

STA 695, CS 685, PPA 784 Spring 2012 Homework 2

Part I.

Analysis 1:

Download file: [lolC introns trm.fasta](#)

Open the file in a text editor. Select all and copy.

In a browser, go to <http://www.phylogeny.fr/>. Click on [Phylogeny analysis](#), and select “[One Click](#)”.

Go to the bottom of the page, and click [Create Workflow](#).

Paste text (from [lolC introns trm.fasta](#)) into the text box.

Scroll to the bottom, and click [Submit](#).

After several seconds, a page will appear with [Tree Rendering results](#).

You will see several options and buttons on this page. Among those options, try those under [Tree style](#). Also click on the [Reroot \(outgroup\)](#) button, then go to the tree and double-click on some branches. Describe what you observed. Feel free to try other options on this page.

Click the [Alignment](#) tab.

Click (under [Outputs](#)), [Alignment in FASTA format](#).

How does "alignment in FASTA format" differ from the input FASTA?

Select all and copy. **Note:** This is output from Analysis 1, which will now serve as input for Analysis 2.

Reformatting data:

In a browser, go to <http://www.ebi.ac.uk/cgi-bin/readseq.cgi>. Paste into the large window near the top of the page. Under “Output sequence format:” choose [Phylip|Phylip4](#). Below that, choose [View in Browser](#), then click the [Submit](#) button above that.

Select all and copy. **Note:** This is output that will now serve as input for Analysis 2.

Analysis 2:

Open <http://www.trex.uqam.ca/>. Click [DNApars](#), then Paste the text into the large window near top.

Click the [Compute](#) button. After a moment, you will have a window with links to [Infile](#), [Outfile](#), and [Outtree](#). Describe what you get when you click those links.

What was the purpose of Reformatting data before Analysis 2? Is there an alternative to that reformatting step?

Additional Question:

Note that on each of the pages in the exercises above (Analysis 1, Reformatting data, and Analysis 2) there will typically be several options. In addition to the Tree style and Reroot options I mentioned above, try two other options. Describe what happens and the reason, if you think you know why.

Part II.

This part of the homework pertains to trees, Newick format and phylogeny inference based on parsimony. Three exercises are listed after a sample tree produced with software available in <http://www.phylogeny.fr>.

Introduction:

A few sample drawings for the (A:0.1,B:0.2,(C:0.3,D:0.4)E:0.5)F; Newick formatted tree taken from http://en.wikipedia.org/wiki/Newick_format.

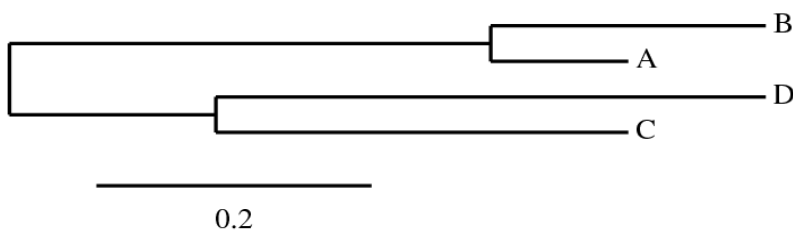
1) From http://www.phylogeny.fr/version2_cgi/data_converter.cgi with output format set to “Nexus”

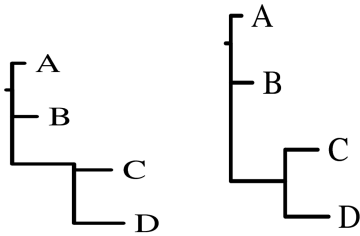
```
#NEXUS
```

```
BEGIN TAXA;  
  DIMENSIONS ntax=4;  
  TAXLABELS A B C D;  
END;
```

```
BEGIN TREES;  
  TREE tree_1 = (A:0.1,B:0.2,(C:0.3,D:0.4)E:0.5)F;  
END;
```

2) http://www.phylogeny.fr/version2_cgi/one_task.cgi?task_type=treedyn

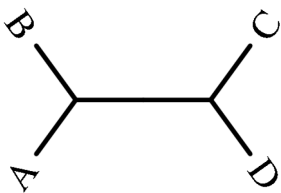


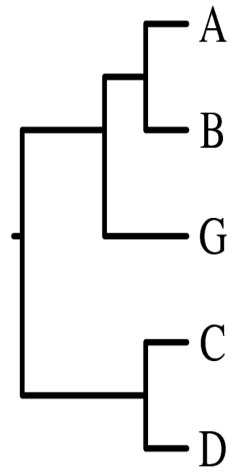
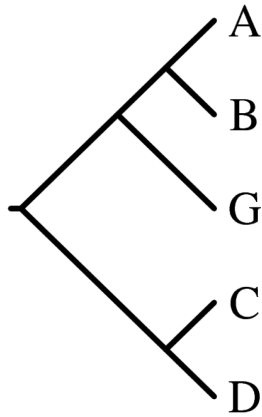
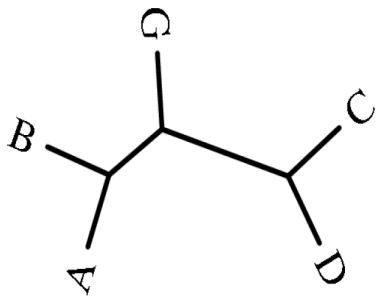
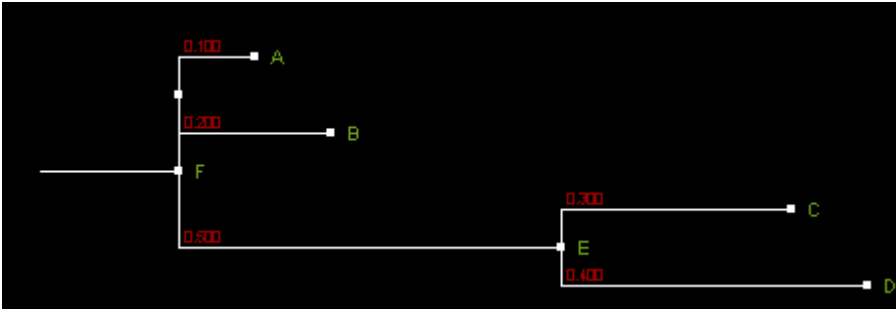
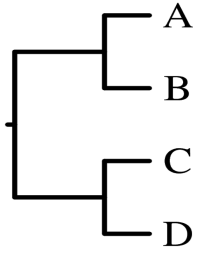


3) http://www.phylogeny.fr/version2_cgi/one_task.cgi?task_type=drawgram (option Phenogram and Cladogram and Phenogram with Advanced Settings V-Shaped)

Questions:

Question II.1 What is the Newick format for the following six trees? Provide the answer and test it with tree drawing software:





Verify all of your answers using drawing tools available at <http://www.phylogeny.fr> and provide the resulting drawings. To differentiate your drawings from the above pictures, use the following taxa names association table (that is, when submitting a Newick file to a drawing program, use the following long names instead of A, B, C, D, E, F, G):

- A --Achlis
- B – Bagiennik
- C – Cadejo
- D – Dilon
- E – Epimeliad
- F – Fuath
- G – Gandaberunda

Question II.2 What is the number of labeled trees

- rooted binary trees with 10 leaves (taxa)? With 11 leaves (taxa)?
- non-rooted trees with each internal node of degree 3 (i.e. strictly bifurcating, as discussed in class), with 10 leaves (taxa)? With 11 leaves (taxa)?
- If it takes 1 sec to draw by hand one rooted labeled tree, how long would it take to draw all of them for 10 leaves? Provide answer using common time units such as minutes, hours, days, etc.

Hint: use the formulas derived in class.

Question II.3 Parsimony:

In this exercise, you are asked to compute parsimony tree scores using the brute-force method (see class notes). Specifically, for the alignment of five sequences given below, provide tree scores for the following state assignments to the internal nodes: A, A, A and T, C, A. Use the cost matrix that assigns five times as much weight to transversions as to transitions. Show all the trees for the considered assignments and identify a tree with the best score.

Analyze the bolded positions in the following alignment:

```
...ACATAAG GATCATG ...
...ACATAAG CATCATG ...
...ACATAAG AATCATG ...
...ACATAAG CATCATG ...
...ACATAAG TATCATG ...
```