Chapter 12 Molecular Phylogeny

Molecular phylogeny is the study of evolutionary relationships among organisms by using techniques of molecular biology. The evolutionary relationships among a group of organism are illustrated by means of a phylogenetic tree.

- **Phenetic data:** Phenetic relationships are defined by Sneath and Sokal (1973) are similarities based on a set pf phenotypic characters for the objects under stuty.
- **Cladistic data:** Cladistic relationships contain information about ancestry and so can be used to study evolutionary pathways.

Phenogram and cladogram:

Species trees and **gene trees:** A tree constructed from the DNA sequence of a short region, for example, is called a gene tree and may not be the same as the species tree.



Figure 10.1 A species tree links species A, B, X, Y, and Z. Three possible gene trees, indicated by solid lines, link the genes indicated by solid circles. Source: Nei (1987).

Two species may carry genes that diverged prior to the species split.

Rooted and unrooted trees:



Figure 10.2 Two of the 15 rooted trees and one of the three unrooted trees for four species.

Character and distance data: Character data provide information about attributes of genes, individuals, populations, or species, whereas distance data (similarity data) refer to pairs of genes, individuals, populations, or species.

Table 10.1 Matrix of distances between all pairs of *n* operational taxonomic units.

		OTU's					
		1	2	3		n	
OTU's	$\begin{array}{c}1\\2\\3\\\ldots\\n\end{array}$	$\begin{array}{c} - & & - & & \\ d_{21} & & & \\ d_{31} & & & & \\ \cdots & & & & \\ d_{t1} \end{array}$	$d_{12} - d_{32} - d_{32} - d_{t2}$	$d_{13} \\ d_{23} \\ - \\ \dots \\ d_{t3}$	· · · · · · · · · ·	d_{1t} d_{2t} d_{3t} 	

Phylogenetic Trees: nodes, branches, topology, branch length. External nodes, internal nodes, Operational taxonomic units (OTUs),Bifurcating and multifurcating nodes. True and inferred trees:

Methods of Tree Reconstruction Distance Matrix Methods: UPGMA (unweighted pair-group method using arithmetic average): UPGMA assumes a constant rate of change along all branches in a tree.

Jukes-Cantor distance:

$$\widetilde{K} = \frac{3}{4} \ln(\frac{3}{4\widetilde{q} - 1})$$

q: the proportion of the bases in the two sequences is the same.

 Table 10.2
 Numbers of differences (below diagonal) and Jukes-Cantor distances (above diagonal) for the five mitochondrial sequences of Figure 1.3.

	Human	Chimpanzee	Gorilla	Orangutan	Gibbon
Human	_	0.015	0.045	0.143	0.198
Chimpanzee	1	_	0.030	0.126	0.179
Gorilla	3	2	-	0.092	0.179
Orangutan	9	8	6	_	0.179
Gibbon	12	11	11	11	



Figure 10.3 UPGMA dendrogram for the mitochondrial sequence data of Figure 1.3.

Fitch-Margoliash Algorithm: The constant rate assumption is removed.



Figure 10.4 Initial step for Fitch-Margoliash algorithm applied to the mitochondrial data of Figure 1.3.



Figure 10.5 Intermediate steps for Fitch-Margoliash algorithm applied to mitochondrial sequence data of Figure 1.3.

Fitch-Margoliash unrooted tree



Figure 10.6 Fitch-Margoliash unrooted tree for mitochondrial sequence data in Figure 1.3.





Figure 10.7 Fitch-Margoliash rooted tree for the mitochondrial sequence data in Figure 1.3.

Neighboring-joining method: Saitou and Nei (1987) described a method for identifying closest pairs, or neighbors, of OUT's in a way to manimized the total length of a tree (for unrooted tree?)

Table 10.4 Neighbor-joining calculations for mitochondrial sequences of Figure 1.3. d_{ij} above diagonal, M_{ij} below diagonal.



Figure 10.9 Neighbor-joining unrooted tree for mitochondrial sequence data in Figure 1.3.

Parsimony Methods: Parsimony methods take explicit notice of the character values observed for each species, rather than working with the distance matrix. Branch lengths are generally not obtained. **Informative sites:** A nucleotide site is

niormative sites: A nucleotide site is

phylogenetically imformative only if it favors some trees over the others.



Figure 10.10 Illustration of procedure for finding maximum parsimony tree for one site in six sequences. Source: Fitch (1971).



Figure 10.11 Maximum parsimony tree for mitochondrial sequence data of Figure 1.3. Figures are the numbers of changes between nodes.

Maximum likelihood method



Figure 10.12 Rooted tree for two sequences. At site j, the two sequences have bases s_1, s_2 and the node has base k.