

Assignment 5 - Fun with Maximum Likelihood

Part I. Likelihood methods using paper and pencil ☹.

- 1) Complete exercise 4(a and b) on page 99 of the textbook.
- 2) Calculate the likelihood for the following tree under the JC69 model

Taxon:	Sequences:	Topology:
A	AG	<pre>graph LR; A --- 0.3 Node1; Node1 --- 0.1 B; Node1 --- 0.1 C;</pre>
B	AC	
C	AG	

Remember that to calculate the likelihood of a site, you have to consider all possible reconstructions for internal nodes and that the edge lengths are given in terms of expected number of substitutions v , where $v = 3\alpha t$ (for the JC model)

Part II. Likelihood methods using RAxML-NG and the dataset from Lab 4.

- 3) Find the NJ (use FastME) and the ML trees using the GTR+G model of sequence evolution. Are the two trees the same? If not, what is the difference in their log-likelihood scores?

Extra point: Print these trees “face to face” aka “tanglegram” using Dendroscope. Submit with your assignment.

- 4) Partition the dataset either by gene or by codon position. What partitioning scheme results in a higher likelihood (use your ML tree to evaluate). Use the best partitioning scheme to infer the ML tree. Does it have a different topology from your original ML tree?

- 5) Write a scripts for the Slurm Workload manager to do a bootstrap analyses (1000 replicates) for the dataset above. Request 1 node, 16 cores. Use `--threads 16` option to utilize all 16 cores for the analysis. How much time would you need for the analysis? Make an estimate based on your lab experience. How much time did it actually take? Print your ML tree with bootstrap support mapped on it and submit it with the assignment.