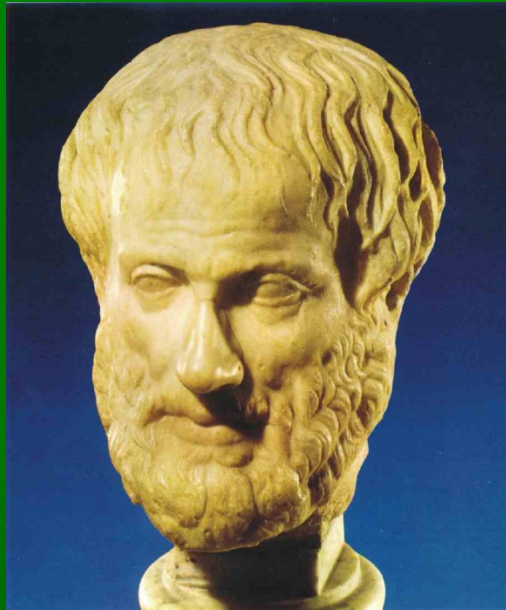


# Phylogenetic Analysis





## Aristotle

- Through classification, one might discover the essence and purpose of species.

Enaima (animals with blood)

Viviparous: Man

Cetacea

Quadrupeds

Oviparous:

Perfect eggs: Birds

Scaly Quadrupeds

Imperfect eggs: Fishes

Anaima (bloodless animals)

Perfect eggs: Malacia

Malacostraca

Special eggs: Insects

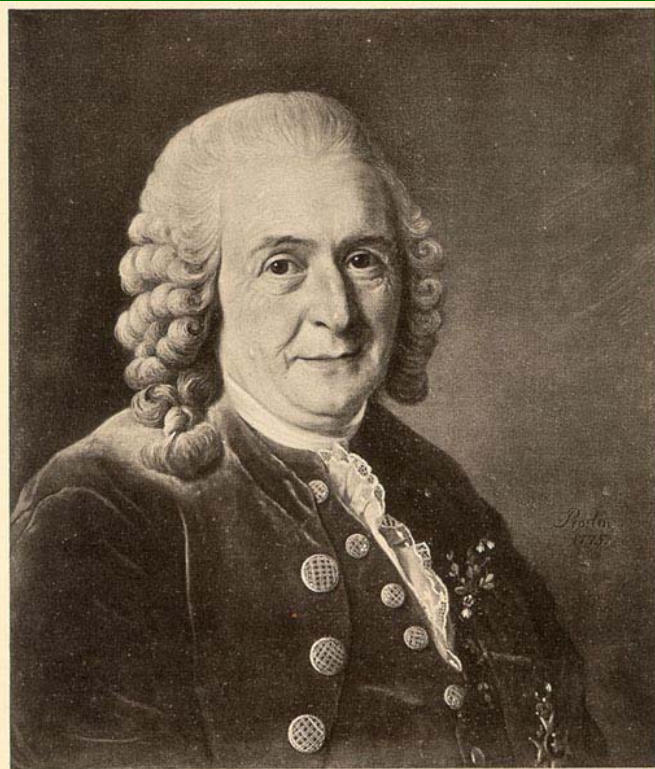
Generative slime: Ostracoderma

Spontaneous generation: Zoophytes

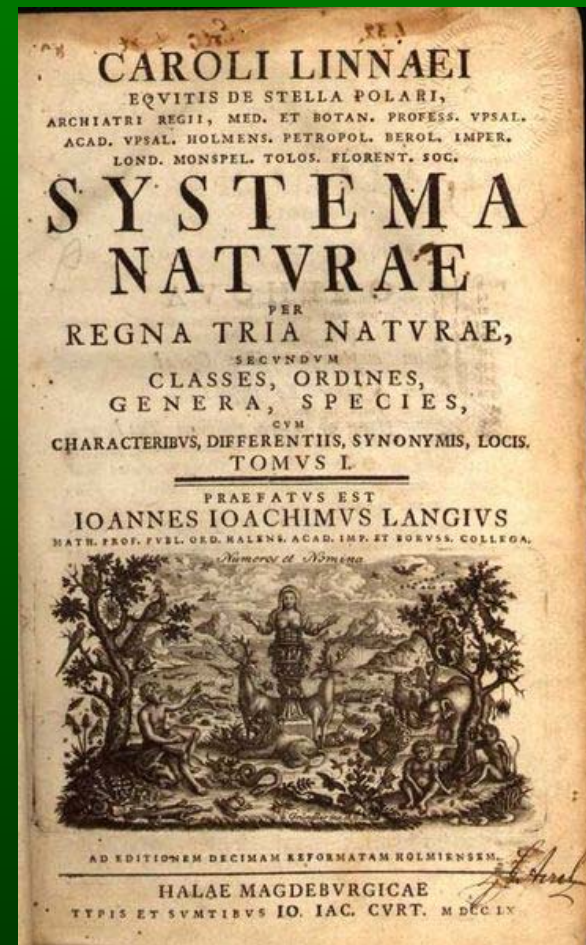
Nelson & Platnick (1981) *Systematics and Biogeography*

# Carl Linnaeus

- Swedish botanist (1700s)
- Listed all known species
- Developed scheme of classification to discover the plan of the Creator



Carl von Linné  
Painting by A. Roslin, 1775

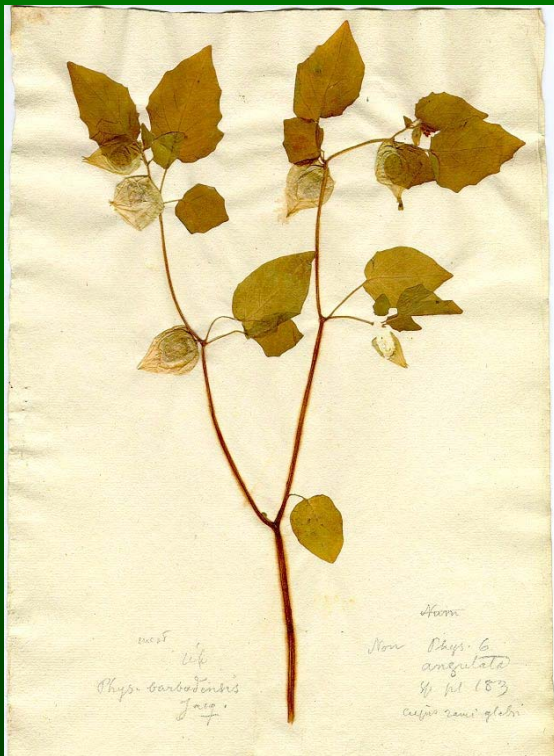


# Linnaeus' Main Contributions

## 1) Hierarchical classification scheme

Kingdom: Phylum: Class: Order: Family: Genus: Species

## 2) Binomial nomenclature



### Before Linnaeus

*physalis amno ramosissime ramis  
angulosis glabris foliis dentoserratis*

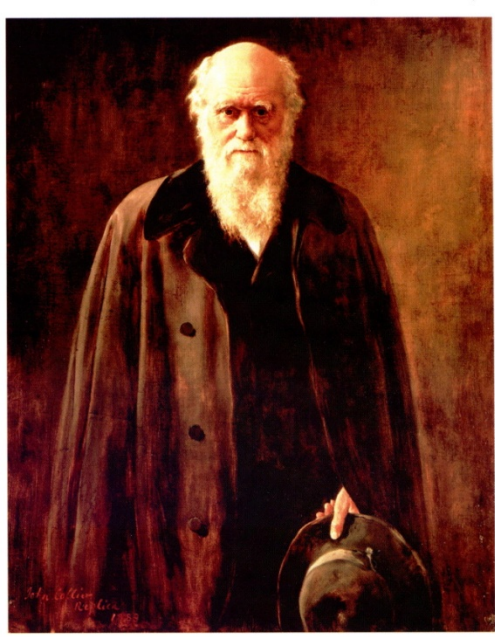
### After Linnaeus

*Physalis angulata*

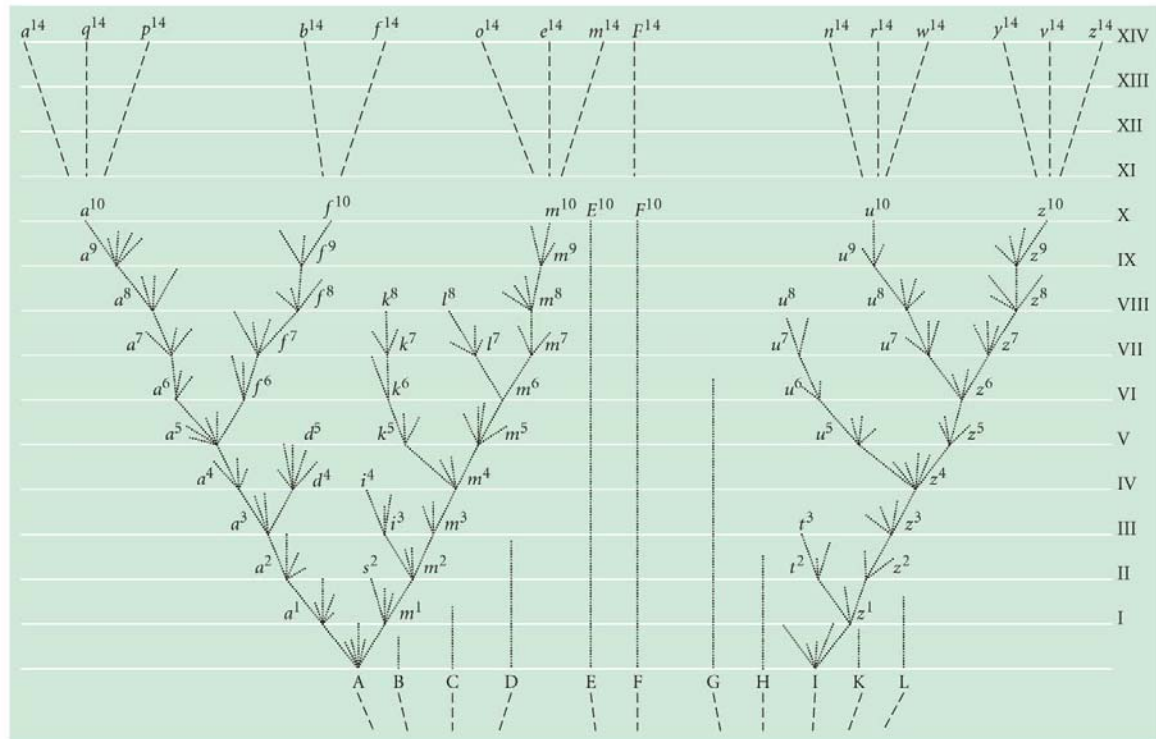
(aka Cutleaf groundcherry)

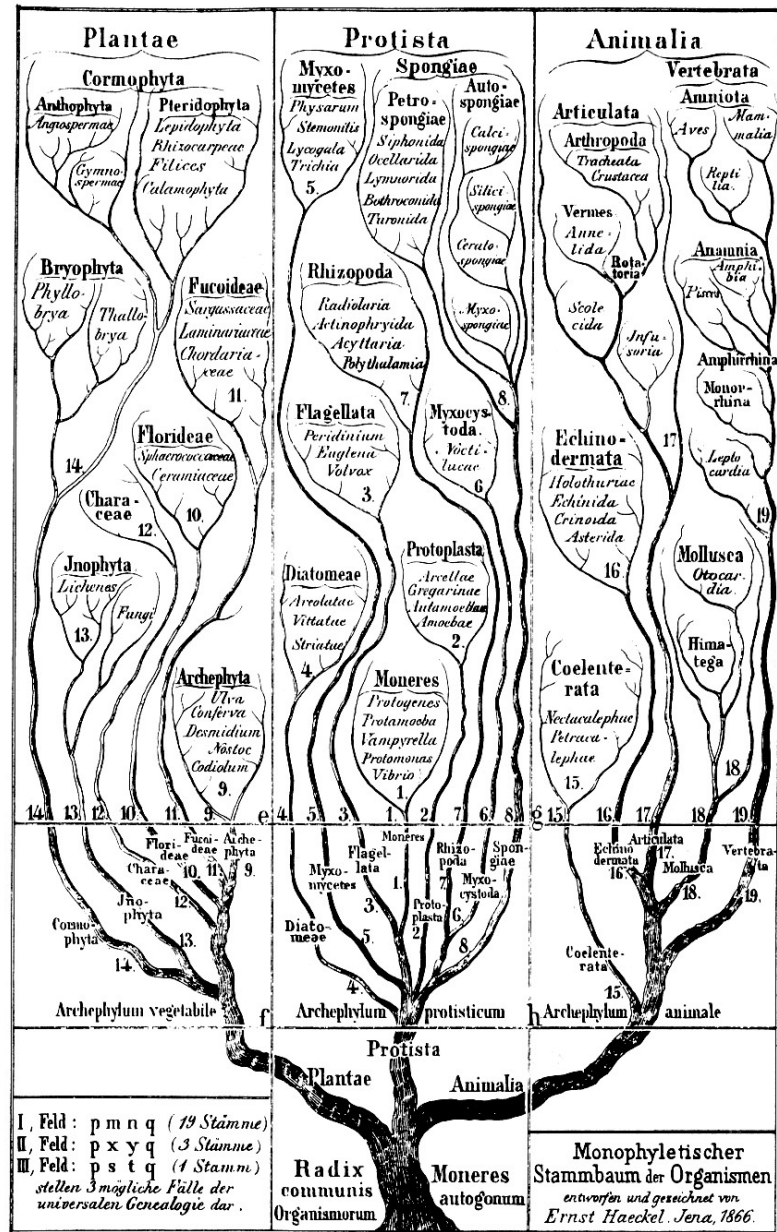
3) Originated the practice of using the ♂ - (shield and arrow) Mars and ♀ - (hand mirror) Venus **glyphs as the symbol for male and female.**

# Charles Darwin



- Species evolved from common ancestors.
- Concept of closely related species being more recently diverged from a common ancestor. Therefore taxonomy *might* actually represent phylogeny!



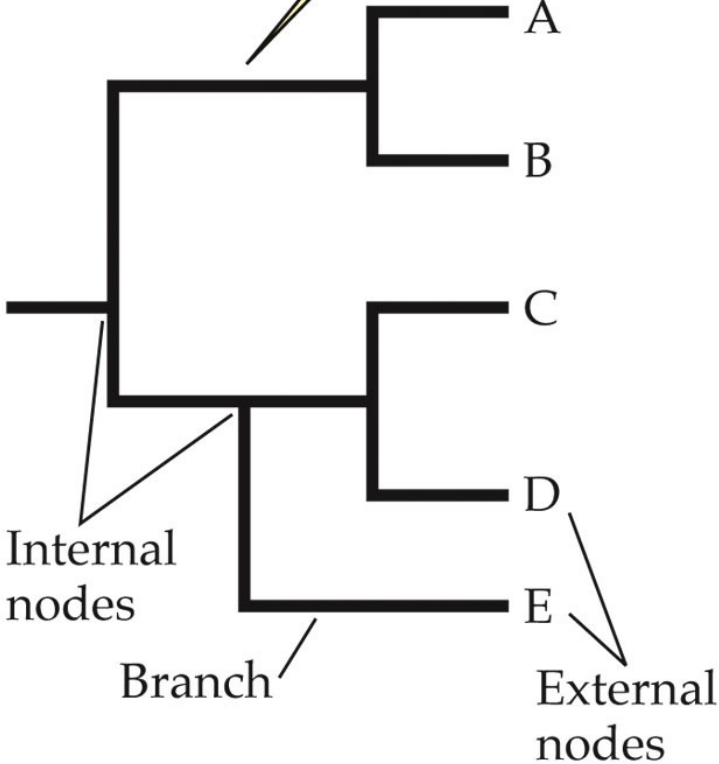
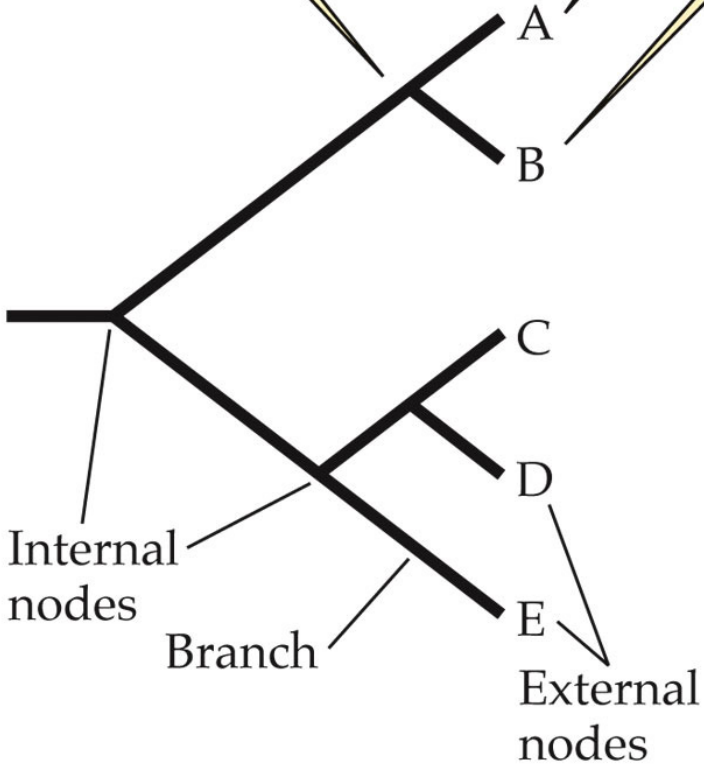


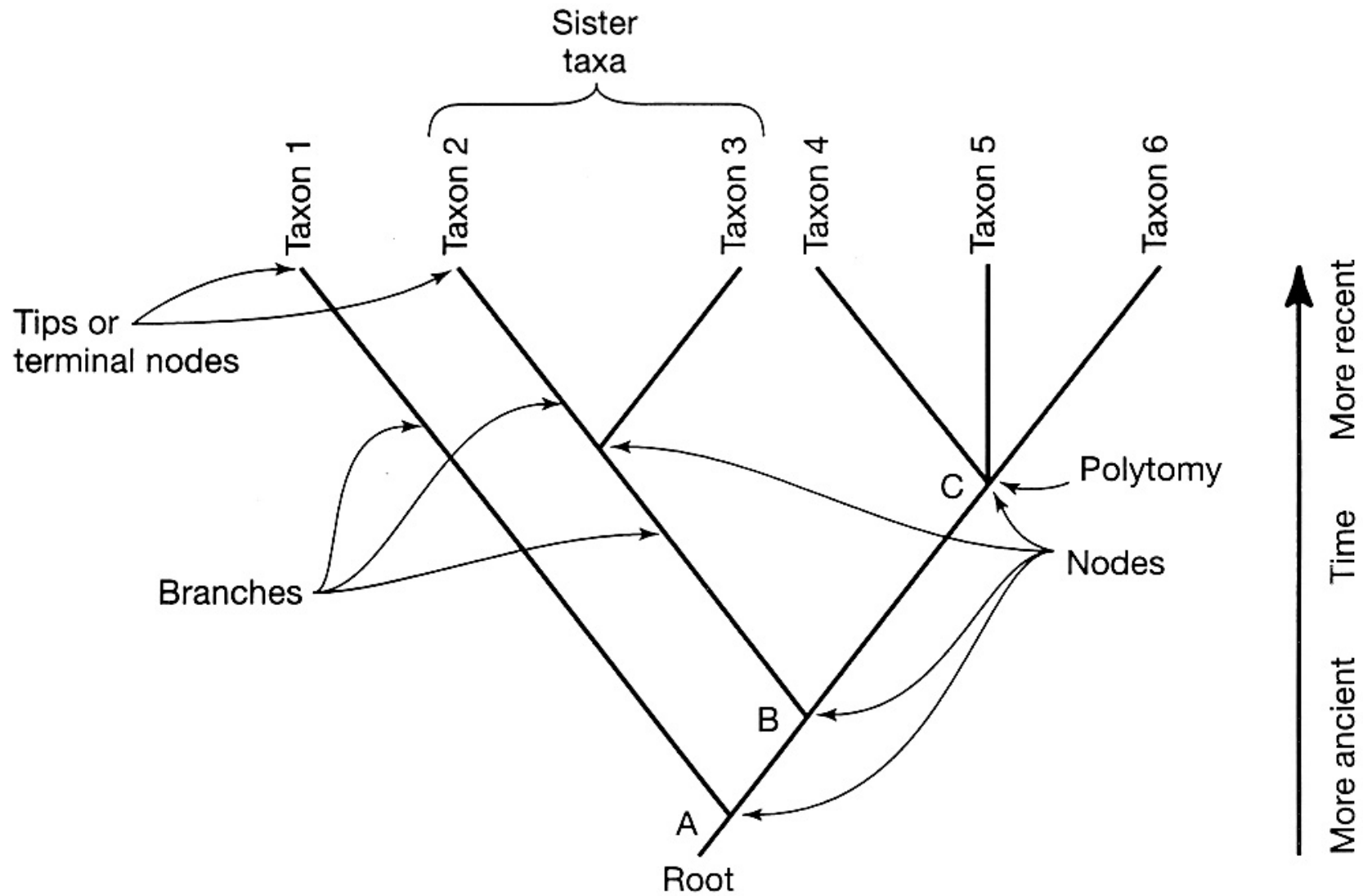
The phylogeny and classification of life as proposed by Haeckel (1866).

**Internal nodes**  
represent ancestor  
species...

...**external nodes**  
represent extant,  
known species...

...**branch** lengths repre-  
sent evolutionary distance  
between species.



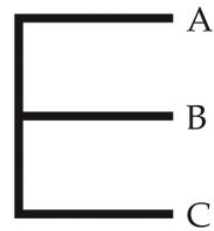
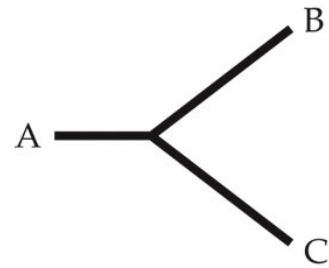


**Parts of a phylogenetic tree** In this hypothetical phylogeny, node A defines a monophyletic group, or clade, comprising taxa 1–6. Node B defines a monophyletic group comprising taxa 2–6. Node C defines a monophyletic group comprising taxa 4–6.

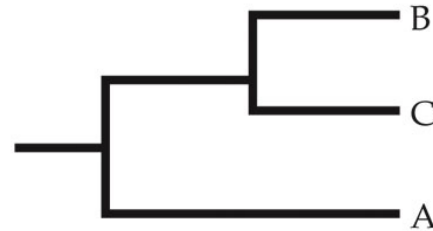
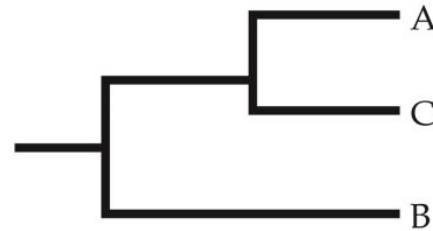
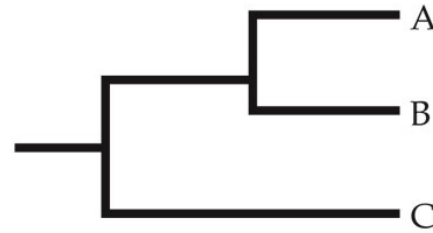


# Trees - Rooted and Unrooted

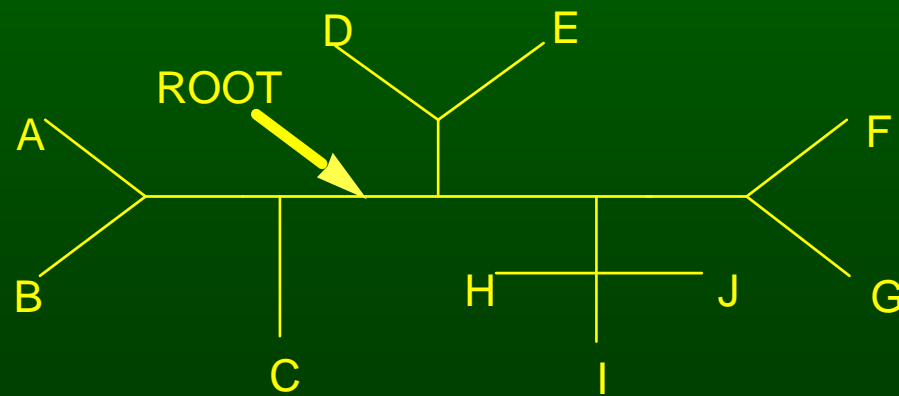
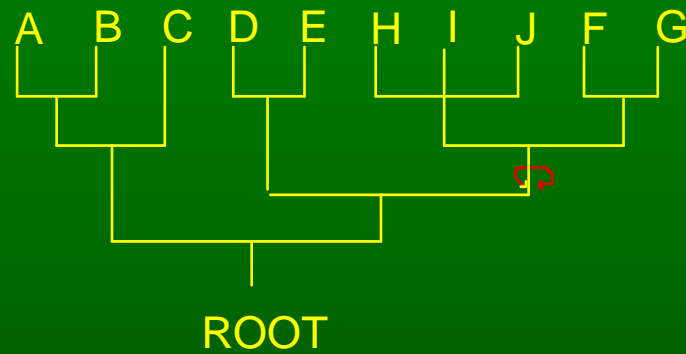
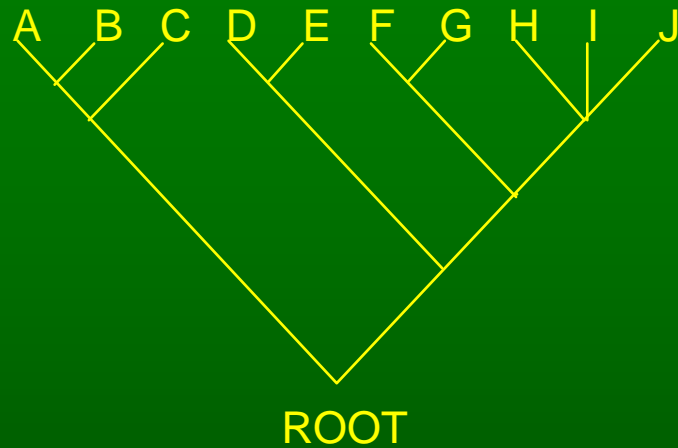
(A) Unrooted trees



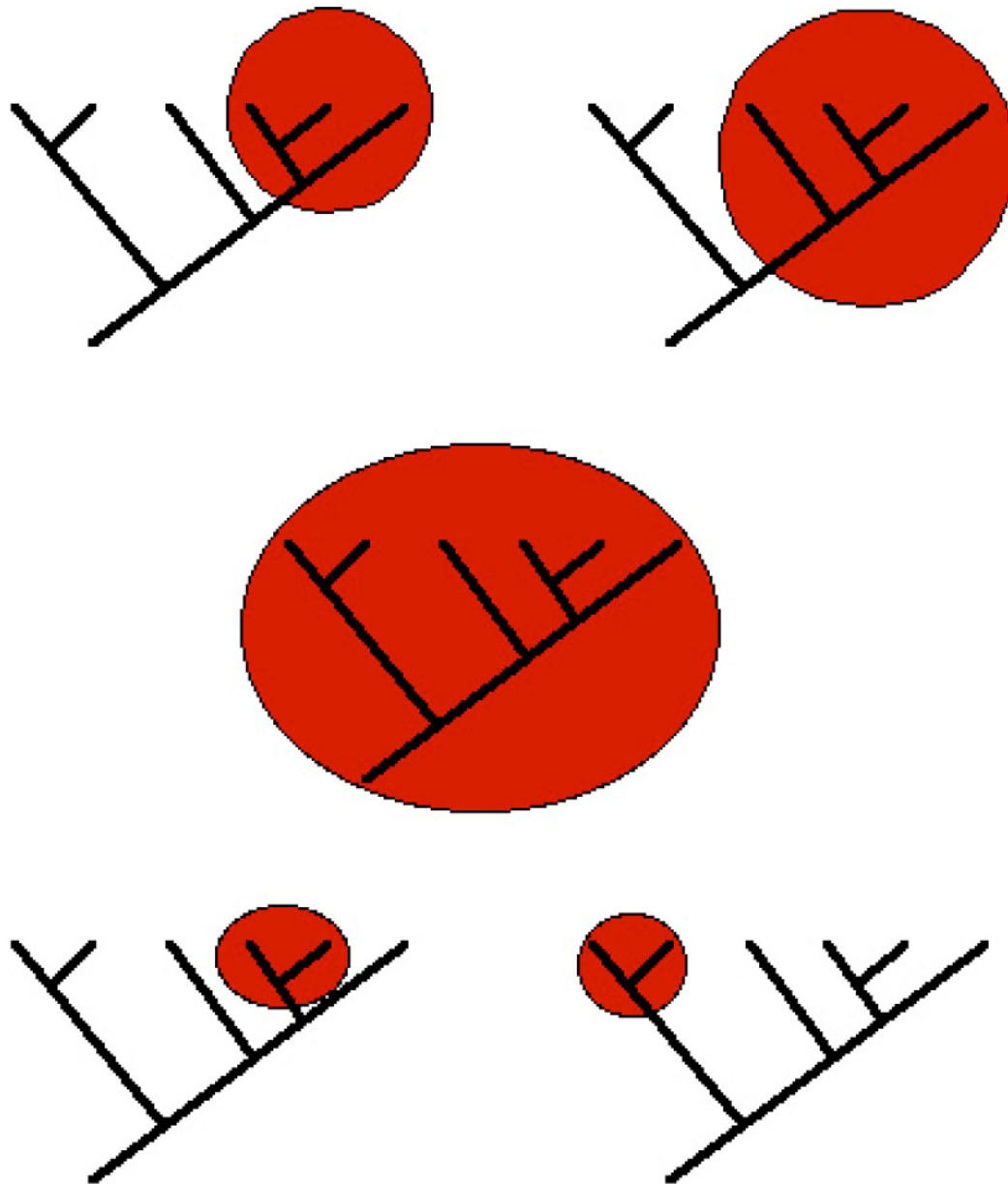
(B) Rooted trees



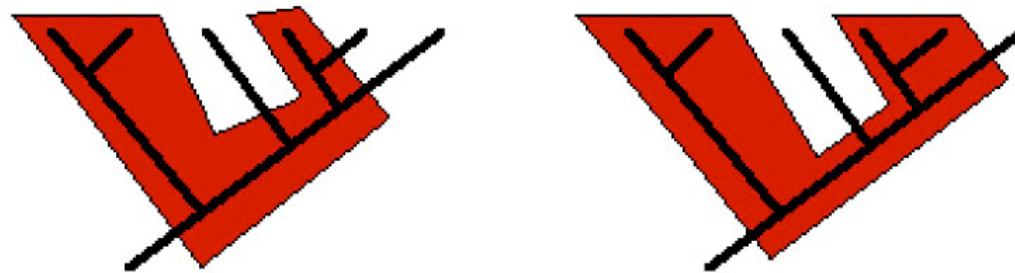
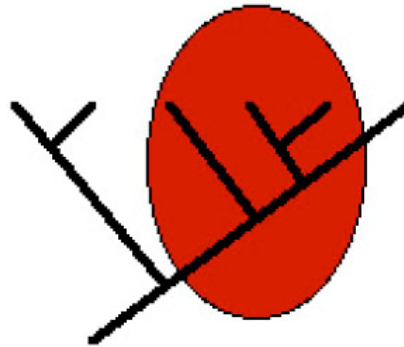
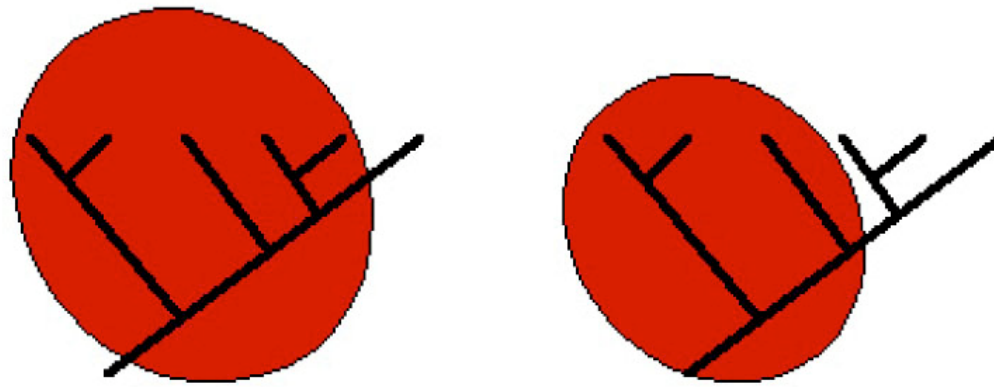
# Trees - Rooted and Unrooted



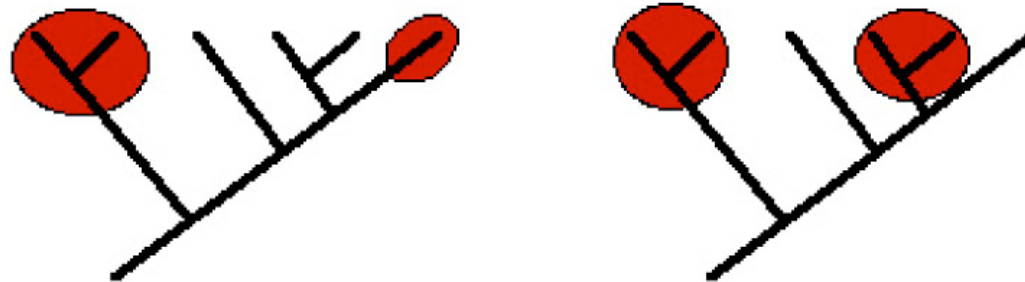
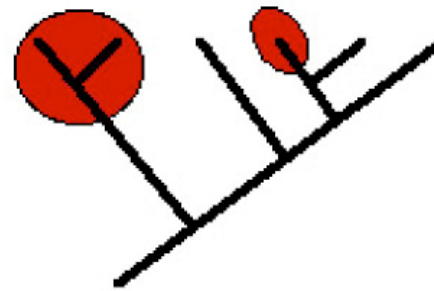
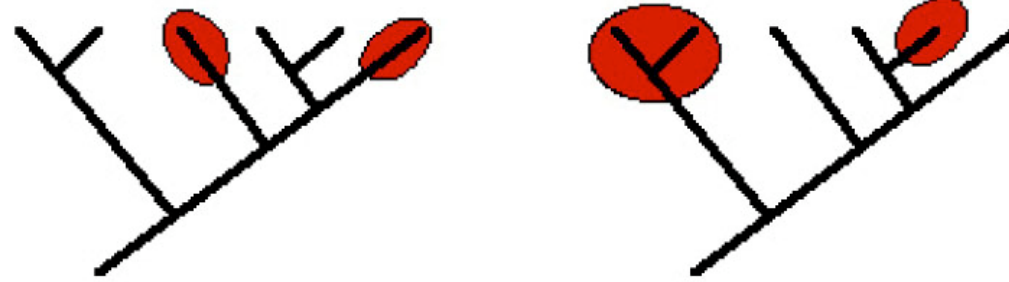
# Monophyletic Groups

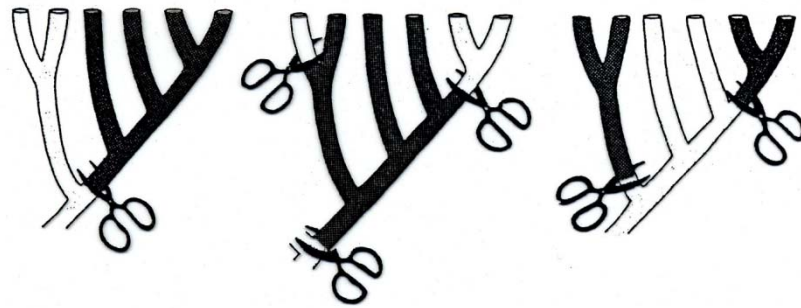


# Paraphyletic Groups



# Polyphyletic Groups





Monophyletic

one and only one cut

One branch

Paraphyletic

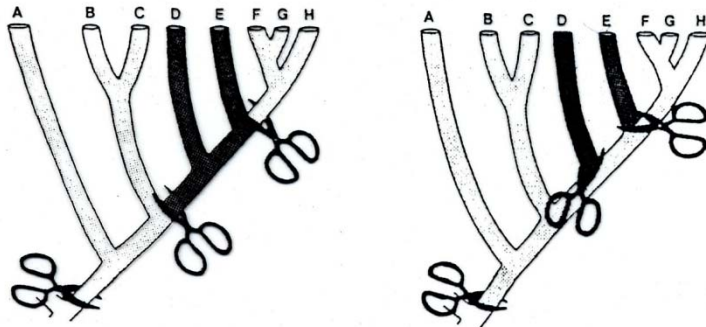
one cut below the group and one or more cuts higher up

A piece of a branch

Polyphyletic

more than one cut below the group

More than one piece of a branch

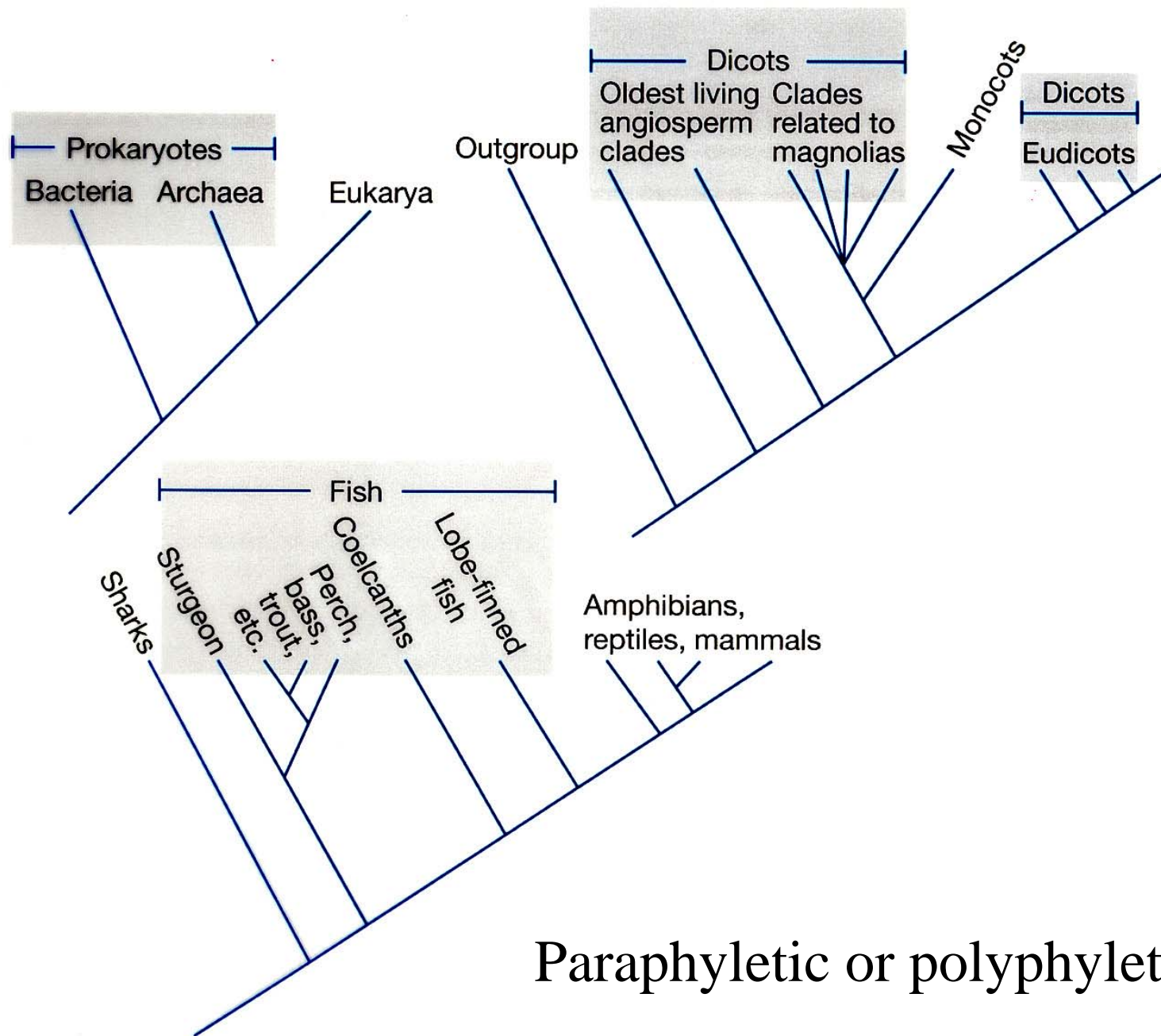


In the <sup>right</sup> ~~left~~-hand classification the group A-B-C-F-G-H is polyphyletic. It consists of two pieces of the tree diagram separated by two cuts above the stem line. The group D-E is paraphyletic. In the ~~right~~ <sup>left</sup>-hand classification, the two monophyletic groups D and E are ranked equal with A-B-C-F-G-H, which then becomes paraphyletic. It is a clade with two subclades taken out but still a coherent piece of the diagrammatic tree.

**Monophyletic:** A group composed of a collection of organisms, including the most recent common ancestor of all those organisms and all the descendants of that most recent common ancestor. A monophyletic taxon is also called a clade.

**Paraphyletic:** A group composed of a collection of organisms, including the most recent common ancestor of all those organisms. Unlike a monophyletic group, a paraphyletic group does not include all the descendants of the most recent common ancestor.

**Polyphyletic:** A group composed of a collection of organisms in which the most recent common ancestor of all the included organisms is not included, usually because the common ancestor lacks the characteristics of the group. Polyphyletic groups are considered "unnatural", and usually are reclassified once they are discovered.



Paraphyletic or polyphyletic?

# Character Evolution

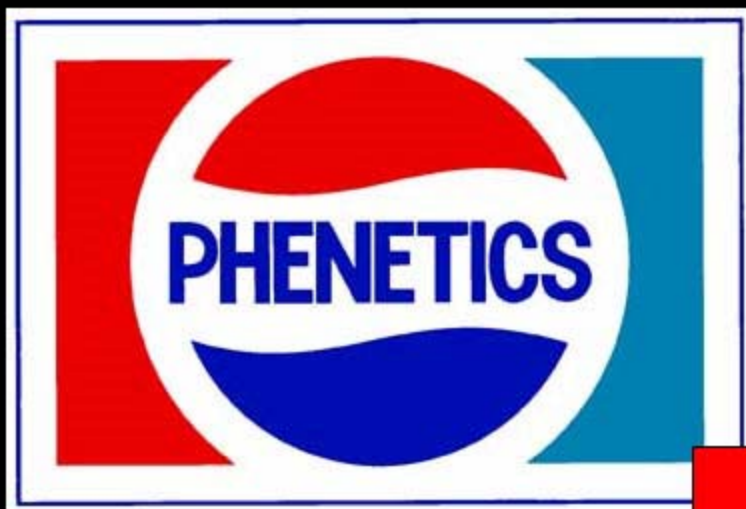
- **Heritable** changes (in morphology, gene sequences, etc.) produce different character states.
- **Variability** in terms of similarities and differences in character states provide the basis for inferring phylogeny (i.e., provide evidence of relationships).
- **Independent** occurrence of different character states is a necessary requirement. The more frequent the better the resolution.



# Characters and Character States

- Organisms comprise sets of features.
- When organisms/taxa differ with respect to a feature (e.g., its presence or absence or different nucleotide bases at specific sites in a sequence) the different conditions are called *character states*.
- The collection of character states with respect to a feature constitute a *character*.

Theoretical advances in areas such as **phenetics** and **cladistics** have provided improved, empirical methods for framing and testing evolutionary hypotheses.



“overall similarity”

R.R. Sokal & P.H.A. Sneath (1963),  
*Principles of numerical taxonomy*

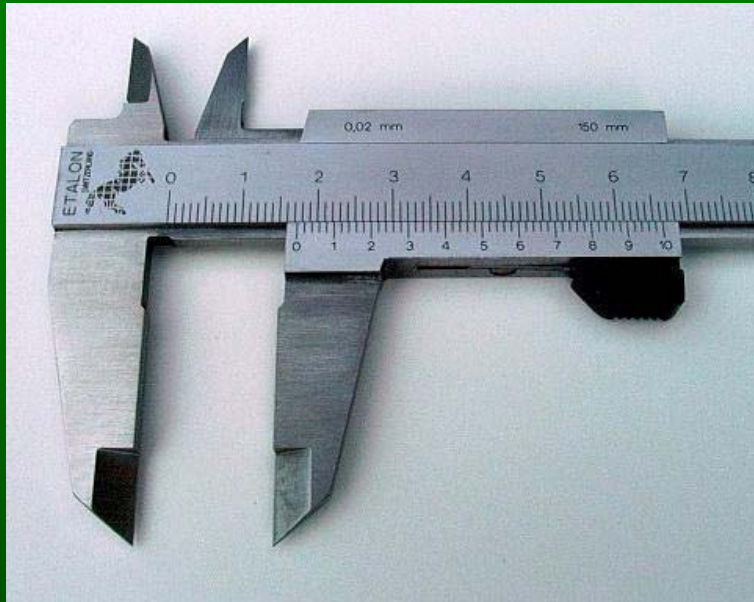
W. Hennig (1963), *Phylogenetic Systematics*

“shared, derived characters”



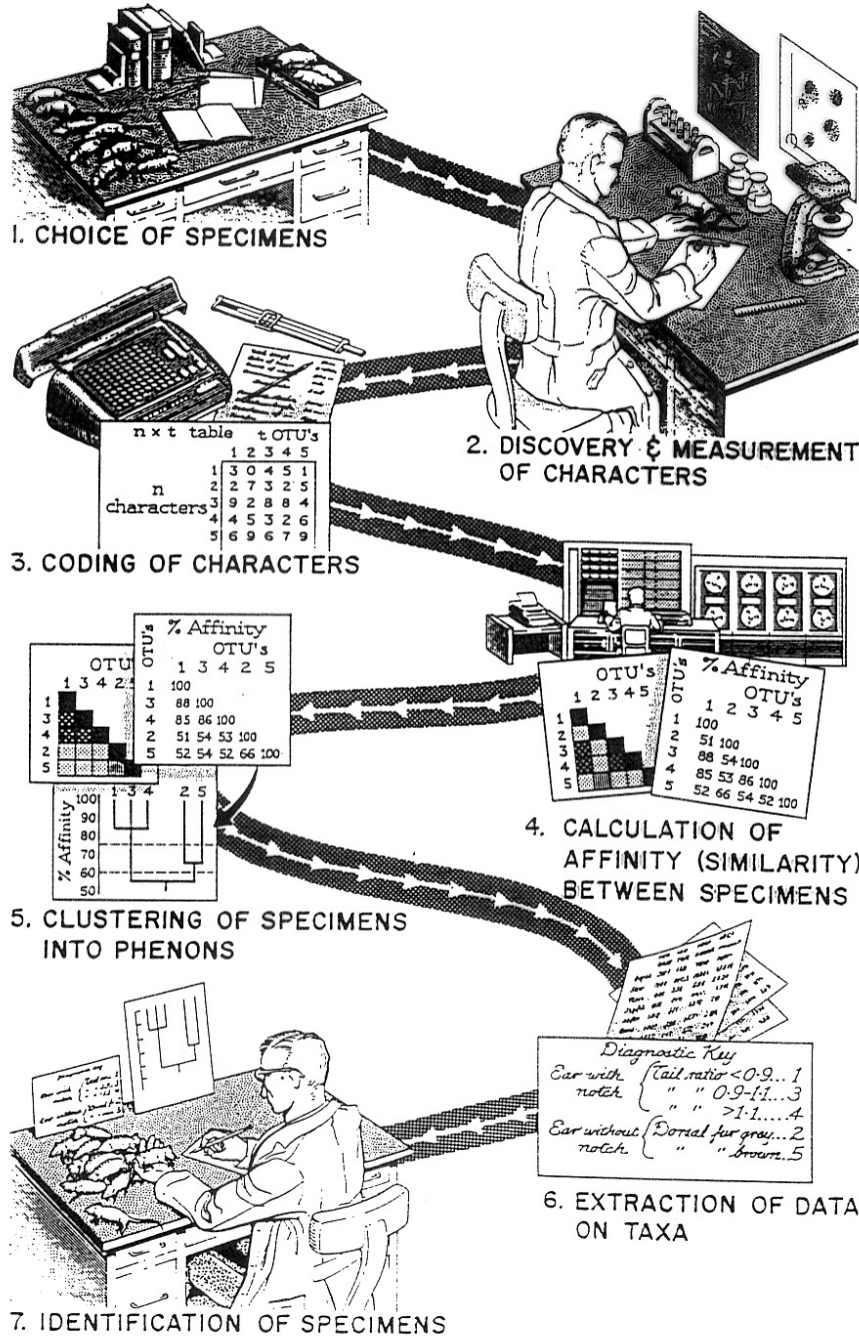
(Designs by D. Maddison)

# Numerical Taxonomy aka Phenetics



- Based on a numerical comparison of similarities and differences of a group of organisms.
- Gave a new sense of rigor.  
(eventually used computers!)
- The source of data: characters and character states.

# A FLOW CHART OF NUMERICAL TAXONOMY



# The Terminology of **Cladistics**

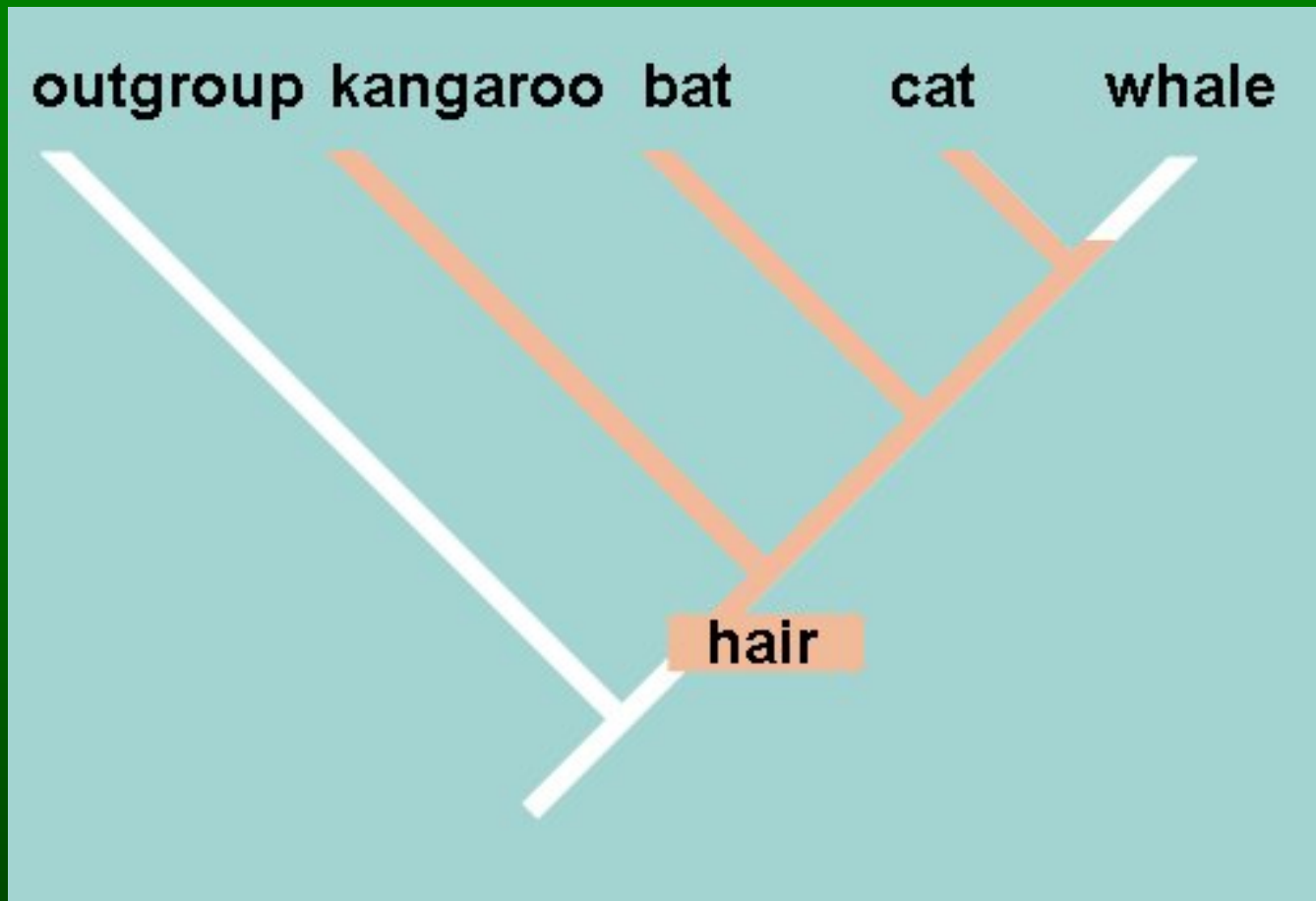
**Synapomorphy:** a shared, derived character state.  
These are the basis of cladistics!

**Autapomorphy:** a unique, derived character state.

**Symplesiomorphy:** a shared, ancestral character state.

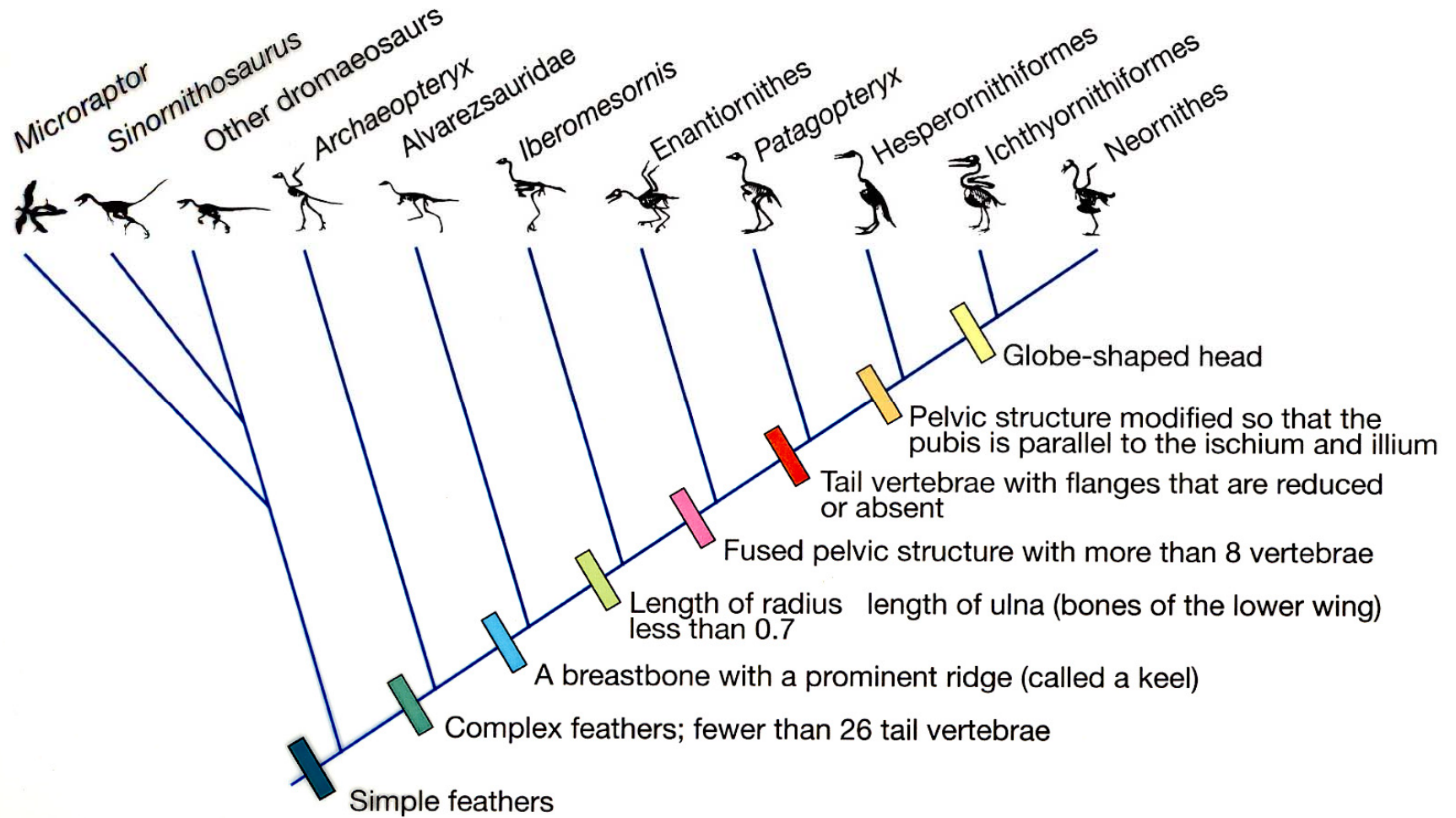
Only synapomorphies are used in order to avoid or minimize homoplasy.

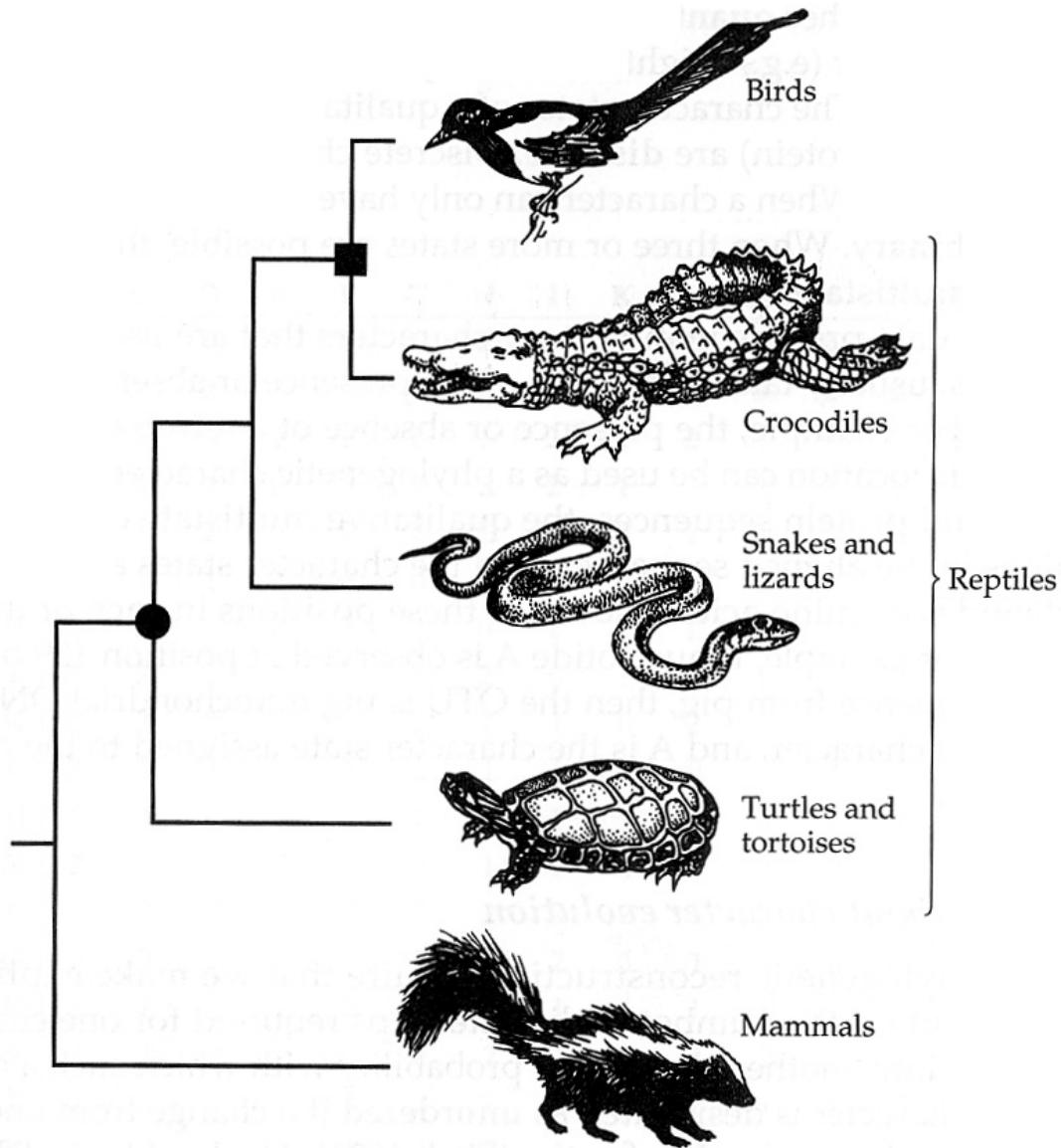
# Symplesiomorphy



Within this taxon set, hair is a **shared ancestral character** (**symplesiomorphy**) of non-cetaceans. It would *not* indicate a close relationship of marsupial and placental mammals.

# Cladogram of Mesozoic Birds





**FIGURE 5.8** Cladogram of birds, reptiles, and mammals. The reptiles do not constitute a natural clade, since their most recent common ancestor (black circle) also gave rise to the birds, which are not included in the original definition of reptile. Birds and crocodiles, on the other hand, constitute a natural clade (Archosauria), since they share a common ancestor (black box) that is not shared by any non-archosaurian organism.



# Evolutionary Taxonomy



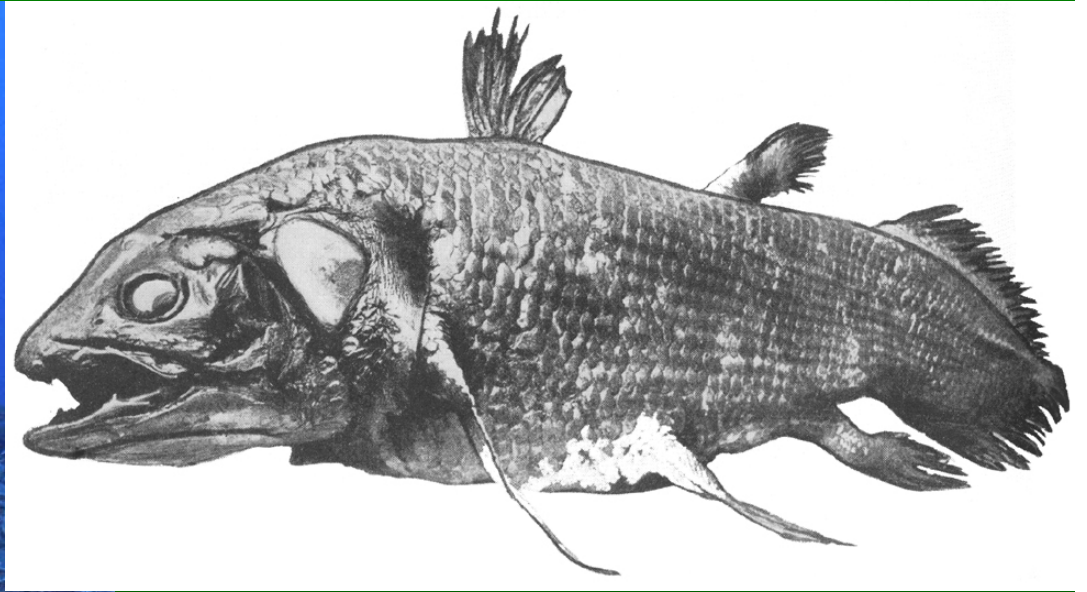
- Based on an overall impression of similarities and differences, borne out of years of study of a group of organisms.
- Seeks to classify organisms using a combination of phylogenetic relationship and overall similarity. It thus allows for paraphyletic taxa.

# Why Might Similarities and Differences Be Unreliable for Inferring Phylogenies?



**Homology** vs. **homoplasy**  
(e.g., convergence, reversal, parallelism)

# Why Might Similarities and Differences Be Unreliable for Inferring Phylogenies?



- Unequal rates of evolution



# Why Might Similarities and Differences Be Unreliable for Inferring Phylogenies?

**Homoplasy:** similarity that is not homologous (not due to common ancestry). Can provide misleading evidence of phylogenetic relationships (if mistakenly interpreted as homology). Remember same as analogous structures or traits.

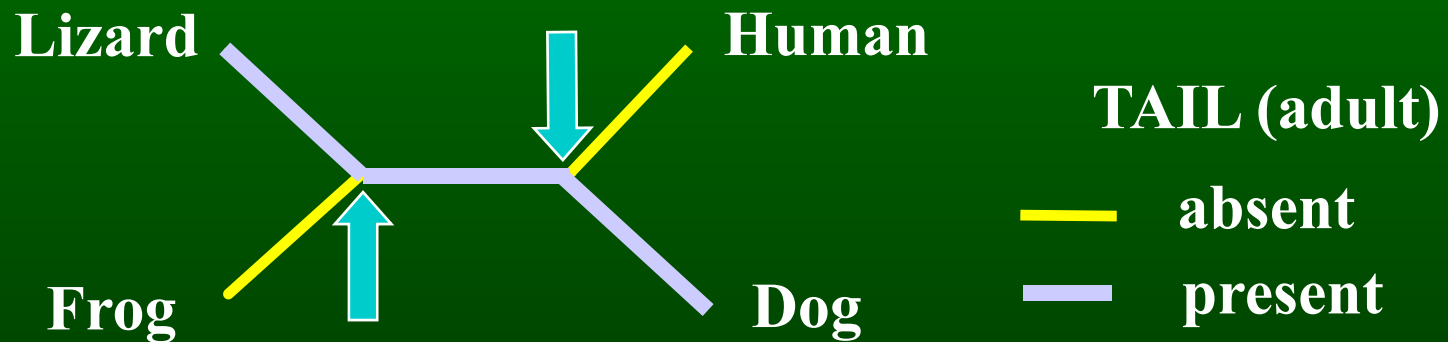
**Parallelism** refers to the independent evolution of the same derived trait via the same developmental changes.

**Convergence** refers to superficially similar traits that have a distinct developmental basis.

**Reversal** refers to return to the ancestral state. More problematic when considering molecular character data.

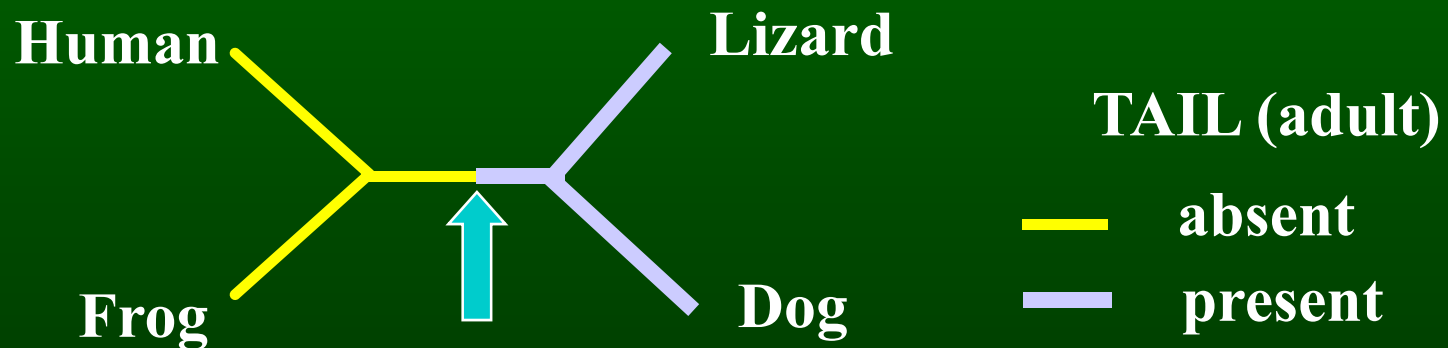
# Homoplasy - independent evolution

- Loss of tails evolved independently in humans and frogs - there are two steps on the true tree.



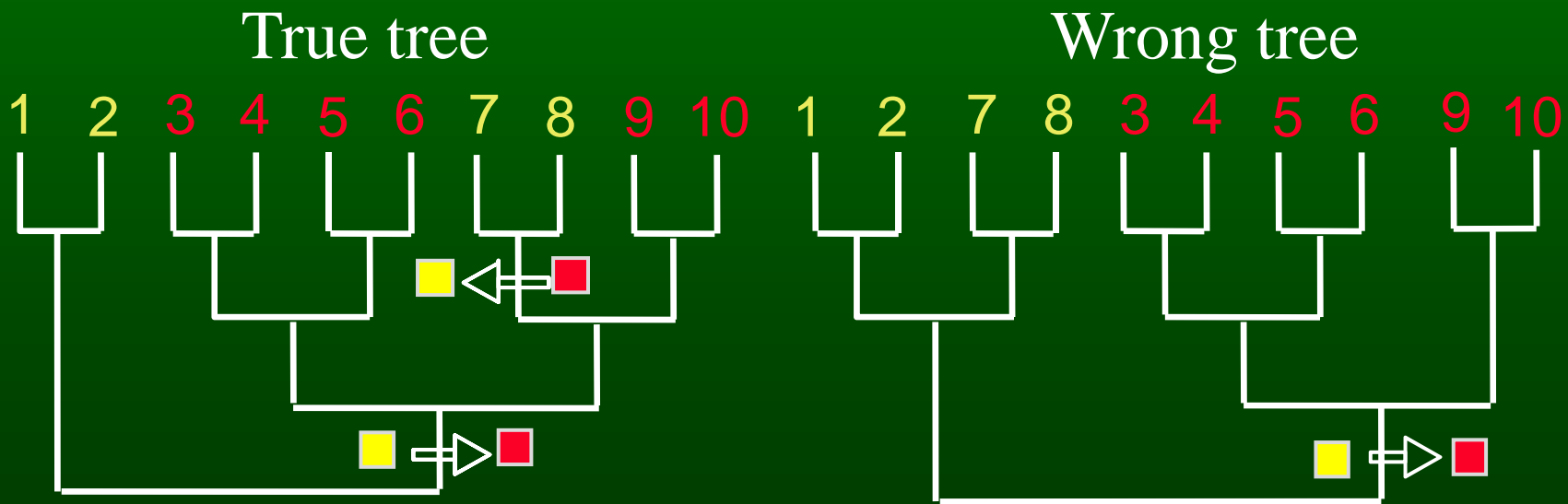
# Homoplasy - misleading evidence of phylogeny

- If misinterpreted as homology, the absence of tails would be evidence for a wrong tree: grouping humans with frogs and lizards with dogs.



# Homoplasy - Reversal

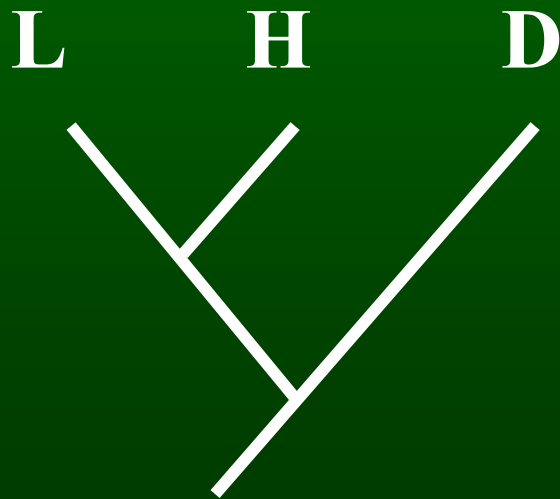
- Reversals are evolutionary changes back to an ancestral condition.
- As with any homoplasy, reversals can provide misleading evidence of relationships.



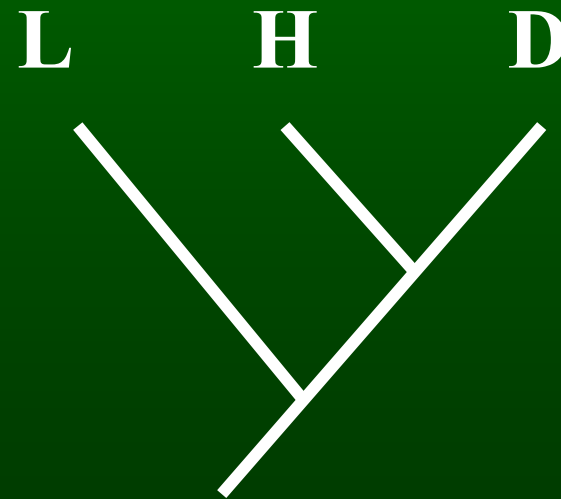
## Ancestral vs. derived character states

For lizards, dolphins, and horses, the **ancestral character state** is 4 limbs. A **derived character state** is the presence of mammary glands.

Phylogeny based on  
number of limbs.



Phylogeny based on  
mammary glands.



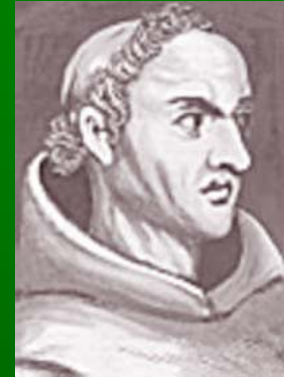


How do we determine if a trait  
(character state) is ancestral or derived?

- **Fossil record**
- **Outgroup analysis**
- **Evolutionary Modeling**

# How Do We Improve on Cladistics?

- **Ockham's Razor** - “entities should not be multiplied unnecessarily” aka the KISS system.



- **Parsimony** - the best estimate of phylogeny requires the fewest evolutionary changes to explain the distribution of shared derived character states (synapomorphies) in the organisms.

- **Maximum Parsimony** – exhaustive examination of all possible trees, where trees with the fewest inferred changes between character states are the maximum parsimony trees.

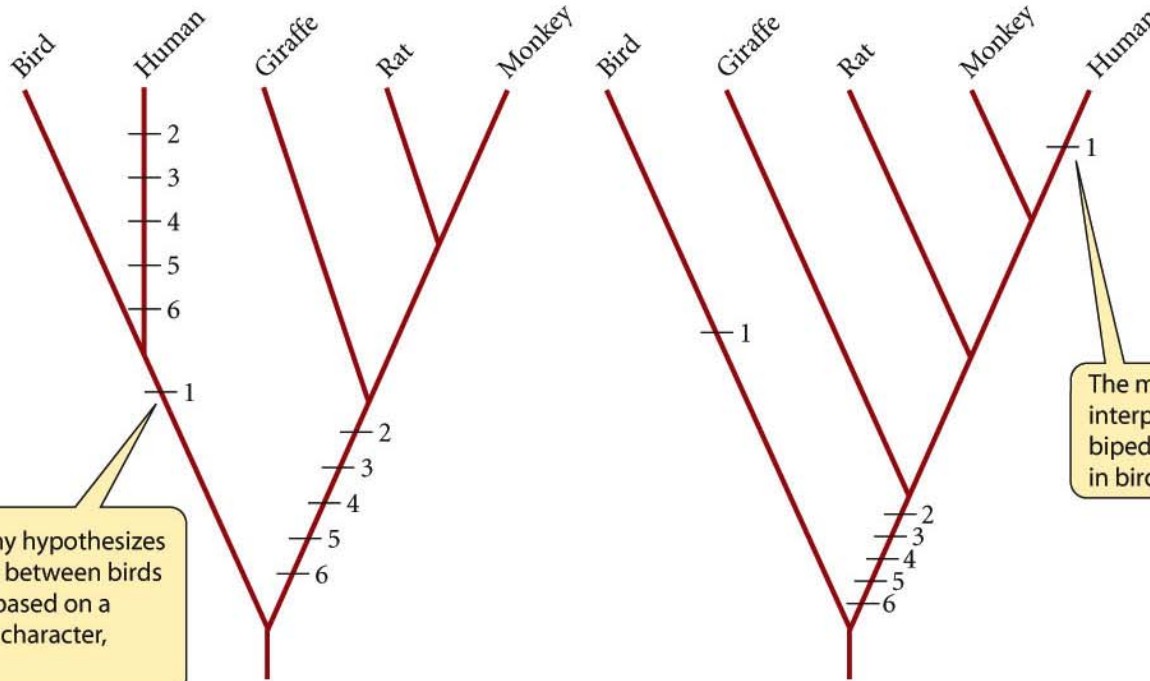
# Inferring a phylogeny by parsimony

(A) Hypothetical phylogeny

(B) Accepted phylogeny

Character key:

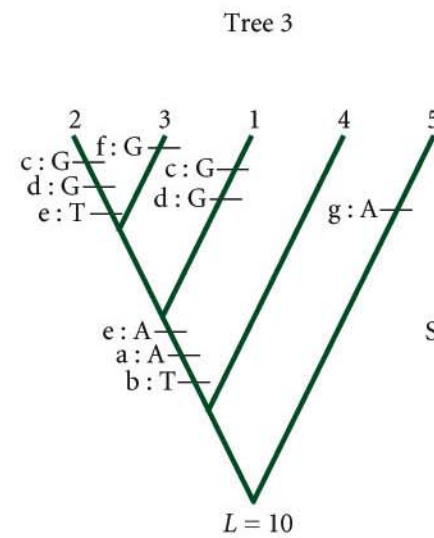
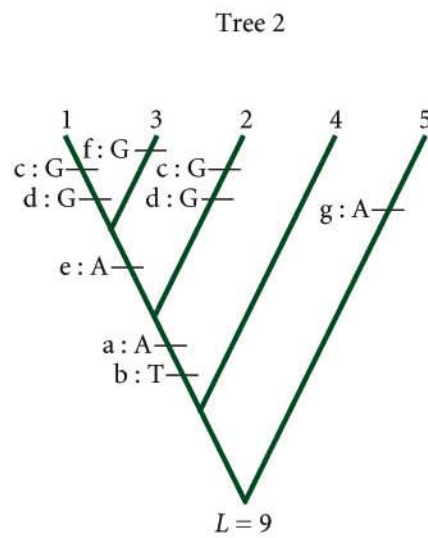
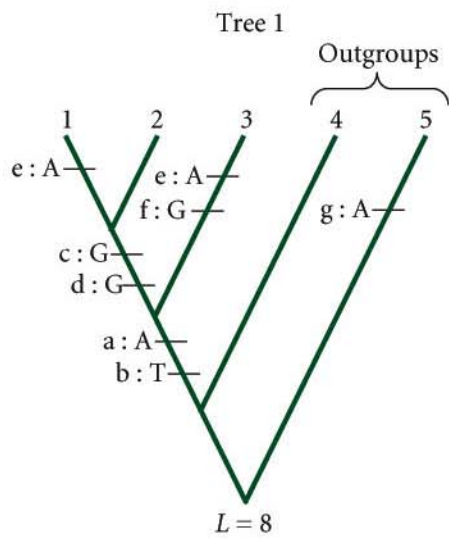
1. Bipedal
2. Hair
3. Milk
4. Diaphragm
5. Molar teeth
6. Three middle-ear bones



This phylogeny hypothesizes a relationship between birds and humans based on a single shared character, bipedalism.

The more parsimonious interpretation is that the bipedalism evolved separately in birds and in humans.

# Inferring a phylogenetic tree by the method of maximum parsimony



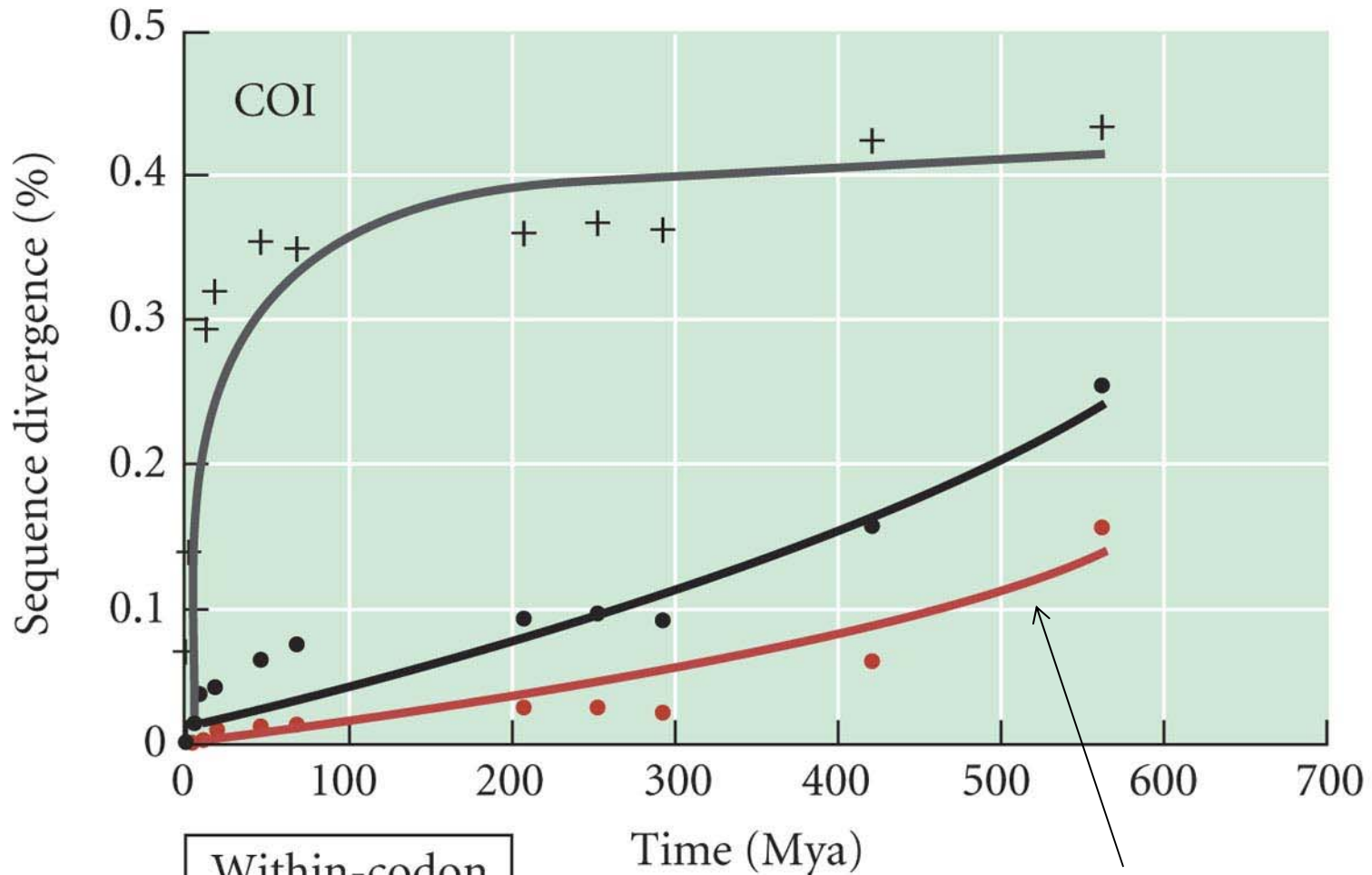
Species	Character (nucleotide base at given site)						
	a	b	c	d	e	f	g
1	A	T	G	G	A	C	T
2	A	T	G	G	T	C	T
3	A	T	T	C	A	G	T
4	C	G	T	C	T	C	T
5	C	G	T	C	T	C	A

# Phylogenetic Analysis: Other Algorithms using various Evolutionary Models

- Maximum likelihood
- Neighbor joining (Distance)
- Quartet Puzzling (Distance)
- Bayesian

# Difficulties in Phylogenetics

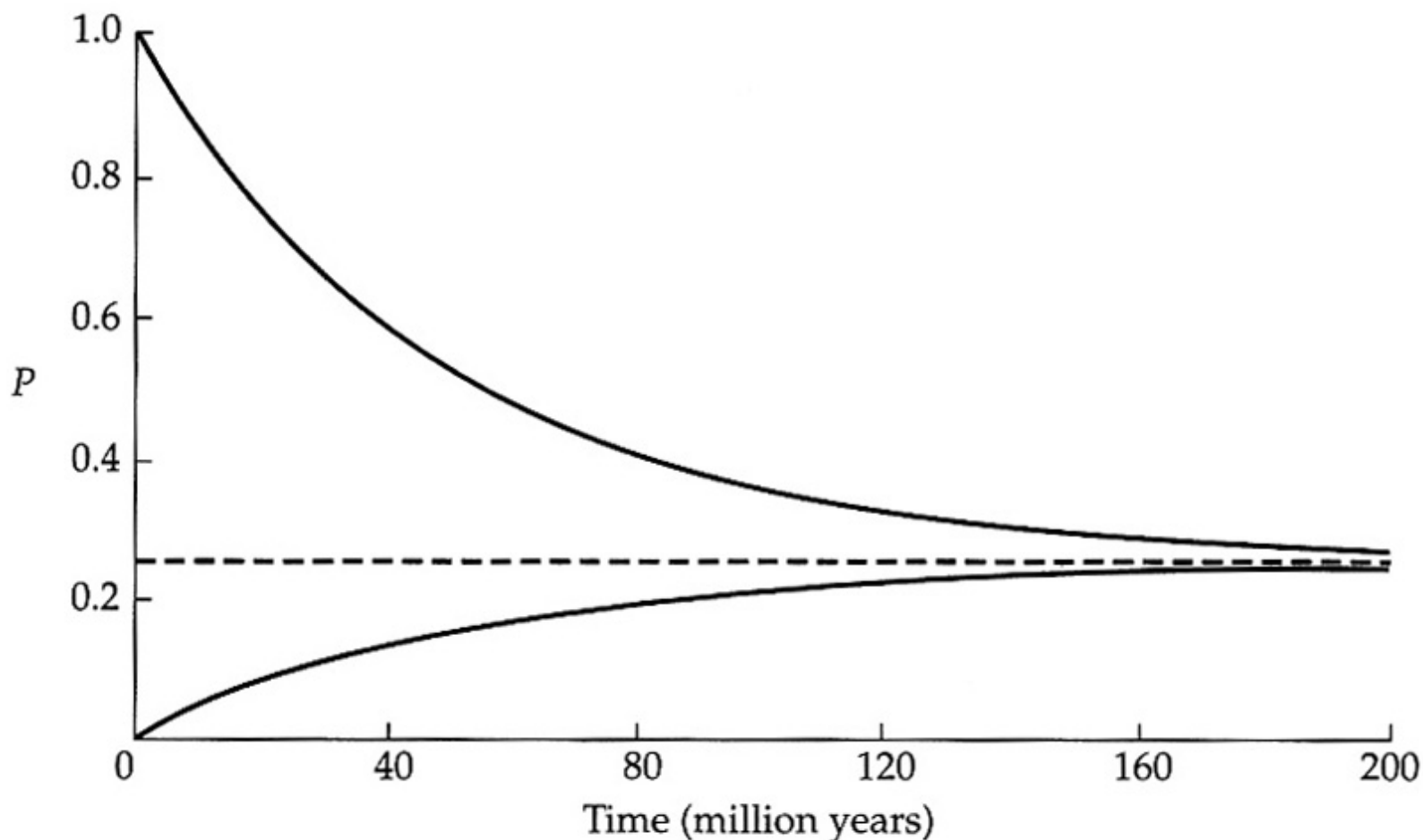
- **Scoring characters can be challenging**
- **Homoplasy may be common**
- **Evolutionary change may erase signs of evolutionary history**
- Rapid radiation from a common ancestor
- Gene trees may imply the wrong phylogeny
- Hybridization and horizontal gene transfer



Within-codon  
base position

- First
- Second
- + Third

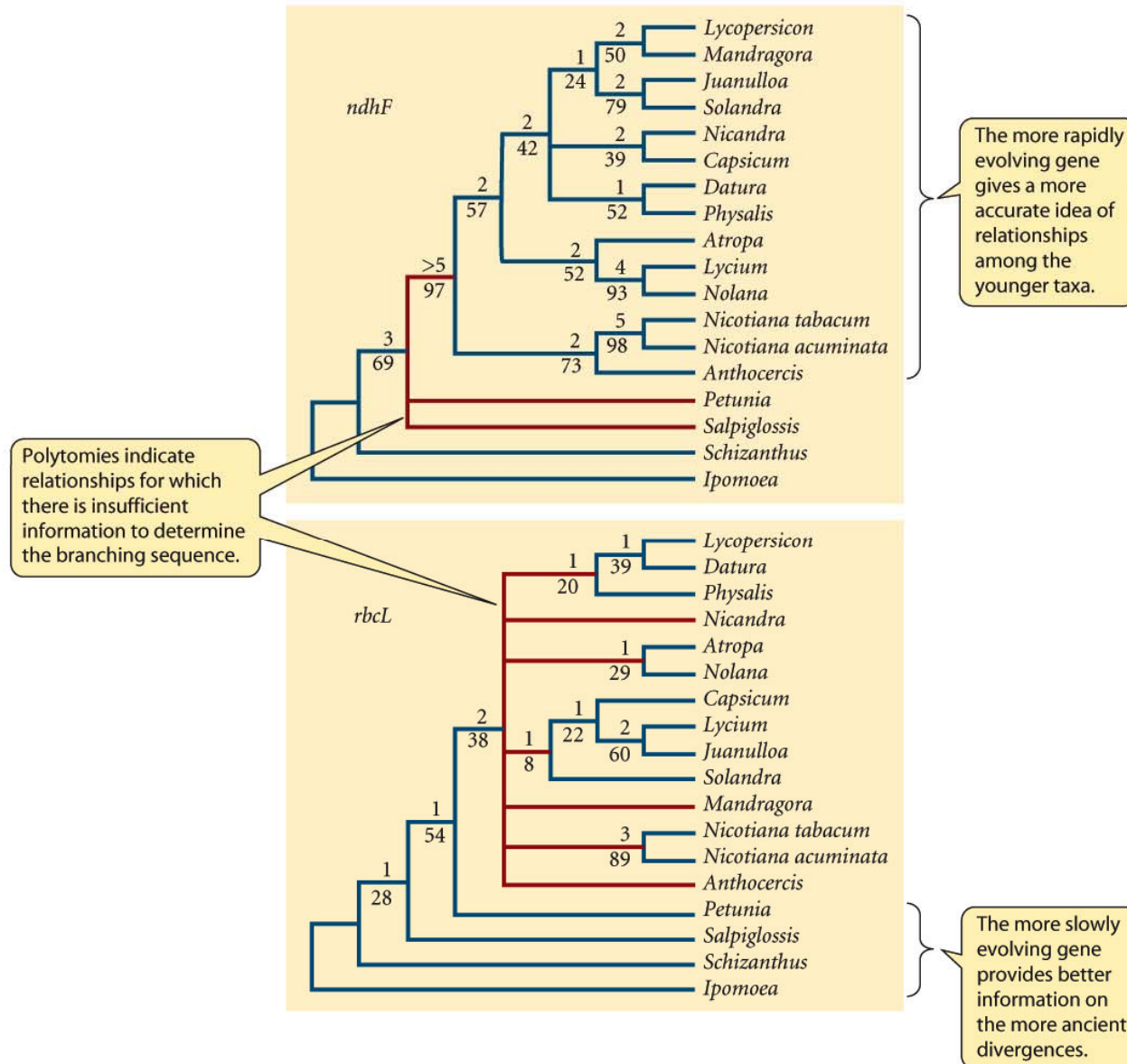
Each position is not equal!



**FIGURE 3.3** Temporal changes in the probability,  $P$ , of having a certain nucleotide at a position starting with either the same nucleotide (upper line) or with a different nucleotide (lower line). The dashed line denotes the equilibrium frequency ( $P = 0.25$ ).  $\alpha = 5 \times 10^{-9}$  substitutions per site per year.



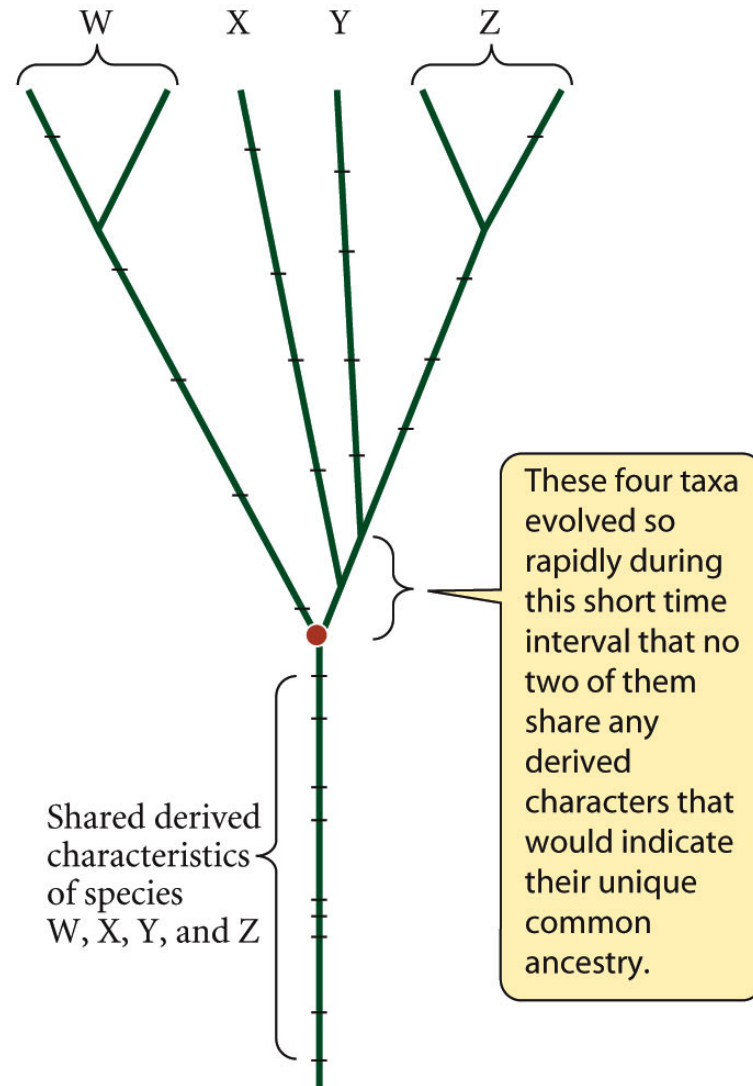
# Chloroplast gene *ndhF* has evolved more rapidly than *rbcL*



Chloroplast gene *rbcL* provides better resolution for ancient divergences

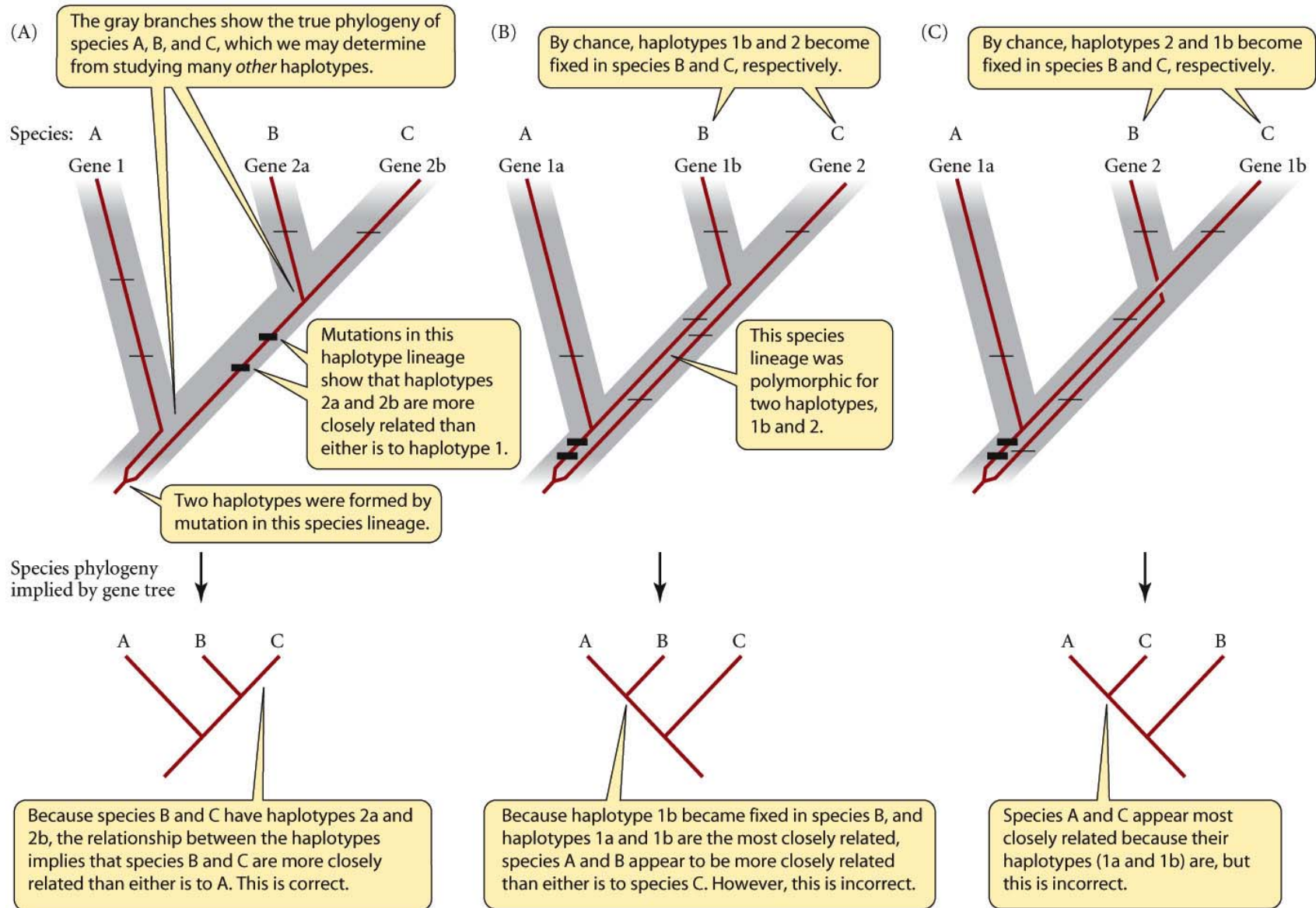
# Difficulties in Phylogenetics

- Scoring characters can be challenging
- Homoplasy may be common
- Evolutionary change may erase signs of evolutionary history
- **Rapid radiation from a common ancestor**
- **Gene trees may imply the wrong phylogeny**
- Hybridization and horizontal gene transfer



Rapid evolutionary radiation driven by adaptive radiation

# Gene trees may or may not reflect the phylogeny of the species.



**EVOLUTION 2e, Figure 2.20**

# A genealogy of cichlid fishes.

## MITOCHONDRIAL-GENE TREES VERSUS NUCLEAR-GENE TREES

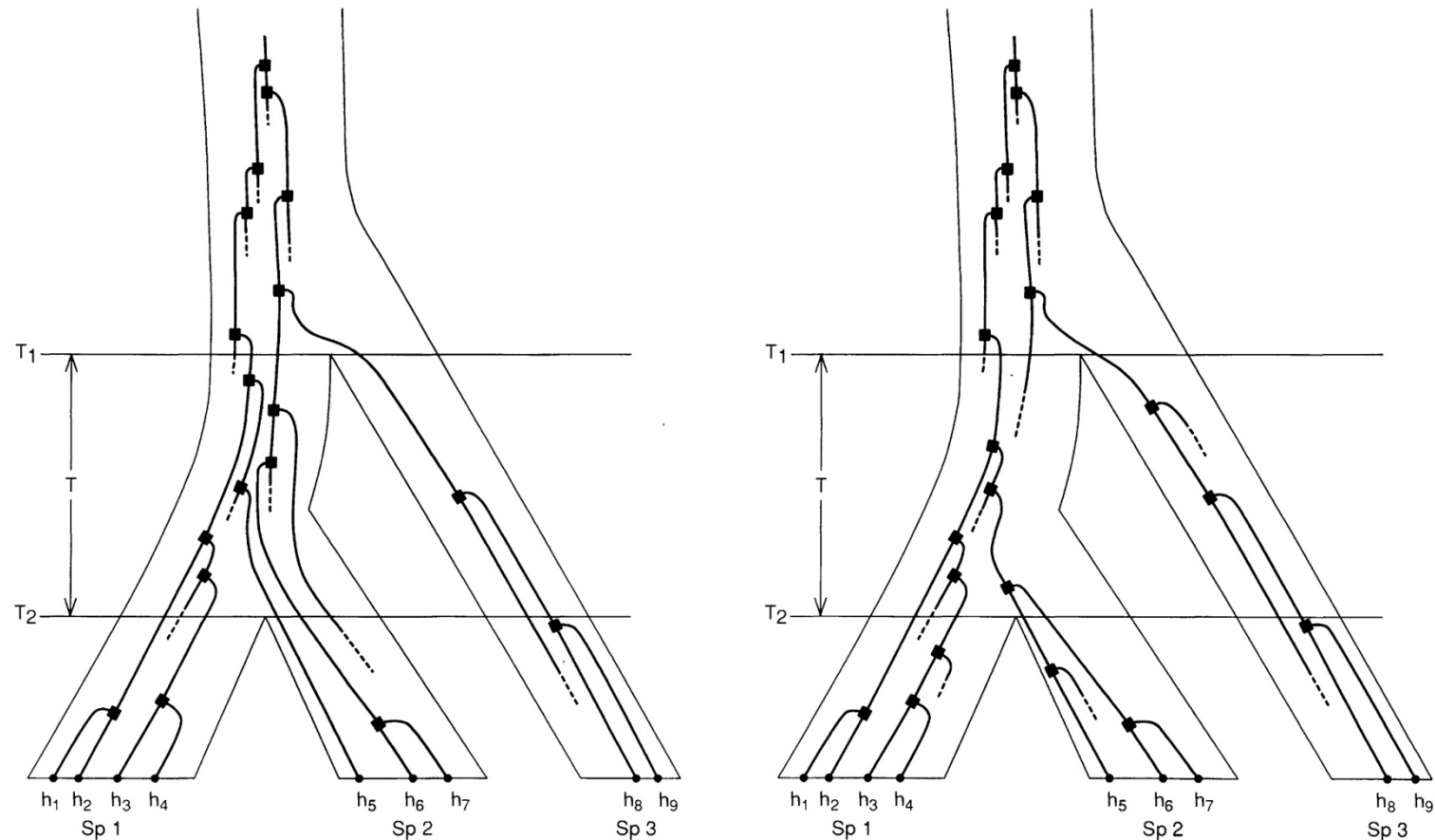
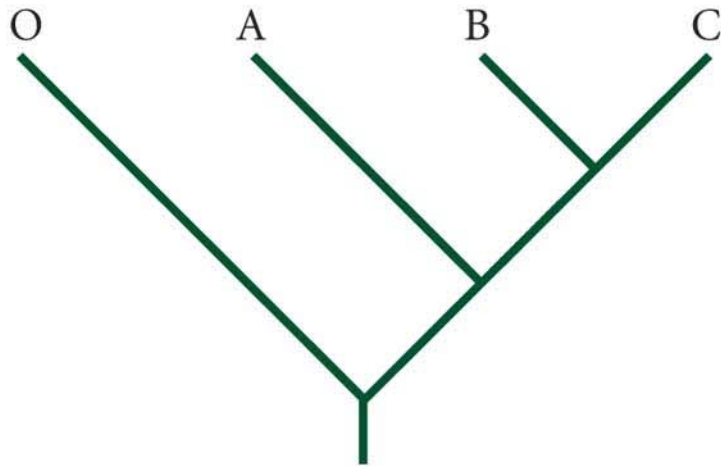


FIG. 1. The relationship between gene trees and species trees. The outer framework represents the species tree.  $T_1$  and  $T_2$  are times at which speciation occurs. The single-trace branches within the species represent branches of the gene tree. Figure 1A illustrates a situation where a polymorphism ancestral to all three species was transmitted through the ancestral stem of species Sp1 and Sp2 and persists in these two species. Even assuming that the gene tree was correctly resolved, it is apparent that if alleles (haplotypes)  $h_6$  or  $h_7$  happened to be the sampled sequences, then Sp2 and Sp3 would appear to be sister species. If, on the other hand, one happened to sequence  $h_5$  rather than  $h_6$  or  $h_7$ , then the gene tree would accurately reflect the species tree. Figure 1B illustrates a situation where the gene tree would correctly reflect the species tree, regardless of which haplotypes were sequenced.

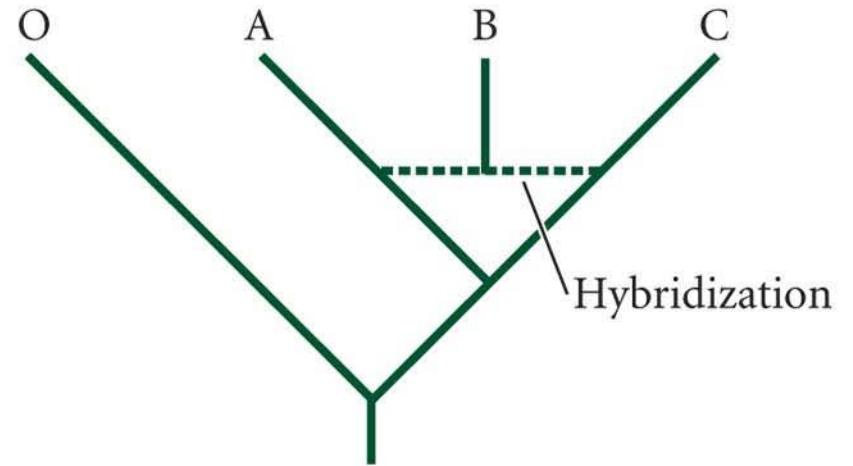
# Difficulties in Phylogenetics

- Scoring characters can be challenging
- Homoplasy may be common
- Evolutionary change may erase signs of evolutionary history
- Rapid radiation from a common ancestor
- Gene trees may imply the wrong phylogeny
- **Hybridization and horizontal gene transfer**

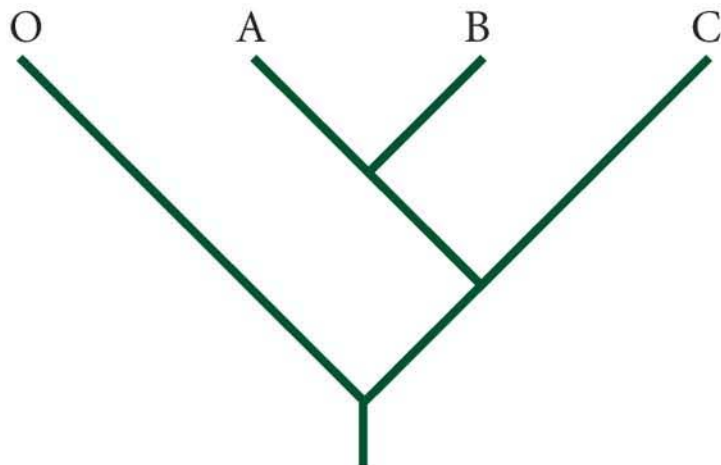
(A) Phylogeny based on sequence of gene 1



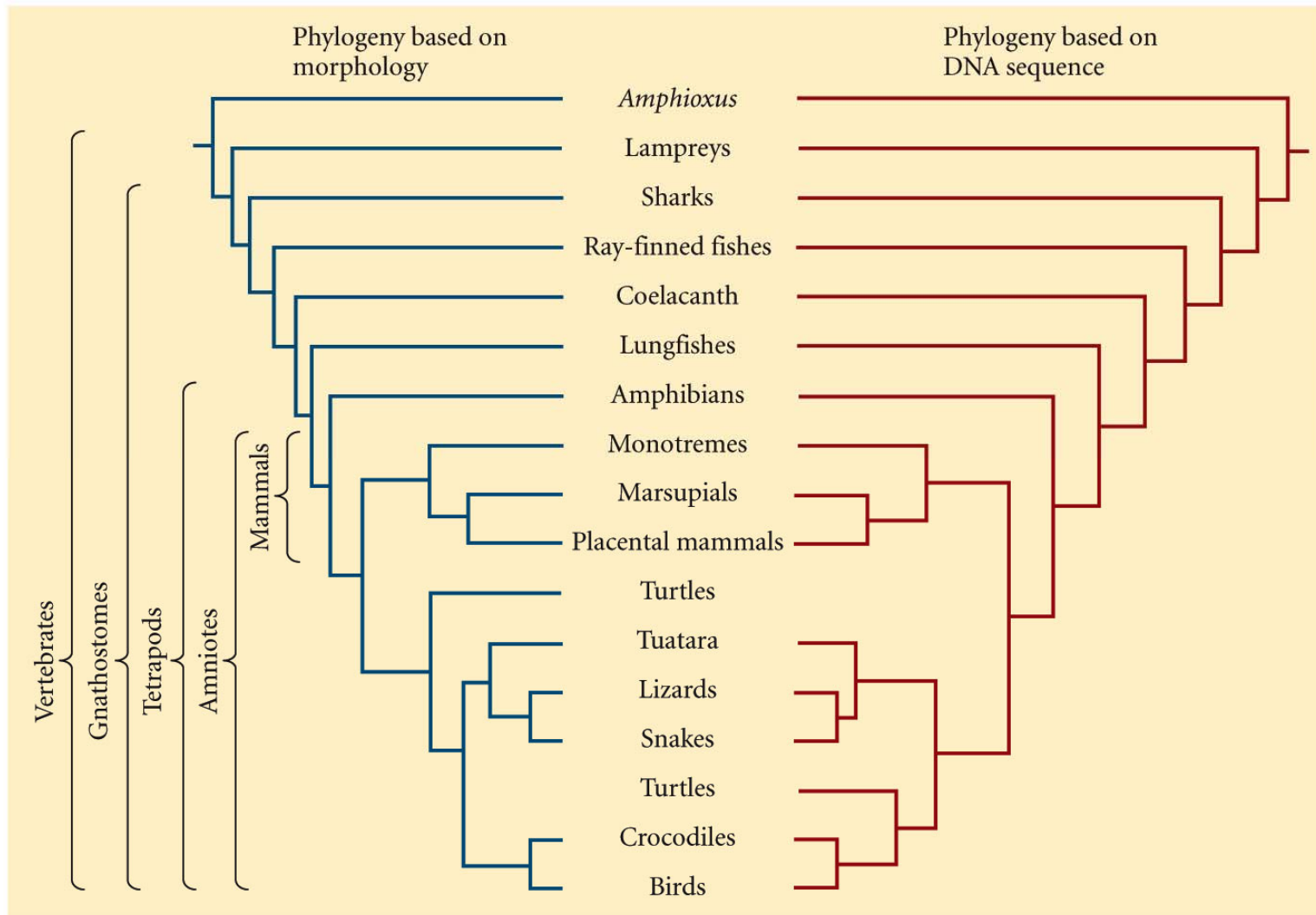
(C) Reticulate phylogeny



(B) Phylogeny based on sequence of gene 2



In spite of such difficulties, independent trees are often similar, unless you're a turtle...





# The Universal Tree of Life

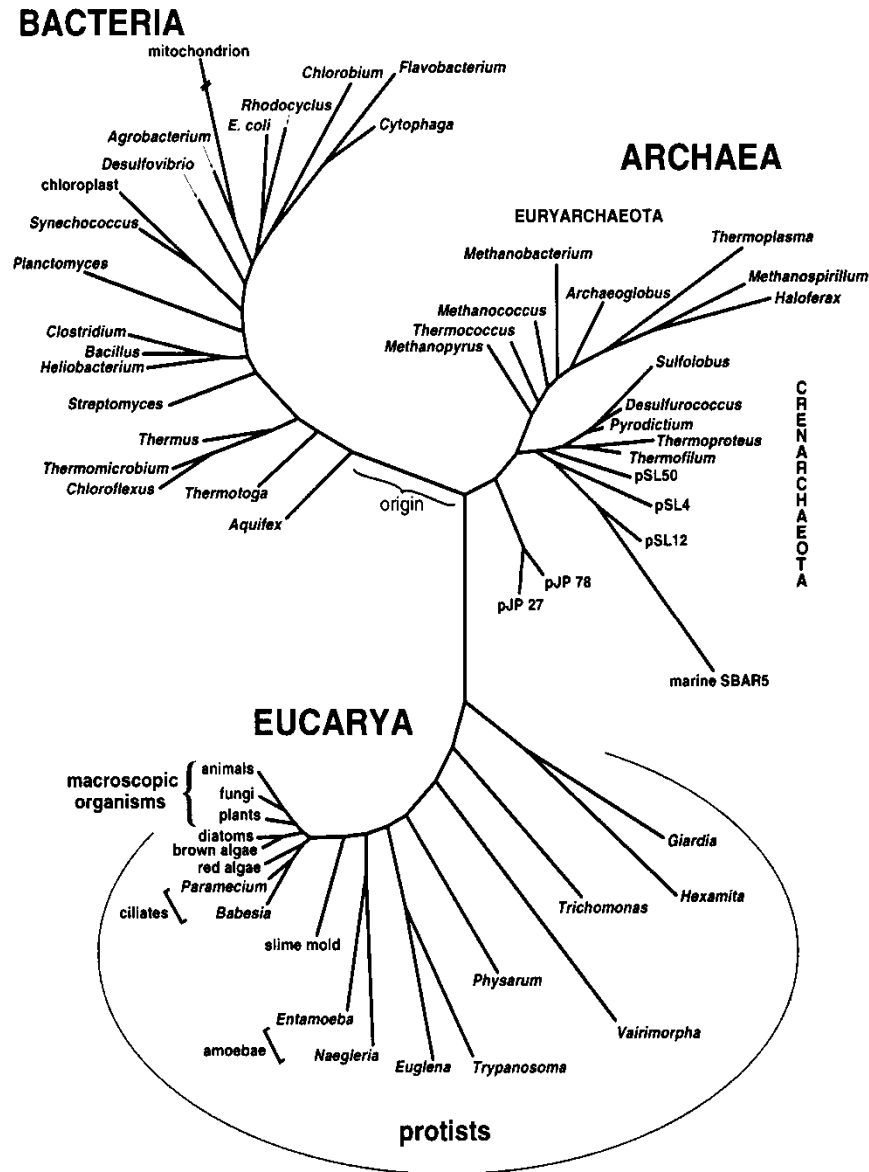
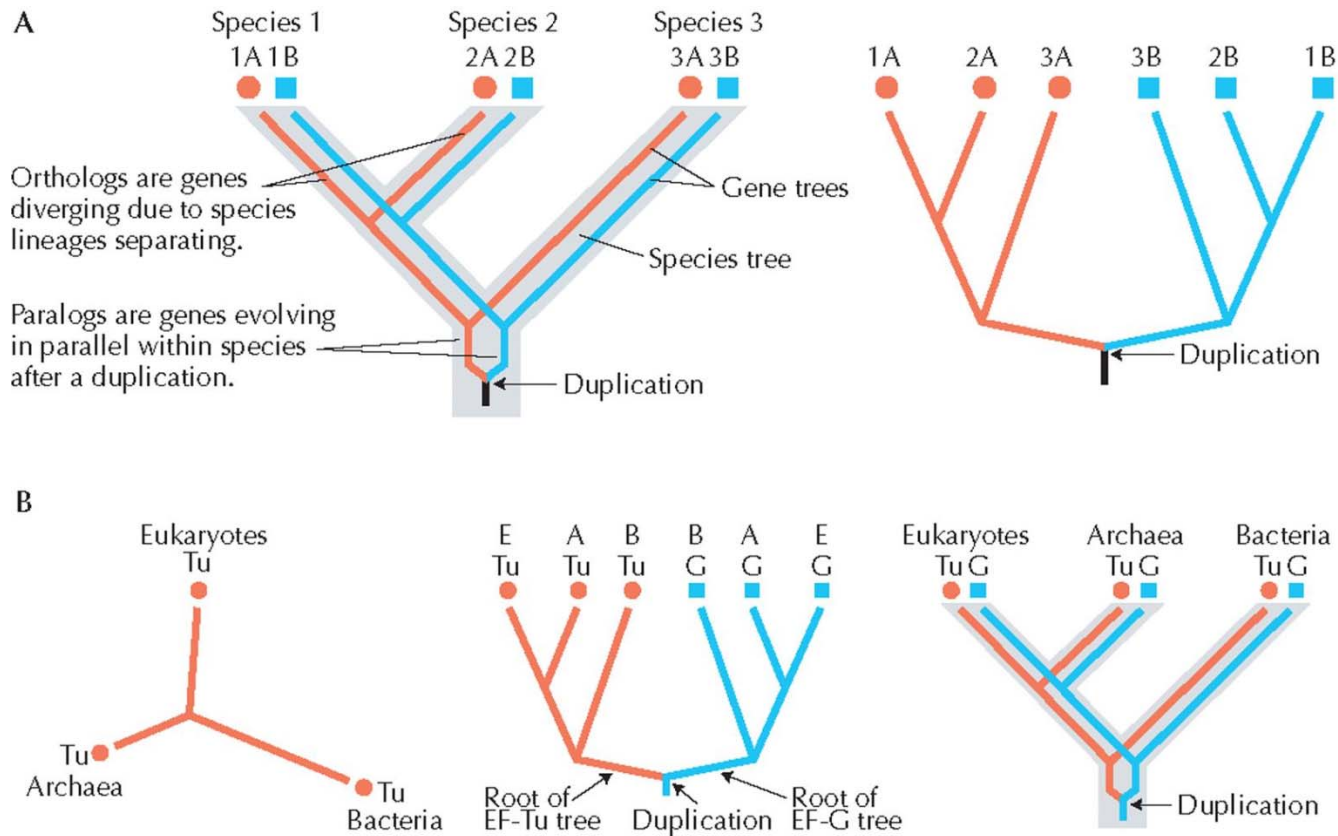


Figure 1. Diagrammatic "Universal" phylogenetic tree of life, based on small-subunit ribosomal RNA sequences. Based on analyses of Barns et al. (1996b), Olsen et al. (1994), and Sogin (1994).



**FIGURE 5.20.** Orthologs, paralogs, and rooting the tree of life. (A) Evolutionary trees of species and genes representing gene duplication events. (Left) The tree includes a species tree (*thick gray lines*) and gene trees (*blue and red lines*). A gene duplication event, leading to the coexistence of the blue and red paralogs in the root of the species tree, is labeled. The species tree subsequently splits twice, producing three species, each of which has inherited the blue and red paralogs. (Right) The gene tree is extracted from the species tree and untangled. The red forms of the gene, which are orthologs of each other, are more closely related to each other than to any of the blue forms. The same is true for the blue forms of the gene. Note that the species relationships among the two groups of orthologs (red and blue) are the same. (B) The same types of trees as in A, but these correspond to the evolution of elongation factors Tu and G across the three domains of life. The *red* and *blue* branches in the rightmost tree each correspond to a Tree of Life, and each is rooted by the paralogous elongation factor.

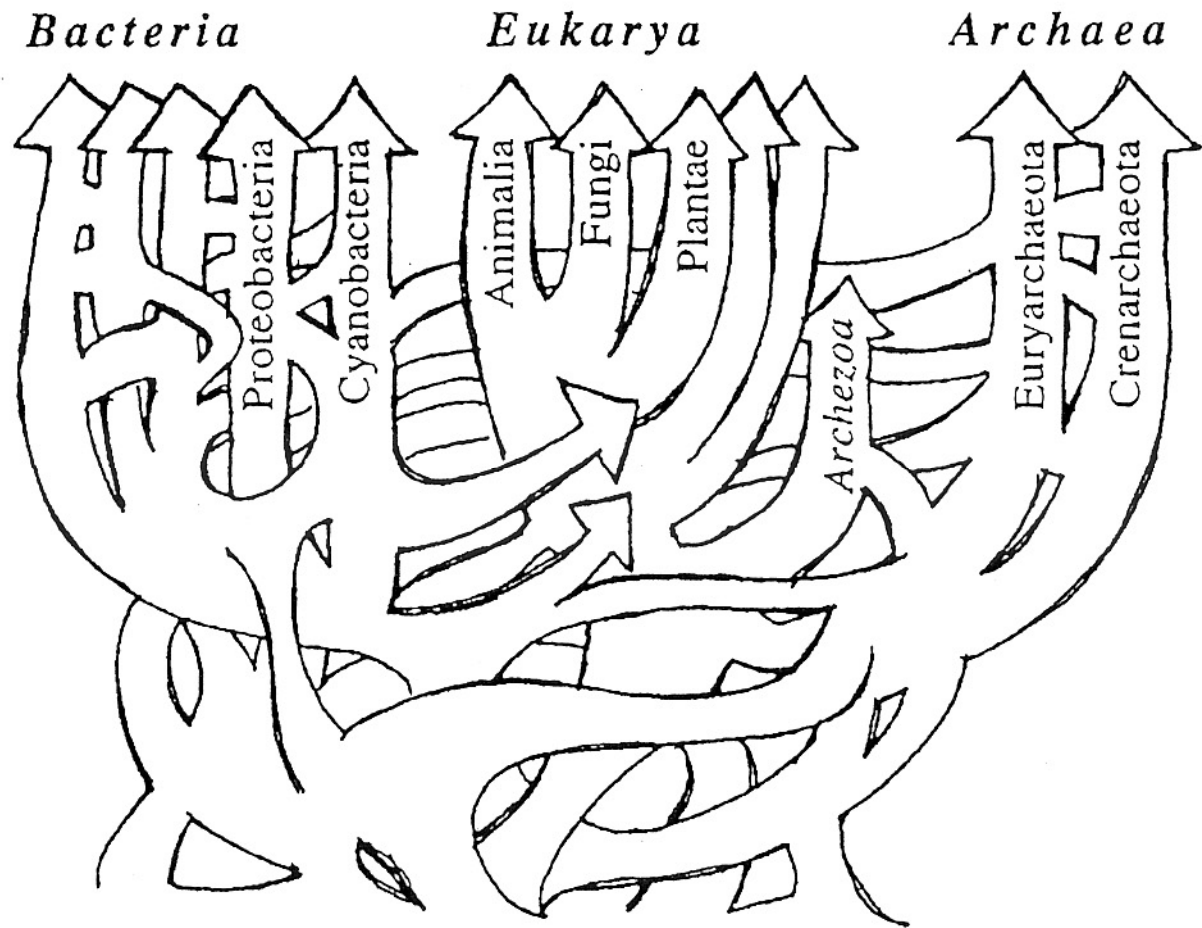
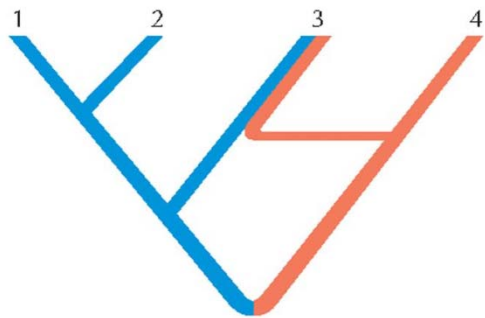
## Some Lessons from the **BIG TREE** of Life: Map of the Biological Record

Evolutionary “clock” is NOT constant among different lineages.

- Terminal nodes NOT all the same length, so not constant for all organisms either!
- Endosymbionts sped up very fast (semi-autonomous organelles).
- Eucarya – Fast clocks
- Archaea – Slow clocks
- Bacteria – Intermediate

# The Universal Tree of Life???

Fig. 3. A reticulated tree, or net, which might more appropriately represent life's history.



## Some Lessons from the BIG TREE of Life: Map of the Biological Record

What does genome sequencing and study of functional genomics add to our perspective?

- The central information processing machinery encompasses core genome.
- Metabolic functions, that's when relationships get murky.
- Endosymbiosis involved more than organelles, i.e., two-way transfer of genes with most going to the nucleus.
- Mitochondria have been at it much longer than chloroplasts.

# Take Home Message

- Phylogeny is right or wrong, we try to infer it the best we can.
- Taxonomy is useful or not, depending upon your point of view.
- Phylogeny allows us to ask testable questions, e.g., hypothesis testing regarding evolutionary relationships.