**Creating Phylogenies Exercise**

The purpose of this exercise is to get familiar with creating phylogenies using MEGA6.

Note of caution about MEGA, including MEGA7:

* MEGA works better on PCs than Macs.
* MEGA works only if the sequences are converted to a MEGA format.

For more information on MEGA there are a number of tutorials on the mega site at:

[www.megasoftware.net](http://www.megasoftware.net)

A number of examples in “Molecular Evolution and Phylogenetics” by Masatoshi Nei and Sudhir Kumar (Oxford University Press) have been created using MEGA.

In addition “Phylogenetic Trees Made Easy: A How-To Manual” (4th edition) by Barry G. Hall (Sinauer Associates, Inc.) uses MEGA to illustrate tree building.

Learning Goals:

* Become familiar with MEGA 7 and be able to use it to make phylogenies
* Be able to interpret the phylogenies created in MEGA

1. Using MEGA to create a multiple sequence alignment.

Open the MEGA program and click on the ‘Align’ tab and choose ‘Edit/Build Alignment’.



A new ‘Alignment Editor’ window will open with three options. Since we are creating a new alignment for this exercise, choose the ‘Create a new alignment’ option. In the next ‘Data type for alignment’ window choose the ‘Protein’ option.



After you have made your initial selections for the alignment, the multiple sequence alignment will be created in the ‘Alignment Explorer’ window. In the ‘Alignment Explorer’ window, click on the ‘Data’ tab, followed by the ‘Open’ and then ‘Retrieve Sequences from File’ button. From your Notebook (Word or Text file), grab your sequences in FASTA format and paste into the ‘Alignment Explorer’ window (CTRL+V for Windows/ Command+V for Mac). All the sequences from your file should now be loaded. To create the alignment, make sure all the sequences are **selected** in the explorer window.



In the explorer window, click on the ‘Alignment’ tab and choose the ‘Align by MUSCLE’ option. A new window with information on the default scoring parameters for MUSCLE will open. Click on the ‘Compute’ button and the program will create a multiple sequence alignment of the sequences.



Because MEGA uses it’s own formatting for files, you will need to save the alignment in a format that MEGA recognizes. Once the alignment is created, click on the ‘Data’ tab and select ‘Save Session’. Save the file by giving it a unique name. MEGA will attach the correct suffix. Close the ‘Alignment Explorer’ window. When prompted, tell MEGA to save your current alignment session as a separate .mas file. Remember where you’ve put it!

If you’ve closed your file, go to ‘File’ – then click on the menu item that says ‘Open a File/Session’ and re-open the .mas file. When asked how you would like to open the MAS file : Choose the Align option



2. Finding the best substitution model for an alignment:

While tree building is probably not that familiar to most of us trained as biochemists or molecular biologists, it’s important to remember that there are criteria for choosing which algorithm to use. One way is built into MEGA – click on “Models” then “Find Best DNA/Protein Models” – then in the dialog box click on your \*.mas file. (For a Windows computer you will need to select the All files option to see your file with the .mas extension)



Using the default parameters in the box that comes up next, click on “Compute” at the bottom of the box. In a few minutes, the program will calculate the BIC (Bayesian Inference Criteria) values – a measure of the posterior probabilities, which come from learning something about the data (Hall). According to MEGA “Models with the lowest BIC scores are considered to describe the substitution pattern the best.”



Based on the return, choose that variation to construct your tree – the method that provides the lowest BIC is at the top of the chart. Here’s a link to explain more: <http://www.bioinf.man.ac.uk/leaphy/Modelling/Modelling.htm>.

However, as implied by the fact that the BIC values for many of the different parameters are similar, you get nearly the same maximum likelihood tree for this data set no matter how the data are calculated. This may not be true for other data sets!

For this dataset the best model was JTT + G

3. To create a tree, click on ‘Phylogeny’ then click on ‘Construct/Test Maximum Likelihood Tree’ (top menu choice). Based on the model test, choose the parameters that are appropriate for your data set then hit compute. If all goes well, you should see a tree in the ‘Tree Explorer’ window.



For this data set, what do you notice? What inferences might you make about the evolutionary history of this protein? Play around with some of the other tree building programs in addition to maximum likelihood. Are the big clusters always the same? What might this tell you?

Phage proteins form one clade, whereas bacterial homologs form a separate clade.

Trees can be saved as PDF files by Clicking on the Image and saving in the desired format. A description of the process and a picture of the resulting tree(s) should be saved in the Laboratory Notebook.