Name:
Due on Wensday, December 9th

Bioinformatics Take Home Exam #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.)
Make sure each answer is only on one page, by using page breaks. Splitting an answer onto two pages leads to grading errors.
Do not write or type in font smaller than 12 point or write in cursive. Doing so will lose you 2 pts.
If you have an emergency and cannot submit a quiz in person, email it in by the start of class on the due date. If you do so, first 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.
Note on Late Quizzes: Late quizzes are an inconvenience and cannot be accepted at all after the answers have been released. If your quiz is submitted within the first 12 hours after the deadline, you will receive 5% off. Each additional 12 hours is an additional 5% off, up until the graded quizzes are returned or the answers released.
All questions worth 1pt unless otherwise stated.

1. What process brought 2 divergent chlorophylls into the ancestor of the cyanobacteria, which enabled oxygen-producing photosynthesis?
A. Gene duplication
B. Genome duplication
C. Meiotic hybridization between species
D. Horizontal gene transfer
E. Long branch attraction

2. Which of the following is are linked to the ability to digest fruit sugars into alcohol in yeast and adaptations of yeast in modern times to human uses?
A. Gene duplication
B. Genome duplication
C. Meiotic hybridization between species
D. Horizontal gene transfer
E. Long branch attraction

3. What happened to allow some Asian people to be able to digest algae?
A. Gene duplication, followed by neofunctionalization
B. HGT from a red algal parasite to Bacteroidetes bacteria, which are symbiont in the human gut
C. HGT to humans from the red algae
D. Humans gained a new symbiont in their gut from the red algae, that can digest algae
E. None of the above

4. True/False: Smelt was the likely beneficiary of an antifreeze gene laterally transferred between fishes.

5. Which of the following is true regarding HGT?
A. It is a process through which genes enter a genome, without being inherited parentally
B. It can lead to important biological innovations
C. A transferred gene can be inherited parentally, so that a clade of organisms all share the same inherited ancient HGT.
D. It is more common in Bacteria than in humans.
E. All of the above.

6. Which processes allow favorable genetic changes to be combined into the same individual, speeding up the rate of evolution?
A. Gene duplication and neofunctionalization
B. Genetic drift
C. Punctuated equilibrium
D. Sex and HGT
E. None of the above.

7. Why is PSI-BLAST more prone to false positives than normal blast, especially as the number of iterations increases?
A. As the program narrows down possible homologous proteins, each new protein is more likely to be recognized as homologous simply because there are less proteins remaining in the pool.
B. With every iteration, there is the possibility that false positives are incorporated into the position specific scoring matix.
C. The program creates a more comprehensive profile-HMM as more sequences added with each iteration, increasing the likelihood of false positive because more protein domains are being scored.
D. All of the above

8. What does the PSI in PSI-BLAST stand for?
A. Phylogeny Sequence Initiative
B. Phylogeny Stimulated Image
C. Position-Specific Iterated
D. Position-Sequence Initiated
E. None of the above

9. What are false negatives?
A) non-homologous sequences that are listed as matches
B) homologous sequences that are not detected
C) homologous sequences that are detected
D) non-homologous sequences that are not listed as matches

10. True/False PSI Blast has more false positives than normal Blast because of profile corruption.

11. Which program is specifically designed to detect distant relationships between proteins? (B. PSI-BLAST)
A) Seaview
B) PSI-BLAST
C) njplot
D) Swiss-pdb
E) None of the above

12. What is a PSSM?

13. Which computing program uses MCMC algorithms? (MrBayes)
A. MrBayes
B. Seaview
C. PSI-BLAST
D. Njplot
E. Swiss pdb

14. True or False Mutation plays little to no role in the evolution of organisms; selective processes account for a the majority of evolution.

15. True/False: There are fewer false negatives with PSI blast than with normal blast.

16. True/False MrBayes is extremely reliable in predicting the correct tree. As a result, the values that the trees produce should be believed if they are comparable to bootstrap values.

17. **True/False** psiBLAST is an algorithm HMMER uses to find distant homologs to a query sequence.

18. Which program can align nucleotide sequences based on a protein alignment?
A. Mr.Bayes
B. Seaview
C. psiBLAST
D. Clustal
E. Cluster

19. Not doing which of the following commands at the beginning of a session on the cluster is considered rude, because it will crash the head node?
A. ls
B qlogin
C qstat
D cd
E qdel

20. Why might Amino acids on the outside of virus capsids be under positive selection?
A. They interact with the immune system and need to change to evade capture
B. These positions are under strong selection to maintain function, because they are really important to the virus
C. They interact with the host DNA and need to change as the host evolves
D. Binding of host antibodies triggers mutations in the virus
E. All of the above

21. What allowed Walter Fitch to beat of the CDC in picking the strain of flu to vaccinate, year after year, until the CDC finally started doing things his way?
A. He had a huge team of researchers on the project, while the CDC just had one retired professor on the project
B. He had enormous computing power at his disposal, while the CDC was using a pocket calculator
C. He had new, modern laboratory equipment, allowing him more to obtain more accurate sequences than the CDC
D. He had decades of knowledge of evolution and with it was able to develop the novel idea of using an evolutionary pattern he spotted in flu phylogenies to predict next years epidemic
E. A, B, and C

22. The Bayesian framework calculates?
A) The probability of the model given the data is assessed
B) The probability of the data given the model is assessed
C) This is the same as maximum likelihood analysis
D) Both a) and c)
E) Both b) and c)

23. The Maximum Likelihood principle calculates?
A) The probability of the model given the data is assessed
B) The probability of the data given the model is assessed
C) This is the same as Bayesian framework analysis
D) Both a) and c)
E) Both b) and c)

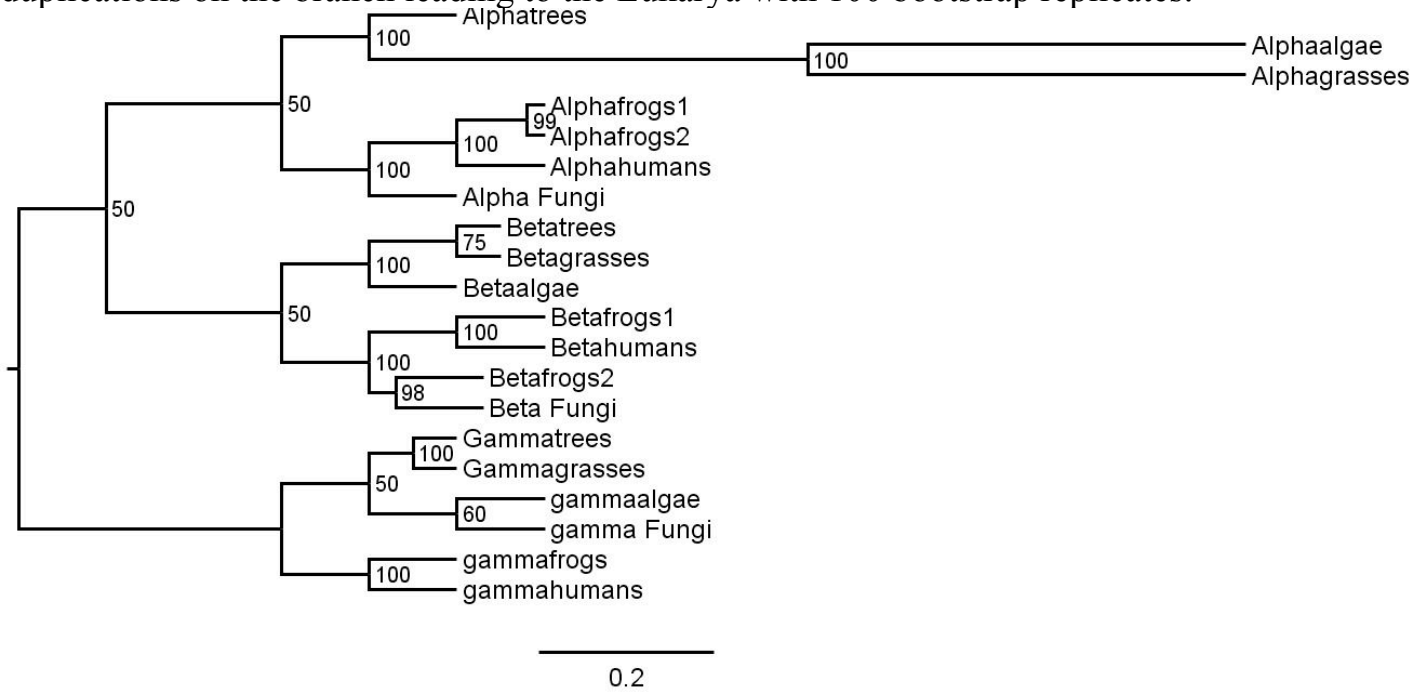
24. When considering data obtained from flipping one coin four times and obtaining all tails, what will maximum likelihood calculate? (Consider that there are three models possible for this coin toss: 1. A fair coin model. 2. A coin with both sides heads. And 3. A coin with both sides tails. Priors are 1. 99.8%, 2. 0.1%, 3. 0.1%)
A. The probability of obtaining all tails, averaged over all possible models
(i.e. ((.5)^4 * 0.998) + (0 * 0.001) + (1.0 * 0.001))
B. The probability of obtaining all tails, given the model that maximizes this probability
(i.e. 100% and it will always chose the third model)
C. The probability of obtaining all tails when using a fair coin (i.e. (.5)^3 * 0.998))

D. The probability of obtaining all tails, 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.

26. If you are given a task to do on the cluster, what program do you need to open? Choose the one that applies to your operating system
A. Command Line
B. Terminal
C. Clustal
D. SSH secure shell client
E. Filezila
F. Firefox

28. What is NOT considered an example of horizontal gene transfer producing new pathways in organisms?
a) The evolution of archaeal ATP synthase
b) Oxygen-producing photosynthesis
c) Acetoclastic methanogenesis
d) Acetyl-CoA assimilation
e) All are examples

For questions 28-37 use the following diagram showing three genes related by ancient gene duplications on the branch leading to the Eukarya with 100 bootstrap replicates:



28. What type of diagram is this?

29. What does the vertical axis represent?
A. The amount of evolution, in substitutions per site
B. The amount of time, in years
C. The rate of evolution in substitutions per site per year
D. A, B, and C
E. This axis is meaningless and is adjusted for visualization purposes

30. 2pts The following is a list of phenomena occurring that makes this tree deviate from the species tree. Match them to the deviations below:
i. In-paralogs resulting from a recent gene duplication
ii. Horizontal gene transfer
ii. Long branch attraction
iv. Difficult to tell, but could be a result of lack of resolution

A. Beta2frongs grouping with Betafungi

B. Alphaalgae grouping with Alphagrasses

C. Gammaalgae grouping with Gammafungi

D. Two copies of the alpha gene being present in frogs

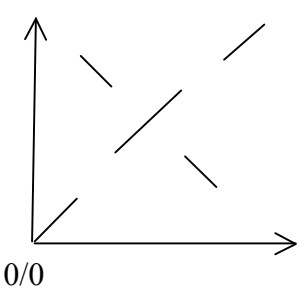
31. Might whole genome duplication have played a role in forming this tree (Yes or No)? If yes, how many rounds is most likely to have occurred?

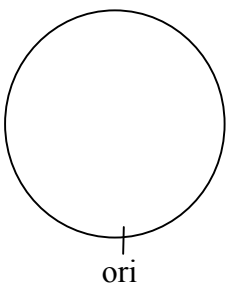
32. Is it possible that gammaalgae groups with the 2 gamma plant sequences in 50% of bootstrap samples? If not, what is the maximum number of samples?

33. According to the Alpha and Beta paralogs, where is the root within the Eukarya?

34. True/False The 60 means that 60% of the time, the two gamma plant sequences group with the two gamma animal sequences and all of the alpha and beta sequences.

35. 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





Extra Credit

1pt You want to find all copies of a transposase gene in a particular microbial genome. A blastp search of the annotated genome resulted in 12 significant hits. A PSI-blast search of the annotated genome using a PSSM calculated from first searching nr for 5 generations resulted in 16 significant matches. A PSI-blast search of the 6 frame translation of the genome gives 42 significant matches.
Explain why there are additional matches obtained in PSI blast searches?

1pt – Why is a Bayesian consideration advantageous, as opposed to the Maximum Likelihood estimate, for the coin toss example used in previous questions?

1pt Molecular evolution of which macromolecule is important in flu vaccine development? (more than one may be correct)
Hemagglutinin
Neuraminidase
Inteins
Lipid bilayer
DNA
ATPsynthase