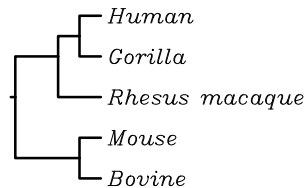


HOMEWORK #1 – DUE IN TWO WEEKS

For this problem set, we'll consider the following DNA sequence data:

Mouse	ACCAAAAAAAAAACATCC
Human	ACCCCACTCACCCAT
Gorilla	ACTATACCCACCCAA
Bovine	ACCAAACCTGTCCCC
Rhesus macaque	ACTTCACCCGTTTAC

- For the tree given below,
 - Find the length using Fitch parsimony (include all the sites).
 - Which of the sites in the alignment are not *parsimony informative* (i.e., have the same length on all possible phylogenies)?
 - Find the most parsimonious reconstruction at the internal nodes for site number 3.



- Note that for Kimura's two parameter model, the transition probabilities are:

$$P_{ij}(t) = \begin{cases} \frac{1}{4} + \frac{1}{4}e^{-\mu t} + \frac{1}{2}e^{-\mu t \frac{\kappa+1}{2}}, & i = j, \\ \frac{1}{4} + \frac{1}{4}e^{-\mu t} - \frac{1}{2}e^{-\mu t \frac{\kappa+1}{2}}, & i \neq j, \text{ transition} \\ \frac{1}{4} - \frac{1}{4}e^{-\mu t}, & i \neq j, \text{ transversion.} \end{cases}$$

Assuming that $\mu = 0.3$, and $\kappa = 1.5$, calculate the probability that over a time period of length $t = 0.1$,

- the base at a particular site remains the same

- (b) a transition occurs at a particular site
 - (c) a transversion occurs at a particular site
3. Next, we'll use our results from the previous two problems to think about the computation of the likelihood for the tree above.
- (a) For the tree shown above, consider the third site in the DNA sequence. Assuming that the lengths of the branches of the tree are all 0.1, that $\mu = 0.3$, and that $\kappa = 1.5$, find the likelihood for the data from site 3 only, assuming that the states of the internal nodes are as defined in the most parsimonious reconstruction that you found in Problem #1.
 - (b) What you found in the previous exercise is the likelihood of that particular tree for site 3, *given specific states at the internal nodes*. How would you find the overall likelihood for the given tree at that site?
 - (c) How would you find the likelihood of the tree given in the figure for *all* of the data above?
 - (d) How would you find the *maximum likelihood tree* for this problem?
4. We'll now consider this data set using the programs we've discussed in class so far. A NEXUS-formatted file containing the data can be downloaded from the course website: <http://www.stat.osu.edu/~lkubatko/stat882/>.
- (a) Carry out a parsimony analysis in PAUP*.
 - (b) Use Modeltest to select the best-fit evolutionary model for these data.
 - (c) Use both PAUP* and GARLI to carry out a maximum likelihood analysis for these data, using the evolutionary model you selected in part (b). To carry out the analysis in PAUP*, you will need to set the criterion to likelihood by issuing the following command after reading in the data: `set criterion=likelihood`. The other options (e.g., `hsearch`) are similar to the parsimony analysis.
 - (d) Remove one of the non-informative sites you identified in Problem #1(b) from the data set. You can do this with the `exclude` command – e.g., issuing the command `exclude 1` will remove the first character from the data set. Recompute the score of the most parsimonious tree and the likelihood of the maximum likelihood tree. Have they changed? If so, how? Why?