

Name:

Are you a graduate or undergraduate student? Please circle one.

Bioinformatics Take Home Test #6

Due Date Monday 11/11/2013 before class

(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.

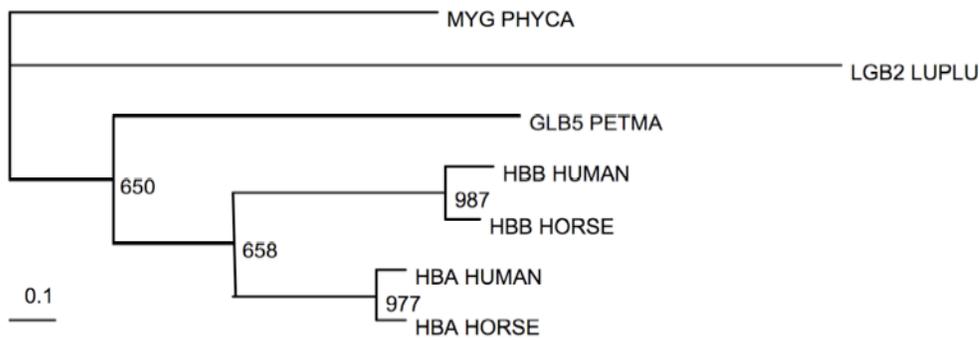
If your answer uses a citation, give the source of the quoted text.)

Notes on Formatting Quizzes: Please make sure each answer is only on one page, by using page breaks. Splitting an answer onto two pages tend to lead to grading errors.

Please do not write or type in font smaller than 12 point or write in cursive.

If you submit your quiz via email, please remove the instructions and extras (blank lines, alternative answers for multiple choice questions) from your document, so that only your answers, a minimal amount white space, and optionally the questions, are left.

1. **1pt True/False** Phylogenetic trees of flu virus strains are used every year in the decision making process with regards to which strain to immunize for, because phylogenetic information is used in identifying which strain will spread.
2. **1pt True/False** Paralogs ARE a type of homolog.
3. **1pt True/False** Orthologs are NOT a type of homolog.
4. **1pt True/False** Xenologs are homologs that have been acquired from an endosymbiont.
5. **1pt True/False** A parametric bootstrap is when datasets are simulated using a model of evolution, based on the original dataset, while a nonparametric bootstrap is your basic resampling of the original dataset.
6. **1pt True/False** Branch lengths ARE considered when calculating the bootstrap support of a branch in a phylogenetic tree.
7. **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.



The above tree 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.

8. **1pt True/False** The number 977 indicates that in 97.7% of the bootstrapped samples “HBB HUMAN” groups with “HBB HORSE”.
9. **1pt True/False** The number 650 indicates that in 65.0% of the bootstrapped samples “MYG PHYCA” , “LGB2 LUPLU” , and “GLB5 PETMA” group together.
10. **1pt True/False** In the above tree, the number 658 does NOT indicates that in 65.8% of the bootstrapped samples “MYG PHYCA” , “LGB2 LUPLU” , and “GLB5 PETMA”
11. **1pt True/False** In the above tree, the number 650 indicates that in 65.0% of the bootstrapped samples “HBA HUMAN” , “HBA HORSE” , “HBB HUMAN” , and “HBB HORSE” group together.
12. **1pt True/False** Based on the above tree it is NOT possible that the GLB5 PETMA groups with MYG PHYCA in 35% of the analyses performed on the bootstrapped samples.
13. **1pt True/False** Based on the above tree is possible that the GLB5 PETMA groups with HBB HUMAN in 35% of the analyses performed on the bootstrapped pseudosamples.

14. **1pt True/False** Based on the above tree is possible that the GLB5 PETMA groups with (HBB HUMAN, HBB HORSE) in 35% of the analyses performed on the bootstrapped pseudosamples.
15. **1pt True/False** Based on the above tree it is possible that HBB Horse groups with (HBA Human, HBA Horse) in 21% of the analyses performed on the bootstrapped samples.
16. **1pt True/False** 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.
17. **1pt True/False** The outgroup of a phylogenetic tree represents the ancestor of the ingroup.
18. **1pt True/False** The outgroup in a phylogenetic tree is used to inform upon the ancestral state of a given character.
19. **1pt True/False** Choosing the wrong outgroup can introduce long branches into a tree and cause Long Branch Attraction.
20. **1pt True/False** Long Branch Attraction can lead to recovering the wrong tree with 100% bootstrap support.
21. **1pt** Which of the following pairs is an example of orthologs? There is more than one answer.
- The ATPsynthase and the RHO termination factor
 - The ATPsynthase and the Flagellar ATPase
 - The catalytic and noncatalytic subunits of the ATPsynthase
 - The Bacterial F-type ATPsynthase and the Vacuolar-type ATPase
 - Alpha and Beta hemoglobin
22. **1pt** What are the reason why a gene tree could be different from the species tree?
- Lack of resolution
 - Lineage sorting
 - Gene duplication or loss
 - Horizontal gene transfer
 - All of the above
23. **1pt** Phylogenetic reconstruction based on the parsimony principle aims to find the phylogenetic tree
- A) that explains the evolutionary history that gave rise to the aligned

sequences with the least number of substitution events.

B) that is most probable given the data.

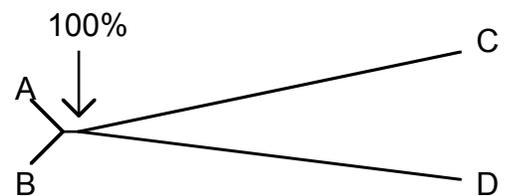
C) under which the data set (e.g., aligned sequences) is most probable.

D) that is in the best possible agreement with the observed number of substitutions observed between the sequences.

24. **1pt** When analyzing phylogenetic trees, what do bootstrap values mean?

- Nothing; it is impossible to pull yourself up by your bootstraps and the process is therefore worthless.
- The number of datasets that support a given clustering, when evolution is re-run.
- The number of datasets created by resampling the columns of an alignment with replacement, which agree with a split.
- The length of time two sequences have been evolving away from each other.
- None of the above.

25. **1pt** When analyzing a quartet of putatively orthologous sequences, the maximum parsimony tree looks like this, with the central branch having 100% bootstrap support:

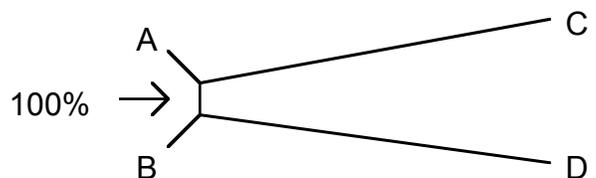


A) Maximum parsimony is not subject to the long branch attraction artifact, rather it always has the tendency to group the long with the short branches. Therefore, the finding that A and B group together is reliable.

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.)

26. **1pt** When analyzing a quartet of putatively orthologous sequences, the maximum parsimony tree looks like this, with the central branch having 100% bootstrap support:



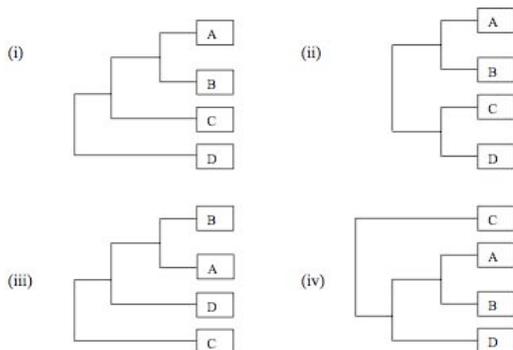
A) Maximum parsimony is not subject to the long branch attraction artifact, but always has the tendency to group the long with the short branches (aka long branch repulsion). Therefore, the finding that A and C group together is unreliable.

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.)

27. **1pt** How is a neighbor joining tree calculated?
- By maximizing the probability of a dataset, given the model.
 - By maximizing the probability of a model, given the data.
 - By minimizing the number of substitution events.
 - By using pairwise comparisons to determine the nearest neighbor and then collapsing a node and recalculating all of the pairs for the node.
 - By programming a little robot to walk around in tree space.
28. **1pt** Phylogenetic reconstruction using Markov chain Monte Carlo sampling aims to find the phylogenetic tree
- that explains the evolutionary history that gave rise to the aligned sequences with the least number of substitution events.
 - that is most probable given the data.
 - under which the data set (e.g., aligned sequences) are most probably.
 - that is in the best possible agreement with the observed number of substitutions observed between the sequences.
29. **1pt** Phylogenetic reconstruction based on the maximum likelihood principle aims to find the phylogenetic tree
- that explains the evolutionary history that gave rise to the aligned sequences with the least number of substitution events.
 - that is most probable given the data.
 - under which the data set (e.g., aligned sequences) is most probable.
 - that is in the best possible agreement with the observed number of substitutions observed between the sequences.
30. **1pt** Which of the following **rooted** trees have identical topology when considered as unrooted?



- Tree (i) and (ii)
- Tree (ii) and (iii)
- Tree (iii) and (iv)
- Tree (i) and (iii)
- All trees a different topology
- Tree (i) and (iv)
- All trees are identical in their topology

Extra credit:

1. **Max of 2pts** Models used to describe sequence evolution frequently use the Gamma distribution, using the alpha parameter. What is the name of the process often described by the Gamma distribution? Why is the Gamma distribution more useful than the normal distribution?
2. **1pt** Often it is said that humans are the highest life form, because we are more highly evolved than all the other animals. Is this correct? Can *any* species be more evolved than another?

For Graduate Students: Short essays please.

1. **3pt** Who was Walter Fitch? What contributions did he make to phylogenetics and bioinformatics (at least one for each)? Why were his contributions important? What is remarkable about his role in the design of a vaccine (hint, one of his contributions should involve vaccine development)?