

Homework no. 5
Due Friday, May 6

Write a program (most conveniently using code from the previous homeworks) to read in a data set of DNA sequences.

1. Compute a table of distances between the DNA sequences using the Jukes-Cantor formula. If you don't have this formula you will find it in the projection for lecture 26 as the first formula in the panel titled "Approximate variances for distances".
2. Then construct a UPGMA tree from these distances.
3. Then (the most fun part) print it out in some interesting and relatively readable pictorial form (you can just use "character graphics" if you want). i.e., not in the Newick form but something more readable by people.

The data set is the one that can be fetched from the link in the description of lecture 29 (it is also available from the link for Homework no. 4) on the course web site. You should be able to save it by using the **Save As** function in the **File** menu of your browser. It is a set of sites from the D-loop region of mitochondrial DNA and adjacent noncoding (third codon position) sites.

Show me (by email) the output of your program, and, as usual, attach the source code for use in case something appears to be wrong.