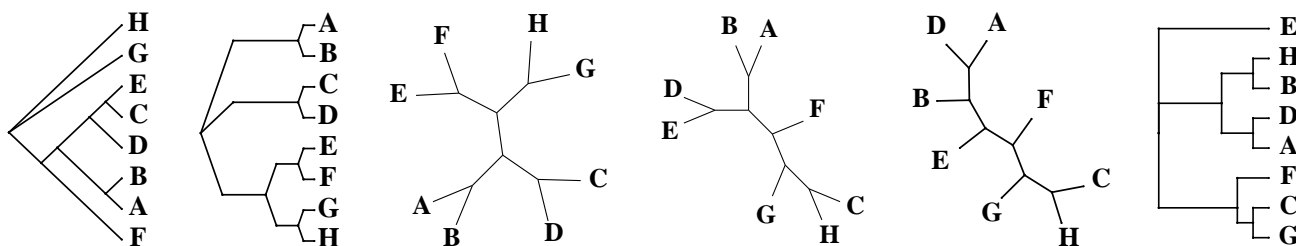


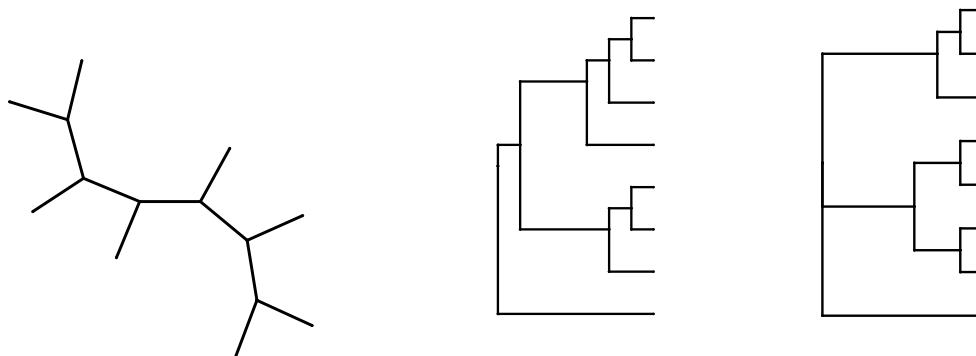
This exam is closed-book and totals 200 points. You can use a calculator if you need to; if you don't have one you can instead leave the calculations as expressions such as $(3.67 \times 234)/1243 + 4.5$. Make sure to put your name on each page as the pages may be separated for grading. If you do any work on the blank backs of the pages, it will be best to work on the back of the same page, so that this work does not become separated from the question during grading. Showing your work may help you get partial credit if the final answer is wrong.

1. (36 points) **Basic Tree Literacy (Arboricity? Dendricity?).**

(a) Two of these trees are the same unrooted tree topology. Circle the two.

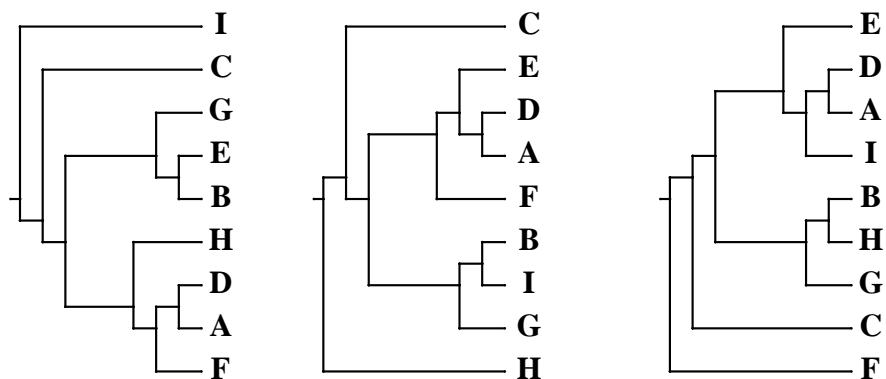


(b) Can all three of these trees have species names (use A through H) placed on their tips so as to be the same tree (when considered as an unrooted tree)? If some can, do it for them; if one can't, explain why it can't be done for it.



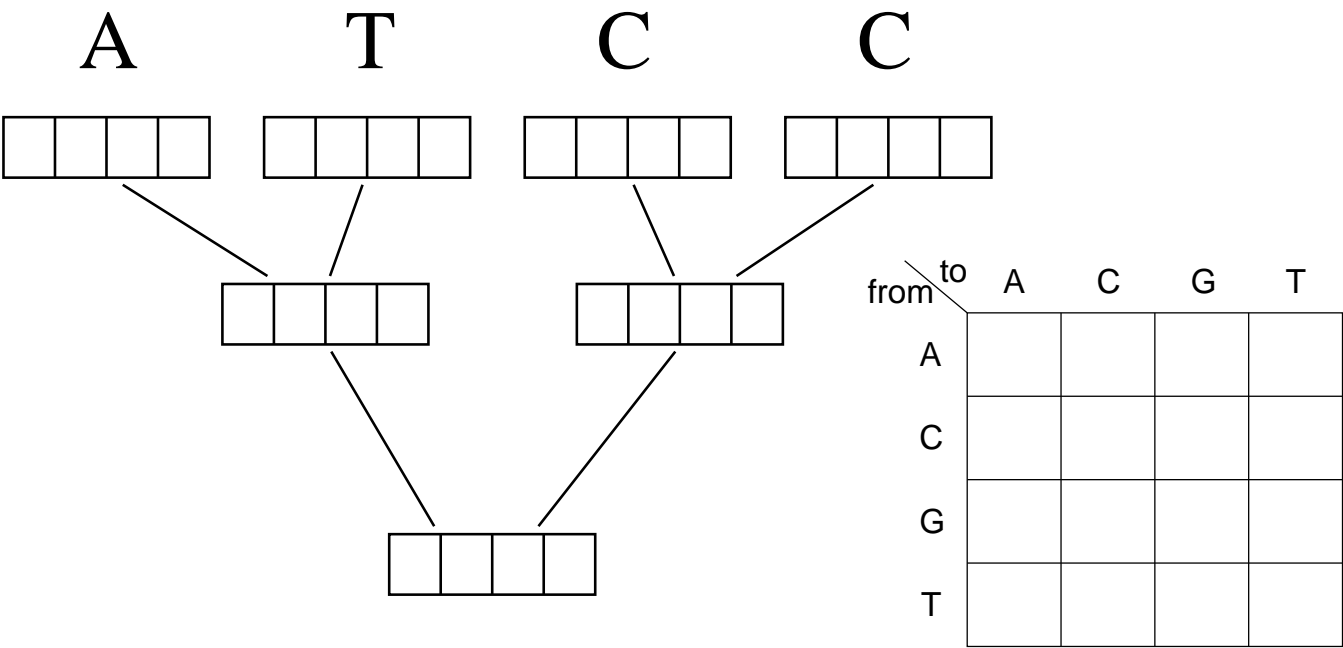
(continued on next page)

- (c) For the three trees shown, toss out species until you get (as large as possible) a subtree that is the same in all three cases. (This is called a *maximum agreement subtree*). Draw it below them. (I am not assuming that you know some algorithm for doing this, just that you can figure out the answer by trial-and-error).



2. (36 points) For this site, use the Sankoff Algorithm to compute the parsimony score if we count 1 for transitions, and 3 for transversions. Fill in the boxes of numbers on the tree, working downwards. A table for the “costs” of changes is provided on the right, which you can fill in (if you wish) and then use as you go down the tree.

(a) If this is tree, with the bases shown at the tips, fill in the boxes on the tree according to the Sankoff Algorithm:



(b) What is the score of the tree?

(c) What set of possible nucleotides at the ancestor of the tree is inferred by this?

3. (33 points) I get emails something like this: “Most esteemed Professor Felsenstein Joe: I am humble student Mr. Nietsneslef Eoj, of the Institute of Entomological Etymology and Hermetic Hermeneutics in Ruritania. I have many DNA sequences from the gene for Congenital Halitosis. Three of my professors have made different suggestions for the analysis:

1. One says that I should assume nothing about variation among sites in the rate of evolution, as to do so would amount to requiring prior knowledge, and that I should therefore use analyses that do not allow rates to vary.
2. A second says that I should use a Gamma distribution for rate variation, or perhaps a Hidden Markov Model. I am confused: if I do, how am I to know what amount of rate variation (the α parameter) to assume?
3. The third says that it will help if I try to estimate a rate of evolution for each site, a separate rate parameter for each site.

Which of these should I do, and why? Please send your response in Microsoft Word format, double-spaced, by next Friday morning.”

Help me advise this poor fellow, giving reasons for your opinions.

4. (30 points) Consider distance matrix methods used on DNA sequences. True or False (circle the choice for each statement):

- | | | |
|---|---|---|
| Neighbor-joining is a distance matrix method | T | F |
| UPGMA is a distance matrix method | T | F |
| Least squares methods include Neighbor-Joining | T | F |
| Least squares methods include the Fitch-Margoliash method | T | F |
| It is not possible to bootstrap if you use the Fitch-Margoliash method | T | F |
| If you are given the distance matrices, but not the sequences,
you cannot bootstrap sample from them | T | F |
| Neighbor-Joining assumes a molecular clock | T | F |
| You can infer branch lengths on a predefined tree with
the Fitch-Margoliash method | T | F |
| You can infer branch lengths on a predefined tree with
the Neighbor-Joining method | T | F |
| UPGMA assumes a molecular clock | T | F |

(if you need room to explain an answer that might be controversial, use this):

5. (30 points) Two people ask you these questions:

- Person 1 asks whether it is true that if you run a Bayesian inference program for a long time, the clade probabilities will get higher the longer you run it.
- Person 2 asks whether it is true that you need many bootstrap replicates to get the bootstrap support values up to 90%.

The correct answers have something in common, but are not entirely the same. Tell me what the answers should be, and why.

6. (35 points) A student from another place actually did write (in part) this:

```
"I have tried to ask the professors at my University, but none of them knew
which programs ... I should use.

Is it at all possible to take the file from [name of program they use] with
codominant data and use it to generate a tree (the file is attached to the
mail and also in the button of the mail)?

Which program ... do I have to use ...

Best regards
(name withheld)

-----txt file-----
1 7 . Codominant data template
7
308 Pop1
1 2.2 0.0 2.2 2.4 2.2 3.3 5.5
2 2.2 0.0 2.2 1.1 1.2 2.3 3.5
3 2.2 3.3 2.2 2.2 1.2 2.3 5.5
4 2.2 2.2 2.2 2.4 2.2 3.4 3.5
... "
```

It is important to notice that each line of this data file is a genotype at seven loci for a single individual from the same particular population ("Pop1").

What advice should I have given this student (you should not confine yourself to answering the "which program" question)? Give reasons as well.