**Trees of Trees: Phylogenetic Tree Clustering for Consensus and Comparison**

Nicholas R. Steffen, Sandy Irani, and Richard H. Lathrop
Department of Information and Computer Science,
University of California, Irvine, CA, 92697-3425 USA
{ nsteffen | irani | rickl } @ics.uci.edu

This work describes a new method for constructing phylogenetic trees with improved accuracy over commonly used methods. This method combines trees produced by maximum parsimony, neighbor joining, and maximum likelihood methods. These multiple phylogenetic trees are clustered using agglomerative hierarchical clustering. The resulting tree of trees, called a meta-tree, estimates relationship among all leaf phylogenetic trees analogous to how a phylogenetic tree estimates relationship among all leaf species. Each internal node of the meta-tree estimates the consensus of its descendant phylogenetic trees by determining a tree based on the average inter-species distance matrix of its immediate children. The root of the meta-tree therefore estimates the consensus of all leaf phylogenetic trees. Two sets of test sequences were generated representing maximally balanced and maximally unbalanced trees with eight leaves. Sequence lengths from 100 to 4000 were used and 1000 trials performed with each sequence length. Using these sequences, trees were built using neighbor-joining, maximum likelihood, maximum parsimony, and the meta-tree methods. The meta-tree method was superior to all of the other methods at estimating the original tree topology for unbalanced trees and was comparable to the best method for balanced trees. Phylogenetic trees built from HIV-1 integrase nucleotide sequences were also clustered using this method. In this case the meta-tree roots have a topology that correctly group the different lineages.

Adams, E. N. 1972. Consensus techniques and the comparison of taxonomic trees. Syst. Zool. 21:390-397.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368-376.

Fitch, W. M. 1971. Toward defining the course of evolution: minimum change for a specified tree topology. Syst. Zool. 20:406-416.

Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131-147.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

Smith, R. F. and T. F. Smith. 1992. Pattern-induced multi-sequence alignment (PIMA) algorithm employing secondary structure-dependent gap penalties for use in comparative protein modelling. Protein Eng 5: 35-41.