

## Analysis of DNA (Deoxyribose Nucleic Acid) sequences

**YOU SHOULD USE YOUR LAPTOP PROVIDED BY THE UNIVERSITY. IF YOU HAVE NOT BROUGHT IT WITH YOU, YOU SHOULD SHARE WITH SOMEONE WHO HAS.**  
\*\*\*\*\*

If you do not have the BioEdit software installed on your laptop (or you wish to install it on another machine), you should install it as follows:

- a). Using your web browser, go to [www.mbio.ncsv.edu/BioEdit/bioedit.html](http://www.mbio.ncsv.edu/BioEdit/bioedit.html)
- b). Create a folder on your PC called BioEdit
- c). On the BioEdit webpage, choose full installation, and download the file BioEdit.zip into your new BioEdit folder.
- d). Unzip the zipfile, so that the unzipped files are in the new BioEdit folder.
- e). Run (by double-clicking) setup
- f). follow on-screen prompts such as 'next' when prompted
- g). Choose TYPICAL when asked
- h). Press FINISH when prompted and LAUNCH PROGRAM (Bioedit will now open)

### **1). Load the test sequences into BioEdit.**

Download the dnadata.phy file from the course website into the new BioEdit folder on your PC. (Beware: if Windows adds '.txt' as an extension, rename the file to dnadata.phy)

Import the DNA base sequences provided (dnadata.phy) into BioEdit. (File/Open/Phylip files)  
Each line is a DNA sequence from a different animal, the names of 2 of which (for the purposes of this exercise) are unknown.

### **2). Identify Unknown Sequences using a BLAST (Basic Local Alignment Search Tool) search**

There will be a frame to the left containing names, and a large window to the right containing nucleotide sequences. Select **about 50 characters** of an unknown nucleotide sequence by using click and drag to highlight.

Press CONTROL-C or use the menu to Copy.

Try to identify that sequence by choosing Accessory Applications/BLAST/MWWW, then Nucleotide-Nucleotide BLAST (blastn) from the menu. Before activating a search, you will have to PASTE the sequence from the clipboard into the search window. Then you can activate a SEARCH.

You will have to *wait a few minutes* for a reply from the database (or if you're luck, a few seconds!). You should (after waiting!) press FORMAT! to receive your search results from the server.

Remember, this is a scientific database, so the names of the identified species will be returned in Latin!!

(You may recall what I said in the lectures about genus and species names in a Latin binomial)

Repeat this exercise for the other unknown sequence.

***DON'T FORGET TO CLEAR THE PREVIOUS SEARCH FROM THE WINDOW FIRST!!!!***

Note if the above doesn't work, try doing a BLAST search from NCBI

<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>

### **3). Performing a Cladistic Analysis from the test sequences**

Now that you have identified the unknown sequences, it is a useful exercise to conduct a cladistic analysis, i.e. to attempt to reconstruct an evolutionary tree from the DNA sequences.

Select by highlighting (click and drag) all the sequence names in the left window.

Select from the menu AccessoryApplication\DNAdist-->neighbour phylogenetic tree, then run the application. This will use the DNAdist program from PHYLIP to produce a cladogram by a *neighbour-based* method. TREEVIEW should automatically be opened to view the resultant tree. If TREEVIEW is not automatically opened, you might need to run TREEVIEW first, before generating the tree.

It is possible to produce a cladogram using a *parsimony* method using the program DNAPARS.exe in PHYLIP (from a previous practical session). It is necessary to rename the DNA datafile used earlier to 'infile', place it in the H:\Phylip directory, then run 'dnapars.exe', accepting all the default settings.

Note: these trees are unrooted. That means that they are best viewed as a radial tree (e.g. in TREEVIEW). When other trees are drawn, the apparent rooting can be misleading. Using TREEVIEW choose what you think is the most 'primitive and ancient' member to be the outgroup and redisplay the tree. Does this produce a more sensible tree?

(If the resultant tree is renamed with a .tre extension, then you can open it in TREEVIEW in the normal way. It is a very useful exercise to reroot the tree with a suitable outgroup. e.g. you can choose the fish as an outgroup.)

#### References:

DNA data from <http://newfish.mbl.edu/Course/Software/FileFormats/index.html>

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids. Symp. Ser.* 41:95-98

BLAST search: <http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>