Bioinformatics Take Home Test #6

(This is an open book exam based on the honors system -- you can use notes, lecture notes, online manuals, and text books. *Teamwork is not allowed on the exams*, write down your own answers, do not cut and paste from webpages. If your answer uses a citation, give the source of the quoted text.)

1. **1pt** How are bootstrap values used in analyzing phylogenetic trees?

2. **1pt** What is a parametric bootstrap?
   a. Resampling a dataset with replacement.
   b. Aligning a sequence dataset using a neighbor joining tree.
   c. Estimating a distribution from the dataset and simulating new datasets based on that estimated distribution.
   d. Calculating the likelihood of a dataset, given the model.
   e. Calculating the probability of a model, given the dataset.

3. **1pt** What is a nonparametric bootstrap?
   a. Resampling a dataset with replacement.
   b. Aligning a sequence dataset using a neighbor joining tree.
   c. Estimating a distribution from the dataset and simulating new datasets based on that estimated distribution.
   d. Calculating the likelihood of a dataset, given the model.
   e. Calculating the probability of a model, given the dataset.

4. **1pt True/False** Branch lengths are considered when calculating the bootstrap support of a branch in a phylogenetic tree.

5. **1pt True/False** Branch lengths should always be shown when available because they are a measure for how much evolution has occurred along a branch.

6. **1pt** Which of the following **rooted** trees have identical topology when considered as unrooted?

   A) Tree (i) and (ii)  B) Tree (ii) and (iii)  C) Tree (iii) and (iv)  D) Tree (i) and (iii)
   E) Tree (i) and (iv)  F) All trees have different topology  G) All trees are identical in their topology
The above tree that was calculated in ClustalX from 1000 bootstrap replicates. It should be considered as unrooted; bootstrap support values are **written to the right of the branch** they pertain to.

MYG denotes a myoglobin (oxygen buffering molecule in the muscle of animals); LGB2 denotes a plant globin that functions as an oxygen buffer in root nodules of leguminous plants; GLB5 PETMA denotes a hemoglobin (oxygen carrier in the blood) from a deep branching fish lineage; HBA and HBB denote the alpha and beta chains of the hemoglobin molecule, respectively.

7. **1pt** The number 987 indicates that in 98.7% of the bootstrapped samples “HBB HUMAN” groups with “HBB HORSE”.
   
   Correct / incorrect

8. **1pt** In the above tree, the number 650 indicates that in 65.0% of the bootstrapped samples “MYG PHYCA”, “LGB2 LUPLU”, and “GLB5 PETMA” group together. Correct / incorrect

9. **1pt** In the above tree, the number 658 indicates that in 65.8% of the bootstrapped samples “MYG PHYCA”, “LGB2 LUPLU”, and “GLB5 PETMA” Correct / incorrect

10. **1pt** In the above tree, the number 650 indicates that in 56.5% of the bootstrapped samples “HBA HUMAN”, “HBA HORSE”, “HBB HUMAN”, and “HBB HORSE” group together. Correct / incorrect

11. **1pt** Based on the above tree it is possible that the GLB5 PETMA groups with MYG PHYCA in 35% of the analyses performed on the bootstrapped samples. Correct / incorrect

12. **1pt** Based on the above tree it is possible that the GLB5 PETMA groups with HBB HUMAN in 35% of the analyses performed on the bootstrapped pseudosamples. Correct / incorrect

13. **1pt** Based on the above tree it is possible that the GLB5 PETMA groups with (HBB HUMAN, HBB HORSE) in 35% of the analyses performed on the bootstrapped pseudosamples. Correct / incorrect

14. **1pt** Based on the above tree it is possible that HBB Horse groups with (HBA Human, HBA Horse) in 23% of the analyses performed on the bootstrapped samples. Correct / incorrect

15. **1pt** Based on the above tree it is possible that HBA Horse groups with (HBB Human, HBB Horse) in 23% of the analyses performed on the bootstrapped samples. Correct / incorrect

16. **1pt** You analyze a quartet of putatively orthologous sequences. The maximum parsimony tree looks like this:

   The central branch has 100% bootstrap support.
   
   A) Maximum parsimony is not subject to the long branch attraction artifact, rather it always has the tendency to group the long with the short branches. Therefore, the finding that A and B group together is reliable.
   
   B) This tree groups the two long branches together. The possibility exists that this result might represent a long branch attraction artifact.
   
   C) The central branch is so strongly supported that one can exclude a long branch attraction artifact. (LBA is a statistical phenomenon and never reaches 100% bootstrap support.)

17. **1pt** You analyze a quartet of putatively orthologous sequences. The maximum parsimony tree looks like this:

   The central branch has 100% bootstrap support.
   
   A) Maximum parsimony is not subject to the long branch attraction artifact, but always has the tendency to group the long with the short branches (aka long branch
repulsion). Therefore, the finding that A and C group together is unreliable. 
B) This tree does not group the two long branches together, indicating that the result is not due to long branch attraction. 
C) The central branch is so strongly supported that one can exclude any artifact that might occur during phylogenetic reconstruction. (The artifacts caused by long branches are a statistical phenomenon and never reached 100% bootstrap support.)

18. 1pt True/False The outgroup, or root, of a phylogenetic tree represents the ancestor of the ingroup.

19. 1pt True/False The outgroup, or root, in a phylogenetic tree is used to inform upon the ancestral state of a given character.

20. 1pt True/False Choosing the wrong outgroup can introduce long branches into a tree and cause Long Branch Attraction.

21. 1pt True/False Long Branch Attraction can NEVER lead to recovering the wrong tree with 100% bootstrap support.

22. 2pt What are the advantages and disadvantages of using the following alignment programs?
   a. Clustal-
   b. Muscle-

23. 1pt How is a neighbor joining tree calculated?
   a. By maximizing the probability of a dataset, given the model.
   b. By maximizing the probability of a model, given the data.
   c. By minimizing the number of substitution events.
   d. By using pairwise comparisons to determine the nearest neighbor and then collapsing a node and recalculating all of the pairs for the node.
   e. By programing a little robot to walk around in tree space.

24. 1pt What does adding ASRV to an evolutionary model do?
   a. Corrects for multiple substitutions to make the model more realistic and therefore produce better trees.
   b. Over parameterizes the model so that it is a poor representation of reality and therefore produces worse trees.
   c. Optimizes the alignment.
   d. Optimizes the Needleman-Wunsch algorithm.
   e. Makes the phylogenetic tree more parsimonious.

25. 1pt Draw the following tree (((A, B), C), D, (E, F), G)). Hint- this is a rooted tree.

26. 1pt Draw the following tree (((A, B), C), D, E). Hint- this is an unrooted tree.

Extra credit:

27. Max of 3pts Models used to describe sequence evolution frequently use the Gamma distribution. What is the name of the parameter used to describe the gamma distribution? What is the name of the process often described by the Gamma distribution? Why is the Gamma distribution more useful than the normal distribution?

For Graduate Students:

28. 2pt What do allele frequencies say about the existence of races and about personalized medicine, according to the papers by Bruce Lahn? Include at least four unique points or supporting details.
29. 3pt Read the paper, Evidence that the adaptive allele of the brain size gene *microcephalin* introgressed into *Homo sapiens* from an archaic *Homo* lineage, by Evans, et al. (found [here](#))? What is the conclusion of the paper? What is the evidence for this conclusion? And are there any alternative explanations for the findings.