Molecular Phylogenetics (EEOB 563)
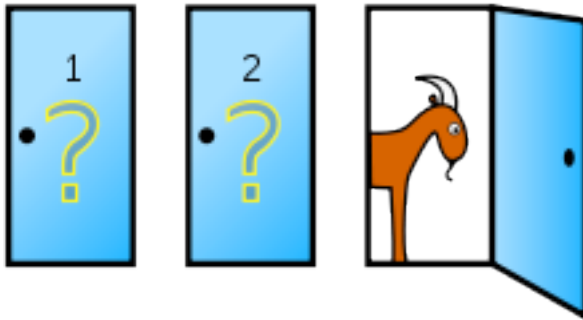Homework #6: Bayesian inference

Part I. Bayes rule using paper and pencil ☺.

1) Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?"

Is it to your advantage to switch the doors? Provide a Bayesian solution by considering posterior probabilities of the price located behind doors 1 and 2 given that the door 3 has been opened.

Part II. Bayesian inference of phylogeny.

The sequence data for this question are available at the class webpage (hiv.nxs).

The data consist of 11 sequences of partial HIV-1 envelope gene fragments. The sequences were collected for a court case where complainant B claims that he was infected by defendant A. A is countersuing B, because he claims that was actually infected by B. Sequences labeled with Ax or Bx are those obtained from A or B, respectively, and sequences labeled Lx are obtained from local controls who are drawn randomly from the population of infected individuals. One sequence is a reference sequence (HIVRF). You can read more about the case in Science 256:1165-1171.

2) Use MrBayes and averaging over the GTR family of models to reconstruct the history of infection. What is the probability, under the models you are using, and given the data you have, that the sequences of B and A form a clade? Give an argument with respect to the direction of infection.

3) What substitution model had the highest posterior probability?

4) Give evidence to show that your MCMC simulation has converged to equilibrium, and that your conclusions are insensitive to the starting tree. State the length of the MCMC simulation you made, and any subsampling, *i.e.* give the (X, Y, Z) values you used in the line mcmcp ngen=X nchains=1 printfreq=Y samplefreq=Z;

Extra point:
5) Test your inference on the direction of infection.