



An Introduction to Molecular Phylogenetic Inference



Kirill Vinnikov

PhD Student,
Department of Biology

ITS Cyberinfrastructure Workshop Series
September 9, 2016

- MCMC diagnostics
- Tree calibration in BEAST

- Tracer (tree.bio.ed.ac.uk/software/tracer)
- BEAST 2.4.3 package of programs (beast2.org)

Bayesian Analysis:

6. Summarize the mcmc search information and build the tree

MrBayes > sump

MrBayes > sumt

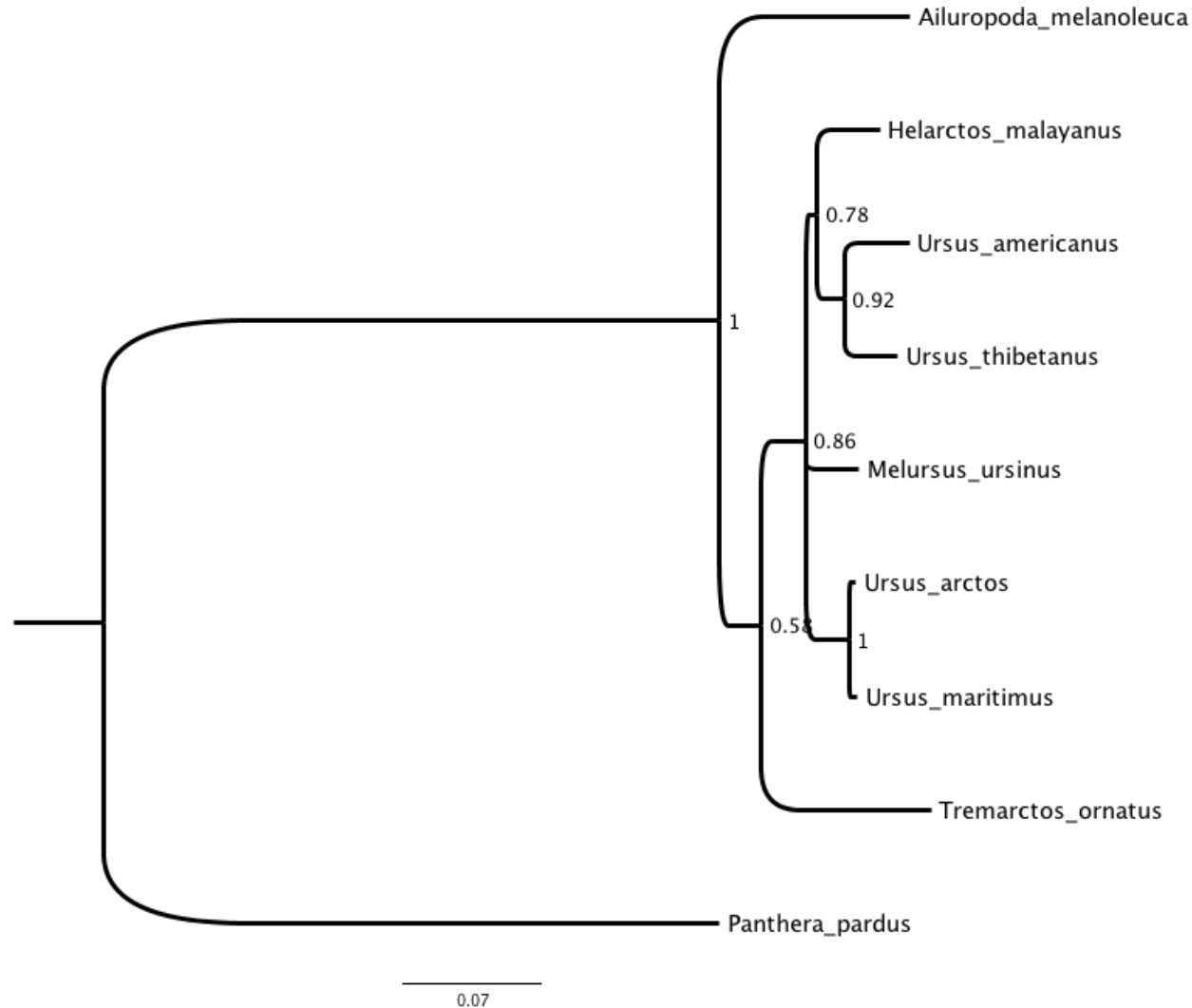
Run	Arithmetic mean	Harmonic mean
1	-8044.99	-8061.31
2	-8045.16	-8060.15
TOTAL	-8045.07	-8060.89

Bayesian Analysis:

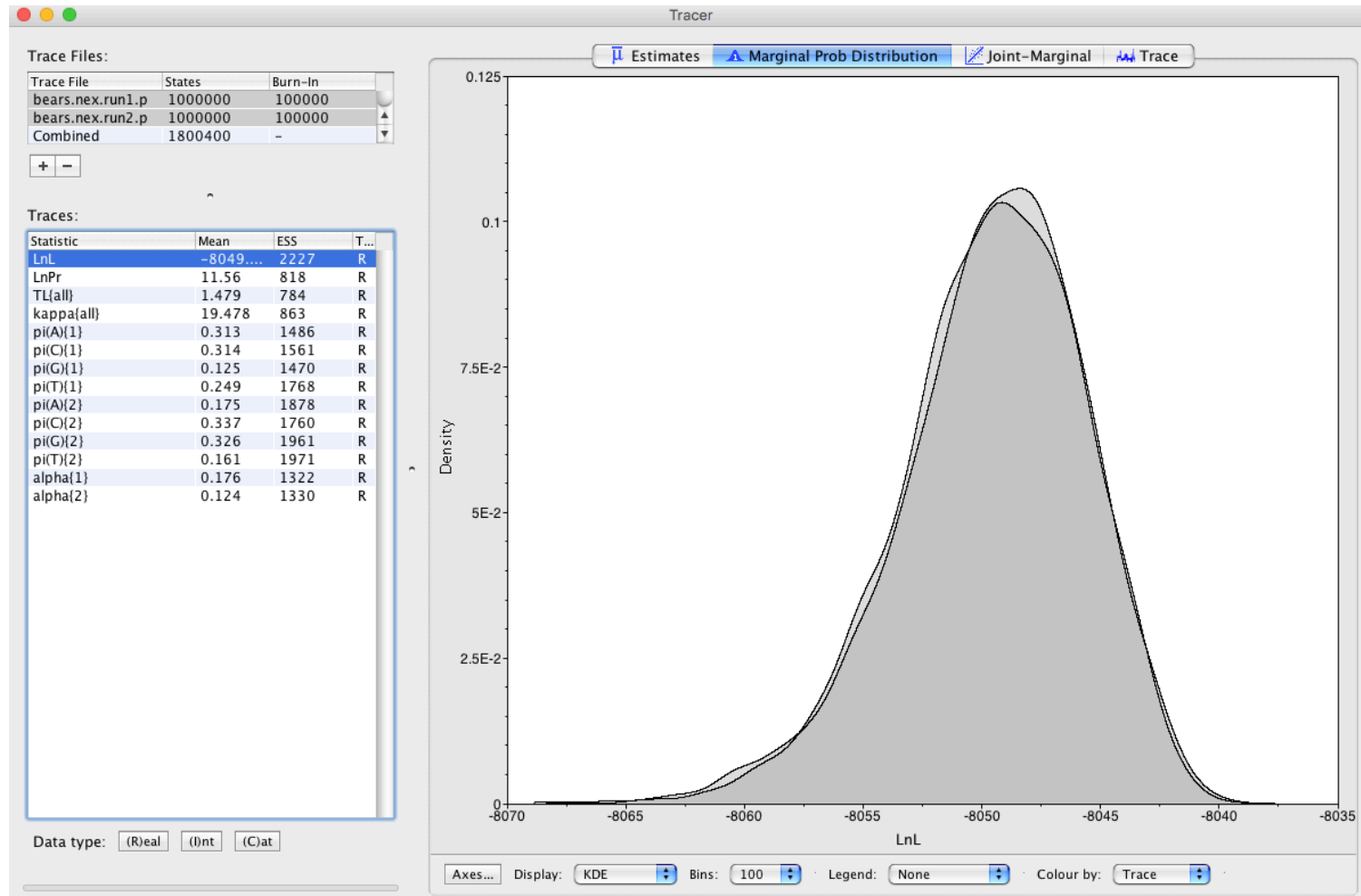
7. Quit the MrBayes

MrBayes > quit

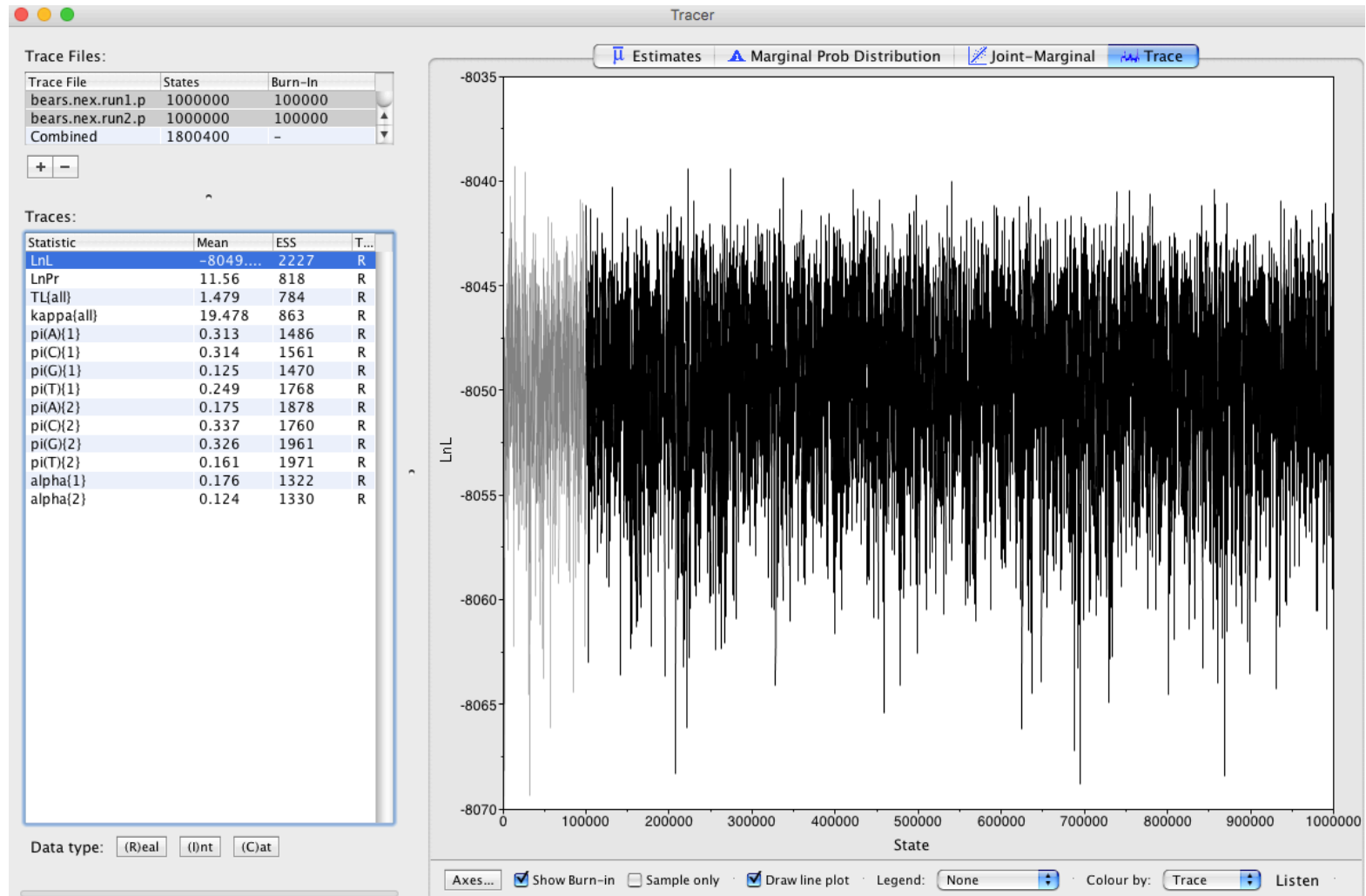
Visualize the bayesian tree in FigTree



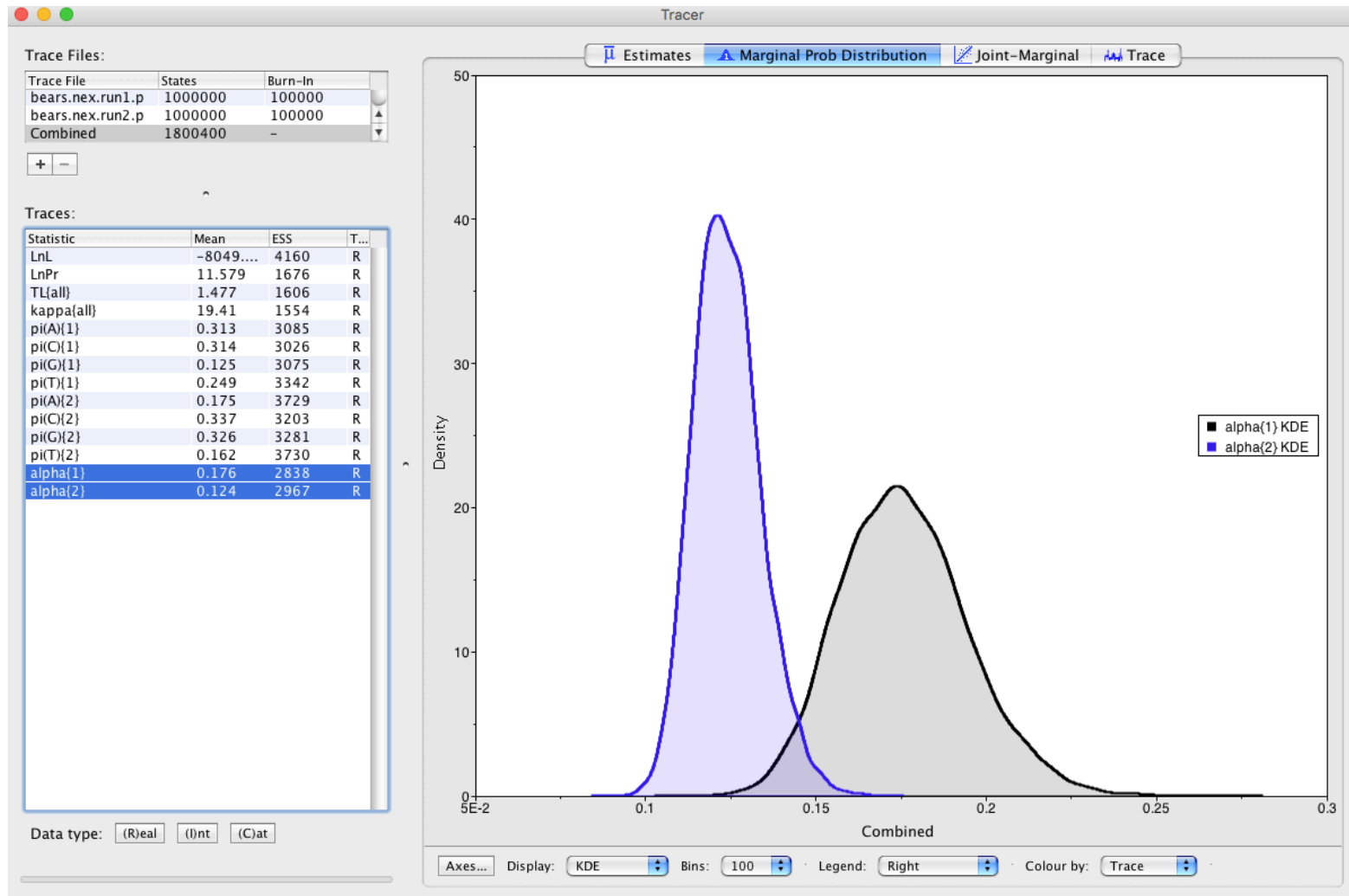
Check the convergence between mcmc runs



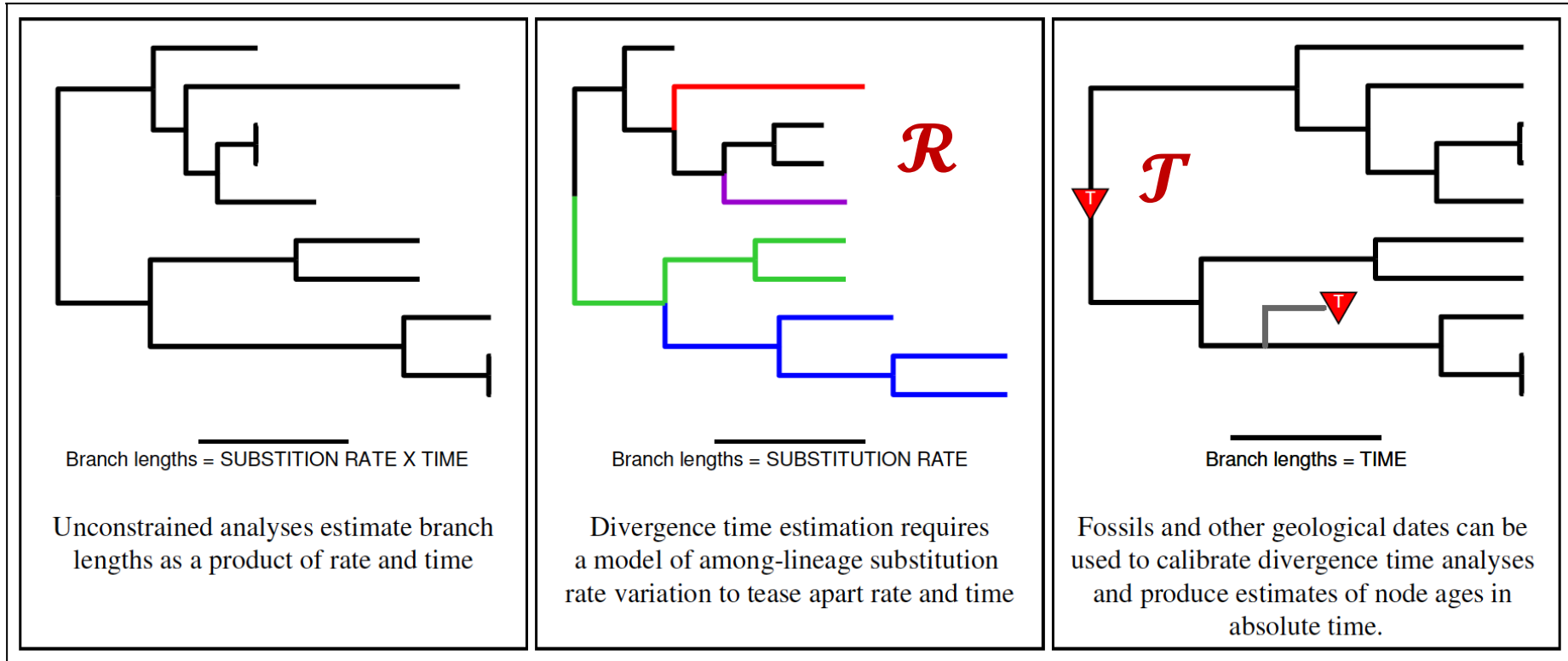
Check the trace of mcmc search



Compare parameter distributions for two genes



Divergence time estimation (adding fossil data)



$$P(\mathcal{R}, \mathcal{T} \mid \text{grid}, \text{fish}) = \frac{P(\text{grid} \mid \mathcal{R}, \mathcal{T}) \times P(\mathcal{R}) \times P(\mathcal{T} \mid \text{fish})}{P(\text{grid} \mid \text{fish})}$$

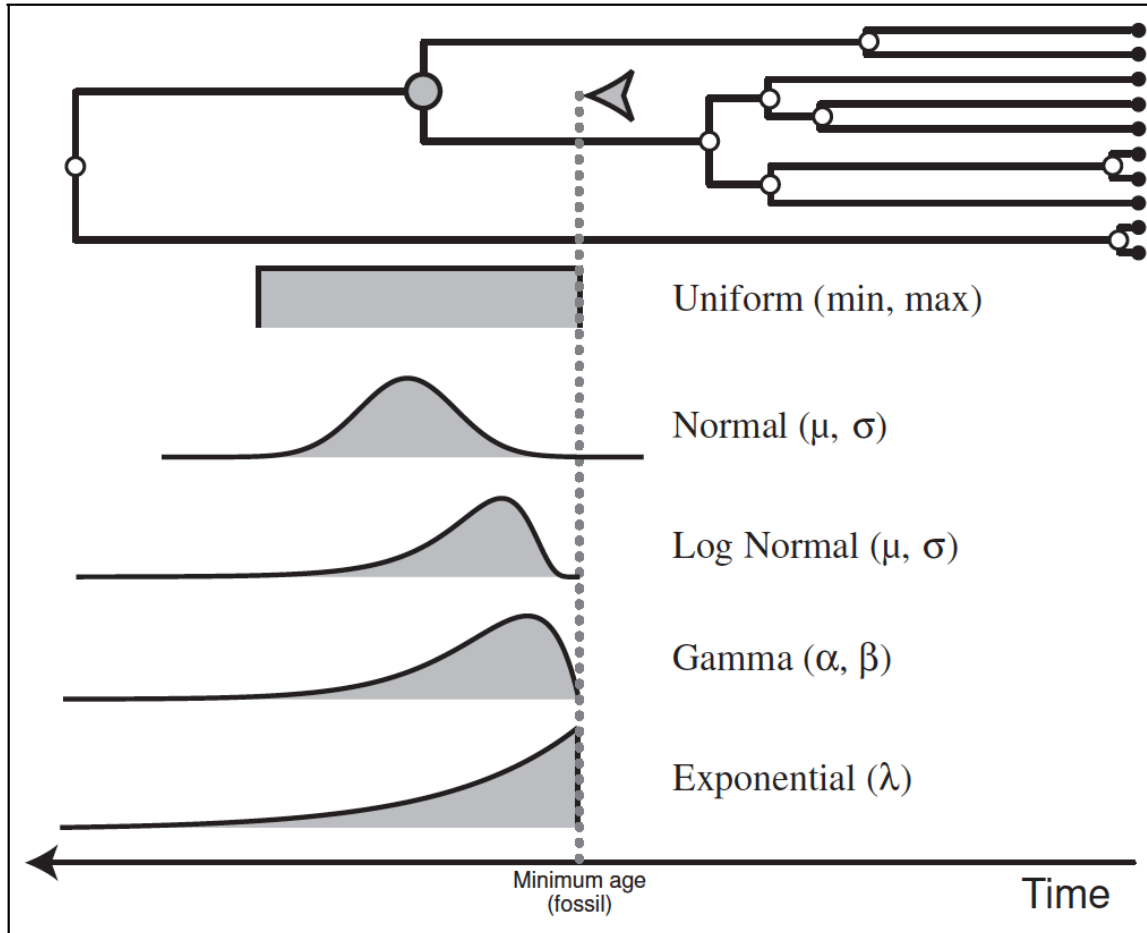
Models of lineage specific rate variation

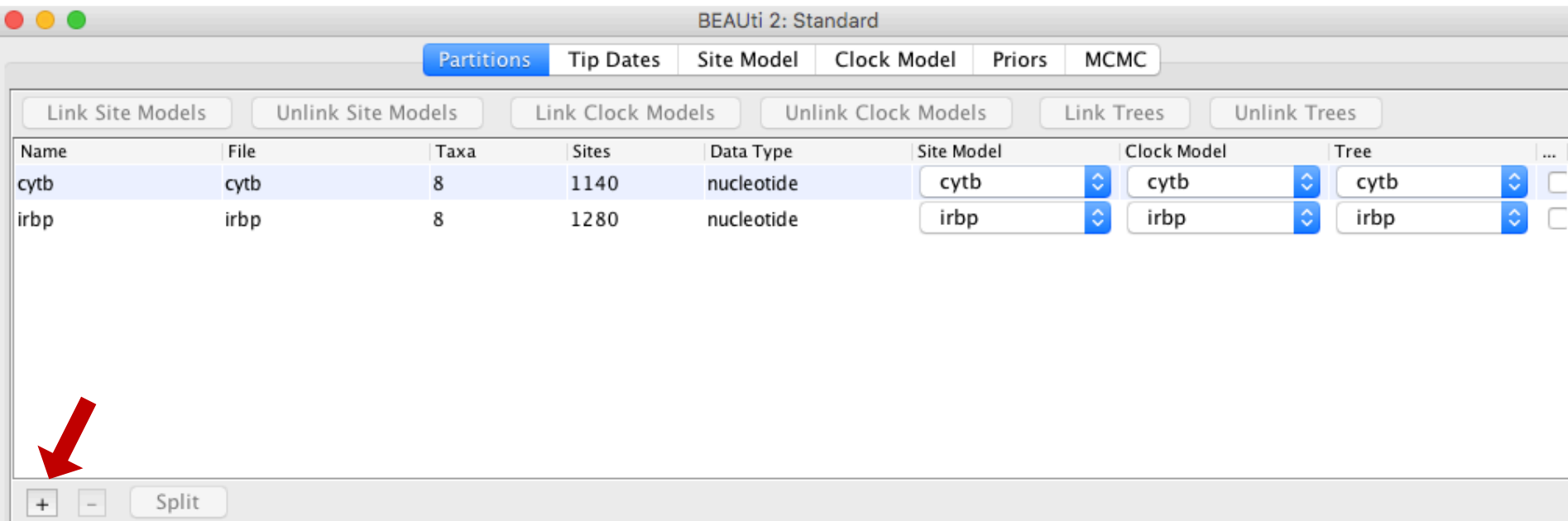
- Global molecular clock
- Local molecular clock
- Compound Poisson process
- Autocorrelated rates
- Uncorrelated rates
- Infinite mixture model on branch lengths

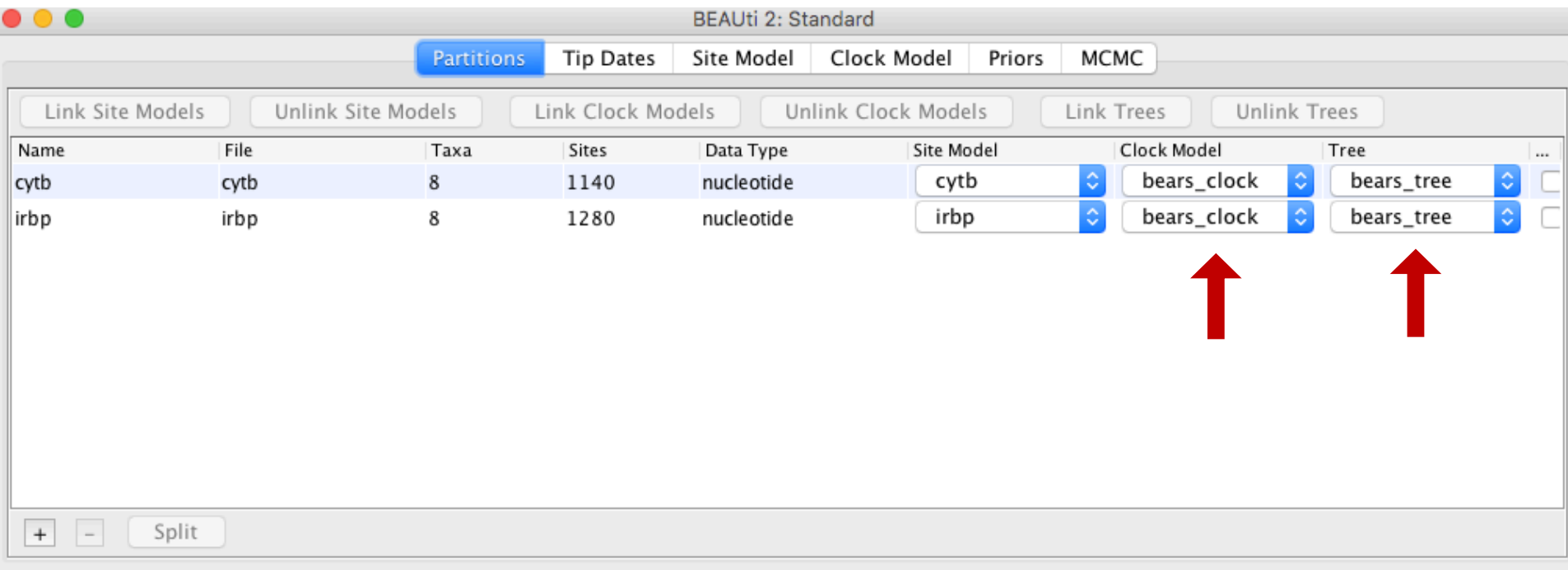
Priors on lineage diversification times (node ages)

1. Node distribution on a tree (e.g., uniform or Dirichlet).
Prior is conditional on the age of the root.
2. Population-level processes (e.g., coalescent prior)
3. Lineage diversification processes:
 - Yule process (speciation rate)
 - Birth-Death process (speciation and extinction rates)

Parametric distributions implemented in BEAST 2







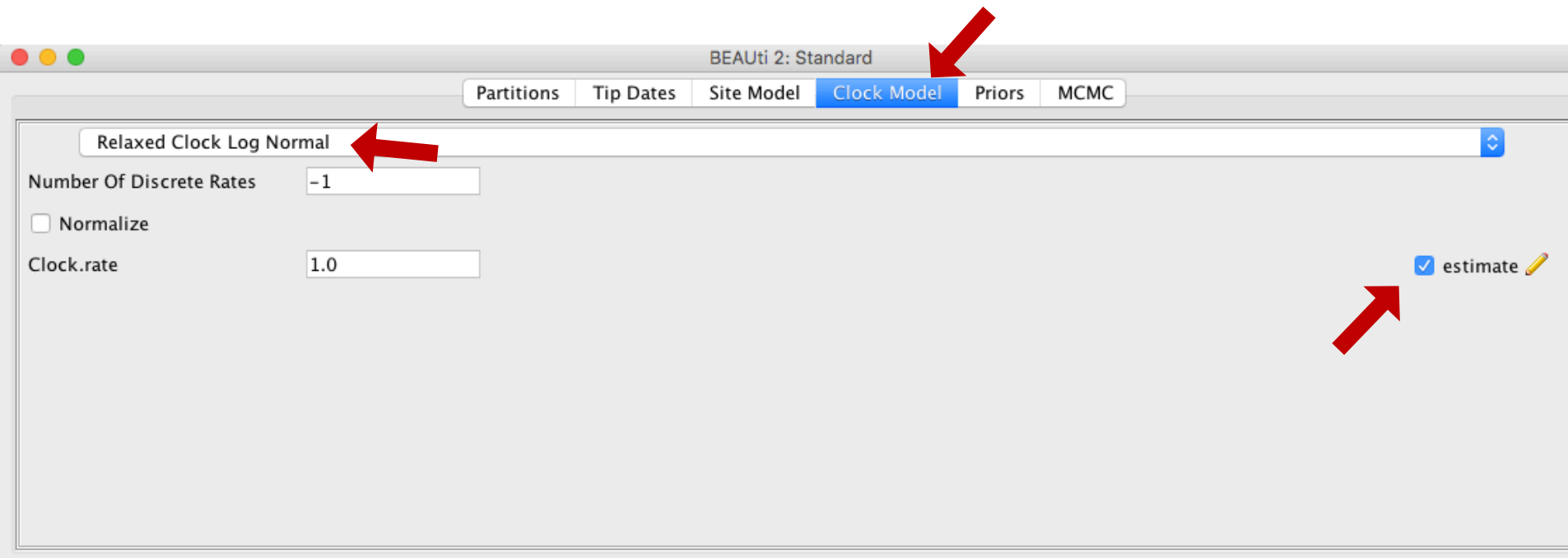
Input file preparation for analysis in BEAST 2

4. Set substitution models for the genes

The screenshot shows the BEAST 2 'Site Model' configuration window. The 'Partitions' tab is selected, and the 'irbp' partition is highlighted in the left sidebar. The 'Gamma Site Model' is selected for this partition. The following parameters are configured:

- Substitution Rate:** 1.0
- Gamma Category Count:** 4
- Shape:** 1.0
- Proportion Invariant:** 0.0
- Subst Model:** HKY
- Kappa:** 2.0
- Frequencies:** Estimated

On the right side, the 'estimate' checkbox is checked for the Substitution Rate, Shape, and Kappa parameters, while it is unchecked for the Proportion Invariant parameter. Red arrows point to the 'irbp' partition, the 'Gamma Site Model' dropdown, the 'Gamma Category Count' field, the 'Subst Model' dropdown, and the 'estimate' checkboxes for Substitution Rate, Shape, and Kappa.



Input file preparation for analysis in BEAST 2

6. Priors settings

The screenshot shows the 'Priors' tab in the BEAST 2 Standard interface. The 'Birth Death Model' is selected for the 'Tree.t:bear_tree' partition. The 'Type' is set to 'unscaled'. The 'Relative Death Rate' is 0.5, and the 'Birth Diff Rate' is 1.0. The 'Origin Height' is set to '[none]'. The 'Conditional On Root' checkbox is checked. The 'birthRate2.t:bear_tree' prior is set to 'Uniform' with an initial value of [1.0] [0.0,10000.0]. The 'gammaShape.s:cytb' prior is set to 'Exponential' with an initial value of [10.0]. The 'gammaShape.s:irbp' prior is set to 'Exponential' with an initial value of [1.0] [-∞,∞]. The 'kappa.s:cytb' prior is set to 'Log Normal' with an initial value of [2.0] [0.0,∞]. The 'kappa.s:irbp' prior is set to 'Log Normal' with an initial value of [2.0] [0.0,∞]. The 'relativeDeathRate2.t:bear_tree' prior is set to 'Uniform' with an initial value of [0.5] [0.0,1.0]. The 'ucldMean.c:bears_clock' prior is set to 'Exponential' with an initial value of [10.0]. The 'ucldStdev.c:bear_clock' prior is set to 'Gamma' with an initial value of [0.1] [0.0,∞].

Red arrows highlight the following settings:

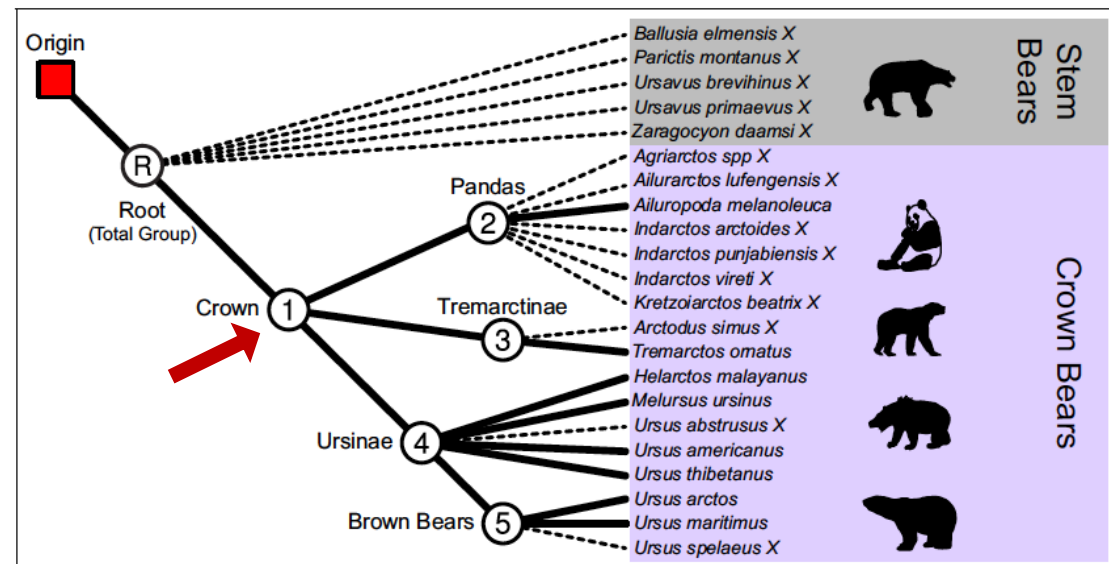
- The 'Priors' tab in the top navigation bar.
- The 'Birth Death Model' dropdown menu.
- The 'unscaled' dropdown menu for the 'Type'.
- The 'Conditional On Root' checkbox.
- The 'initial = [10.0]' value for the 'gammaShape.s:cytb' prior.
- The 'initial = [10.0]' value for the 'ucldMean.c:bears_clock' prior.
- The '+' button at the bottom to add more priors.



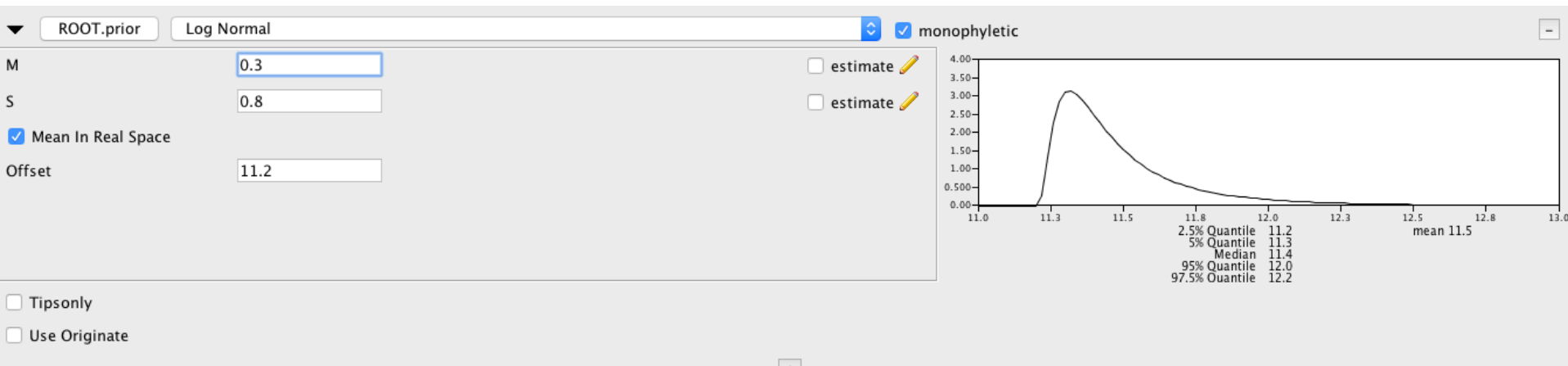
7. Choose the oldest fossil for the node(s)

Fossil species	Age range (My)	Citation
<i>Parictis montanus</i>	33.9–37.2	Clark and Guensburg 1972; Krause et al. 2008
<i>Zaragocyon daamsi</i>	20–22.8	Ginsburg and Morales 1995; Abella et al. 2012
<i>Ballusia elmensis</i>	13.7–16	Ginsburg and Morales 1998; Abella et al. 2012
<i>Ursavus primaevus</i>	13.65–15.97	Andrews and Tobien 1977; Abella et al. 2012
<i>Ursavus brevihinus</i>	15.97–16.9	Heizmann et al. 1980; Abella et al. 2012
<i>Indarctos vireti</i>	7.75–8.7	Montoya et al. 2001; Abella et al. 2012
<i>Indarctos arctoides</i>	8.7–9.7	Geraads et al. 2005; Abella et al. 2012
<i>Indarctos punjabiensis</i>	4.9–9.7	Baryshnikov 2002; Abella et al. 2012
<i>Ailurarctos lufengensis</i>	5.8–8.2	Jin et al. 2007; Abella et al. 2012
<i>Agriarctos spp.</i>	4.9–7.75	Abella et al. 2011; Abella et al. 2012
<i>Kretzoiarctos beatrix</i>	11.2–11.8	Abella et al. 2011; Abella et al. 2012
<i>Arctodus simus</i>	0.012–2.588	Churcher et al. 1993; Krause et al. 2008
<i>Ursus abstrusus</i>	1.8–5.3	Bjork 1970; Krause et al. 2008
<i>Ursus spelaeus</i>	0.027–0.25	Loreille et al. 2001; Krause et al. 2008

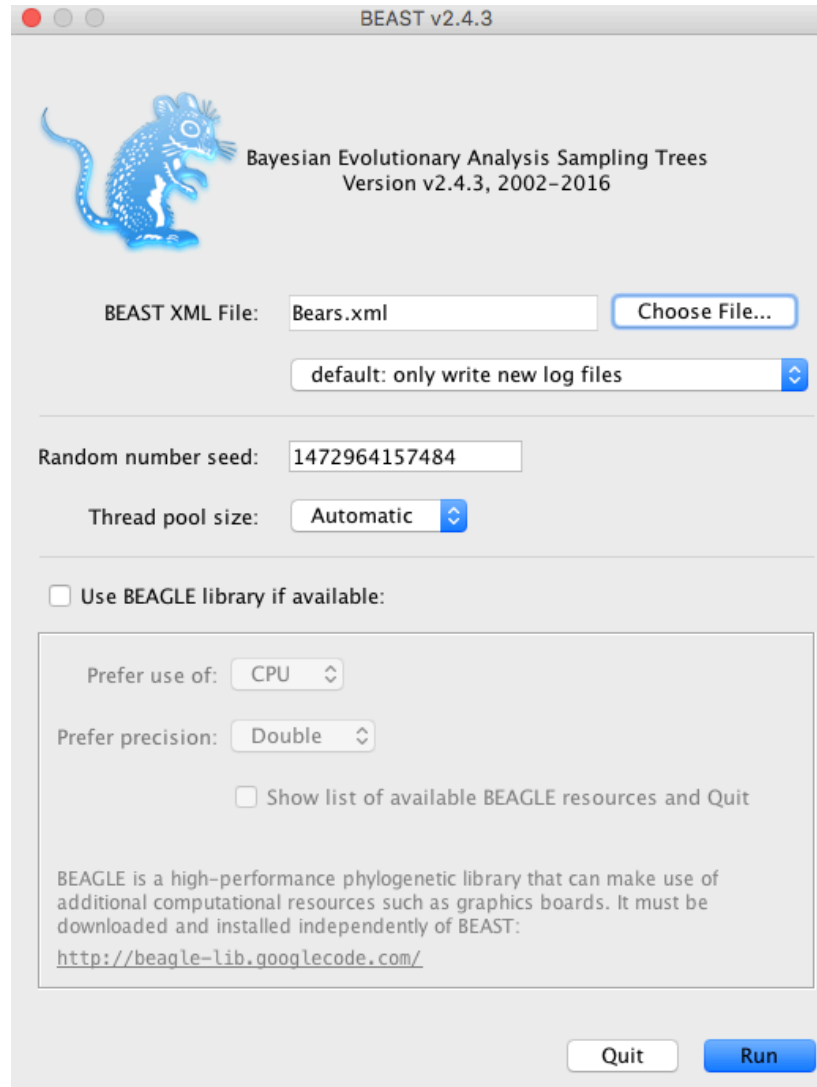
Crown →



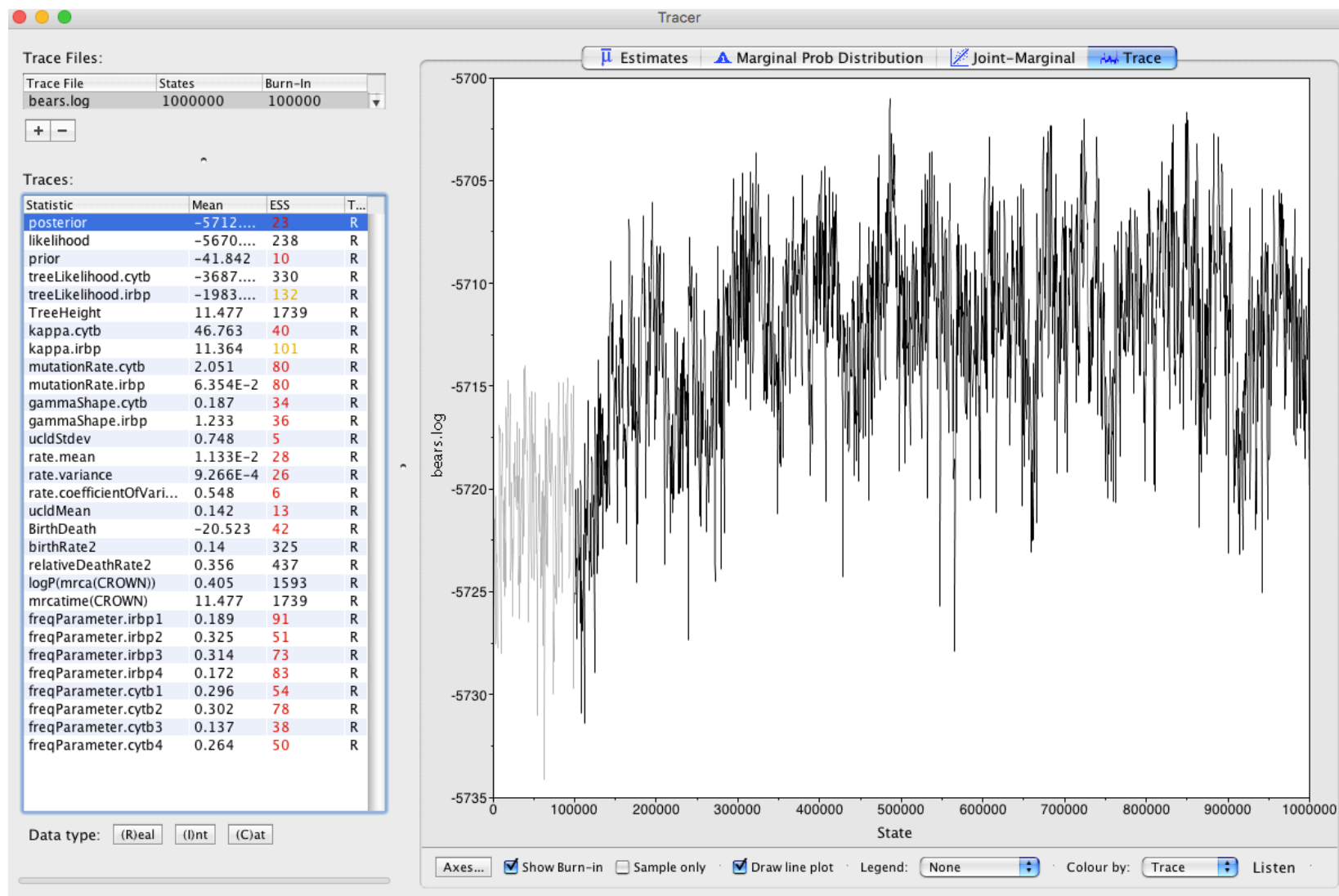
8. Set prior distributions for node time calibrations

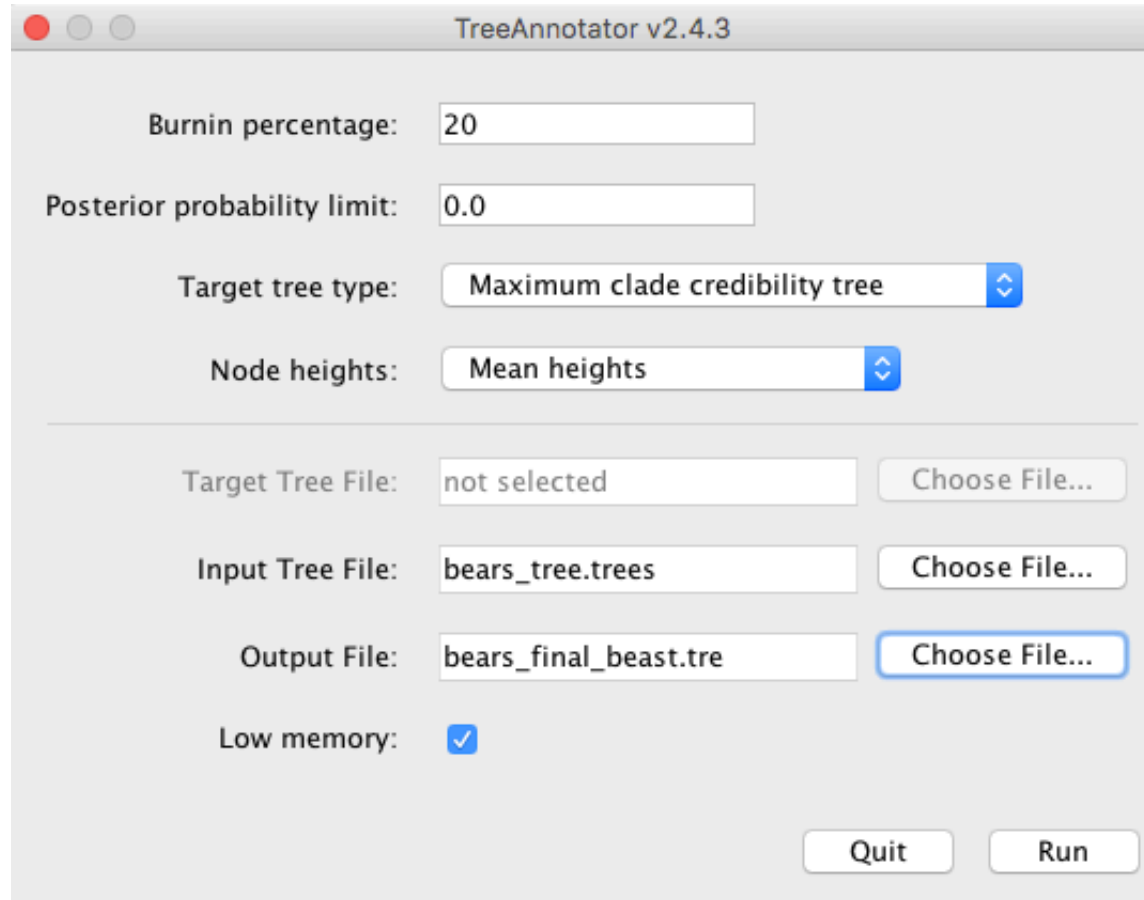


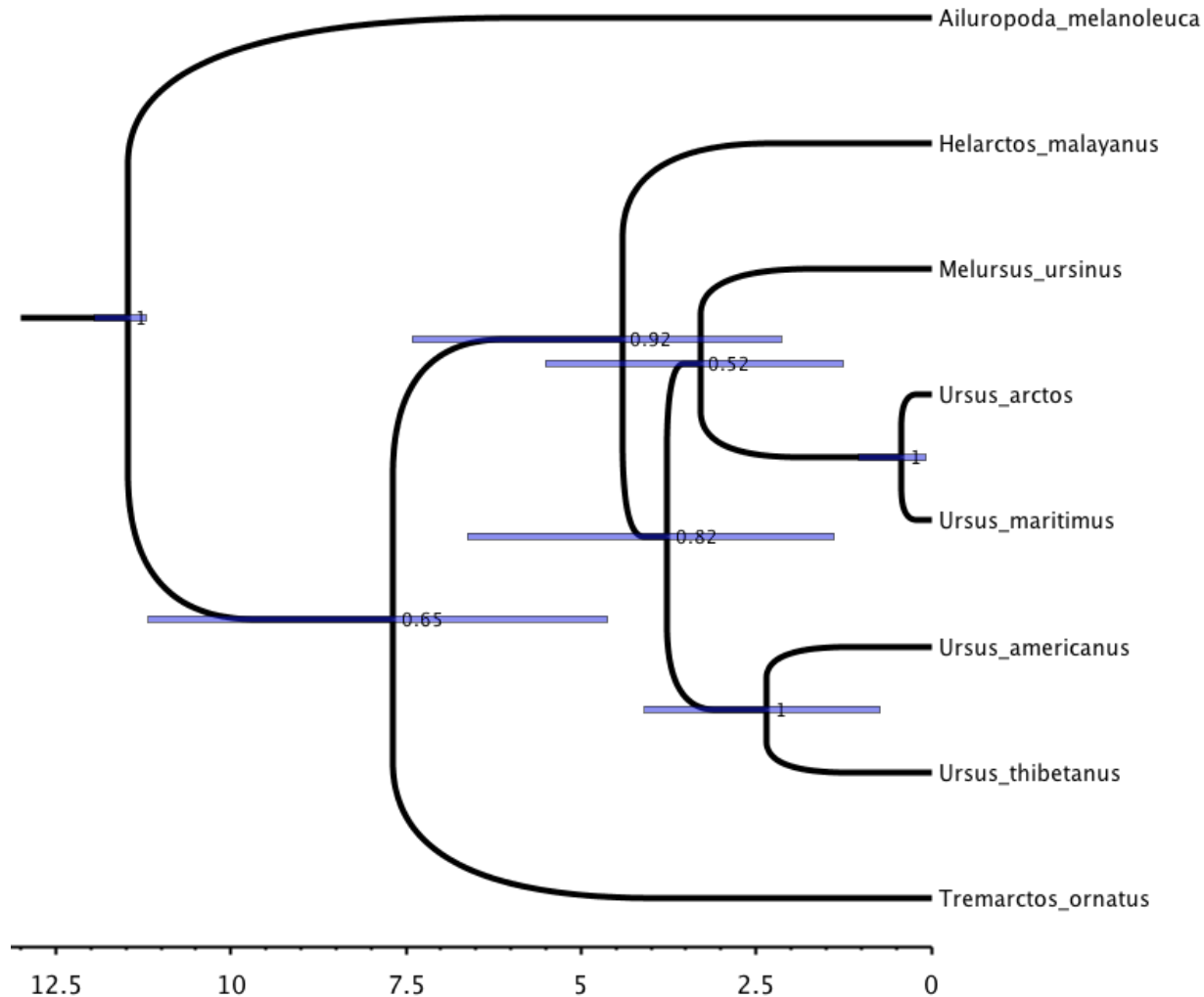
10. Save file `bears.xml`



12. Check the posterior probability distributions in Tracer





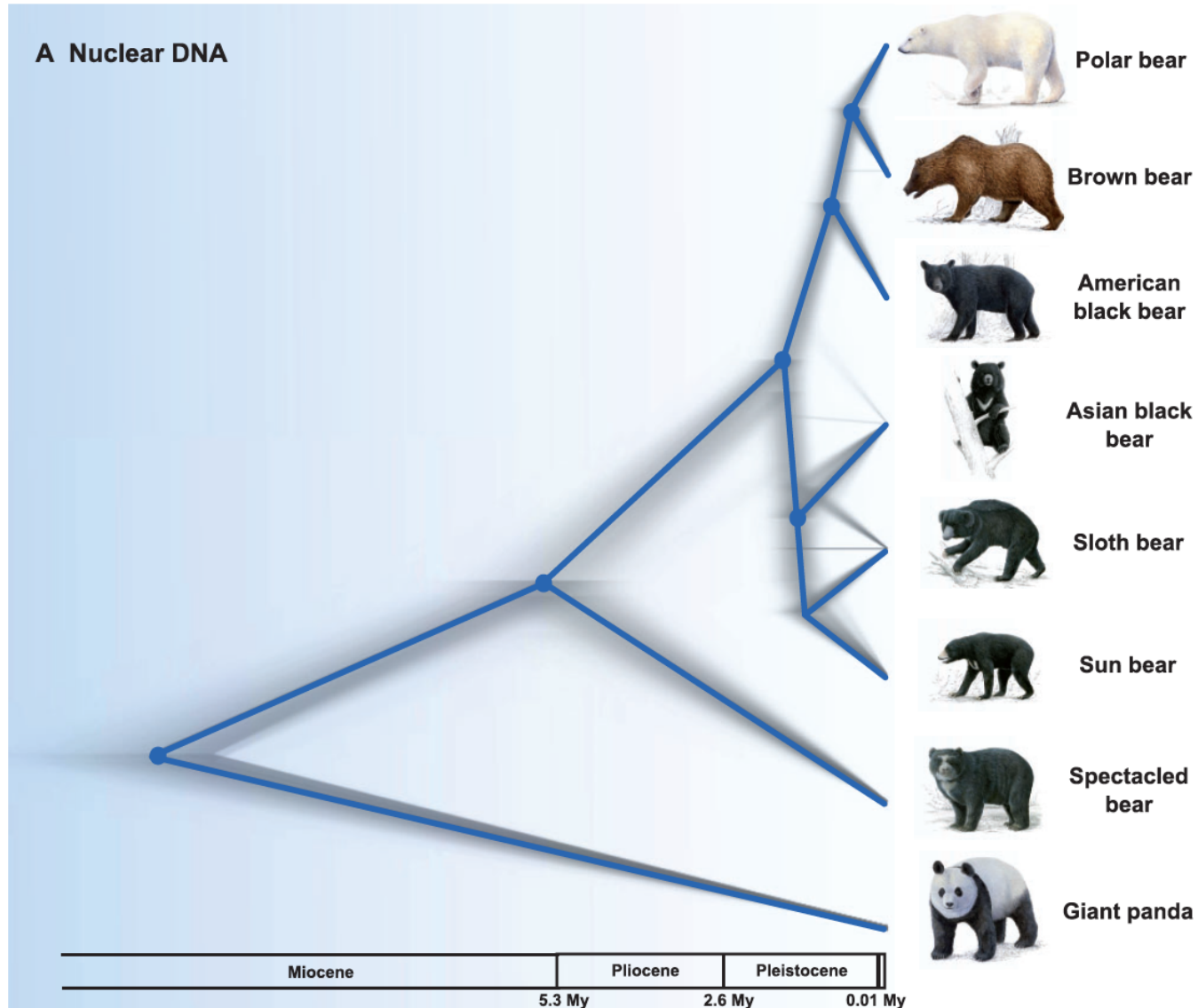




HOMWORK QUESTIONS

1. Compare the topology, branch lengths and node support values between RAxML, MrBayes and BEAST trees. How do they differ?
2. For BEAST analysis, increase the number of mcmc steps in a chain (e.g., #10,000,000). Does it change the high posterior densities (HPDs) for nodes? Are the effective sample sizes (ESS) for parameter estimates above 200 (check new log file in Tracer)?
3. Repeat the previous run one more time and compare both log files in the Tracer. Did their distributions converge?
4. Add more fossil calibrations in the BEAST analysis. Did the high posterior density intervals (HPDs) for node calibrations become narrower or wider?

Phylogenomic Cloudogram



Kutschera et al., 2014