

Homework no. 2
Due Monday, May 5

- Using a nucleotide or protein sequence data set (which can be the same as the one used for the last homework but a new one would be nice too), 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 (more would be good). If you do not have your own data sets here are some great sources of aligned molecular data sets (it will take some struggle):

<http://www.pfam.sanger.ac.uk> (lots of good protein sequence alignments)

<http://www.arb-silva.de> (RNA small and large subunit sequences)

<http://systbio.org/?q=node/35> (see the links at left side)

If you get a sequence alignment from any of these in FASTA format you can use this web server:

<http://www-bimas.cit.nih.gov/molbio/readseq/>

to convert it into PHYLIP format and some others too. There are many other databases you can use – if you find one that is a particularly good source, let me know.

- 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 (3-5 pages or so) report. Show some results if needed. Report:
 1. Which distance method and its model of evolution was used, and why.
 2. The results of the distance analysis.
 3. How the results of Least Squares and NJ compared.
 4. The results of the bootstrap.
 5. How well the programs functioned and how easy or hard they were to use.

e-mail me ([joe \(at\) gs.washington.edu](mailto:joe@gs.washington.edu)) with a report in PDF or MS Word format on the results, by the end of the day.