Phylogeographic inference in discrete space
A hands-on practical

This chapter provides a step-by-step tutorial for reconstructing the spatial dynamics of Eurasian influenza H5N1 based on a set of 190 hemagglutinin (HA) sequences isolated at different points in time (1996-2005) in 10 different Eurasian locations (Wallace et al., 2007, PNAS, 104 (11): 4473:4478; Lemey et al., 2009, PLoS Comput Biol 5(9): e1000520). The aim of this tutorial is to estimate the ancestral locations of the virus using a Bayesian discrete phylogeographic approach and to infer the most significant epidemiological links using a Bayesian Stochastic Search Variable Selection (BSSVS) procedure.

The first step will be to convert a NEXUS file with a DATA or CHARACTERS block into a BEAST XML input file. This is done using the program BEAUti (this stands for Bayesian Evolutionary Analysis Utility). This is a user-friendly program for setting the evolutionary model and options for the MCMC analysis. The second step is to actually run BEAST using the input file that contains the data, model and settings. The final step is to explore the output of BEAST in order to diagnose problems and to summarize the results.

To undertake this tutorial, you will need to download the following software packages in a format that is compatible with your computer system (all are available for Mac OS X, Windows and Linux/UNIX operating systems):

- **BEAST** - this package contains the BEAST program, BEAUti and a couple of utility programs. At the time of writing, the current version is v1.7.5. It is available for download from http://beast.bio.ed.ac.uk/.

BEAST v 1.10

- **Tracer** - this program is used to explore the output of **BEAST** (and other Bayesian MCMC programs). It graphically and quantitively summarizes the empirical distributions of continuous parameters and provides diagnostic information. At the time of writing, the current version is v1.5. It is available for download from http://beast.bio.ed.ac.uk/.

- **FigTree** - this is an application for displaying and printing molecular phylogenies, in particular those obtained using **BEAST**. At the time of writing, the current version is v1.4. It is available for download from http://tree.bio.ed.ac.uk/.

- **SPREAD** - this is an application for the visualization of phylogeographic analyses performed with **BEAST**. At the time of writing, the current version is v1.0.5. The application and its tutorial are available for download from http://www.phylogeography.org/SPREAD.

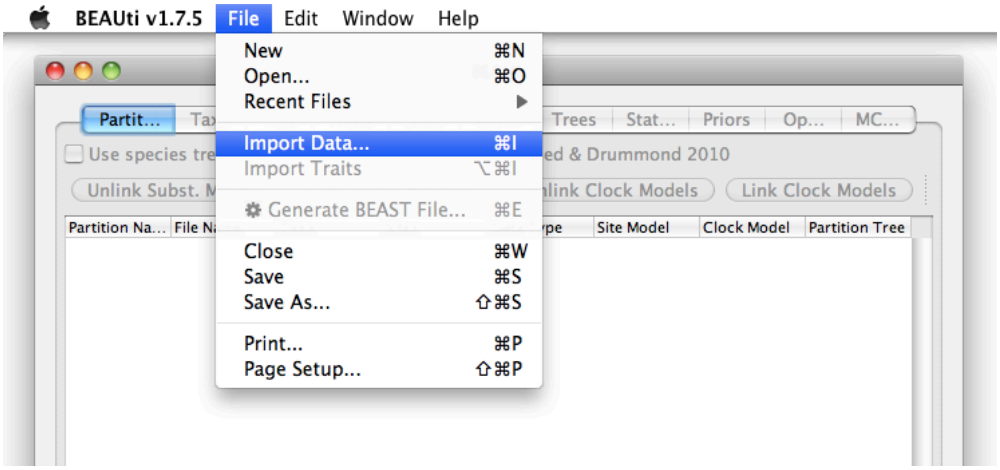
- **Google Earth** - this is a freely available virtual globe software that can be used to visualize KML output from SPREAD in an interactive fashion. **Google Earth** is available at http://earth.google.com.

Running BEAUti

The program **BEAUti** is a user-friendly program for setting the model parameters for BEAST. Run BEAUti by double clicking on its icon.

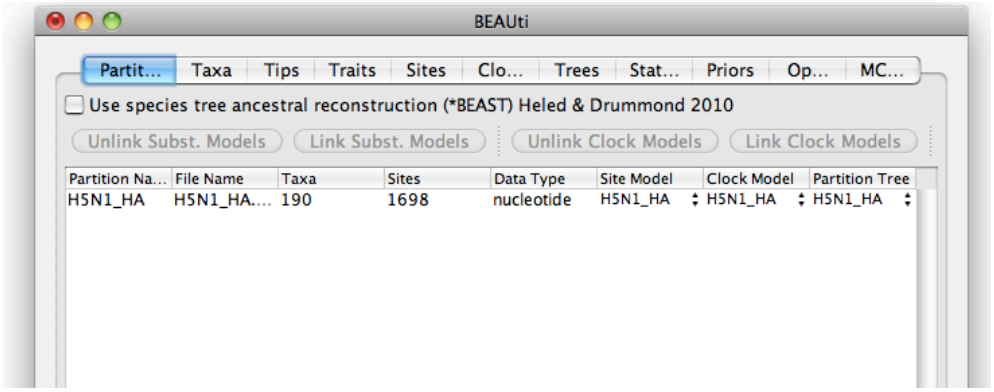
Loading the NEXUS file

To load a NEXUS format alignment, simply select the **Import Data..**. option from the **File** menu.



The NEXUS alignment

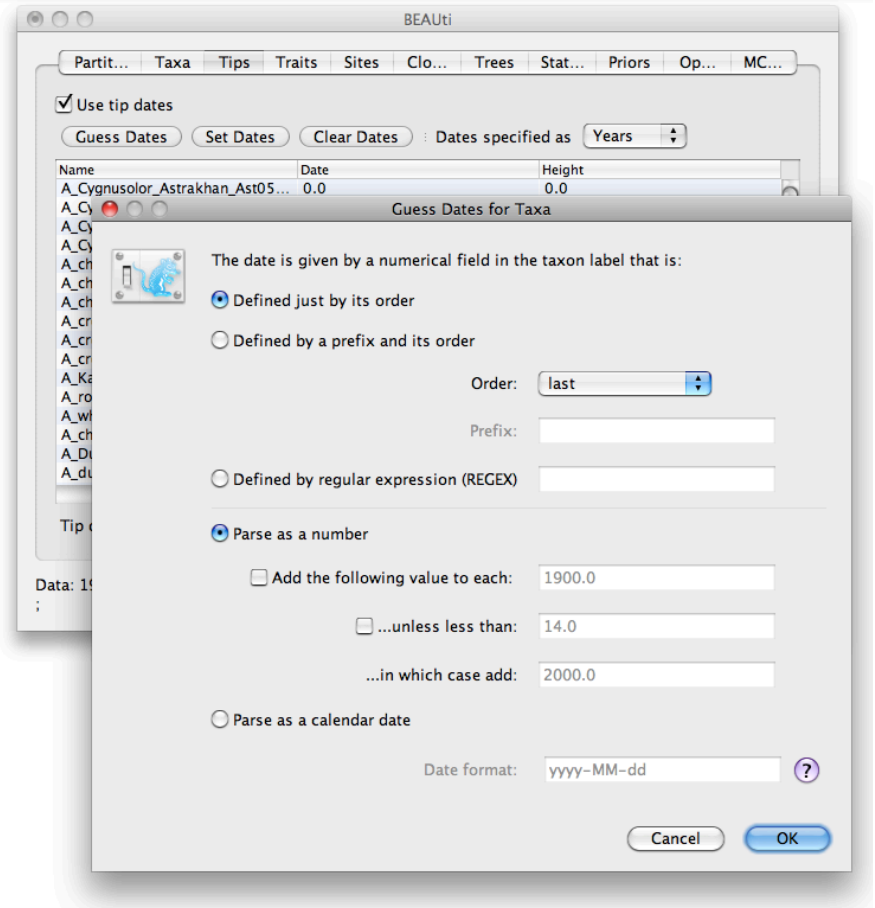
Select the file called `H5N1_NA.nex`. This file contains an alignment of 190 *HA* gene sequences of influenza A H5N1, 1698 nucleotides in length. Once loaded, the sequence data will be listed under **Data Partitions**:



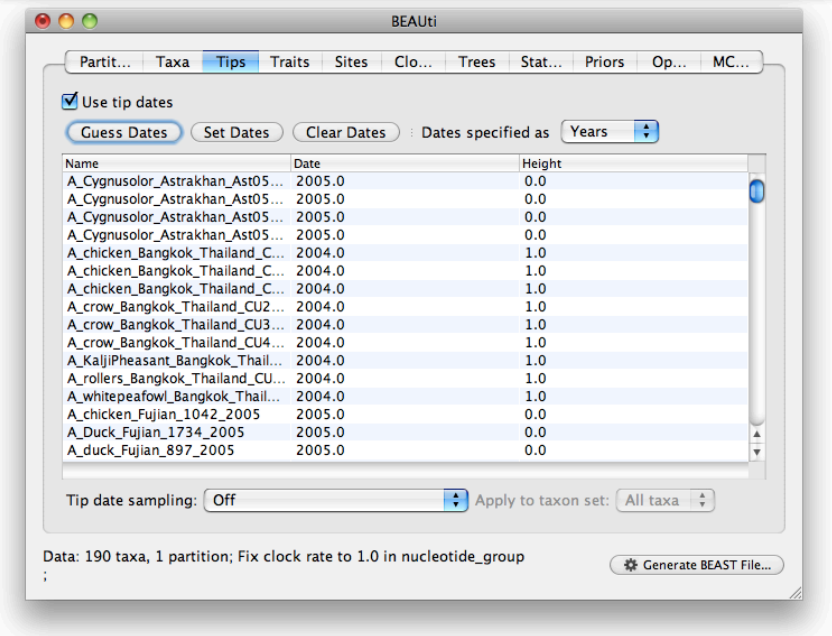
Specifying the sampling date information

To inform BEAUti/BEAST about the sampling dates of the sequences, go to the **Tips** tab and select the `Use tip dates`"option. By default all the taxa are assumed to have a date of zero (i.e. the sequences are assumed to be sampled at the same time). In this case, the H5N1 sequences have been sampled at various dates going back to 1996. The actual year of sampling is given in the name of each taxon and we could simply edit the value in the Date column of the table to reflect these. However, if the taxa names contain the calibration information, then a convenient way to specify the dates of the sequences in BEAUti is to use the `Guess Dates` button at the top of the Data tab. Clicking this will make a dialog box appear:

"Parse Dates"



This operation attempts to guess what the dates are from information contained within the taxon names. It works by trying to find a numerical field within each name. If the taxon names contain more than one numerical field then you can specify how to find the one that corresponds to the date of sampling. You can (1) specify the order that the date field comes (e.g., first, last or various positions in between) or (2) specify a prefix (some characters that come immediately before the date field in each name) and the order of the field, or (3) define a regular expression (**REGEX**). For the H5N1 sequences you can keep the default **Defined just by its order** and set the **Order**: to **last**.

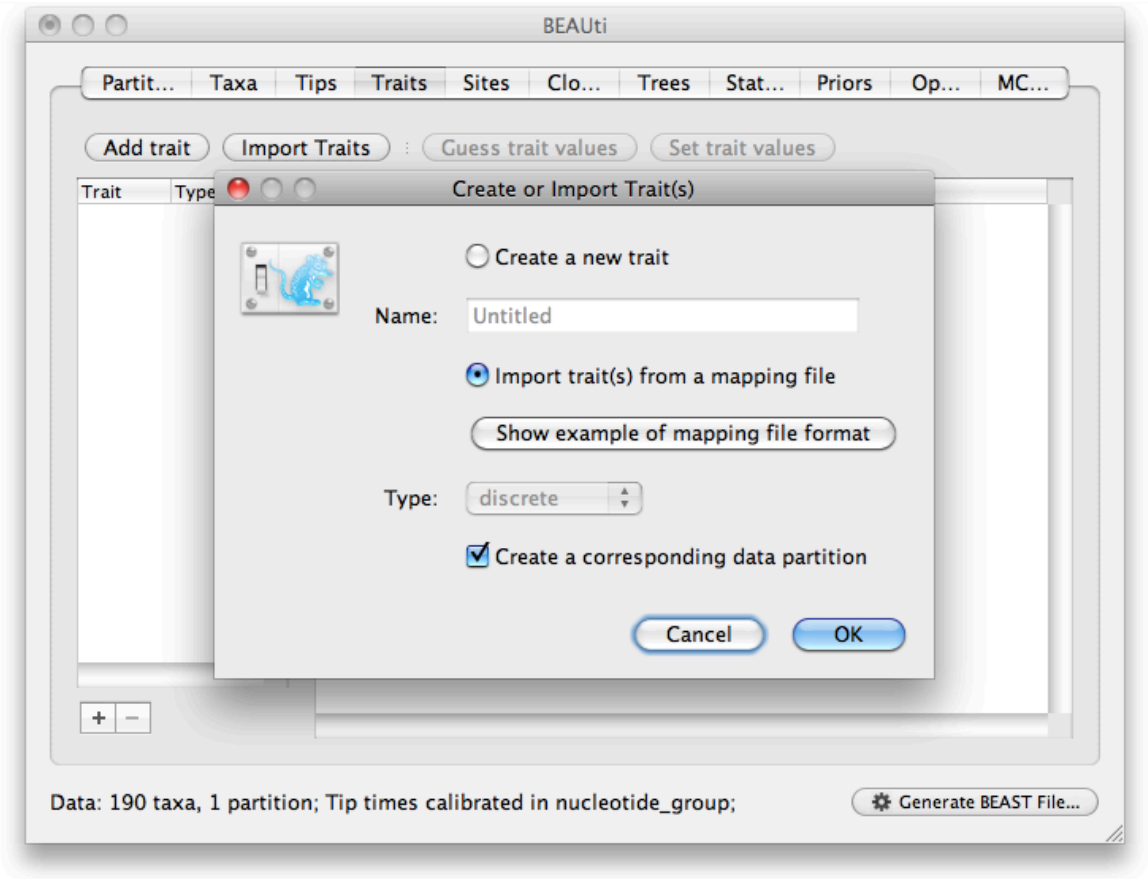


When parsing a number, you can ask BEAUti to add a fixed value to each guessed date. For example, the value "1900" can be added to turn the dates from 2 digit years to 4 digit. Any dates in the taxon names given as "00" would thus become "1900". However, if these "00" or "01", etc. represent sequences sampled in 2000, 2001, etc., '2000' needs to be added to those. This can be achieved by selecting the **unless less than**: .. and **..in which case add**:.. option adding for example 2000 to any date less than 13. There is also an option to parse calendar dates should this be required which is not needed here. After pressing **OK**, the dates will be set to their appropriate sampling year. At the top of the window you can set the units that the dates are given in (years, months, days) and whether they are specified relative to a point in the past (as would be the case for years such as 1996) or backwards in time from the present (as in the case of radiocarbon ages).

Specifying the spatial information for the sequences

The next thing to do is to click on the **Traits** tab at the top of the main window. A trait can be any characteristic that is inherent to the specific taxon, for example, geographical location or species host. This step will assign a specific geographical location to each taxa.

To associate the sequences with the sampling locations, we need to add a new trait under the **Traits** tab (click **Add trait**). This will open a new window to **Create or Import Trait(s)**:

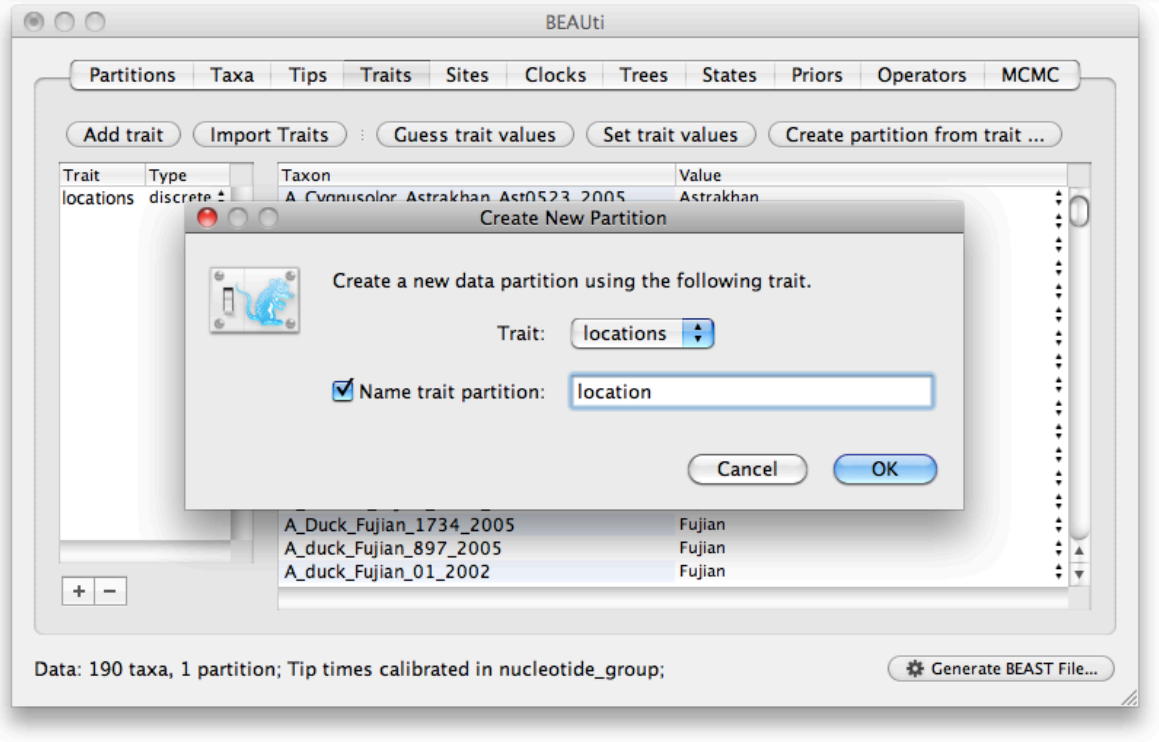


Select **Import trait(s) from a mapping file format**. Browse to and load the **TaxaLocations.txt** tab-delimited file which contains the discrete location for each sequence (in this example there are a total of k=20 locations).

```
traits    locations
A_Cygnusolor_Astrakhan_Ast0523_2005        Astrakhan
A_Cygnusolor_Astrakhan_Ast0524_2005        Astrakhan
A_Cygnusolor_Astrakhan_Ast0525_2005        Astrakhan
A_Cygnusolor_Astrakhan_Ast0527_2005        Astrakhan
A_chicken_Bangkok_Thailand_CU20_2004       Bangkok
                ....
```

A_Mallardduck_Vietnam_133_2004 Vietnam
A_muscovyduck_Vietnam_MdGL_2004 Vietnam
A_quail_Vietnam_36_2004 Vietnam
A_VietNam_1203_2004 Vietnam
A_chicken_Vietnam_27_2003 Vietnam

After clicking OK, select create partition from trait..'. Enter the name location for this partition:

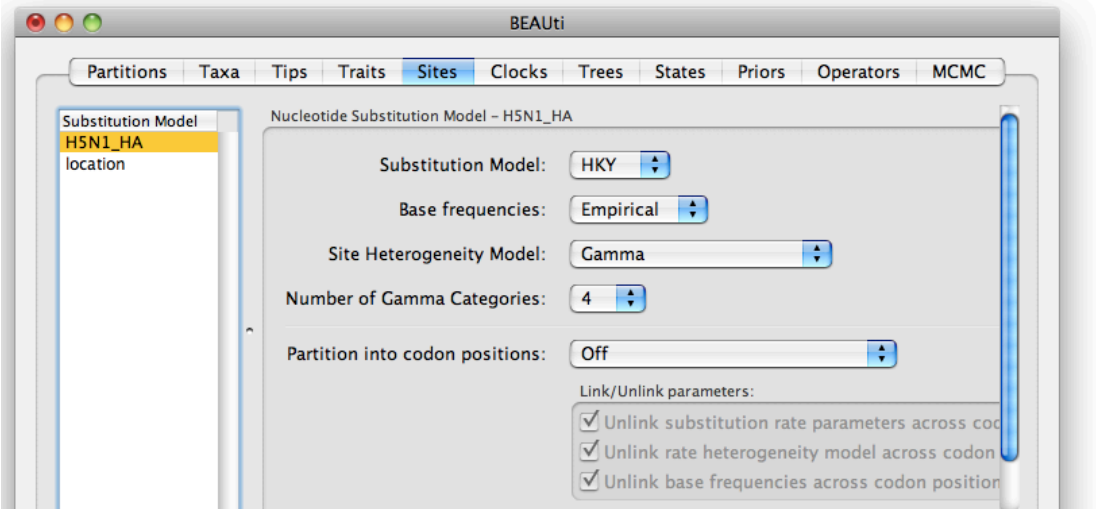


Setting the evolutionary and diffusion model

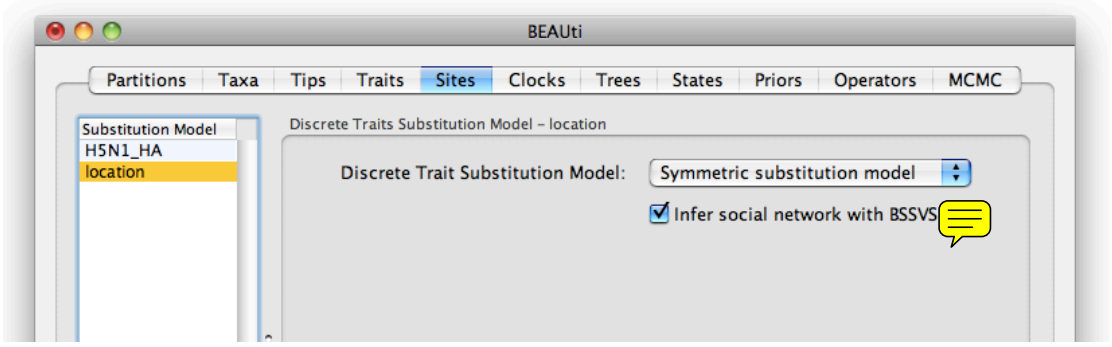
The next thing to do is to click on the **Sites** tab at the top of the main window. This will reveal the evolutionary model settings for BEAST. Exactly which options appear depend on whether the data are nucleotides, amino acids or traits. This tutorial assumes that you are familiar with the evolutionary models available; however there are a couple of points to note about selecting a model in **BEAUti**:

- Selecting the Partition into codon positions option assumes that the data are aligned as codons. This option will then estimate a separate rate of substitution for each codon position, or for 1+2 versus 3, depending on the setting.
- Selecting the Unlink substitution model across codon positions will specify that BEAST should estimate a separate transition-transversion ratio or general time reversible rate matrix for each codon position.
- Selecting the Unlink rate heterogeneity model across codon positions will specify that BEAST should estimate set of rate heterogeneity parameters (gamma shape parameter and/or proportion of invariant sites) for each codon position.

For the nucleotide model in this tutorial, keep the default HKY substitution model, set base frequencies to Empirical, and use Gamma-distributed rate variation among sites.

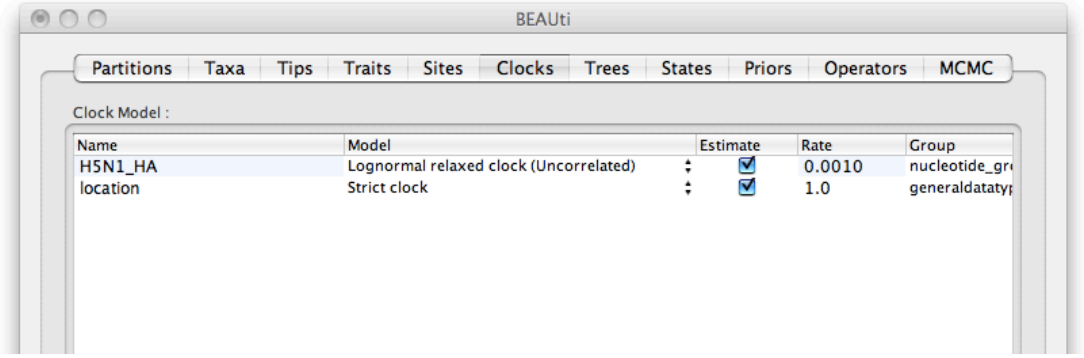


Click on **location** in the **Substitution model** window and set the **Discrete Trait Substitution Model** to **Symmetric substitution model** and select the option to perform **BSSVS (Infer social network with BSSVS)**. The **Symmetric substitution model** specifies a discrete phylogeographic analysis using a standard continuous-time Markov chain (CTMC), in which the transition rates between locations are symmetric. Keep this model for this analysis. The **Asymmetric substitution model** specifies a discrete phylogeographic analysis using a nonreversible CTMC, or in other words, transition probabilities are not symmetrical. Selecting the **BSSVS** option enables the Bayesian Stochastic Search Variable Selection procedure. This procedure will attempt to invoke a limited number of rates (at least k-1) to adequately explain the phylogenetic diffusion process.



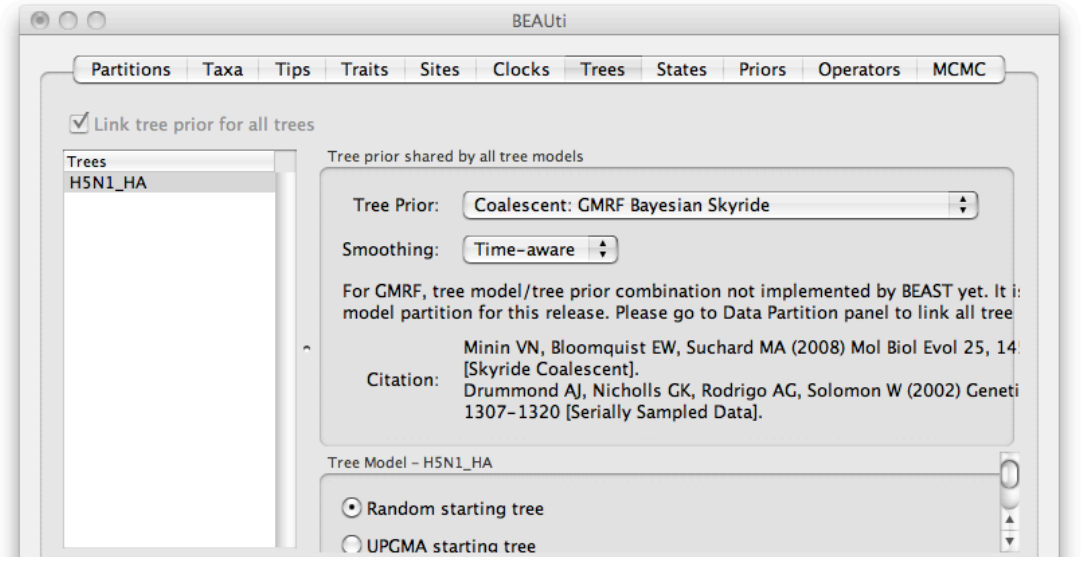
Setting the clock model

Click on the **Clocks** tab at the top of the main window. We will perform our run using the **Lognormal relaxed molecular clock (Uncorrelated)** model and set the **Rate** (initial value) to 0.001 In the **Clocks** tab, select the **Lognormal relaxed molecular clock (Uncorrelated)** model for the nucleotide partition (**H5N1_HA**) and set the **Rate** (initial value) to 0.001.



Setting the starting tree and tree prior

Click on the **Trees** tab at the top of the main window. We will select the Bayesian skyride model as a flexible demographic tree prior (**Coalescent: GMRF Bayesian Skyride**) and keep the default random starting tree. In the **Trees** tab, select the Bayesian skyride model as a flexible demographic tree prior (**Coalescent: GMRF Bayesian Skyride**) and keep the default random starting tree.



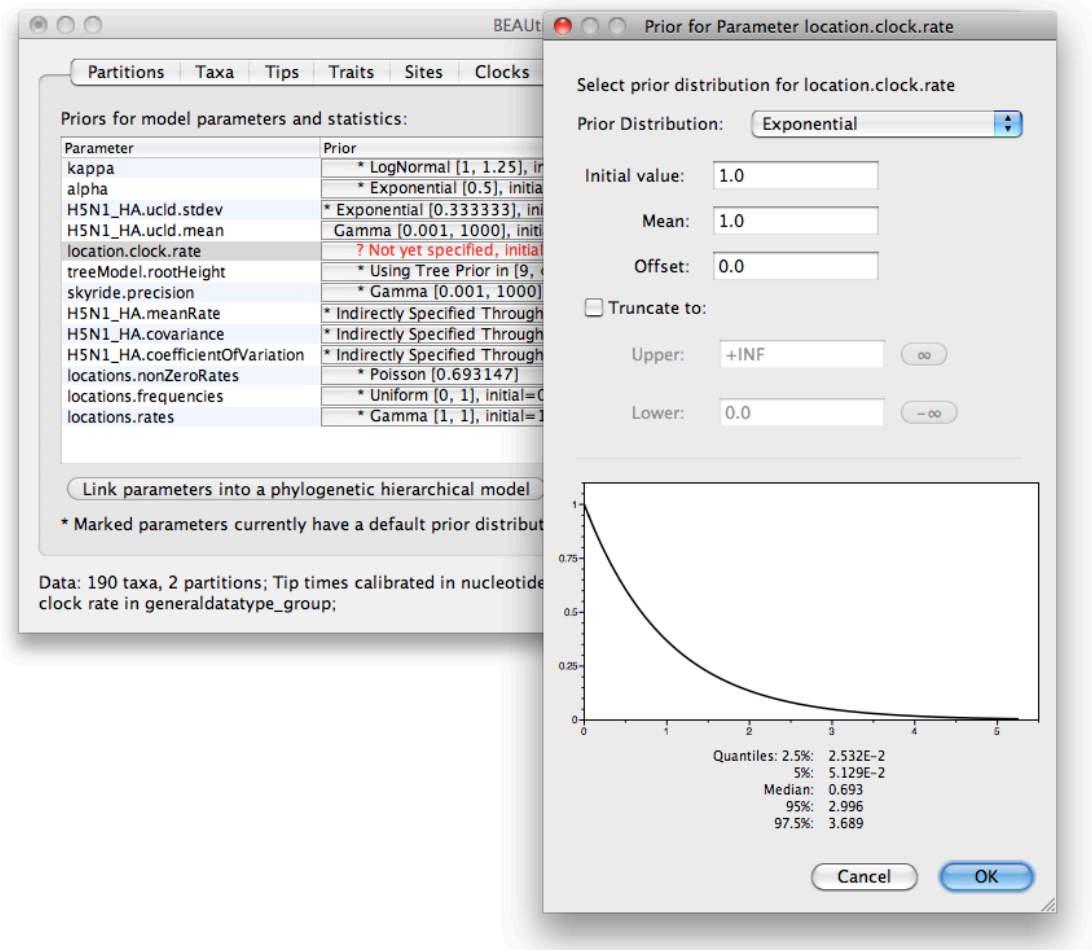
Choose Coalescent Constant Size instead…. GMRF Bayesian Skyride coalescent gives poor results

In addition, in the **States** tab, check that for the **locations** partition the option to **Reconstruct states at all ancestors** is selected (by default).

Setting up the priors

Review the prior settings under the **Priors** tab. Priors that have not been set yet appear in red (**ucld.mean** and **location.clock.rate**). Click on the prior for the **ucld.mean** and a prior selection window will appear. Set the prior to a gamma distribution with shape = 0.001 and scale = 1000. The graphical representation of this prior distribution indicates that most prior mass is put on small values, but the density remains sufficiently diffuse. Notice that the prior setting turns black after confirming this setting by clicking **OK**.

For the discrete location state rate (**locations.clock.rate**), we will use a somewhat more informative prior; we recommend to set this to an exponential prior with **Mean** = 1.0 (and **Offset** = 0).



Setting up the operators

Each parameter in the model has one or more "operators" (these are variously called *moves* and *proposals* by other MCMC software packages such as **MrBayes** and **LAMARC**). The operators specify how the parameters change as the MCMC runs. The operators tab in **BEAUti** has a table that lists the parameters, their operators and the tuning settings for these operators. In the first column are the parameter names while the next column has the type of operators that are acting on each parameter. For example, the scale operator scales the parameter up or down by a random proportion and the uniform operator simply picks a new value uniformly within a range. Some parameters relate to the tree or to the divergence times of the nodes of the tree and these have special operators.

The next column, labelled Tuning, gives a tuning setting to the operator. Some operators don't have any tuning settings so have **n/a** under this column. The tuning parameter will determine how large a move each operator will make which will affect

how often that change is accepted by the MCMC which will affect the efficiency of the analysis. For most operators (like the subtree slide operator) a larger tuning parameter means larger moves. However for the scale operator a tuning parameter value closer to 0.0 means bigger moves. At the top of the window is an option called **Auto Optimize** which, when selected, will automatically adjust the tuning setting as the MCMC runs to try to achieve maximum efficiency. At the end of the run a table of the operators, their performance and the final values of these tuning settings can be written to standard output.

The next column, labelled **Weight**, specifies how often each operator is applied relative to the others. Some parameters tend to be sampled very efficiently - an example is the kappa parameter - these parameters can have their operators down-weighted so that they are not changed as often. We can keep the default operator settings for the current analysis.

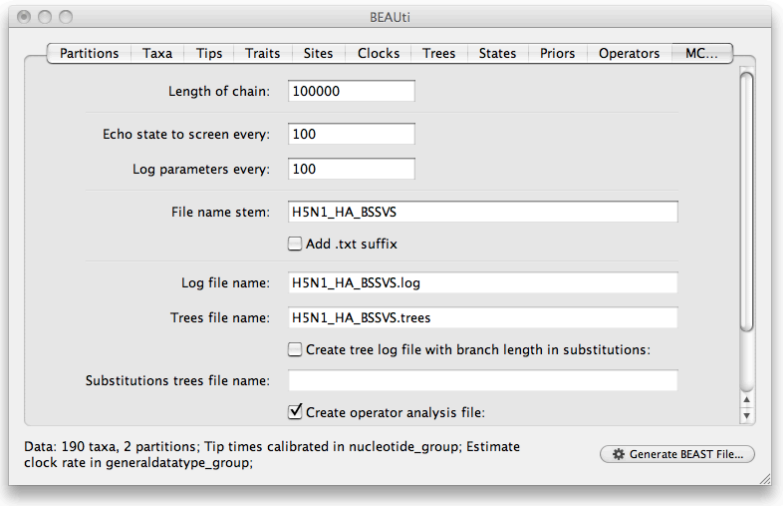
Setting the MCMC options

The **MCMC** tab in BEAUti provides settings to control the MCMC chain. Firstly we have the **Length of chain**. This is the number of steps the MCMC will make in the chain before finishing. How long this should depend on the size of the dataset, the complexity of the model and the precision of the answer required. The default value of 10,000,000 is entirely arbitrary and should be adjusted according to the size of your dataset. We will see later how the resulting log file can be analyzed using Tracer in order to examine whether a particular chain length is adequate.

The next couple of options specify how often the current parameter values should be displayed on the screen and recorded in the log file. The screen output is simply for monitoring the program's progress so can be set to any value (although if set too small, the sheer quantity of information being displayed on the screen will slow the program down). For the log file, the value should be set relative to the total length of the chain. Sampling too often will result in very large files with little extra benefit in terms of the precision of the estimates. Sample too infrequently and the log file will not contain much information about the distributions of the parameters. You probably want to aim to store no more than 10,000 samples so this should be set to something >= chain length / 10,000. For this dataset let's initially set the chain length to 100,000 as this will run quickly on most modern computers. Although the suggestion above would indicate a lower sampling frequency, in this case set both the sampling frequencies to 100.

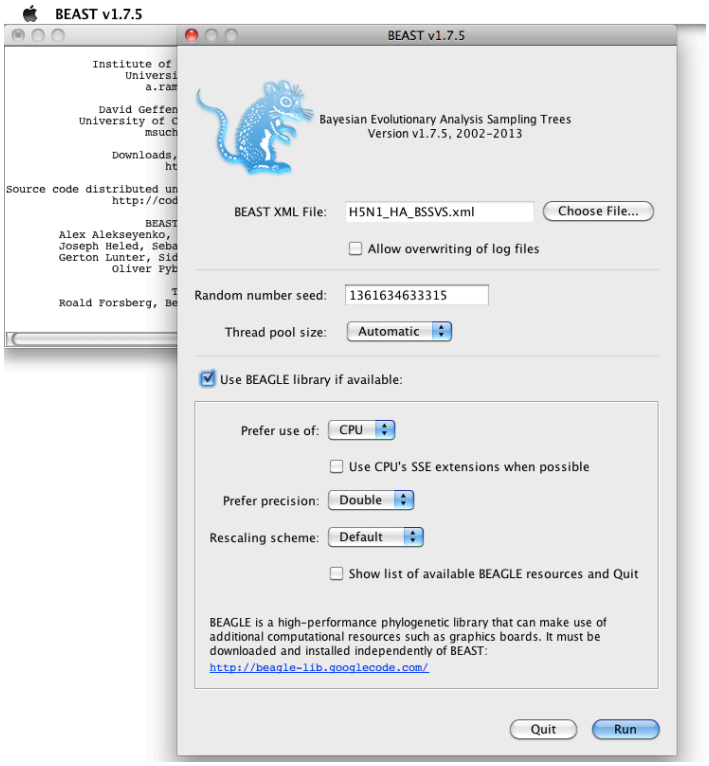
The next option allows the user to set the File stem name; set to **H5N1_HA_BSSVS**, you can type this in here. The next two options give the file names of the log files for the parameters and the trees. These will be set to a default based on the file stem name. Let's also create an operator analysis file by selecting the relevant option. Finally, an option is available to sample from the prior only, which can be useful to evaluate how divergent our posterior estimates are when information is drawn from the data. Here, we will not select this option, but analyze the actual data.

At this point we are ready to generate a BEAST XML file and to use this to run the Bayesian evolutionary analysis. To do this, either select the **Generate BEAST File...** option from the File menu or click the similarly labelled button at the bottom of the window. BEAUti will ask you to review the prior settings one more time before saving the file (and indicate that some are improper). Continue and choose a name for the file (for example, **H5N1_HA_BSSVS.xml**) and save the file. For convenience, you can leave the *BEAUti* window open so that you can change the values and re-generate the *BEAST* file if necessary.



Running BEAST

Once the **BEAST** XML file has been created the analysis itself can be performed using **BEAST**. The exact instructions for running **BEAST** depends on the computer you are using, but in most cases a standard file dialog box will appear in which you select the XML file: If the command line version is being used then the name of the XML file is given after the name of the **BEAST** executable. Select the option Use BEAGLE library if available (see Tutorial X to install BEAGLE library) and use the default settings. When pressing Run, the analysis will be performed with detailed information about the progress of the run being written to the screen. When it has finished, the log file and the trees file will have been created in the same location as your XML file.

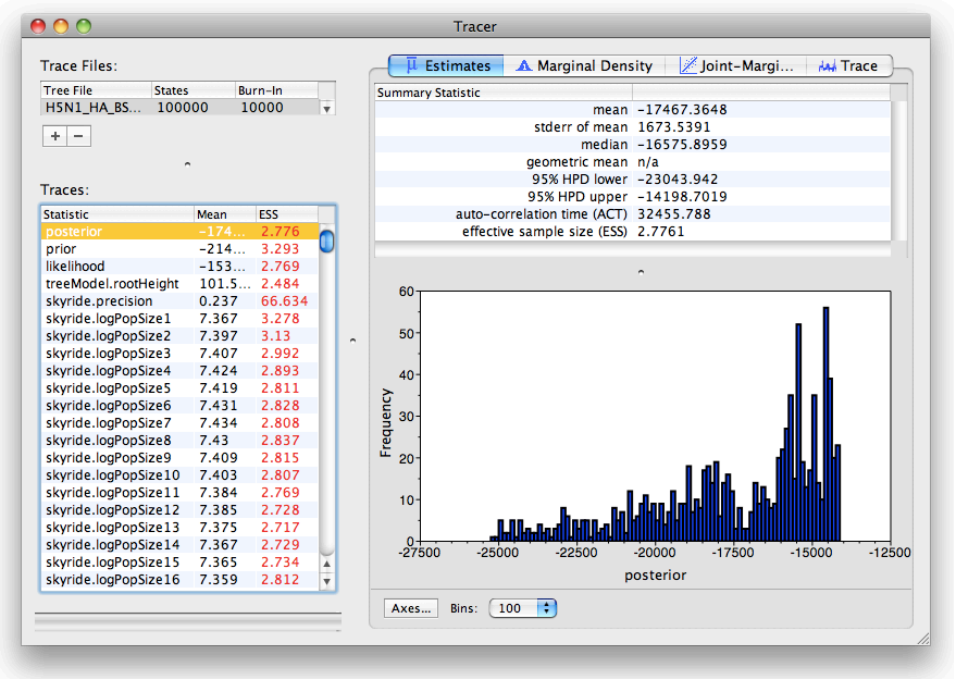


Analysing the BEAST output

To analyze the results of running BEAST we are going to use the program **Tracer**. The exact instructions for running Tracer differs depending on which computer you are using. Please see the README text file that was distributed with the version you downloaded. Once running, Tracer will look similar irrespective of which computer system it is running on.

Select the Import Trace File... option from the File menu. If you have it available, select the log file that you created in the previous section. The file will load and you will be presented with a window similar to the one below. Remember that MCMC is a stochastic algorithm so the actual numbers will not be exactly the same.

On the left hand side is the name of the log file loaded and the traces that it contains. There are traces for a quantity proportional to posterior (this is the product of the data likelihood and the prior probabilities, on the log-scale), and the continuous parameters. Selecting a trace on the left brings up analyses for this trace on the right hand side depending on tab that is selected. When first opened, the posterior trace is selected and various statistics of this trace are shown under the Estimates tab.



In the top right of the window is a table of calculated statistics for the selected trace. The statistics and their meaning are described in the table below.

Mean - The mean value of the samples (excluding the burn-in).

Stdev of mean - The standard error of the mean. This takes into account the effective sample size so a small ESS will give a large standard error.

Median - The median value of the samples (excluding the burn-in).

Geometric mean - The central tendency or typical value of the set of samples (excluding the burn-in).

95% HPD Lower - The lower bound of the highest posterior density (HPD) interval. The HPD is the shortest interval that contains 95% of the sampled values.

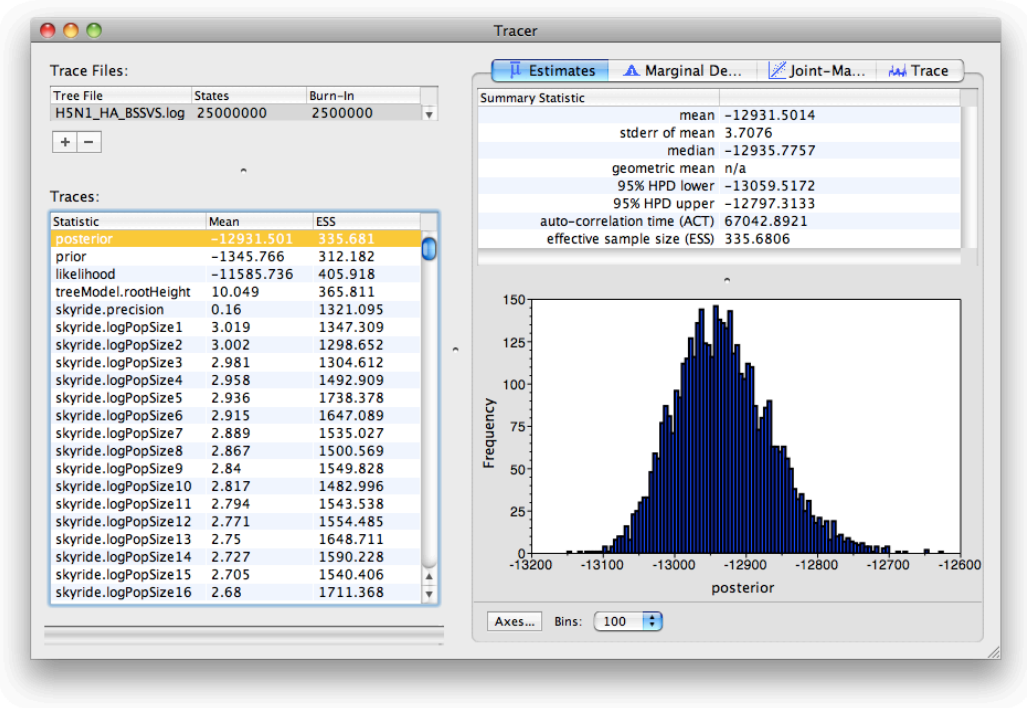
95% HPD Upper - The upper bound of the highest posterior density (HPD) interval.

Auto-Correlation Time (ACT) - The average number of states in the MCMC chain that two samples have to be separated by for them to be uncorrelated (i.e. independent samples from the posterior). The ACT is estimated from the samples in the trace (excluding the burn-in).

Effective Sample Size (ESS) - The effective sample size (ESS) is the number of independent samples that the trace is equivalent to. This is calculated as the chain length (excluding the burn-in) divided by the ACT.

Note that the effective sample sizes (ESSs) for all the traces are small (ESSs less than 100 are highlighted in red by Tracer and values > 100 but < 200 are in yellow). This is not good. A low ESS means that the trace contained a lot of correlated samples and thus may not represent the posterior distribution well. In the bottom right of the window is a frequency plot of the samples which is expected given the low ESSs is extremely rough. Inspecting the **Trace** of many continuous parameters shows that the default burn-in of 10% of the chain length is inadequate for this example (the posterior values are still increasing over most of the chain). Not excluding enough of the start of the chain as burn-in will bias the results and render estimates of ESS unreliable.

The simple response to this situation is that we need to run the chain for longer. The example below was run for 25 million steps, sampling every 5,000th step, which means that 5,000 samples where stored in the log file. In this case, the MCMC run has reached stationarity: there are no obvious trends in the plot which would suggest that the MCMC has not yet converged, and there are no large-scale fluctuations in the trace which would suggest poor mixing.

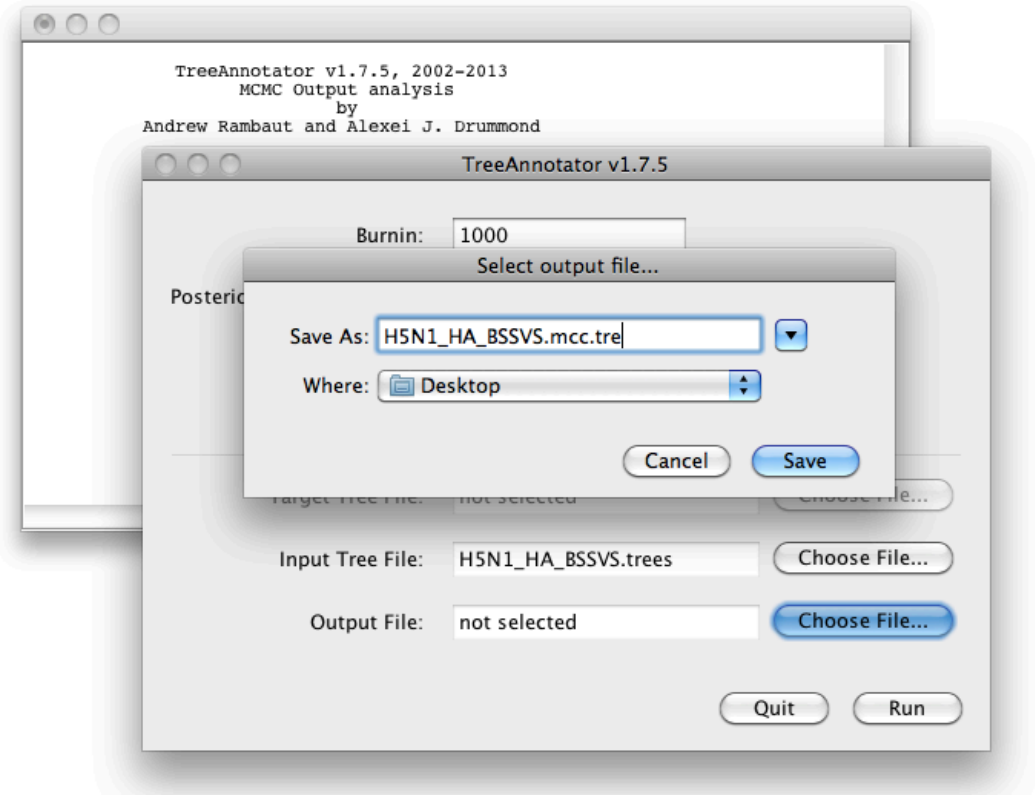


Once are happy with the behavior of posterior probability we can move on to summarize the annotated phylogeographic tree inferred with the BSSVS procedure and to estimate the most significant rates of diffusion. Note that the BSSVS procedure might affect root estimation. If you are only interested in summarizing the Bayes Factor rates from the BSSVS analysis and not in summarizing the tree from your run, jump to the last section of this tutorial entitled **Identifying well-supported BF rates using Bayes factor test in SPREAD**. If you are also interested in summarizing the tree, continue to next section.

Summarizing the trees

We have seen how we can diagnose our MCMC run using Tracer and produce estimates of the marginal posterior distributions of parameters of our model. However, BEAST also samples trees (either phylogenies or genealogies) at the same time as the other parameters of the model. These are written to a separate file called the `trees' file. This file is a standard NEXUS format file. As such it can easily be loaded into other software in order to examine the trees it contains. One possibility is to load the trees into a program such as MrBayes or PAUP* and construct a consensus tree in a similar manner to summarizing a set of bootstrap trees. In this case, the support values reported for the resolved nodes in the consensus tree will be the posterior probability of those clades.

In this tutorial, however, we are going to use a tool that is provided as part of the BEAST package to summarize the information contained within our sampled trees. The tool is called **TreeAnnotator** and once running, you will be presented with a window like the one below.



TreeAnnotator takes a single `target' tree and annotates it with the summarized information from the entire sample of trees. The summarized information includes the average node ages (along with the HPD intervals), the posterior support and the average rate of evolution on each branch (for models where this can vary). The program calculates these values for each node or clade observed in the specified 'target' tree.

- Burnin - This is the number of trees in the input file that should be excluded from the summarization. This value is given as the number of trees rather than the number of steps in the MCMC chain. Thus for the example above, with a chain of 1,000,000 steps, sampling every 500 steps, there are 10,000 trees in the file. To obtain a 10% burn-in, set this value to 1,000.

  We have 1.000.000 states, generations, sampled every 100th. Therefoe, there are 10.000 trees; we can remove 25%: Burnin = 2500 trees

- Posterior probability limit - This is the minimum posterior probability for a node in order for TreeAnnotator to store the annotated information. The default is 0.0 so every node, no matter what its support, will have information summarized. Make sure this value remains 0.0 as every node will require location annotation for further visualization.

- Target tree type - This has two options "Maximum clade credibility" or "User target tree". For the latter option, a NEXUS tree file can be specified as the Target Tree File, below. For the former option, TreeAnnotator will examine every tree in the Input Tree File and select the tree that has the highest sum of the posterior probabilities of all its nodes.

- Node heights - This option specifies what node heights (times) should be used for the output tree. If the ``Keep target heights'' is selected, then the node heights will be the same as the target tree. The other two options give node heights as an average (Mean or Median) over the sample of trees. Keep the default median node heights for the time being.
- Target Tree File - If the "User target tree" option is selected then you can use "Choose File..." to select a NEXUS file containing the target tree.
- Input Tree File - Use the "Choose File..." button to select an input trees file. This will be the trees file produced by BEAST.
- Output File - Select a name for the output tree file (e.g., H5N1_HA_BSSVS_mcc.tre).

Once you have selected all the options above, press the "Run" button. TreeAnnotator will analyze the input tree file and write the summary tree to the file you specified. This tree is in standard NEXUS tree file format so may be loaded into any tree drawing package that supports this. However, it also contains additional information that can only be displayed using the FigTree program.

Viewing the annotated tree

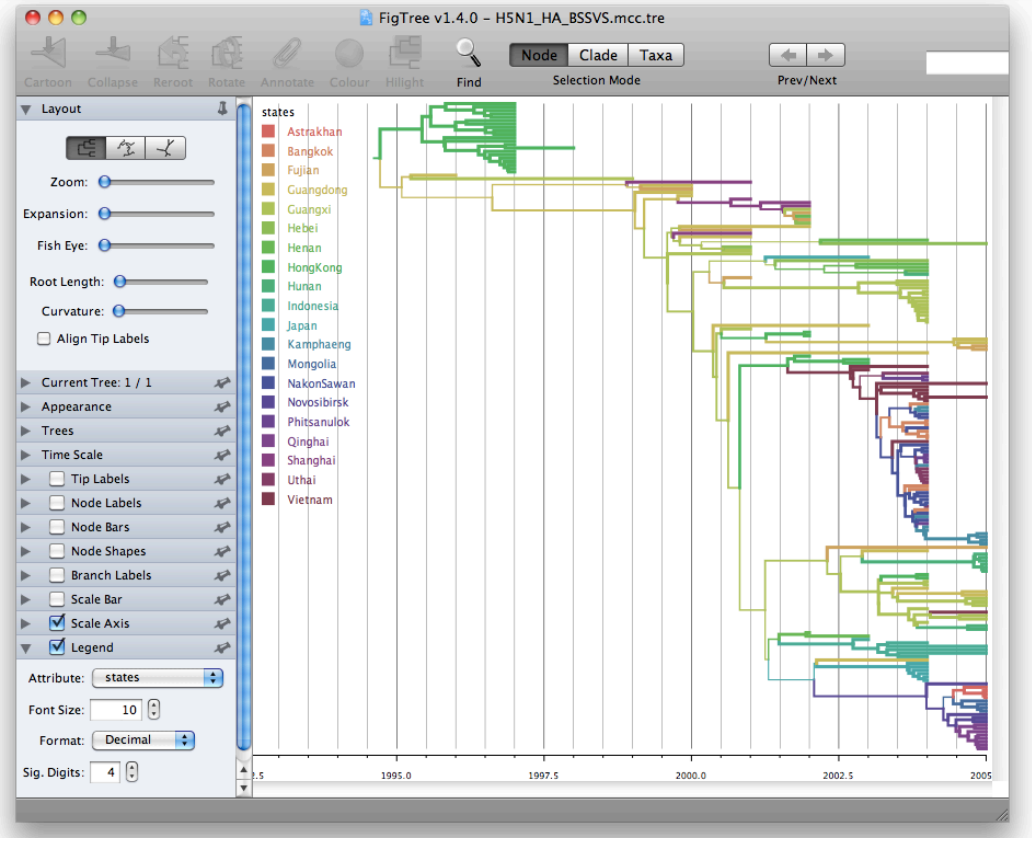
Run FigTree and select the Open... command from the File menu. Select the tree file you created using TreeAnnotator in the previous section. The tree will be displayed in the FigTree window. On the left hand side of the window are the options and settings which control how the tree is displayed. In this case we want to display the posterior probabilities of each of the clades present in the tree and estimates of the age of each node. In order to do this you need to change some of the settings.

First open the Branch Labels section of the control panel on the left. Now select posterior from the Display popup menu. The posterior probabilities won't actually be displayed until you tick the check-box next to the Branch Labels title.

We now want to display bars on the tree to represent the estimated uncertainty in the date for each node. TreeAnnotator will have placed this information in the tree file in the shape of the 95% highest posterior density (HPD) intervals (see the description of HPDs, above). Open the Node Bars section of the control panel and you will notice that it is already set to display the 95% HPDs of the node heights so all you need to do is to select the check-box in order to turn the node bars on. We can also plot a time scale axis for this evolutionary history (select 'Scale Axis' and deselect 'Scale bar'). For appropriate scaling, open the 'Time Scale' section of the control panel, set the 'Offset' to 2005 (date of the most recent sample), the scale factor to -1.0. and 'Reverse Axis' under 'Scale Axis'.

It is +1.0

Open the Appearance panel and alter the Line Weight to 2 in oder to draw the tree with thicker lines. Under the same panel, alter Colour by and select States. Click on Setup Colours and set Primary to Saturation and Secondary count to 20, which is the total number of locations. Still in the Appearance panel, select Width by states.prob, with a Min weight of 1. This will allow the width of tree branches to be proportional to the inferred ancestral state probability. Unselect the Tip Labels and the Scale Bar option. Finally, in the Legend panel select Attribute states. None of the options actually alter the tree's topology or branch lengths in anyway so feel free to explore the options and settings. You can also save the tree and this will save all your settings so that when you load it into FigTree again it will be displayed exactly as you selected.



Identifying well-supported BF rates using Bayes factors test in SPREAD

SPREAD (Spatial Phylogenetic Reconstruction of EvolutionAry Dynamics) is a software to visualize the output from Bayesian phylogeographic analysis. SPREAD comes with its own map and virtual globe software, and it is able to generate KML files to visualize the output in GoogleEarth. Some of the functionalities of **SPREAD** that relate to the discrete phylogeographic analysis performed previously include visualizing location-annotated MCC trees, generation of KML output files for **Google Earth** and identification of well-supported rates using Bayes Factor test. The later option takes as input the rate matrix file (H5N1_HA_BSSVS.locations.rates.log) generated under the analysis using the Bayesian Stochastic Search Variable Selection (**BSSVS**) procedure. This test aims at identifying frequently invoked rates to explain the diffusion process and visualize them using SPREAD's own map and in virtual globe software. A detailed tutorial for this particular step is available at http://www.phylogeography.org/tutorial/spread_tutorial.html#toc-Section-3.

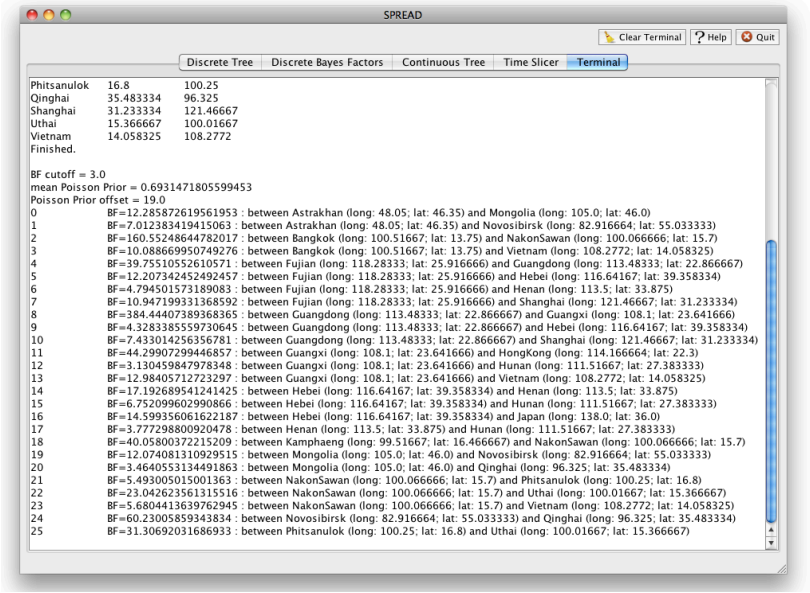
Briefly, go to the **Discrete Bayes Factor** menu and using *Load log file* upload the output BEAST file containing the spatial rates and rate indicators (H5N1_HA_BSSVS.locations.rates.log). To visualize the results in the log file, a tab delimited file with location names and corresponding latitude and longitude coordinates needs to be uploaded using the *Load locations file*. The locations file should look like this:

```
Astrakhan       46.35   48.05
Bangkok         13.75   100.5166667
Fujian  25.91666667     118.2833333
Guangdong       22.86666667     113.4833333
```

...
Qinghai 35.48333333 96.325
Shanghai 31.23333333 121.4666667
Uthai 15.36666667 100.0166667
Vietnam 14.058325 108.277197

Click done to upload your location file (locations.txt). The next step is to set up visualization attributes. Here you can specify the burn-in (default is 10%), the Poisson prior mean and offset, the Bayes factor cut-off, the color of the mapped rates, the KML name and more technical attributes. Once this is done, go to *Generate KML / Plot map* and click on *Plot* to visualize the Bayes factor rates in a map and click on *Generate* to save a KML output file that can be further inspected using Google Earth.

Finally, go to the menu **Terminal** to visualize the values of the Bayes factors for the rates that achieve a support beyond the specified cut-off and the respective locations involved for these rates. Note that for the symmetrical model the order of the locations, between 'X' and 'Y', is arbitrary as there is no directionality in this case.



Conclusion and Resources

This tutorial only scratches the surface of the analyses that are possible to undertake using BEAST. It has hopefully provided a relatively gentle introduction to the fundamental steps that will be common to all BEAST analyses and provide a basis for more challenging investigations. BEAST is an ongoing development project with new models and techniques being added on a regular basis. The BEAST website provides details of the mailing list that is used to announce new features and to discuss the use of the package. The website also contains a list of tutorials and recipes to answer particular evolutionary questions using BEAST as well as a description of the XML input format, common questions and error messages.

- The BEAST website: **http://beast.bio.ed.ac.uk**/
- Tutorials: **http://beast.bio.ed.ac.uk/Tutorials/**
- Phylogeography: **http://www.phylogeography.org** (includes **SPREAD** and tutorial)
- Frequently asked questions: **http://beast.bio.ed.ac.uk/FAQ/**