Lab 3: Build a Tree

Background

For this lab, we will estimate a multiple sequence alignment for the given data file 'sequence.fasta' that can be found on Blackboard, then estimate and interpret your phylogenetic estimates.

Learning Objectives

1. Gain further experience in multiple sequence alignment approaches.
2. Develop an understanding of the parameters associated with multiple sequence alignment.
3. Gain experience estimating phylogenies and understanding the underlying approaches.
4. Develop an understanding of models of evolution as used in phylogeny estimation.
5. Gain experience in interpreting phylogenetic output, including identification of monophyletic groups and interpreting results.

Tasks

Part 0 - Regular Expressions

Use a REGEX to rename your sequence data to something short and informative: From: AY156743.1 HIV-1 clone P.ENV f . . . .. To: AY156743_P etc.

Part 1 - Multiple Sequence Alignment

Using the data set provided, estimate a multiple sequence alignment. Be sure to include details that answer the following questions in your methods section:

1. What software did you use to construct your multiple sequence alignment?

2. What parameters could you adjust in your software for nucleotides?

3. What parameters could you adjust for amino acids?

4. What values did you use for these parameters and **why**?

Part 2 - Phylogeny Estimation

1. Take your nucleotide alignment from Part 0 and determine the best-fit model of evolution for your data. What is the model? What are the parameters of the model? How does this model compare to that used in the alignment assignment?
2. Using your alignment and best-fit model, build a phylogenetic tree. Show an image of your tree? Are you confident in your tree? Why or why not?
3. Write your lab report in the typical style (below) with the introduction including the literature associated with your alignment and phylogeny estimation, details on your methodology for alignment and phylogeny estimation, and a discussion of your phylogenetic results. Be sure in your results and discussion to address the questions above. **Hint** Remember your GenBank skills to link the data to PubMed records and discuss the phylogeny in the context of where the data came from! Brownie points if you can bring the original context of the data into the discussion or overall paper. Remember to cite the software that you use!

Guidelines

Begin the paper with an original title, followed by your name, the course name, and the date. All write-ups should be single-spaced and in 12 pt font. Your paper should have all of the following sections.

Sections

0. **Title:** A concise, but informative quick summary or description of the report.
1. **Introduction:** Introduce the general problem or issue you are addressing. Provide the background context for the reader. You should prepare the reader for the following sections of the report as well as convince the reader that you are tackling a problem worth addressing in your work.
2. **Materials and Methods:** Describe the methods used to obtained the data, analyze the data, and test hypotheses associated with the data. You should explicitly mention and cite any bioinformatic tools used. You may wish to provide the parameters (settings) of any software that you used, if relevant or important.
3. **Results:** Provide and describe the results of the data analysis and hypothesis testing. This may include presenting charts, tables, pictures, and other figures when appropriate for effectively communicating your results. This section is primarily for *only* presenting the results. The following section is for discussing them.
4. **Discussion:** Interpret your results and discuss their implications with the reader. This section should include a synthesis of ideas.

5. **Conclusion:** Wrap up and summarize the paper. You may also wish to discuss the future direction of similar research here, if relevant. The conclusion does not need to be its own section—you may choose to put a concluding section at the end of discussion. Your paper, however, should include some piece of writing the neatly wraps up your report.

6. **References:** List the relevant literature you have read and used to support your arguments/analyze your data. The literature cited should be in the format of the journal *Bioinformatics*. If you are unfamiliar with this citation style, look up the guidelines for it in the journal's "Instructions to Authors" guide. Failure to cite work used may be considered plagiarism, if the offense is serious.