**DUE: Monday Apr 29 at the BEGINNING of class.**
Hand In: Answer the questions on paper, number your answers, show all work (as necessary), identify key assumptions, and indicate all collaborations and assistance received or given.

Note that showing work means that if you utilize software for assistance (programs you write or stock software such as Excel or R), you should indicate as such and provide sufficient details so that I can judge the work. That may mean sending me (by email) source code or associated Excel files.

Your work must be legible -- if your handwriting isn't great, type it up and print it.

Questions

1. Consider the following phylogenetic tree:
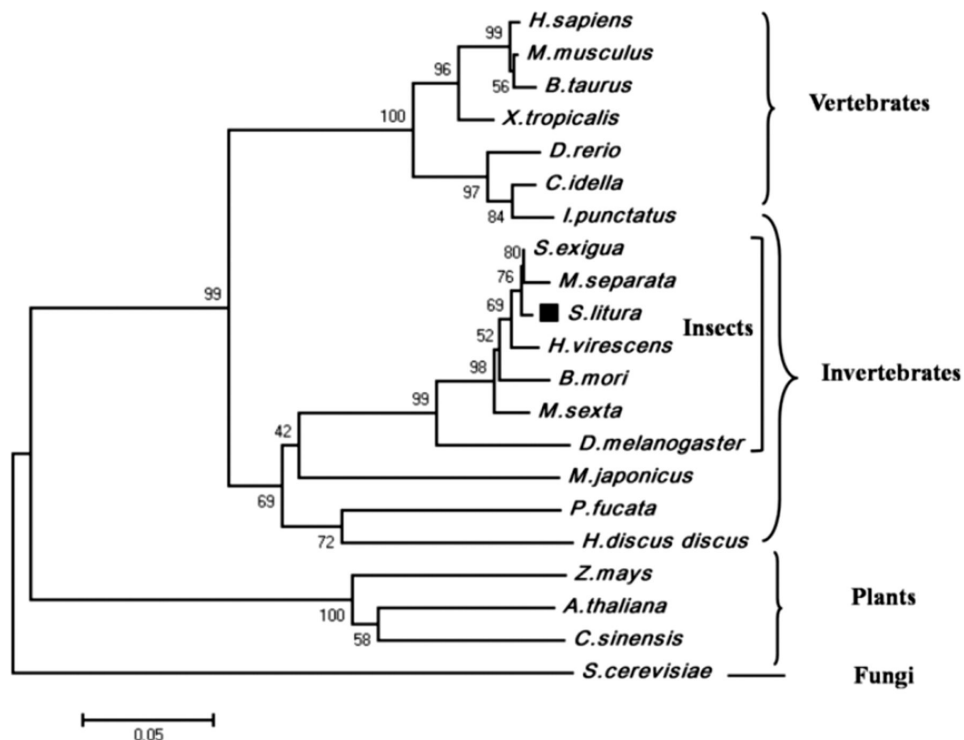
a) Is this a cladogram or a phylogram?
b) Which sequence(s) is/are presumably the outgroup?
c) Which sequence is most closely related to A.thaliana?
d) Circle (on the tree above) the last common ancestor of M. musculus and D. rerio.
e) Which branch(es) do you have the least confidence in? Why?

2. Consider the following distance matrix:

	A	B	C	D	E

A	-				
B	90	-			
C	20	100	-		
D	80	30	90	-	
E	50	40	60	50	-

Calculate a rooted tree using the UPGMA method of tree construction. For full credit you must show the final topology of the tree, the calculated branch lengths, the location of the root, and **ALL** intermediate matricies utilized in its construction.

3. Consider this unrooted tree:
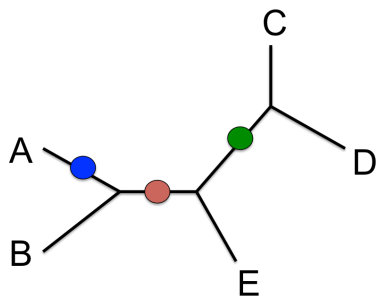
a) (Ignore the colored dots for this part.) How many unrooted and rooted trees are possible for this many operational taxonomic units (OTUs)?

b) Draw the three rooted trees that arise by placing the root at each of the three labeled colored dots (blue, red, green).

4. Explain the difference between an algorithm and a scoring scheme. Use an example from class (anytime during the semester) to clarify your definition.

5. (Advanced) Read "The what, where, how and why of gene ontology -- a primer for bioinformaticians" (PDF is available on D2L as GOprimer.pdf) and answer the following questions:

a) What are the three ontologies in GO?

b) The relationships within GO form what kind of graph?

c) What is the main difference between the full and filtered GO?

d) Who does most of the current GO annotation?

e) According to Figure 5, what is the most common type of evidence for information within GO?

f) Give one pro and one con of the information content measure of GO terms.

g) List 3 major criticism of GO.