

Biol5705
Module: Gene Sequence Analysis
Assignment 4

Download the the list of protein sequences in fasta format:
<http://morganlangille.com/teaching/biol5705/assignment4.fasta>

Using MEGA

Download and install MEGA: <http://www.megasoftware.net/>

Load the sequences into MEGA: (Align->Edit/Build Alignment->Retrieve Sequences From File)

Create a multiple sequence alignment using Muscle with default parameters (Alignment->Muscle Alignment)

Get the data ready to build trees (Data->Phylogenetic Analysis)

Create a Maximum Parsimony tree using defaults (Phylogeny)

Show the character ancestral states (Ancestral->show all)

Find the box that allows you to enter the character position that you are interested in (also contains an up and down arrow to iterate through positions).

- 1) What is the first position in the MSA that contains different characters across the sequences (e.g not conserved). What are the different characters for this site? What is the minimum number of ancestral character changes needed to explain this tree?

Create a Maximum Likelihood tree in MEGA with 100 bootstraps

- 2) Is the topology of the parsimony tree and the ML tree the same?
- 3) Describe the clades with the highest and lowest bootstrap support.
- 4) What two taxa are the most closely related (shortest branch length between them)?
- 5) Export the tree in Newick format, open the file, and copy/paste the tree into your answers.

Using Phylogeny.fr

Visit <http://www.phylogeny.fr/>

Create your own pipeline (Phylogeny->"A la Carte") using T-Coffee as your multiple alignment program, using Gblocks to automatically edit your MSA, PhyML to construct the tree, and any program for visualization.

Load the assignment4.fasta file into the pipeline, browse through the options and run the pipeline. Note that Phylogeny section there is an option to choose bootstrap values. You can either choose to bootstrap your tree OR use the faster default "SH-like" Approximate Likelihood Ratio Test.

Browse the "Curation" tab to view what GBlocks edited in your MSA.

- 6) How many characters/positions did GBlocks select? Where in the MSA are these characters at (give start and end position).
- 7) Get the tree in newick format and copy newick string into your answers.

Install and use any of the tree viewers mentioned in my slides to view either the ML tree that you created with MEGA or phylogeny.fr. Display an unrooted tree in your tree viewer and take a screen capture showing your tree and the program you are using.

- 8) Attach or embed your screen capture.