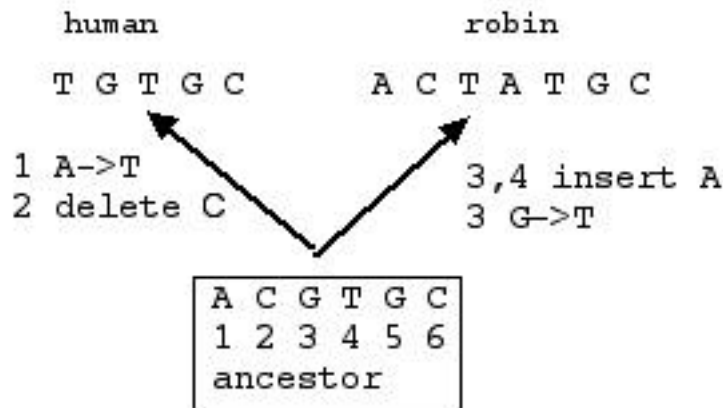
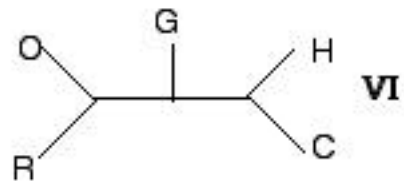
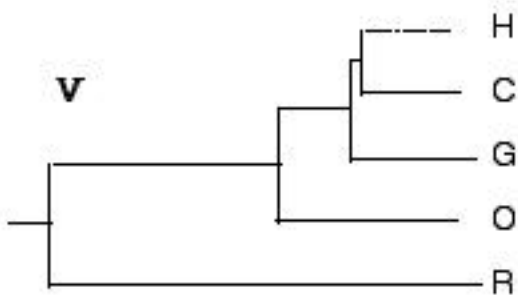
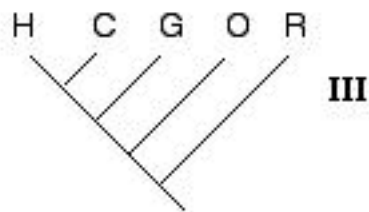
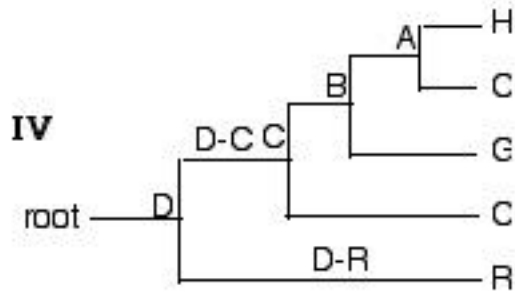
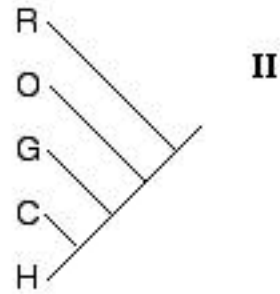
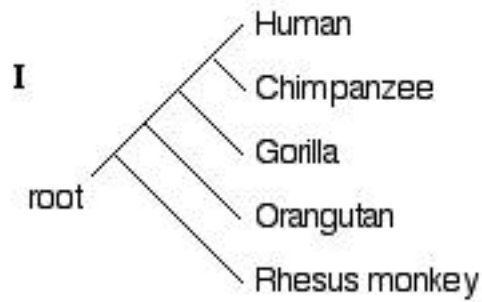
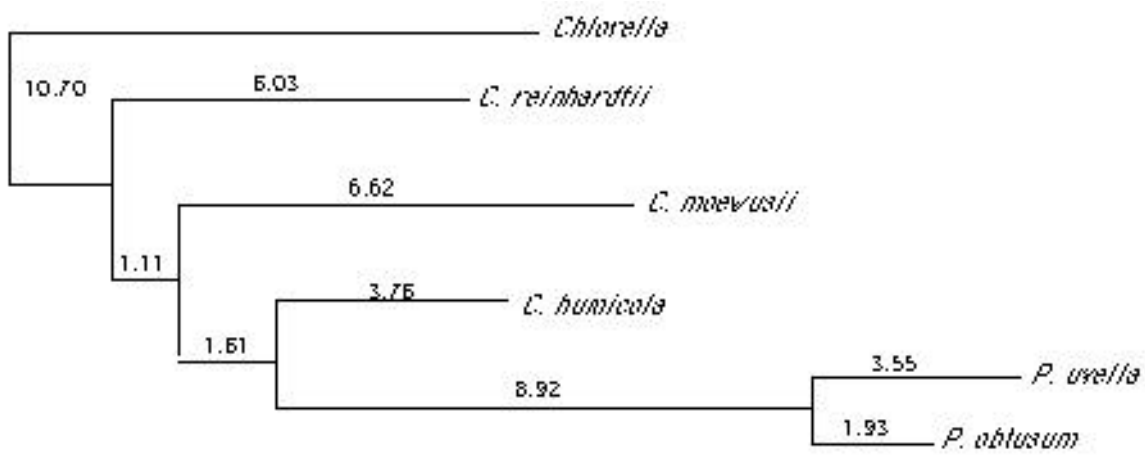
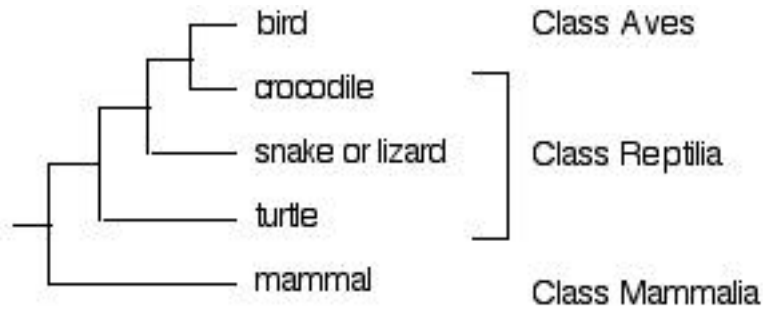


pairwise d

	green	blue	black
green	-	0.4	0.4
blue		-	0.2
black			-

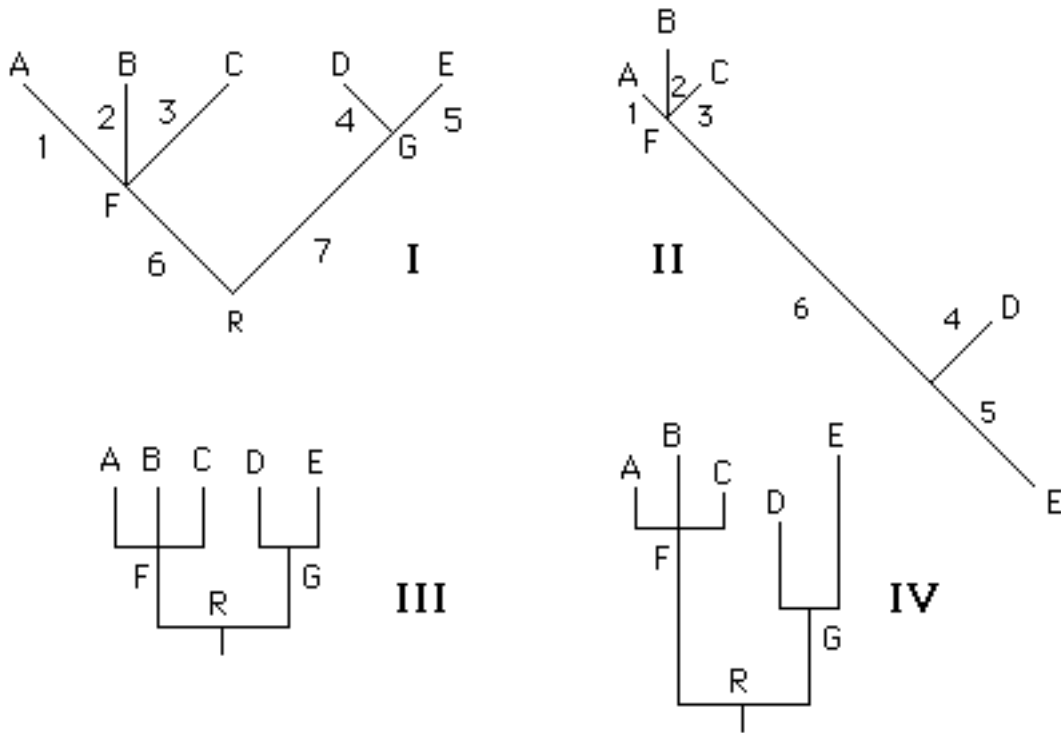






**Exercise:**

1.



What is the most recent common ancestor of D and E?

Could all of these trees be constructed from the same data? How do you know?

Which trees are phylograms?

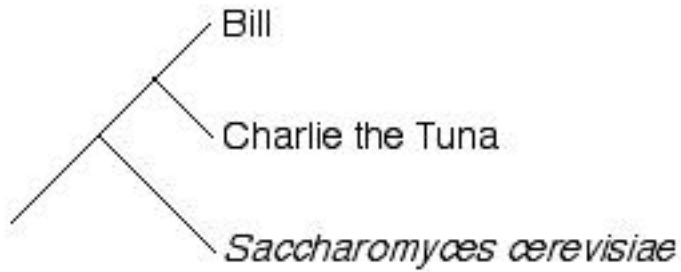
Which trees are cladograms?

What clades are shown?

If G underwent speciation 50 Mya, how much time elapsed along the branch G-D? G-E?

Why is FB longer than FA or FC? GE longer than GD?

2.



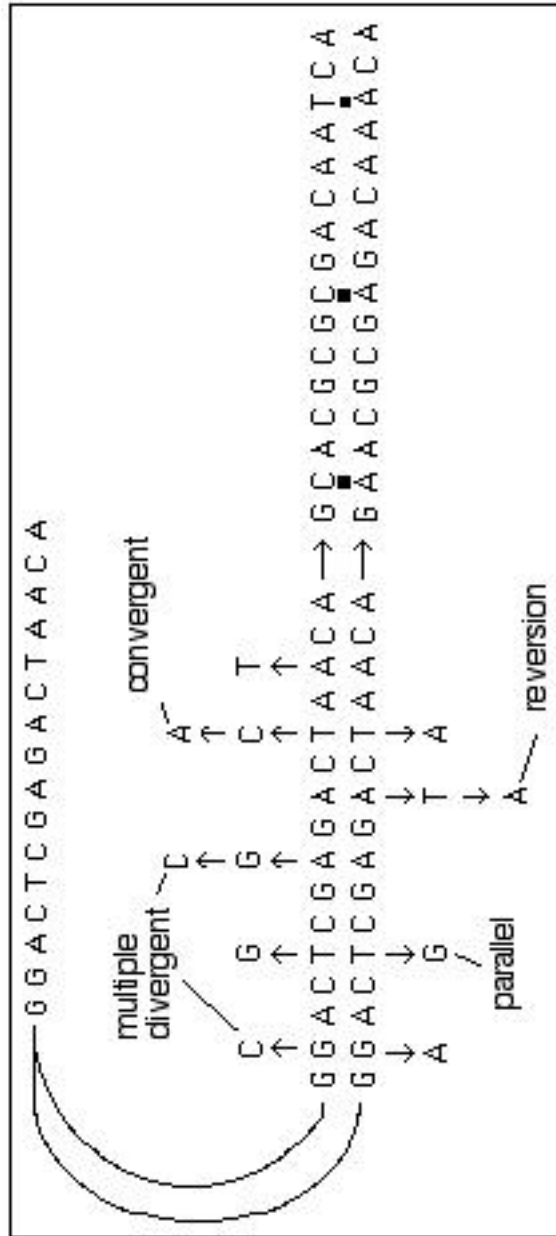
Whose 18S rRNA sequence is more similar to that of *Saccharomyces*, Bill's or Charlie's?

Who is more closely related to a fungus, Bill or Charlie?

Which species is older, the tuna (I don't know the scientific name) or *Saccharomyces cerevisiae*?

3. True or false: all organisms are related.

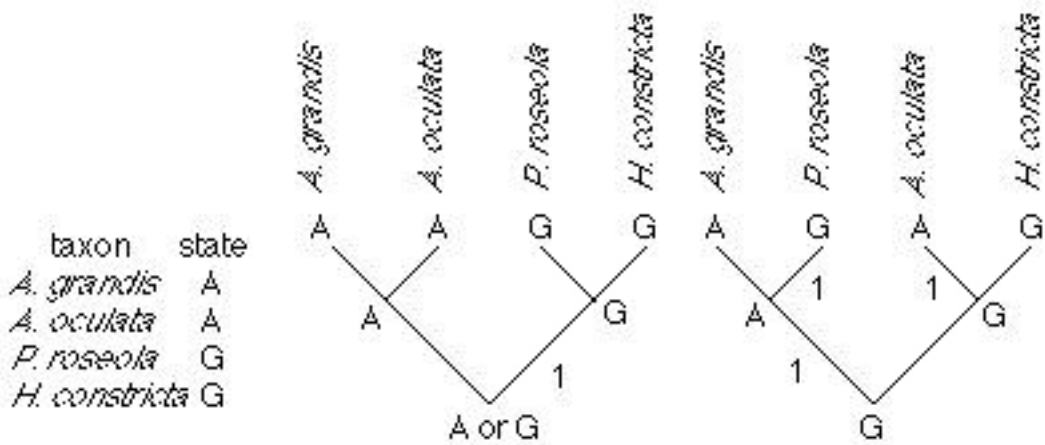
4. True or false: any two species of organisms have a common ancestor.



**Informative and uninformative sites**

- |     |   |   |   |   |   |   |   |   |
|-----|---|---|---|---|---|---|---|---|
|     | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| (a) | A | T | G | A | C | T | A | A |
| (b) | G | T | G | A | T | T | G | A |
| (c) | A | C | G | G | A | T | A | A |
| (d) | A | T | G | C | A | T | T | A |
| (e) | A | C | G | C | A | T | C | A |

**Parsimony**



**Exercise**

1. Draw all the unrooted trees for 4 taxa.
2. Now explain what you would have to do to draw all the rooted trees for 5 taxa.

## STUDY GUIDE TO METHODS FOR MAKING PHYLOGENETIC TREES

There are four widely-used methods for making trees: Parsimony, Maximum Likelihood, Neighbor-Joining, and Bayesian analysis. We will focus on the first three.

All methods make phylogram, and all can also provide the corresponding cladograms.

All methods can be used with bootstrapping which gives information about the confidence we can have in each clade.

### 1. Parsimony

- Uses parsimonious inference to reconstruct ancestral states; you should understand how this works. Best tree is shortest tree (most parsimonious, with smallest number of changes).
- Emphasizes finding clades; uses only informative sites and discards the rest of the data.
- No correction for multiple hits, less reliable when there are long branches in the real tree.
- Slow.

The following are called distance methods, to emphasize that they use all the data to estimate differences between taxa. Consequently they can be used with evolutionary models that correct for multiple hits.

### 2. Maximum likelihood (ML)

- For each tree, calculates likelihood of tree, given data and model; highest likelihood is best.
- Can be used to estimate values of parameters in any evolutionary model.
- Can reconstruct ancestral states.
- Very slow.

### 3. Neighbor-joining (NJ)

- Algorithm that seeks to put the most similar sequences closest together.
- Finds one tree, usually good approximation of true tree.
- Must be bootstrapped for confidence.
- Very fast.

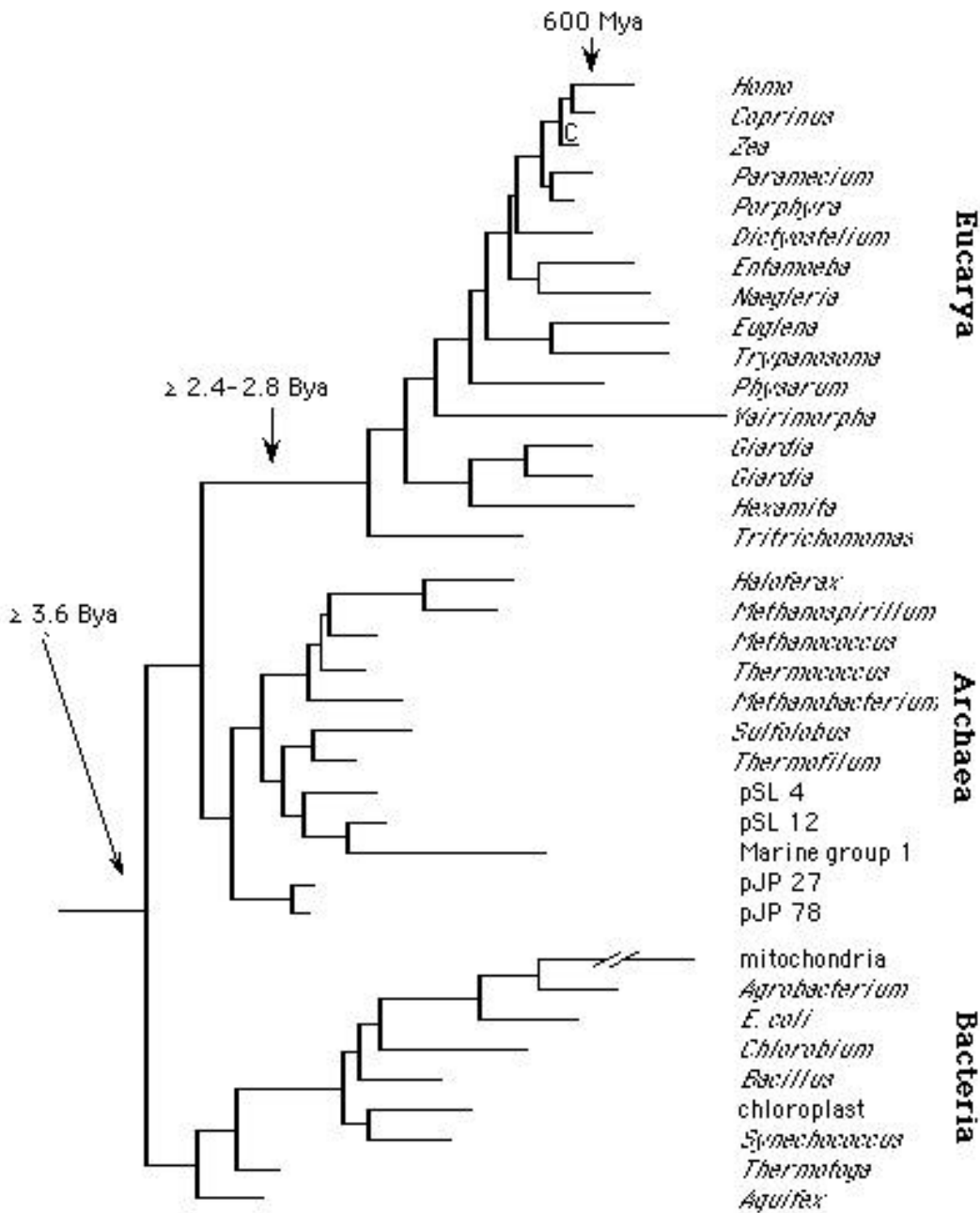
Parsimony and ML can use any of three ways of searching for the best tree.

1. Exhaustive searches: estimates number of changes for all possible trees, therefore very slow. But it lets you see how much difference there is between the best (shortest) tree, second best tree, etc.
2. Branch-and-bound: guaranteed to find shortest tree, but only finds that one tree; faster but still slow.
3. Heuristic search: very fast algorithm that may not find best tree, unless it is repeated many times with some kind of randomization (bootstrapping or randomizing order of input).

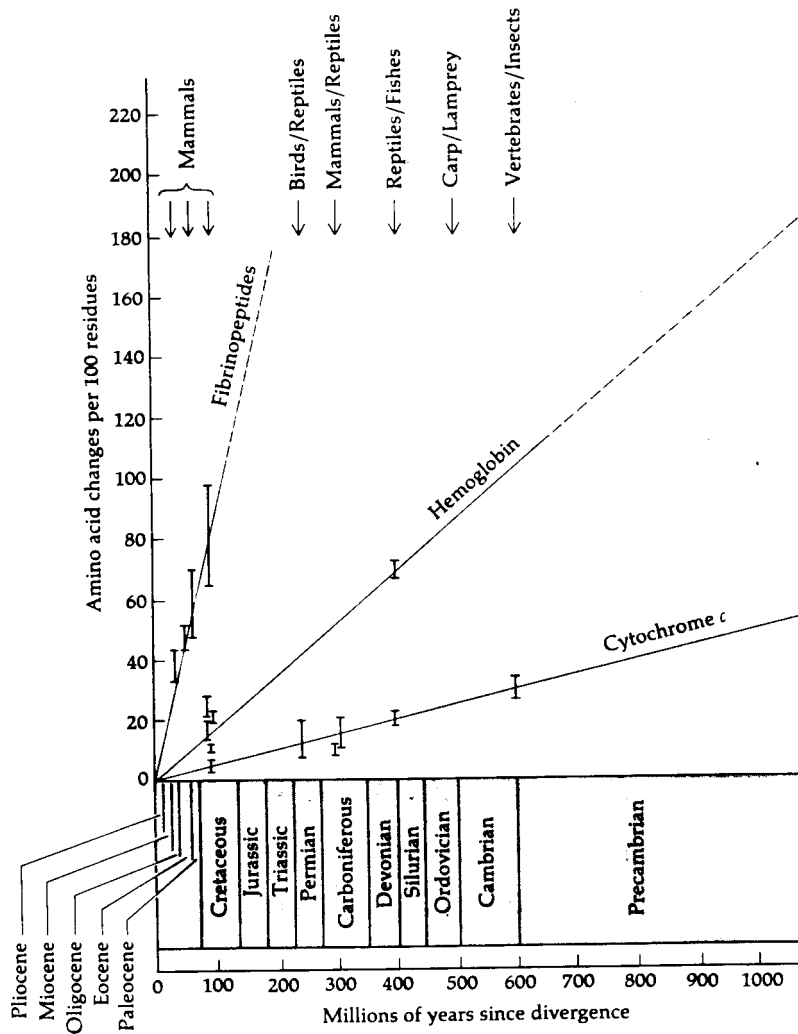
Suggested quick and dirty procedure if you have to do it yourself:

1. Make an NJ tree with no correction to see if there are long branches. If all are  $< 0.1$ , you probably don't need to correct for multiple hits. Use parsimony, NJ, and ML with bootstrapping, and pray that they agree!
2. If there are long branches, use ML to select best model and estimate parameters, then use that model with NJ if you have  $>10$  taxa. If you have  $<10$  taxa, use NJ and ML with the best model.

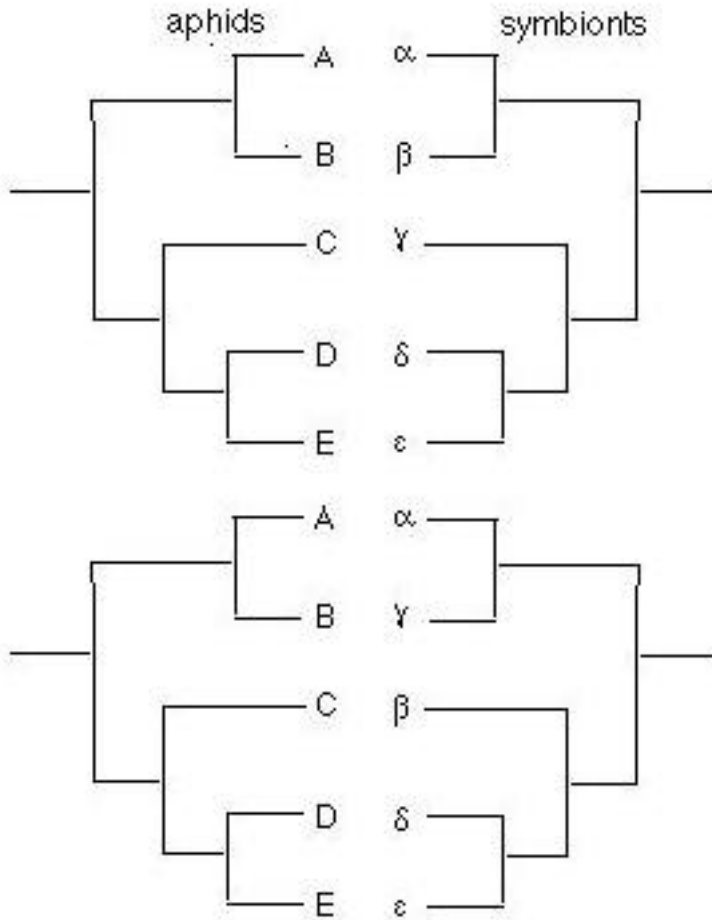
The Tree of Life based on SSU rRNA gene sequences



Evidence for molecular clock from protein sequences



Coevolution of aphid and *Buchnera* endosymbionts



Possible result of starch gel electrophoresis of alcohol dehydrogenase from 8 individual *Drosophila melanogaster*. S and F are inferred *AdhS* and *AdhF* alleles.

