

Phylogenetic Tree (continued....)

Different methods of phylogenetic tree construction

Phylogenetics is the study of genetic relatedness of individuals of the same, or different, species. Through phylogenetics, evolutionary relationships can be inferred. A phylogenetic tree may be rooted or unrooted, depending on whether the ancestral root is known or unknown, respectively. A phylogenetic tree's root is the origin of evolution of the individuals studied. Branches between leaves show the evolutionary relationships between sequences, individuals, or species, and branch length represents evolutionary time. When constructing and analyzing phylogenetic trees, it is important to remember that the resulting tree is simply an estimate and is unlikely to represent the true evolutionary tree of life. Various methods can be used to construct a phylogenetic tree. The two most commonly used and most robust approaches are maximum likelihood and Bayesian methods.

Distance-matrix methods

Distance-matrix methods are rapid approaches that measure the genetic distances between sequences. After having aligned the sequences through multiple sequence alignment, the proportion of mismatched positions is calculated. From this, a matrix is constructed that describes the genetic distance between each sequence pair. In the resulting phylogenetic tree, closely-related sequences are found under the same interior node, and the branch lengths represent the observed genetic distances between sequences.

Neighbor joining (NJ), and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) are two distance-matrix methods. NJ and UPGMA produce unrooted and rooted trees, respectively. NJ is a bottom-up clustering algorithm. The main advantages of NJ are its rapid speed, regardless of dataset size, and that it does not assume an equal rate of evolution amongst all lineages. Despite this, NJ only generates one phylogenetic tree, even when there are several possibilities. UPGMA is an unweighted method, meaning that each genetic distance contributes equally to the overall average. UPGMA is not a particularly popular method as it makes various assumptions, including the assumption that the evolutionary rate is equal for all lineages.

Maximum parsimony

Maximum parsimony attempts to reduce branch length by minimizing the number of evolutionary changes required between sequences. The optimal tree would be the shortest tree with the fewest mutations. All potential trees are evaluated, and the tree with the least amount of homoplasy, or convergent evolution, is selected as the most likely tree. Since the most-parsimonious tree is always the shortest tree, it may not necessarily best represent the evolutionary changes that have occurred. Also, maximum parsimony is not statistically consistent, leading to issues when drawing conclusions.

Maximum likelihood

Despite being slow and computationally expensive, maximum likelihood is the most commonly used phylogenetic method used in research papers, and it is ideal for phylogeny construction from sequence data. For each nucleotide position in a sequence, the maximum likelihood algorithm estimates the probability of that position being a particular nucleotide, based on whether the ancestral sequences possessed that specific nucleotide. The cumulative probabilities for the entire sequence are calculated for both branches of a bifurcating tree. The likelihood of the whole tree is provided by the sum of the probabilities of both branches of the tree. Maximum likelihood is based on the concept that each nucleotide site evolves independently, enabling phylogenetic relationships to be analysed at each site. The maximum likelihood method can be carried out in a reasonable amount of time for four sequences. If more than four sequences are to be analysed, then basic trees are constructed for the initial four sequences, and further sequences are subsequently added, and maximum likelihood is recalculated. Bias can be introduced into the calculation as the order in which the sequences are added, and the initial sequence used, play pivotal roles in the outcome of the tree. Bias can be avoided by repeating the entire process multiple times at random so that the majority rule consensus tree can be selected.

Bayesian inference

Bayesian inference methods assume phylogeny by using posterior probabilities of phylogenetic trees. A posterior probability is generated for each tree by combining its prior probability with the likelihood of the data. A phylogeny is best represented by the tree with the highest posterior probability. Not only does Bayesian inference produce results that can be easily interpreted; it can also

incorporate prior information and complex models of evolution into the analysis, as well as accounting for phylogenetic uncertainty.

Reference links:

<https://www.khanacademy.org/science/high-school-biology/hs-evolution/hs-phylogeny/a/phylogenetic-trees>

<http://phylobotanist.blogspot.com/2016/01/types-of-phylogenetic-tree-diagrams.html>