Due on Friday, December 9th

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

Bioinformatics Take Home Test #9

(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 What OSX application can be used to access the cluster through a command line interface?
   a. Cluster
   b. Clustal
   c. Terminal
   d. Apple File-sharing Protocol
   e. Muscle

2. 1pt Which of the following is the host name of the cluster in the biotech center?
   a. mcb221u016
   b. bbcxsrv1.biotech.uconn.edu
   c. mcb221u016@bbcxsrv1.biotech.uconn.edu
   d. It is variable, because it is defined by the user once you log on for the first time
   e. It doesn’t have one, because the computer automatically remembers the cluster

3. 1pt When using the cluster, what is the first thing that has to be typed after establishing a connection and entering your password?
   a. cd folder_name (i.e. whatever you named the folder your files are in)
   b. qrsh
   c. your user name and password
   d. mcb221u016@bbcxsrv1.biotech.uconn.edu
   e. Nothing, you can go directly to running your program

4. 1pt What does the command qrsh do?
   a. Tells the cluster to reset your password
   b. Tells the cluster to use a secure connection to your computer, so that other people cannot hack the connection
   c. Log you onto a subnode, so that your programs are not being run on the head node; running things on the headnode is very bad, because it will bog down the entire cluster, or worse, crash it.
   d. Log you onto the headnode, so that your programs are not being run on a subnode. Programs must be run on the headnode, where they are installed.
   e. Allows you to move from your home directory to the folder where your files are stored.

5. 1pt What do the commands cd and ls do?
   a. cd allows you to change directories and ls lists the contents of the directory you are currently in
   b. cd lists the contents of the directory you are currently in and ls allows you to change directories
   c. cd logs you onto a subnode and ls logs you onto the headnode
   d. cd logs you onto the headnode and ls logs you onto a subnode
   e. cd transfers a file from your computer to the cluster and ls transfers a file from the cluster to your computer
6. 4 pt

A

What tree topology does the probability vector corresponding to point 1 support?

\((1,2),3,4\)  \((1,3),2,4\)  \((1,4),2,3\)

How reliable is the support?

**Strong / Weak**

B) What tree topology does the probability vector corresponding to point 2 support?

\((1,2),3,4\)  \((1,3),2,4\)  \((1,4),2,3\)

How reliable is the support?

**Strong / Weak**

7. 1pt When considering data obtained from flipping one coin three time and obtaining all heads, what will maximum likelihood calculate? (Hint- Consider that there are three models possible for this coin toss and that the first one is twice as likely as the other two: 1. A fair coin model. 2. A coin with both sides heads. And 3. A coin with both sides tails.)

a. The probability of obtaining all heads, averaged over all possible models
   (i.e. \((.5)^3 * .5\) + (.25 * 1.0) + (.25 * 0))

b. The probability of obtaining all heads, given the model that maximizes this probability
   (i.e. 100% and it will always chose the second model)

c. The probability of obtaining all heads when using a fair coin (i.e. \((.5)^3 * .5\))

d. The probability of obtaining all heads, without considering possible models. This is possible because a robot is used to explore probability space.

e. Maximum likelihood is not applicable to coin toss data, only nucleotide or amino acid sequence data can be used.
8. 2pt - grad students only What could be possible goals of a Bayesian consideration of the coin toss example?

9. 1pt When considering the above coin toss experiment in a Bayesian framework, which other quantity do you need to know to arrive at a conclusion?
   A) the posterior probability
   B) the prior probability
   C) the joint probability of the three possible models

10. 4 pt Your health care provider performed a test for a rare genetic disease on you. The test gives a false positive result only in 1 out of a 1000 cases. The rate of false negatives is zero (if you have the disease, the test will detect it). About 1 in a million Americans have the disease. Your test is returned positive. What is the probability that you actually have the disease? (You might get partial credit, if you show your reasoning.)

11. 1pt Why do small world networks have a low the degree of separation?
   A) All nodes have many connections
   B) A few nodes with many connections act as hubs
   C) The probability of a node to have n connections is constant for all n

12. 2 pt The selfish operon theory explains the assembly of operons through which process?
   A) Tandem gene duplications lead to similar enzymes located next to one another in the genome. Because of the similarity between substrate and product, these duplications often lead to extension of the metabolic pathway at the same time as the operon is extended
   B) Following the transfer of genes that encode part of a selected function (e.g. a metabolic pathway that allows the utilization of the new substrate), intervening genes that do not contribute to this function and that were useful in the donor for other reasons, become useless in the recipient. These new useless genes will be deleted in the recipient, thereby moving the genes encoding the selected function closer together.
   C) Genome rearrangements occur frequently in prokaryotes. When two genes encoding parts of the same function happen to land next to each other in the genome so that they can share the same regulatory promoter, then the benefit from shared regulation provides a selective advantage that will keep the genes next to one another.

13. 3 pt You compare two genomes from closely related organisms using a Genome plot. Which processes that could have given rise to the following:
   (Genome 1 is on the ordinate (X), Genome 2 is on the abscissa (Y))
Panel I:
A) Genome 1 has an insertion that is not present in genome 2, or genome 2 had a deletion.
B) Genome 2 has an insertion that is not present in genome 1, or genome 1 had a deletion.
C) Genome 1 possesses a region that is duplicated.
D) Genome 1 or genome 2 underwent an inversion.

Panel II:
A) Genome 1 has an insertion that is not present in genome 2, or genome 2 had a deletion.
B) Genome 2 has an insertion that is not present in genome 1, or genome 1 had a deletion.
C) Genome 1 possesses a region that is duplicated.
D) Genome 1 or genome 2 underwent an inversion.

Panel III:
A) Genome 1 has an insertion that is not present in genome 2, or genome 2 had a deletion.
B) Genome 2 has an insertion that is not present in genome 1, or genome 1 had a deletion.
C) Genome 1 possesses a region that is duplicated.
D) Genome 1 or genome 2 underwent an inversion.

14. 2 pt Assuming that the origin of replication is at (0/0) in the coordinate system, and that you compare two closely related genomes. The organisms in question have circular genomes. In the sketched genomes on the right, indicate where the rearrangement event(s) took place that could have given rise to the genome plot on the left.