Divergence Time Estimation using BEAST v2.1.3

Central among the questions explored in biology are those that seek to understand the timing and rates of evolutionary processes. Accurate estimates of species divergence times are vital to understanding historical biogeography, estimating diversification rates, and identifying the causes of variation in rates of molecular evolution.

This tutorial will provide a general overview of divergence time estimation and fossil calibration in a Bayesian framework. The exercise will guide you through the steps necessary for estimating phylogenetic relationships and dating species divergences using the program BEAST v2.1.3.

Background: Divergence time estimation

Estimating branch lengths in proportion to time is confounded by the fact that the rate of evolution and time are intrinsically linked when inferring genetic differences between species. A model of lineage-specific substitution rate variation must be applied to tease apart rate and time. When applied in methods for divergence time estimation, the resulting trees have branch lengths that are proportional to time. External node age estimates from the fossil record or other sources are necessary for inferring the real-time (or absolute) ages of lineage divergences (Figure 1).

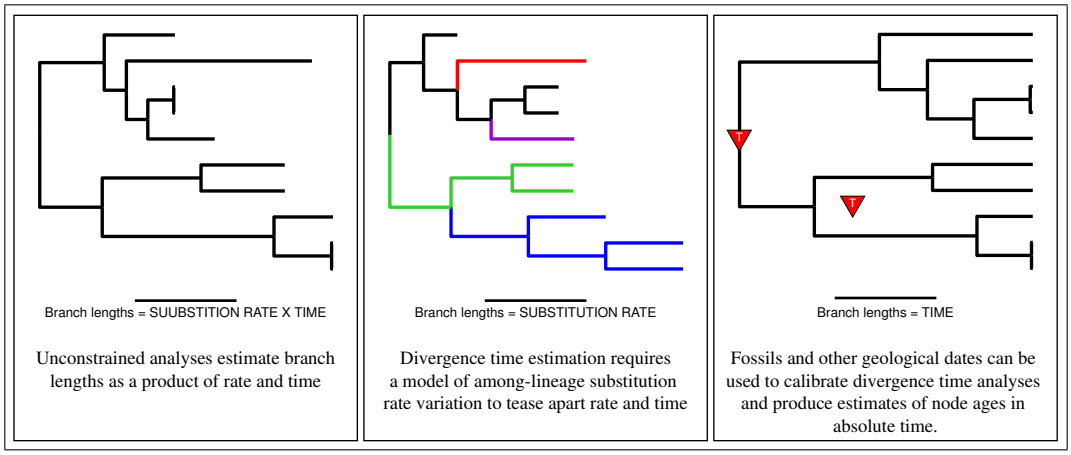

Figure 1: Estimating branch lengths in units of time requires a model of lineage-specific rate variation, a model for describing the distribution of speciation events over time, and external information to calibrate the tree.

Ultimately, the goal of Bayesian divergence time estimation is to estimate the joint posterior probability, $\mathbb{P}(\mathcal{R}, \mathcal{T}|\mathcal{S}, \mathcal{C})$, of the branch rates ($\mathcal{R}$) and times ($\mathcal{T}$) given a set of sequences ($\mathcal{S}$) and calibration information ($\mathcal{C}$):

$$\mathbb{P}(\mathcal{R}, \mathcal{T} \mid \mathcal{S}, \mathcal{C}) = \frac{\mathbb{P}(\mathcal{S} \mid \mathcal{R}, \mathcal{T})\ \mathbb{P}(\mathcal{R})\ \mathbb{P}(\mathcal{T} \mid \mathcal{C})}{\mathbb{P}(\mathcal{S} \mid \mathcal{C})},$$

where $\mathbb{P}(\mathcal{S}|\mathcal{R}, \mathcal{T})$ is the likelihood, $\mathbb{P}(\mathcal{R})$ is the prior probability of the rates, $\mathbb{P}(\mathcal{T}|\mathcal{C})$ is the prior probability of the times, and $\mathbb{P}(\mathcal{S}|\mathcal{C})$ is the marginal probability of the data. We use numerical methods—Markov chain Monte Carlo (MCMC)—to eliminate the difficult task of calculating the marginal probability of the data. Thus, our primary focus, aside from the tree topology, is devising probability distributions for the prior on the rates, $\mathbb{P}(\mathcal{R})$, and the prior on the times, $\mathbb{P}(\mathcal{T}|\mathcal{C})$.

Modeling lineage-specific substitution rates

Many factors can influence the rate of substitution in a population such as mutation rate, population size, generation time, and selection. As a result, many models have been proposed that describe how substitution rate may vary across the Tree of Life.

The simplest model, the molecular clock, assumes that the rate of substitution remains constant over time (Zuckerkandl and Pauling, 1962). However, many studies have shown that molecular data (in general) violate the assumption of a molecular clock and that there exists considerable variation in the rates of substitution among lineages.

Several models have been developed and implemented for inferring divergence times without assuming a strict molecular clock and are commonly applied to empirical data sets. Many of these models have been applied as priors using Bayesian inference methods. The implementation of dating methods in a Bayesian framework provides a flexible way to model rate variation and obtain reliable estimates of speciation times, provided the assumptions of the models are adequate. When coupled with numerical methods, such as MCMC, for approximating the posterior probability distribution of parameters, Bayesian methods are extremely powerful for estimating the parameters of a statistical model and are widely used in phylogenetics.

Some models of lineage-specific rate variation:

- Global molecular clock: a constant rate of substitution over time (Zuckerkandl and Pauling, 1962)
- Local molecular clocks (Kishino, Miyata and Hasegawa, 1990; Rambaut and Bromham, 1998; Yang and Yoder, 2003; Drummond and Suchard, 2010)
  - Closely related lineages share the same rate and rates are clustered by sub-clades
- Compound Poisson process (Huelsenbeck, Larget and Swofford, 2000)
  - Rate changes occur along lineages according to a point process and at rate-change events, the new rate is a product of the old rate and a Γ-distributed multiplier.
- Autocorrelated rates: substitution rates evolve gradually over the tree
  - Log-normally distributed rates: the rate at a node is drawn from a log-normal distribution with a mean equal to the parent rate (Thorne, Kishino and Painter, 1998; Kishino, Thorne and Bruno, 2001; Thorne and Kishino, 2002)
  - Cox-Ingersoll-Ross Process: the rate of the daughter branch is determined a non-central χ^2 distribution. This process includes a parameter that determines the intensity of the force that drives the process to its stationary distribution (Lepage et al., 2006).
- Uncorrelated rates
  - The rate associated with each branch is drawn from a single underlying parametric distribution such as an exponential or log-normal (Drummond et al., 2006; Rannala and Yang, 2007; Lepage et al., 2007).
- Mixture model on branch rates
  - Branches are assigned to distinct rate categories according to a Dirichlet process (Heath, Holder and Huelsenbeck, 2012).

The variety of models for relaxing the molecular clock assumption presents a challenge for investigators interested in estimating divergence times. Some models assume that rates are heritable and autocorrelated over the tree, others model rate change as a step-wise process, and others assume that the rates on each branch are independently drawn from a single distribution. Furthermore, studies comparing the accuracy (using simulation) or precision of different models have produced conflicting results, some favoring uncorrelated models (Drummond et al., 2006) and others preferring autocorrelated models (Lepage et al., 2007).

Because of this, it is important for researchers performing these analyses to consider and test different relaxed clock models (Lepage et al., 2007; Ronquist et al., 2012; Li and Drummond, 2012; Baele et al., 2013). It is also critical to take into account the scale of the question when estimating divergence times. For example, it might not be reasonable to assume that rates are autocorrelated if the data set includes very distantly related taxa and low taxon sampling. In such cases, it is unlikely that any signal of autocorrelation is detectible.

Priors on node times

There are many component parts that make up a Bayesian analysis of divergence time. One that is often overlooked is the prior on node times, often called a *tree prior*. This model describes how speciation events are distributed over time. When this model is combined with a model for branch rate, Bayesian inference allows you to estimate *relative* divergence times. Furthermore, because the rate and time are confounded in the branch-length parameter, the prior describing the branching times can have a strong effect on divergence time estimation.

We can separate the priors on node ages into different categories:

- **Phenomenological**—models that make no explicit assumptions about the biological processes that generated the tree. These priors are conditional on the age of the root.
  - Uniform distribution: This simple model assumes that internal nodes are uniformly distributed between the root and tip nodes (Lepage et al., 2007; Ronquist et al., 2012).
  - Dirichlet distribution: A flat Dirichlet distribution describes the placement of internal nodes on every path between the root and tips (Kishino, Thorne and Bruno, 2001; Thorne and Kishino, 2002).

- **Mechanistic**–models that describe the biological processes responsible for generating the pattern of lineage divergences.
  - Population-level processes—models describing demographic processes (suitable for describing differences among individuals in the same species/population)
    - Coalescent—These demographic models describe the time, in generations, between coalescent events and allow for the estimation of population-level parameters (Kingman, 1982a;b;c; Griffiths and Tavare, 1994).
  - Species-level processes—stochastic branching models that describe lineage diversification (suitable for describing the timing of divergences between samples from different species)
    - Yule (pure-birth) process: The simplest branching model assumes that, at any given point in time, every living lineage can speciate at the same rate, λ. Because the speciation rate is constant through time, there is an exponential waiting time between speciation events Yule (1924); Aldous (2001). The Yule model does not allow for extinction.
    - Birth-death process: An extension of the Yule process, the birth-death model assumes that at any point in time every lineage an undergo speciation at rate λ or go extinct at rate μ (Kendall, 1948; Thompson, 1975; Nee, May and Harvey, 1994; Rannala and Yang, 1996; Yang and Rannala, 1997; Popovic, 2004; Aldous and Popovic, 2005; Gernhard, 2008). Thus, the Yule process is a special case of the birth-death process where $\mu = 0$.

In BEAST, the available tree priors for divergence time estimation using inter-species sequences are variants of the birth-death prior. Extensions of the birth-death model include the calibrated Yule (Heled and

Drummond, 2012), the birth-death model with incomplete species sampling (Rannala and Yang, 1996; Yang and Rannala, 1997; Stadler, 2009), and serially-sampled birth-death processes (Stadler, 2010). Other programs also offer speciation priors as well as some alternative priors such as a uniform prior (*PhyloBayes*, *MrBayes v3.2*, *DPPDiv*), a Dirichlet prior (*multidivtime*), and a birth-death prior with species sampling (*MCMCTree*).

Tree priors based on the coalescent which are intended for population-level analyses or time-stamped virus data are also available in BEAST. The effect of different node-time priors on estimates of divergence times is not well understood and appears to be dataset-dependent (Lepage et al., 2007). Accordingly, it is important to account for the characteristics of your data when choosing a tree prior. If you know that your sequences are from extant species, each from different genera, then it is unlikely that a coalescent model adequately reflects the processes that generated those sequences. And since you do not have any samples from lineages in the past, then you should not use the serial-sampled birth-death model. Furthermore, if you have prior knowledge that extinction has occurred, then a pure-birth (Yule) prior is not appropriate.

Calibration to absolute time

Without external information to calibrate the tree, divergence time estimation methods can only reliably provide estimates of relative divergence times and not absolute node ages. In the absence of adequate calibration data, relative divergence times are suitable for analyses of rates of continuous trait evolution or understanding relative rates of diversification. However, for some problems, such as those that seek to uncover correlation between biogeographical events and lineage diversification, an absolute time scale is required. Calibration information can come from a variety of sources including "known" substitution rates (often secondary calibrations estimated from a previous study), dated tip sequences from serially sampled data (typically time-stamped virus data), or geological date estimates (fossils or biogeographical data).

Age estimates from fossil organisms are the most common form of divergence time calibration information. These data are used as age constraints on their putative ancestral nodes. There are numerous difficulties with incorporating node age estimates from fossil data including disparity in fossilization and sampling, uncertainty in dating, and correct phylogenetic placement of the fossil. Thus, it is critical that careful attention is paid to the paleontological data included in phylogenetic divergence time analyses. With an accurately dated and identified fossil in hand, further consideration is required to determine how to apply the node-age constraint. If the fossil is truly a descendant of the node it calibrates, then it provides a reliable minimum age bound on the ancestral node time. However, maximum bounds are far more difficult to come by. Bayesian methods provide a way to account for uncertainty in fossil calibrations. Prior distributions reflecting our knowledge (or lack thereof) of the amount of elapsed time from the ancestral node to is calibrating fossil are easily incorporated into these methods.

A nice review paper by Ho and Phillips (2009) outlines a number of different parametric distributions appropriate for use as priors on calibrated nodes. In this exercise we will use the uniform, normal, log-normal, and exponential distributions (Figure 2).

Uniform distribution – Typically, you must have both maximum and minimum age bounds when applying a uniform calibration prior (though some methods are available for applying uniform constraints with soft bounds). The minimum bound is provided by the fossil member of the clade. The maximum bound may come from a bracketing method or other external source. This distribution places equal probability across all ages spanning the interval between the lower and upper bounds.

Normal distribution – The normal distribution is not always appropriate for calibrating a node using fossil

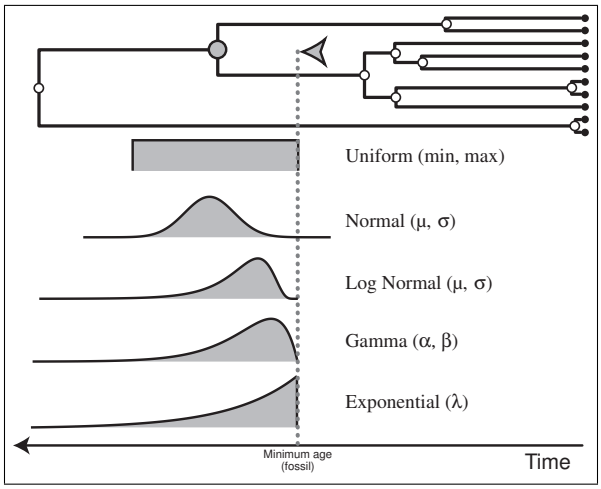

Figure 2: Five different parametric distributions that can be applied as priors on the age of a calibrated node. (figure adapted from Heath, 2012)

information (though some methods allow for assigning a truncated normal prior density). When applying a biogeographical date (e.g. the Isthmus of Panama) or a secondary calibration (a node age estimate from a previous study), the normal distribution can be a useful calibration prior. This distribution is always symmetrical and places the greatest prior weight on the mean (μ). Its scale is determined by the standard deviation parameter (σ).

Probability distributions restricted to the interval $[0, \infty)$, such as the log-normal, exponential, and gamma are appropriate for use as zero-offset calibration priors. When applying these priors on node age, the fossil age is the origin of the prior distribution. Thus, it is useful to consider the fact that the prior is modeling the amount of time that has elapsed since the divergence event (ancestral node) until the time of the descendant fossil (Figure 3).

Gamma distribution – The gamma distribution is commonly used as a prior on scalar variables in Bayesian inference. It relies on 2 parameters: the scale parameter (α) and a rate parameter (λ). More specifically, the gamma distribution is the sum of α independently and identical exponentially distributed random variables with rate λ. As α becomes very large ($\alpha > 10$), this distribution approaches the normal distribution.

Exponential distribution – The exponential distribution is a special case of the gamma distribution and is characterized by a single rate parameter (λ) and is useful for calibration if the fossil age is very close to the age of its ancestral node. The expected (mean) age difference under this distribution is equal to λ^{-1} and the median is equal to $\lambda^{-1} * ln(2)$. Under the exponential distribution, the greatest prior weight is placed on node ages very close to the age of the fossil with diminishing probability to ∞. As λ is increased, this prior density becomes strongly informative, whereas very low values of λ result in a fairly non-informative prior (Figure 3a).

Log-normal distribution – An offset, log-normal prior on the calibrated node age places the highest probability on ages somewhat older than the fossil, with non-zero probability to ∞. If a variable is log-normally distributed with parameters μ and σ, then the natural log of that variable is normally distributed with a mean of μ and standard deviation of σ. The median of the lognormal distribution is equal to e^{μ} and the mean is equal to $e^{\mu+\frac{\sigma^2}{2}}$ (Figure 3b).

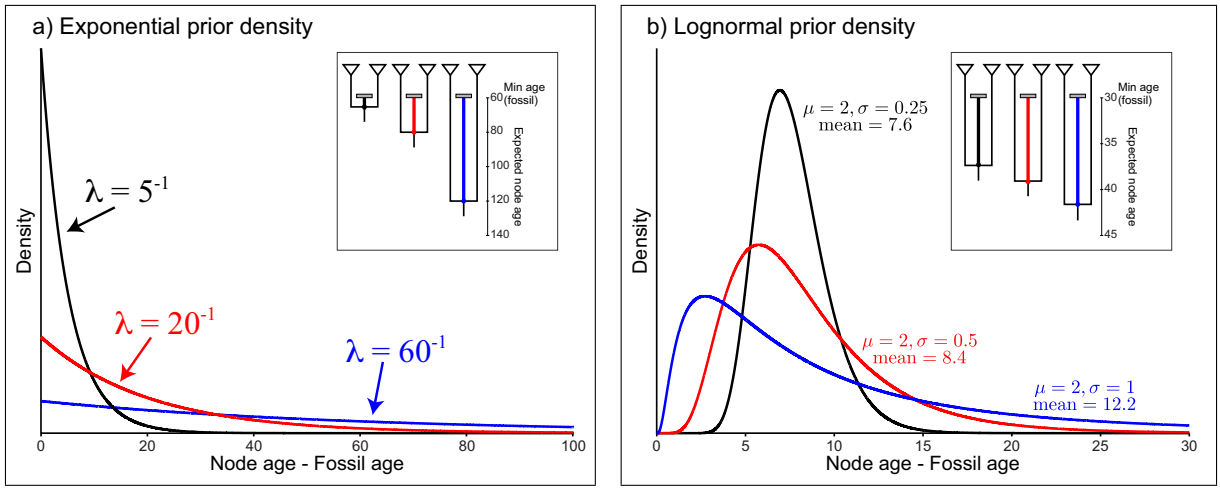

Figure 3: Two common prior densities for calibrating node ages. a) The exponential distribution with three different values for the rate parameter, λ. As the value of the λ rate parameter is decreased, the prior becomes less informative (the blue line is the least informative prior, $\lambda = 60^{-1}$). The inset shows an example of the three different priors and their expected values placed on the same node with a minimum age bound of 60. b) The lognormal distribution with 3 different values for the shape parameter, σ. For this distribution, even though μ is equal to 2.0 for all three, expected value (mean) is dependent on the value of σ. The inset shows an example of the three different priors and their expected values placed on the same node with a minimum age bound of 30.

Integrating Fossil Occurrence Times in the Speciation Model

Calibrating Bayesian divergence-time estimates using parametric densities is ultimately a difficult and unsatisfactory approach, particularly if the calibration information comes from fossil occurrence times. The calibration densities are typically applied in a multiplicative manner such that the prior probability of a calibrated node age is the product o the probability coming from the tree-wide speciation model and the probability under the calibration density (Heled and Drummond, 2012). This approach leads to an incoherence and induces a prior that is inconsistent with the described calibration density. This statistical incoherence has been corrected by conditional tree prior models (Yang and Rannala, 2006; Heled and Drummond, 2012; 2013). These conditional approaches are an important contribution to the field, particularly when non-fossil data are used to calibrate an analysis. However, when using fossil information, it is more appropriate to account for the fact that the fossils are part of the same diversification process (*i.e.*, birth-death model) that generated the extant species. Thus, new models that jointly estimate the tree of extant and fossil species and divergence times (Lee, Oliver and Hutchinson, 2009; Pyron, 2011; Ronquist et al., 2012) and integrate the fossil observations into the diversification model (Stadler, 2010; Heath, Huelsenbeck and Stadler, 2014; Gavryushkina et al., 2014) are an important direction for estimating speciation times on an absolute timescale.

Programs used in this exercise

BEAST – Bayesian Evolutionary Analysis Sampling Trees

BEAST 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 (Drummond et al., 2006; Drummond and Rambaut, 2007; Bouckaert et al., 2014). The development and maintenance of BEAST is a large, collaborative effort and the program includes a wide array of different types of analyses:

- Phylogenetic tree inference under different models for substitution rate variation
  - Constant rate molecular clock (Zuckerkandl and Pauling, 1962)
  - Uncorrelated relaxed clocks (Drummond et al., 2006)
  - Random local molecular clocks (Drummond and Suchard, 2010)
- Estimates of species divergence dates and fossil calibration under a wide range of branch-time models and calibration methods
- Analysis of non-contemporaneous sequences
- Heterogenous substitution models across data partitions
- Population genetic analyses
  - Estimation of demographic parameters (population sizes, growth/decline, migration)
  - Bayesian skyline plots
  - Phylogeography (Lemey et al., 2009)
- Gene-tree/species-tree inference (*BEAST; Heled and Drummond, 2010)
- and more...

BEAST is written in java and its appearance and functionality are consistent across platforms. Inference using MCMC is done using the BEAST program, however, there are several utility applications that assist in the preparation of input files and summarize output (BEAUti, LogCombiner, and TreeAnnotator are all part of the BEAST software bundle).

There are currently two available versions of the BEAST package:

BEAST v1.8 http://beast.bio.ed.ac.uk (BEAST 1)

BEAST v2.1.2 http://www.beast2.org/ (BEAST 2).

BEAST 2 is a complete re-write of BEAST 1, with different design choices (Bouckaert et al., 2014). The BEAST 2 package allows for implementation and distribution of new models and methods through *add-ons* (also called "plugins"). Add-ons include SNAPP (phylogenetic analysis using SNP and AFLP data) and BDSSM (a birth-death skyline model for serially-sampled data), as well as several others that are available or in development. It is important to note, however, that the set of analyses and models available in the BEAST 2 package do not completely overlap with the set of analyses with BEAST 1 (though this should not be the case in the near future). I strongly encourage you to learn more about BEAST and the BEAST v2 software by reading the book provided online by the developers: *Bayesian Evolutionary Analysis with BEAST 2* (Drummond and Bouckaert, 2014).

BEAUti – Bayesian Evolutionary Analysis Utility

BEAUti is a utility program with a graphical user interface for creating BEAST and *BEAST input files which must be written in the eXtensible Markup Language (XML). This application provides a clear way to specify priors, partition data, calibrate internal nodes, etc.

LogCombiner – When multiple (identical) analyses are run using BEAST (or MrBayes), LogCombiner can be used to combine the parameter log files or tree files into a single file that can then be summarized using Tracer (log files) or TreeAnnotator (tree files). However, it is important to ensure that all analyses reached convergence and sampled the same stationary distribution before combining the parameter files.

TreeAnnotator – TreeAnnotator is used to summarize the posterior sample of trees to produce a maximum clade credibility tree and summarize the posterior estimates of other parameters that can be easily visualized on the tree (e.g. node height). This program is also useful for comparing a specific tree topology and branching times to the set of trees sampled in the MCMC analysis.

Tracer – Tracer is used for assessing and summarizing the posterior estimates of the various parameters sampled by the Markov Chain. This program can be used for visual inspection and assessment of convergence and it also calculates 95% credible intervals (which approximate the 95% highest posterior density intervals) and effective sample sizes (ESS) of parameters (http://tree.bio.ed.ac.uk/software/tracer).

FigTree – FigTree is an excellent 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) in a visually appealing way (http://tree.bio.ed.ac.uk/software/figtree).

The eXtensible Markup Language

The eXtensible Markup Language (XML) is a general-purpose markup language, which allows for the combination of text and additional information. In BEAST, the use of the XML makes analysis specification very flexible and readable by both the program and people. The XML file specifies sequences, node calibrations, models, priors, output file names, etc. BEAUti is a useful tool for creating an XML file for many BEAST analyses. However, typically, dataset-specific issues can arise and some understanding of the BEAST-specific XML format is essential for troubleshooting. Additionally, there are a number of interesting models and analyses available in BEAST that cannot be specified using the BEAUti utility program. Refer to the BEAST web page (http://beast.bio.ed.ac.uk/XML_format) for detailed information about the BEAST XML format. Box 1 shows an example of BEAST XML syntax for specifying a birth-death prior on node times.

```
<!- An exponential prior distribution on the gamma shape parameter of gene1                    ->
<prior id="GammaShapePrior.s:gene1" name="distribution" x="@gammaShape.s:gene1">
    <Exponential id="Exponential.0" name="distr">
        <parameter id="RealParameter.0" lower="0.0" name="mean" upper="0.0">1.0</parameter>
    </Exponential>
</prior>
```

Box 1: BEAST 2 XML specification of an exponential prior density on the shape of a gamma distribution.

Practical: Divergence time estimation

For this exercise, we will estimate phylogenetic relationships and date the species divergences of the ten simulated sequences in the file called `divtime.nex`. This simple alignment contains two genes, each 500 nucleotides in length.

- Download all of the compressed directories and place them in a directory you've created named: `divtime_beast`. After uncompressing, you should have the files listed in Box 2. (Tutorial url: http://treethinkers.org/divergence-time-estimation-using-beast/)

```
• divtime_beast/data/
    – divtime.nex
    – xml_files/
        * divtime_100m.1.xml
        * divtime_100m.2.xml
        * divtime.xml

• divtime_beast/output1/
    – divtime_100m.1.log
    – divtime_100m.2.log
    – divtime_100m.prior.log

• divtime_beast/output2/
    – divtime_100m.1.trees
    – divtime_100m.2.trees
    – divtime.comb.trees
```

Box 2: The data files required for this exercise.

- Open the NEXUS file containing the sequences in your text editor. The `ASSUMPTIONS` block contains the commands for partitioning the alignment into two separate genes (Box 3). Tests for model selection indicated that `gene1` and `gene2` evolved under separate GTR+Γ models.

```
#NEXUS
BEGIN DATA;
    DIMENSIONS NTAX=10 NCHAR=1000;
    FORMAT DATATYPE = DNA GAP = - MISSING = ?;
    MATRIX
    T1   CTACGGGAGGGCAACGGGGCTAGATGGTAAACGCGCCATCGATCGCAAG...
    T2   CTACGGGAGGGCGACGGGGCTAGATGGTAAACGCGCCCTCGATCGCAAG...
    T3   CAGCGTGAGGGCCACGGGGCTGGCAGGTACTCCGGCCCACGAGTGGAAG...
    T4   CAGCGTGGGGGCCACGGGGCTAGAAGTTACTCCGGCCCACGAGTGGAAG...
    T5   CAGCGTGGGGGCCACGGGGCTAGAAGTTACTCCGGCCCACGAGTGGAAG...
    T6   CAGCGAGAAGCCGACGGGGATGGAAGGGACTCAGACGCACGAGTCCATG...
    T7   CATCGCGAGGGGGACGGGGCTCGTAGATTATCGTTCATGCAAGCTGAAG...
    T8   CATCGCGAGGGGGACGGGGCTCGTAGTTTATCGTTCAGGCAAGCTGAAG...
    T9   CAGCGTGACCACGACGGGGCTGGGGGTGATTCCCGCTGACAAGATGAAG...
    T10  CTGCGTGACAACGACGGGGCTGGGAGTTGTTCCCGCTCACAAGAGGAAG...
;
END;

BEGIN ASSUMPTIONS;
    charset gene1 = 1-500;
    charset gene2 = 501-1000;
END;
```

Box 3: A fragment of the NEXUS file containing the sequences for this exercise. The data partitions are defined in the `ASSUMPTIONS` block..

These 10 sequences consist of 6 ingroup taxa: `T1, T2, T3, T4, T5, T6` and 4 outgroup taxa: `T7, T8, T9, T10`. After performing an unconstrained analysis using maximum likelihood, we get the topology in Figure

4. The branch lengths estimated under maximum likelihood are indicative of variation in substitution rates. Divergence time estimation and fossil calibration require that you have some prior knowledge of the tree topology so that calibration dates can be properly assigned.

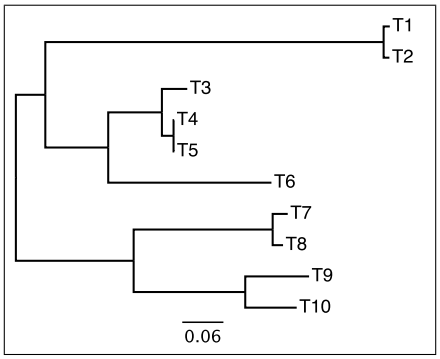

Figure 4: A maximum likelihood estimate of the phylogenetic relationships of `divtime.nex`. (Constructed using PAUP* v4.0a125; Swofford, 1998)

Fossil node calibrations are often difficult to obtain for every node and for many groups they are simply unavailable. In such cases, constraints can be applied to outgroup nodes. There are four calibration points for this data set. These are illustrated in Figure 5. The oldest fossil belonging to the ingroup can calibrate the age of that clade. This fossil was identified as a member of the clade, falling within the *crown* group. Two fossils calibrate nodes within the outgroup clade and a well supported estimate of the root age from a previous study allows us to place a prior distribution on that node (Figure 5).

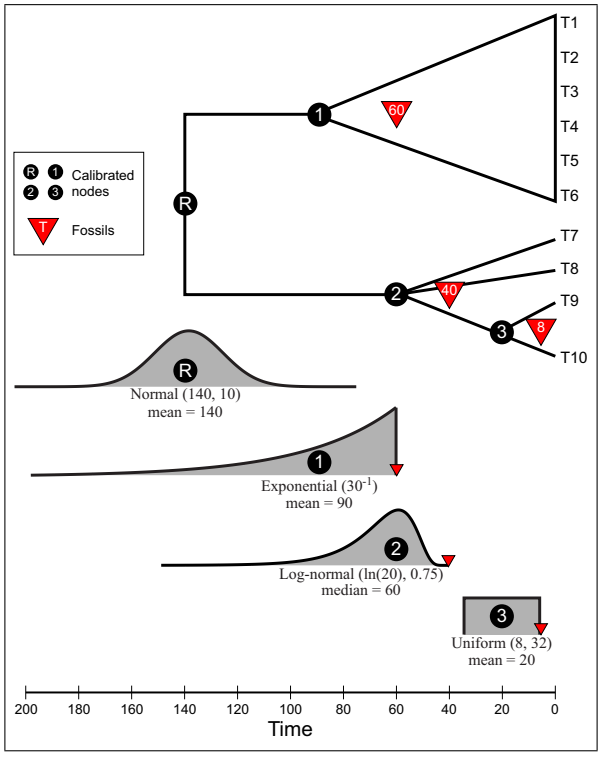

Figure 5: Four nodes with fossil calibration for the `divtime.nex` data set. There are calibrations on the root and 3 other internal nodes. Each calibration point is assumed to have a different prior density.

Getting started with BEAUti

Creating a properly-formatted BEAST XML file from scratch is not a simple task. However, BEAUti provides a simple way to navigate the various elements specific to the BEAST XML format.

- Begin by executing the BEAUti program. For Mac OSX and Windows, you can do this by double clicking on the application. For Unix systems (including Mac OSX), it is convenient to add the entire `BEAST/bin` directory to your path.

- Import the sequences from `divtime.nex` using the pull-down menu: ***File→Import Alignment***.

This example data set contains 2 different partitions labeled `gene1` and `gene2`. When the NEXUS file is imported into BEAUti, the ***Partitions*** tab lists each partition and their currently assumed substitution model, clock model, and tree.

- Double click on the file name (`divtime`) next to one of the data partitions. This will bring up a window allowing you to visually inspect your alignment.

We would like to analyze each gene under separate substitution models, while assuming the clock and tree are linked. By default, BEAUti imports a partitioned alignment with the site models, clock models, and trees unlinked. We would like to link the clock models and trees; and for convenience, we will also link the site models so that we only need to specify the GTR+Γ model once.

- Select both `gene1` and `gene2`. While both partitions are highlighted click the ***Link Site Models*** button. Then click the ***Link Clock Models*** and ***Link Trees*** buttons. You will notice that options for `gene2` have now all changed to `gene1`.

- For the sake of clarity, let's rename the ***Clock Model*** and ***Tree*** that are shared by each gene. Click on the label "`gene1`" in the ***Clock Model*** column. This will create a textbox for you to change the name. Change the ***Clock Model*** name to `divtimeClock`. You will notice when you click on the ***Clock Model*** for `gene2`, this automatically changes the name to `divtimeClock`.

- Similarly, change the ***Tree*** name to `divtimeTree`. [Figure 6]

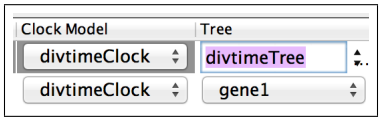

Figure 6: Rename the clock model and tree.

Now your ***Partitions*** tab should show that the two genes are assumed to have a single substitution model called `gene1`, a single clock model called `divtimeClock`, and a single tree called `divtimeTree`. [Figure 7]

For this dataset, we assume that each gene evolved under separate GTR+Γ models. We will specify the site model while the partitions are linked sot hat we do not have to perform the actions multiple times.

- Go to the ***Site Model*** tab.

- Specify a ***GTR*** model in the options for the ***Subst Model***. This initializes all of the exchangeability rates and base frequencies of the GTR model.

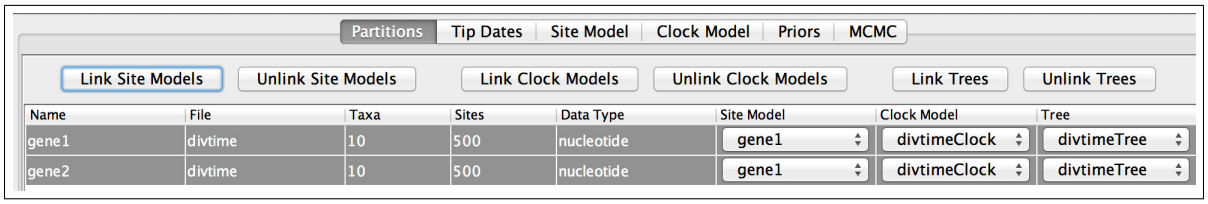

Figure 7: The data partitions with all *linked* models.

- To set up the gamma-distributed rates, change the ***Gamma Category Count*** to `4`. When you enter this value, a new parameter becomes available. This is the shape of the mean-one gamma distribution on site rates. Since we want to put a prior on this, check the ***estimate*** box for the ***Shape*** parameter.

- Now check ***estimate*** for the ***Substitution Rate*** and mark the box next to ***Fix mean substitution rate***. This will allow for estimation of the relative rates for each partition once the genes are unlinked. [Figure 8]

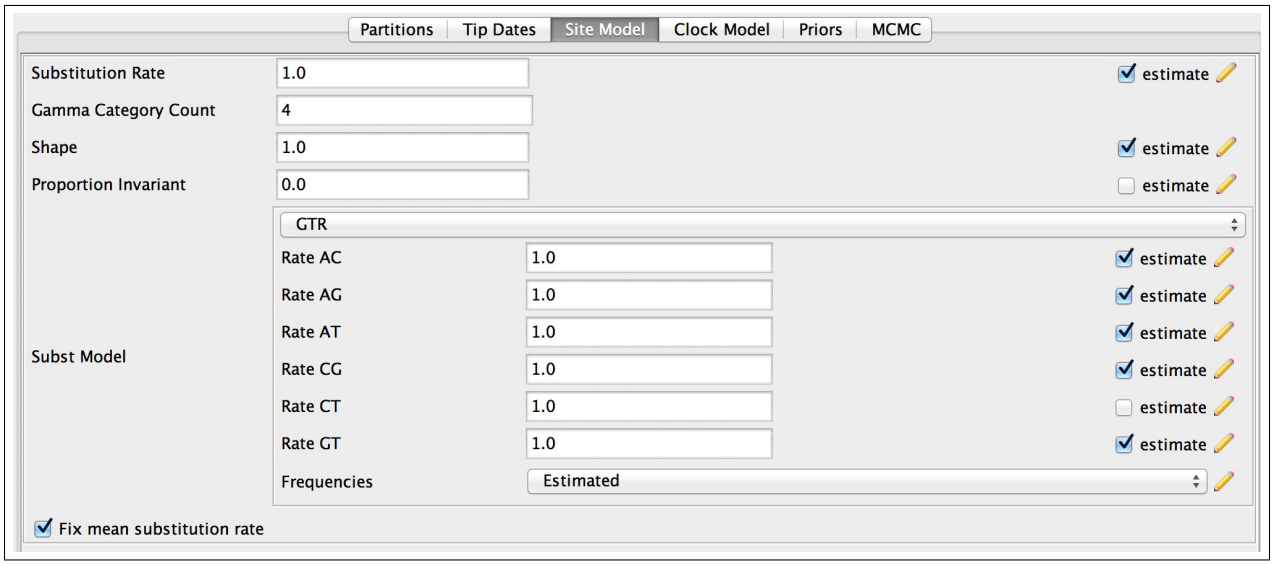

Figure 8: The ***Site Model*** panel with a GTR+Γ model.

Since we are assuming that the two genes evolved under independent GTR+Γ models, we must now unlink the site models for our partitions.

- Return to the ***Partitions*** panel.

- Highlight both of the data partitions and click the ***Unlink Site Models*** button. Now, under the ***Site Model*** column, each gene has a different model name. [Figure 9]

- If you navigate back to the ***Site Model*** pane, you should have a GTR+Γ model specified for `gene1` and for `gene2`.

The ***Tip Dates*** menu contains the options necessary for analyses of data sets containing non-contemporaneous tips. This is primarily for serial sampled virus data and not applicable to this exercise. We will skip this menu and move on to specify the ***Clock Model***.

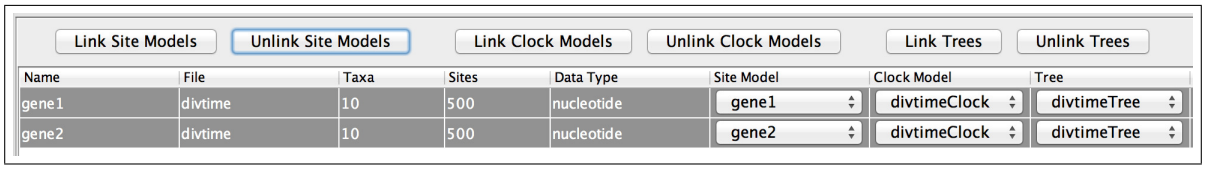

Figure 9: The data partitions with unlinked site models and linked clock model and tree.

- Move on to the ***Clock Model*** menu to set up the relaxed clock analysis.

Here, we can specify the model of lineage-specific substitution rate variation. The default model in BEAUti is the ***Strict Clock*** with a fixed substitution rate equal to `1`. Three models for relaxing the assumption of a constant substitution rate can be specified in BEAUti as well. The ***Relaxed Clock Log Normal*** option assumes that the substitution rates associated with each branch are independently drawn from a single, discretized lognormal distribution (Drummond et al., 2006). Under the ***Relaxed Clock Log Normal*** model, the rates associated with each branch are exponentially distributed (Drummond et al., 2006). The ***Random Local Clock*** uses Bayesian stochastic search variable selection to average over random local molecular clocks (Drummond and Suchard, 2010).

- Change the model from ***Strict Clock*** to ***Relaxed Clock Log Normal***. By default, the ***Clock.rate*** parameter is fixed and the ***estimate*** box is not available as an option.

- To allow for estimation of the ***Clock.rate***, navigate to the ***Mode*** pull-down menu above and select ***Automatic set clock rate***. Then check the box that allows you to ***estimate*** the ***Clock.rate***. [Figure 10]

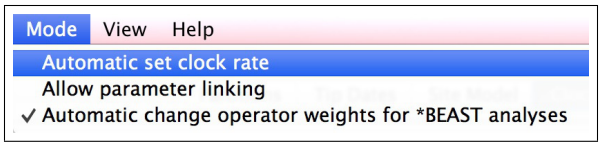

Figure 10: Allow estimation of the clock rate parameter by unchecking this option.

The fully specified ***Clock Model*** assumes that the rates for each branch are drawn independently from a single lognormal distribution. The mean of the rate distribution will be estimated, thus we can account for uncertainty in this parameter by placing a prior distribution on its value. Note that there is an option to ***Normalize*** the average clock rate. We will leave this unchecked. [Figure 11]

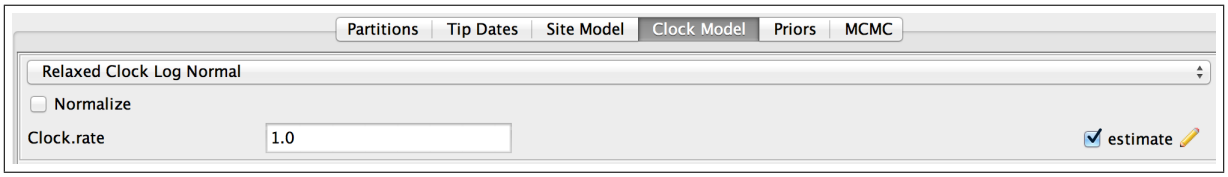

Figure 11: The ***Clock Model*** panel with an uncorrelated, log normal relaxed clock model specified.

- Move to the ***Priors*** menu panel.

In the ***Priors*** window, all of the parameters and hyperparameters (and hyper-*hyper*parameters, etc.) specific to the models defined in the ***Site Model*** and ***Clock Model*** windows are listed. Here you can set up

the prior distribution on these parameters, as well as define calibration nodes and calibration densities and specify a tree model. One convenient feature of BEAUti is that the list of parameters changes dynamically as you change the models. Thus, if you missed a step along the way, you would notice at this point because something might be missing here. For example, if you did not unlink the site models for each gene in the ***Partitions*** window, then you would only see the exchangeability rates, gamma shape, and base frequency parameters listed for `gene1`. [Figure 14]

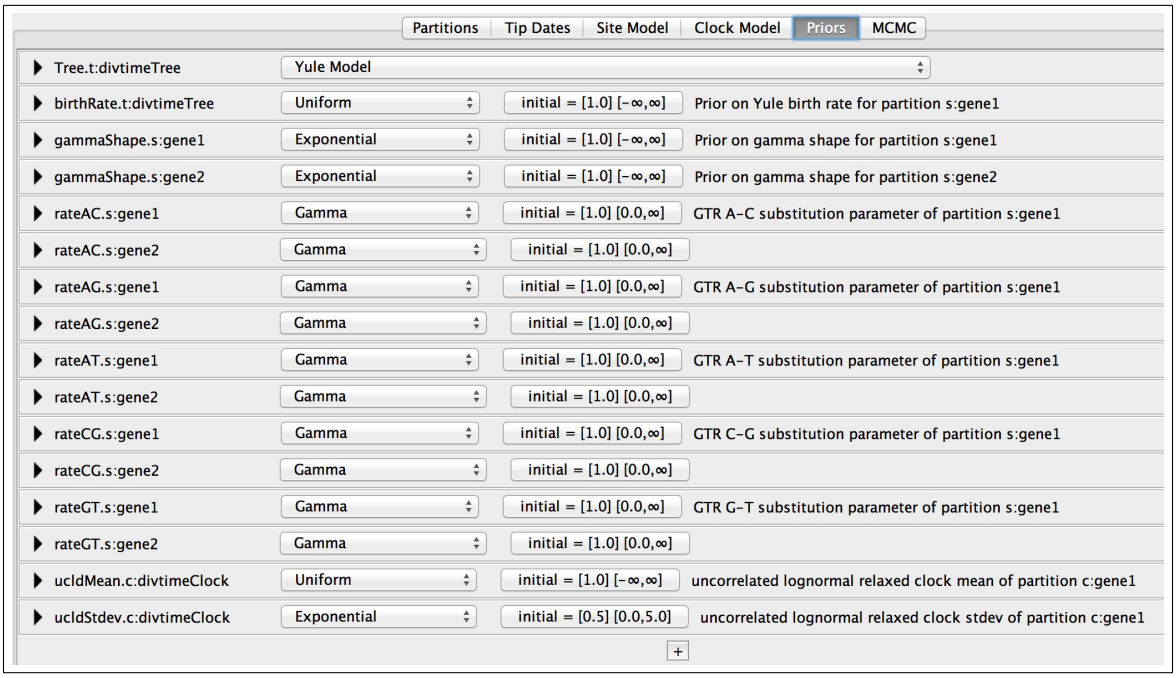

Figure 12: The parameters and hyperparameters specific to this analysis and default priors.

Let's start by specifying the tree model, which will define the prior distribution on tree topologies and node ages.

- Change the ***Tree.t:divtimeTree*** option to the ***Birth Death Model***.

This model and the ***Yule Model*** and ***Calibrated Yule Model*** are appropriate for analyses of inter-species relationships. These speciation models are stochastic branching processes with a constant rate of speciation (λ) and a constant rate of extinction (μ). In the case of the Yule models, the extinction rate is equal to 0. The ***Calibrated Yule Model*** defines a specific parameterization of the pure-birth model that conditions on a single calibration density (Heled and Drummond, 2012). An important extension of the calibrated Yule model was recently developed (Heled and Drummond, 2013), conditioning the birth-death model on multiple calibration nodes. This calibrated birth-death model is available in BEAST2, but not through the BEAUti interface. For this exercise, however, we will use the reconstructed birth-death process.

Previous versions of BEAUti displayed the relevant citation for each of the different tree priors, once they were selected (now this is displayed in the BEAST screen output). Both the Yule and birth-death models are implemented in BEAST following Gernhard (2008) with hyper-prior distributions on the parameters of the model. Thus, for the birth-death model used in this analysis, our runs will sample the net diversification rate ($\lambda - \mu$) and the relative rate of extinction or "turnover" ($\frac{\mu}{\lambda}$).

You will notice several other options for the tree prior available in BEAUti. The coalescent tree priors are appropriate for population-level analyses. Conversely, when you are estimating relationships and divergence

times of interspecies data, it is best to employ a speciation prior. Choosing a coalescent prior for estimating deep divergences, or a speciation prior for intra-species data, can often lead to problematic results due to interactions between the prior on the node ages and the prior on the branch rates. Thus, it is critical that these priors are chosen judiciously. Furthermore, it is important to note that our understanding of the statistical properties of speciation prior densities combined with calibration densities is somewhat incomplete, particularly when the tree topology is considered a random variable (Heled and Drummond, 2012; Warnock, Yang and Donoghue, 2012).

- Now specify the options of the ***Birth Death Model*** by clicking on the ▶ to the left of the ***Tree.t:divtimeTree***. This will reveal the hyperparameters of the model and various options.

- Leave the ***Type*** set to ***unscaled***. The other options here define the type of tree on which this model acts.

- We will also be sure to ***estimate*** both of the birth-death process hyperparameters. These are the ***Relative Death Rate*** ($\frac{\mu}{\lambda}$) and the ***Birth Diff Rate*** (this is the diversification rate, $\lambda - \mu$). The text boxes next to each parameter allow you to specify a starting value. If you click on the ✎ symbol next to ***estimate***, you alter some options for each parameter, though leave these unchanged for this exercise.

The probability of the reconstructed birth-death process can be written down in different ways by conditioning on different components of the tree (Gernhard, 2008). Often, the probability of a given tree and times is conditioned on the time of origin of the tree. The time of origin is different from the root age and indicates the origin of the lineage that diverged at the root of the tree. This concept is often not intuitive for most analyses of species-level relationships and we would prefer to condition on the root age.

- Select the option to parameterize this model such that the probability is ***Conditional On Root***.

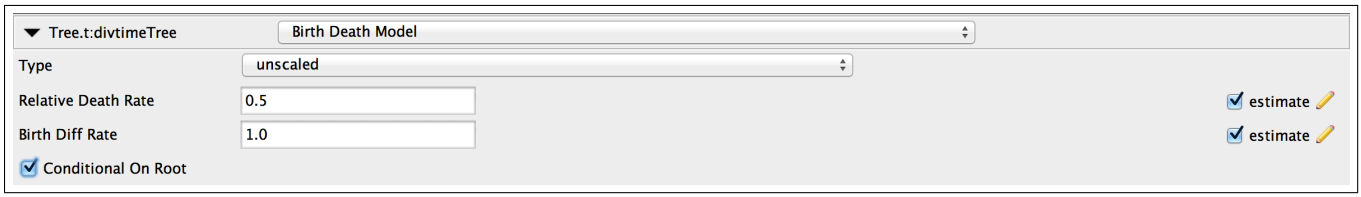

Figure 13: The specification of the ***Birth Death Model***.

After selecting the ***Birth Death Model***, you may have noticed that a new parameter, appeared in the list: ***relativeDeathRate2.t***. This parameter along with ***birthRate2.t*** are the hyperparameters of the birth-death process, which are given uniform distributions by default. In Bayesian terminology these are called *hyperparameters* because they are parameters describing a prior distribution and not direct parameters of the data model (like base frequencies or branch lengths). In Bayesian inference a prior distribution can be placed on a hyperparameter, and this is called a *hyperprior*. By allowing the value of these hyperparameters to vary, we are freed from the responsibility of fixing the diversification rate and relative extinction rates of the birth-death model. Additionally, the Markov chain will sample both hyperparameters along with the other parameters directly associated with the models on our data, providing us with an estimate of their posterior distributions.

- Reveal the options for the prior on the ***relativeDeathRate2.t*** by clicking the ▶ to the left and verify that the prior density is a uniform distribution between 0 and 1.

This parameter will always be on the interval [0,1) because it is assumed that $\mu < \lambda$. Thus a prior distribution bounded on this interval is appropriate. Notice that this the bounds of the uniform distribution can be altered and that BEAUti provides a visualization of the prior density, which isn't very interesting for this parameter. But, if we had prior knowledge that the rate of extinction was very high, then we could choose a different prior density, such as a Beta(1, 4). For this exercise, however, assume uniform prior densities for both birth-death parameters.

- Inspect the default priors on the parameters of the substitution models on each gene. You will notice that the ***estimate*** box might be checked for the hyperparameters of the prior distributions on the ***gammaShape.s*** and each of the exchangeability rates (*e.g.*, ***rateAC.s***). It isn't entirely clear why these are checked. If you do not change this, BEAUti creates parameters in the XML file, but does not instantiate operators to change those values. If you do uncheck the ***estimate*** boxes for these, the resulting analysis will be the same as if you never touched them. *However*, if you uncheck them, then recheck them, new hyperparameters appear in the list in the ***Priors*** menu. So do not change these or uncheck them all, because we do not wish to place additional priors on these parameters.

- Leave the priors on the substitution model parameters set to their default values.

Now we will consider the hyperparameters of the uncorrelated lognormal relaxed clock model. Since we are assuming that the branch rates are drawn from a lognormal distribution, this induces two hyperparameters: the mean and standard deviation (***ucldMean.c*** and ***ucldStdev.c*** respectively). By default, the prior distribution on the ***ucldMean.c*** parameter is an improper, uniform distribution on the interval $(-\infty, \infty)$. Now, this may seem odd to you because the substitution rate cannot be less than 0, and leaving the default prior may be problematic. Thus, it is important to specify a prior that reflects our understanding of the value of this parameter. Note that this type of prior is called improper because the prior density of a *uniform* distribution with infinite bounds does not integrate to 1. Although improper priors can sometimes lead to proper posterior distributions, they may also have undesired effects and cause problems with mixing and convergence.

- Reveal the options for the prior on ***ucldMean.c*** by clicking on the ▶. Change the prior density to an ***Exponential*** with a mean of `10.0`.

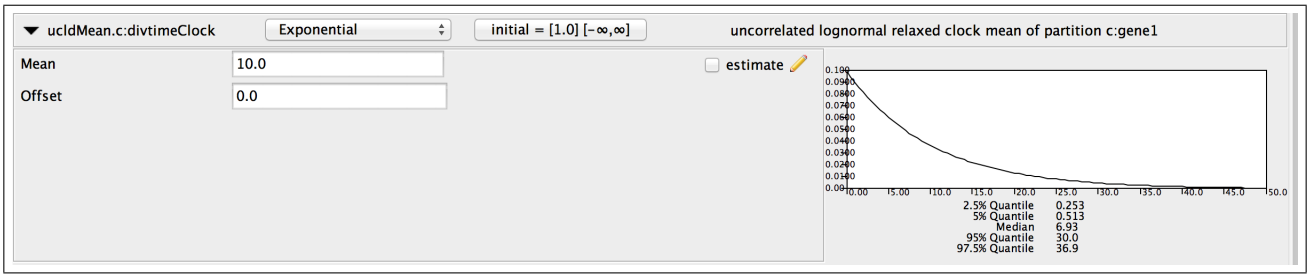

Figure 14: The exponential prior distribution on the mean of the log normal relaxed clock model.

For this exercise, we have four different calibration ages for four nodes in our tree (Figure 5). The calibration nodes are defined in the ***Priors*** panel in this version of BEAUti. The process for revealing a new ***Taxon set editor*** is not exactly obvious.

- Create a new taxon set, which defines a calibration node by clicking on the [+] at the bottom of the list of parameters in the ***Priors*** panel. This will bring up a ***Taxon set editor*** for you to define the terminal taxa that belong in the calibration node.

- Begin by defining the root node. In the ***Taxon set editor*** specify the ***Taxon set label*** for this node: `rootAge`. Then, highlight *all* of the taxon names in the right-hand box and move them to the left-hand box using the >> button. Then, click ***OK***.

This will create a new parameter in the list called ***rootAge*** (sometimes BEAUti adds on ***.prior***). Notice that you can click on the button labeled ***rootAge*** and reveal the ***Taxon set editor*** defining this node.

- Reveal the options for the ***rootAge*** prior. Currently, the type of prior is set to ***[none]***, so there are no available options.

- Change the prior density to a normal distribution and specify the values based on our prior uncertainty in the root age found in Figure 5. To do this, set the ***Mean*** to `140` and ***Sigma*** to `10`. And for the sake of completeness (although it should be unnecessary), check the box labeled ***monophyletic*** to indicate that the root is monophyletic. [Figure 15].

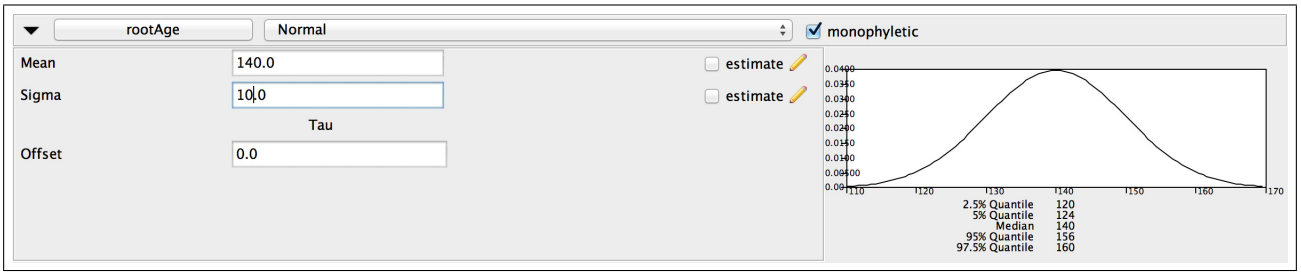

Figure 15: Assume a normal prior distribution on the age of the root node.

The normal distribution is not always appropriate for representing fossil age constraints, but is useful for imposing a prior requiring soft minimum and maximum bounds. Typically, this type of prior on the calibrated node age is based on biogeographical data or when a secondary calibration date is used.

- Create a new taxon set for calibration node 1 by clicking the + at the bottom of the parameter list. Change the ***Taxon set label*** to `mrca1`. These taxa form our "ingroup" clade, and thus are assumed monophyletic. Then, select the taxa descended from calibration node 1 (`T1, T2, T3, T4, T5, T6`). When you have highlighted each of these taxon names, move them from the left column to the right column by clicking the >> button. Then click ***OK***. [Figure 16]

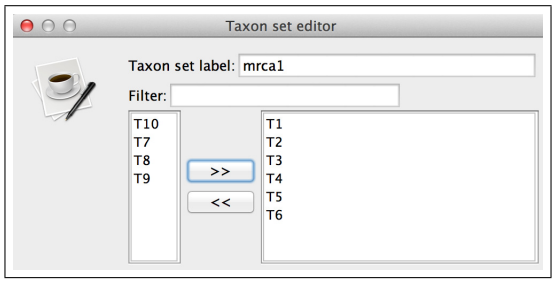

Figure 16: Create a taxon set for node 1.

Note that when dealing with very large trees, the ***Filter*** text box provides an easy way to define calibration nodes. This works only if the labels of the taxa you wish to select have character strings in common (*e.g.*, a genus name).

A fossil age constraint is available for the most recent common ancestor of our ingroup, `mrca1` (`T1, T2, T3, T4, T5, T6`). This minimum age estimate will serve as a hard lower bound on the age that node. Thus, the prior distribution on this node will be offset by the age of the fossil (`60`). When applying off-set priors on nodes, it is perhaps easiest to consider the fact that the distribution is modeling the time difference between the age of the calibrated node and the age of the fossil (see Figure 3). We are applying an exponential prior on the age of `mrca1`. The exponential distribution is characterized by a single rate parameter (λ), which controls both the mean and variance. The expected (mean) age difference under this distribution is equal to λ^{-1} and the median is equal to $\lambda^{-1} * ln(2)$. Specifying the exponential prior on a node age in BEAST requires that you set the *expected* age difference between the node and fossil (***Mean***) and the hard lower bound (***Offset***).

- Reveal the options for the prior on ***mrca1*** and specify that these taxa are ***monophyletic***. Change the prior distribution to an ***Exponential*** with a ***Mean*** equal to `30` and ***Offset*** equal to `60`. The quantiles for this prior are provided below the graph of the prior density. And you can see that with an offset of 60 My, 97.5% of the prior probability will be below 170.67 My. [Figure 17]

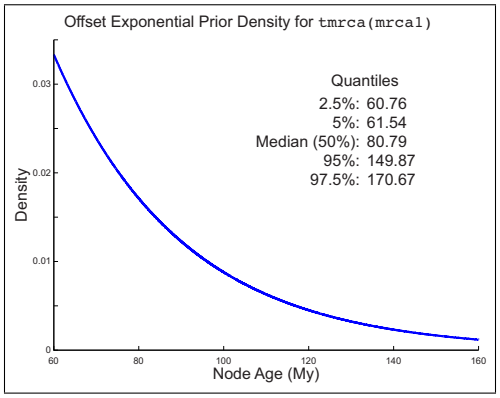

Figure 17: An offset exponential prior on the calibration node `mrca1`.

Perhaps one of the biggest challenges for most biologists conducting time-calibrated divergence time analyses is the specification of calibration density parameters. Importantly, the parameterization of these priors strongly influence the estimates of node ages because the sequence data are not informative on an absolute time scale. A Bayesian would simply put a prior on it and account for her uncertainty in the value of the calibration-density hyperparameter (Heath, 2012). This approach is now easily done in the current version of BEAUti.

- Check the ***estimate*** box next to the ***Mean*** parameter of the exponential calibration density on ***mrca1***. A new parameter called: ***parameter.hyperExponential-mean-mrca1*** should appear in the list, alphabetically.

- Reveal the options for the prior on the ***parameter.hyperExponential-mean-mrca1*** hyperparameter.

Keep in mind that we are specifying a prior on the *expected age difference* between `mrca1` and the fossil occurrence time (we will denote this $\mathcal{D}_{mrca1}$). We want to construct a prior that will give us an expected value of `30`, $\mathbb{E}(\mathcal{D}_{mrca1}) = 30$, with the variance of that distribution describing our uncertainty in the value of the parameter. Furthermore, the distribution must be on positive, real-valued quantities.

- Change the distribution on the ***parameter.hyperExponential-mean-mrca1*** hyperparameter to a ***Gamma*** prior. The parameterization used in BEAUti specifies a shape parameter (***Alpha*** or α) and a scale parameter (***Beta*** or β), where $\mathbb{E}(\mathcal{D}_{mrca1}) = \alpha\beta$. Parameterize the ***Gamma*** distribution with ***Alpha*** = `10` and ***Beta*** = `3`. [Figure 18]

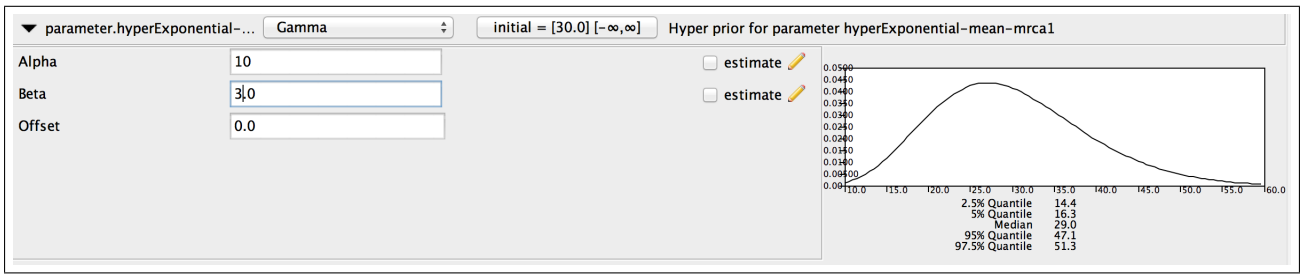

Figure 18: Specify a gamma hyperprior on the *mean* of the exponential calibration density on the age of `mrca1`.

- Create a new taxon set for calibration node 2 (`T7, T8, T9, T10`; Figure 5) and rename it `mrca2`. These taxa make up the "outgroup" and should also be monophyletic.

The fossil calibration for the MRCA of the outgroup (`T7, T8, T9, T10`) provides a minimum age bound for the age of `mrca2`. The log-normal distribution is often used to describe the age of an ancestral node in relation to a fossil descendant. If a random variable, χ, is log-normally distributed with parameters μ and σ: $\chi \sim \text{LN}(\mu, \sigma)$, then the natural log of that variable is normally distributed with a mean of μ and standard deviation of σ: $\log(\chi) \sim \text{Norm}(\mu, \sigma)$. The median of the lognormal distribution is equal to e^{μ} and the mean (or expectation) is:

$$\mathbb{E}(\chi) = e^{\mu + \frac{\sigma^2}{2}}.$$

When applying the lognormal offset prior distribution on node age in BEAST, first consider the expected age difference between the MRCA and the fossil. Generally, it is difficult to know with any certainty the time lag between the speciation event and the appearance of the fossil and, typically, we prefer to specify prior densities that are not overly informative (Heath, 2012).

For the prior density on `mrca2`, we will specify a lognormal prior density with an expected value equal to 20. Thus, if we want the expectation of the lognormal distribution to equal 20, we must determine the value for μ using the equation above solve for μ:

$$\begin{aligned} \mu &= \ln(20) - \frac{0.75^2}{2} \\ &= 2.714482, \end{aligned}$$

where `0.75` is the standard deviation parameter of the lognormal distribution for this particular fossil calibration.

- Reveal the prior options for ***mrca2*** and specify a ***Log Normal*** prior distribution for `mrca2` and set ***M*** to $\mu = 2.714482$. The age of the fossil is `40` time units; use this date to set the ***Offset*** for the lognormal prior distribution. Finally, set the ***S*** to `0.75`, so that 97.5% of the prior probability is below `105.7`. [Figure 19]

Notice that the options for the ***Log Normal*** prior distribution allow you to specify ***Mean in Real Space***. If you choose this option, then the mean value you enter is the expected value of the lognormal distribution.

You will specify the exact same distribution as above if you set the ***M*** parameter to `20` while this option is selected (Figure 19). It is important that you are very careful when specifying these parameters. If, for example, ***Mean in Real Space*** was checked and you provided a value of `2.714482` for ***M***, then your calibration prior density would be overly informative. In the case of the prior on ***mrca2***, this would place 95% of the prior density below 47.03566. Conversely, if you set ***M*** `= 20`, your calibration prior would be extremely diffuse, with a *median* value of 485,165,235!

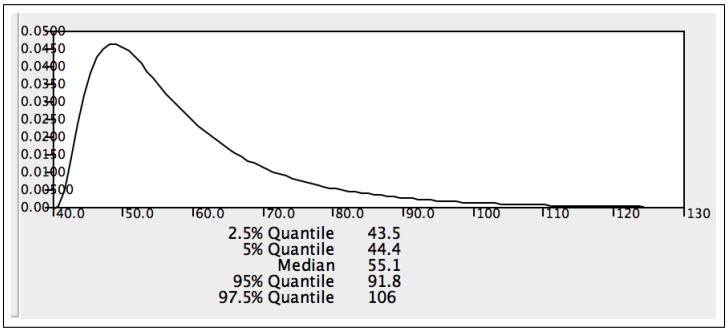

Figure 19: The log normal calibration prior density on the age of `mrca2`.

- Create a new taxon set for node 3 and change the ***Taxon set label*** to `mrca3`. Include only tips `T9` and `T10`. Since we aren't certain that these taxa are monophyletic, we will leave that option unchecked.

- Set the ***Uniform*** prior distribution on the time of `mrca3` with a ***Lower*** limit of `8` and an ***Upper*** limit of `32`. Leave the ***Offset*** set to `0.0`. [Figure 20]

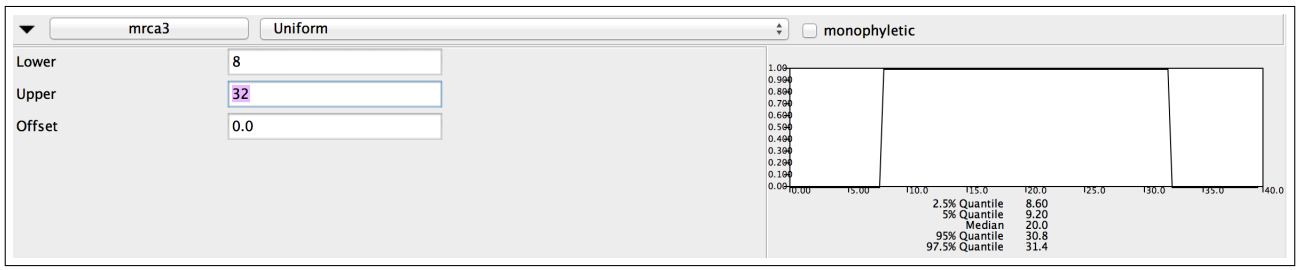

Figure 20: The parameters of the uniform calibration prior on `mrca3`.

The remaining unmodified priors in the ***Priors*** window can be left at their default values for this exercise. All of the parameters that can be manipulated in BEAUti are now set (Figure 21).

There are two additional windows that are hidden in BEAUti by default. You can reveal them by selecting ***View*** $\Rightarrow$ ***View All*** from the pull-down menu above. This will reveal the ***Initialization*** and ***Operators*** panels. The ***Initialization*** options allow you to change the starting values for the various parameters and specify if you want them estimated or fixed. The ***Operators*** menu contains a list of the parameters and hyperparameters that will be sampled over the course of the MCMC run. In this window, it is possible to turn off any of the elements listed to fix a given parameter to its starting value. For example, if you would like to estimate divergence times on a fixed tree topology (using a starting tree that you provided), then disable proposals operating on the `Tree`. For this exercise, leave both windows unmodified.

Now that you have specified all of your data elements, models, priors, and operators, go to the ***MCMC*** tab to set the length of the Markov chain, sample frequency, and file names. By default, BEAST sets the number of generations to 10,000,000.

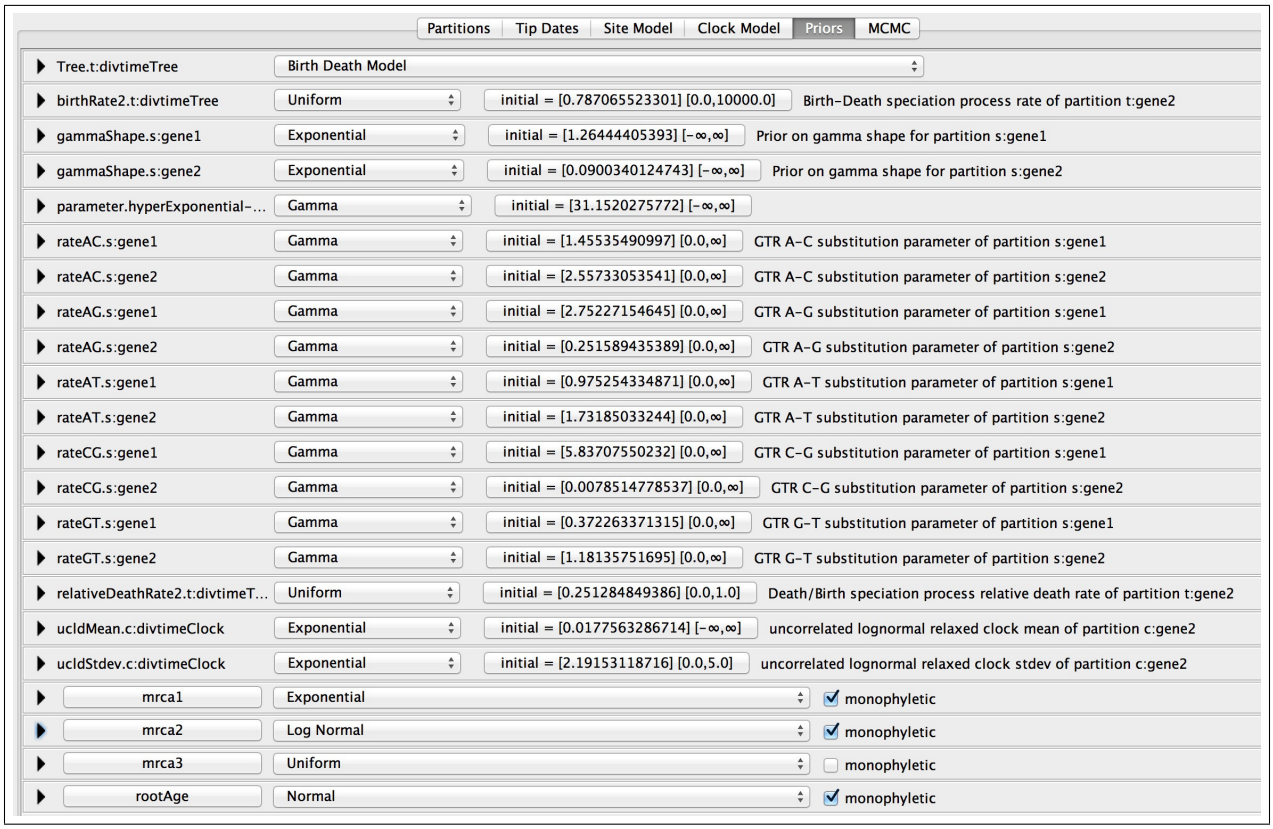

Figure 21: The list of parameters and hyperparameters with fully specified priors.

- Since we have a limited amount of time for this exercise, change the ***Chain Length*** to 1,000,000. (Runtimes may vary depending on your computer, if you have reason to believe that this may take a very long time, then change the run length to something smaller.)

- Reveal the options for the ***tracelog*** using the ▶ to the left.

The frequency parameters are sampled and logged to file can be altered in the box labeled ***Log Every***. In general, this value should be set relative to the length of the chain to avoid generating excessively large output files. If a low value is specified, the output files containing the parameter values and trees will be very large, possibly without gaining much additional information. Conversely, if you specify an exceedingly large sample interval, then you will not get enough information about the posterior distributions of your parameters.

- Change ***Log Every*** to `100`. Also specify a name for the log file by changing ***File Name*** to `divtime.log`.

- Reveal the options for the ***treelog.t:divtimeTree*** file. Change the ***File Name*** to `divtime.trees` and ***Log Every*** to `100`.

Now we are ready to save the XML file!

- In the pull-down menu save the file by going to ***File*** ⇒***Save As*** and save the file: `divtime.xml`.

For the last step in BEAUti, create an XML file that will run the analysis by sampling under the prior.

- Check the box labeled ***Sample From Prior*** at the bottom of the ***MCMC*** panel. We will want to change the names of the output files as well, so change the ***tracelog*** – ***File Name*** to `divtime.prior.log` and the ***treelog.t:divtimeTree*** – ***File Name*** to `divtime.prior.trees`.

- Save these changes by going to ***File*** ⇒***Save As*** and name the file `divtime.prior.xml`.

Making changes in the XML file

BEAUti is a great tool for generating a properly-formatted XML file for many types of BEAST analyses. However, you may encounter errors that require modifying elements in your input file and if you wish to make small to moderate changes to your analysis, altering the input file is far less tedious than generating a new one using BEAUti. Furthermore, BEAST is a rich program and all of the types of analyses, models, and parameters available in the core cannot be specified using BEAUti. Thus, some understanding of the BEAST XML format is essential.

The options for specifying the metadata save to the ***.trees*** file do not seem to be working in older versions of BEAUti (v2.0.1). Thus, the trees are not saved with the branch-rate values annotated on them. We can do this, however, by simply changing the XML files.

- Open the `divtime.xml` file generated by BEAUti in your text editor and glance over the contents. BEAUti provides many comments describing each of the elements in the file.

As you look over the contents of this file, you will notice that the components are specified in an order similar to the steps you took in BEAUti. The XML syntax is very verbose. This feature makes it fairly easy to understand the different elements of the BEAST input file. If you wished to alter your analysis or realized that you misspecified a prior parameter, changing the XML file is far simpler than going through all of the steps in BEAUti again. For example, if you wanted to change bounds of the uniform prior on `mrca3` from (8, 32) to (12, 30), this can be done easily by altering these values in the XML file (Box 4), though leave these at 8 and 32 for this exercise.

```
<distribution id="mrca3" spec="beast.math.distributions.MRCAPrior" tree="@Tree.t:divtimeTree">
    <taxonset id="mrca31" spec="TaxonSet">
        <taxon idref="T10"/>
        <taxon idref="T9"/>
    </taxonset>
    <Uniform id="Uniform.02" lower="8.0" name="distr" upper="32.0"/>
</distribution>
```

Box 4: The XML syntax for specifying a uniform prior distribution on `mrca3`. Changing the parameters (highlighted) of this prior is simply done by altering the XML file.

To save the branch-specific rates in the NEXUS trees, we need to change the `logger` for the trees file and add the `branchratemodel` to the `TreeWithMetaDataLogger`.

```
<logger fileName="divtime.trees" id="treelog.t:divtimeTree" logEvery="100" mode="tree">
    <log branchratemodel="@RelaxedClock.c:divtimeClock" id="TreeWithMetaDataLogger.t:divtimeTree"
         spec="beast.evolution.tree.TreeWithMetaDataLogger" tree="@Tree.t:divtimeTree"/>
<logger>
```

Box 5:

- Modify the `logger` for the trees file by adding `branchratemodel="@RelaxedClock.c:divtimeClock"` as in Box 5 above. Do this for both `divtime.xml` and `divtime.prior.xml`. Look over the elements specified in each of the XML files and verify that everything is satisfactory. Save and close the input files. If you are using the most recent version of BEAUti (BEAST v2.1.1 release or later), then you can skip this step.

Although running multiple, independent analyses is an important part of any Bayesian analysis, BEAST does not do this by default. However, setting up multiple runs is trivial once you have a complete XML file in hand and only requires that you make a copy of the input file and alter the names of the output files in the XML (it's also best to change the initial states for all of your parameters, including the starting tree).

Running BEAST

Now you are ready to start your BEAST analysis. BEAST allows you to use the BEAGLE library if you already have it installed. BEAGLE is an application programming interface and library that effectively takes advantage of your computer hardware (CPUs and GPUs) for doing the heavy lifting (likelihood calculation) needed for statistical phylogenetic inference (Ayers et al., 2012). Particularly, when using BEAGLE's GPU (NVIDIA) implementation, runtimes are significantly shorter.

- Execute `divtime.prior.xml` and `divtime.xml` in BEAST. You should see the screen output every 1,000 generations, reporting the likelihood and several other statistics.

- Once you have verified that your XML file was properly configured and you see the likelihood update, feel free to kill the run. I have provided the output files for this analysis and you can find them in `divtime_beast/output*`.

Summarizing the output

Once the run reaches the end of the chain, you will find three new files in your analysis directory. The MCMC samples of various scalar parameters and statistics are written to the file called `divtime.log`. The tree-state at every sampled iteration is saved to `divtime.trees`. The tree strings in this file are all annotated in extended Newick format with the substitution rate from the uncorrelated lognormal model at each node. The file called `divtime.xml.states` summarizes the performance of the proposal mechanisms (operators) used in your analysis, providing information about the acceptance rate for each move. Reviewing this file can help identify operators that might need adjustment if their acceptance probabilities are too high.

The main output files are the `.log` file and `.trees` file. It is not feasible to review the data contained in these files by simply opening them in a spreadsheet program or a tree viewing program. Fortunately, the developers of BEAST have also written general utility programs for summarizing and visualizing posterior samples from Bayesian inference using MCMC. Tracer is a cross-platform, java program for summarizing posterior samples of scaler parameters. This program is necessary for assessing convergence, mixing, and determining an adequate burn-in. Tree topologies, branch rates, and node heights are summarized using the program TreeAnnotator and visualized in FigTree.

Tracer

This section will briefly cover using Tracer and visual inspection of the analysis output for MCMC convergence diagnostics.

- Open Tracer and import the `divtime.log` file in the ***File→Import New Trace File***.

The fist statistic loaded will be the ***posterior***, in the ***Estimates*** tab, and you can see the list of statistics and variables that you can navigate through and visualize the summaries. You will notice that many items in the ***ESS*** column are red. The MCMC runs you have performed today are all far too short to produce adequate posterior estimates of divergence times and substitution model parameters and this is reflected in the ESS values. The ESS is the *effective sample size* of a parameter. The value indicates the number of effectively independent draws from the posterior in the sample. This statistic can help to identify autocorrelation in your samples that might result from poor mixing. It is important that you run your chains long enough and sufficiently sample the stationary distribution so that the ESS values of your parameters are all high (over 200 or so).

- Click on a parameter with a low ESS and explore the various windows in Tracer. It is clear that we must run the MCMC chain longer to get good estimates.

Provided with the files for this exercise are the output files from analyses run for 100,000,000 iterations. The files can be found in the `divtime_beast/output2/` directory and are all labeled with the file stem: `divtime_100m*`.

- Close `divtime.log` in Tracer using the [-] button and open `divtime_100m.1.log`, `divtime_100m.2.log` and `divtime_100m.prior.log`.

These log files are from much longer runs and since we ran two independent, identical analyses (with different starting values for each parameter), we can compare the log files in Tracer and determine if they have converged on the same stationary distribution. Additionally, analyzing an empty alignment allows you to compare your posterior estimates to the prior distributions used for each parameter.

- Select and highlight all three files (`divtime_100m.1.log`, `divtime_100m.2.log` and `divtime_100m.prior.log`) in the ***Trace Files*** pane (do not include ***Combined***). This allows you to compare all three runs simultaneously. Click on the various parameters and view how they differ in their estimates and 95% credible intervals for those parameters.

The 95% credible interval is a Bayesian measure of uncertainty that accounts for the data. If we use the 95% credible interval, this means that the probability the true value of a parameter lies within this interval is 0.95, given our model and data. This measure is often used to approximate the 95% highest posterior density region (HPD).

- Find the parameter `ucldStdev` and compare the estimates of the standard deviation of the uncorrelated log-normal distribution.

The `ucldStdev` indicates the amount of variation in the substitution rates across branches. Our prior on this parameter is an exponential distribution with $\lambda = 3.0$ ($mean = 0.33333$). Thus, there is a considerable amount of prior weight on `ucldStdev` $= 0$. A standard deviation of 0 indicates support for no variation in substitution rates and the presence of a molecular clock.

- With `ucldStdev` highlighted for all three runs, go to the ***Marginal Density*** window, which allows you to compare the marginal posterior densities for each parameter.

- Color (or "colour") the densities by ***Trace File*** next to ***Colour by*** at the bottom of the window (if you do not see this option, increase the size of your Tracer window). You can also add a ***Legend*** to reveal which density belongs to which run file. [Figure 22]

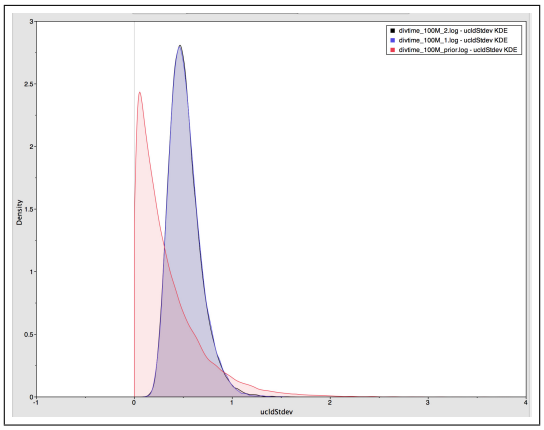

Figure 22: Comparing the marginal densities (using the kernel density estimate) of the `ucldStdev` parameter from 2 independent runs (red and gray) and the prior (blue) in Tracer.

The first thing you will notice from this plot is that the marginal densities from each of our analysis runs (`divtime_100m.1.log` and `divtime_100m.2.log`) are nearly identical. If you click through the other sampled parameters, these densities are the same for each one. This indicates that both of our runs have converged on the same stationary distribution. Since some of the other parameters might not have mixed well, we may want to run them longer, but we can have good confidence that our runs have sampled the same distribution.

Second, notice how the marginal densities for the `ucldStdev` parameter from each of the analysis runs are quite different from the marginal density of that parameter sampled from the prior. The signal in the data is not overwhelmed by the prior on this parameter. Moreover, our analysis runs do not have any significant density at zero, indicating no support for a constant rate of substitution (*e.g.*, strict molecular clock). This is also evident if you view the 95% credible intervals (95% HPD) for each of the runs. When the analyses are run with data, a `ucldStdev` of 0 does not fall within the credible interval.

When calibrating divergence time estimates using off-set parametric prior densities, it is *very* important to evaluate (and report) the marginal densities of both the prior and posterior samples of calibrated node heights.

- Highlight only the trace file containing MCMC samples under the prior (`divtime_100m.prior.log`). Then inspect the ***Marginal Density*** for each calibrated node and the hyperparameter of the calibration density on `mrca1`. The prior densities for the root age, node 2, and node 3 are quite close to the calibration priors we specified in BEAUti and described in Figure 5. For node 1, the marginal prior density reflects the fact that the node age is sampled from a mixture of exponential distributions because of the gamma-distributed hyperprior on the rate of the calibration density.

Because the two analysis runs (`divtime_100m.1.log` and `divtime_100m.2.log`) sampled from the same posterior distribution, particularly for the parameters we are interested in, they can be combined. Tracer does this when you import files and you will see a file called ***Combined*** in the ***Trace Files*** window. To use this option, however, you must first remove the prior trace file.

- Highlight only the prior file (`divtime_100m.prior.log`) in the ***Trace Files*** pane. Click the [-] button below the window to remove the file.

- Now highlight the ***Combined*** trace file. Navigate through the sampled parameters and notice how the ESSs have improved.

Continue examining the options in Tracer. This program is very, *very* useful for exploring many aspects of your analysis.

Summarizing the trees

After reviewing the trace files from the two independent runs in Tracer and verifying that both runs converged on the posterior distributions and reached stationarity, we can combine the sampled trees into a single tree file and summarize the results.

- Open the program LogCombiner and set the ***File type*** to ***Tree Files***. Next, import the two `trees` files in the `divtime_beast/output3/` directory (`divtime_100m.1.trees` and `divtime_100m.2.trees`) using the [+] button.

These analyses were each run for 100,000,000 iterations with a sample frequency of 10,000. Therefore, each of the trees files contains 100,000 trees.

- Set a burn-in percentage of 25 for each file, thus discarding the first 25% of the samples in each tree file. Then click on the ***Choose file ...*** button to create an output file (call it `divtime.comb.trees`) and run the program.

- Alternatively, use LogCombiner in unix with the command:

```
> logcombiner -log divtime_100m.1.trees -log divtime_100m.2.trees -b 25 -o divtime.comb.trees
```

Once LogCombiner has terminated, you will have a file containing 10,000 trees called `divtime.comb.trees` which can be summarized using TreeAnnotator. TreeAnnotator takes a collection of trees and summarizes them by identifying the topology with the best support, calculating clade posterior probabilities, and calculating 95% credible intervals for node-specific parameters. All of the node statistics are annotated on the tree topology for each node in the Newick string.

- Open the program TreeAnnotator. Since we already discarded a set of burn-in trees when combining the tree files, we can leave ***Burnin*** set to `0` (though, if TreeAnnotator is taking a long time to load the trees, set the burnin to 10–60% to reduce the number of trees).

- For the ***Target tree type***, choose ***Maximum clade credibility tree***.

The ***Maximum clade credibility tree*** is the topology with the highest product of clade posterior probabilities across all nodes. Alternatively, you can select the ***Maximum sum of clade credibilities*** which sums all of the clade posteriors. Or you can provide a target tree from file.

The ***Posterior probability limit*** option applies to summaries on a target tree topology and only calculates posteriors for nodes that are above the specified limit.

- Choose ***Median heights*** or ***Mean heights*** for ***Node heights*** which will set the node heights of the output tree to equal the median or mean height for each node in the sample of trees.

- Choose `divtime.comb.trees` as your ***Input Tree File***. Then name the ***Output File:*** `divtime.fig.tre` and click ***Run***.

Once the program has finished running, you will find the file `divtime.fig.tre` in your directory.

- Open `divtime.fig.tre` in your text editor. The tree is written in NEXUS format. Look at the tree string and notice the annotation. Each node in the tree is labeled with comments using the `[&parameter_name=<value>]` format.

An alternative program to LogCombiner and TreeAnnotator is **SumTrees**, a program in the **DendroPy** package (Sukumaran and Holder, 2010). SumTrees is very flexible and allows more options than TreeAnnotator and it provides a way to summarize sets of trees from a number of different programs and analyses.

The tree summaries produced by TreeAnnotator or SumTrees can be opened with FigTree. FigTree reads the comments for each each node and can display them in a variety of ways.

- Open FigTree and open the file `divtime.fig.tre`.

FigTree is a great tree-viewing program and it also allows you to produce publication-quality tree figures. This summary tree is shown in Figure 23. The posterior probabilities are labeled on each branch (they are almost all equal to 1). The branches of the tree are colored by the average substitution rate estimated for each lineage. On each node, bars are displayed representing the node age 95% credible interval.

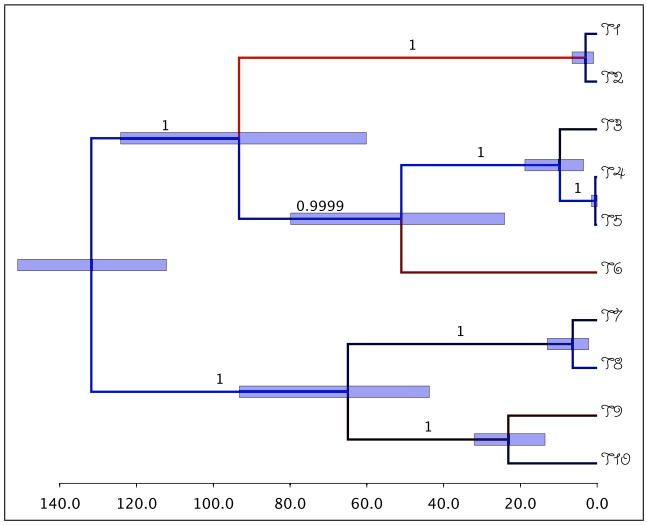

Figure 23: The final tree with node bars representing the age 95% credible intervals and posterior probabilities of each clade labeled on the branches. Each branch is colored according to the average substitution rate sampled by the MCMC chain.

- Explore the various options for creating a tree figure and recreate the tree in Figure 23. The figure you create can be exported as a PDF or EPS file.

Divergence time estimation for this simulated data set is very straight-forward. To generate the sequences for this exercise I first simulated a tree topology and divergence times under a constant-rate birth-death process with 10 extant taxa (`T1, T2, T3, T4, T5, T6, T7, T8, T9, T10`). The simulated time-tree is shown in Figure 24A.

With a tree topology and branch lengths in units of time, I simulated lineage-specific substitution rate variation under an uncorrelated model such that the rate associated with each branch was drawn from a log-normal distribution (Figure 24B). The tree with branch lengths in units of $rate * time$ was then used to

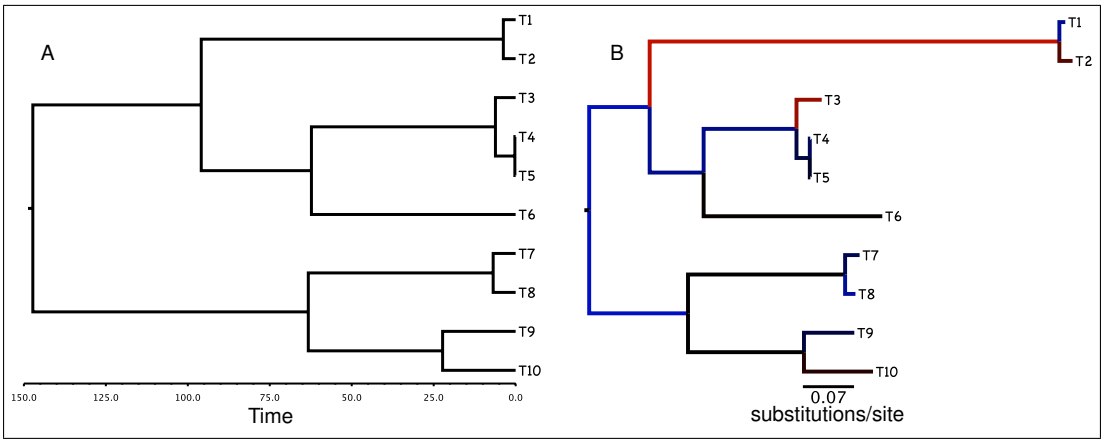

Figure 24: A) The true tree and branching times. The age of the root is equal to 147.28. B) The true tree and branch lengths in units of $rate * time$. The branches are colored according to their substitution rate.

simulate two separate genes each under a different GTR+Γ model. For the most part, the models assumed in this analysis matched the models used to generate the data, therefore our estimates of divergence time and tree topology are quite accurate. We inferred the correct tree topology and the true divergence times all fall within the node age 95% credible intervals.

Useful Links

- *Bayesian Evolutionary Analysis with BEAST 2* (Drummond and Bouckaert, 2014)
- BEAST 2 website and documentation: http://www.beast2.org/
- BEAST 1 website and documentation: http://beast.bio.ed.ac.uk
- Join the BEAST user discussion: http://groups.google.com/group/beast-users
- RevBayes: https://github.com/revbayes/code
- DPPDiv: http://phylo.bio.ku.edu/content/tracy-heath-dppdiv
- PhyloBayes: www.phylobayes.org/
- multidivtime: http://statgen.ncsu.edu/thorne/multidivtime.html
- MCMCtree (PAML): http://abacus.gene.ucl.ac.uk/software/paml.html
- BEAGLE: http://code.google.com/p/beagle-lib/
- A list of programs: http://evolution.genetics.washington.edu/phylip/software.html
- The Paleobiology Database: http://www.paleodb.org
- The Fossil Record & Date A Clade: http://www.fossilrecord.net

Source URL: http://treethinkers.org/divergence-time-estimation-using-beast/.

Version dated: August 2, 2014

Relevant References

Aldous D. 2001. Stochastic models and descriptive statistics for phylogenetic trees, from yule to today. Statistical Science. 16:23–34.

Aldous D, Popovic L. 2005. A critical branching process model for biodiversity. Advances in Applied Probability. 37:1094–1115.

Aris-Brosou S, Yang Z. 2002. Effects of models of rate evolution on estimation of divergence dates with special reference to the Metazoan 18S ribosomal RNA phylogeny. Systematic Biology. 51:703–714.

Aris-Brosou S, Yang Z. 2003. Bayesian models of episodic evolution support a late precambrian explosive diversification of the Metazoa. Molecular Biology and Evolution. 20:1947–54.

Ayers DL, Darling A, Zwickl DJ, et al. (12 co-authors). 2012. BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. Systematic Biology. 61:170–173.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Molecular Biology and Evolution. 30:2157–2167.

Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. 2013. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. Molecular Biology and Evolution. 30:239–243.

Benton MJ, Ayala FJ. 2003. Dating the tree of life. Science. 300:1698–1700.

Benton MJ, Wills MA, Hitchin R. 2000. Quality of the fossil record through time. Nature. 403:534–537.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. Beast 2: A software platform for bayesian evolutionary analysis. PLoS computational biology. 10:e1003537.

Brown R, Yang Z. 2011. Rate variation and estimation of divergence times using strict and relaxed clocks. BMC Evolutionary Biology. 11:271.

Cutler DJ. 2000a. Estimating divergence times in the presence of an overdispersed molecular clock. Molecular Biology and Evolution. 17:1647–1660.

Cutler DJ. 2000b. Understanding the overdispersed molecular clock. Genetics. 154:1403–1417.

Darriba D, Aberer AJ, Flouri T, Heath TA, Izquierdo-Carrasco F, Stamatakis A. 2013. Boosting the performance of Bayesian divergence time estimation with the phylogenetic likelihood library. IEEE 27th International Symposium on Parallel & Distributed Processing. p. doi:10.1109/IPDPSW.2013.267.

dos Reis M, Inoue J, Hasegawa M, Asher R, Donoghue P, Yang Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proceedings of the Royal Society B: Biological Sciences. 279:3491–3500.

dos Reis M, Yang Z. 2011. Approximate likelihood calculation on a phylogeny for bayesian estimation of divergence times. Molecular Biology and Evolution. 28:2161–2172.

dos Reis M, Yang Z. 2012. The unbearable uncertainty of Bayesian divergence time estimation. Journal of Systematics and Evolution. 51:30–43.

Doyle JA, Donoghue MJ. 1993. Phylogenies and angiosperm diversification. Paleobiology. 19:141–167.

Drummond AJ, Bouckaert RR. 2014. Bayesian evolutionary analysis with BEAST 2. Cambridge University Press.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology. 4:e88.

Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. 2003. Measurably evolving populations. Trends in Ecology & Evolution. 18:481–488.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology. 7:214.

Drummond AJ, Suchard MA. 2010. Bayesian random local clocks, or one rate to rule them all. BMC Biology. 8:114.

Fan Y, Wu R, Chen MH, Kuo L, Lewis PO. 2011. Choosing among partition models in Bayesian phylogenetics. Molecular Biology and Evolution. 28:523–532.

Gandolfo MA, Nixon KC, Crepet WL. 2008. Selection of fossils for calibration of molecular dating models. Annals of the Missouri Botanical Garden. 95:34–42.

Gaut BA, Weir BS. 1994. Detecting substitution-rate heterogeneity among regions of a nucleotide sequence. Molecular Biology and Evolution. 11:620–629.

Gavryushkina A, Welch D, Stadler T, Drummond AJ. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. arXiv preprint arXiv:1406.4573. .

Gernhard T. 2008. The conditioned reconstructed process. Journal of Theoretical Biology. 253:769–778.

Graur D, Martin W. 2004. Reading the entrails of chickens: Molecular timescales of evolution and the illusion of precision. Trends in Genetics. 20:80–86.

Griffiths R, Tavare S. 1994. Simulating probability distributions in the coalescent. Theoretical Population Biology. 46:131–159.

Guindon S. 2010. Bayesian estimation of divergence times from large sequence alignments. Molecular Biology and Evolution. 27:1768–1781.

Hasegawa M, Kishino H, Yano T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution. 22:160–174.

Hasegawa M, Kishino H, Yano T. 1989. Estimation of branching dates among primates by molecular clocks of nuclear DNA which slowed down in Hominoidea. Journal of Human Evolution. 18:461–476.

Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 57:97–109.

Heath TA. 2012. A hierarchical Bayesian model for calibrating estimates of species divergence times. Systematic Biology. 61:793–809.

Heath TA, Hedtke SM, Hillis DM. 2008. Taxon sampling and the accuracy of phylogenetic analyses. Journal of Systematics and Evolution. 46:239–257.

Heath TA, Holder MT, Huelsenbeck JP. 2012. A Dirichlet process prior for estimating lineage-specific substitution rates. Molecular Biology and Evolution. 29:939–255.

Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences, USA. 111:E2957–E2966.

Heath TA, Zwickl DJ, Kim J, Hillis DM. 2008. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. Systematic Biology. 57:160–166.

Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution. 27:570–580.

Heled J, Drummond AJ. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Systematic Biology. 61:138–149.

Heled J, Drummond AJ. 2013. Calibrated birth-death phylogenetic time-tree priors for bayesian inference. arXiv preprint arXiv:1311.4921. .

Himmelmann L, Metzler D. 2009. Treetime: an extensible c++ software package for bayesian phylogeny reconstruction with time-calibration. Bioinformatics. 25:2440–2441.

Ho SYW. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. Journal of Avian Biology. 38:409–414.

Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper A. 2011. Time-dependent rates of molecular evolution. Molecular Ecology. 20:3087–3101.

Ho SYW, Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Systematic Biology. 58:367–380.

Höhna S, Drummond AJ. 2012. Guided Tree Topology Proposals for Bayesian Phylogenetic Inference. Systematic Biology. .

Höhna S, Stadler T, Ronquist F, Britton T. 2011. Inferring speciation and extinction rates under different sampling schemes. Molecular Biology and Evolution. 28:2577–2589.

Huelsenbeck JP, Larget B, Swofford DL. 2000. A compound Poisson process for relaxing the molecular clock. Genetics. 154:1879–1892.

Hug LA, Roger AJ. 2007. The impact of fossils and taxon sampling on ancient molecular dating analyses. Molecular Biology and Evolution. 24:1889–1897.

Hugall AF, Foster R, Lee MSY. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. Systematic Biology. 56:543–63.

Hull DL. 1988. Science as a Process. An Evolutionary Account of the Social and Conceptual Development of Science. Chicago: University of Chicago Press.

Inoue J, Donoghue PC, Yang Z. 2010. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. Systematic Biology. 59:74–89.

Kendall DG. 1948. On the generalized "birth-and-death" process. Annals of Mathematical Statistics. 19:1–15.

Kingman JFC. 1982a. The coalescent. Stochastic Processes and their Applications. 13:235–248.

Kingman JFC. 1982b. Exchangeability and the evolution of large populations. In: Koch G, Spizzichino F, editors, Exchangeability in Probability and Statistics. North-Holland, pp. 97–112.

Kingman JFC. 1982c. On the genealogy of large populations. In: Gani J, Hannan EJ, editors, Essays in Statistical Science: Papers in Honour of P. A. P. Moran, Journal of Applied Probability, Special Volume 19A. Applied Probability Trust, pp. 27–43.

Kishino H, Hasegawa M. 1990. Converting distance to time: Application to human evolution. Methods in Enzymology. 183:550–570.

Kishino H, Miyata T, Hasegawa M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. Journal of Molecular Evolution. 31:151–160.

Kishino H, Thorne JL, Bruno W. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Molecular Biology and Evolution. 18:352–361.

Lartillot N. 2013. Interaction between selection and biased gene conversion in mammalian protein-coding sequence evolution revealed by a phylogenetic covariance analysis. Molecular biology and evolution. 30:356–368.

Lartillot N, Delsuc F. 2012. Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. Evolution. 66:1773–1787.

Lartillot N, Poujol R. 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Molecular Biology and Evolution. 28:729–744.

Lee MSY, Oliver PM, Hutchinson MN. 2009. Phylogenetic uncertainty and molecular clock calibrations: A case study of legless lizards (Pygopodidae, Gekkota). Molecular Phylogenetics and Evolution. 50:661–666.

Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. PLoS Computational Biology. 5:e1000520.

Lepage T, Bryant D, Philippe H, Lartillot N. 2007. A general comparison of relaxed molecular clock models. Molecular Biology and Evolution. 24:2669–2680.

Lepage T, Lawi S, Tupper P, Bryant D. 2006. Continuous and tractable models for the variation of evolutionary rates. Mathematical Biosciences. 199:216–233.

Li WLS, Drummond AJ. 2012. Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. Molecular Biology and Evolution. 29:751–761.

Lloyd GT, Young JR, Smith AB. 2012. Taxonomic structure of the fossil record is shaped by sampling bias. Systematic Biology. 61:80–89.

Lukoschek V, Keogh JS, Avise JC. 2012. Evaluating fossil calibrations for dating phylogenies in light of rates of molecular evolution: A comparison of three approaches. Systematic Biology. 61:22–43.

Magallón S. 2009. Using fossils to break long branches in molecular dating: A comparison of relaxed clocks applied to the origin of angiosperms. Systematic Biology. 59:384–399.

Marshall CR. 1990. Confidence intervals on stratigraphic ranges. Paleobiology. 16:1–10.

Marshall CR. 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. The American Naturalist. 171:726–742.

Morlon H, Parsons T, Plotkin J. 2011. Reconciling molecular phylogenies with the fossil record. Proceedings of the National Academy of Sciences. 108:16327–16332.

Muse SV, Weir BS. 1992. Testing for equality of evolutionary rates. Genetics. 132:269–276.

Nee S, May RM, Harvey PH. 1994. The reconstructed evolutionary process. Philosophical Transactions of the Royal Society B. 344:305–311.

Nylander J, Wilgenbusch J, Warren D, Swofford D. 2008. Awty (are we there yet?): a system for graphical exploration of mcmc convergence in bayesian phylogenetics. Bioinformatics. 24:581.

Parham JF, Donoghue PCJ, Bell CJ, et al. (25 co-authors). 2012. Best practices for justifying fossil calibrations. Systematic Biology. 61:346–359.

Popovic L. 2004. Asymptotic genealogy of a critical branching process. Annals of Applied Probability. 14:2120–2148.

Pyron RA. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Systematic Biology. 60:466–481.

Rambaut A, Bromham L. 1998. Estimating divergence dates from molecular sequences. Molecular Biology and Evolution. 15:442–448.

Rannala B, Yang Z. 1996. Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. Journal of Molecular Evolution. 43:304–311.

Rannala B, Yang Z. 2007. Inferring speciation times under an episodic molecular clock. Systematic Biology. 56:453–466.

Robinson M, Gouy M, Gautier C, Mouchiroud D. 1998. Sensitivity of the relative-rate test to taxonomic sampling. Molecular Biology and Evolution. 15:1091–1098.

Rodrigo AG, Goode M, Forsberg R, Ross HA, Drummond A. 2003. Inferring evolutionary rates using serially sampled sequences from several populations. Molecular biology and evolution. 20:2010–2018.

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology. 61:973–999.

Rutschmann F, Eriksson T, Salim KA, Conti E. 2007. Assessing calibration uncertainty in molecular dating: the assignment of fossils to alternative calibration points. Systematic Biology. 56:591–608.

Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Molecular Biology and Evolution. 14:1218–1231.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution. 19:101–109.

Shapiro B, Ho SYW, Drummond AJ, Suchard MA, Pybus OG, Rambaut A. 2011. A Bayesian phylogenetic method to estimate unknown sequence ages. Molecular Biology and Evolution. 28:879–887.

Stadler T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. Journal of Theoretical Biology. 261:58–66.

Stadler T. 2010. Sampling-through-time in birth-death trees. Journal of Theoretical Biology. 267:396–404.

Stadler T. 2011a. Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Sciences, USA. 108:6187–6192.

Stadler T. 2011b. Simulating trees on a fixed number of extant species. Systematic Biology. 60:668–675.

Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. 2013. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proceedings of the National Academy of Sciences. 110:228–233.

Suchard MA, Weiss RE, Sinsheimer JS. 2003. Testing a molecular clock without an outgroup: Derivations of induced priors on branch-length restrictions in a bayesian framework. Systematic Biology. 52:48–54.

Sukumaran J, Holder MT. 2010. DendroPy: A Python library for phylogenetic computing. Bioinformatics. 26:1569–1571.

Swofford DL. 1998. PAUP*: Phylogenetic Analysis Using Parsimony and Other Methods. Sunderland, Massachusetts: Sinauer Associates, Inc.

Tajima F. 1993. Simple methods for testing molecular clock hypothesis. Genetics. 135:599–607.

Thompson EA. 1975. Human Evolutionary Trees. Cambridge, England: Cambridge University Press.

Thorne J, Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. Systematic Biology. 51:689–702.

Thorne J, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. Molecular Biology and Evolution. 15:1647–1657.

Thorne JL, Kishino H. 2005. Estimation of divergence times from molecular sequence data. In: Nielsen R, editor, Statistical Methods in Molecular Evolution. New York: Springer, pp. 235–256.

Thorpe JP. 1982. The molecular clock hypothesis: Biochemical evolution, genetic differentiation and systematics. Annual Review of Ecology and Systematics. 13:139–168.

Warnock RCM, Yang Z, Donoghue PCJ. 2012. Exploring the uncertainty in the calibration of the molecular clock. Biology Letters. 8:156–159.

Weir J, Schluter D. 2008. Calibrating the avian molecular clock. Molecular Ecology. 17:2321–2328.

Welch JJ, Fontanillas E, Bromham L. 2005. Molecular dates for the "cambrian explosion": The influence of prior assumptions. Systematic Biology. 54:672–678.

Wilkinson RD, Steiper ME, Soligo C, Martin RD, Yang Z, Tavaré S. 2011. Dating primate divergences through an integrated analysis of palaeontological and molecular data. Systematic Biology. 60:16–31.

Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Systematic Biology. 60:150–160.

Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method. Molecular Biology and Evolution. 14:717–724.

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Molecular Biology and Evolution. 23:212–226.

Yang Z, Yoder AD. 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. Systematic Biology. 52:705–716.

Yoder AD, Yang Z. 2000. Estimation of primate speciation dates using local molecular clocks. Molecular Biology and Evolution. 17:1081–1090.

Yule GU. 1924. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Wills, F. R. S. Philosophical Transactions of the Royal Society of London, Biology. 213:21–87.

Zuckerkandl E, Pauling L. 1962. Molecular disease, evolution, and genetic heterogeneity. In: Kasha M, Pullman B, editors, Horizons in Biochemistry. Academic Press, New York, pp. 189–225.

Zuckerkandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. Evolving Genes and Proteins. pp. 97–166.