**Bayes Factor and Discrete Trait Analysis**

**Required programs:**

BEAUTI v.1.10

Tracer v1.7.1

SpreaD3: https://rega.kuleuven.be/cev/ecv/software/SpreaD3

The program SpreaD3 can be used to analyze the transmission rates between different states in a phylogeographic model. The discrete trait analysis can be done for non-geographic data to look at the relationships and networks between different character states such as Host or Age group as non-geographic examples. In this tutorial, we will show you how to use the output from BEAST to calculate the Bayes Factor and the posterior probability for a transition between two states.

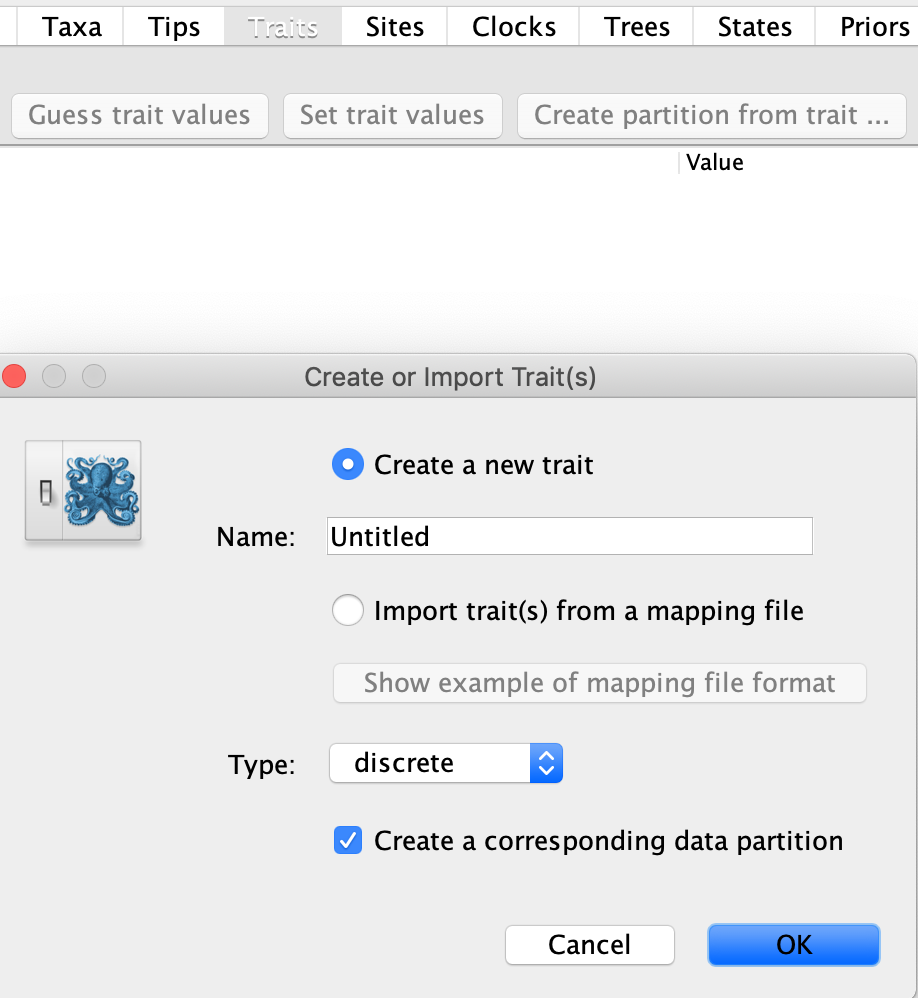
In this tutorial, the files being used can be found in the CEIRS-Training-Taiwan-2019/CIERStraining-Asia2017HA/BEAST/Phylogeography folder on GitHub. This contains a simple fasta file and a coordinates file for the geographic states found in the associated fasta file.

**Pre-BEAST**

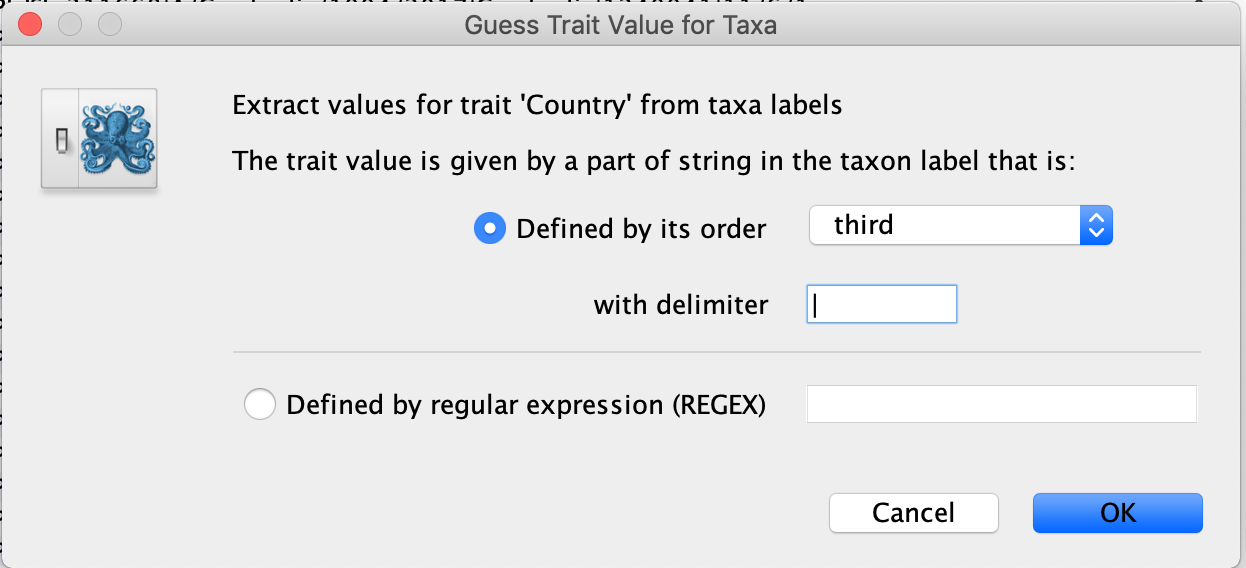
When creating your XML in BEAUTI the procedure is very similar to when creating it for a normal BEAST run, there are different options that can be chosen that will allow for a discrete trait analysis and an ancestral state reconstruction. It is important that your sequence labels are properly labeled with the metadata that is important because we will parse this data when creating a new XML, review the Molecular Epidemiology pipeline handout for further information on descriptive label creation.

The file that will be used in this analysis from the Phylogeography folder on the tutorial Github:

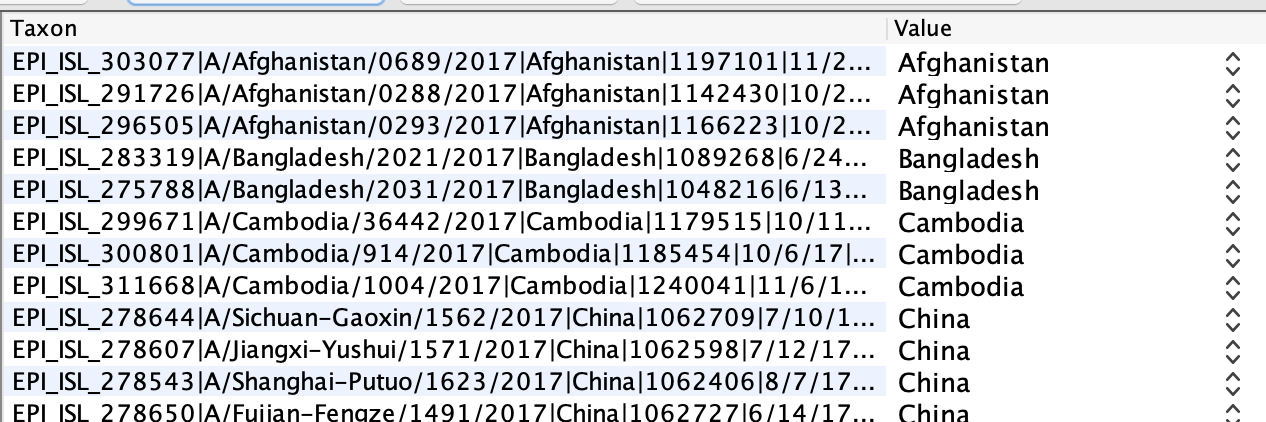
Asia2017H3N2\_subset.pdacut.newstates.fa



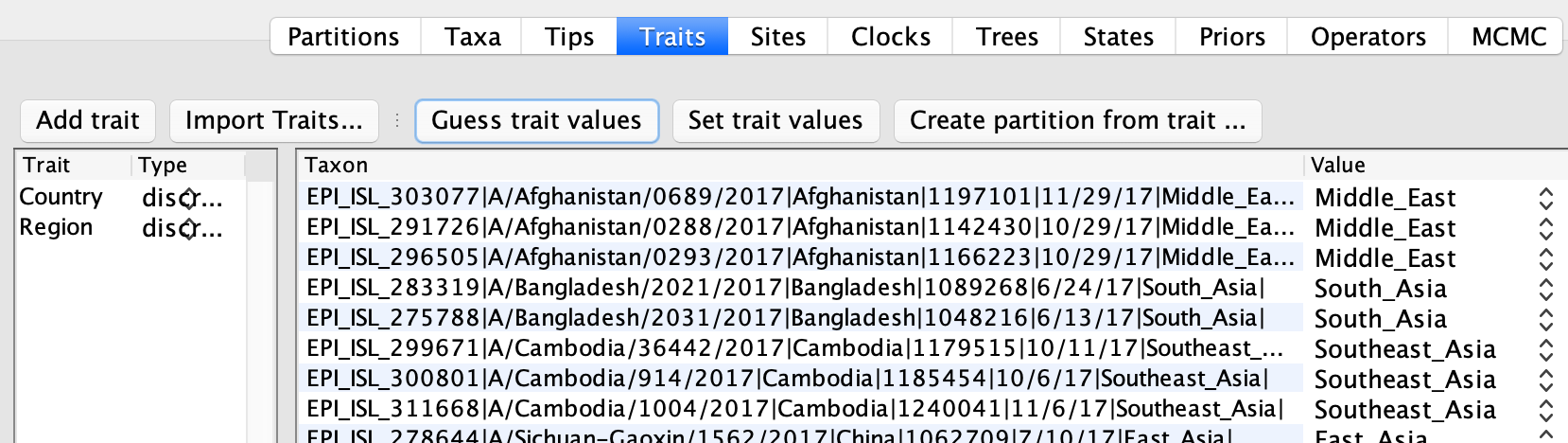
In the traits tab of BEAUTI, select the option for Add trait, this will prompt you to name the trait that you desire to study as seen above. In this tutorial, our data labels give us options to look at both the region and country, for the first trait we will do country. Replace the name “Untitled” with “Country” your trait and click OK. You will then guess the trait, based on the delimited sequence labels we created previously, in this tutorial the provided fasta file has the country name in the third data label position. An example of how to do this is seen below.



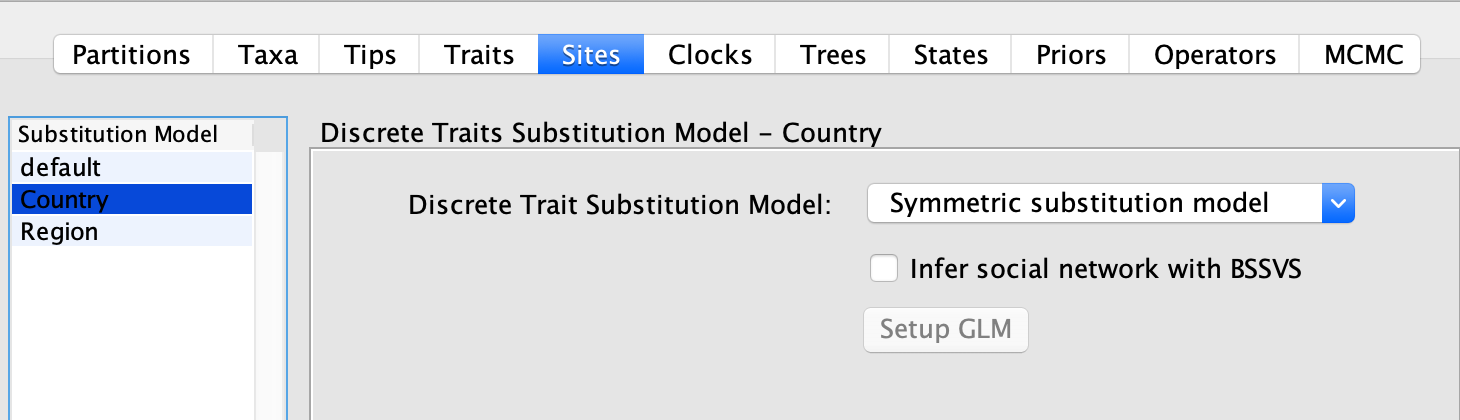
Once the data is parsed the values for the discrete trait will be displayed as seen below.



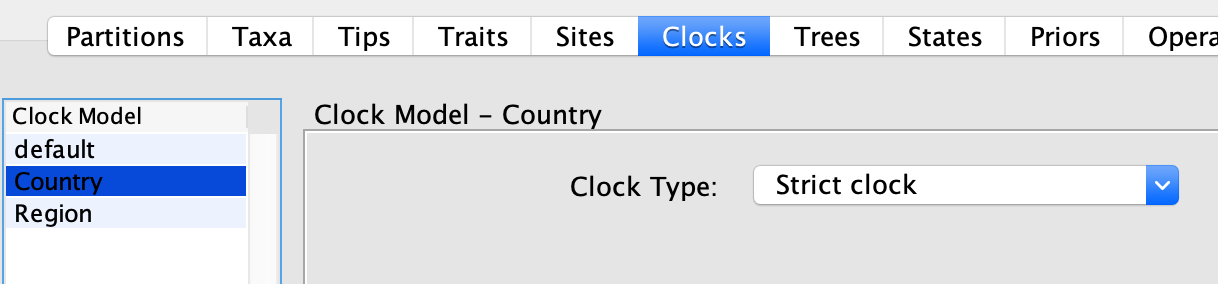
You will now repeat this for our second discrete trait that we are interested in, the “region” category. Repeat the steps above and create the “Region” trait, parse the values based on the correct delimiter (in this file it will be the last data label position) and you should see a window that looks like the one below.



The process of setting up your XML file is very similar to doing a standard BEAST run. In the Tips window make sure that you parse the dates from your sequence labels. In the Sites window you will be able to choose each of your partitions and choose the substitution model. In the default you can choose a standard substitution model such as a GTR or HKY, for the specific discrete traits we will have a different option as seen below. Choose the option seen below for both the Country and Region traits.



In the Clocks window, the default will be an uncorrelated relaxed clock as we have used before. For both the Country and Region you will choose a strict clock as seen below.

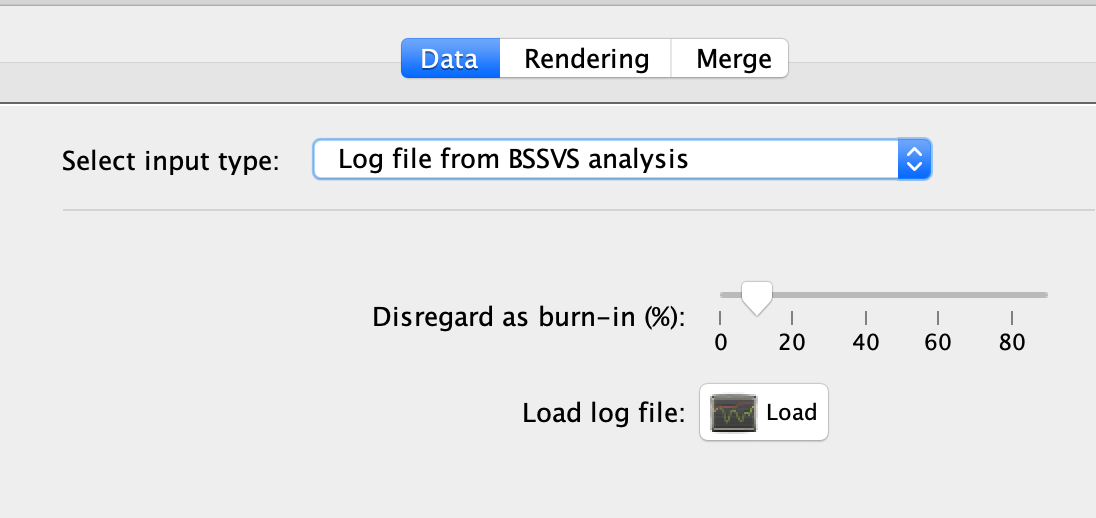


The Trees Window will only give an option to choose a coalescent for your default partition, for this choose any coalescent you want, we have used GMRF Skyride before so choose this if you would like to use a coalescent you have seen before. In the States window, you will have the option to perform an ancestral state reconstruction, this can be a useful analysis to determine what the most probable prior character state taxa was based on the taxa throughout the tree. In the Priors window, the priors will stay the same, we will change the ucld.mean as we have done before to the init values of 0.0033 and Upper and Lower of 1.0 and 0.0 respectively. In the MCMC window choose your chain-length and logging, generate the XML file

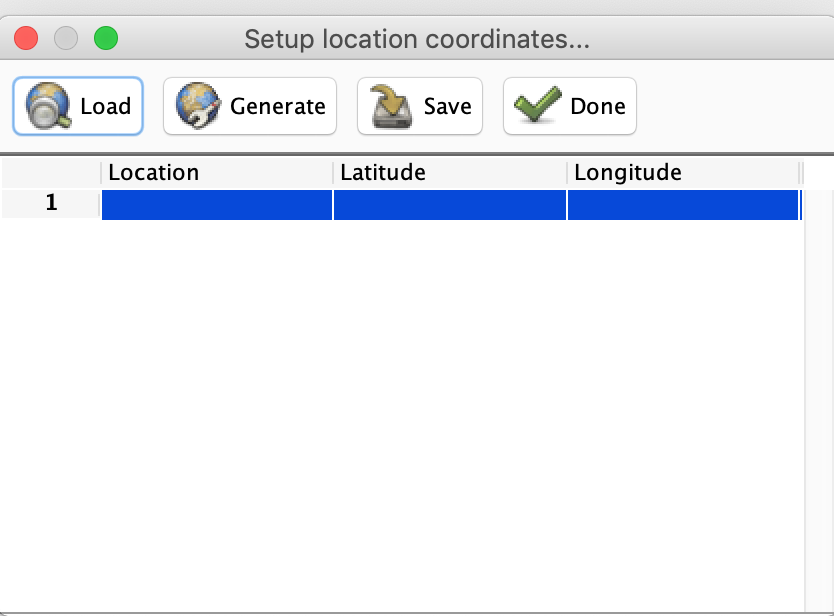
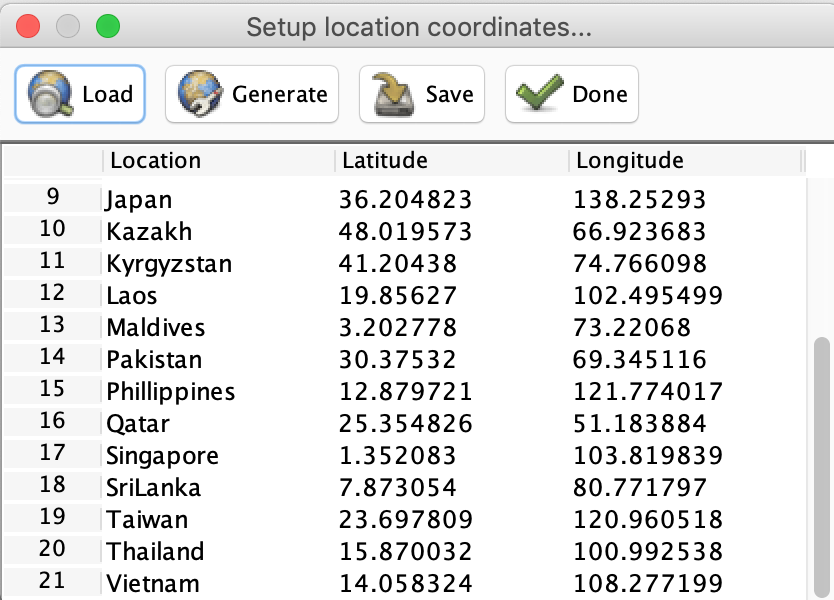
**Calculation of Bayes Factor using SpreaD3**

Before this step ensure that your MCMC chain and BEAST run have converged and that adequate ESS values were achieved using Tracer. If the runs have not converged or achieved adequate ESS values increase chain length or increase number of runs. When checking the convergence of the runs make sure that you are using the .log file, the .rates.log file will be used in SpreaD3 for calculation of Bayes Factors.

If doing multiple runs, combine the .rates.log files for use in SpreaD3 using the program LogCombiner v 1.10 . When opening program SpreD3 choose the drop-down option seen below:

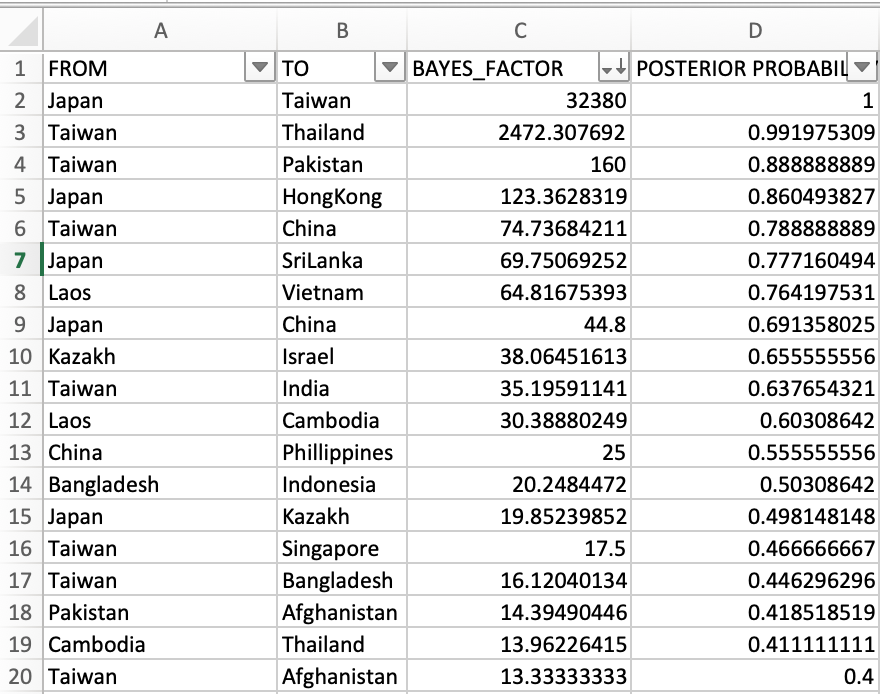


After the rates.log file has been loaded, ensure that the appropriate burn in is set, if burn in was removed for a combined log file from LogCombiner, set the burn in to 0, otherwise set burn in to desired burn in (usually 10%). When prompted setup the location coordinates for your data. Choose the Load option and upload the provided text file with the coordinates for the states being studying in the rate.log file. If only trying to calculate the Bayes Factors this is not necessary, you can choose the generate option and when asked for a number of locations, use the number of states, then enter the names of the sates in the locations column manually in the order that they are found in the states section of your XML file. The latitude and longitude values can be random if only trying to calculate the Bayes factor, if trying to visualize phylogeography on a map, the coordinate data is important.

Once the location data is set, there will be options for the Poisson prior and Poisson prior offset. The relevant values for this can be found in your XML file used for BEAST by searching for “Poisson” and looking for the prior and offset data values. Once this is done, you will be able to generate an output that will come in the form of tab separated text file.

Open the tab file in excel and filter by the Bayes factor as seen below (filter in descending order).



The following scale be used when determining whether a Bayes Factor is significant and supports a certain transition:

No support BF < 3

Substantial support BF = 3–10

Strong support BF = 11–30

Very strong support BF = 31–100

Decisive support BF > 100