

Lab 4 – MrBayes

1. As in previous labs, please log in to any of
mordor.stat.osu.edu
rohan.stat.osu.edu
gondor.stat.osu.edu
rivendell.stat.opsu.edu or
shire.stat.osu.edu
2. MrBayes is a popular Bayesian phylogenetics program written by Fred Ronquist, John Huelsenbeck, Paul van der Mark and Bret Larget (see <http:// mrbayes.scs.fsu.edu/wiki>) the program is run from the command line on our machines by entering “mb” at the Unix command. To learn what commands are available type “help” and to learn about subcommands type “help command.”
3. filename.xxx output files
.mcmc characteristics of the chain (acceptance rates, swaps etc...)
.con consensus tree
.t sampled trees (one file per run)
.p sampled parameter values (one file per run)
.parts (partitions and their estimated probabilities)
.trprobs estimated posterior probabilities for trees found
4. You can set up a complete MrBayes analysis in a MrBayes block at the end of a nexus file. For example, the file yeast.nex on the course website provides a complete annotated analyses illustrating many popular options in a “partitioned” analysis of four loci across six yeast species. Note that many commands are currently “commented out” in the file so you will need to remove brackets around commands you want to execute before running.
5. In running MrBayes interactively, it is important to keep track of the current settings you have for important parameters of your model and your Markov chain. Use “showmodel” and “help mcmc” often.
6. Try running the primates data on your own for 100,000 generations using the first 10,000 generations as a burn-in and using the model suggested by ModelTest in lab 2. What is the estimated posterior probability of the most likely tree? Was the burn-in period long enough? Did the Markov chain appear to converge overall? How might you estimate the variability in the percentages given for the splits by the sumt command?
7. Try running the hpv data on your own with enough generations and runs to feel comfortable about convergence (this will go much slower than the primates data and will need to be done outside of class). From lab 3 we know that there are at least two islands of high probability. How does that affect what you need to look at in checking for convergence?