

Homework no. 1  
Due Monday, April 21

- Get a computer program that can do parsimony on nucleotide or protein sequences, or on discrete characters. It should also be able to do branch-and-bound search as well as heuristic search. For a list of programs, many of them free, see <http://evolution.gs.washington.edu/phylip/software.html> or type “Phylogeny Programs” into Google. Some possibilities include PAUP\*, PHYLIP, MEGA, TNT. Some of the other excellent general-purpose programs there may not have both heuristic search and branch-and-bound.
- Get a data set, either molecular sequences or discrete characters. 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://pfam.sanger.ac.uk/>  
You may have to hand-edit some of these into the proper format.
- Run heuristic searches for the most parsimonious trees on data sets of size 10, 15, 20, ... species (or sequences). Explain exactly what choice of strategy you used. How much time do these take? (For some data file formats you can do different size data sets just by deleting whole species).
- Run branch and bound searches on some of those same data set sizes (start with the smallest ones and stop when it takes too long). How much time do these take? Does the run become impossibly slow? How do the trees found compare in number and parsimony score?
- Report the results to me in a short (3-5 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.

You should e-mail me (joe (at) gs.washington.edu) with a report on the results.