



# BEAST v1.X tutorial, in a case of four cicada genera

Soichi Osozawa<sup>1</sup>

<sup>1</sup>KawaOso Molecular Bio-Geology Institute

Jan 05, 2021

1 Works for me dx.doi.org/10.17504/protocols.io.bq6mmzc6

KawaOso

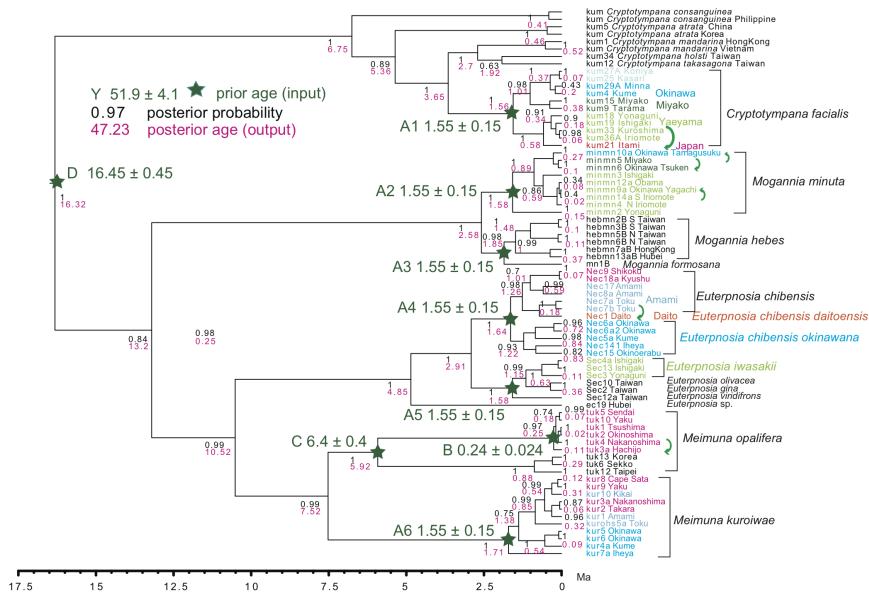
kawaoso

## ABSTRACT

A Bayesian inference (BI) tree is constructed using the software BEAST v.1X (Bayesian Evolutionary Analysis Sampling Trees; Suchard et al., 2018; released on 10th June, 2018), running BEAUti, BEAST, TreeAnnotator, and FigTree, in ascending order. Before operating the BEAST software, the BEAGLE Library must be downloaded. Platform software is BEAST, and the calculation time is less than 30 minutes for the present case, but usually calculation time is up to 2 when specimens are increased.

BEAST v1.X has an important function of calibration in the associated software of BEAUti (Bayesian Evolutionary Analysis Utility). The calibration protocol of BEAUti is to input a time of the most recent common ancestor (tMRCA = minimum age) of the ingroup species, either by fossil calibration or by the geological event calibration. Using BEAST v1.X, without fixing maximum age constraint, we can simply build a reliably-dated phylogeny.

Another important function is combined gene analysis of the mitochondrial COI (1,534 bp) and the nuclear 28S rRNA (874 bp) in the present case, and the analysis can be performed by simply setting "Partitions" of these genes in BEAUti. The resolution was sufficient, but note that nuclear gene does not severely affect on the final topology. Concatenation and re-partitioning using PartitionFinder is not required. Note that BEAST2 analyzes each gene separately, and later combined tree by LogCombiner, but the consequent tree is sometimes unacceptable.



DOI

dx.doi.org/10.17504/protocols.io.bq6mmzc6

## PROTOCOL CITATION

Soichi Osozawa 2021. BEAST v1.X tutorial, in a case of four cicada genera. [protocols.io](https://doi.org/10.17504/protocols.io.bq6mmzc6)  
<https://dx.doi.org/10.17504/protocols.io.bq6mmzc6>



## KEYWORDS

null, BEAST v1X. calibration. combined gene analysis

## LICENSE

This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

## CREATED

Jan 05, 2021

## LAST MODIFIED

Jan 05, 2021

## PROTOCOL INTEGER ID

45997

## DISCLAIMER:

### DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK

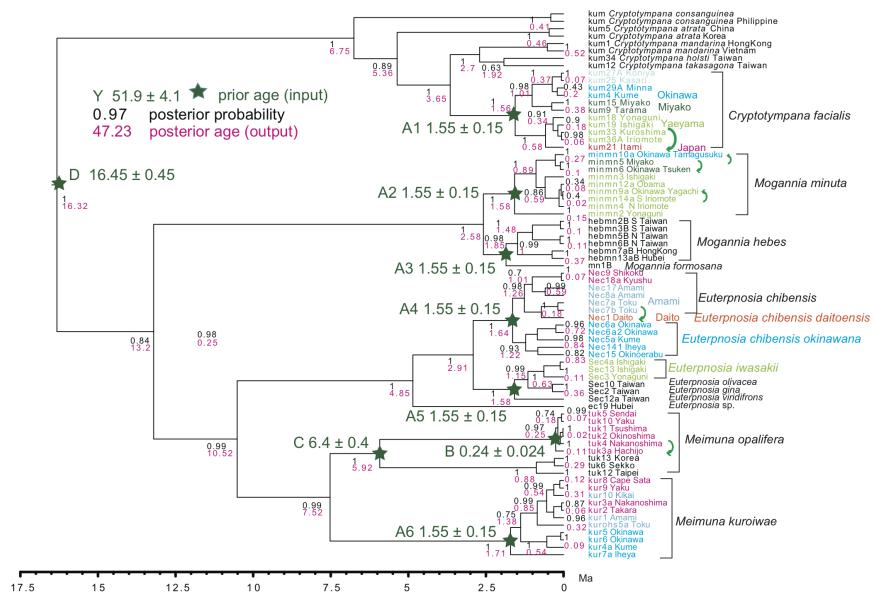
The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to [protocols.io](#) is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with [protocols.io](#), can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

## ABSTRACT

A Bayesian inference (BI) tree is constructed using the software BEAST v.1X (Bayesian Evolutionary Analysis Sampling Trees; Suchard et al., 2018; released on 10th June, 2018), running BEAUti, BEAST, TreeAnnotator, and FigTree, in ascending order. Before operating the BEAST software, the BEAGLE Library must be downloaded. Platform software is BEAST, and the calculation time is less than 30 minutes for the present case, but usually calculation time is up to 2 when specimens are increased.

BEAST v1.X has an important function of calibration in the associated software of BEAUti (Bayesian Evolutionary Analysis Utility). The calibration protocol of BEAUti is to input a time of the most recent common ancestor (tMRCA = minimum age) of the ingroup species, either by fossil calibration or by the geological event calibration. Using BEAST v1.X, without fixing maximum age constraint, we can simply build a reliably-dated phylogeny.

Another important function is combined gene analysis of the mitochondrial COI (1,534 bp) and the nuclear 28S rRNA (874 bp) in the present case, and the analysis can be performed by simply setting “Partitions” of these genes in BEAUti. The resolution was sufficient, but note that nuclear gene does not severely affect on the final topology. Concatenation and re-partitioning using PartitionFinder is not required. Note that BEAST2 analyzes each gene separately, and later combined tree by LogCombiner, but the consequent tree is sometimes unacceptable.



## 1 General flow

A Bayesian inference (BI) tree is constructed using the software BEAST v.1X (Bayesian Evolutionary Analysis Sampling Trees; Suchard et al., 2018; released on 10th June, 2018), running BEAUti, BEAST, TreeAnnotator, and FigTree, in ascending order. Before operating the BEAST software, the BEAGLE Library must be downloaded. Platform software is BEAST, and the calculation time is less than 30 minutes for the present case, but usually calculation time is up to 2 when specimens are increased.

BEAST v1.X has an important function of calibration in the associated software of BEAUti (Bayesian Evolutionary Analysis Utility). The calibration protocol of BEAUti is to input a time of the most recent common ancestor (tMRCA = minimum age) of the ingroup species, either by fossil calibration or by the geological event calibration. Using BEAST v1.X, without fixing maximum age constraint, we can simply build a reliably-dated phylogeny.

Another important function is combined gene analysis of the mitochondrial COI (1,534 bp) and the nuclear 28S rRNA (874 bp) in the present case, and the analysis can be performed by simply setting “Partitions” of these genes in BEAUti. The resolution was sufficient, but note that nuclear gene does not severely affect on the final topology. Concatenation and re-partitioning using PartitionFinder is not required. Note that BEAST2 analyzes each gene separately, and later combined tree by LogCombiner, but the consequent tree is sometimes unacceptable.

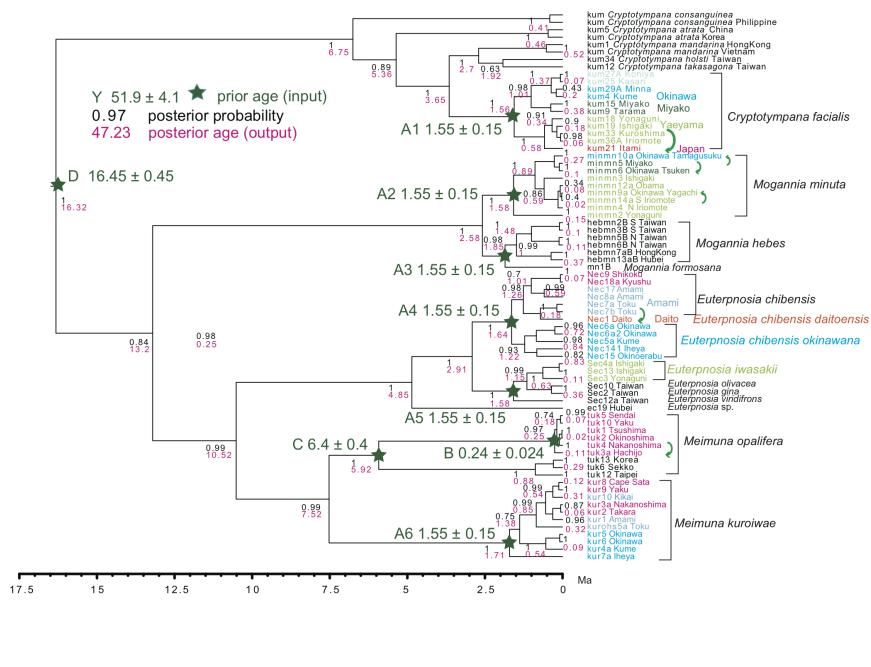
Calibrations points are shown on the consequent BI tree, and these dates were input in “Priors” in BEAUti; they are summarized below. See references in our paper appeared in PlosOne.

Calibration point A (A1 to A6) is after our geological event calibration that adopts a  $1.55 \pm 0.15$  Ma date based on multiple biostratigraphic and supporting radio-isotopic dates connected to various geologic relationships in the Ryukyu islands region(Osozawa et al., 2012). This geologic event calibration was also used in a previous study of *Platyleura* cicadas by our group (Osozawa et al. 2017a).

Calibration point B: Hachijo oceanic island is a part of the Izu volcanic arc, and we recently estimated the emergent time of Hachijo as an island at 0.24 Ma (Osozawa et al., 2021). This date is applicable for *Meimuna opalifera* on the Hachijo-jima, Izu oceanic islands, and the Japan continental islands.

Calibration point C: Crown *Meimuna opalifera*: Fossil *M. protopalifera* was found from the Itamuro Formation, Tochigi, Japan (Fujiyama 1969; Yoshikawa 2005), and the fission track age of the correlative terrestrial strata of the Nashino Formation, Sendai, is  $6.4 \pm 0.4$  Ma (Fujiwara et al. 2008).

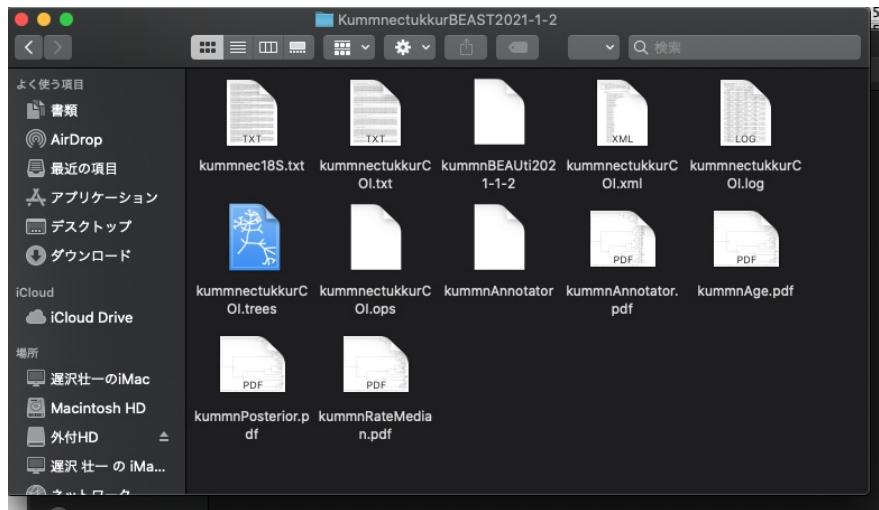
Calibration point D: Crown *Cryptotympana*: Fossil *C. incasa* and *C. miocenica* were found from Shanwang, Shandong, China (Moulds, 2018), and these strata are considered to be time correlative to the European MN5 mammalian stage ( $16.45 \pm 0.45$  Ma; Roc'ek et al. 2011).



BI dated tree

## 2 Prepare .txt files of COI and 28S rRNA, respectively.

Important: label (e.g., >kum12) must be coincident each other between COI and 28S rRNA files. We recommend copy-past commands for labeling.



### 3 Partitions: Loading of COI and 28S rRNA is by using the plus button.

Each sequence of COI and 28S rRNA represents partition.

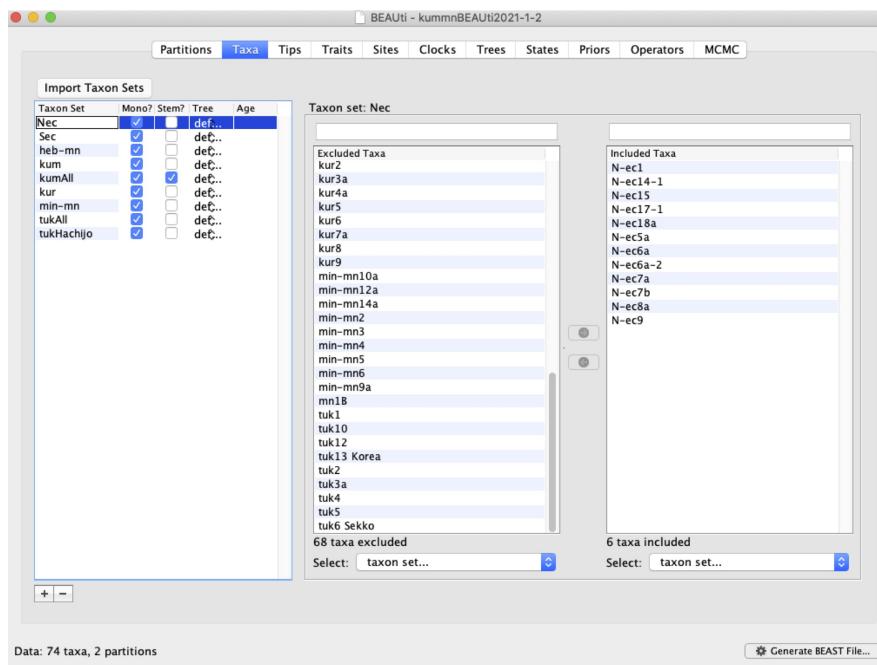
The screenshot shows the BEAUTi software interface with the title bar "BEAUTi - kummnBEAUTi2021-1-2". The top menu bar includes "Partitions", "Taxa", "Tips", "Traits", "Sites", "Clocks", "Trees", "States", "Priors", "Operators", and "MCMC". Below the menu is a toolbar with buttons for "Unlink Subst. Models", "Link Subst. Models", "Unlink Clock Models", "Link Clock Models", "Unlink Trees", and "Link Trees". The main window displays a table of partitions:

| Partition Name  | File Name        | Taxa | Sites | Data Type  | Site Model | Clock Model | Partition Tree |
|-----------------|------------------|------|-------|------------|------------|-------------|----------------|
| kummnecukkurCOI | kummnecukkurC... | 74   | 1534  | nucleotide | default ▲  | default ▲   | default ▲      |
| kummnec18S      | kummnec18S.txt   | 74   | 874   | nucleotide | default ▲  | default ▲   | default ▲      |

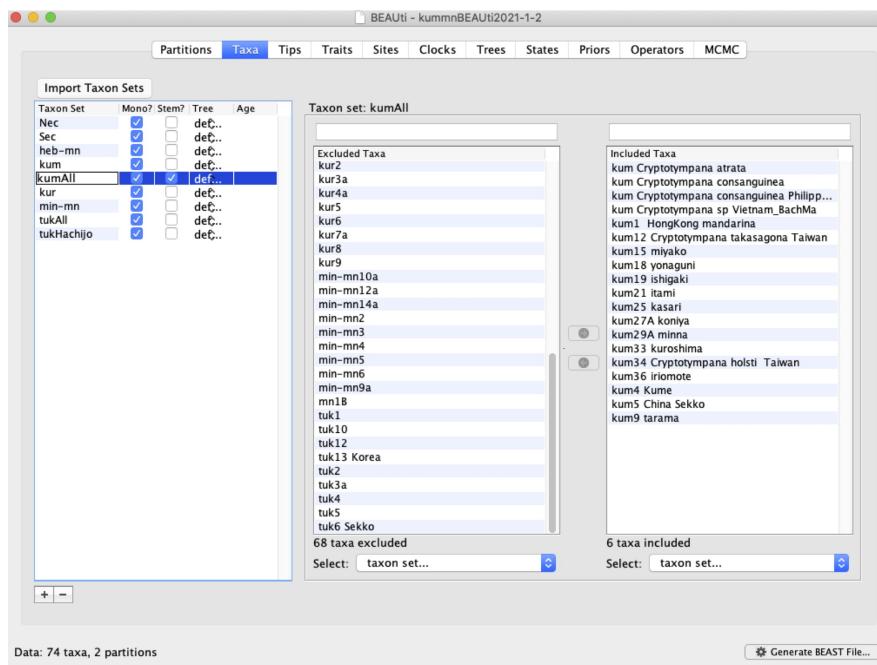
At the bottom of the window, there are buttons for "+", "-", "View Partition ...", "Create partition from trait ...", and "Generate BEAST File...". A status message "Data: 74 taxa, 2 partitions" is displayed at the bottom left, and a "Generate BEAST File..." button is at the bottom right.

### 4 Taxa: Loading of taxa as ingroup was by using the plus button.

The left screen: Taxon Set (monophyletic boxes were checked for all in the present case, and stem box were checked in case by case but solely for kumAll), and the right screen: Included Taxa.



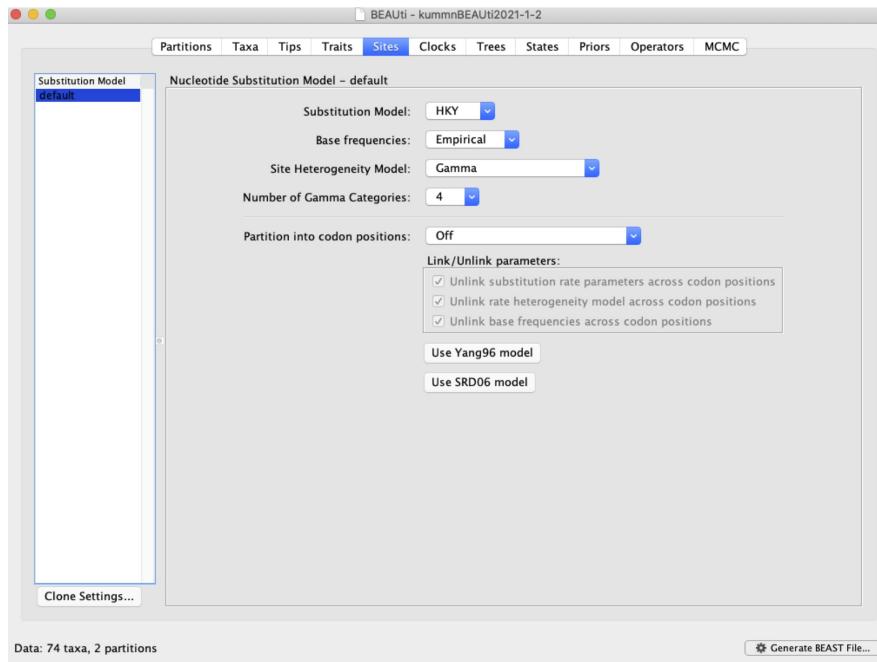
Data: 74 taxa, 2 partitions Generate BEAST File...



Data: 74 taxa, 2 partitions Generate BEAST File...

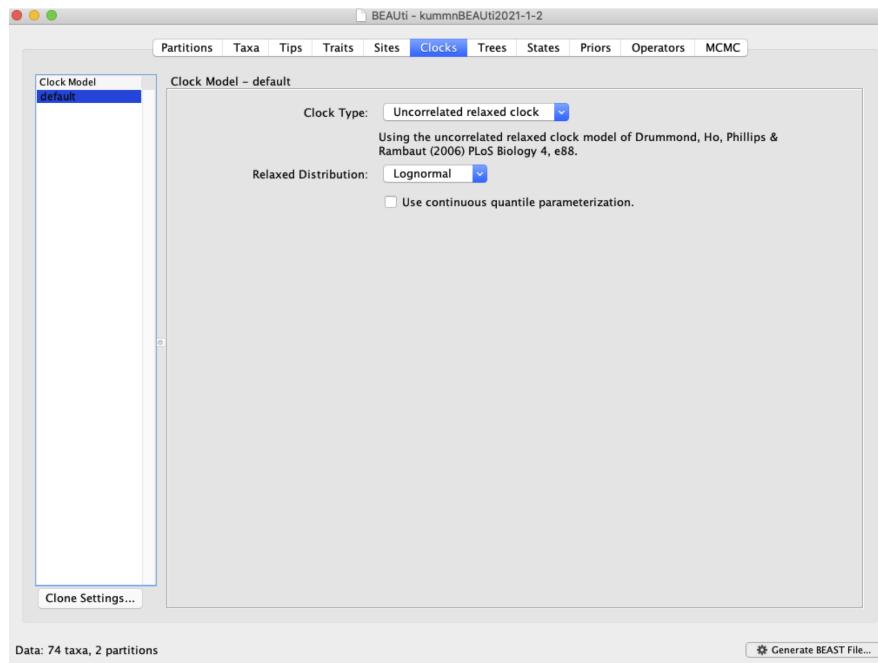
## 5 Sites: Substitution Model: HKY (Hasegawa, Kishino and Yano) model.

Base frequencies: Empirical, Site Heterogeneity Model: Gamma, Number of Gamma Categories: 4, Partition into codon positions: Off. The GTR model generates similar topology.



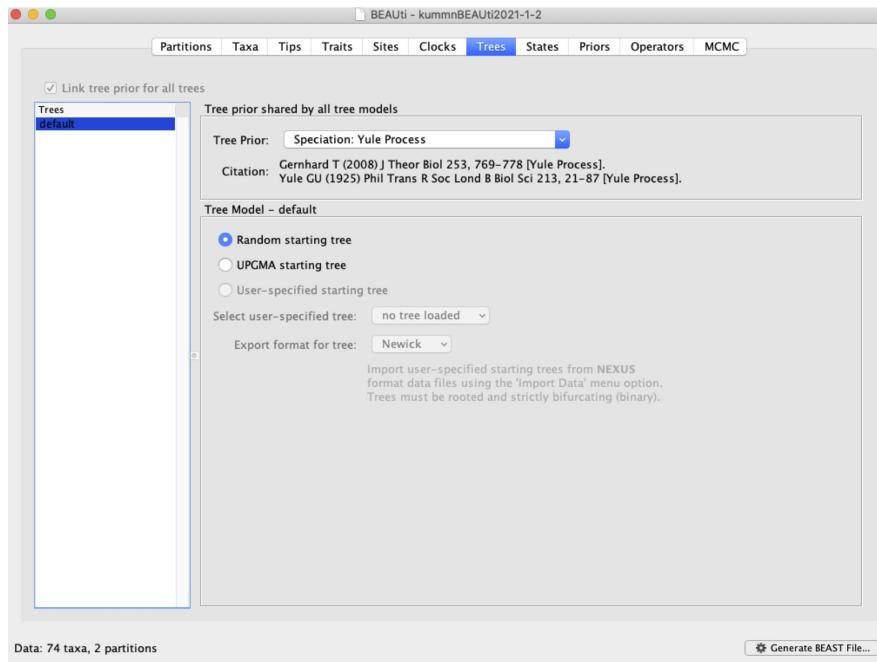
Data: 74 taxa, 2 partitions      [Generate BEAST File...](#)

- 6 Clocks: Clock Type: Uncorrected relaxed clock, Relaxed Distribution: Lognormal.** Uncorrelated relaxed clocks allow each branch of a phylogenetic tree to have its own evolutionary rate under log-normal distribution, and the node rate is the rate median of three branches (Drummond *et al.*, 2006).



Data: 74 taxa, 2 partitions      [Generate BEAST File...](#)

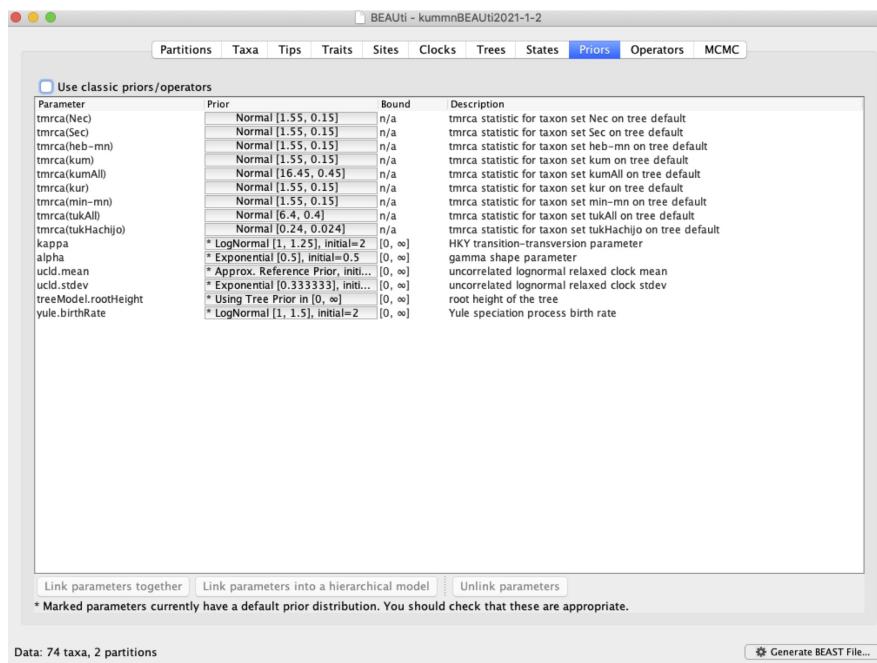
- 7 Trees: Tree Prior: Speciation: Yule Process.**



Data: 74 taxa, 2 partitions

[Generate BEAST File...](#)

## 8 Priors: tmrca (time of MRCA) was input from the calibration point date noted above as Prior Distribution: Normal, and the Mean and Standard deviation.



[Link parameters together](#)

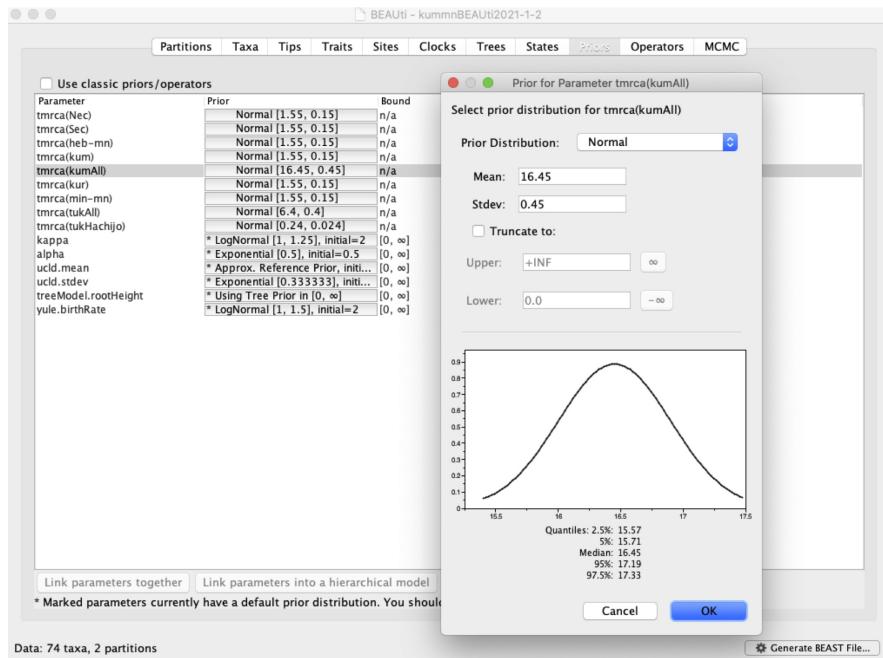
[Link parameters into a hierarchical model](#)

[Unlink parameters](#)

\* Marked parameters currently have a default prior distribution. You should check that these are appropriate.

Data: 74 taxa, 2 partitions

[Generate BEAST File...](#)



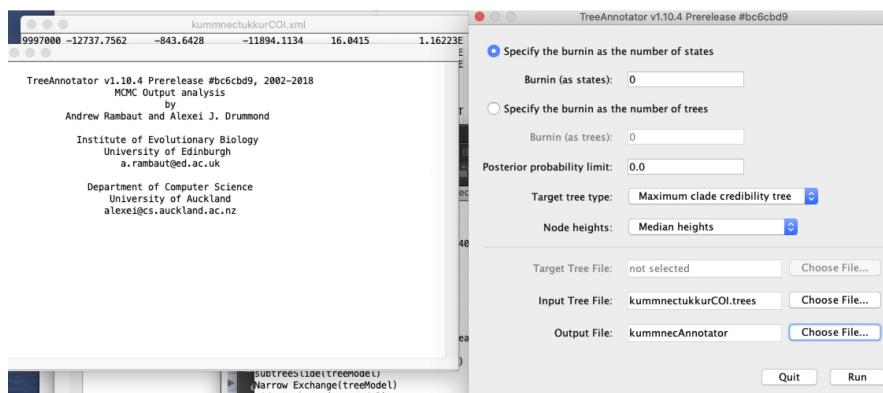
Data: 74 taxa, 2 partitions

[Generate BEAST File...](#)

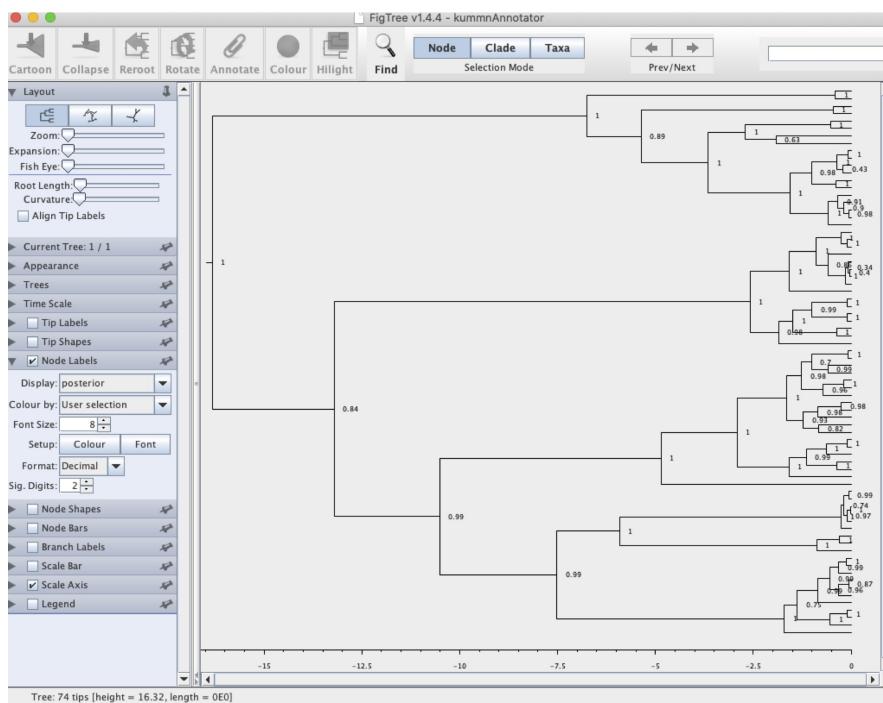
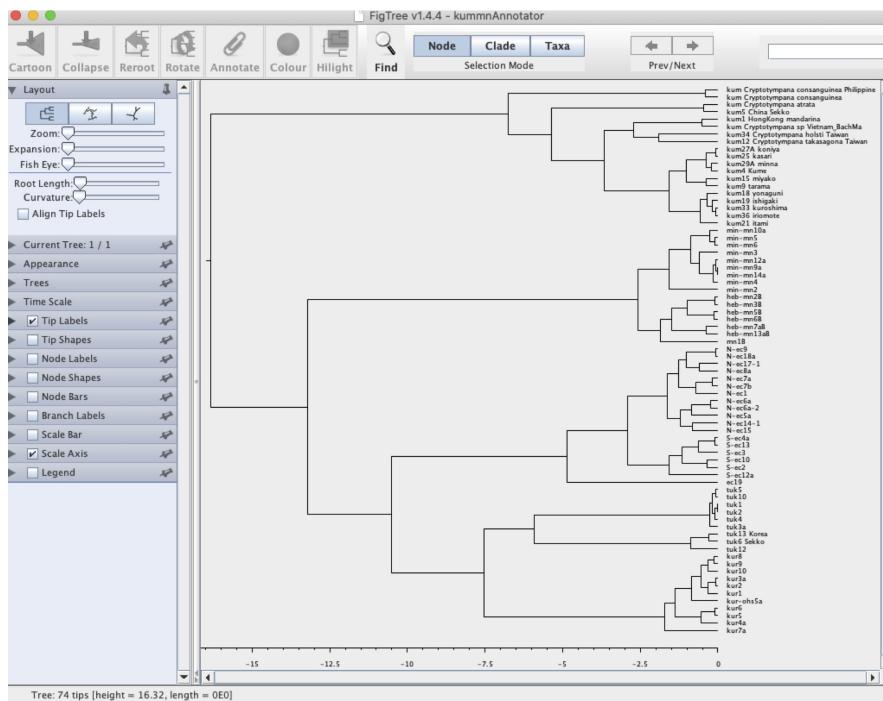
**MCMC: Length of chain: 10,000,000.** Default setting.

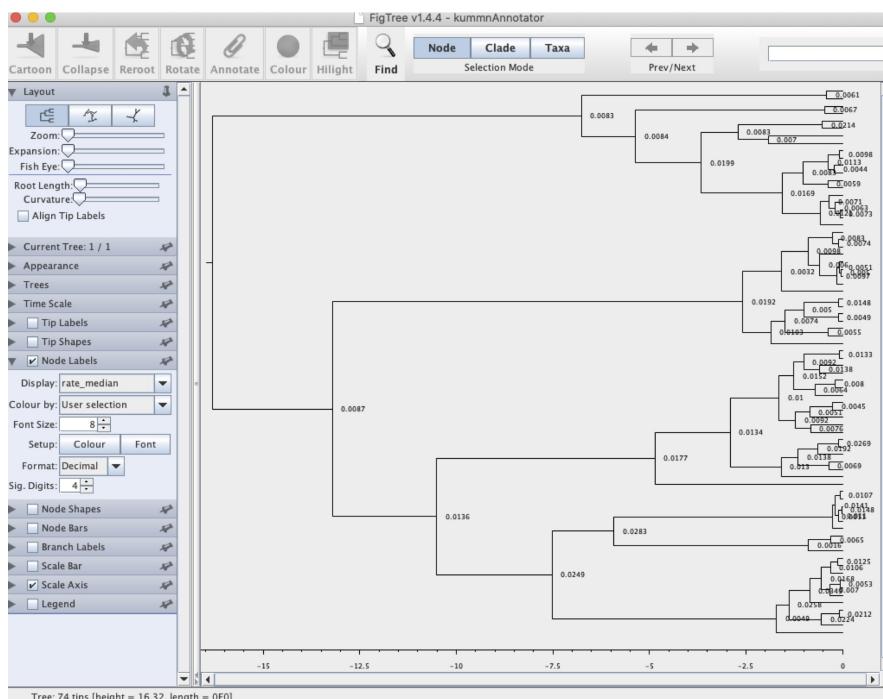
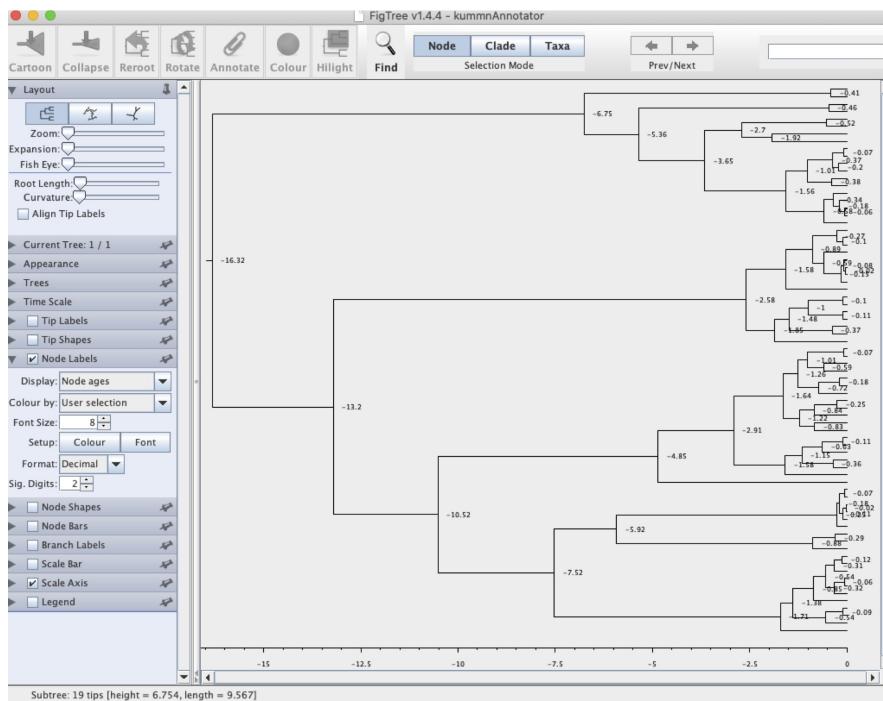
## 9 Running BEAST was done by incorporating xml input file made by BEAUTi.

The consequent tree (~3 hours) was drawn by FigTree v1.4.2, for that, the tree files were input into TreeAnnotator.



- 10 The 95% highest posterior density for confidence intervals of ages can be output in FigTree, but not shown in BI tree above to avoid confusion. In FigTree, posterior probability ("posterior"), posterior age ("Node ages"), and "rate median" (not constant) can be output, and two values are shown at each node in BI tree. This function was not fully used in any previous paper, and we found in our other papers the inconsistent rates through the time as suspected by the relaxed clock model of BEAST (Drummond *et al.*, 2012). Consequently, we made base substitution rate ("rate median" shown at each node in FigTree) vs age ("Node age" shown at each node in FigTree) diagram using function of Excel (not in the present paper).





## 11 Adobe Illustrator