

Homework no. 2
Due Friday, May 5

- Using a nucleotide or protein sequence data set (which can be the same as the one used for the last homework), infer a Fitch-Margoliash least squares tree from distances computed from a distance method that you consider appropriate (you should defend its appropriateness).
- As before, it should have at least 20 sequences (or species) and of course these should be aligned. For molecular data sets there should be at least 200 sites. If you do not have your own data sets here are some sources of aligned molecular data sets:
<http://systbio.org/?q=node/35> (see the links at left side)
<http://www.sanger.ac.uk/Software/Pfam/>
You may have to hand-edit some of these into the proper format.
- Also run a neighbor-joining (NJ) analysis on the same distances. How do the speeds compare?
- Do a bootstrap analysis with at least 100 data sets, using NJ. (With distance matrices one bootstraps on the original data set, then computes a distance matrix and then a tree on that for each replicate. The bootstrap analysis includes making a majority-rule consensus tree.
- Report the results to me in a short (5 pages or so) report. Show some results if needed.
- Comment on the program you used, how well it worked and how easy it was to use.

e-mail me ([joe \(at\) gs.washington.edu](mailto:joe@gs.washington.edu)) with a report in PDF or MS Word form on the results, or give them to me in a printed form in class or in my hallway mailbox on 3rd floor (east end) of K wing, by the end of the day.