

Table 1:

Dolphin	ATG ACC AAC ATC CGA AAA ACA CAC CCT CTA
Hippopotamus	ATG ACA AAC ATC CGA AAA TCT CAC CCC TTA
Camel	ATG ACA AAC ATC CGA AAA TCA CAC CCA CTT
Cow	ATG ACT AAC ATT CGA AAG TCC CAC CCA CTA
Giraffe	ATG ATC AAC ATC CGA AAG TCC CAC CCA CTA

Please **complete three of the following four problems**. If you skip problem (1), you may use the distance formula from *Felsenstein* equation (11.18) to complete problems 2 and 3.

1. If the current base is x and we wait a time for there to be t base substitutions per site on average, the Jukes-Cantor model probability that the base is y is

$$p_{xy}(t) = \begin{cases} \frac{1}{4} + \frac{3}{4}e^{-\frac{4}{3}t} & \text{if } x = y \\ \frac{1}{4} - \frac{1}{4}e^{-\frac{4}{3}t} & \text{if } x \neq y \end{cases}$$

In comparing two aligned sequences, suppose that there are $n = n_1 + n_2$ sites where n_1 share the same base while n_2 are different. Write the likelihood of the observed data in terms of n_1 , n_2 , and t . Find the maximum likelihood estimate for t under the Jukes-Cantor model. (The answer is in *Felsenstein* in equation 11.18 on page 158, but expressed in terms of a distance D_S rather than n_1 and n_2 .) I want you to derive an expression from the direct likelihood expression.) Find conditions for which the maximum likelihood estimate of t is infinite.

2. Table 1 shows the first 30 bp of 1140bp of the mitochondrial cytochrome b gene for five mammals. Find the Jukes-Cantor maximum likelihood distance for each pair of species. Then, apply the algorithm on page 162 to these distances to find the UPGMA tree.
3. For the same data and distances from the previous algorithm, apply the neighbor-joining algorithm on page 167 to find a tree. Compare this tree with the UPGMA tree. How are they different?
4. Use software to find the UPGMA and neighbor-joining trees. Do one of the following three options.
 - (1) If you have PAUP* and know how to use it, go ahead.
 - (2) Alternatively, you can use the free software package PHYLIP written by Felsenstein and co-workers. The main web page is

<http://evolution.genetics.washington.edu/phylip.html>

and you can find links to either the source code or executables for several platforms. Specific instructions appropriate for your computer system are on this page.

<http://evolution.genetics.washington.edu/phylip/install36.html>

You will want to use the program `dnadist` to create a distance matrix (use the Jukes-Cantor option to compare to results from earlier problems) and then use the program `neighbor`

- (3) Or, write your own code in R or a programming language such as C or C++ to find the UPGMA and neighbor-joining trees for this data.