

CMSC 858P Algorithms for Biosequence Analysis
Project 2 Proposal

Phylogenetic tree construction using neighbor-joining*

Problem:

Given a distance matrix between a set of sequences, construct and display the phylogenetic tree for the sequences using the neighbor-joining algorithm.

Background, relevance to biology and class:

Phylogenetic trees are used to show the evolutionary relationships among known sequences of organisms [1]. Typically the leaves in phylogenetic trees correspond to the sequences; each node with descendants represents the most recent common ancestor of those descendants; and edge lengths are used to represent the time estimates for one node to evolve into another.

In terms of multiple alignment problem, phylogenetic trees are used to extract out sequences, which are in turn used to find an optimal alignment to a set of related sequences.

Algorithm:

Different methods have been used to construct phylogenetic trees; among these are neighbor-joining, minimum evolution, maximum parsimony, and maximum likelihood [2]. Although some of them result in more accurate trees (i.e. maximum likelihood), they also have more runtime costs. Distance matrix based methods such as neighbor-joining provide great efficiency and have been used by many tools so far. On the other hand, they always end up in a single possible tree and mostly depend on the evolution model used [3].

References:

- [1] http://en.wikipedia.org/wiki/Phylogenetic_tree
- [2] http://en.wikipedia.org/wiki/Neighbour_joining
- [3] <http://www.icp.ucl.ac.be/~opperd/private/neighbor.html>

*Taken from the undergraduate class project suggestions
(<http://www.cbcu.umd.edu/confcour/Spring2007/CMSC423-materials/Project%202.pdf>).