

Assignment II – 12 Oct 2004

You'll need to use PAUP\* for the following problems. PAUP is installed on iNquiry (<http://inquiry.egg.isu.edu>). Data required for the following problems may be obtained from the course website but I strongly suggest using your own data. Each problem requires the generation of some data – you will also provide a brief interpretation (1-2 paragraphs) of the final results. **The assignment is due on Tuesday, October 19.**

1. Given the CO1 genes of vertebrates, consisting of 10 taxa and 705bp (or your favorite data set of about the same size or larger), identify the best, unbiased estimate of substitution rates for the given data set:
  - a) If you are using your own data, make certain the data are appropriately aligned in Nexus format.
  - b) Insert the "ModelTest" PAUP command block in your Nexus file and execute it on iNquiry. (The ModelTest block may be obtained from the course web site and from the ModelTest program site on iNquiry – click on the program name for links to the PAUP block, the manual, and the ModelTest homepage.)
  - c) From the results page of your PAUP run there should be a file called "model.scores" – under this, click on the link piping the file to the ModelTest program. Execute ModelTest (click "run").
  - d) For both likelihood ratio test (LRT) and Akaike Information Criterion (AIC), list the model identified and (if applicable) the shape of the gamma parameter and the proportion of invariant sites. Choose the simpler model; how do you know which is simpler?
2. Using the Jukes-Cantor distance and your unbiased estimate of substitution rates, identify the following trees. Provide the tree and statistics related to the tree. Use "set criterion=distance; dset distance=JC;" for Jukes-Cantor (and the appropriate abbreviation for your unbiased estimator when setting the substitution estimator to that distance).
  - a) The UPGMA tree; hints:
    - i. Use "UPGMA brlens=yes;" to order this type of tree and have it return length estimates for each branch.
    - ii. Use "DeroorTrees;" after the UPGMA command to remove the UPGMA root, allowing distance scores to be calculated.
    - iii. Use "Dscores;" to report tree distance using the current objective function (default: ME).
  - b) The NJ tree; hints:
    - i. Use "NJ brlens=yes;" to order this type of tree and have it return length estimates for each branch.
    - ii. Use "Dscores;" to report tree distance using the current objective function (default: ME).
  - c) The LS tree(s) via branch & bound search; hints:
    - i. Use "dset objective=lsfit negbrlen=allow;" to set the objective function of the search to LS and allow negative branch lengths (req'd for B&B).
    - ii. Use "set increase=auto;" to allow it to continue searching even when the maximum number of trees retained in memory has been reached.
  - d) The ME tree(s) via heuristic; hints:
    - i. Use "dset objective=me negbrlen=allow;" to set the objective function to ME.
    - ii. Use "set increase=auto;" to allow it to continue searching even when the maximum number of trees retained in memory has been reached.
3. Interpret the results in 1-2 paragraphs:
  - a) Compare/contrast the NJ, UPGMA, LS, and ME trees with respect to topology and branch length estimates.
  - b) Compare/contrast the JC distance with the unbiased distance estimate with respect to topology and branch length estimates.