

## Making Phylogenetic Trees using the Computer

You will now learn how to make a phylogenetic tree using bioinformatics tools that are freely available on the web. You will use a program called ClustalW to make this tree. ClustalW takes an input of different sequences, aligns the sequences to each other, and calculates how similar each sequence is to each other sequence. The program then takes those numbers and turns them into a phylogenetic tree. The distance along the branches between each pair of organisms shows how related those two organisms are to each other. A shorter distance means a closer degree of relatedness.

### **Instructions:**

1. Let's start by making some predictions from your last set of data. Look back at the scores you obtained from your BLAST search for human and boar. Think about the scores these two organisms got when you used an opossum DNA sequence to do your search. Draw a phylogenetic tree below that contains opossum, cow, horse, macaque, & platypus. (Do this using your previous tree made from DNA sequences.) Then, add in human and boar where you predict they would fit in the tree.

2. Now let's make a tree with these organisms to see if your prediction is correct. To begin, open the file called "seven mammals." This Word document contains the sequence of the LDL receptor gene from opossum (i.e. the sequence you looked up in the previous section of this exercise). It also contains the LDL receptor gene sequences from the six other species matches that you found in your BLAST search and recorded in your chart in the previous section of this exercise.

3. Edit this Word document so that your sequences are in the correct format for the ClustalW program. The correct format looks like this example:

>StaphylococcusCondementi

TGCAAGTCGAGCGAACAGATAAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGG  
TGAGTAACACGTGGGTAACCTACC

>StaphylococcusXylosus

GTCACATGGAGCGAACCGTTAAGAGGCCTACTCCTTTGACGTTACGGCGGACAGGT  
GAGTCATACGTGGGTAACCTATG

To put the sequences in this format, you must follow these rules:

- There must be no spaces within your sequence, before it begins, or after it ends.
- There must be no blank spaces or lines between the names and the sequences.
- The name of each species must be in the format shown above, where there is a right-facing carat followed by the name of the organism, with no spaces in between any of the characters.

**4.** Now go to the ClustalW website:

<http://www.ebi.ac.uk/clustalw/>

**5.** Look for the empty box under the line that reads:

“Enter or Paste a set of Sequences in any supported format.”

**6.** Paste in the sequences of the seven mammals, entered in the correct format shown above, where there are no blank spaces or lines between any of the sequences. Make sure there is also no blank line at the end of your list after you paste it into this box.

**7.** Once your sequences are pasted into the box, hit the button that says “Run.” Wait as the job completes; the page will automatically reload when it is finished. This may take 2 or 3 minutes.

**8.** Once the page is reloaded, scroll all of the way down to the bottom of the page. Click on the button that says “Show as Phylogram Tree.” When the page reloads, scroll all the way down to the bottom, where the picture of your tree will be.

**9.** To create a file of your phylogenetic tree on your desktop that you can save or print, first press “Apple-shift-4”. The cursor will now look like an X with a circle in the middle. Place the cursor at the upper left-hand corner of the part of the webpage where your tree is. While holding down the mouse button, drag cursor so that whole tree is highlighted in gray. Let go of the mouse button such that the gray box disappears. Look on the desktop of your computer, and there should be a file called “Picture 1”. Open that file, and it should be a picture of your tree.

**10.** Now that you have your tree, analyze it in order to answer the following questions:

**a)** You have now made two trees containing opossum, cow, horse, macaque, & platypus. You made one tree by counting letters of DNA by hand, and the other by the ClustalW program. Are these two trees identical to each other? If yes, explain why you think they turned out identically. If no, explain why you think they didn't turn out identically.

b) Look back at your prediction from question #1 above. Did your predictions for where human and boar would fit into the tree match with the tree made by the ClustalW program?

11. Once you are done with your tree, scroll back up on the web page from the phylogram until you get to a section with the heading “**Alignment**” that has a lot of rows that look like this:

```

BacillusMojavensis      AGAACAAATAGGAAA-AATTGAATATTTAGTACATAGTATTCTCAGTTTAC 197
BacillusVallismortis    AAAACAAAAGAAGGGAATTGAATATTTAGTACATCATATTTTGTAGTATCA 56
BacillusAtrophaeus      AAAA-AAGAGGGGGGAATTGAATATTTAGTACATAGTGTTCAGTT--A 193
AnoxybacillusFlavithermus -----TCAACTAAAATTATC 15
GeobacillusThermodenitrificans -----CTAGCTATAGTTGTC 15
                                     * *

```

Here you can see the sequences from all seven mammals aligned to each other.

\* A star means that the identical letter is found at that position in all seven organisms.

- A dash means that letter is missing in that organism.

The numbers show that letter’s position in the full DNA sequence that you entered.

a) Look through this alignment and answer the following question:

The first time one sees a stretch of letters that are identical between all seven species is the sequence: GGCTG. Write out a DNA sequence from LDL receptor gene in which 8 letters of DNA in a row are identical in all of these seven mammals.

b) The opossum DNA sequence we used in this exercise is 749 letters long. If you look across this alignment at these 749 letters, 60% of them are identical in all of these 7 mammals: opossum, cow, horse, macaque, platypus, boar, & human. Is this surprising to you? Why or why not?