Phylogenetic Analysis of Molecular Data (Botany 563)
Computer Lab 07: Molecular Clock Analysis

Learning objectives:
- Familiarize with Molecular Clock analysis in a ML framework using PAUP*

Data files: **hibbett_mt.nex**; **hibbett_tree1.tre**

Task 1. Get familiar with some of the models of molecular evolution available for maximum likelihood analysis.

a) Open and execute the dataset **hibbett_mt.nex**

b) Load the tree from **hibbett_tree.tre** into memory.

c) Optimize parameters for settings corresponding to the HKY+gamma model of evolution (ti-tv = estimate; state frequencies = empirical; Across sites rates = Gamma, shape = estimate). Calculate the likelihood of the tree, and save it to a tree file <u>with likelihood branch lengths included</u> (click on the "option" button when saving the tree).

> **-ln L$_{non\text{-}clock}$=**

d) Root the tree with an appropriate outgroup (*Dacrymyces chrysospermus*). Look at a phylogram and save it as a pdf.

e) Determine the likelihood score of the same tree enforcing a molecular clock: Likelihood settings>Other (Thorne parametrization is fine).

> **-ln L$_{clock}$=**

Is the Likelihood score higher or lower than that generated in step c?

Which is the "simpler" model?

f) Conduct a likelihood ratio test to determine if the difference in likelihoods is significant.
> 2x lnL difference=
> degrees of freedom=
> p-value=
> Does this result mean you should use the clock or nobn-clock model? Why?

g) Look at a phylogram and compare it with the one you generated in step c). How are they different?

Task 2. See the effect of rooting under a molecular clock.

    a) Re-root the tree using a different rooting option or outgroup.

           Outgroup used =

    b) Calculate the likelihood score enforcing a molecular clock.

           -ln L =

    c) Which rooting is favored by the clock model?

       Would it be appropriate to do a likelihood ratio test to assess this difference? How many

       degrees of freedom?

Task 3. See the effect of parameter estimation on the likelihood ratio test of clock vs. non-clock
      models.

    a) Return to your optimally rooted tree.
    b) Recalculate the parameters, estimating state frequencies, under the clock model and
       obtain the tree's likelihood score

       -ln L$_{clock}$=

    c) Fix the parameters (using "previous"), then deroot the tree and calculate the score of the
       tree <u>without</u> a molecular clock

       -ln L$_{non-clock}$=

    d) Is the likelihood ratio lower or higher than that of Task 1? Why do you think this is the
       case? Would it be more likely to reject a clock model using estimated parameters
       enforcing a clock model or using fixed parameters as estimated under a non-clock model?

Task 4. Conduct a ML search enforcing a molecular clock. If time allows, conduct a ML search
under a non-clock model. For each search report the following:

    a) your search parameters

    b) likelihood score of best tree found: **-ln L$_{clock}$=**

How do the likelihood scores compare to the ones obtained in Task 1? Would this be expected?