**GS 373 Homework 5 KEY**

**Bioinformatics Questions (80 points)**

1. (25 points) Primates and rodents both have two copies of gene A, called A1 and A2. You have sampled sequences of A1 and A2 from humans (H), chimps (C), rats(R), and mice(M).
   a. (11 points) Draw the rooted phylogenetic tree you would expect if A1 and A2 arose from an ancient gene duplication in the ancestors of all mammals and have evolved independently since. Your tree should have eight tips (H-A1, H-A2, C-A1, C-A2, R-A1, R-A2, M-A1, and M-A2). You don't have to worry about branch lengths.
   b. (11 points) Draw the rooted phylogenetic tree you would expect if A1 and A2 arose in the ancestor of all mammals but are next to each other on the chromosome and experience strong concerted evolution.
   c. (3 points) Explain why the trees in parts a and b are the same or different.
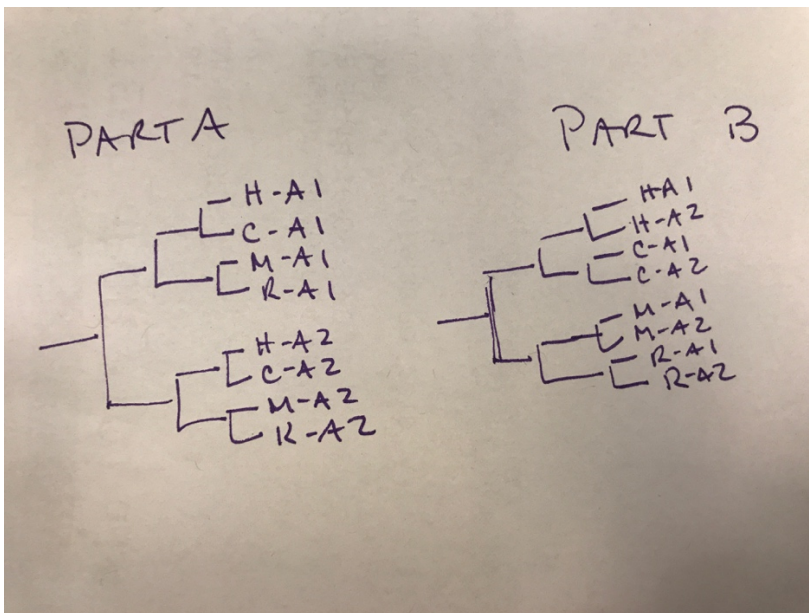
In part A, A1 and A2 experience independent evolution, which means the evolutionary history of A1 is more similar to other A1s than the A2 within the same species. In part B, A1 and A2 experience strong concerted evolution, which means their evolutionary history more closely matches the species tree.

2. (25 points)
Consider the following majority consensus tree. Please give the bootstrap values as proportions for the subtrees marked by i, ii, and iii.

majority
consensus
tree

bootstrap trees
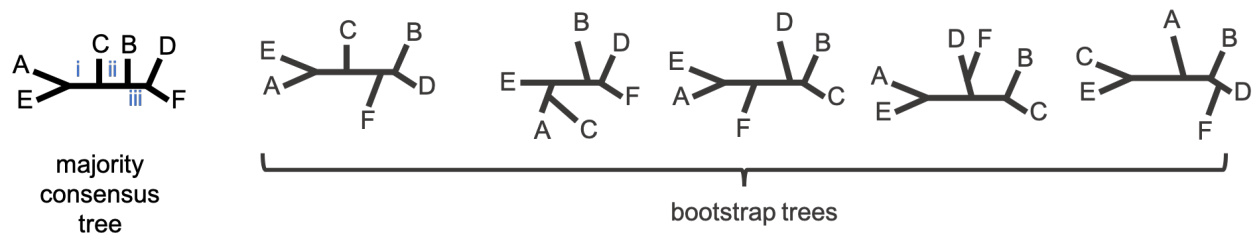
Hint:
Try constructing a table like this

subtree split name	subtree split description	number of trees
i	*AE | BCDF*	*3*
ii	*ACE | BDF*	*3*
iii	*ACBE | DF*	*3*

0.6 for all three nodes

3. (20 points) Read Kellis *et al.* (2004). Specifically, the introduction up until but not including "Genome sequencing and alignment" and "Evolutionary Analysis" sections "Pattern of gene loss", "Accelerated protein divergence", and "Ancestral and derived functions."
   a. (5 points) Name one concept or algorithm we discussed in class which was crucial for this paper.
      Many different answers (alignment, phylogenetic trees, gene duplication, etc)
   b. (15 points) What are the competing hypotheses for post-duplication divergence of gene pairs? Which hypothesis does this paper support and what is their evidence?
      Hypoth #1: After duplication, one copy would preserve original function and one would diverge. Hypoth #2: Both copies diverge and acquire new functions.
      In 95% of the genes they analyzed, the authors found accelerated evolution in only one of the duplication pairs. This supports hypothesis #1.

4. (10 points)
   a. (5 points) Explain why you want to build a species tree with orthologues rather than paralogues. Make sure you define each term in your response.
      Paralogs are gene duplications within a species. Orthologs arise via speciation. You should build your species tree with orthologs because they more closely match the speciation pattern.
   b. (5 points) Explain the relationship between gene duplication and gene family.
      When a gene duplicates, one or more copies could diverge in function. This leads to a gene family, a collection of genes with a common ancestor but slightly different functions