**Course: Advance Bio Informatics**

**Module Title: Tree Evaluation**

**Module No: 63**

**Stage 5: Criteria for tree evaluation**

After you have constructed a phylogenetic tree, how can you assess its accuracy? The main criteria by which accuracy may be assessed are consistency, efficiency, and robustness. One may study the accuracy of a tree-building approach or the accuracy of a particular tree.

**Bootstrap Analysis**

The most common approach is bootstrap analysis. Bootstrapping is not a technique to assess the accuracy of a tree. Instead, it describes the robustness of the tree topology. Givena particular branching order, how consistently does a tree-building algorithm ﬁnd that branching order using a randomly permuted version of the original data set? Bootstrapping allows an inference of the variability in an unknown distribution from which the data were drawn.

**Procedure**

Nonparametric bootstrapping is performed as follows. A multiple sequence alignment is used as the input data to generate a tree using some tree-building method. The program then makes an artiﬁcial data set of the same size as the original data set by randomly picking columns from the multiple sequence alignment. This is usually performed with replacement, meaning that any individual column may appear multiple times (or not at all). A tree is generated from the randomized data set. A large number of bootstrap replicates are then generated; typically, between 100 and 1000 new trees are made by this process. The bootstrap trees are compared

to the original, inferred tree(s). The information you get from bootstrapping is the frequency with which each clade in the original tree is observed.

Maximum likelihood approaches report the tree with the greatest likelihood, and they also report the likelihood for internal branches. For Bayesian inference of phylogeny, the result is typically the most probable tree (called a maximum a posteriori probability estimate). The results are often summarized using a majority rule consensus tree in which the values represent the posterior probability that each clade is true.

**MEGA4:** percentage of times that a given clade is supported in original tree is provided based on how often the bootstraps supported the original tree topology.

>70% are sometimes considered to provide support for the clade designations.

