

# bModelTest in BEAST 2.4.x

## Site model averaging without tears

Remco Bouckaert

June 22, 2016

One of the decisions to make when performing a phylogenetic analysis using nucleotide data is setting up the site model and associated substitution model. Instead of fixing these settings, bModelTest allows site model averaging through bModelTest [1], which makes it easy to set up the site model: just choose bModelTest. In this tutorial, we go through an analysis using bModelTest in BEAST v2.4.x[2], and look into how to interpret the results. This tutorial assumes you already have done one of the other tutorials, and are familiar with BEAUti, BEAST and Tracer, but if you are not, you may try to start with the Divergence dating tutorial, available from <http://beast2.org/tutorials/>.

For this tutorial, you need

- BEAST version 2.4.x, available from <http://beast2.org/>
- Tracer version 1.6, available from <http://tree.bio.ed.ac.uk/software/tracer/>

We will run through the following steps:

- Install bModelTest package
- Set up analysis in BEAUti
- Run analysis with BEAST
- Analyse using Tracer
- Analyse using BModelAnalyser

## Install bModelTest package

Make sure that you have the bModelTest version 1.6 package installed. It depends on the BEASTlabs package, which should be installed when you install bModelTest. See <http://beast2.org/managing-packages/> for details on how to install packages.

## Set up analysis in BEAUti

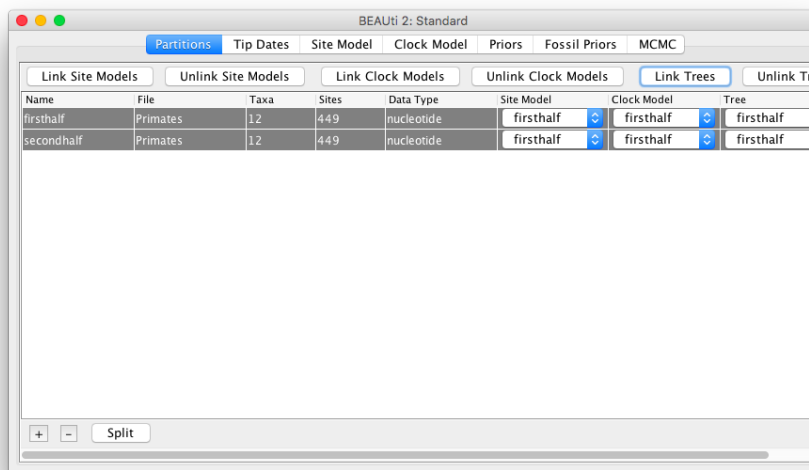
We will analyse an alignment of 12 primates with 898 sites [3].

### Import alignment

First, start BEAUti (restart, if you just installed the bModelTest package through BEAUti), and we need to import the alignment. The alignment is in the file `examples/neuxs/Primates.nex` which you can find in the BEAST directory. It is probably easiest to select the menu **File/Set working dir/BEAST** to start in the BEAST directory, then select menu **File/Import alignment**, and browse to `example/nexus/Primates`.

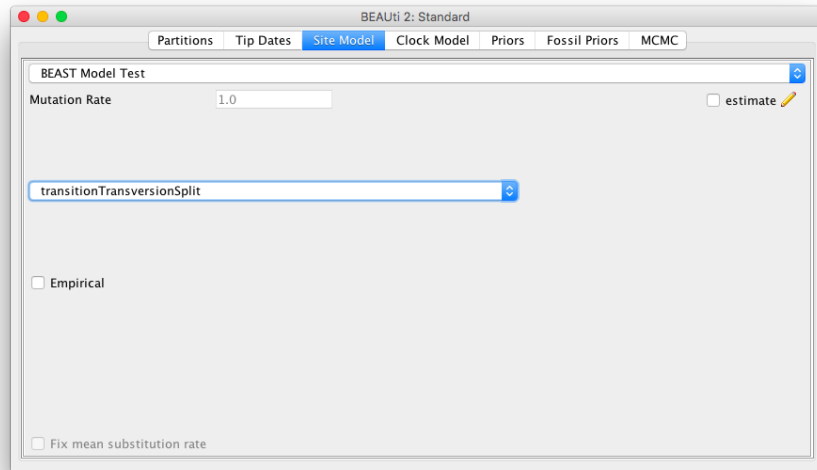
### Link partitions

After importing the alignment, two partitions will be shown. We want to link everything as if it were a single alignment for this tutorial. To do this, select both alignments in the table, and click the **Link Site Models** button, then **Link Clock Models** and finally **Link Trees**. The screen should look something like this:



### Set up site model

Click the **Site Model** tab in BEAUti. Select the drop down box at the top which says **Gamma Site Model** and change to **BEAST Model Test**. That's all. The screen should look something like this:



## Set up clock model

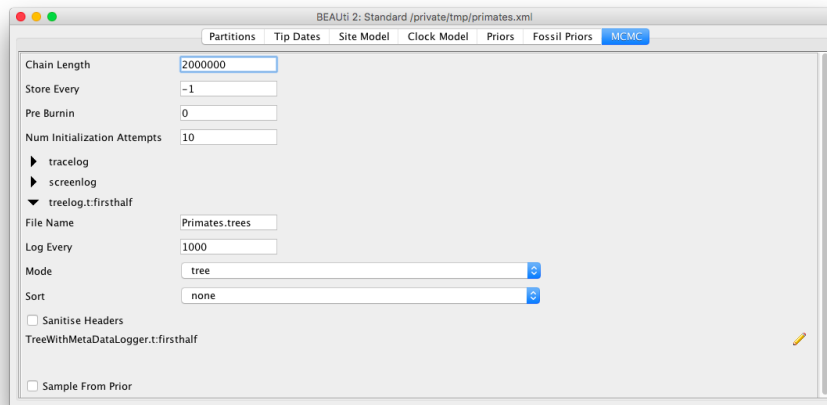
Click the **Clock model** tab. For this analysis, we use a strict clock, which is the default, so no further changes needed.

## Set up priors

Click the **Priors** tab to show the priors. We will go with default priors for this tutorial.

## Set up MCMC

Click the **MCMC** tab, and change the **chain length** to 2 million, and the file name for the tree trace to **Primates.trees**:



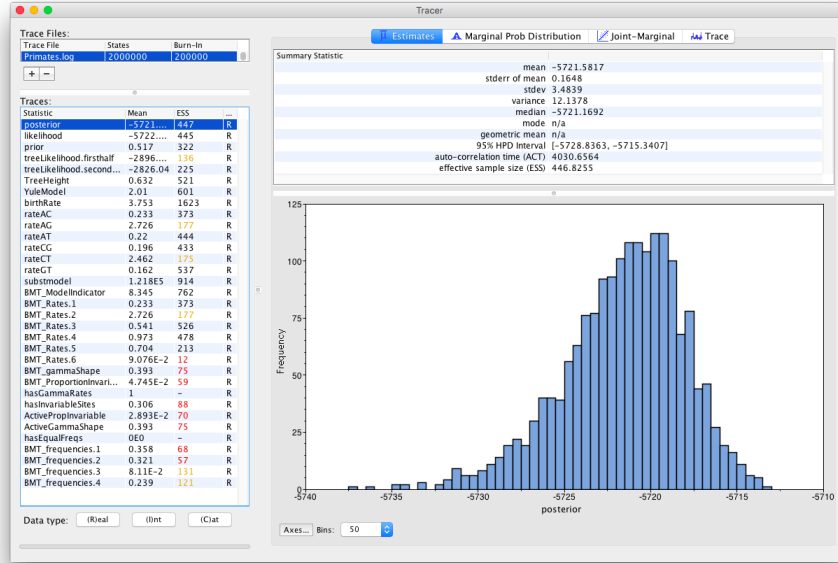
Save the file using the File/Save menu in say `primates.xml`

## Run analysis with BEAST

Run `primates.xml` using BEAST, as you are used to. Since this is a small alignment, it should take only a minute or two. If ESSs are not satisfactory, you can resume the run for another million samples or so.

## Analyse using Tracer

Open `Primates.log` in Tracer. You will see the following entries:

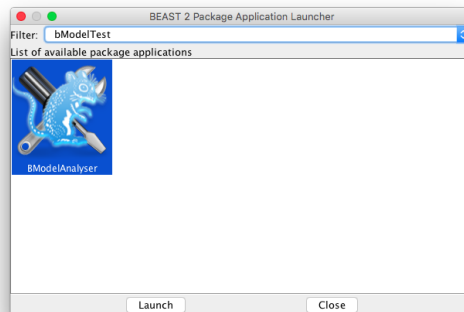


- BMT\_ModelIndicator is the index of the substitution model as listed in the Appendix.
- substmodel is the model represented as a 6-digit number, where the position of the digit refers to rates ac, ag, at, cg, ct and gt respectively, and equal digits indicates that rates are shared, so 111111 is Jukes Cantor (if frequencies are kept equal), 121121 is HKY, 123456 is GTR etc.
- rateAC,...,rateGT are the rates according to the model. ESSs should be good for these rates, but if you plot joint-marginals of pairs you may find high correlation between some of these rates.
- BMT\_Rates.1 to 6 are the rates used to build up the rate matrix. If only low parameter models are samples, the higher rates will be sampled very infrequently, and you should expect low ESSs for them. Correlation between pairs of rates should be low.
- BM\_gammaShape is the gamma shape parameter as it is being sampled. For parts of the chain that gamma rate heterogeneity is switched off, the parameter will not be sampled, and the trace will show periods where the parameter is stuck.
- hasGammaRates indicates whether gamma rate heterogeneity it used (1) or not used (0). The mean can be interpreted as the proportion of time that gamma rate heterogeneity is switched on.

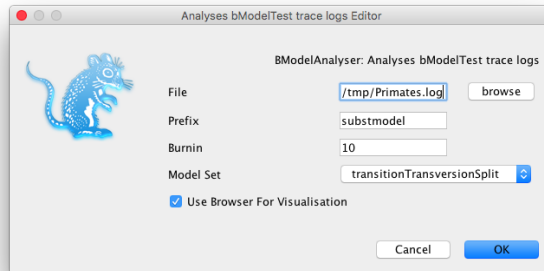
- ActiveGammaShape is the gamma shape parameter when it is sampled, but it is zero when it is not sampled. To get the estimate of the mean of the shape parameter, divide the mean ActiveGammaShape by the mean of hasGammaRates.
- BMT\_ProportionInvariable, hasInvariableSites and ActivePropInvariable are the value for proportion invariable similar to BMG\_gammaShape, hasGammaRates and ActiveGammaShape respectively.
- hasEqualFreqs indicates whether equal frequencies are used and the mean can be interpreted as the proportion of time that equal frequencies is used. When empirical frequencies are used, this parameter is not reported.

## Analyse using BModelAnalyser

BModelAnalyser is a utility that comes with the bModelTest package. To start it, in BEAUti select the **File/Launch Apps** menu. A window pops up where you can select the BModelAnalyser App (filter on bModelTest at the top of the dialog if you have many packages).



When you click the **Launch** button, after a little while, a dialog appears where you can select the log file:



- **File** is the trace log file produced by BEAST.
- **Prefix** is the prefix of the entry in the log file containing the substitution model trace (default 'substmodel', which is fine when the analysis is set up in BEAUti)
- **burnin** is percentage of the log file to disregard as burn-in.
- **Model Set** is the set of models to choose from, should be the same as used in the BEAST XML that generated the log file, so if you selected a non-default substitution model set, you need to change the model set.
- **Use Browser For Visualisation** use default web browser for visualising the resulting dot graph.

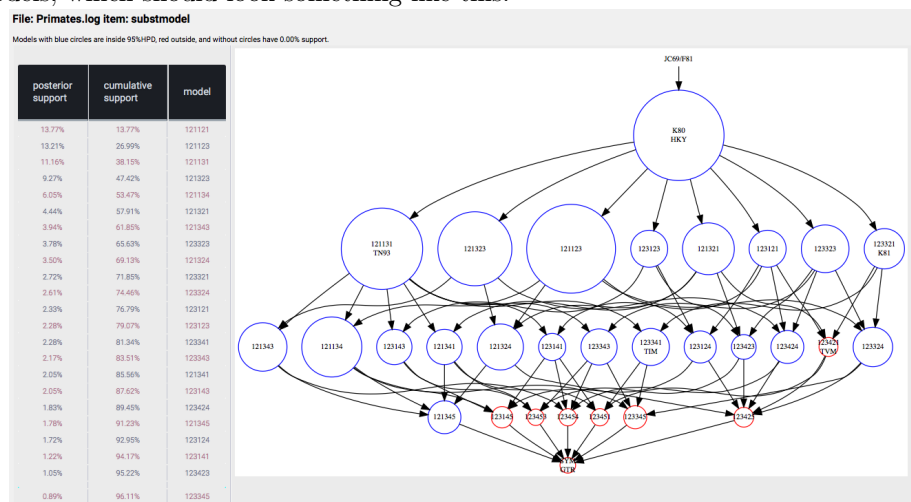
Click OK, and a console pops up with model coverage results.

```
BEAST v2.4.2 Prerelease BModelAnalyser
REPLACE FOR V2.4.1 in BModelAnalyser.run(): LogAnalyser analyser = new LogAnalyser(file.getAbsolutePath(), burnin, false, false);
Loading /tmp/Primates.log, burnin 10%, skipping 200 log lines

|-----|-----|-----|-----|-----|-----|
*****
substmodel
13.77% 13.77% 121121
13.21% 26.99% 121123
11.16% 38.15% 121131
9.27% 47.42% 121323
6.05% 53.47% 121134
4.44% 57.91% 121321
3.94% 61.85% 121343
3.78% 65.63% 123323
3.50% 69.13% 121324
2.72% 71.85% 123321
2.61% 74.46% 123324
2.33% 76.79% 123121
2.28% 79.07% 123123
2.28% 81.34% 123341
2.17% 83.51% 123343
2.05% 85.56% 121341
2.05% 87.62% 123143
1.83% 89.45% 123424
1.78% 91.23% 121345
1.72% 92.95% 123124
1.22% 94.17% 123141
1.05% 95.22% 123423

Got it!
```

After a little while, a page is opened in your web browser containing the same information, but better formatted, and it shows a graph containing the models, which should look something like this:



Arrows indicate that the model at the tail is nested inside the model at the head and can be obtained by adding one parameter to the model.

The area of the circles around models are proportional to the posterior support for these models. In this case, we see a lot of support for HKY, 121123 and TN93.



Blue circles indicate the model is contained in the 95% credible set, red circles indicate the model is outside the 95% credible set, and no circles indicate there is hardly any support (if at all) in the posterior.

## References

- [1] Remco Bouckaert and Alexei Drummond. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *bioRxiv*, page 020792, 2015.
- [2] Remco R. Bouckaert, Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh-Hsi Wu, Dong Xie, Marc A Suchard, Andrew Rambaut, and Alexei J Drummond. BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput Biol*, 10(4):e1003537, Apr 2014.
- [3] Kenji Hayasaka, Takashi Gojobori, and Satoshi Horai. Molecular phylogeny and evolution of primate mitochondrial dna. *Molecular Biology and Evolution*, 5(6):626–644, 1988.