Notes Assisting BEAST Tutorials

This notes have some extra details about aspects of some of the programs.

Guess 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 (such as the RSVA sequences in the MEP tutorial) then you can specify how to find the one that corresponds to the date of sampling. You can either specify the order that the date field comes (e.g., first, last or various positions in between) or specify a prefix (some characters that come immediately before the date field in each name). For the RSVA sequences you can select 'last' from the drop-down menu for the order or use the prefix option and specify '_' (underscore) as the prefix.

In this dialog box, you can also get BEAUti to add a fixed value to each guessed date. In this case the value "1900" has been 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". Some of the sequences in the example file actually have dates after the year 2000 so selecting the will option would convert them correctly, adding 2000 to any date less than 09. When you press OK the dates will appear in the appropriate column of the main window. You can then check these and edit them manually as required. 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 1984) or backwards in time from the present (as in the case of radiocarbon ages).

Tracer statistics

The statistics reported in Tracer for each logged quantity are:

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

- Stdev - 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). 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.

TreeAnnotator

The sampled trees in BEAST 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 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.

TreeAnnotator is a software program distributed with BEAST that can summarize the tree file. It 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. The options in TreeAnnotator are detailed below:

- **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 1000 steps, there are 1000 trees in the file. To obtain a 10% burnin, set this value to 100.

- **Posterior probability limit** - This is the minimum posterior probability for a node in order for TreeAnnotator to store the annoted information. The default is 0.5 so only nodes with this posterior probability or greater will have information summarized (the equivalent to the nodes in a majority-rule consensus tree). Set this value to 0.0 to summarize all nodes in the target tree.

- **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.

- **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.

Once you have selected all the options, above, press the "Run" button.