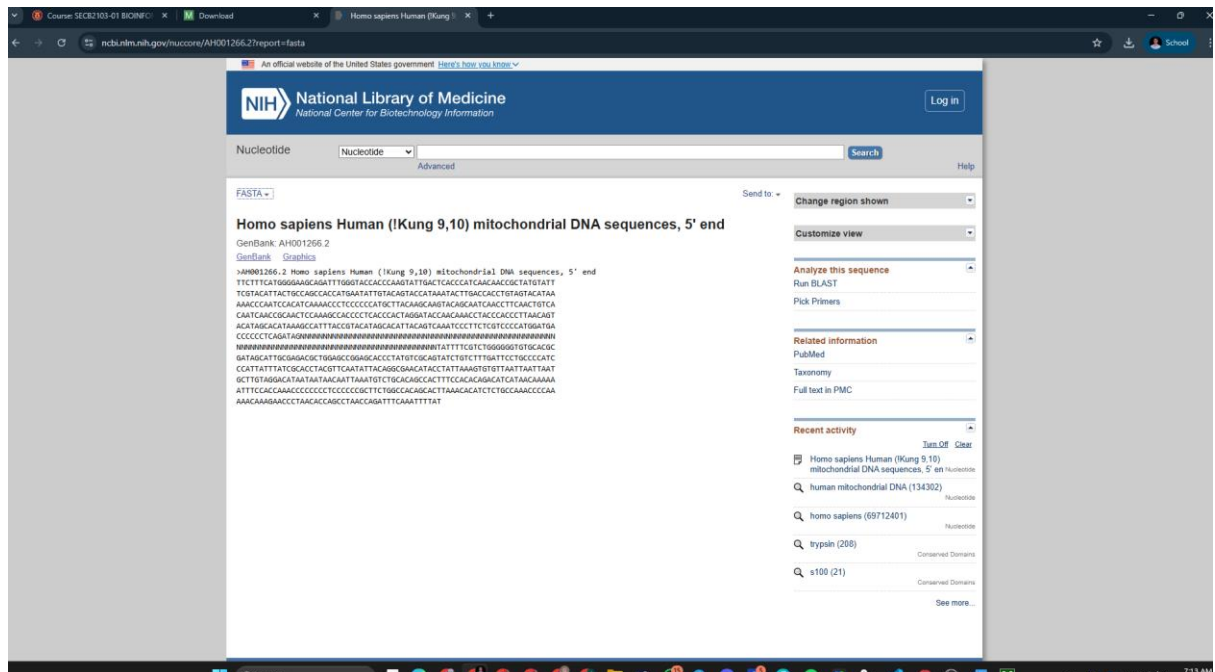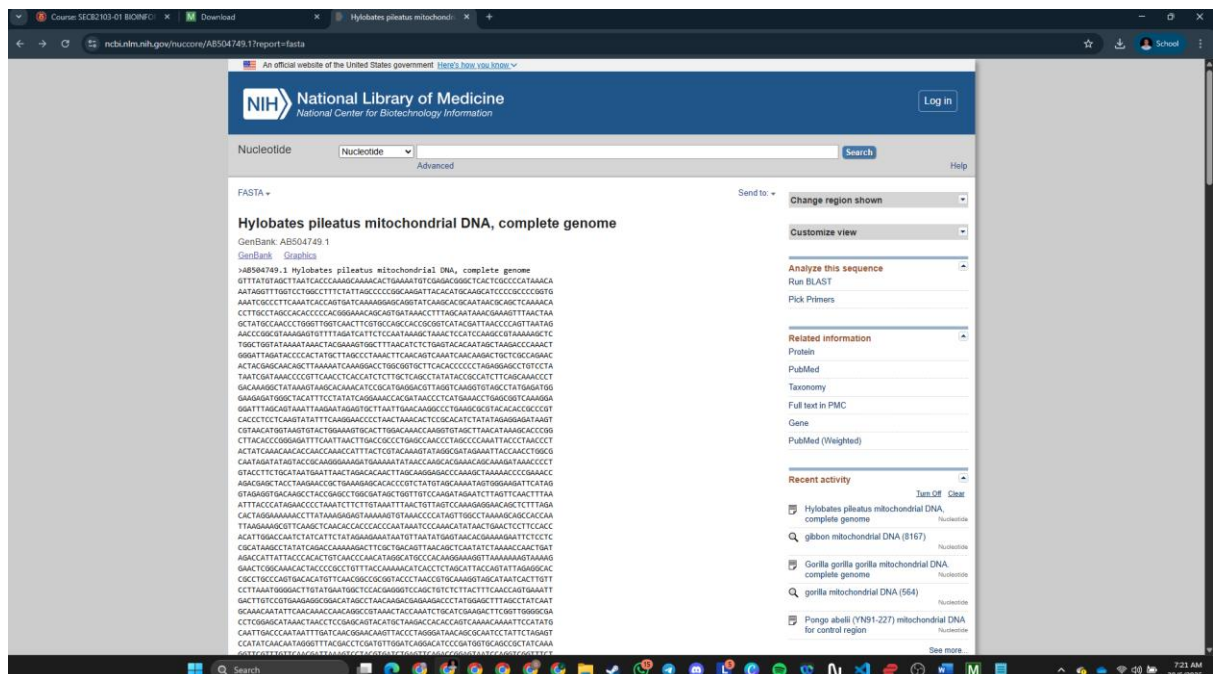

Figure 1.2: Download of Human FASTA
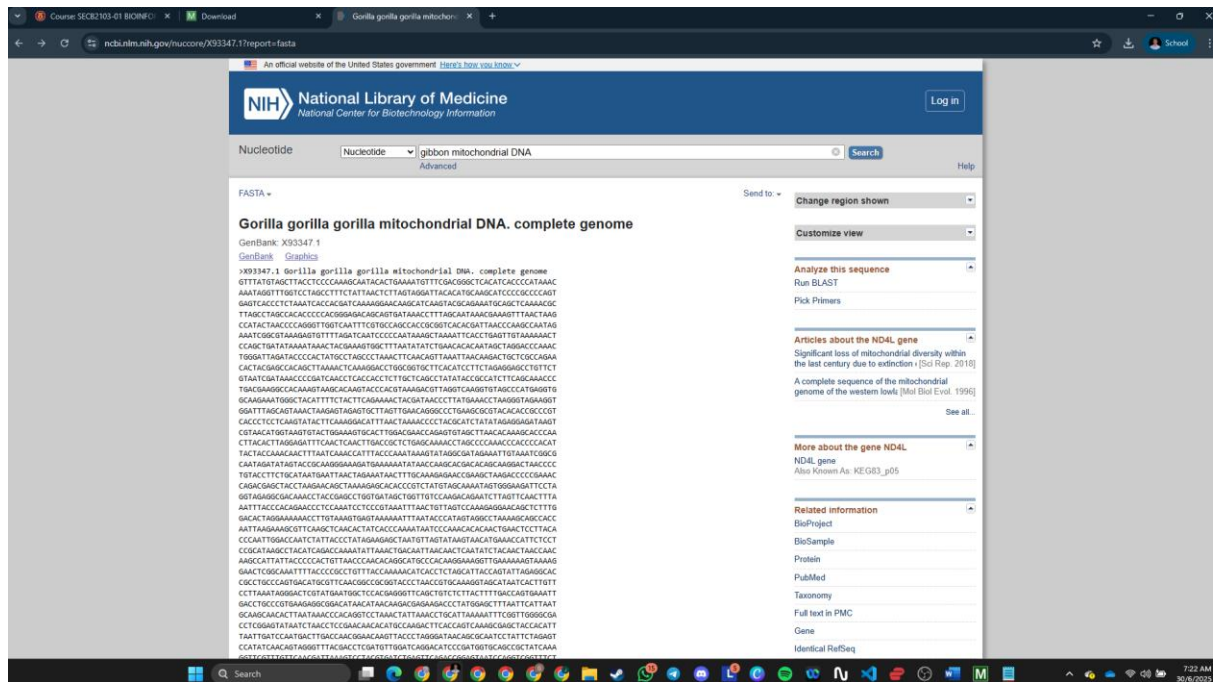

Figure 1.3: Download of Gibbon FASTA

Figure 1.4: Download of Gorilla FASTA

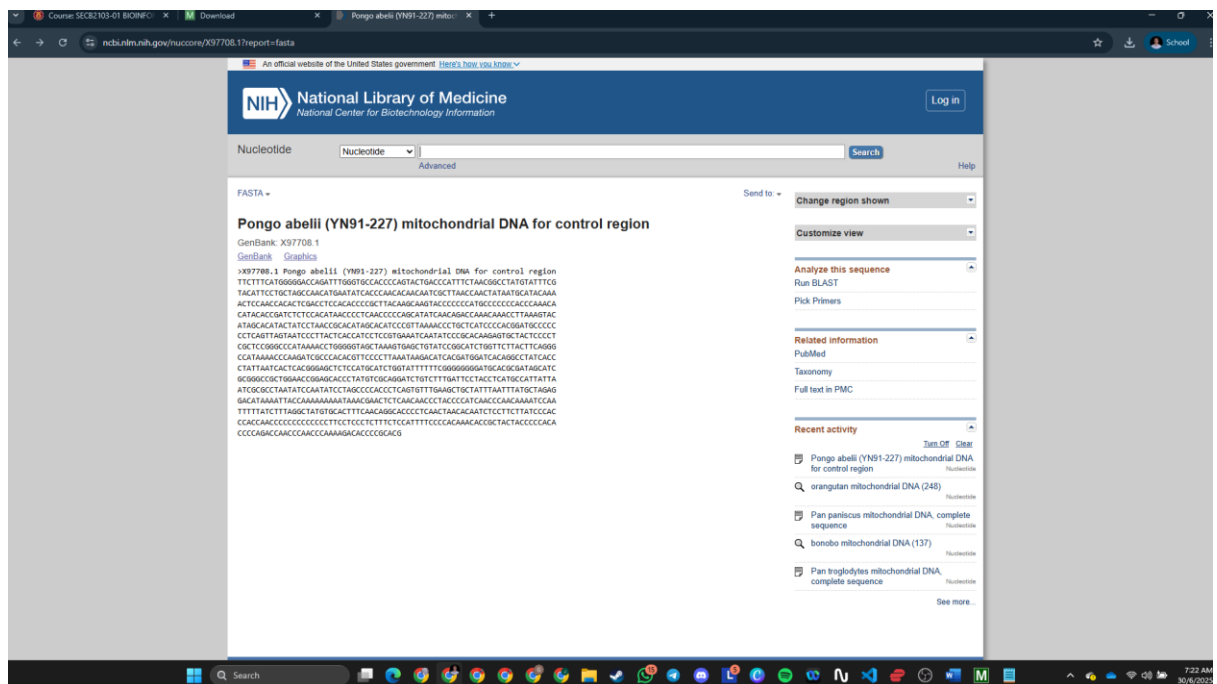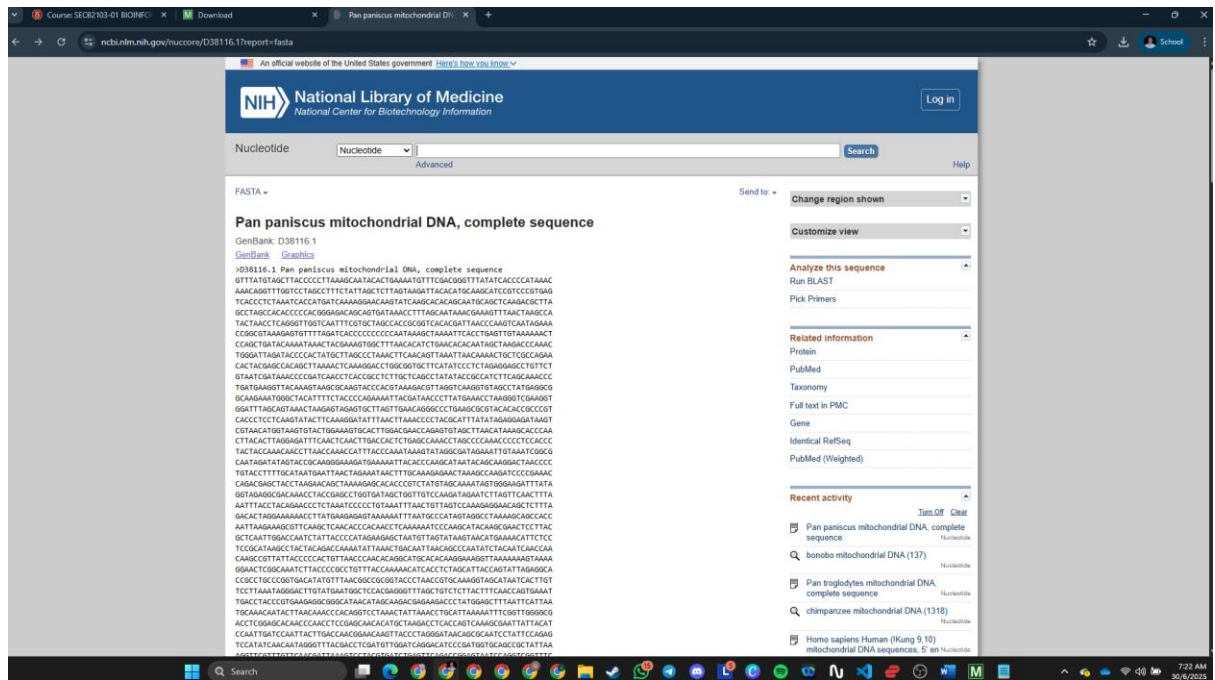*Figure 1.5: Download of Orang Utan FASTA*

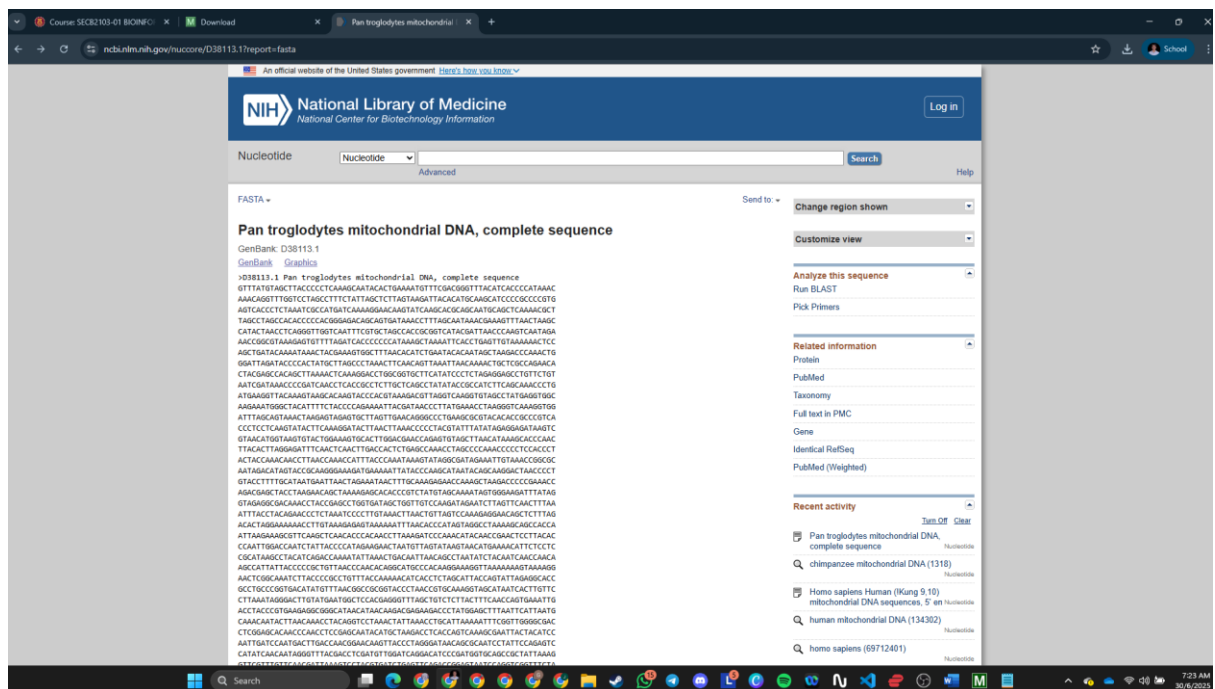*Figure 1.6: Download of Bonobo FASTA*

Figure 1.7: Download of Chimpanzee FASTA

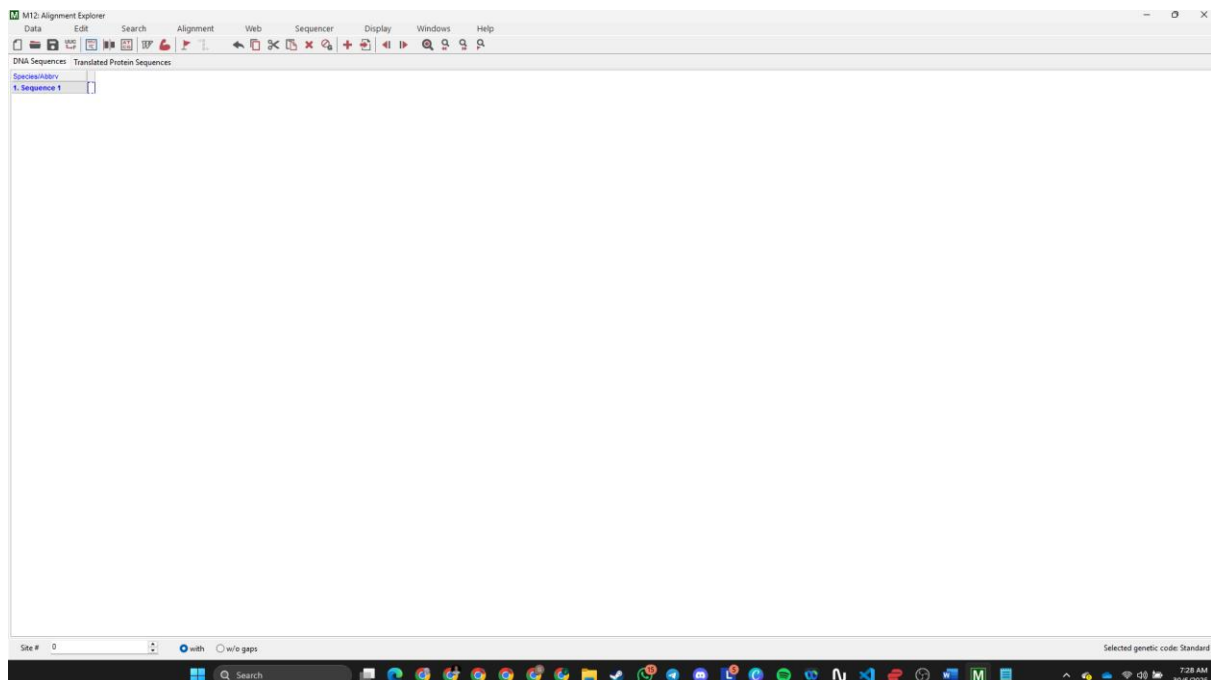iii) Apply Tajima's test using an appropriate outgroup. Is the probability value significant (<0.05)?
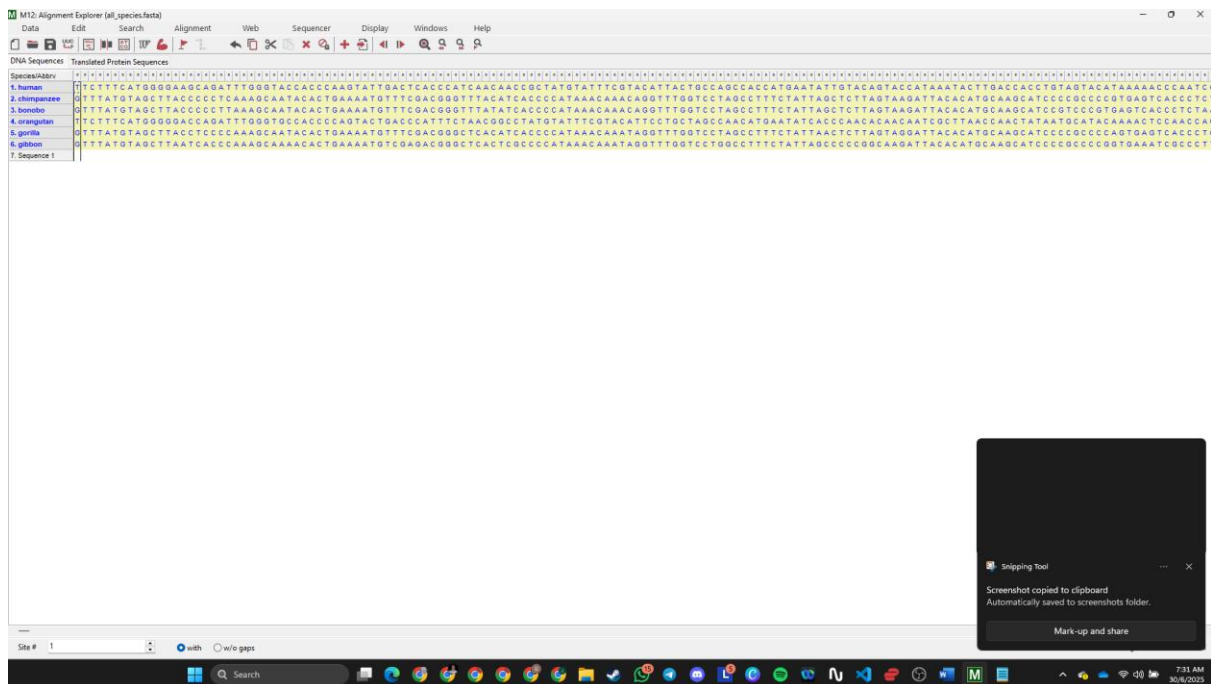

Figure 1.8: Create New at MEGA (Align DNA)
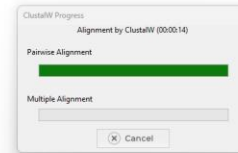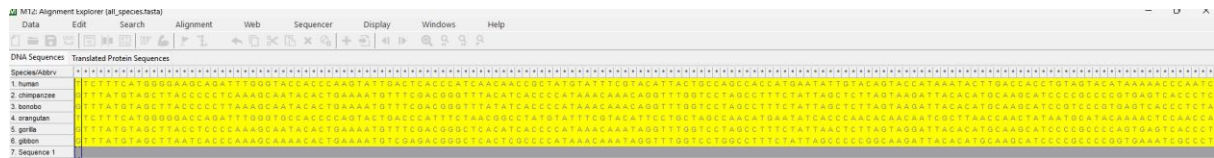

Figure 1.9: Import All FASTA
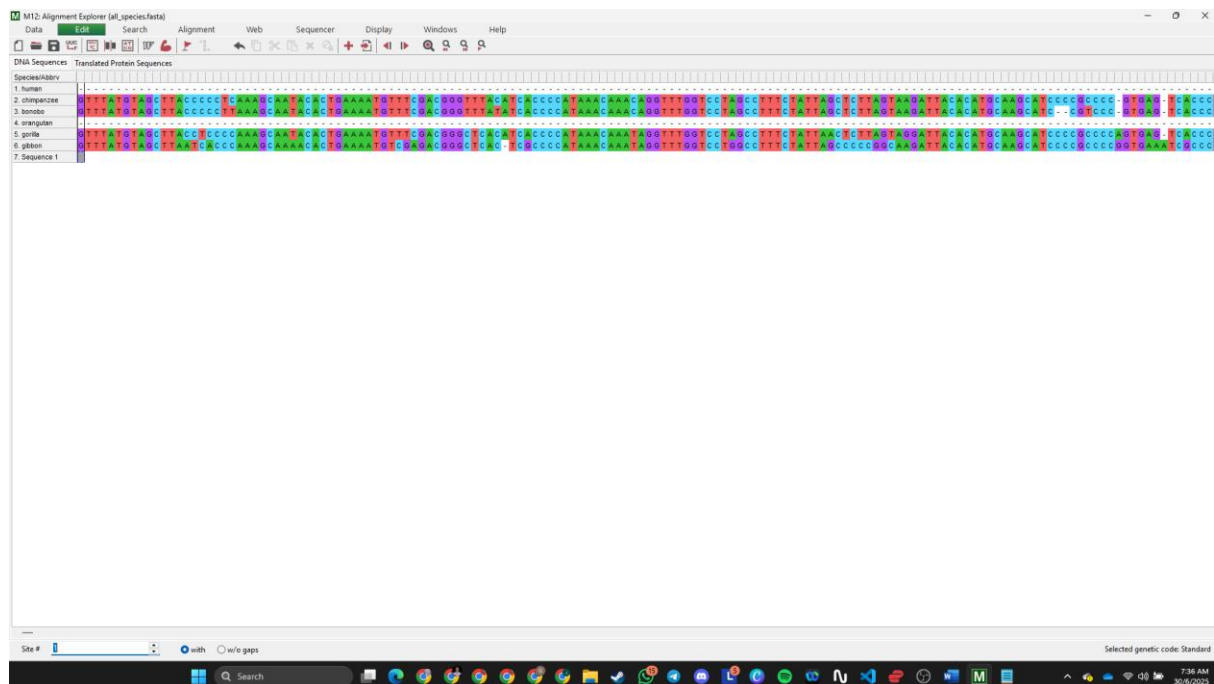
Figure 1.10: Align using ClustalW
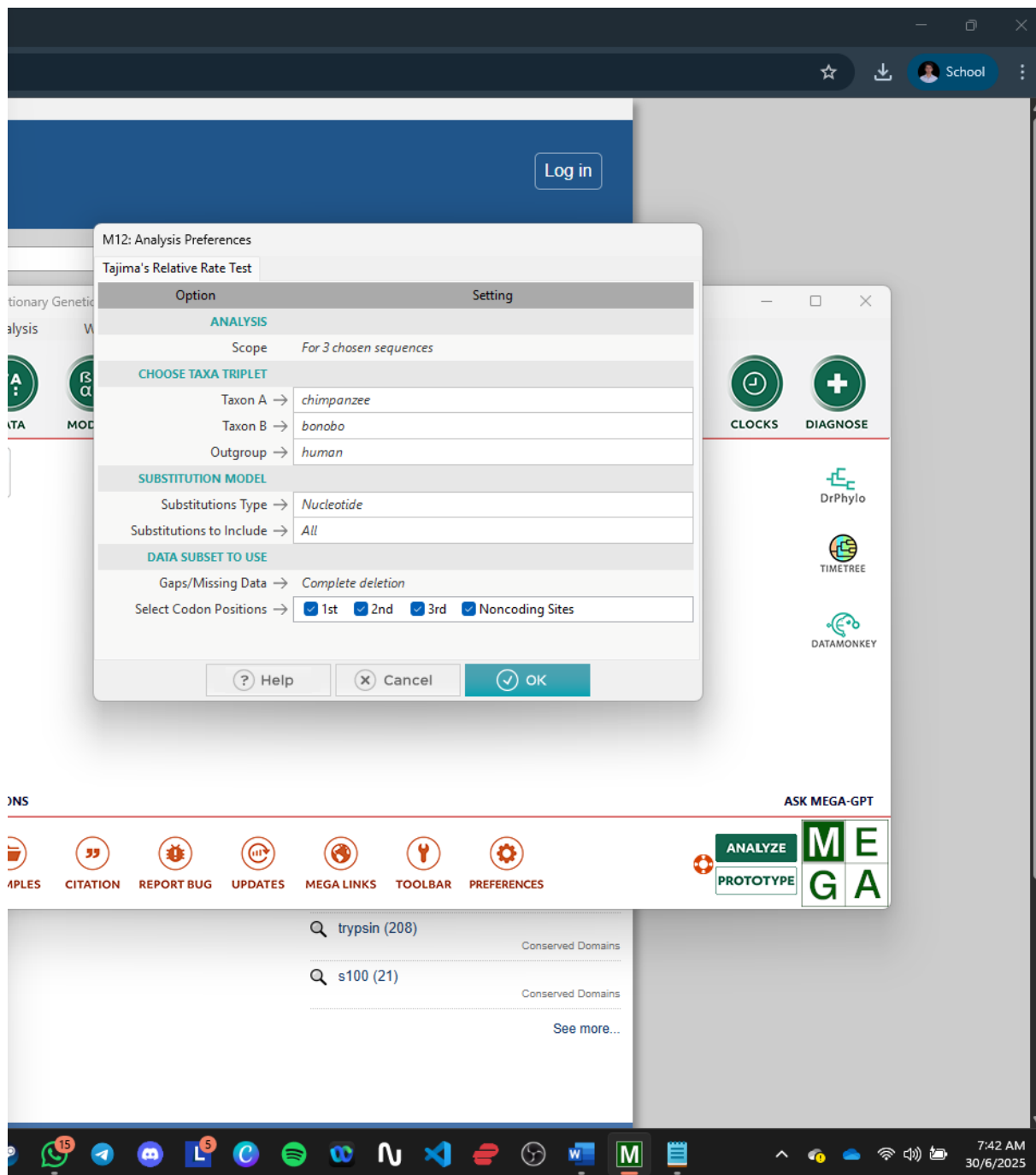

Figure 1.11: Result of ClustalW

Figure 1.12: Tajima's Test

Table. Results from the Tajima's clock test for 3 Sequences.

Configuration	Count
Identical sites in all three sequences	553
Divergent sites in all three sequences	9
Unique differences in Sequence A	38
Unique differences in Sequence B	45
Unique differences in Sequence C	60

NOTE.-- The comparative analysis investigated the equality of evolutionary rates between sequences **A** (*chimpanzee*) and **B** (*bonobo*), utilizing sequence **C** (*human*) as an outgroup in Tajima's relative rate test [1]. The χ^2 test yielded a value of 0.59 (*P* = 0.44228 with 1 degree[s] of freedom). Conventionally, a P-value less than 0.05 is employed to test the null hypothesis of equal rates between the selected lineages. The analytical procedure encompassed 3 coding nucleotide sequences using 1st, 2nd, 3rd, and non-coding positions. The complete deletion option was applied to eliminate positions containing gaps and missing data resulting in a final data set comprising 705 positions. Evolutionary analyses were conducted in MEGA12 [2].

1. Tajima F. (**1993**). Simple methods for testing molecular clock hypothesis. *Genetics* 135:599-607.
2. Kumar S., Stecher G., Suleski M., Sanderford M., Sharma S., and Tamura K. (**2024**). Molecular Evolutionary Genetics Analysis Version 12 for adaptive and green computing. *Molecular Biology and Evolution* 41:1-9.

Disclaimer: While every effort has been made to ensure the accuracy of the caption, it is provided "as is" without warranty of any kind. Users are advised to thoroughly review the caption before its use for any purpose and report any errors or issues to the authors at www.megasoftware.net. The authors and their employers disclaim any liability for damages, including but not limited to special or consequential damages. Additionally, the authors expressly disclaim all other warranties, whether expressed or implied, including the suitability of the caption text for a specific purpose, use, or application.
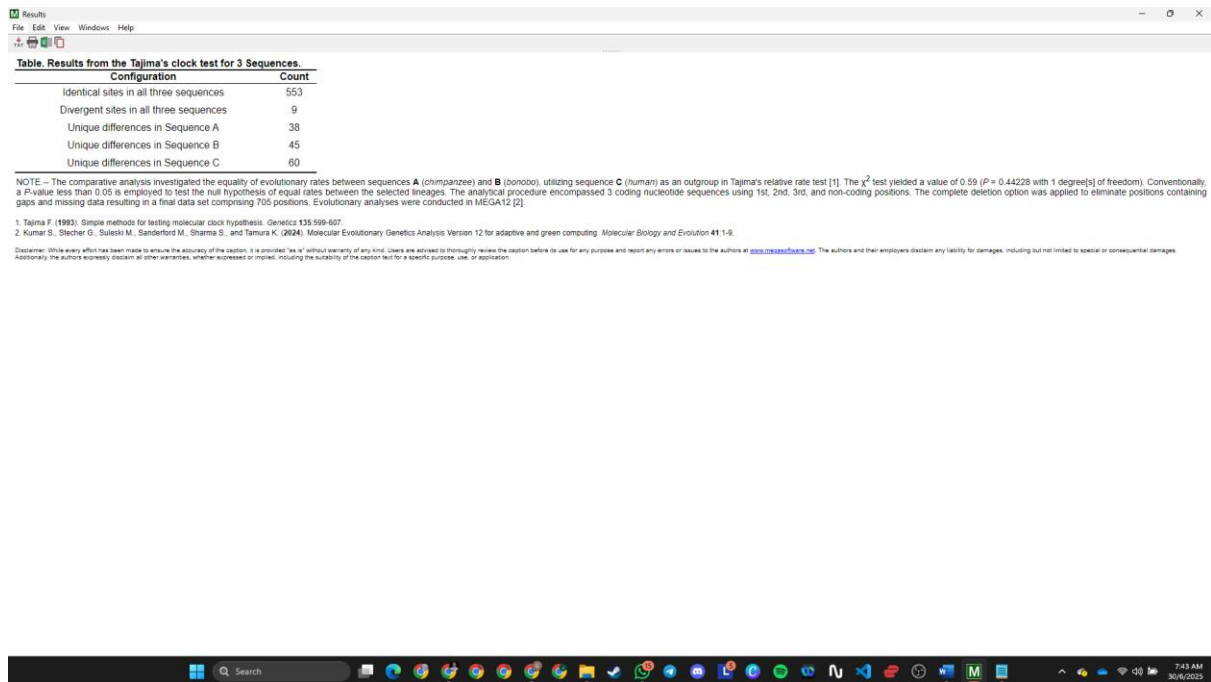
Figure 1.13: Result of Tajima's Test

Based on the result of Tajima's Relative Rate Test, the χ^2 value obtained was 0.59 with a corresponding P-value of 0.44228. Since this P-value is greater than the significance threshold of 0.05, we fail to reject the null hypothesis that assumes equal evolutionary rates between the selected lineages. Therefore, we conclude that there is no statistically significant difference in the evolutionary rates between the mitochondrial DNA sequences of chimpanzee and bonobo when human is used as the outgroup. This suggests that both species are evolving at a relatively similar rate over time.

2) Perform phylogenetic analyses using MEGA software

i)   Go to the conserved domain Database (http://www.ncbi.nlm.nih.gov/cdd) at NCBI. Click the box to set the view to default tracks.

ii) Enter lipocalins (or another family of your choice; you can also begin at Ensembl, HomoloGene, or Pfam).

iii) Select the mFasta format then click "Reformat." The result is a multiple sequence alignment. Copy this into a text editor (such as NotePad++), and simplify the names of the sequences.

iv) Import the file (or paste the sequences) into MEGA as shown in Fig. 7.9. Align the sequences, save in the .mas and .meg formats.

v)   Choose Phylogeny > Construct/Test to create neighbor-joining, maximum likelihood, or other trees.

vi) For each tree you create, read the caption. Try the tree tools (e.g. placing a root, flipping nodes, showing or hiding branch lengths, interconverting display formats).

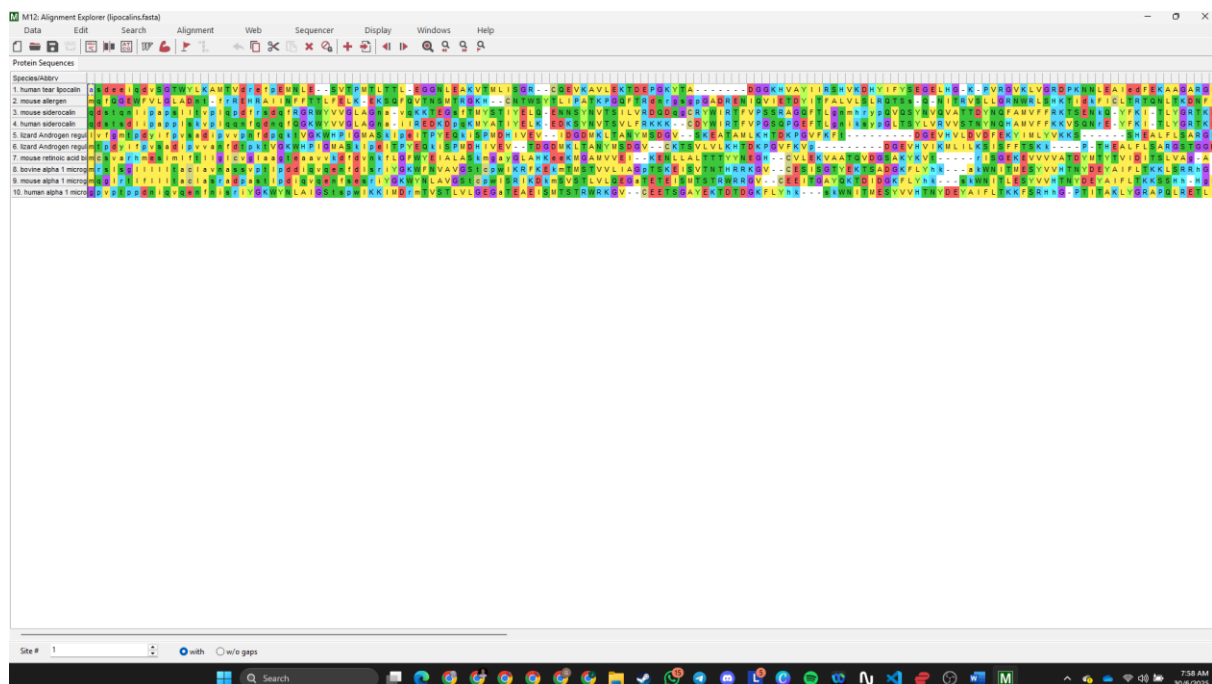vii) Perform bootstrapping. Identify clades having low levels of support. Why does this occur?

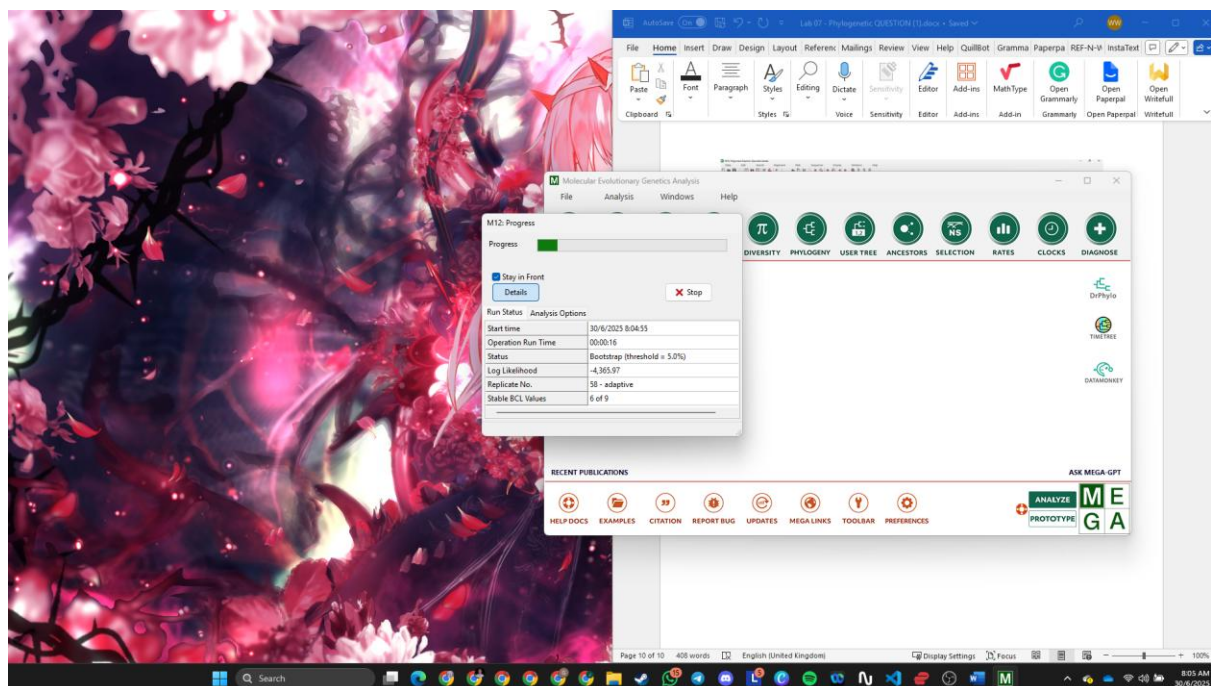*Figure 2.1: Align in mFASTA*

Figure 2.2: Result of ClustalW

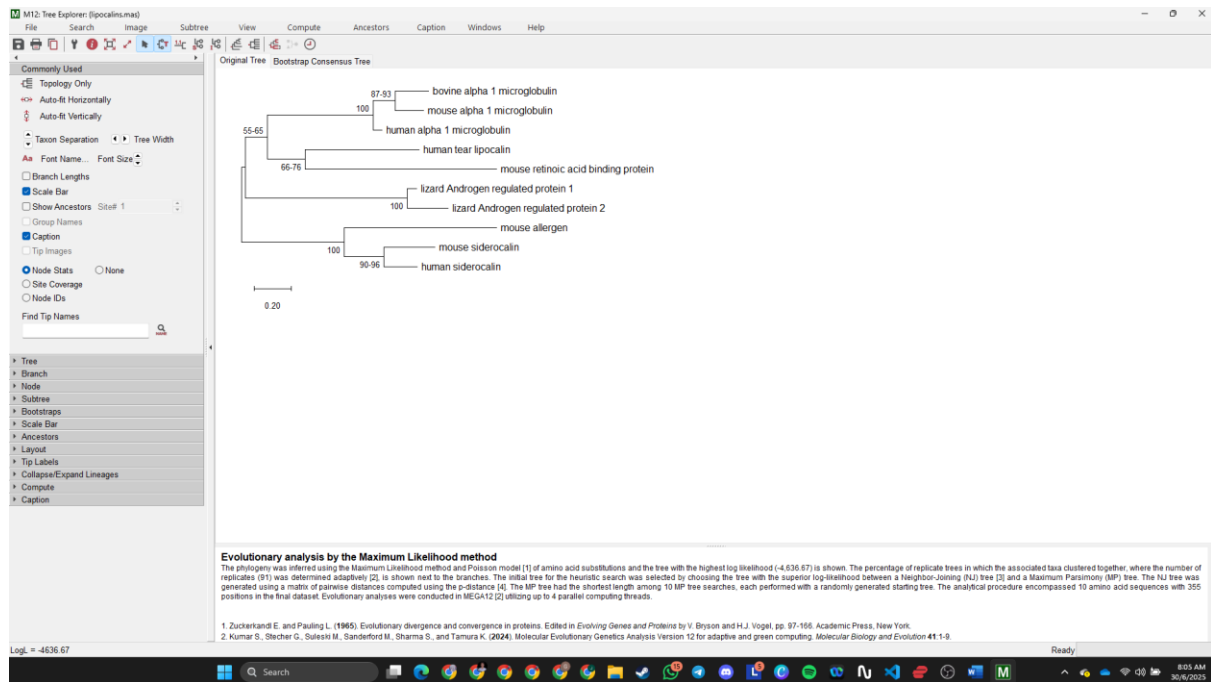*Figure 2.3: Construct Phylogenetic Tree*

Figure 2.3: Result of Phylogenetic Tree

The phylogenetic tree constructed using the Maximum Likelihood method with 1000 bootstrap replications reveals the evolutionary relationships among 10 lipocalin protein sequences from various species. Strong bootstrap support was observed in several clades, such as the grouping of lizard androgen-regulated proteins (100%), mouse allergen with this group (100%), and the close relationship between mouse and human siderocalin (96–100%). These high-confidence branches suggest robust evolutionary relatedness. In contrast, some internal nodes, such as the clade connecting human tear lipocalin and retinoic acid binding protein (66.7%) and the node connecting alpha 1 microglobulins with this group (55.6%), showed moderate to weak support. This may be due to conserved sequences or limited informative sites. Overall, the analysis supports the evolutionary conservation and divergence patterns within the lipocalin family and demonstrates the usefulness of phylogenetic analysis and bootstrapping in assessing clade confidence.