

Homework no. 4
Due Friday, April 16

Write a program (most conveniently using code from the previous homework) to read in a Newick-format tree, and a data set of DNA sequences.

Then use the Fitch algorithm to count the number of steps a parsimony method would get for this tree. Just for the fun of it, you might want to alter the tree (by hand or using your previous program) and see what that does to the number of steps. Does rerooting the tree change it?

The tree we will use is the one you used for the previous assignment. It has branch lengths, though those will not be made use of here.

The data set is the one that can be fetched from the link in the description of lecture 26 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.

The data set format is the PHYLIP format. The first line has two integers, separated by blanks, the number of species and the number of sites. Each species starts on a new line, with 10 characters of species name (keep in mind that species that have a blank in their name have this represented by an underscore character in the tree file). Then the sequence continues, going to new lines as needed. The sequences are aligned, and all are of the same length.

The sequences are A's, T's, C's, and G's, as well as some “-” (minus sign) characters that mean “gap”. In the Fitch algorithm, “-” can be represented by the set of all four bases, {A,C,G,T}.

Show me (by email) the output of your program, and whether rerooting the tree changed the number of steps. As with the previous homework, attach your source code as well.