MrBayes practical (with primates!)

Thomas Guillerme
t.guillerme@imperial.ac.uk

February 4, 2016

Contents

1 Installing MrBayes

You can download **MrBayes** (Ronquist et al., 2012) for any platform (here). A simple tutorial is available (here) along with a manual (here).

2 The data

For this practical we are going two files, one matrix containing phylogenetic information (`NEXUS` format; `.nex`) from Hayasaka et al. (1988) ad Stevens et al. (2013) and a second file being a MrBayes executable containing the phylogenetic information and MrBayes instructions:

- `TotalEvidence_Primates.nex`: alignment containing 898 molecular characters and 191 morphological ones for 20 taxa.
- `TotalEvidence_Primates_exe.nex`: the same alignment but with more detailed MrBayes instructions.

3 Starting MrBayes

To start MrBayes, simply double click on the executable or, for terminal users, enter `mb` in the terminal. Then start running a log file to save all your analysis by using the `log` command:

```
log start filename=MrBayes_practical.txt
```

4 Running the analysis

4.1 Reading the data

We can read the data into MrBayes by using the command `execute` (make sure you're in the right working directory!).

```
execute TotalEvidence_Primates.nex
```

We can now manually manipulate some characteristic of our data file, for example by setting an outgroup:

```
outgroup Lemur_catta
```

Or by creating the group that contains only the fossil taxa (here we define the taxa number 13 to 20 to be fossils).

```
taxset fossils=13-20
```

And some topological constraints to make the analysis a bit faster and to allow to calibrate these groups in the node calibration analysis.

```
constraint root=1-.
constraint Haplorhini=2-.
constraint Simiiformes=3-.
constraint Catarrhini=3-11
constraint Hominidae=3-7
constraint Cercopithecidae=8-11
```

4.2 Setting the data partitions

We can then partition our data by using `charset`. We need a DNA and a morphology partition.

```
charset DNA=1-898
charset morphology=899-1089
```

Now that the partitions are identified we need to specify that we want to use both partitions.

```
partition my_parition=2: DNA, morphology
set partition=my_parition
```

4.3 Setting the evolutionary models

Now that our data are ready to use, we have to parametrise our phylogenetic model. We need to set up a substitution model for each partition and (later on) a clock model for dating our phylogeny.

For the molecular data, we can go with a **GTR** model (General Time Reversible). This model allows each transition rate between each nucleotide to be different (see here for more models and more (easy) explanations). To simplify the model, we can also set a finite number of transition rates categories (for example, really slow, slow, medium, fast, etc.) that will follow a certain distribution. Here we are going to use the **Gamma** (Γ) distribution with 4 distinct categories.

Model for the molecular data (GTR + 4Γ)

```
lset applyto=(1) nucmodel=4by4 nst=6 rates=gamma Ngammacat=4 covarion=no
```

For the morphological data, we are going to use the M*k* model that assumes an equal transition rate between each state for each character. Again, to simplify the model, we will set a 4 transition rates categories following a Gamma distribution.

Model for the molecular data (M*k* + 4Γ)

```
lset applyto=(2) nst=1 rates=gamma Ngammacat=4
```

We can look at our model settings at any time by typing the following:

```
help lset
```

4.4 Setting the priors

We can also set some Bayesian priors for the analysis. Most of the priors can be set to be flat (i.e. uninformative: the information is only a prior in a Bayesian sense and is not giving an *a priori* to our analysis).

```
help prset
```

We are just going to change the rate priors to be variable.

```
prset applyto=(all) ratepr=variable
```

4.5 Summarising the parameters

Finally, we can summarise our whole model parameters by typing the following:

```
showmodel
```

Because all these manipulations only concern the data set, it might be a good idea to save them directly in the `TotalEvidence_Primates.nex` file so that we don't have to retype them each time prior to any dating analysis. This can be done between the following tags at the end of your nexus document :

```
begin mrbayes;
    ...
end;
```

Using the indentation is a good practice for emphasizing that the `...` will be executed within the `begin mrbayes` tag. You can add all the commands below in place of the `...` and **finish all command lines by a semi-colon (;)**. You can also use the square brackets (`[]`) to write down some comments, they will be ignored by MrBayes.

4.6 The tip-dated analysis

Set up the tip ages

The first step in a Tip-dating analysis is to set up the age of the tips. We can only inform the fossils ages since the living species have an aged assumed to be 0 million years old by default.

```
calibrate Catopithecus_browni=Fixed(36)
calibrate Aegyptopithecus_zeuxis=Fixed(33)
calibrate Dendropithecus_macinnesi=Fixed(20)
calibrate Noropithecus_bulukensis=Fixed(18)
calibrate Victoriapithecus_macinnesi=Fixed(15)
calibrate Proconsul_major=Fixed(20)
calibrate Afropithecus_turkanensis=Fixed(16)
calibrate Morotopithecus_bishopi=Fixed(18)
```

We can also set up some calibrations to fix the age of the root and the Haplorhini group. Because we are not sure of the age of fossils, we can use a simple uniform distribution between the oldest and the youngest age estimate.

```
calibrate root=uniform(60.99,76.72)
calibrate Haplorhini=uniform(57.62,69.59)
```

Set up the clock model

Now we can set up the clock model by specifying that the branch length should represent time which is uniformly distributed along the branches (i.e. one unit of branch length = one unit of time across the whole tree; `brlenspr=clock:uniform`). We can then set the clock to be relaxed by using the **IGR** (Independent Gamma rates) model for the clock rate variation. This model assumes an independent rate on each branch distributed following a Gamma distribution (`clockvarpr=igr`). Finally we can set some node calibration using the calibration we set up before.

```
prset brlenspr=clock:uniform
prset clockvarpr=igr
prset nodeagepr=calibrated
prset clockratepr=normal(0.01,0.005)
```

Along with our topological constraint:

```
prset topologypr=constraints(root, Haplorhini)
```

Run the analysis

Finally we can run the analysis by setting up some of the MCMC parameters: the number of chains (`nchains=4`), the number of independent runs (`nruns=2`), which generations to sample (`samplefreq=1000`), which ones to print (`printfr=100`) and when to run the convergence diagnosis between both chains (`diagnfreq=2500`). We can finally run the MCMC for a certain number of generations (1500000!) and see if the two chains converge.

```
mcmcp temp=0.1 nchain=4 samplefreq=1000 printfr=100 nruns=2 diagnfreq=2500
mcmcp filename=Primates-tip_dating
mcmc ngen=150000 Stoprule=YES stopval=0.01
```

Save the tree

Ones the analysis is over, we can save the parameters estimations by typing:

```
sump filename=Primates-tip_dating relburnin=YES burninfrac=0.25
```

Each parameters are saved in the `Primates-tip_dating.pstat` file.
As well as the tree by typing:

```
sumt filename=Primates-tip_dating relburnin=YES burninfrac=0.25
```

The consensus tree is saved in the `Primates-tip_dating.con.tre` file.

4.7 The calibrated analysis

Secondly we want to compare our results to the node calibration analysis (without fossils). Most parameters are the same so they will not be described in details.

Remove the fossil species

We first have to remove our fossil partition:

```
delete fossils
```

Set up the calibrations

We then need to set up our calibration (with uniform distributions for accounting for uncertainty).

```
calibrate root=uniform(60.99,76.72)
calibrate Haplorhini=uniform(57.62,69.59)
calibrate Simiiformes=uniform(33.55,49.48)
calibrate Catarrhini=uniform(57.62,69.59)
calibrate Hominidae=uniform(11.04,20.8)
calibrate Cercopithecidae=uniform(8.93,18.27)
```

Set up the clock model

We can use the same parameters as for the tip-dating analysis.

```
prset brlenspr=clock:uniform
prset clockvarpr=igr
prset nodeagepr=calibrated
prset clockratepr=normal(0.01,0.005)
```

Along with our topological constraint:

```
prset topologypr=constraints(root, Haplorhini)
```

Run the analysis

```
mcmcp temp=0.1 nchain=4 samplefreq=1000 printfr=100 nruns=2 diagnfreq=2500
mcmcp filename=Primates-node_dating
mcmc ngen=150000 Stoprule=YES stopval=0.01
```

Save the tree

Summarising the parameters:

```
sump filename=Primates-node_dating relburnin=YES burninfrac=0.25
```

Summarising the trees:

```
sumt filename=Primates-node_dating relburnin=YES burninfrac=0.25
```

5 Comparing both trees

We can now use R to compare both trees and see what is the effect of the both methods.

5.1 Load the packages

```
Theses are the pacakges we will be using.
## Installing the packages
if(!require(ape)) install.packages("ape")
if(!require(devtools)) install.packages("devtools")
library(ape, devtools)
install_github("TGuillerme/dispRity", ref = "release")
```

5.2 Import the trees

First we need to import the trees in R. Note that here I am using two pre-made trees, the ideal would be to use your own trees!

```
## Reading the trees
## WARNING: Make sure you're in the right folder! Use setwd(), getwd() and
## list.files() to navigate.
node_calibrated_tree <- read.nexus("Primates-node_dating.bkp.tre")
tip_dated_tree       <- read.nexus("Primates-tip_dating.bkp.tre")

## Removing the fossils from the tip dated tree
tip_dated_tree_living <- drop.tip(tip_dated_tree, tip = c("Catopithecus_browni",
  "Aegyptopithecus_zeuxis", "Dendropithecus_macinnesi",
  "Noropithecus_bulukensis", "Victoriapithecus_macinnesi", "Proconsul_major",
  "Afropithecus_turkanensis", "Morotopithecus_bishopi"))
```

5.3 Comparing the trees

Then we can compare the branch length and the topology from each tree:

```
## Calculate the age difference between both trees for each nodes
library(dispRity)
node_calibrated_ages  <- tree.age(node_calibrated_tree)
tip_dated_living_ages <- tree.age(tip_dated_tree_living)
tip_dated_ages        <- tree.age(tip_dated_tree)
```

We can now plot the two trees (including the two versions of the tip dated tree):

```
## Setting the graphical parameters
op <- par(mfrow = (c(1,3)))

## Plotting the node calibrated tree
plot(node_calibrated_tree, main = "node calibrated", cex=0.7)
## Adding the age of the nodes calculates using tree.age()
nodelabels(text = round(node_calibrated_ages$ages[13:23], 2), cex = 0.5)
## Adding the time ruler
axisPhylo()

## Plotting the tip dated tree without fossils (with the labels facing the
## calibrated one)
plot(tip_dated_tree_living, main = "tip-dated (living only)", direction = "l",
  cex=0.7)
## Adding the age of the nodes calculates using tree.age()
nodelabels(text = round(tip_dated_living_ages$ages[13:23], 2), cex = 0.5)
## Adding the time ruler
axisPhylo()

## Plotting the tip dated tree with fossils
plot(tip_dated_tree, main = "tip-dated", cex=0.7)
## Adding the age of the nodes calculates using tree.age()
nodelabels(text = round(tip_dated_ages$ages[21:38], 2), cex = 0.5)
## Adding the time ruler
axisPhylo()

## Resetting the graphical parameters to default
par(op)
```

6 Questions

1. What are the main differences between the trees? Are their also differences in parameters estimations?
2. Which one of these trees is the best? For which purpose is it better suited?
3. Can you see some caveats for each method?
4. Which method took longer and why?

References

Hayasaka, K., T. Gojobori, and S. Horai. 1988. Molecular phylogeny and evolution of primate mitochondrial dna. Molecular Biology and Evolution 5:626–644.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539–42.

Stevens, N. J., E. R. Seiffert, P. M. O'Connor, E. M. Roberts, M. D. Schmitz, C. Krause, E. Gorscak, S. Ngasala, T. L. Hieronymus, and J. Temu. 2013. Palaeontological evidence for an oligocene divergence between old world monkeys and apes. Nature 497:611–614.