Tutorial using BEAST v2.7.7
contraband Tutorial

Rong Zhang and Fábio K. Mendes

Phylogenetic inference using continuous traits based on Brownian motion models

1 Background

Time-scaled phylogenetic trees are an ultimate goal of evolutionary biology and a necessary ingredient in comparative studies. The accumulation of genomic data has resolved the tree of life to a great extent, yet timing evolutionary events remains challenging if not impossible without external information such as fossil ages and morphological characters.

Methods for incorporating morphology in tree estimation have lagged behind their molecular counterparts, especially in the case of continuous characters that are scored at a resolution and variable within and across species. Popular continuous character phylogenetic models are based on Brownian motion (BM) and can incorporate correlated evolution among traits, which are assumed to evolve as a random walk whose diffusion rate is the evolutionary rate. Using continuous characters in total-evidence tip dating thus not only has the potential to improve phylogenetic inference by enhancing morphological data sets but also provides natural workarounds for the issues observed under discrete-character models.

While many computational methods exist for the study of morphological character evolution, tools capable of jointly modeling molecular and morphological characters are still lacking, particularly those that simultaneously account for uncertainty in species tree topology and branch lengths. One way forward should be easily visible in the joint evolutionary modeling of all available data, whereby different sources of data inform on each other's model parameters and on the phylogeny itself.

In this tutorial, we introduce the **contraband** (**con**tinuous **tra**its **b**rowni**an** mo**d**els) BEAST 2 package that provides a suite of models for the study of continuous trait evolution in a Bayesian framework. By compiling a data set comprised of contemporaneous and ancient molecular sequences, and discrete and continuous characters from living and extinct Carnivora taxa, we show how to leverage continuous morphology in Phylogenetic inference.

2 Programs used in this Exercise

2.0.1 BEAST2 - Bayesian Evolutionary Analysis Sampling Trees2

BEAST2 (http://www.beast2.org) is a free software package for Bayesian evolutionary analysis of molecular sequences using MCMC and strictly oriented toward inference using rooted, time-measured phylogenetic trees. This tutorial is written for BEAST v2.7.7 (**bouckaert2019beast**).

2.0.2 BEAUti2 - Bayesian Evolutionary Analysis Utility

BEAUti2 is a graphical user interface tool for generating BEAST2 XML configuration files.

Both BEAST2 and BEAUti2 are Java programs, which means that the exact same code runs on all platforms. For us it simply means that the interface will be the same on all platforms. The screenshots

used in this tutorial are taken on a Mac OS X computer; however, both programs will have the same layout and functionality on both Windows and Linux. BEAUti2 is provided as a part of the BEAST2 package so you do not need to install it separately.

2.0.3   TreeAnnotator

TreeAnnotator is used to produce a summary tree from the posterior sample of trees using one of the available algorithms. It can also be used to summarise and visualise the posterior estimates of other tree parameters (e.g. node height).

TreeAnnotator is provided as a part of the BEAST2 package so you do not need to install it separately.

2.0.4   Tracer

Tracer (http://tree.bio.ed.ac.uk/software/tracer) is used to summarise the posterior estimates of the various parameters sampled by the Markov Chain. This program can be used for visual inspection and to assess convergence. It helps to quickly view median estimates and 95% highest posterior density intervals of the parameters, and calculates the effective sample sizes (ESS) of parameters. It can also be used to investigate potential parameter correlations. We will be using Tracer v1.7.2.

2.0.5   FigTree

FigTree (http://tree.bio.ed.ac.uk/software/figtree) is a program for viewing trees and producing publication-quality figures. It can interpret the node-annotations created on the summary trees by TreeAnnotator, allowing the user to display node-based statistics (e.g. posterior probabilities). We will be using FigTree v1.4.4.

3   Practical Part I: Data preparation

4   Practical Part II: Parameter and State inference using contraband

In this tutorial we will estimate evolutionary rates, trait correlations, ancestral states and phylogenetic trees using the Brownian motion implemented in BEAST2, contraband.

The aim is to:

- Learn how to infer phylogenetic trees with continuous traits/characters
- Get to know how to choose the set-up of such an analysis
- Learn how to read the output of a "contraband" analysis

4.1   Setting up an analysis in BEAUti

4.1.1   Download contraband

First, we have to download the package contraband using the BEAUTi package manager. Go to *File >> Manage Packages* and download the package contraband.

contraband will only be available in BEAUti once you close and restart the program.

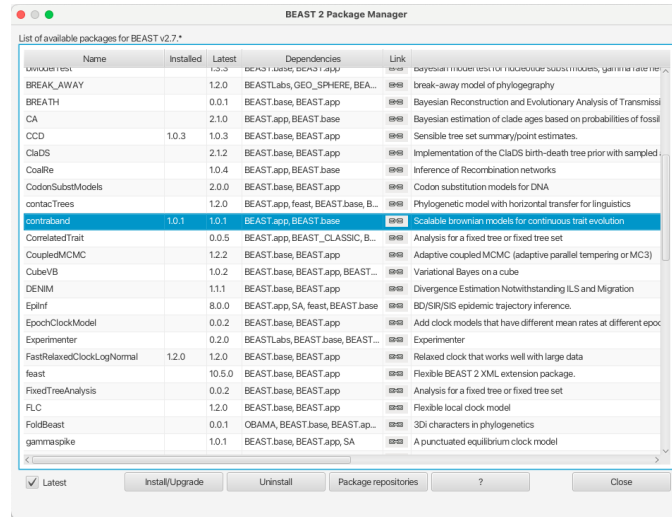### 4.1.2   Loading the Carnivoran Molecular Sequences

Figure 1: Download the contraband package.

4.1.3 Loading DATA the (Partitions)

The sequences from the *data* folder name *XXX.nexus* can be either drag and dropped into BEAUti or added using BEAUti's menu system via *File >> Import Alignment*. Once the sequences are added, we need to specify the sampling dates.

4.1.4 Get the sampling times (Tip Dates)

Open the "Tip Dates" panel and then select the "Use tip dates" checkbox.

The sampling times are encoded in the sequence names. We can tell BEAUti to use these by clicking the *Auto-configure* button. The sampling times appear following the third vertical bar "|" in the sequence name. To extract these times, select "split on character", enter "|" (without the quotes) in the text box immediately to the right, and then select "3" from the drop-down box to the right, as shown in the figure below. Clicking "Ok" should now populate the table with the sample times extracted from the sequence names: the column **Date** should now have values between 2000 and 2002 and the column **Height** should have values from 0 to 2. The heights denote the time difference from a sequence to the most recently sampled sequence. If everything is specified correctly, the sequence with Height 0.0 should have Date 2001.9.

4.1.5 Specify the Brownian motion Model

4.1.6 Set the clock model (Clock Model)

4.1.7 Specify the priors (Priors)

Now, we need to set the priors for the various parameters of the model. You can find the parameter priors below the tree prior.

4.1.8 Specify the MCMC chain length (MCMC)

Here we can set the length of the MCMC chain and after how many iterations the parameter and trees a logged. For this dataset, 2 million iterations should be sufficient. In order to have enough samples but not create too large files, we can set the logEvery to 2000, so we have 1001 samples overall. Next, we have to save the *.xml file under *File >> Save as.*

4.1.9 Run the Analysis using BEAST2

Run the *.xml using BEAST2 or use finished runs from the *precooked-runs* folder. The analysis should take about 6 to 7 minutes.

5 Practical Part III: Post analysis

5.0.1 Analyse the log file using Tracer

First, we can open the *.log file in tracer to check if the MCMC has converged. The ESS value should be above 200 for almost all values and especially for the posterior estimates.

5.0.2 Make the summary tree using TreeAnnotator

Producing MCC tree

Open **TreeAnnotator** and then set the options as in the Figure 1 below. You have to specify the **Burnin percentage**, **Target tree type**, **Node heights**, **Input Tree File** and the **Output File**. Use the typed trees in the file H3N2.H3N2.trees as **Input Tree File**. Name output file H3N2.mcc.tree. After clicking **Run** the program should summarize the trees.

5.0.3 Analyse and compare the MCC trees

5.0.4 Errors that can occur (Work in progress)

Version dated: June 12, 2025