**Course: Advanced Bioinformatics**

**Module title: Phylogeny tree**

**Module no. : 44**

Phylogeny Tree Space

Space of phylogeny tree is exponential.

For n sequences, the number of un-rooted tree is (2n-5)!

For n sequences, the number of rooted tree is (2n-3)!

Rooted Tree

Definition of a tree:

Edge num: E

Internal node num: I

Leaf node num: L

E = I + L - 1

E = 2 \* I (degree)

E = I + L - 1

2I = I + L - 1

I = L - 1

(# internal node= #leaf-1)

E = 2L - 2

Unrooted Tree

Total degree =

(L+3\*I) = 2E

E = L + I - 1

I = L - 2

E = 2L - 3

Total No of Unrooted Tree

Given n species (n >= 3), there are (2i-5)!

Unrooted bifurcating trees.

#leaf node #edge #tree

3 3 1

4 5 1\*3

5 7 1\*3\*5

...

n 2n-3 1\*3\*5\*...\*2n-5

Total number of rooted tree for n leaf nodes:

Total number of un-rooted tree \* total number of edge

= (2n-5)! \* (2n-3) = (2n-3)!!

Unrooted Rooted

3 1 3

4 3 15

5 15 105

6 105 945

7 945 10395

8 10395 135135

9 135135 2027025

10 2027025 34459425

**Tree of Life**

On one level, it is interesting to understand and study how the evolution of species has occurred. There are many different resources discussing the evolution of species. This includes the NCBI taxonomy web sites, and the University of Arizona’s tree of life project. We’ll take a look at both of these web sites in order to get a better appreciation for the evolution of species relative to one another.

NCBI Taxonomy <http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/>

Tree of Life <http://tolweb.org/tree/>

**Evolutionary Trees**

An evolutionary tree is a two dimensional graph showing the evolutionary relationship among a set of items being compared. This set can be organisms, genes, or dna sequences. Consider for the moment that each of the units in the set are referred to as a taxon. Each taxon will be defined by a distinct unit on the tree.

An evolutionary tree is composed of outer branches or leaves that represent the taxa and nodes and branches representing the relationships among the taxa. Two taxa that are derived from the same common ancestor will share a node in the graph. In general, approaches to designing evolutionary trees attempt to define the length of each branch to the next node according to the number of sequence level changes that occurred. One thing to be careful of in phylogenetic analysis is that this distance may not be in direct relation to evolutionary time. Analyses that prescribe to the theory of a uniform rate of mutation are known as the molecular clock hypothesis.

**Rooted Trees**

In a rooted tree topology, one sequence (the root) is defined to be the common ancestor of all of the other sequences. A unique path leads from the root node to any other node, and the direction of the path indicates evolutionary time. The root is chosen by including a sequence from an organism that is thought to have branched off earlier than the other sequences. If the molecular clock hypothesis holds, it is also possible to predict a root. As the number of sequences increase, the number of possible rooted trees increases very rapidly. In some cases, a bifurcating binary tree is the best model to simulate evolutionary events in which case one species branches off into two separate species.

Example of a rooted tree:

