

Due on Monday, November 18th

Name:

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

Bioinformatics Take Home Test #7

(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 does the likelihood value of a tree refer to?
 - A) the probability of the data given the tree and a model of evolution
 - B) the probability of the tree given a model and the data
 - C) the maximized probability assuming that a given tree topology is true
2. **1pt** How is does the program MrBayes assess posterior probabilities?
 - A) Through quartet puzzling
 - B) Through calculating P_i as $L_i/(L_1+L_2+L_3)$, where L_i is the maximum likelihood of tree with topology i
 - C) Through a biased random walk through tree and parameter space.
3. **1pt True/False** Phylogenetics has been used to trace HIV history in court cases.
4. **1pt** Molecular evolution of which macromolecule is important in flu vaccine development? (more than one may be correct)
 - a. Hemagglutinin
 - b. Neuraminidase
 - c. Lipid bilayer
 - d. DNA
 - e. ATPsynthase
5. **1pt True/False** We can use phylogenies to infer functions.
6. **1pt** What is the impact of population size on genetic drift?
 - a. Large populations are more susceptible to genetic drift than small ones.
 - b. Small populations are more susceptible to genetic drift than large ones.
 - c. Large populations eliminate the possibility of genetic drift occurring.
 - d. Small populations eliminate the possibility of genetic drift occurring.
 - e. Population size in not correlated with genetic drift.
7. **1pt** The size of successive populations is 500, 8 billion, 8 billion, 8 billion, 8 billion. What is the "effective population size" for the 5 generations (ignoring spatial heterogeneity, mating etc.) ?
8. **1pt True/False** Mutations refer to a change in the nucleotide sequence of a gene, in contrast a substitution refers to a change in the amino acid sequence.
9. **1pt** 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 50 diploid organisms?

10. **1pt** 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 500 haploid organisms?
11. **1pt True/False** For mutations that provide a selective advantage, the mutation rate and the substitution rate are the same.
12. **1pt True/False** For a mutation that does not provide a selective advantage, the mutation rate and the substitution rate are the same.
13. **1pt** You consider mutations that provide no selective advantage or disadvantage. The mutation rate is assumed to be constant. You compare two populations of different size over long periods of time.
 A) The substitution rate in the smaller population would be higher, because the mutations that occur are more easily fixed in the smaller population due to genetic drift.
 B) The substitution rate is approximately the same in the two populations.
 C) The larger population generates more mutations and therefore has higher diversity and a higher substitution rate.
14. **2pt** On average it takes $4 * N_e$ generations (N_e is the effective population size) until a neutral mutation is fixed in a diploid population. An advantages mutation is fixed already after $(2/s) \ln(2N_e)$ generations.
 How many generations will it take a neutral mutation **and** a mutation with $s=.01$ to become fixed in populations of 1000, 10000, 100 000, and 1 million individuals.

Average time to fixation in number of generations:

Effective population size:	1000	10000	100000	1,000,000
advantages mutation with $s=0.01$				
Neutral				

15. **1pt** Most of the mutations one observes in the population are
 A) under diversifying selection
 B) neutral or nearly neutral
 C) under positive selection
 D) under negative selection
 E) under balancing selection
 F) under purifying selection
16. **1pt** Kimura's theory of neutral evolution says that
 A) Most mutation are selectively neutral
 B) Most differences between alleles coexisting in a population are selectively neutral
 C) Evolution by natural selection does not occur
17. **1pt True/False** Different parts of a protein experience substitutions with different rates. This shows that the mutations in different parts of the protein are experiencing selection pressure at different strengths. This observation is incompatible with Kimura's theory.

18. **1pt** The time till fixation for a single neutral mutation in a population of 10 individuals as compared to a population of 200,000 individuals is
A) the same, B) shorter, C) longer
19. **1pt** The probability for fixation for a single neutral mutation in a population of 10 individuals as compared to a population of 200,000 individuals is
A) the same, B) lower, C) higher

A new mutant allele has just arisen in a population of 100 haploid individuals.

20. **1pt True/False** - If the mutant is neutral with respect to the original allele, there is a 1% probability that the mutant allele will replace the original allele.
21. **1pt True/False** The mutation is very likely to disappear in a few generations due to random drift.
22. **1pt True/False** If the mutant allele reaches a frequency of 50%, it will almost always go on to fixation, even if the mutation does not provide a selective advantage.
23. **2pt** Regarding the ratio of non-synonymous to synonymous substitutions (d_n/d_s) which statement is correct?
A) If the ratio =1, no selection is evident.
B) A ratio <1 corresponds to purifying selection.
C) A ratio >1 corresponds to positive or diversifying selection.
D) A, B, and C are correct.
E) None of the above.
24. **1pt** The d_N/d_S ratio expected for a pseudogene is 1.
25. **1pt True/False** Genes in Eukaryotes always contain more exon than intron sequences (measured in number of base pairs)?
26. **1pt** Which of the following is NOT a program used to view phylogenetic trees?
a. NJ Plot
b. FigTree
c. Protpars
d. TreeView
e. None of the above
27. **1pt** Which of these is the correct spelling of the plural and singular forms of a domain name?
a. Archae and Archaeum
b. Archaea and Archaeum
c. Archaeobacteria and Archaeobacterium
d. Archaeobacteria and Archaeobacterium
e. Archaea and Archaeon

(extra credit: If you think more than one may be considered correct, give a short explanation why the pair should be considered)

Extra credit:

28. **2pt** The fixation probability for an advantageous mutation in a large **diploid** population is given as $2s$. In the following table enter the probability for fixation for a neutral mutation and for a mutation that provides a selective advantage s of 0.0025. 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	10	100	1000	10,000
Fixation probability for advantageous mutation with $s=0.0025$ using $2s$				
Fixation probability for Neutral mutation				
Fixation probability for advantageous mutation with $s=0.0025$ (using $2s/(1-e^{-4Ns})$)				

For Graduate Students:

29. **2pt** Discuss different approaches to infer genes under positive selection. On what are the different approaches based, what are possible advantages/disadvantages for the different approaches?