Project DNA6 Using MEGA to Analyze Aligned Sequence Data

Web Page http://megasoftware.net

"Molecular Evolution and Phylogenetics" (Nei and Kumar, Oxford University Press, 2000).

A very useful textbook on phylogeny.

This program allows the mathematical analysis of aligned sequences. It calculates similarity based on different algorithms. DNA or Protein sequences can be evaluated. However the sequences must be aligned and in a proscribed text format for MEGA to analyze them. CLUSTALW is an alignment program available in BioEdit, on the web and in other DNA alignment software that can be used to align two or more sequences.

NOTE: The sequences must represent homologous regions of the protein or DNA (i.e. be nearly the same length) or the alignment program will introduce gaps to fit them together. This alignment doesn’t provide useful data when analyzed. MEGA contains alignment programs, analysis algorithms and tree-drawing and editing software.

We will analyze the alignment data generated in exercise D5.

Procedure:

1. Start the MEGA program. First we must import the Glycoprotein B sequences and align them.
2. Click File > Open a File/Session and select the Glycoprotein\_B\_raw.mas file
3. When it asks if you how you would like to open the file, choose Align
4. Hopefully your sequence will open in Alignment Explorer window
5. Click Edit>SelectA to select all your sequences for aligning
6. Click Alignment>Align by ClustalW and a window with a lot of choices opens. Use the default selection and click OK
7. A progress window will open showing pairwise alignment and multiple alignment
8. Save the alignment by Clicking Data>Save Session and using Glycoprotein\_B aligned.mas
9. Close the alignment explorer window

*-you’ve already saved your data so check no on the windows that ask if you want to save your data*

1. Now open the sequence in MEGA by clicking File > Open a file/session then selecting your Glycoprotein\_B\_short.mas file
2. Choose Analyze not align. Now you can use the different DNA analysis options to compare DNA’s and plot the results as matrices or trees.

-*Your data is the TA file, the other button closes your data*

1. Lets compare the different DNA’s using a distance algorithm.

Normally, you would need to decide on a model. Computers are now so powerful that the analysis can be run using different models then the best model will be used. For real analysis it is best to run the data using several different models and algorithms and then comparing results. If different models and measurement methods give similar trees then the relationship inferred is said to be “robust”.

-*the various models and algorithms have particular usages. See the help file or the MEGA textbook for more information*

1. Click Models > Find best DNA/protein model
2. Click Phyoleny > Construct/test Neighbor Joining Tree
3. When it asks if it should use the current data set, click yes
4. A selection window will open.

Make sure the Substitutions Non-synonomous

*-this looks at nucleotide changes that don’t change the coding sequence*

Click on Test of Phylogeny and select Bootstrap method then click Compute

-*there are a lot of parameters that can be selected or changed.*

1. The Tree Explorer window opens and displays the phylogenic relationship among the sequences as a an unrooted tree.
2. Save the tree by clicking File > Save current session and naming it NJ tree.mts then clicking Save
3. You can copy the tree as a bitmap or a windows metafile for use as a figure in posters, manuscripts and slides by clicking Image > copy to clipboard or save as an enhanced metafile (EMF) or a TIFF file.

*-the windows metafile format (EMF) is a vector graphic format that can be scaled etc.*

1. Using different models produce more trees that show the relationship that can be inferred among these viruses based on the glycoprotein B gene DNA sequence

Data to be produced:

Computer files Acrobat documents

Glycoprotein\_B aligned.mas (\*.meg) Two DNA Phylogeny Trees