BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024,
Tamil Nadu, India.

Programme : M.Sc., Biomedical Science

Course Title : Bioinformatics
Course Code : BM35S1BI

Unit - III

TOPIC: Method for Construction and Representation of Phylogenetic tree using MEGA Software.

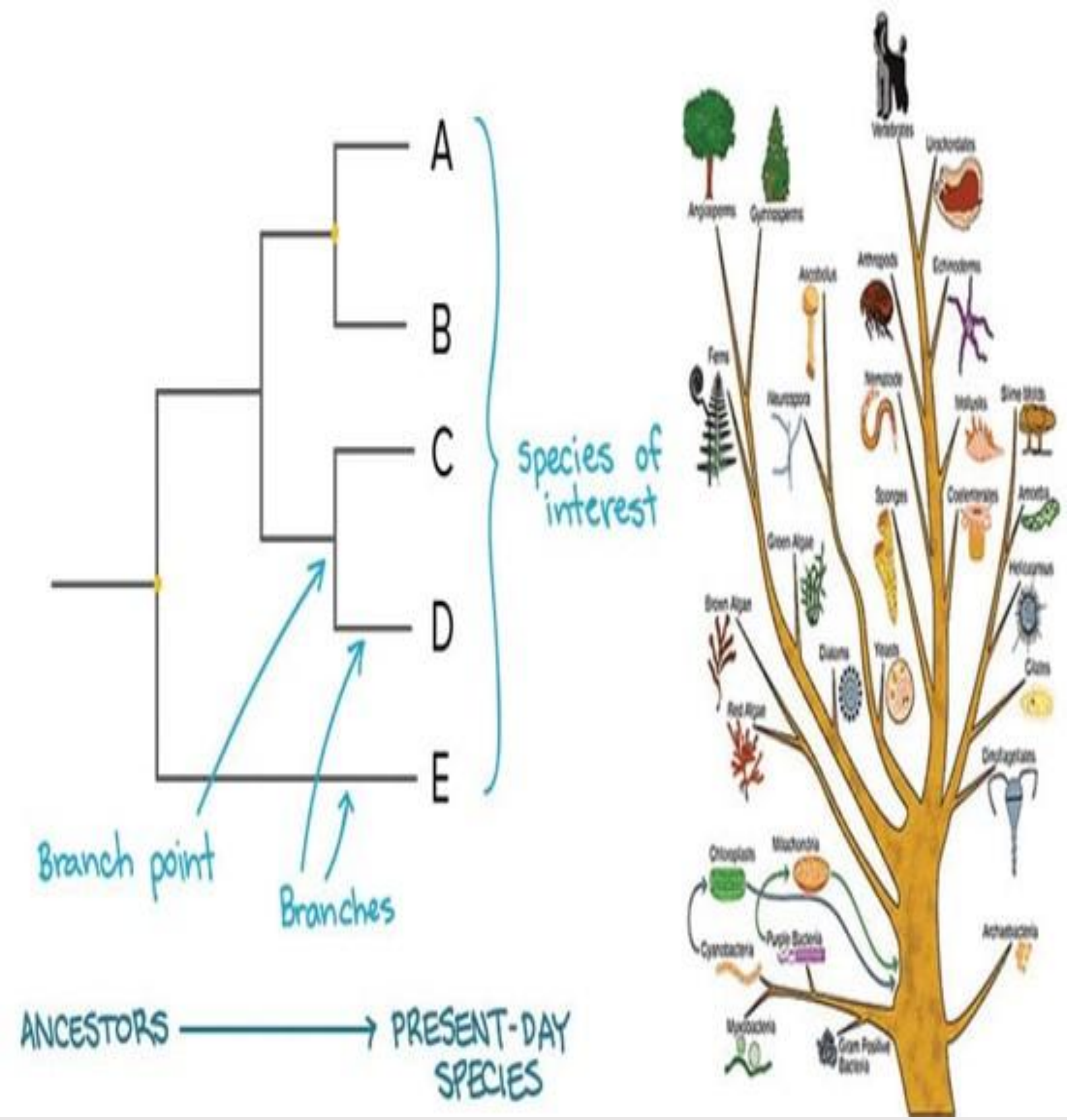
Dr. P. JEGANATHAN
Guest Lecturer
Department of Biomedical Science

Method for construction and representation of phylogenetic tree using MEGA software

MEGA (Molecular Evolutionary Genetics Analysis)

- It is a computer software for conducting statistical analysis of molecular evolution and for constructing phylogenetic tree.
- MEGA 5 is an integrated program that carries out all four steps in a single environment, with a single user interface eliminating the need for interconverting the formats.
- MEGA 5 is sufficiently flexible to permit using other programs for particular steps.

- In MEGA software phylogenetic tree can be constructed using four steps they are :
    1 .Acquiring the sequences
    2 .Aligning the sequences
    3 .Estimate the tree
    4 .Present the tree



1. Acquiring the sequences :

- If the investigator is interested in a particular gene or protein that has been the subject of investigation and wishes to determine the relationship of that gene or protein to its homologs.
- If the sequences are not actually descended from a common ancestor , then it will be meaningless and may be quite misleading.
- MEGA 5 opens its own browser window to show a nucleotide BLAST page from National Center for Biotechnology Information (NCBI).
- There is a set of five tabs near the top of that page (blastn, blastp, blastx, tblastn and tblastx).
- If your sequence is that of a protein click the blastp tab to show the standard protein **BLAST** page.

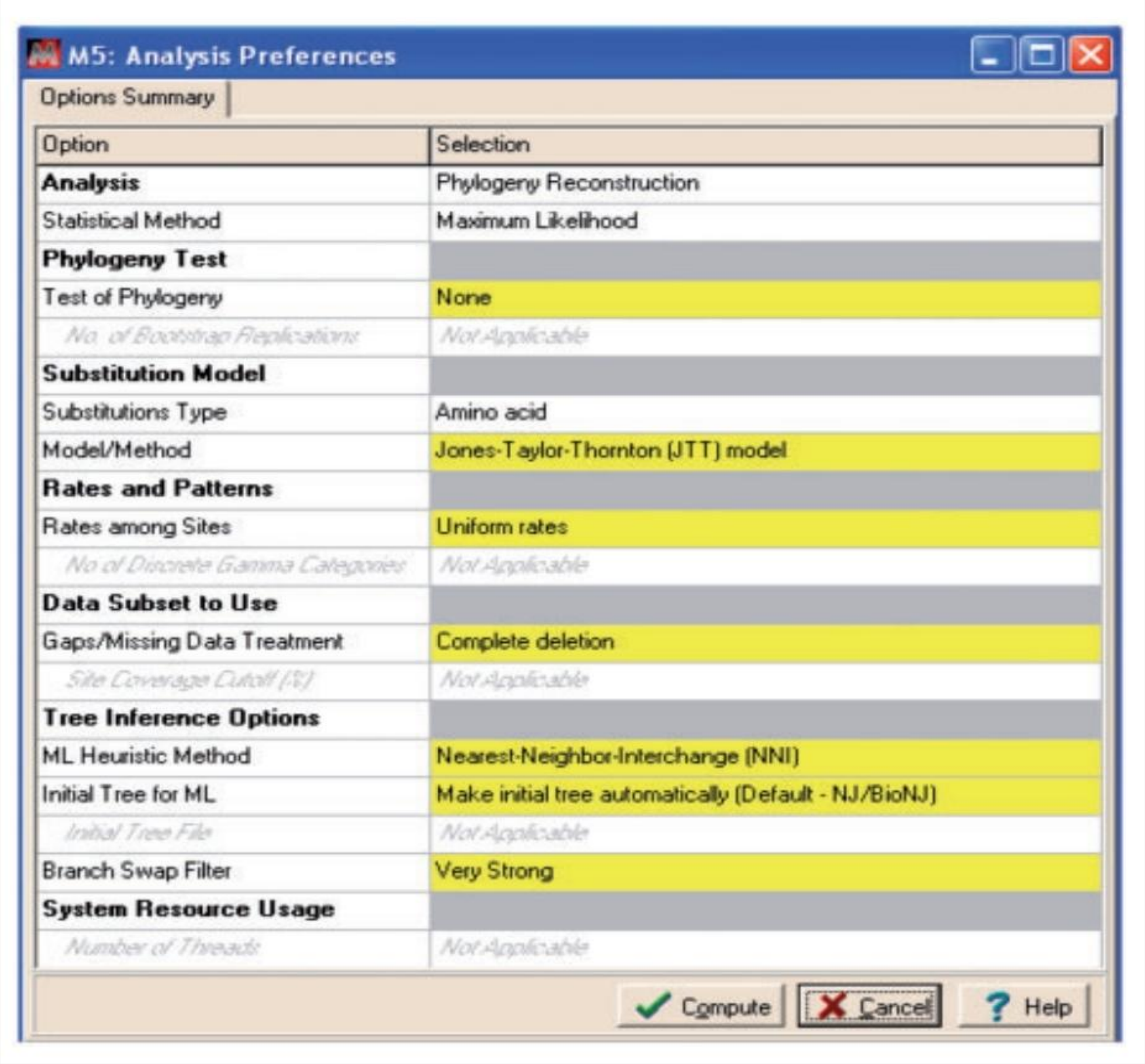
2. Aligning the sequences :

- The alignment explorer shows a name for each sequence at the left, followed by the sequence with coloured residues.
- If the name is very long, that name would eventually appear on the tree, and long names are generally undesirable. But we could edit the names only at this stage . Simply double click each name and change it to something more suitable.
- If your sequence is DNA you will see two tabs: **DNA Sequences** and **Translated protein sequences**.
- The DNA tab is chosen by default. Click the translated protein sequences tab to see the corresponding protein sequence.

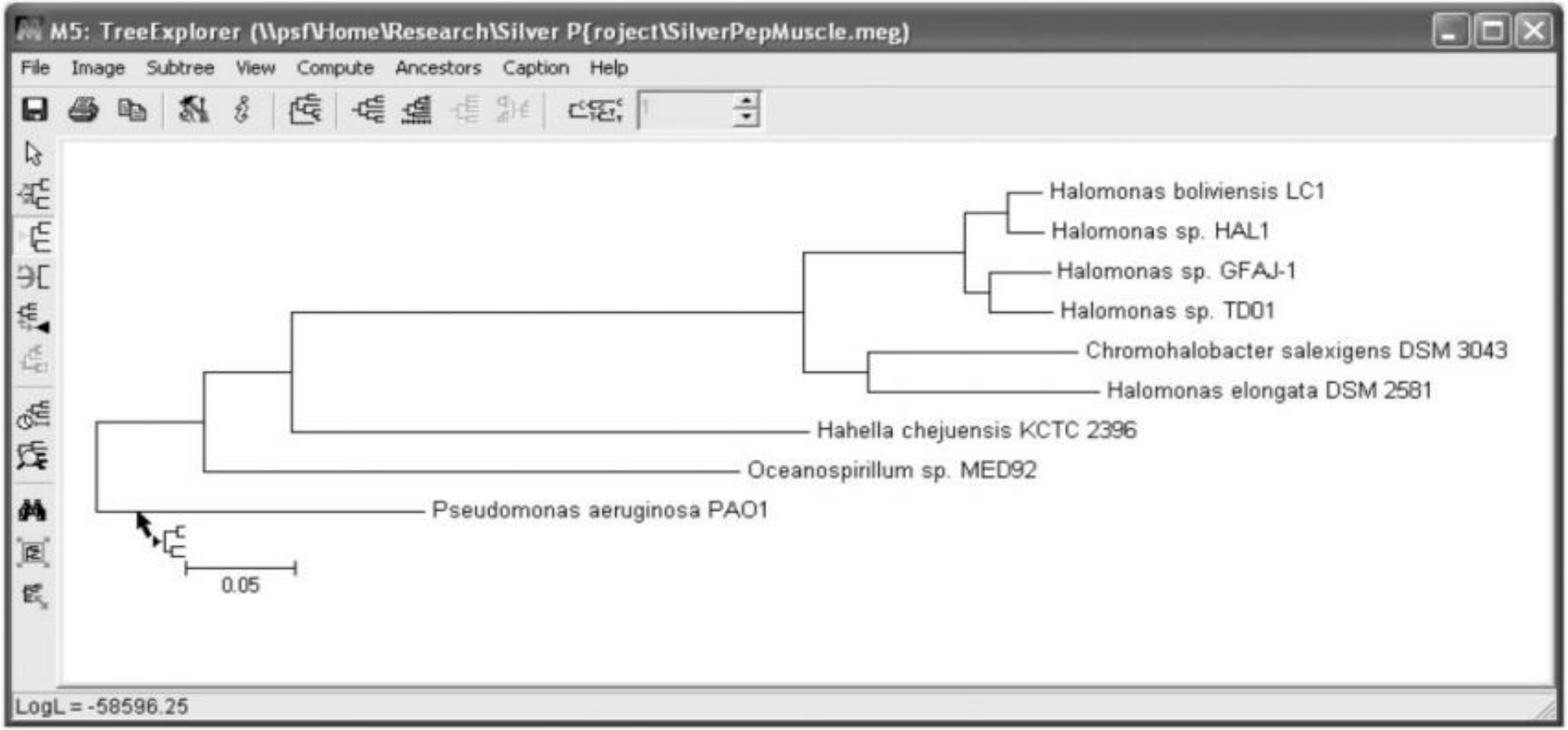
3. Estimate the tree :

- There are several methods for estimating phylogenetic trees (Neighbor joining, Maximum parsimony, Maximum Likelihood), but this is about Maximum Likelihood (ML).
- ML uses a variety of substitution models to correct for multiple at the same site during the evolutionary history of the sequences.
- When a complete window appears that lists the models in order of preference. Note the preferred model, then estimate the tree using that model.

- The yellow areas are parameters that we can modify:
  1. **Model/ Method**
  2. **Rates among sites**
  3. **Gap / missing data treatment.**



Option	Selection
Analysis	Phylogeny Reconstruction
Statistical Method	Maximum Likelihood
Phylogeny Test	
Test of Phylogeny	None
No. of Bootstrap Replications	*Not Applicable*
Substitution Model	
Substitutions Type	Amino acid
Model/Method	Jones-Taylor-Thornton (JTT) model
Rates and Patterns	
Rates among Sites	Uniform rates
No of Discrete Gamma Categories	*Not Applicable*
Data Subset to Use	
Gaps/Missing Data Treatment	Complete deletion
Site Coverage Cutoff (%)	*Not Applicable*
Tree Inference Options	
ML Heuristic Method	Nearest-Neighbor-Interchange (NNI)
Initial Tree for ML	Make initial tree automatically (Default - NJ/BioNJ)
Initial Tree File	*Not Applicable*
Branch Swap Filter	Very Strong
System Resource Usage	
Number of Threads	*Not Applicable*



```
LogL = -58596.25
```

When the preferences are set, click the compute button to compute the tree. Eventually , a tree explorer window will open that displays the tree.

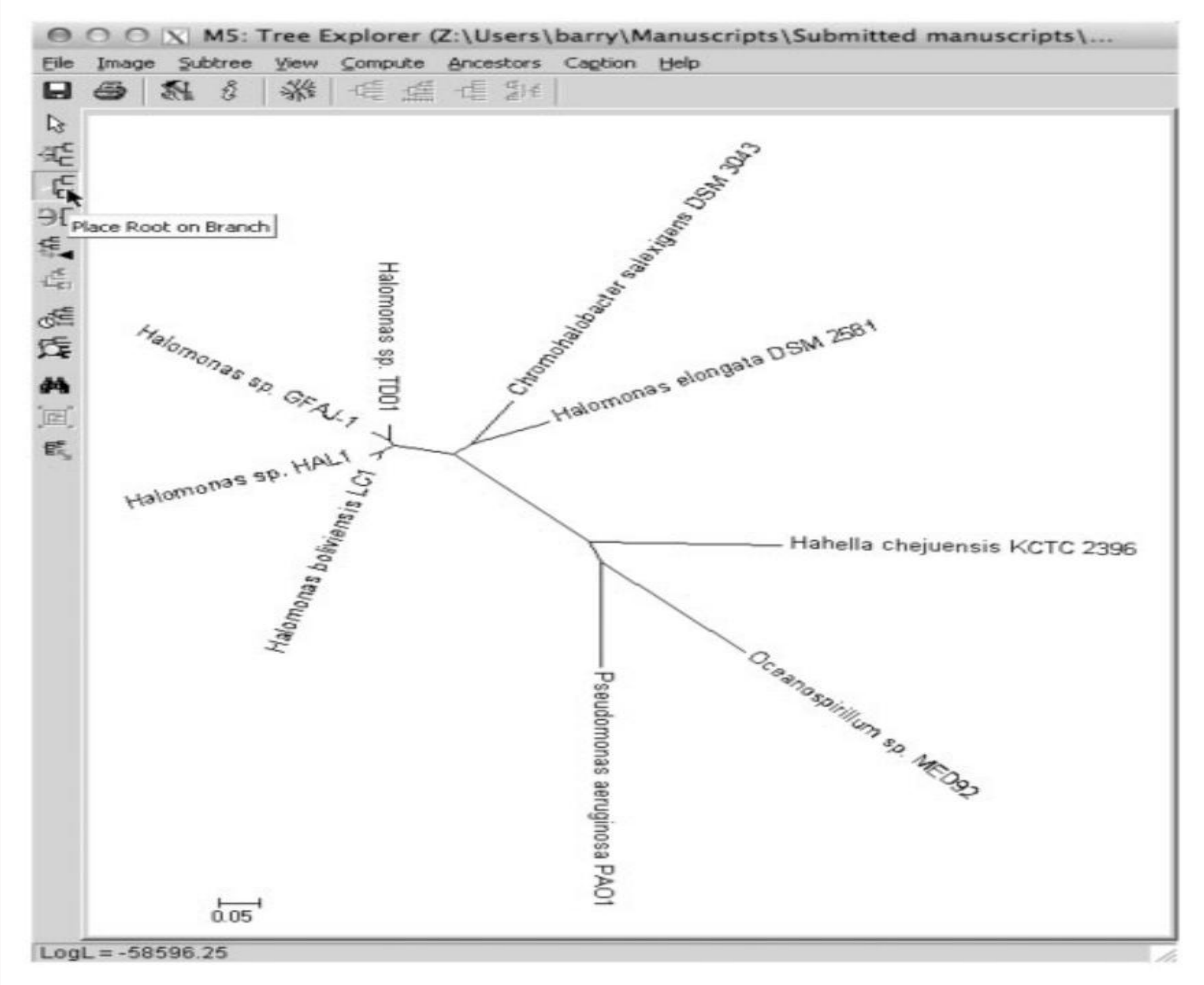
4. Present the tree :

- A phylogenetic tree consists of external nodes that represent the actual sequences that exist today.
- Internal nodes that represent hypothetical ancestors, and branches that connects nodes to each other.
- The length of the branches represent the amount of change that is estimated to have occurred between a pair of nodes.
- The trees can be presented in two forms , they are:

1. Unrooted tree
2. Rooted tree

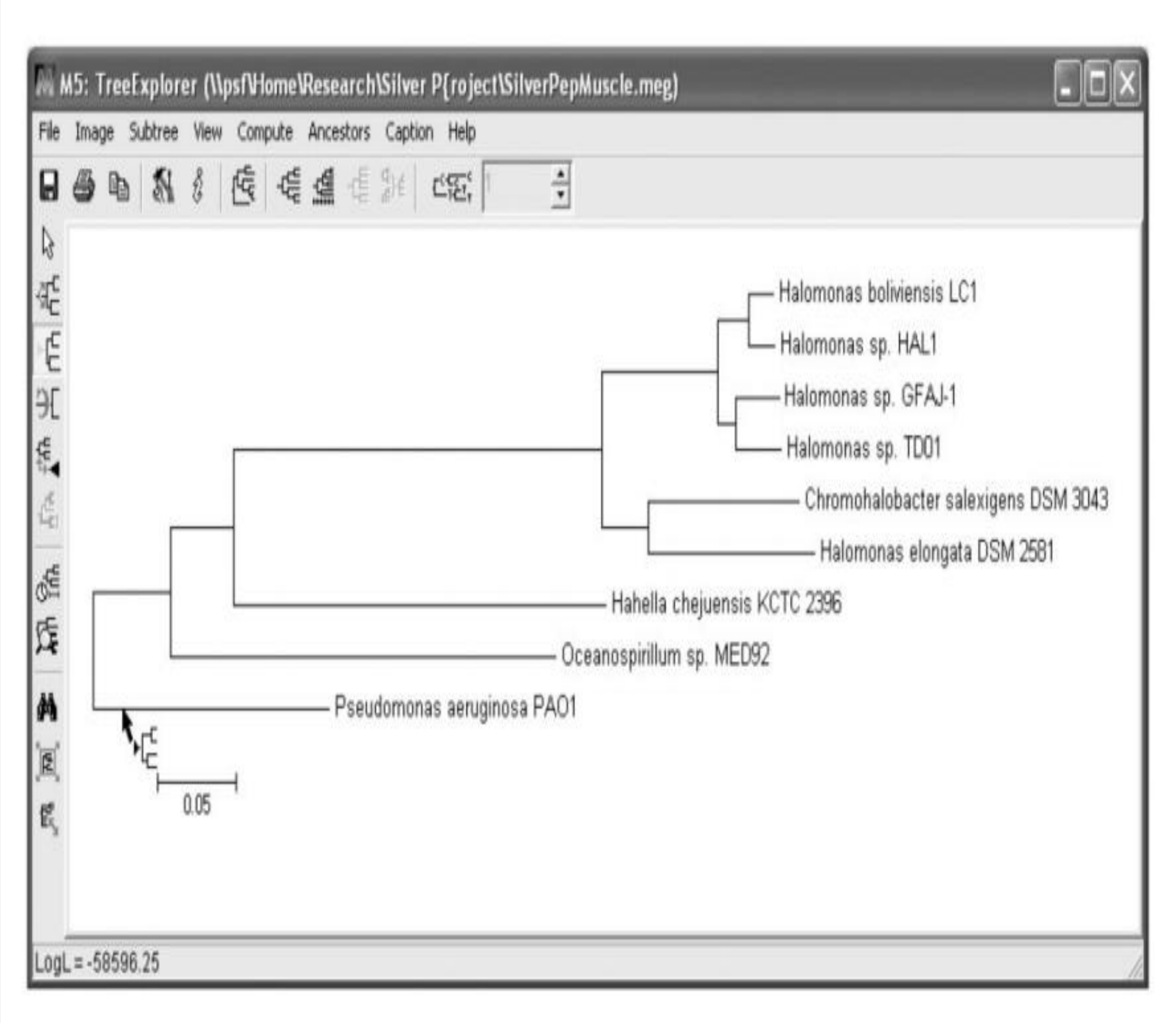
UNROOTED TREE

- The radiation or unrooted is a better way to draw because it does not allow the viewer to assume a root that is known.
- To avoid the unjustified implication of directionality , it is important to specify in the figure that the tree is unrooted.



Rooted Tree

- A rooted tree provides direction to the evolutionary process, with the order of descent from the root towards the tip.



LogL = -58596.25

Reference

Hall, B.G., 2013. Building phylogenetic trees from molecular data with MEGA. *Molecular biology and evolution*, *30*(5), pp.1229-1235.

THANK YOU