

# Phylogenetic Trees

# Phylogenetics

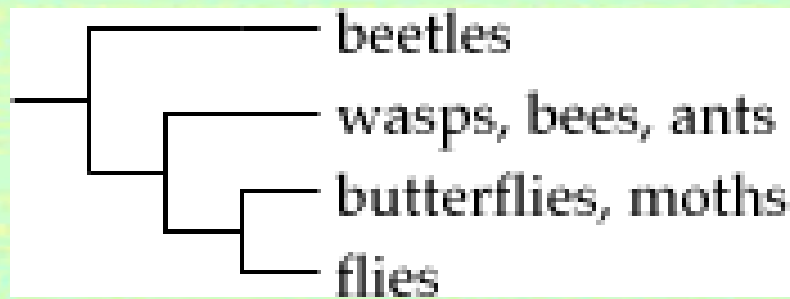
phylogenetics is the study of evolutionary relatedness among various groups of organisms (e.g., species, populations).

# Cladistics

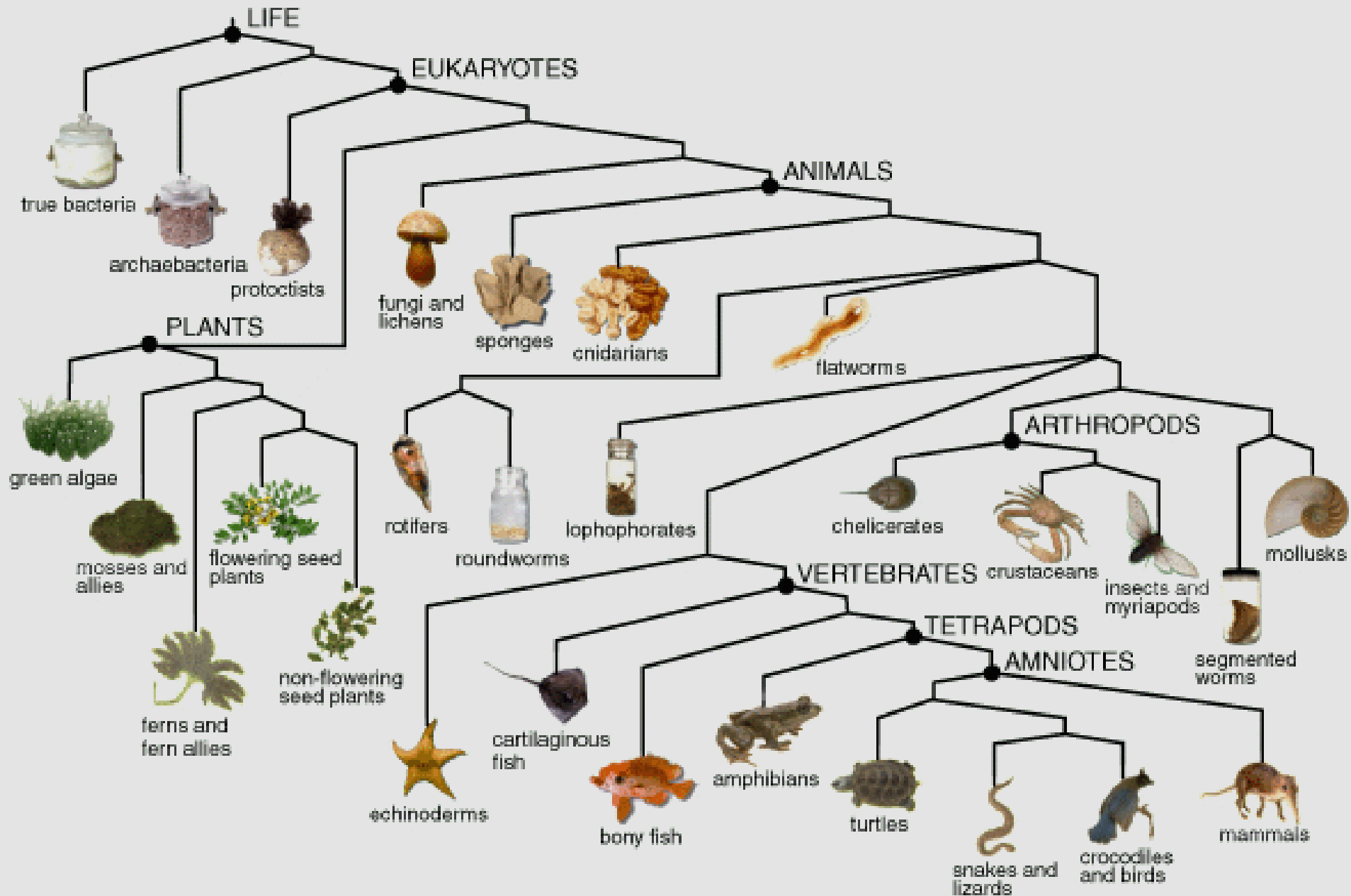
As treelike relationship-diagrams called "cladogram" is drawn up to show different hypotheses of relationships.

A cladistic analysis is typically based on morphological data.

# Cladistics



# Cladistics: *tree of life*



# Phylogenetic Trees

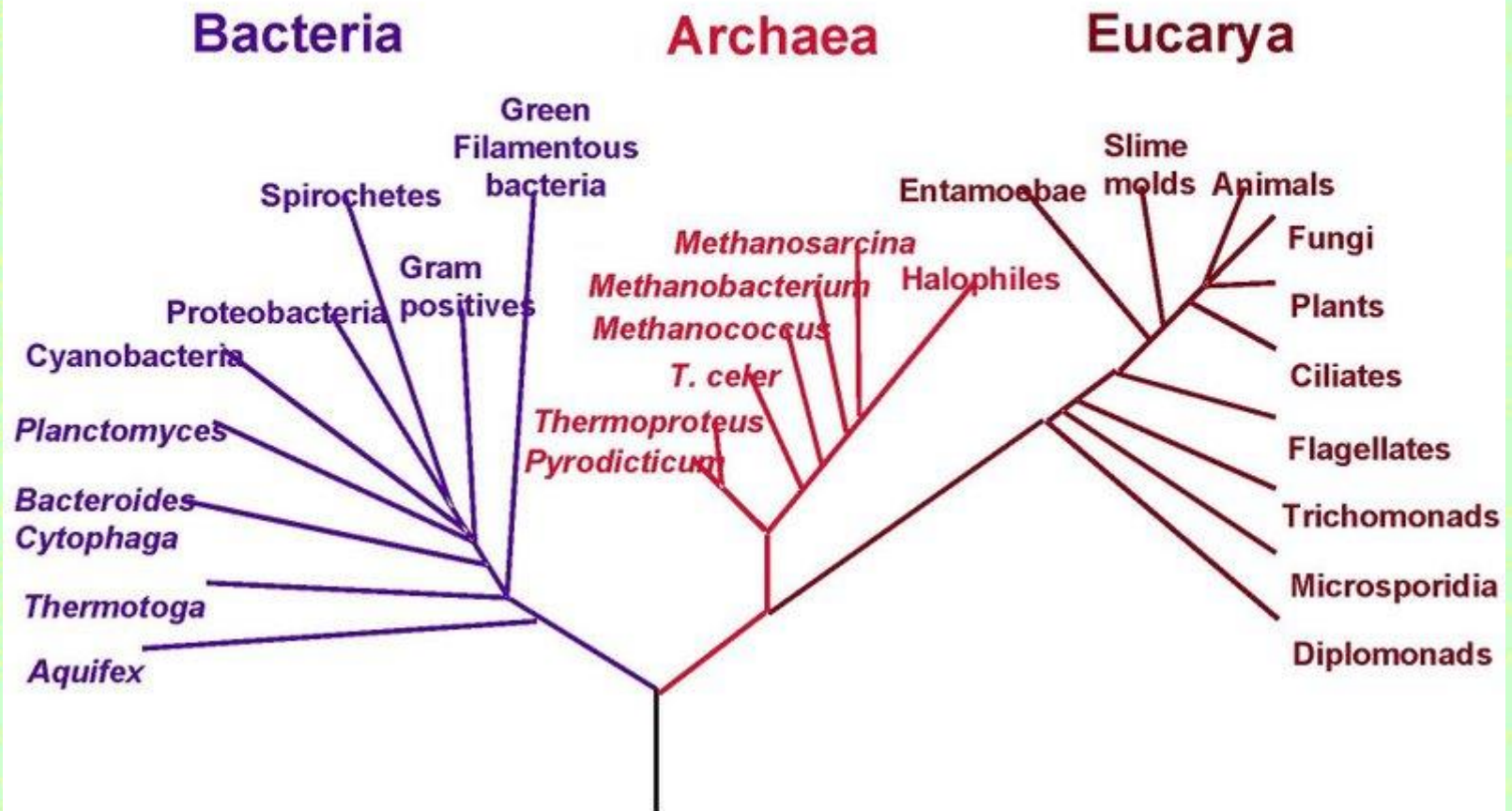
A phylogenetic tree is a tree showing the evolutionary interrelationships among various species or other entities that are believed to have a common ancestor. A phylogenetic tree is a form of a cladogram. In a phylogenetic tree, each node with descendants represents the most recent common ancestor of the descendants, and edge lengths correspond to time estimates.

Each node in a phylogenetic tree is called a taxonomic unit. Internal nodes are generally referred to as Hypothetical Taxonomic Units (HTUs) as they cannot be directly observed.

# Rooted and Unrooted Trees

A **rooted phylogenetic tree** is a directed tree with a unique node corresponding to the (usually imputed) most recent common ancestor of all the entities at the leaves of the tree. Figure 1 depicts a rooted phylogenetic tree, which has been colored according to the three-domain system (Woese 1998). The most common method for rooting trees is the use of an uncontroversial outgroup - close enough to allow inference from sequence or trait data, but far enough to be a clear outgroup.

# Rooted Phylogenetic Tree Phylogenetic Tree of Life

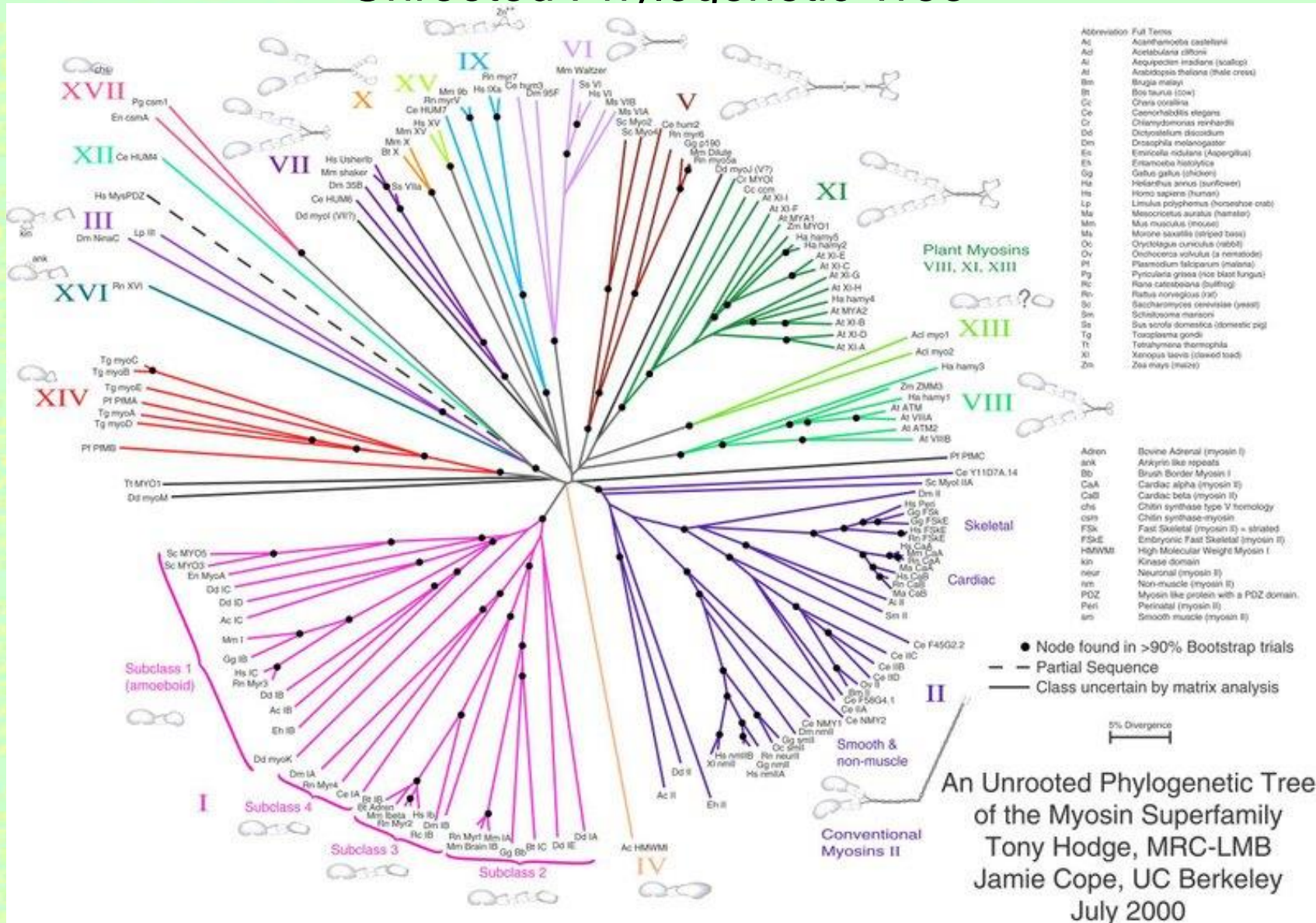




# Rooted and Unrooted Trees

**Unrooted phylogenetic trees** can be generated from rooted trees by omitting the root from a rooted tree, a root cannot be inferred on an unrooted tree without either an outgroup or additional assumptions (for instance, about relative rates of divergence). Figure 2 depicts an unrooted phylogenetic tree<sup>1</sup> for myosin, a superfamily of proteins. Links to other pictures are given in the pictures on the web subsection below.

# Unrooted Phylogenetic Tree



# Distance and Character

A tree can be based on

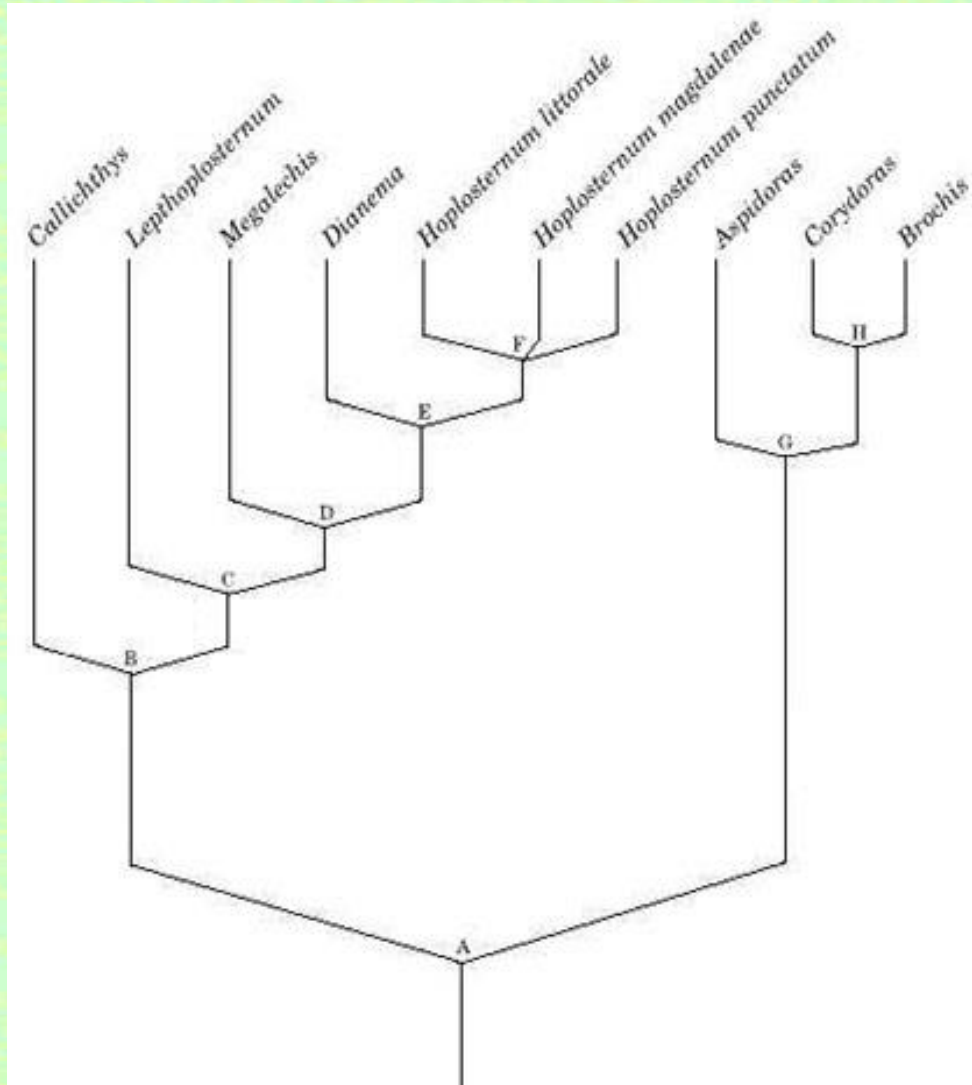
1. **quantitative measures** like the **distance** or **similarity** between species, or
2. based on **qualitative aspects** like **common characters**.

# Trees and Branch Length

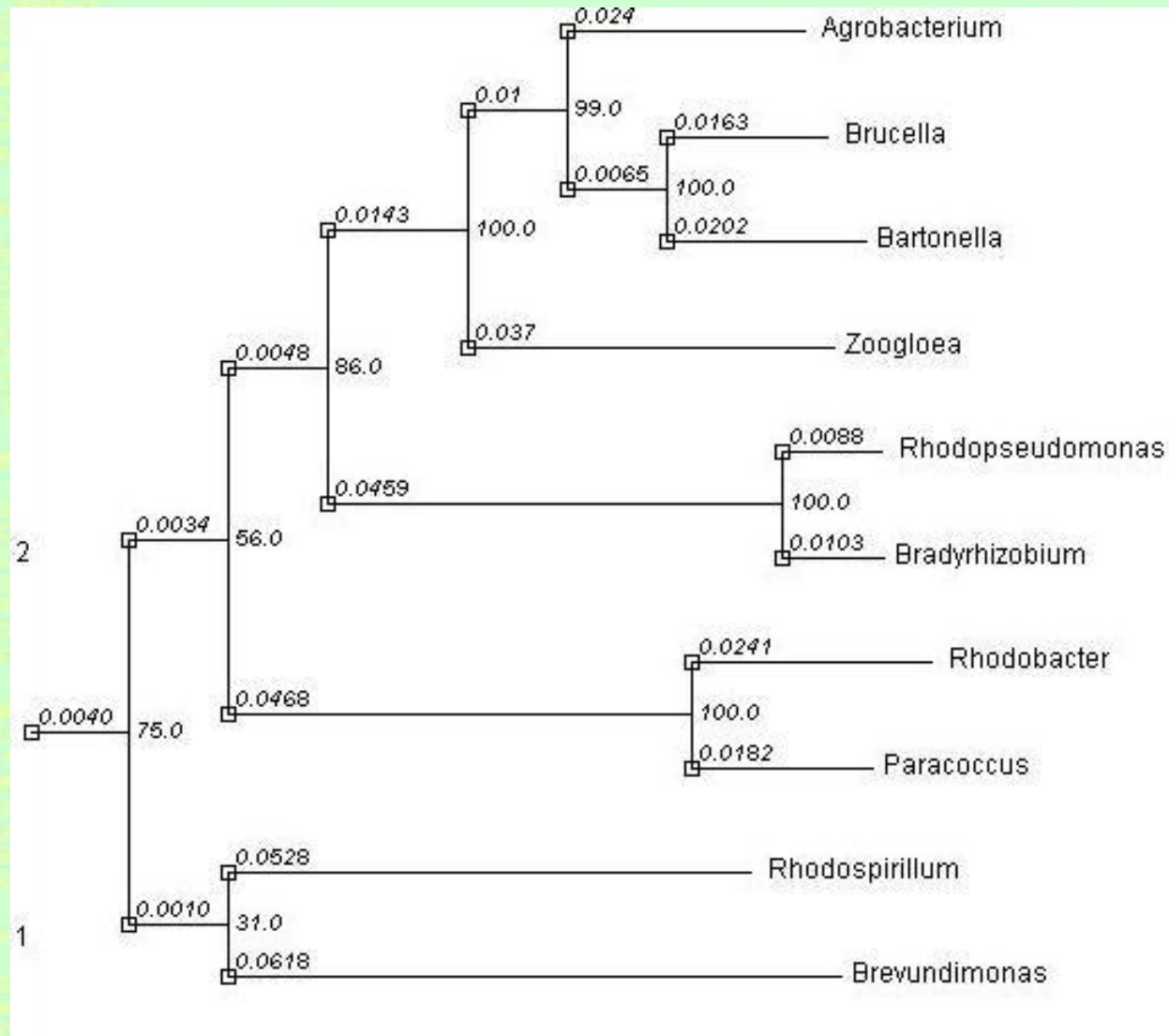
A tree can be a branching tree-graph where branches indicate close phylogenetic relations.

Alternatively, branches can have length that indicate the phylogenetic closeness.

# Tree without Branch Length



# Tree with Branch Length



# Constructing Phylogenetic Trees

There are three main methods of constructing phylogenetic trees:

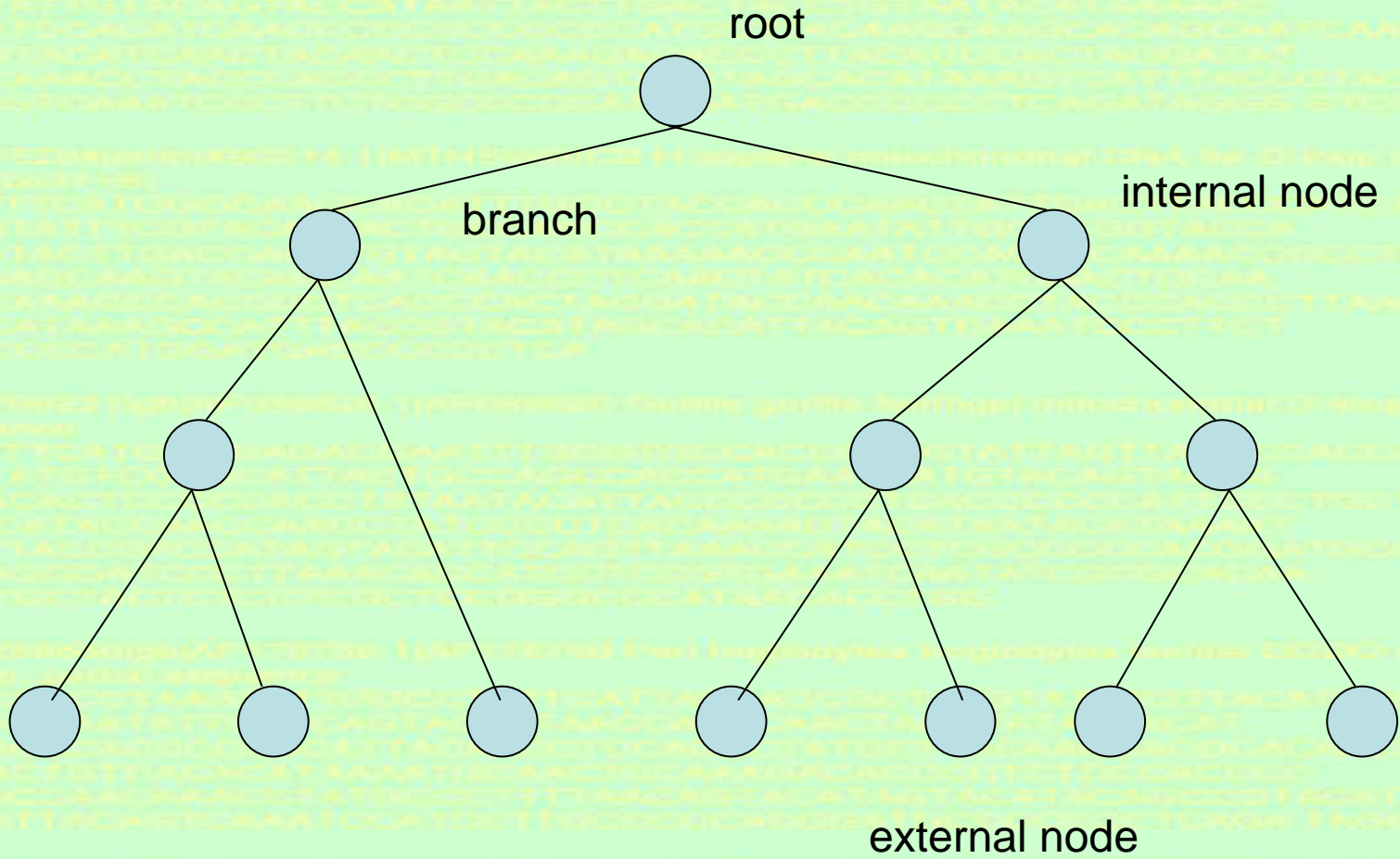
- \* **distance-based methods** such as UPGMA and neighbour-joining,
- \* **parsimony-based methods** such as maximum parsimony, and
- \* **character-based methods** such as maximum likelihood or Bayesian inference.

**Parsimony** is a 'less is better' concept of frugality, economy, stinginess or caution in arriving at a hypothesis or course of action. The word derives from Latin *parsimonia*, from *parcere*: **to spare**.

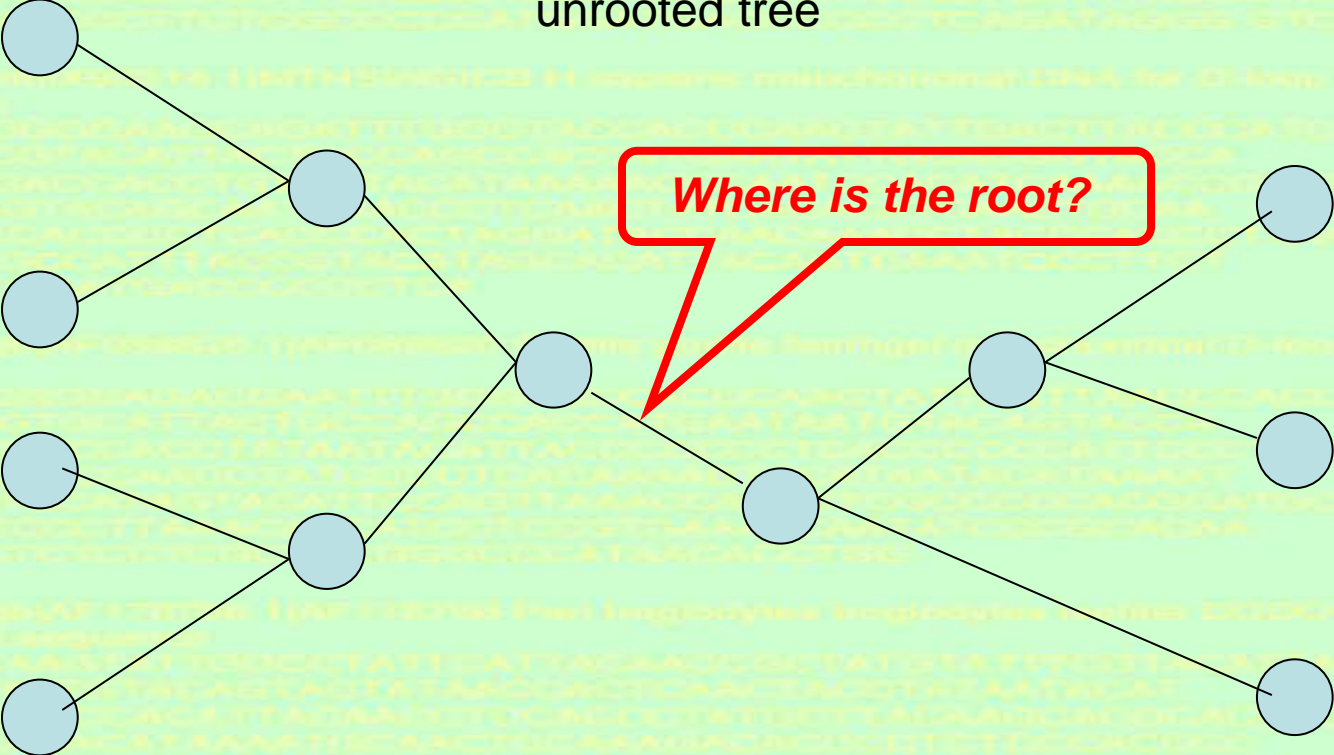
## 7.2 *On trees and evolution*

- \* Relation between “taxa”
- \* Internal nodes and external nodes (leafs)
- \* Branches connects nodes
- \* Bifurcating tree: **internal** nodes have **degree: 3**, **external** nodes degree: **1**, root **degree: 2**.
- \* Root connects to ‘outgroup’
- \* Multifurcating trees

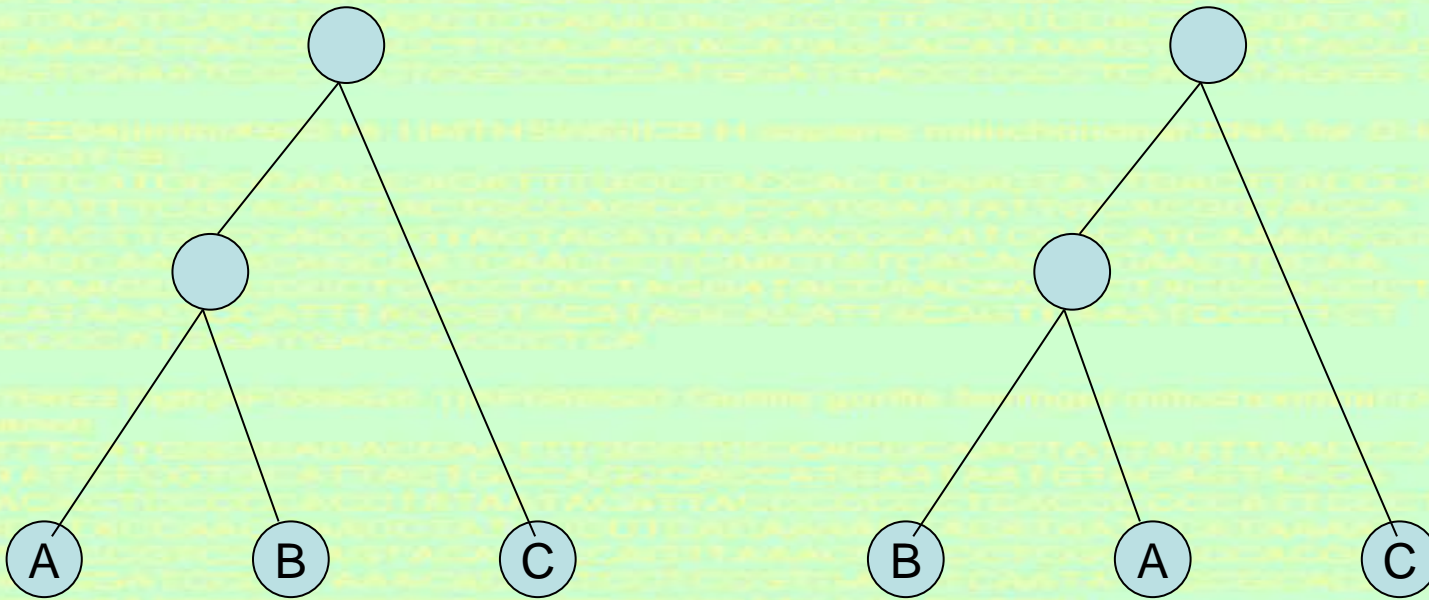




unrooted tree



\* Any rotation of the internal branches of a tree keeps the the phylogenetic relations intact



rotation invariant

# Number of possible trees

- \*  $t$  is number of taxa
- \* unrooted trees for  $t > 2$ :  $(2t - 5)! / (2t - 3(t-3))!$
- \* rooted trees for  $t > 1$ :  $(2t - 3)! / (2t - 2(n-2))!$
- \*  $t = 5$ : rooted trees = 105
- \*  $t = 10$  : rooted trees = 34,459,425

# Representing trees

- \* Various possibilities
- \* Listing of nodes
- \*  $t$  taxa =  $t$  external nodes:  $(t - 1)$  internal nodes
- \* internal nodes with children:  $(t - 1) \times 3$  matrix
- \* ( internal node, daughter\_1, daughter\_2)
- \* Newick format: see next slide for example

# Finding Branch lengths:

Three-point formula:

$$L_x + L_y = d_{AB}$$

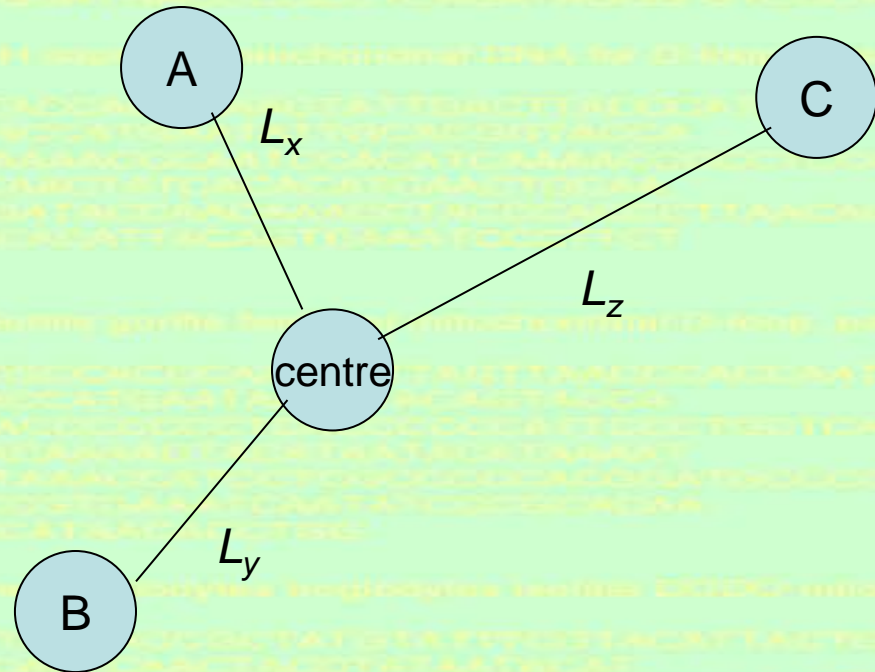
$$L_x + L_z = d_{AC}$$

$$L_y + L_z = d_{BC}$$

$$L_x = (d_{AB} + d_{AC} - d_{BC})/2$$

$$L_y = (d_{AB} + d_{BC} - d_{AC})/2$$

$$L_z = (d_{AC} + d_{BC} - d_{AB})/2$$



## Four-point formula:

when (1,2) and (i,j) are neighbor-couples!  
Four-point condition

$$d(1,2) + d(i,j) < d(i,1) + d(2,j)$$

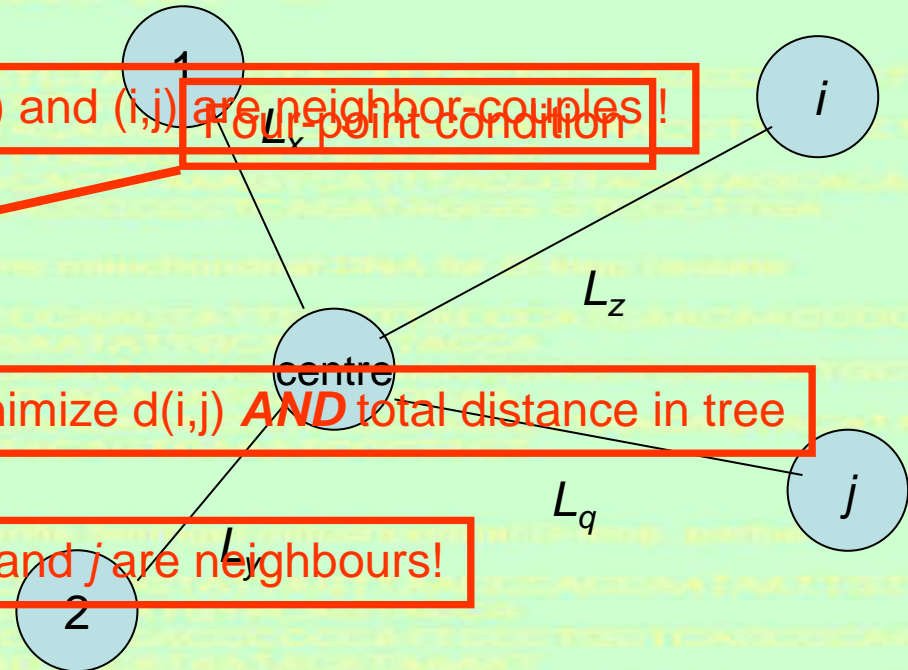
$$R_i = \sum_j d(t_i, t_j)$$

Minimize  $d(i,j)$  **AND** total distance in tree

$$M(i,j) = (n-2)d(i,j) - R_i - R_j$$

If  $i$  and  $j$  are neighbours!

$$M(i,j) < M(i,k) \text{ for all } k \text{ not equal to } j$$





# NJ algorithm:

Input:  $n \times n$  distance matrix  $D$  and an outgroup

Output: rooted phylogenetic tree  $T$

**Step 1:** Compute new table  $M$  using  $D$  – select smallest value of  $M$  to select two taxa to join

**Step 2:** Join the two taxa  $t_i$  and  $t_j$  to a new vertex  $V$  - use 3-point formula to calculate the updated distance matrix  $D'$  where  $t_i$  and  $t_j$  are replaced by  $V$ .

**Step 3:** Compute branch lengths from  $t_k$  to  $V$  using 3-point formula,  $T(V,1) = t_i$  and  $T(V,2) = t_j$  and  $TD(t_i) = L(t_i, V)$  and  $TD(t_j) = L(t_j, V)$ .

**Step 4:** The distance matrix  $D'$  now contains  $n - 1$  taxa. If there are more than 2 taxa left go to step 1. If two taxa are left join them by a branch of length  $d(t_i, t_j)$ .

**Step 5:** Define the root node as the branch connecting the outgroup to the rest of the tree. (Alternatively, determine the so-called “**mid-point**”)

## UPGMA and ultrametric trees:

If the distance from the root to all leafs is equal the tree is **ultrametric**

In that case we can use  $D$  instead of  $M$  and the algorithm is called UPGMA (Unweighted Pair Group Method)

**Ultrametricity** must be valid for the real tree, but due to noise this condition will in practice generate erroneous trees.