

Due: April 8, 2014

Total points: 25

1. (5 points) Consider the evolution of a twenty-base sequence along an evolutionary lineage. *Briefly* explain the difference between (1) the mutation rate, μ ; (2) the branch length, ν ; and (3) the quantity that serves as the x -axis of the simulation demonstration shown on page 16 of the “CalculatingLikelihood.pdf” slides.

2. (6 points) Assuming that you use the Jukes-Cantor model of sequence evolution, and you randomly select a site in the genome, answer the following questions:

(A) what is the probability that a human sequence will display a ‘C’ at the site and a chimp sequence will also show at ‘C’, if the path length between the species is 0.03 expected changes per site?

(B) What is the probability that human would have an ‘A’ and chimp would have a ‘C’?

(C) If the path length separating human and orangutan is three times as long, what is the probability that human would show ‘A’ and orangutan would show a ‘C’?

3. (5 points) *Briefly* describe the biological phenomenon which explains why the transition probability functions used in phylogenetics are not linear functions of the branch length.

4. (9 points) (A) Imagine that the Jukes-Cantor model of evolution was correct, and that you have randomly-selected 1000 sites in the genome of several mammals. 168 of the sites differ in between a pair of rodents (*Mus musculus* and *Rattus norvegicus*), what would be the best guess for the path length between this pair of species?

(B) A researcher calculates two pairwise distance matrices for these taxa. The first matrix is a distance matrix of p -distances. The second is a distance matrix using “corrected” path lengths (of the type that you calculated in part A). Which distance matrix would you expect would be closer to being an additive distance matrix? *Briefly* explain your answer.