

## Mathematical Origins of Life

## Phylogenetic tree construction

Go to the biology workbench website

<http://workbench.sdsc.edu/>

Register for an account.

Once you are registered scroll to the bottom of the window and click on Protein tools button. This is where we will select the protein sequence for different taxa.

Choose Ndjinn multiple database search and then click run. This sends you to a list of databases containing protein sequences.

Enter the protein *myoglobin* in the space, choose 100 hits per page and then choose the SWISSPROT database then click search. Myoglobin is a protein similar to hemoglobin responsible for pigment in muscles (among other things).

After you click search you should get a list of different *myoglobin* sequences for different species. The actual sequences are not shown, but the species are. Select 10-15 species. Choose some that you assume are closely related and some that are not. Choose one that you think is distantly related to the others (the outgroup).

Import the sequences.

On the next page select all the sequences and then align them using the CLUSTALW aligning algorithm. After you click run, choose the rooted tree option and then click submit. Scroll down the page and see how the sequences align and where the mutations occur. Notice these are sequences of amino acids and not bases, but the algorithms are essentially similar to what we discussed in class. Further down the list you'll see a rooted tree. Check that the relationships make sense and the root seems to be in the right place. Are there any surprises? Using one protein may not be the most accurate way to do this. Usually you would repeat the process for other proteins or genes to find the best tree.