



www.esaunggul.ac.id

PENGANTAR BIOINFORMATIKA

IBT 431

By Seprianto S.Pi, M.Si

Pertemuan 10

Analisis Filogenetik

Sasaran Perkuliahan

- Mahasiswa mampu menganalisis dan membuat pohon Filogenetik dari hasil sekuen yang didapatkan
- Mahasiswa mampu melakukan analisis kekerabatan sampai tingkat spesies
- Mampu menginterpretasikan nilai Bootstrap dan Neighbour Joining

Phylogenetic analyses

Phylogenetics

- Study of evolutionary relationship
- Phylogenetic tree is a graphical representation of the evolutionary relationships
- Phylogeny of...
 - Species
 - Strains
 - Genes
 - Metabolic Pathways
- Trees can be inferred by morphology or molecular information

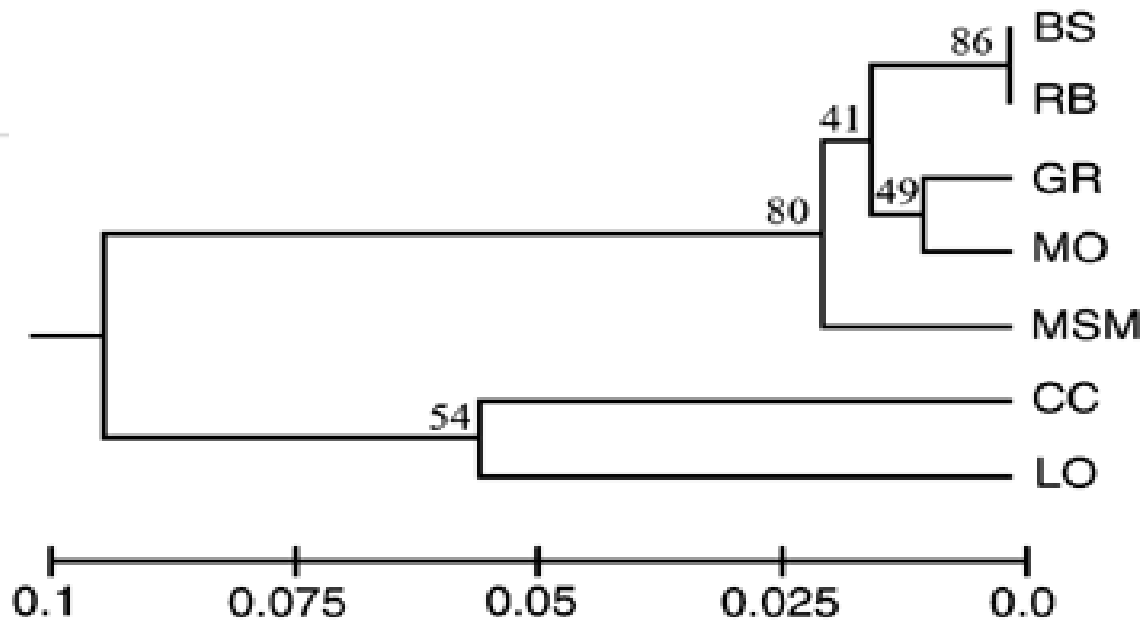
The aim:

To construct a visual representation (a tree) to describe the assumed evolution occurring between and among different groups (individuals, populations, species, etc.) and to study the reliability of the consensus tree.

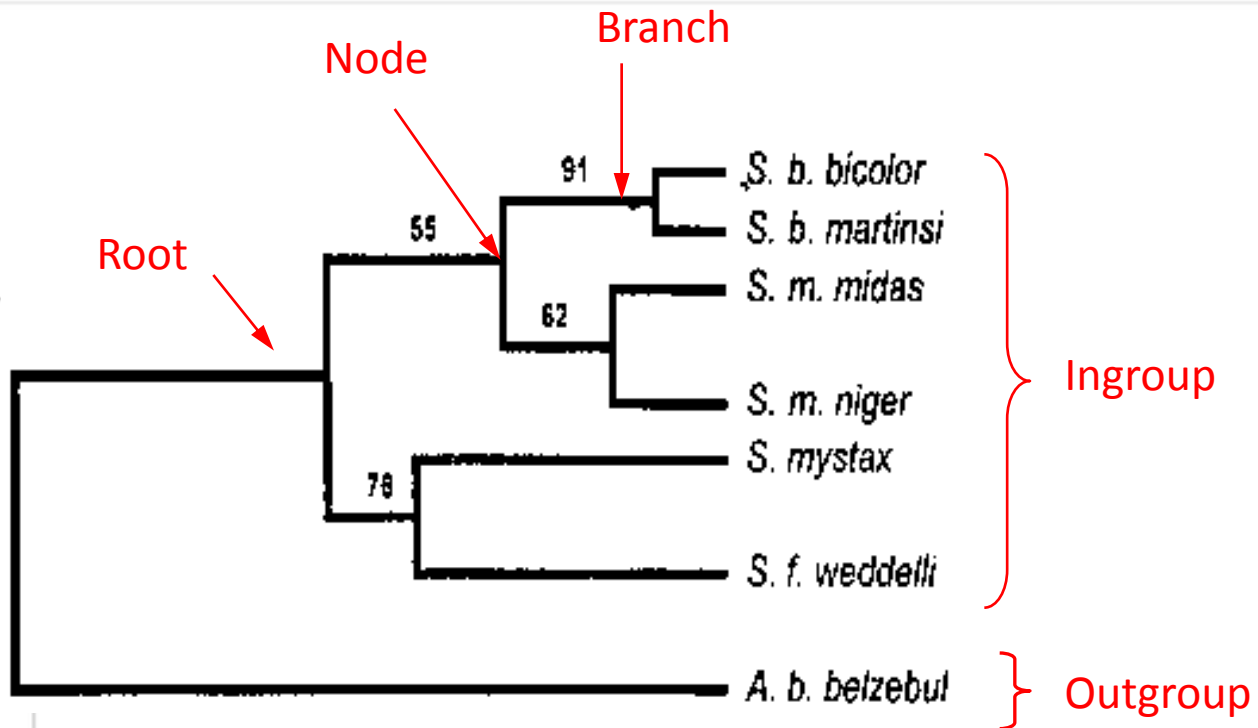
Assumptions

- Evolution produces dichotomous branching
- Evolution is simple – the best explanation assumes least mutations

A phylogeographic tree is a mathematical model of evolution

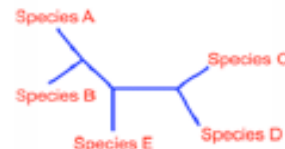
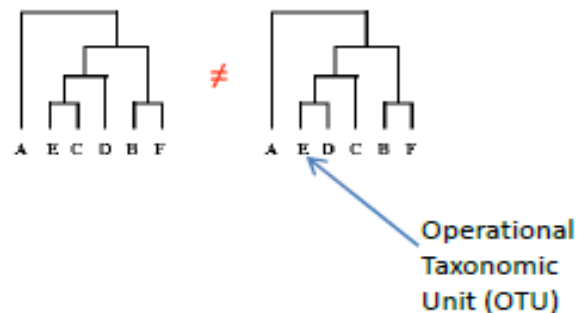


Parts of a phylogenetic tree



Parts of a Tree

- Nodes: taxonomic units (eg. genes, species, etc.)
 - Internal: ancestral state
 - Bifurcating
 - Multifurcating
 - External: Operational Taxonomic Units (OTUs)
- Branches: relationships among the taxonomic units (ie. ancestor-descendent relationship).
 - Clade
- Branch Length: number of changes that have occurred
- Topology: branching pattern
- Rooted vs Unrooted tree



Steps to Create Phylogenetic Trees

- Identify and acquire the sequences that are to be included on the tree
- Align the sequences (MSA using ClustalW, T-Coffee, MUSCLE, etc.)
- Estimate the tree by one of several methods
- Draw the tree and present it

Tree structure

- A tree can be also presented in a text format:
(A(B(C,D)))
- The graphic structure can be difficult to interpret (2-dimentional)

Analyses

1. Choosing the sequence type
2. Alignment of sequence data
3. Search for the best tree
4. Evaluation of tree reproducibility

Analyses can be based on:

- Differences in DNA-sequence structure
- Distance matrix between sequences
- Restriction data
- Allele data

Methods

- Distance matrix
- Maximum parsimony
- Minimum distance

Distance matrix

- A distance matrix is calculated from the sequence dataset
- Algorithms: Fitch-Margoliash, Neighbor-Joining or UPGMA in tree building
- Simple, finds only one tree
- Somewhat old-fashioned (OK if your alignment is good and evolutionary distances are short)

Maximum parsimony

- Finds the optimum tree by minimizing the number of evolutionary changes
- No assumptions on the evolutionary pattern
- May oversimplify evolution
- May produce several equally good trees

Maximum likelihood

- The best tree is found based on assumptions on evolution model
- Nucleotide models more advanced at the moment than aminoacid models
- Programs require lot of capacity from the system

Algorithms used for tree searching

- Exhaustive search: all possibilities → best tree → requires lots of time and computer resources
- Branch and Bound: a tree is built according to the model given → the tree is compared to the next tree while its constructed → if the first tree is better the second tree is abandoned → third tree... → best possible tree
- Heuristic Search: only the most likely options → saves time and resources, does not always result in the best tree

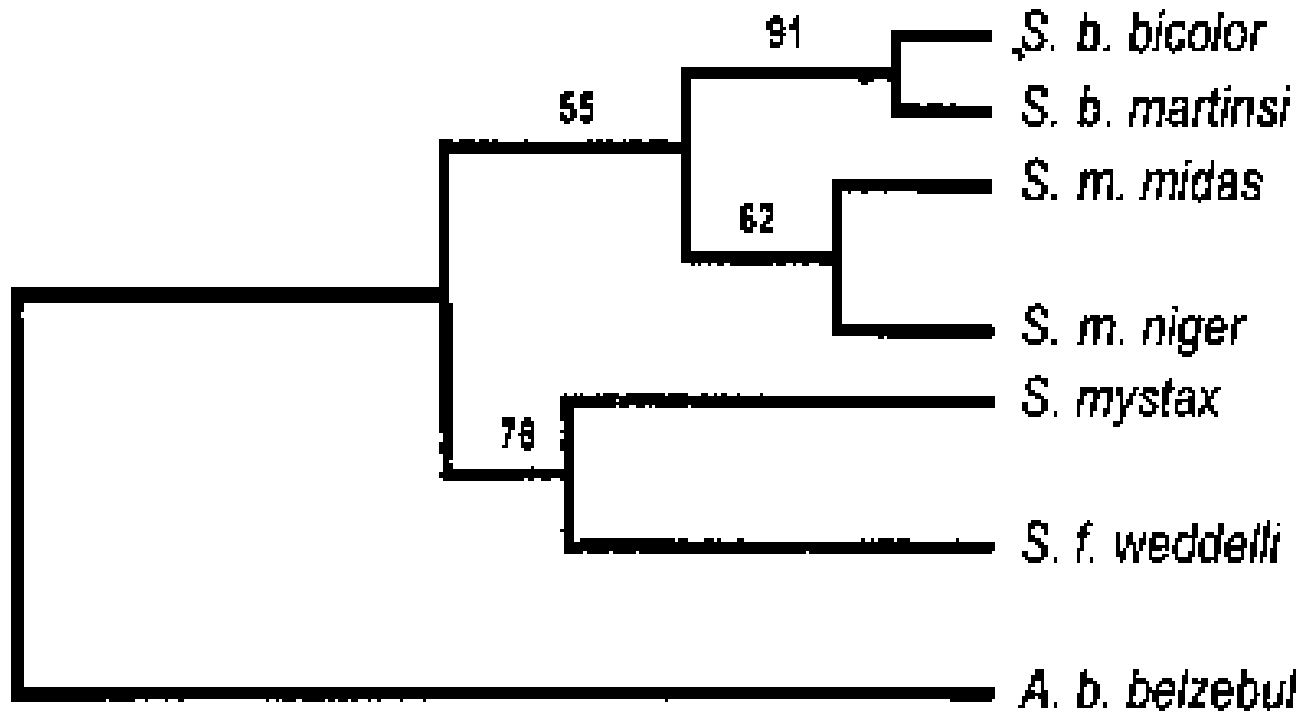
Bootstrapping

- Evaluation of the tree reliability
 - n number of trees are built
($n=100/1000/5000$)
- How many times a certain branch is reproduced

Values between 1-100 (%)

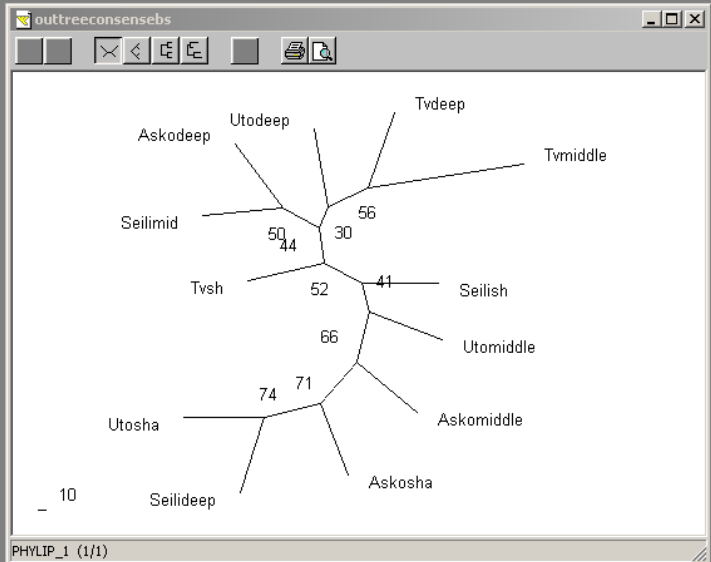
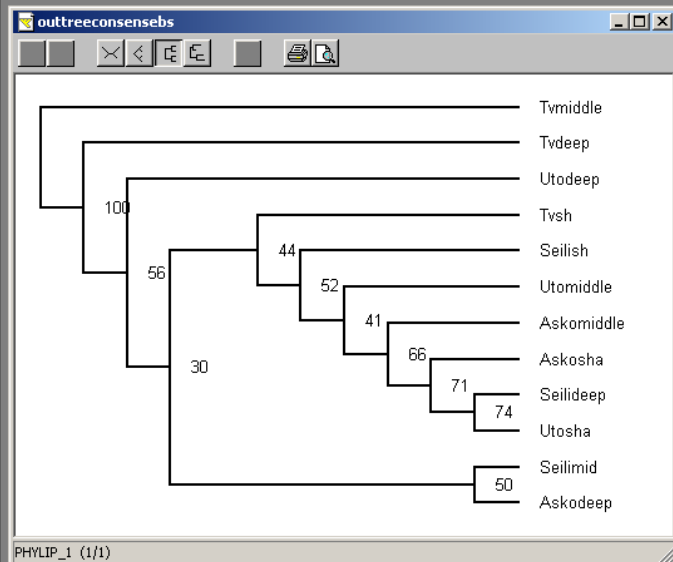
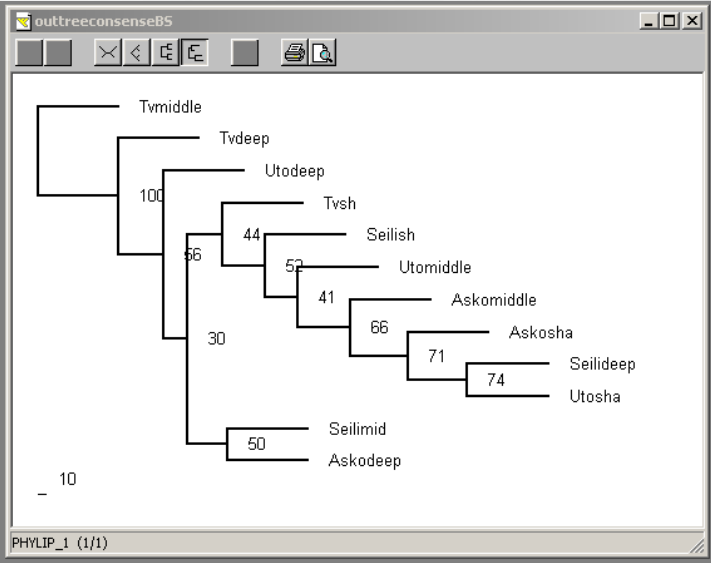
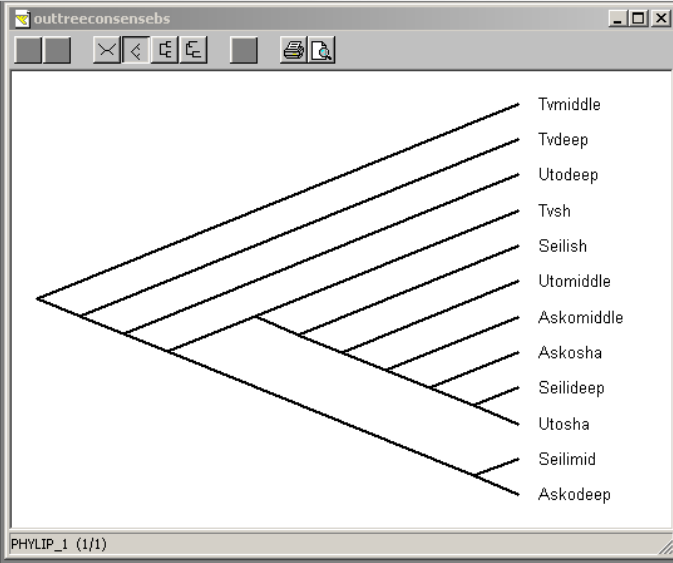
Tree Reliability: Bootstrapping

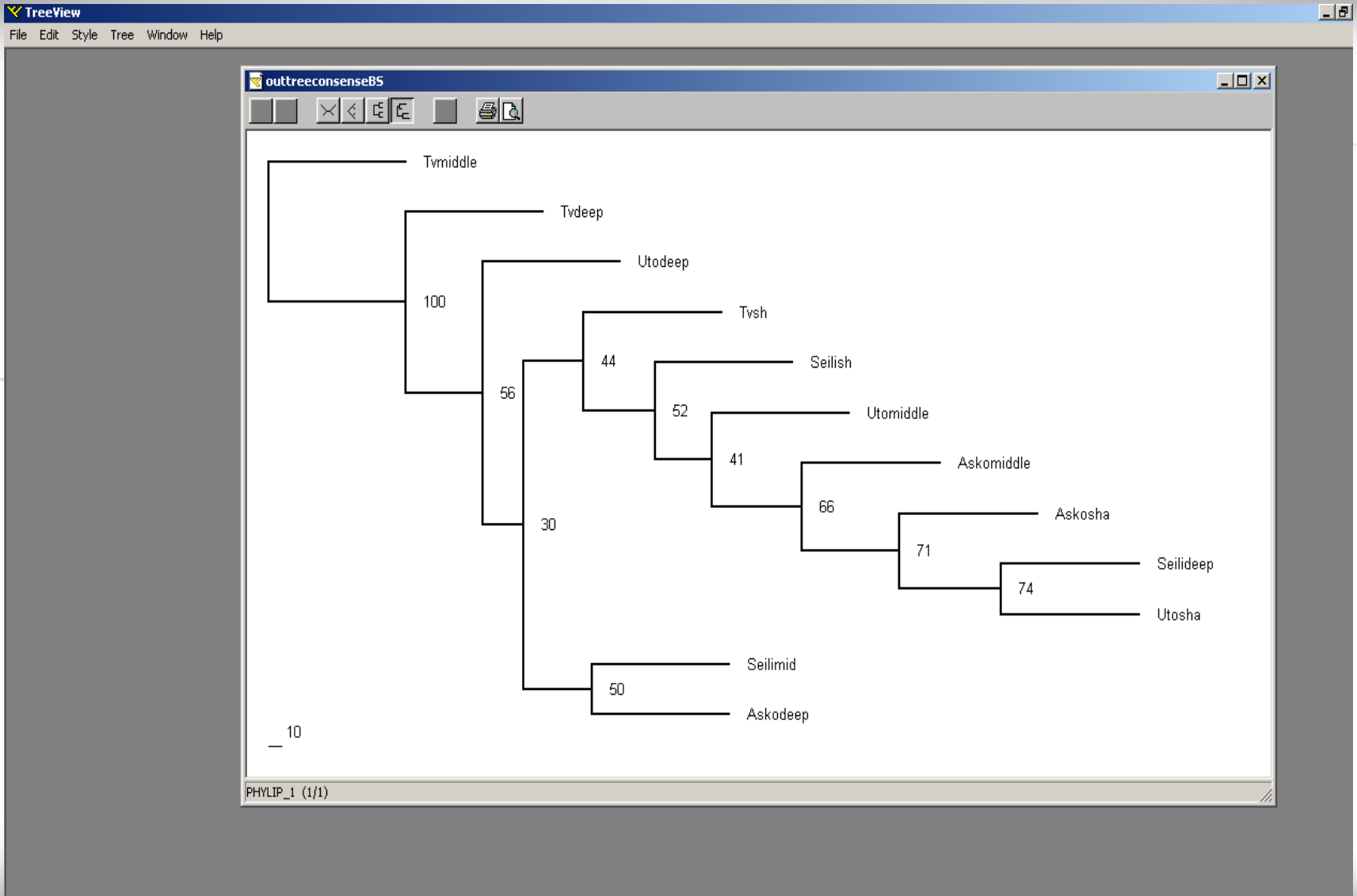
- A statistical re-sampling procedure commonly used for providing confidence to branches in phylogenetic trees
- A measure of repeatability, the probability that the branch would be recovered if the taxa were sampled again
- Bootstrapping values are typically presented from 1000 repeated calculations
- Bootstrap values of $>70\%$ is recommended



Visualising trees

- Treeview
- You can change the graphic presentation of a tree (cladogram, rectangular cladogram, radial tree, phylogram), but not change the structure of a tree





Some Tips

- Use more than one method
- Use more than one software package
- Examine more than one tree if multiple trees are generated
- Bootstrap your data
- Homology vs Homoplasy
- Consider another intermediate taxa to resolve relationship, if needed

Phylogenetic Tree Software

- PHYLIP (the *PHY*Logeny Inference Package):
<http://evolution.genetics.washington.edu/phylip.html>
- PAUP*:
<http://paup.csit.fsu.edu>
- MrBayes:
<http://mrbayes.csit.fsu.edu>
- MEGA(Molecular Evolutionary Genetic Analysis)
<http://www.megasoftware.net>

A More Comprehensive Listing of Phylogeny Programs:
<http://evolution.genetics.washington.edu/phylip/software.html>

MEGA

- Easy-to-use software with multiple features
- Features:
 - Aligning sequences
 - Estimating evolutionary distances
 - Building trees using several methods
 - Testing tree reliability
 - Marking Genes/Domains
 - Testing for selection
 - Computing sequence statistics

TUGAS PRAKTIKUM ---- KUMPULKAN MINGGU DEPAN

1. Buat Langkah langkah dalam analisis sekuens dengan teknik BLAST
2. Interpretasikanlah hasil BLAST berdasarkan E-Value, Max score, Total Score, Max Identified and Query Coverage

THANK
YOU



607132.wordpress.com

Noviani's Blog

