Name:

Bioinformatics Take Home Test #7

Due Date Monday 11/28/2016 before class

MORE THAN ONE ANSWER MIGHT BE CORRECT

1. I slice an alignment up by columns, put each column in a hat, and pick a column from the hat at random, writing it into a new dataset. I put the column back in the hat and randomly pick another, repeating this process until I have the same number of columns as the original dataset. What have I created?
A. Crap
B. A jackknife sample
C. A single nonparametric bootstrap sample
D. A phylogenetic tree
E. A single parametric bootstrap sample

2. If evolutionary parameters are estimated from a dataset and associated tree and those parameters are then used to simulate a new dataset, what is this dataset?
A. Crap
B. A jackknife sample
C. A single nonparametric bootstrap sample
D. A phylogenetic tree
E. A single parametric bootstrap sample

3. True/False Parsimony does a better job handling gaps than Neighbor Joining, but Neighbor Joining can do better with long branches (provided a correction for multiple substitutions is applied).

4. Which of the following are reasons a gene tree may not match the species tree?
A. Incomplete lineage sorting
B. Symplesiomorphies are mistaken for synapomorphies
C. Unrecognized gene duplication followed by gene loss
D. Long branch attraction
E. Insufficient phylogenetic signal
F. Horizontal gene transfer

5. Which of the following is a tree reconstruction artifact?
A. Incomplete lineage sorting
B. Horizontal gene transfer
C. Insufficient phylogenetic signal
D. Long branch attraction
E. Unrecognized paralogy

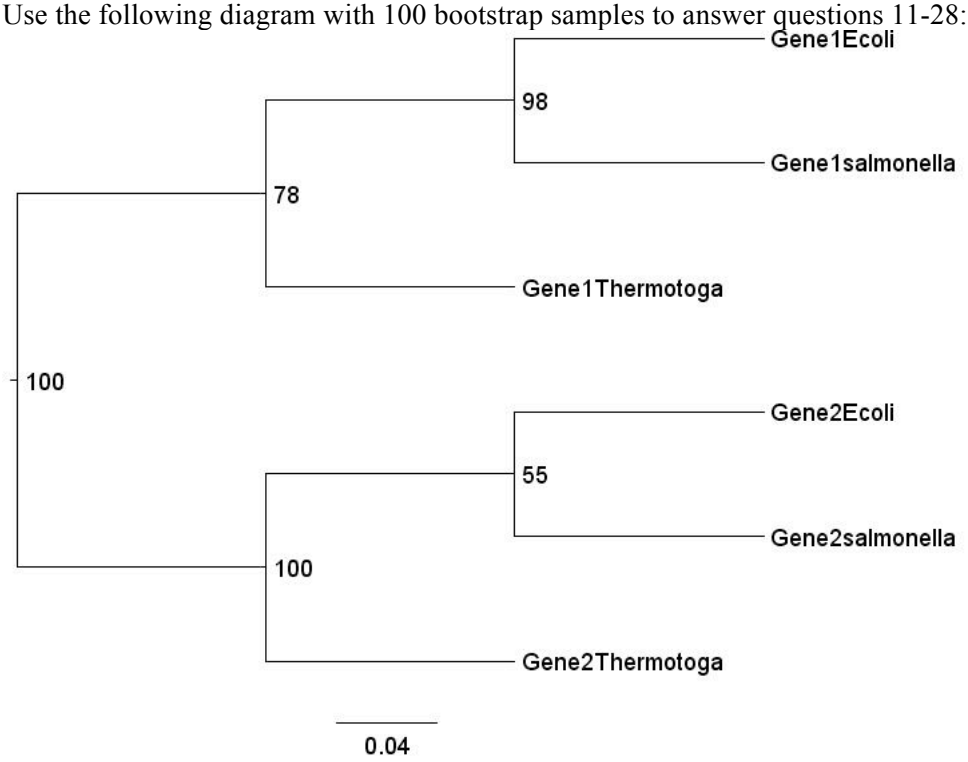
6. Long Branch Attraction is caused by which of the following?
A. Homoplasies resulting from the long branches independently acquiring the same substitution.
B. Alignment programs misalign sequences to maximize similarity
C. Tree building programs underestimating the number of substitutions occurring
D. All of the above.
E. None of the above.

7. Showing Branch lengths on a tree gives what sort of information?
A. How much evolution has occurred along a branch
B. An indication if long branches might be a problem
C. An indication that a strictly bifurcating tree is not the best model, due to a polytomy
D. None of the above

8. Which of the following is true with respect to the outgroup of a phylogenetic tree?
A. To minimize LBA artifacts, it should be as closely related to the ingroup as possible
B. To minimize LBA artifacts, more than one sequence could be used in the outgroup.
C. It can be a random sequence
D. It does not matter how distantly related it is to the ingroup; distantly related outgroups are just as informative as closely related ones.

9. Which of the following is a factor when calculating the bootstrap support of a branch in a phylogenetic tree?
A. Branch lengths
B. The number of times a split is recovered in a set of bootstrap samples
C. The Gamma parameter and among site rate variation
D. Lineage sorting

Use the following diagram with 100 bootstrap samples to answer questions 11-28:



10. What does the line at the bottom, babeled with 0.04 represent?
A. This is meaningless, other than to separate the estimate of 0.04 from the tree
B. The value of the gamma shape parameter estimated for this tree
C. The percent of bootstrap samples expected to produce that split if the sequences were random
D. The scale bar, used to indicate branch length in the tree

11. What does the number 0.04 refer to?
A. The percent of bootstrap samples expected to produce that split if the sequences were random
B. The average number of substitutions per site in a branch of that length
C. The degree to which among site rate variation occurs in the dataset
D. The percent of long branch attraction occurring in the tree
E. None of the above

12. True/False The number 55 indicates that in 55 % of the bootstrapped samples the
A) Gene2Thermotoga groups with Gene1
B) three Gene2 sequences group together.
C) Gene2Ecoli and Gene2salmonella group together.

Use the following sequence labels for the following bifurcation table

1. Gene1Ecoli
2. Gene1Salmonella
3. Gene1Thermotoga
4. Gene2Ecoli
5. Gene2Salmonella
6. Gene12Thermotoga.

13. Which bipartitions are represented in the above tree?
A) ***…
B) ..****
C) ….**
D) …***

14. Which of the following bipartitions is incompatible with the above tree?
A) .*…*
B) *….*
C) ..*.**
D)..**..

15. True/False Gene2 can be used to root Gene1 and vice versa.

16. True/False It is not possible that the grouping of the three Gene2 sequences together is due to Long Branch Attraction, because the bootstrap support is 100%.

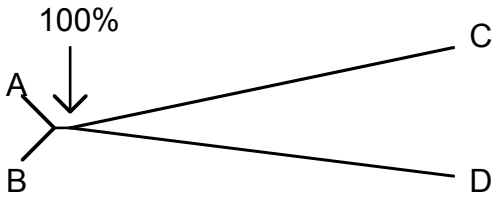
17. The topology of Gene 1 compared to the topology of Gene2 indicates that which of the following has most likely occurred in these sequences?
A. Incomplete lineage sorting
B. Long Branch Attraction
C. Long Branch Repulsion
D. Vertical inheritance
E. Horizontal Gene Transfer

18. What is a Xenolog?
A. A homolog resulting from gene duplication.
B. A gene with the same function as another, but evolved independently.
C. A homolog resulting from the hybridization of two species.
D. A homolog resulting from Horizontal Gene Transfer.
E. A homolog resulting from a speciation event.

19. In the PHYLIP package, Trees written onto "outtree" are in the:
A. Newick format
B. Mathlab Format
C. Java Format
D. C++ Format
E. PHYLIP Format

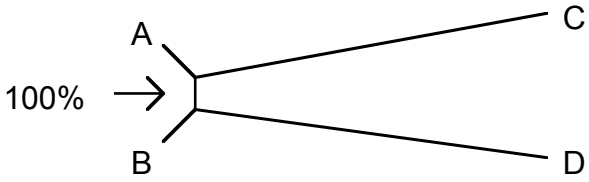
20. Parsimony aims to build the tree that?
A) under which the data set (e.g., aligned sequences) is most probable.
B) is most probable given the data.
C) explains the evolutionary history that gave rise to the aligned sequences with the least number of substitution events.
D) is in the best possible agreement with the observed number of substitutions observed between the sequences.

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



A. This tree groups the two long branches together. The possibility exists that this result might represent a long branch attraction artifact.
B. 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.)
C. 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.

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

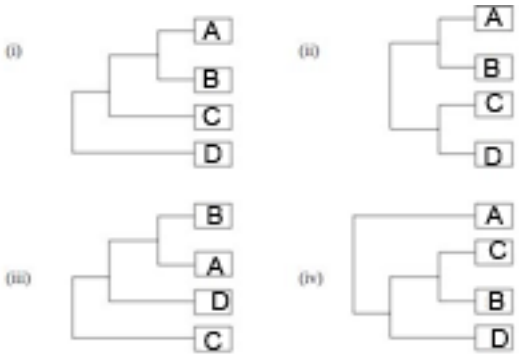


A) Maximum parsimony is subject to the long long branch repulsion artifact. Therefore, the finding that A and C group together is unreliable.
B) This tree does not group the two long branches together, indicating that the result is not due to long branch attraction.
C) The central branch is so strongly 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.)

23. True/False Phylogenetic reconstruction using Markov chain Monte Carlo sampling aims to find the phylogenetic tree that is most probable given the data by walking around in tree space with a biased walk and sampling the trees.

24. Maximum likelihood aims to build the tree
A) that is most probable given the data.
B) that explains the evolutionary history that gave rise to the aligned sequences with the least number of substitution events.
C) that is in the best possible agreement with the observed number of substitutions observed between the sequences.
D) that makes the data set (e.g., aligned sequences) most probable.

25. Which of the following rooted trees does NOT have an identical topology when considered as unrooted?



A) Tree (i)
B) Tree (ii)
C) Tree (iii)
D) Tree (iv)
E) All trees are identical in their topology

26 Models used to describe sequence evolution frequently use the Gamma distribution, using the alpha parameter.
A. What is the name of the process often described by the Gamma distribution?

B. Why is the Gamma distribution more useful than the normal distribution?

27 True/False A substitution is a mutation that was fixed in a population.

28 The mutation rate is _____ the substitution rate for a mutation that provides a selective **advantage**?
A. equal to
B. less than
C. greater than
D. proportional to
E. unrelated to

29 The mutation rate _____ the substitution rate for a mutation that provides a selective **disadvantage**?
A. Is equal to
B. Is less than
C. Is greater than
D. proportional to
E. Is unrelated to

30 The mutation rate _____ the substitution rate for a selective **neutral** mutation?
A. Is equal to
B. Is less than
C. Is greater than
D. proportional to
E. Is unrelated to

31. True/False The neutral theory states that all evolution is neutral and everything is only due to genetic drift.

32. In a **haploid** population, what is the probability that the mutant allele will replace the original allele, if the mutant is **neutral** with respect to the original allele?
A. 1/N
B. 1/2N
C. 1/4N
D. 2*s/(1-e ^(-4*N*s))
E. 2S

33. In a **diploid** population, what is the probability that the mutant allele will replace the original allele, if the mutant is **neutral** with respect to the original allele?
A. 1/N
B. 1/2N
C. 1/4N
D. 2*s/(1-e ^(-4*N*s))
E. 2S

34. True/False - Most mutations disappear in a few generations due to random drift.

35. True/False - If the mutant allele reaches a frequency of 50% in a population, it will almost always go on to fixation, even if the mutation does not provide a selective advantage.

36. The size of successive populations is 500, 10000, 1 billion, 4 billion, 8 billion. What is the "effective population size" for the 5 generations (ignoring spatial heterogeneity, mating etc.) ?

37 What is the chance (probability) that a mutation that arose in a single copy and that provides no selective advantage or disadvantage is fixed in population of 125 haploid organisms?

38. The **time till** fixation for a single neutral mutation in a population of 15 million individuals as compared to a population of 1,000 individuals is
A) the same, B) shorter, C) longer

39. The **probability for fixation** for a single neutral mutation in a population of 15 million individuals as compared to a population of 1,000 individuals is
A) the same, B) lower, C) higher

33. Which of the following bipartitions is not compatible with the others?
A) ......***
B) ....*****
C) ***......
D) ....**
E) ...**....
F) ....**...

34 On average it takes 4*Ne generations (Ne is the effective population size) until a neutral mutation is fixed in a diploid population. An advantages mutation is fixed already after (2/s) ln (2Ne) generations.
How many generations will it take a neutral mutation and a mutation with s=.15 to become fixed in populations of 5000, 50000, 500 000, and 5 million individuals.

Average time to fixation in number of generations:

Effective population size:	5000	50000	500000	5,000,000
advantages mutation with s=0.15				
Neutral				

Extra credit:
The fixation probability for an advantages mutation (that provides a selective advantage s) in a large diploid population is given as 2*s. In the following table enter the probability for fixation for a neutral mutation and for a mutation that provides a selective advantage s of 0.005. Obviously the formula for the fixation probability of an advantageous population does not work for a small population. In the fourth row use the indicated exact formula to calculate the fixation probability.

Effective population size	20	200	2000	20,000
Fixation probability for advantages mutation with s=0.005 using 2*s				
Fixation probability for Neutral mutation				
Fixation probability for advantages mutation with s=0.005 (using 2*s/(1-e ^(-4*N*s))				