

# \*BEAST in BEAST 2

BEAST – Bayesian Evolutionary Analysis Sampling Trees

Estimating Species Trees from Multilocus Data

This is a program for a full Bayesian framework for species tree estimation. The \*BEAST (pronounced "star beast") estimates the species tree that is most probable given the (multi-individual) multi-locus sequence data and to co-estimates gene trees embedded in a shared species tree.

The following software is needed

- BEAST - contains the BEAST program, BEAUti, TreeAnnotator, LogCombiner and DensiTree. It is available at <http://www.beast2.org>.
- Tracer – program to explore BEAST output. Graphical and quantitative summary of the distributions of parameters, gives diagnostic info. It is available at <http://beast.bio.ed.ac.uk/>.
- FigTree – program to display phylogenetic trees, especially those from BEAST. It is available at <http://tree.bio.ed.ac.uk/>.

## 1. convert NEXUS file into a BEAST XML file using BEAUti (Bayesian Evolutionary Analysis Utility)

Here you set up input data (alignments) evolutionary models and options for MCMC analysis

- set the template using: File -> Template -> StarBeast
- load the alignments (one locus per file): File -> Import Alignment (select all NEXUS files holding Shift key)
- the partitions are displayed in the main window
- here you can link/unlink site and clock models and trees, tree models should be linked when linked genetically (e.g., all loci from cpDNA)
- **Taxon sets** – specify here what samples are the same species and the name of the species (either manually, by guessing or uploading a mapping file)
- **Site Model** – set substitution model for each locus, see below how to set particular models
- **Clock Model** – set clock model, 'Estimate' check box is unchecked for the first locus and checked for the rest (i.e., substitution rate is estimated relatively to the first locus)
  - strict clock – every branch in a phylogenetic tree evolves according to the same evolutionary rate (single parameter)
  - uncorrelated relaxed clock – each branch of a phylogenetic tree has its own evolutionary rate. Uncorrelated means that the rate at one branch does not depend upon the rate at any of the neighboring branches, i.e. abrupt changes are permitted and there is no need for steadily increase or decrease. Rates are sampled from a probability distribution (Exponential or Log Normal here).
  - random local clock – more than the strict clock (no variation) but less than the relaxed clock (different rate for each branch)

- **Multi Species Coalescent** – settings to the multi species coalescent model. Ploidy of nuclear loci should be ‘autosomal\_nuclear’ and ploidy of chloroplast loci ‘Y or mitochondrial’. Three population size models are available: linear and constant root, linear, constant. Usually set the first one.
- **Priors** – priors for each parameter in the model. Keep ‘Yule Model’, i.e., constant birth-death prior. This specifies the distribution of the speciation times.
- **MCMC** – more general settings to control the MCMC and files
  - Length of chain – number of steps the MCMC will make, depend on the size of the dataset
  - Store Every – the frequency of saving the parameters
  - File Name – set the file names
  - Log Every – the saving frequency for tracelog and TreeLogger
- save the XML file: File -> Save As

## 2. Run the BEAST analysis

- Choose File.. to select the XML file
- click Run
- wait until the analysis is finished (the message about Total calculation time and End likelihood appears)

## 3. Analyze and explore BEAST output using Tracer

- File -> Import Trace File... and select the \*.log file
- on the left side – different parameters that have been logged
- right side – analysis of the trace depending on the tab selected (select multiple by holding Shift key)
  - Estimates – various statistics, incl. mean, median and HPD (highest posterior density interval, i.e., Bayesian analog to a confidence interval)
  - Marginal Density – (marginal posterior) distribution for the selected parameter
  - Trace – logged values of the MCMC

## 4. Estimate of the phylogenetic tree using TreeAnnotator

- summarize sampled trees – takes the set of trees and identifies a single tree that best represents the posterior distribution, then annotates each clade (mean age, 95% HDP, posterior clade probability etc.)
- **Burnin percentage** – percentage of trees to remove
- **Posterior probability limit** – annotate only nodes above the limit (0 = annotate all nodes)
- **Target tree type** – Maximum clade credibility tree – finds the tree with the highest product of the posterior probability of all its nodes
- **Node heights** – sets the heights (ages) of each node in the tree to the mean/median height across the entire sample of trees for that clade
- **Input Tree File** – select the \*.trees file
- **Output File** – name the resulting species tree (e.g., with the suffix \*.tre)
- press Run to start the calculation

## 5. View the resulting species tree in FigTree

- Open the \*.tre file with File -> Open...
- select Branch Labels to display ‘posterior’ – these are posterior probabilities for each node

## DNA substitution models in BEAST/BEAUti

How to set up specific model in BEAUti: set a base model (one of four available) and set additional settings

<i>Model</i>	<i>Base Model</i>	<i>Additional specifications in BEAUti 2</i>
• JC69	JC69	none
• TrN	TN93	none
• TrNef	TN93	base Frequencies set to "All Equal"
• K80 (K2P)	HKY	base Frequencies set to "All Equal"
• F81	GTR	fix all rate parameters to 1.0 (uncheck the "estimate" box)
• HKY	HKY	none
• SYM	GTR	base Frequencies set to "All Equal"
• TIM	GTR	fix CT and AG rate parameters to 1.0 (uncheck the "estimate" box)
• TVM	GTR	fix the AG rate parameter to 1.0 (uncheck the "estimate" box)
• TVMef	GTR	fix the AG rate parameter to 1.0 (uncheck the "estimate" box), and also set base Frequencies to "All Equal"
• GTR	GTR	none

For +G (rate heterogeneity) edit 'Gamma Category Count' (usually to 4) and set the 'Shape' (or keep it to be estimated).

For +I set the 'Proportion Invariant'.

These comments are based on:

[https://beast2-dev.github.io/beast-docs/beast2/STARBEAST/StarBEAST\\_tutorial.html](https://beast2-dev.github.io/beast-docs/beast2/STARBEAST/StarBEAST_tutorial.html)

<https://justinbagley.rbind.io/2016/10/11/setting-dna-substitution-models-beast/>

<https://beast.community/clocks>