Phylogenetic Analysis of Molecular Data (Botany 563)

Computer Lab 10: Bayesian Analysis II

For useful information refer to the MrBayes manual, which you can find at http://mrbayes.csit.fsu.edu/manual.php

Learning objective:
Exploring the posterior to assess support for alternative hypotheses in a Bayesian framework.

Dataset: pereskiabay4.nex

A few weeks ago, you rejected monophyly of *Pereskia* based on a Templeton test. We will go back to the same dataset, now in a Bayesian framework, to investigate if there is any support for monophyly.

<u>**Task 1a.**</u> Run a Bayesian MCMC analysis of this data.

To do this you only need to analyze the dataset in the file pereskiabay4.nex.
Look at each command line at the end of the bayes block and make sure you know what it is telling MrBayes to do:

```
BEGIN mrbayes;
log start filename=pereskiabay4.log append;
CHARSET phyC = 5018 – 6150 ;
CHARSET chlormito = 1 – 5017 ;
partition genome = 2:chlormito, phyC;
set partition = genome;
lset applyto=(1) nst=2;
lset applyto=(2) nst=6 rates=invgamma;
mcmcp ngen=10000 temp=0.2 nruns=2 nchains=4 samplefreq=10 savebrlens=yes;

showmodel;
END;
```

To run the analysis, TYPE: **mcmc**.

Q: How many samples (lines) should there be in the .p file for this analysis? _____
Explain

Q: Is there good mixing of the chains? What are the pertinent values that tell you this?

If your answer is negative for the previous question, how would you improve the mixing?

1b) You can try improving the mixing on your own, if so, briefly describe what you did and whether it worked. Alternatively, skip over this section and use the files in the folder pereskia7, which were generated with good mixing, for the rest of your analyses ☺.

Once you have good mixing:

<u>**Task 2**</u>: Determine an appropriate burnin, and summarize your results (sump & sumt commands).

Report the burnin used: _____ How did you decide on this value?

How many trees (topologies) are there in the sample from the posterior distribution? _____ and in the 99% credibility interval? _____

<u>**Task 3**</u>. Assessing support for alternative hypotheses.

As a result of our Bayesian MCMC analysis, we have an approximation of the posterior. We are interested in knowing what proportion of the posterior supports one hypothesis (or its alternative). To input our different hypotheses you can make use of constraints.

We will test the following hypotheses:
1. monophyly of *Pereskia*
2. monophyly of the Core Cacti
3. monophyly of the south American Pereskia (South America includes the Andes)

The following translation table might be useful.

```
begin trees;
translate
1 Anacampseros_telephiastrum,
2 Austrocylindropuntia_subulata,
3 Blossfeldia_liluputana,
4 Brasilopuntia_brasiliensis,
5 Calymmanthium_substerile,
6 Ceraria_fruticulosa,
7 Cereus,
8 Echinocactus_platyacanthus,
9 Grahamia_bracteata,
10 Grahamia_coahuilensis,
11 Maihuenia_patagonica,
12 Maihuenia_poeppigii,
13 Opuntia_dilleni,
14 Pereskia_aculeata,
15 Pereskia_aureiflora,
16 Pereskia_bahiensis,
17 Pereskia_bleo,
18 Pereskia_diaz_romeroana,
19 Pereskia_grandifolia_var._grandifolia,
20 Pereskia_guamacho,
21 Pereskia_horrida_var._horrida,
22 Pereskia_lychnidiflora,
23 Pereskia_marcanoi,
24 Pereskia_nemorosa,
25 Pereskia_portulacifolia,
26 Pereskia_quisqueyana,
27 Pereskia_sacharosa,
28 Pereskia_stenantha,
29 Pereskia_weberiana,
30 Pereskia_zinniiflora,
31 Pereskiopsis_A,
32 Pereskiopsis_porteri,
33 Portulaca_oleracea,
34 Portulacaria_afra,
35 Quiabentia_verticillata,
36 Quiabentia_zehntneri,
37 Talinum_paniculatum,
38 Tephrocactus_articulatus;
```

Write the constraints for the three hypotheses above using taxa numbers:
1.
2.
3.

Execute the Pereskia dataset in PAUP.
Use "Trees>Get trees from file" to load to memory the trees sampled in your MrBayes analysis (which reside in two different **.t** files); remember to discard the samples corresponding to the burnin (hint: Trees>Options>Trees; input tree numbers). Say "NO" to unroot trees. Always check that you have the number of trees you are expecting ☺.
You can load constraints simply by typing them to the command line in PAUP* (or you could write them into the data file in a PAUP block before executing) .

To see the proportion of the total trees that satisfy a given constraint use: Trees>Filter trees>New filter. You should select the box that reads "compatible with constraints", and, of course, select the appropriate constraint. A useful option (same window) is the one that reads "retain trees that do NOT match the above criteria".

How many total trees did you load to memory? _____
How many trees satisfied each of the three hypotheses outlined above?
1.
2.
3.
Explore the other options under "Trees>Filter trees".

What is the posterior probability of each of the three hypotheses? What could be a problem with taking these values as good estimates of the "true" PP?