Estimating divergence times using the fossilised birth-death process in BEAST2

Analytical Paleobiology workshop

Rachel Warnock

27.08.2025

Objectives

Lecture

- Brief introduction to the BEAST2 pipeline

Tutorial

- Divergence dating under the fossilised birth-death process
- Choose your own adventure

Intro to BEAST2

BEAST2

Beast2

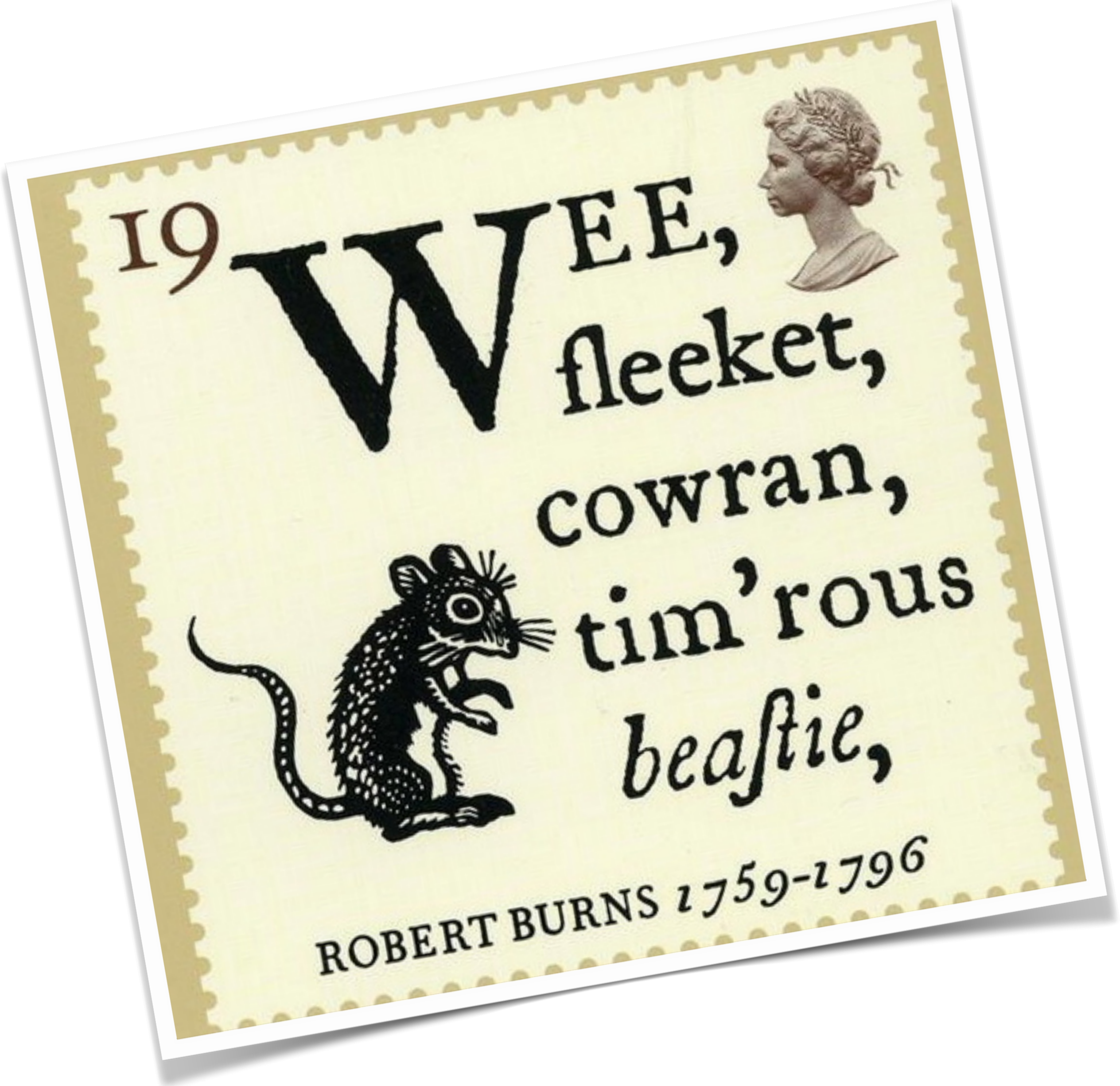
Bayesian evolutionary analysis by sampling trees

Also designed with extendability and flexibility in mind

Also developed and supported by a large international team of developers

Has a suite of apps that can used to generate input files and analysis the output

www.beast2.org



Scots poem - also the BEAST2 logo!

BEAST2 toolkit and work flow

Step 1. generate the xml file in **BEAUti**

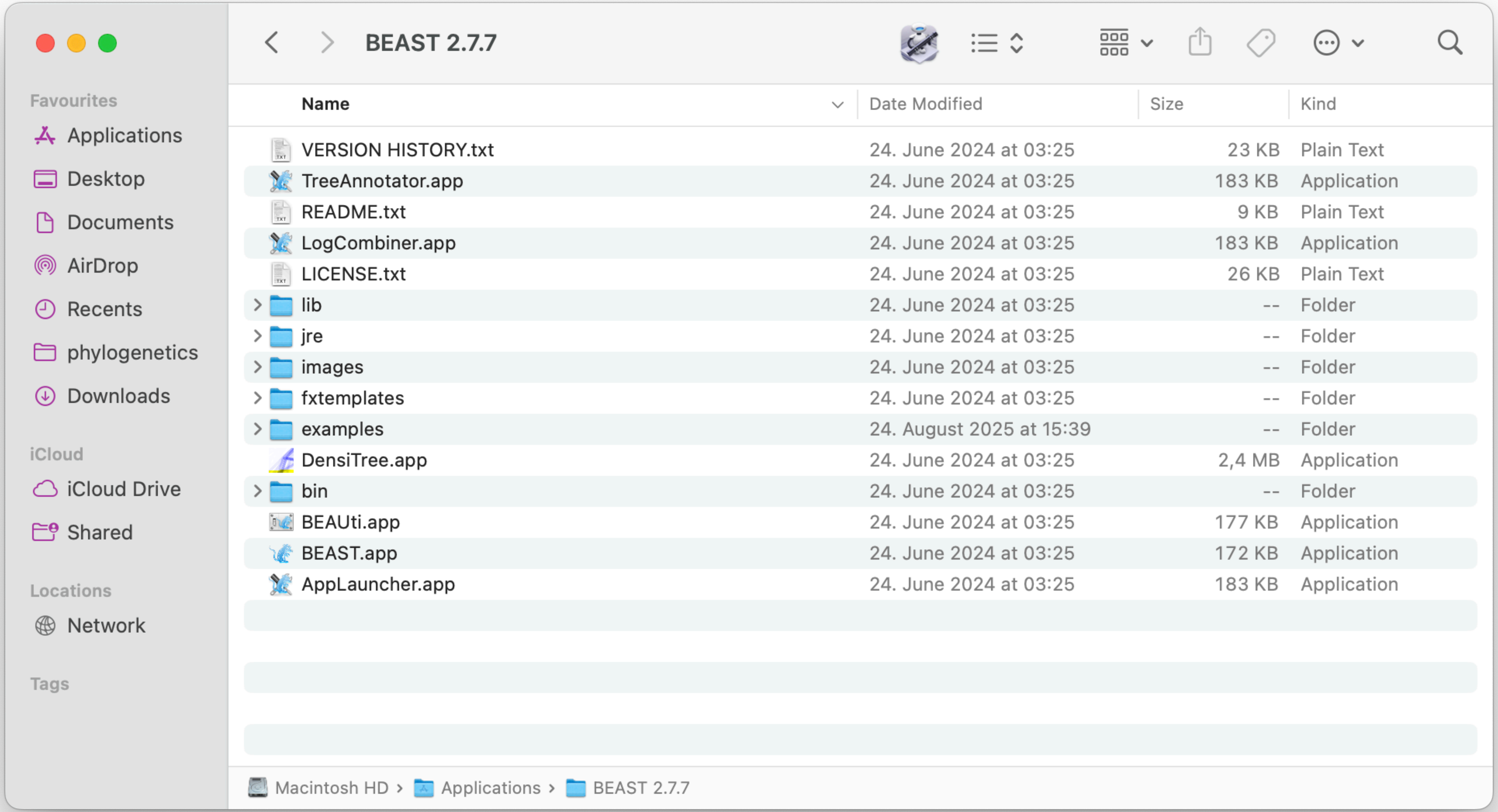
Step 2. run your analysis in **BEAST**

Step 3a. Examine you log files using **Tracer**

Step 3b. Generate a summary tree using **TreeAnnotator**

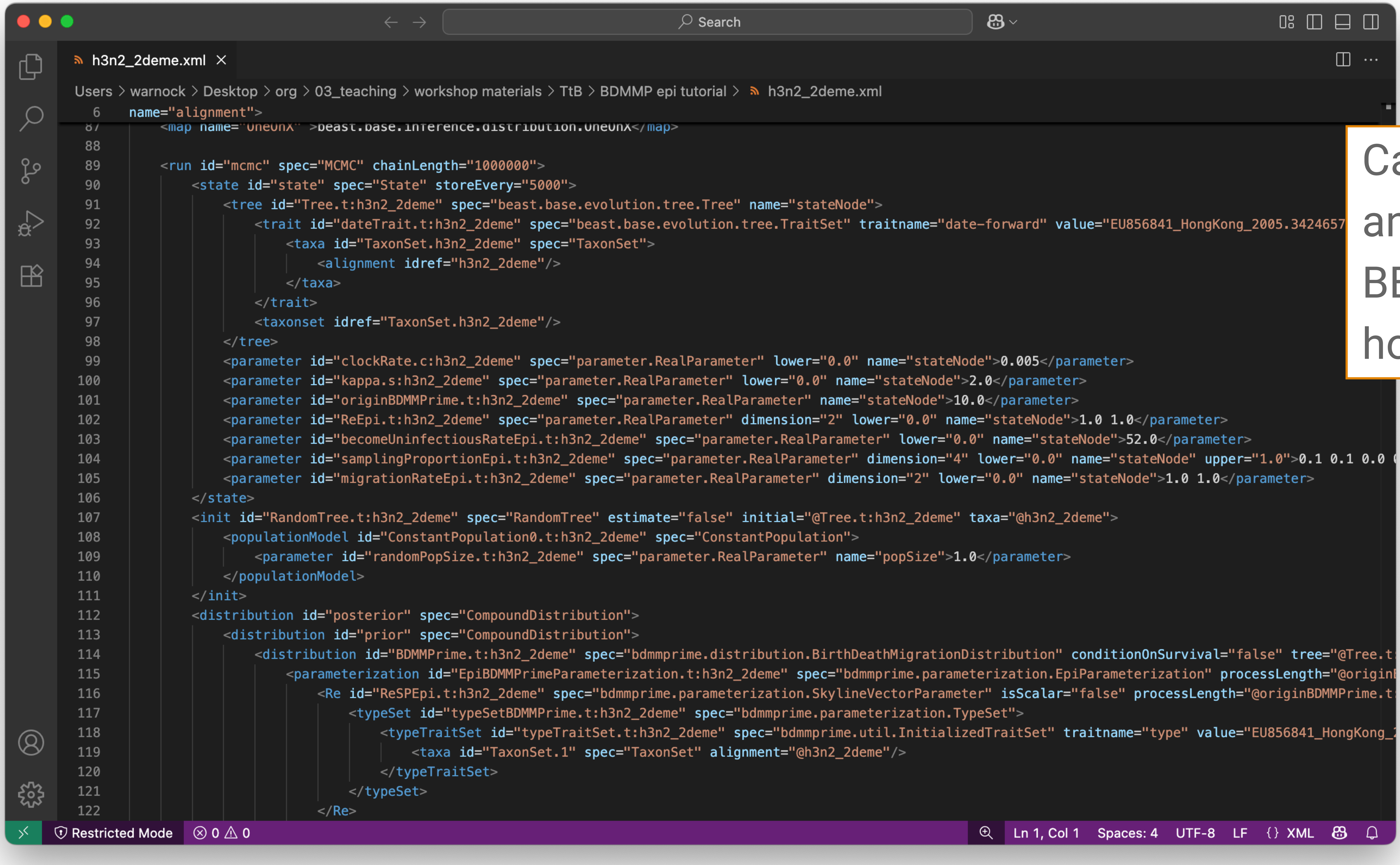
Step 4. Examine your summary tree in **FigTree**

Step… any other downstream analysis



Name	Date Modified	Size	Kind
VERSION HISTORY.txt	24. June 2024 at 03:25	23 KB	Plain Text
TreeAnnotator.app	24. June 2024 at 03:25	183 KB	Application
README.txt	24. June 2024 at 03:25	9 KB	Plain Text
LogCombiner.app	24. June 2024 at 03:25	183 KB	Application
LICENSE.txt	24. June 2024 at 03:25	26 KB	Plain Text
lib	24. June 2024 at 03:25	--	Folder
jre	24. June 2024 at 03:25	--	Folder
images	24. June 2024 at 03:25	--	Folder
fxtemplates	24. June 2024 at 03:25	--	Folder
examples	24. August 2025 at 15:39	--	Folder
DensiTree.app	24. June 2024 at 03:25	2,4 MB	Application
bin	24. June 2024 at 03:25	--	Folder
BEAUti.app	24. June 2024 at 03:25	177 KB	Application
BEAST.app	24. June 2024 at 03:25	172 KB	Application
AppLauncher.app	24. June 2024 at 03:25	183 KB	Application

BEAST2 input: the XML file



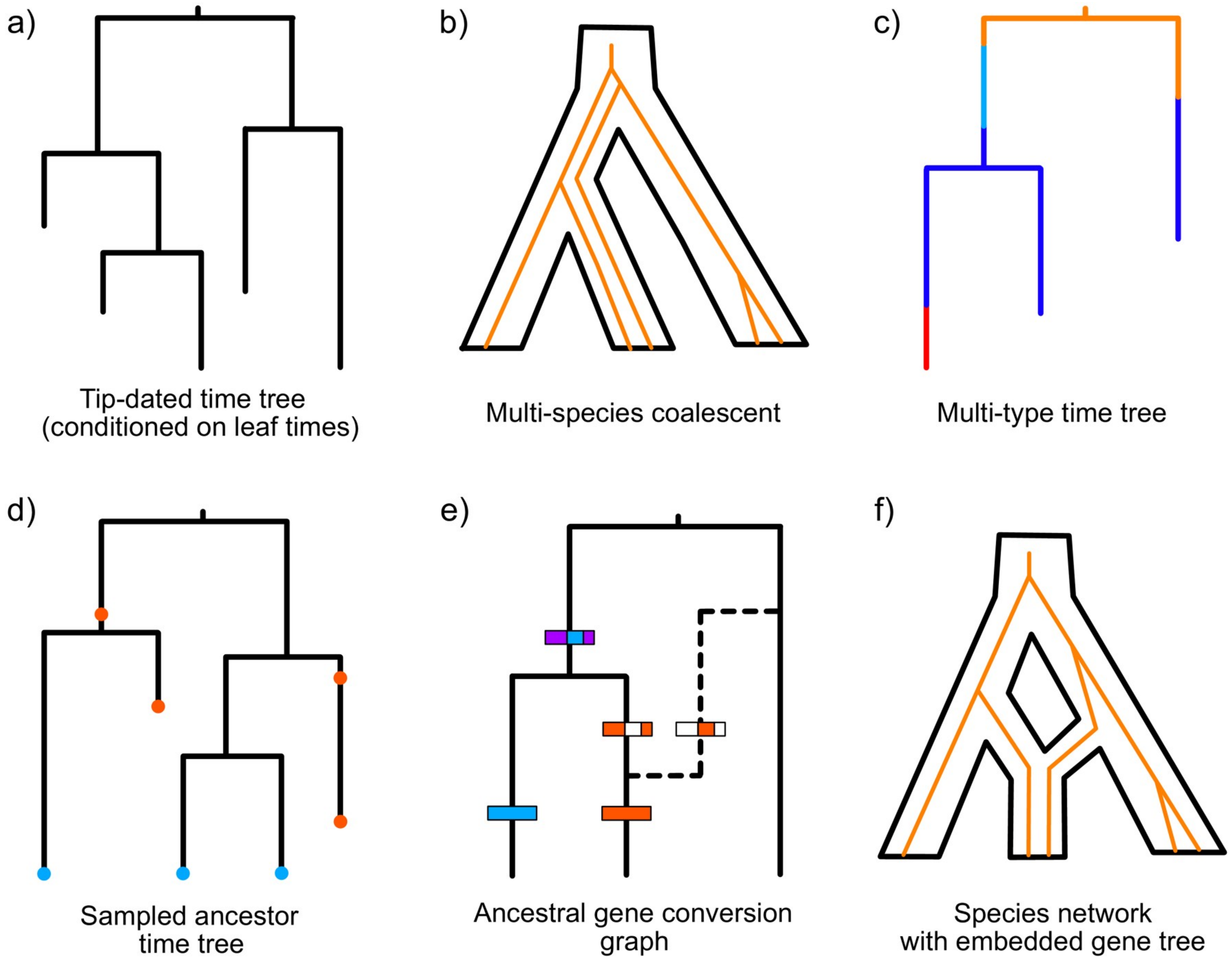
```xml
    name="alignment">
        <map name="OneOnX" >beast.base.inference.distribution.OneOnX</map>

        <run id="mcmc" spec="MCMC" chainLength="1000000">
            <state id="state" spec="State" storeEvery="5000">
                <tree id="Tree.t:h3n2_2deme" spec="beast.base.evolution.tree.Tree" name="stateNode">
                    <trait id="dateTrait.t:h3n2_2deme" spec="beast.base.evolution.tree.TraitSet" traitname="date-forward" value="EU856841_HongKong_2005.3424657
                        <taxa id="TaxonSet.h3n2_2deme" spec="TaxonSet">
                            <alignment idref="h3n2_2deme"/>
                        </taxa>
                    </trait>
                    <taxonset idref="TaxonSet.h3n2_2deme"/>
                </tree>
                <parameter id="clockRate.c:h3n2_2deme" spec="parameter.RealParameter" lower="0.0" name="stateNode">0.005</parameter>
                <parameter id="kappa.s:h3n2_2deme" spec="parameter.RealParameter" lower="0.0" name="stateNode">2.0</parameter>
                <parameter id="originBDMMPrime.t:h3n2_2deme" spec="parameter.RealParameter" name="stateNode">10.0</parameter>
                <parameter id="ReEpi.t:h3n2_2deme" spec="parameter.RealParameter" dimension="2" lower="0.0" name="stateNode">1.0 1.0</parameter>
                <parameter id="becomeUninfectiousRateEpi.t:h3n2_2deme" spec="parameter.RealParameter" lower="0.0" name="stateNode">52.0</parameter>
                <parameter id="samplingProportionEpi.t:h3n2_2deme" spec="parameter.RealParameter" dimension="4" lower="0.0" name="stateNode" upper="1.0">0.1 0.1 0.0
                <parameter id="migrationRateEpi.t:h3n2_2deme" spec="parameter.RealParameter" dimension="2" lower="0.0" name="stateNode">1.0 1.0</parameter>
            </state>
            <init id="RandomTree.t:h3n2_2deme" spec="RandomTree" estimate="false" initial="@Tree.t:h3n2_2deme" taxa="@h3n2_2deme">
                <populationModel id="ConstantPopulation0.t:h3n2_2deme" spec="ConstantPopulation">
                    <parameter id="randomPopSize.t:h3n2_2deme" spec="parameter.RealParameter" name="popSize">1.0</parameter>
                </populationModel>
            </init>
            <distribution id="posterior" spec="CompoundDistribution">
                <distribution id="prior" spec="CompoundDistribution">
                    <distribution id="BDMMPrime.t:h3n2_2deme" spec="bdmmprime.distribution.BirthDeathMigrationDistribution" conditionOnSurvival="false" tree="@Tree.t
                        <parameterization id="EpiBDMMPrimeParameterization.t:h3n2_2deme" spec="bdmmprime.parameterization.EpiParameterization" processLength="@origin
                            <Re id="ReSPEpi.t:h3n2_2deme" spec="bdmmprime.parameterization.SkylineVectorParameter" isScalar="false" processLength="@originBDMMPrime.t
                                <typeSet id="typeSetBDMMPrime.t:h3n2_2deme" spec="bdmmprime.parameterization.TypeSet">
                                    <typeTraitSet id="typeTraitSet.t:h3n2_2deme" spec="bdmmprime.util.InitializedTraitSet" traitname="type" value="EU856841_HongKong_
                                        <taxa id="TaxonSet.1" spec="TaxonSet" alignment="@h3n2_2deme"/>
                                    </typeTraitSet>
                                </typeSet>
                            </Re>
```

Caveat: if you want to use an option not available in BEAUti you have to learn how to edit the XML

A wide range of models and tree structures

Note all tree models in BEAST2 incorporate a temporal component



Tip-dated time tree (conditioned on leaf times)

Multi-species coalescent

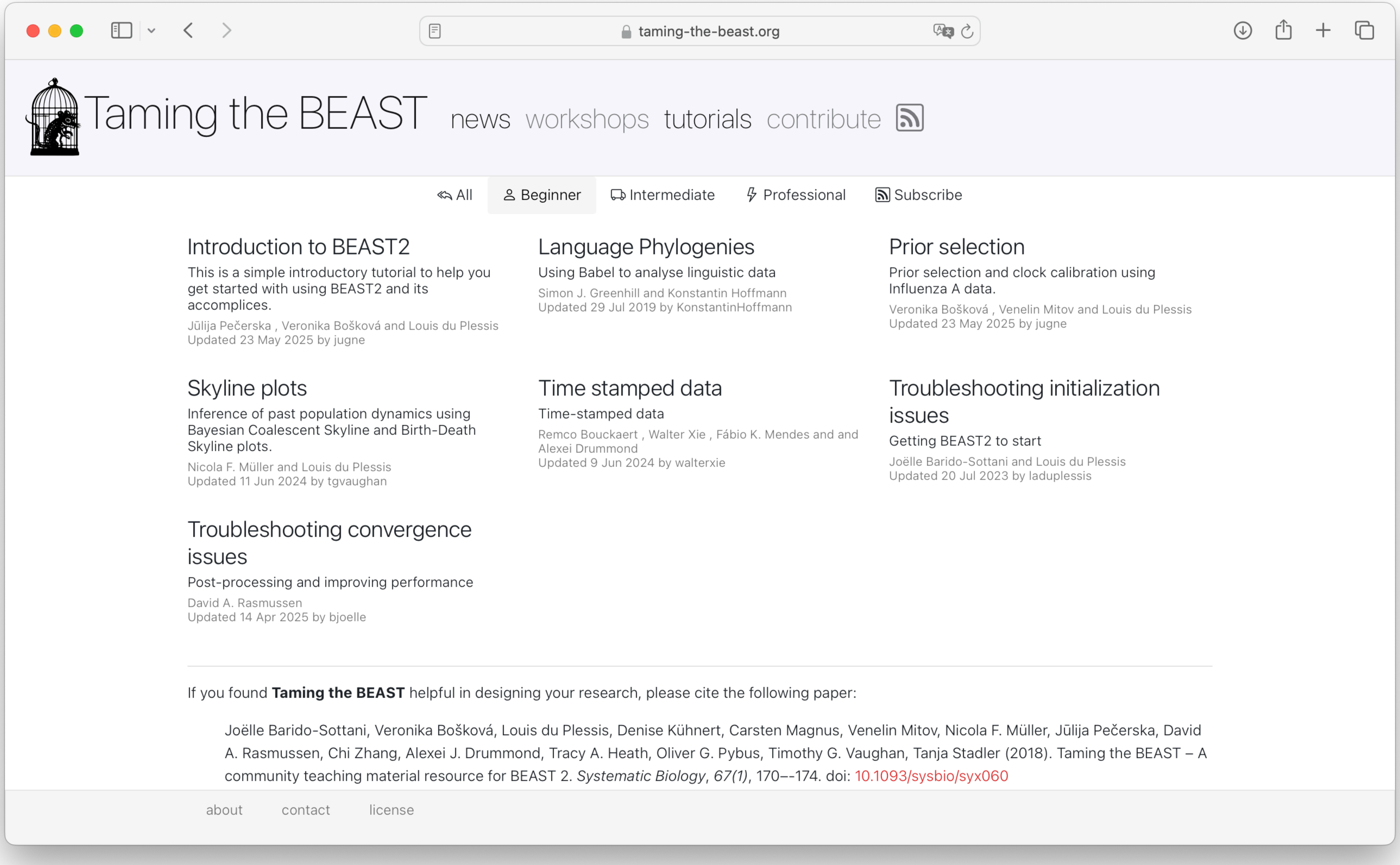
Multi-type time tree

Sampled ancestor time tree

Ancestral gene conversion graph

Species network with embedded gene tree

Bouckaert et al. (2019) PLOS Comp Bio



Introduction to BEAST2

This is a simple introductory tutorial to help you get started with using BEAST2 and its accomplices.

Jūlija Pečerska , Veronika Bošková and Louis du Plessis
Updated 23 May 2025 by jugne

Language Phylogenies

Using Babel to analyse linguistic data

Simon J. Greenhill and Konstantin Hoffmann
Updated 29 Jul 2019 by KonstantinHoffmann

Prior selection

Prior selection and clock calibration using Influenza A data.

Veronika Bošková , Venelin Mitov and Louis du Plessis
Updated 23 May 2025 by jugne

Skyline plots

Inference of past population dynamics using Bayesian Coalescent Skyline and Birth-Death Skyline plots.

Nicola F. Müller and Louis du Plessis
Updated 11 Jun 2024 by tgvaughan

Time stamped data

Time-stamped data

Remco Bouckaert , Walter Xie , Fábio K. Mendes and and Alexei Drummond
Updated 9 Jun 2024 by walterxie

Troubleshooting initialization issues

Getting BEAST2 to start

Joëlle Barido-Sottani and Louis du Plessis
Updated 20 Jul 2023 by laduplessis

Troubleshooting convergence issues

Post-processing and improving performance

David A. Rasmussen
Updated 14 Apr 2025 by bjoelle

If you found **Taming the BEAST** helpful in designing your research, please cite the following paper:

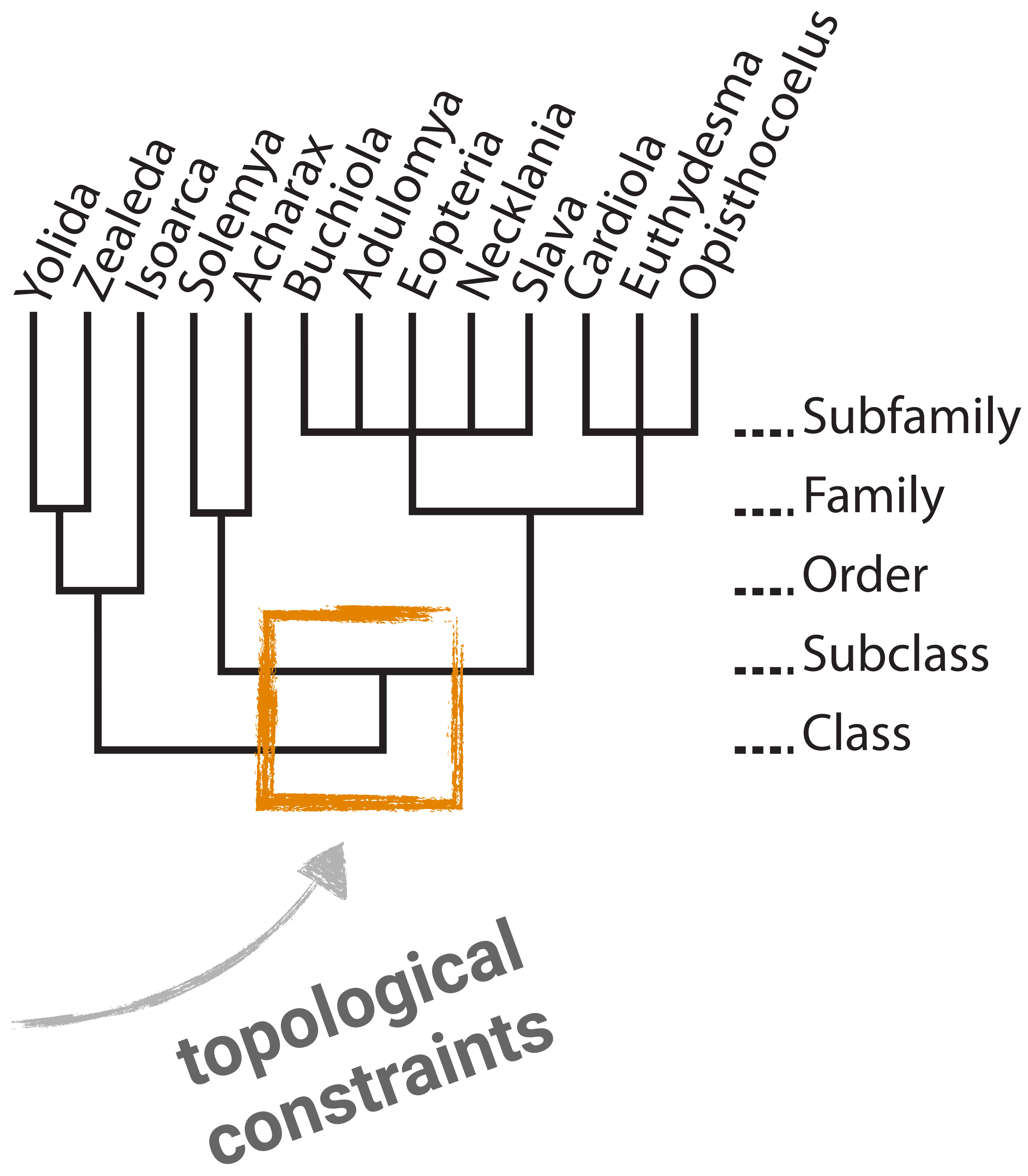
> Joëlle Barido-Sottani, Veronika Bošková, Louis du Plessis, Denise Kühnert, Carsten Magnus, Venelin Mitov, Nicola F. Müller, Jūlija Pečerska, David A. Rasmussen, Chi Zhang, Alexei J. Drummond, Tracy A. Heath, Oliver G. Pybus, Timothy G. Vaughan, Tanja Stadler (2018). Taming the BEAST – A community teaching material resource for BEAST 2. *Systematic Biology*, *67(1)*, 170–-174. doi: 10.1093/sysbio/syx060

https://taming-the-beast.org/

Exercise

Fossils can be incorporated via **taxonomy** or **character data** (total-evidence)

- Bivalvia
  - Palaeotaxodonta
    - Nuculoida
      - Nuculanacea
        - Nuculanidae
          - Yolida
          - Zealeda
        - Isoarcidae
          - Isoarca
  - Cryptodonta
    - Solemyoida
      - Solemyacea
        - Solemyidae
          - Solemya
          - Acharax
    - Praecardioida
      - Praecardiacea
        - Praecardiidae
          - Praecardiinae
            - Buchiola
            - Adulomya
            - Eopteria
            - Necklania
            - Slava
          - Cardiolinae
            - Cardiola
            - Euthydesma
            - Opisthocoelus

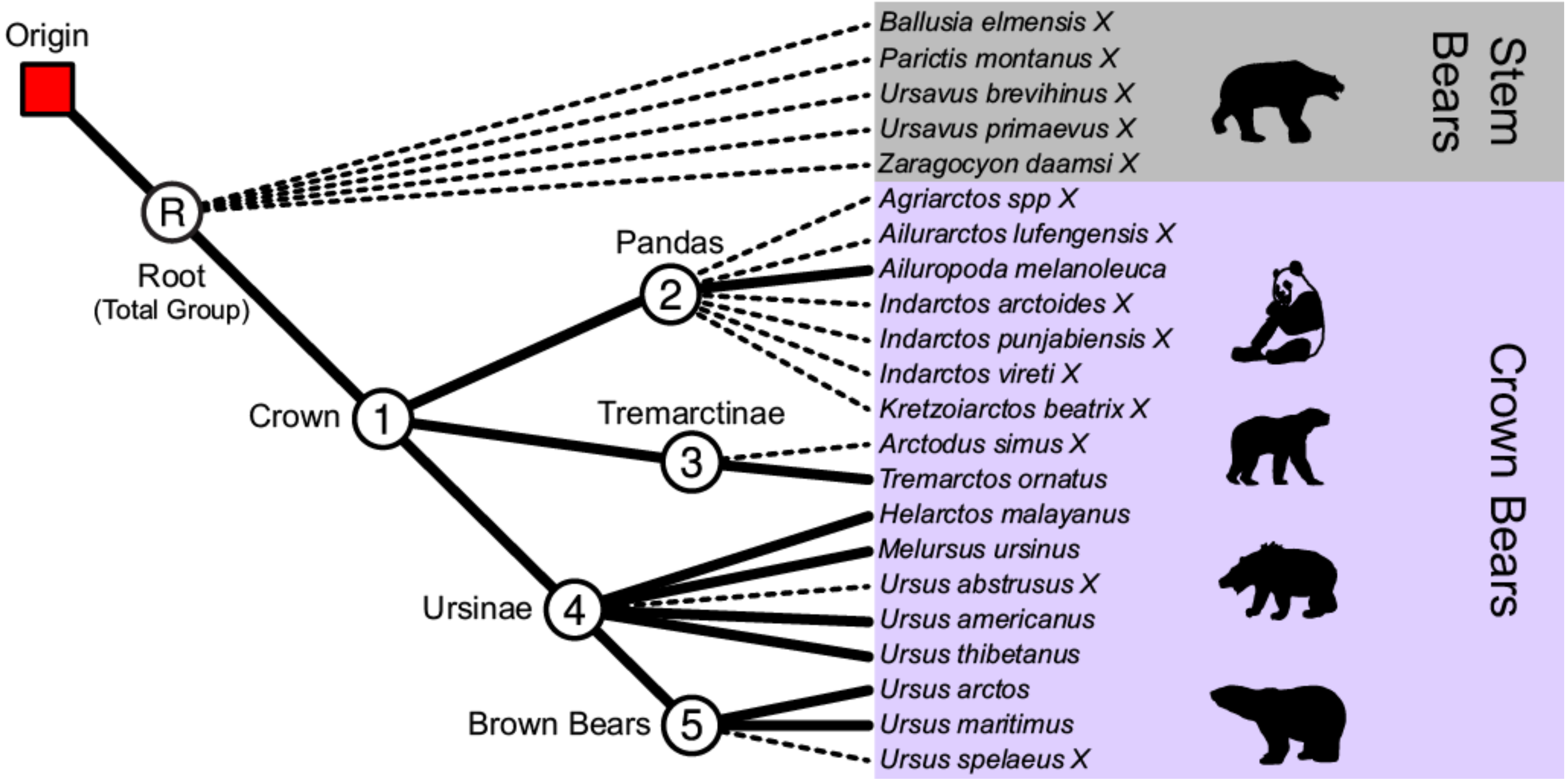


OR

ATAT...
TCAC...
????...

1001...
1101...
0100...

Image source Soul & Friedman ([2015](2015))



Origin

R

Root

(Total Group)

Crown

1

Pandas

2

Tremarctinae

3

Ursinae

4

Brown Bears

5

Ballusia elmensis X

Parictis montanus X

Ursavus brevihinus X

Ursavus primaevus X

Zaragocyon daamsi X

Stem Bears

Agriarctos spp X

Ailurarctos lufengensis X

Ailuropoda melanoleuca

Indarctos arctoides X

Indarctos punjabiensis X

Indarctos vireti X

Kretzoiarctos beatrix X

Arctodus simus X

Tremarctos ornatus

Helarctos malayanus

Melursus ursinus

Ursus abstrusus X

Ursus americanus

Ursus thibetanus

Ursus arctos

Ursus maritimus

Ursus spelaeus X

Crown Bears

Data

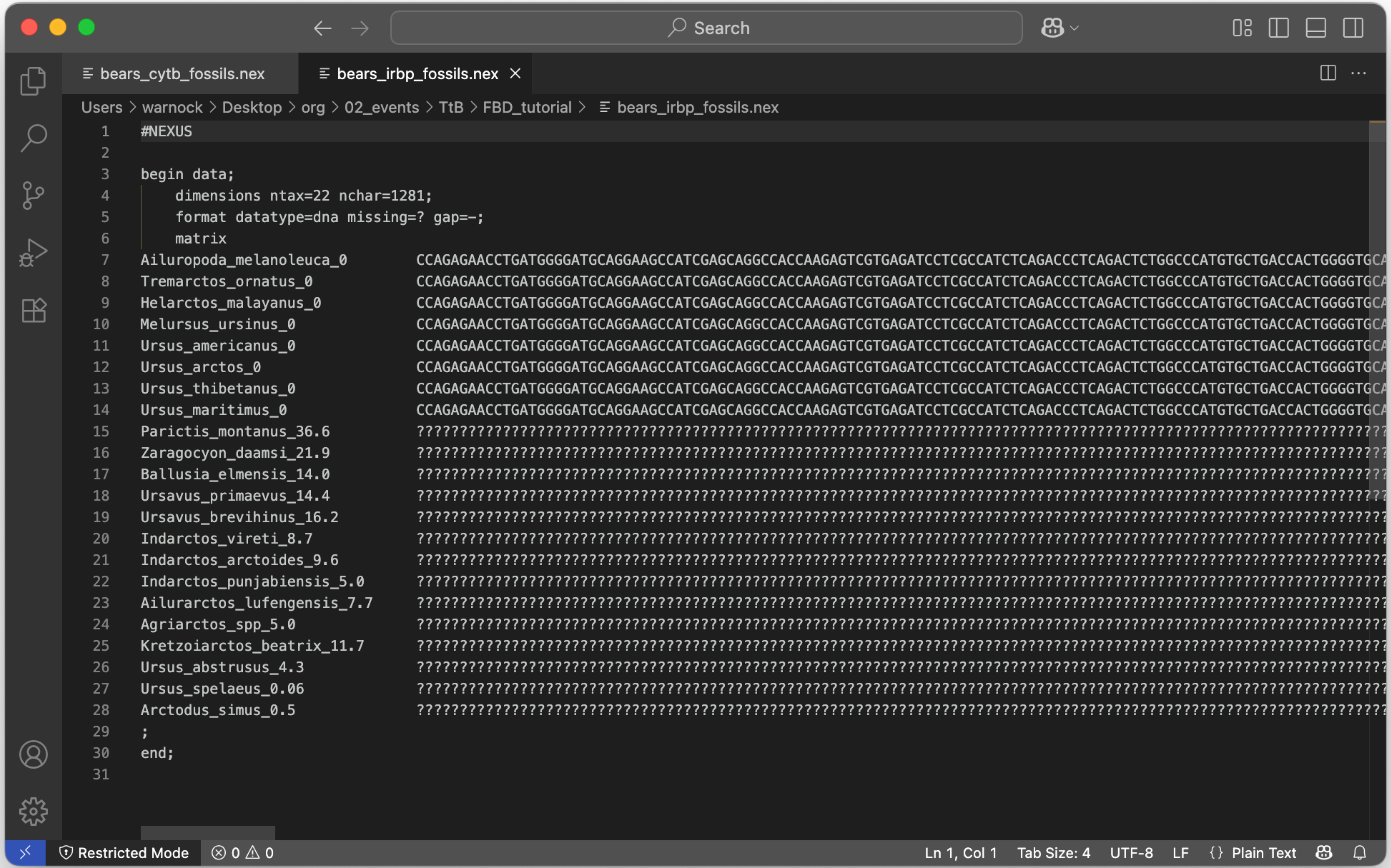
2 genes (irbp, cytb) for 8 extant species

Occurrence times of 14 fossil species

Taxonomic constraints

Note: this analysis includes includes **no character data** for our fossils

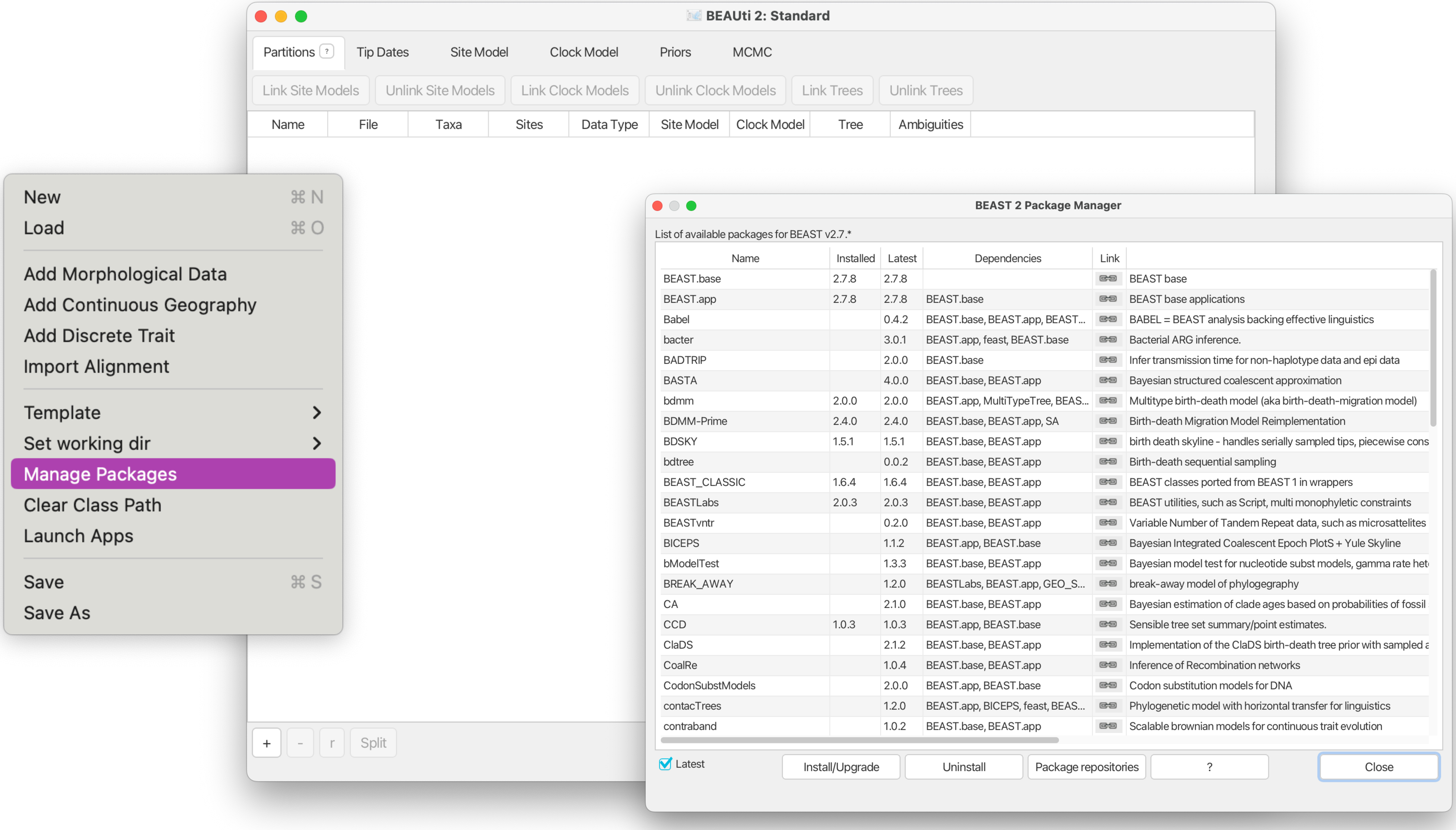
More about models of morphological evolution / TE dating on Thu / Fri



```
#NEXUS

begin data;
    dimensions ntax=22 nchar=1281;
    format datatype=dna missing=? gap=-;
    matrix
Ailuropoda_melanoleuca_0        CCAGAGAACCTGATGGGGATGCAGGAAGCCATCGAGCAGGCCACCAAGAGTCGTGAGATCCTCGCCATCTCAGACCCTCAGACTCTGGCCCATGTGCTGACCACTGGGGTGCA
Tremarctos_ornatus_0            CCAGAGAACCTGATGGGGATGCAGGAAGCCATCGAGCAGGCCACCAAGAGTCGTGAGATCCTCGCCATCTCAGACCCTCAGACTCTGGCCCATGTGCTGACCACTGGGGTGCA
Helarctos_malayanus_0           CCAGAGAACCTGATGGGGATGCAGGAAGCCATCGAGCAGGCCACCAAGAGTCGTGAGATCCTCGCCATCTCAGACCCTCAGACTCTGGCCCATGTGCTGACCACTGGGGTGCA
Melursus_ursinus_0              CCAGAGAACCTGATGGGGATGCAGGAAGCCATCGAGCAGGCCACCAAGAGTCGTGAGATCCTCGCCATCTCAGACCCTCAGACTCTGGCCCATGTGCTGACCACTGGGGTGCA
Ursus_americanus_0              CCAGAGAACCTGATGGGGATGCAGGAAGCCATCGAGCAGGCCACCAAGAGTCGTGAGATCCTCGCCATCTCAGACCCTCAGACTCTGGCCCATGTGCTGACCACTGGGGTGCA
Ursus_arctos_0                  CCAGAGAACCTGATGGGGATGCAGGAAGCCATCGAGCAGGCCACCAAGAGTCGTGAGATCCTCGCCATCTCAGACCCTCAGACTCTGGCCCATGTGCTGACCACTGGGGTGCA
Ursus_thibetanus_0              CCAGAGAACCTGATGGGGATGCAGGAAGCCATCGAGCAGGCCACCAAGAGTCGTGAGATCCTCGCCATCTCAGACCCTCAGACTCTGGCCCATGTGCTGACCACTGGGGTGCA
Ursus_maritimus_0               CCAGAGAACCTGATGGGGATGCAGGAAGCCATCGAGCAGGCCACCAAGAGTCGTGAGATCCTCGCCATCTCAGACCCTCAGACTCTGGCCCATGTGCTGACCACTGGGGTGCA
Parictis_montanus_36.6          ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Zaragocyon_daamsi_21.9          ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Ballusia_elmensis_14.0          ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Ursavus_primaevus_14.4          ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Ursavus_brevihinus_16.2         ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Indarctos_vireti_8.7            ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Indarctos_arctoides_9.6         ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Indarctos_punjabiensis_5.0      ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Ailurarctos_lufengensis_7.7     ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Agriarctos_spp_5.0              ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Kretzoiarctos_beatrix_11.7      ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Ursus_abstrusus_4.3             ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Ursus_spelaeus_0.06             ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
Arctodus_simus_0.5              ?????????????????????????????????????????????????????????????????????????????????????????????????????????????????
;
end;

```



List of available packages for BEAST v2.7.*

Name	Installed	Latest	Dependencies	Link	
BEAST.base	2.7.8	2.7.8		🔗	BEAST base
BEAST.app	2.7.8	2.7.8	BEAST.base	🔗	BEAST base applications
Babel		0.4.2	BEAST.base, BEAST.app, BEAST...	🔗	BABEL = BEAST analysis backing effective linguistics
bacter		3.0.1	BEAST.app, feast, BEAST.base	🔗	Bacterial ARG inference.
BADTRIP		2.0.0	BEAST.base	🔗	Infer transmission time for non-haplotype data and epi data
BASTA		4.0.0	BEAST.base, BEAST.app	🔗	Bayesian structured coalescent approximation
bdmm	2.0.0	2.0.0	BEAST.app, MultiTypeTree, BEAS...	🔗	Multitype birth-death model (aka birth-death-migration model)
BDMM-Prime	2.4.0	2.4.0	BEAST.base, BEAST.app, SA	🔗	Birth-death Migration Model Reimplementation
BDSKY	1.5.1	1.5.1	BEAST.base, BEAST.app	🔗	birth death skyline - handles serially sampled tips, piecewise cons
bdtree		0.0.2	BEAST.base, BEAST.app	🔗	Birth-death sequential sampling
BEAST_CLASSIC	1.6.4	1.6.4	BEAST.base, BEAST.app	🔗	BEAST classes ported from BEAST 1 in wrappers
BEASTLabs	2.0.3	2.0.3	BEAST.base, BEAST.app	🔗	BEAST utilities, such as Script, multi monophyletic constraints
BEASTvntr		0.2.0	BEAST.base, BEAST.app	🔗	Variable Number of Tandem Repeat data, such as microsattelites
BICEPS		1.1.2	BEAST.app, BEAST.base	🔗	Bayesian Integrated Coalescent Epoch PlotS + Yule Skyline
bModelTest		1.3.3	BEAST.base, BEAST.app	🔗	Bayesian model test for nucleotide subst models, gamma rate het
BREAK_AWAY		1.2.0	BEASTLabs, BEAST.app, GEO_S...	🔗	break-away model of phylogegraphy
CA		2.1.0	BEAST.base, BEAST.app	🔗	Bayesian estimation of clade ages based on probabilities of fossil
CCD	1.0.3	1.0.3	BEAST.app, BEAST.base	🔗	Sensible tree set summary/point estimates.
ClaDS		2.1.2	BEAST.base, BEAST.app	🔗	Implementation of the ClaDS birth-death tree prior with sampled
CoalRe		1.0.4	BEAST.base, BEAST.app	🔗	Inference of Recombination networks
CodonSubstModels		2.0.0	BEAST.app, BEAST.base	🔗	Codon substitution models for DNA
contacTrees		1.2.0	BEAST.app, BICEPS, feast, BEAS...	🔗	Phylogenetic model with horizontal transfer for linguistics
contraband		1.0.2	BEAST.base, BEAST.app	🔗	Scalable brownian models for continuous trait evolution

☑ Latest

Make sure you have the latest version of ORC package installed (v 1.2.1)

Bayesian divergence time estimation

The data

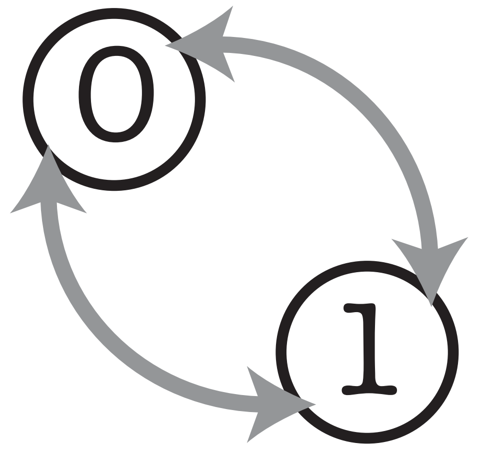
and / or

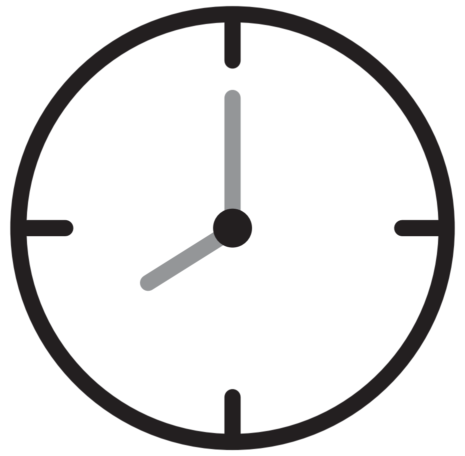
0101... ATTG...
1101... TTGC...
0100... ATTC...

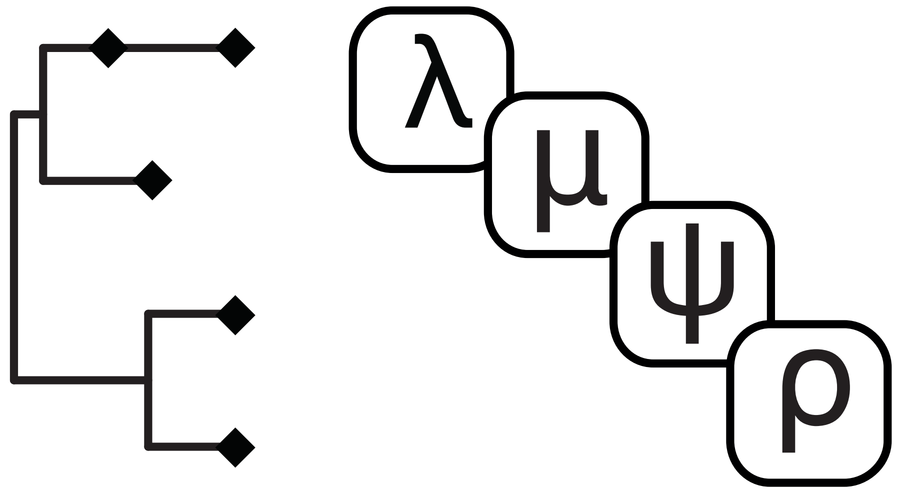
phylogenetics characters

sample ages

3 model components





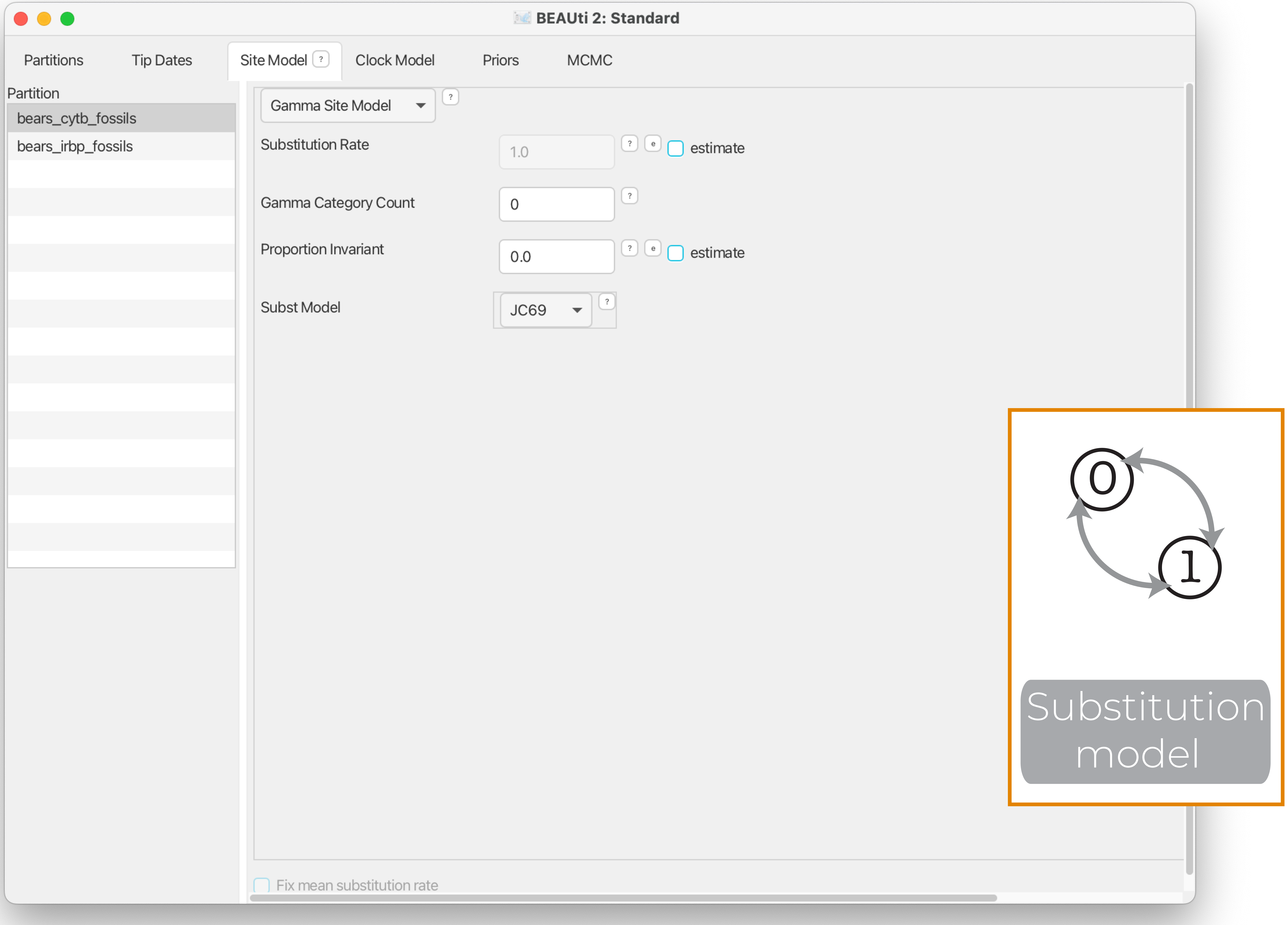


substitution model

clock model

tree and tree model

Understanding the tripartite approach to Bayesian divergence time estimation
Warnock and Wright (2020)



BEAUti 2: Standard

Partition
bears_cytb_fossils
bears_irbp_fossils

Gamma Site Model

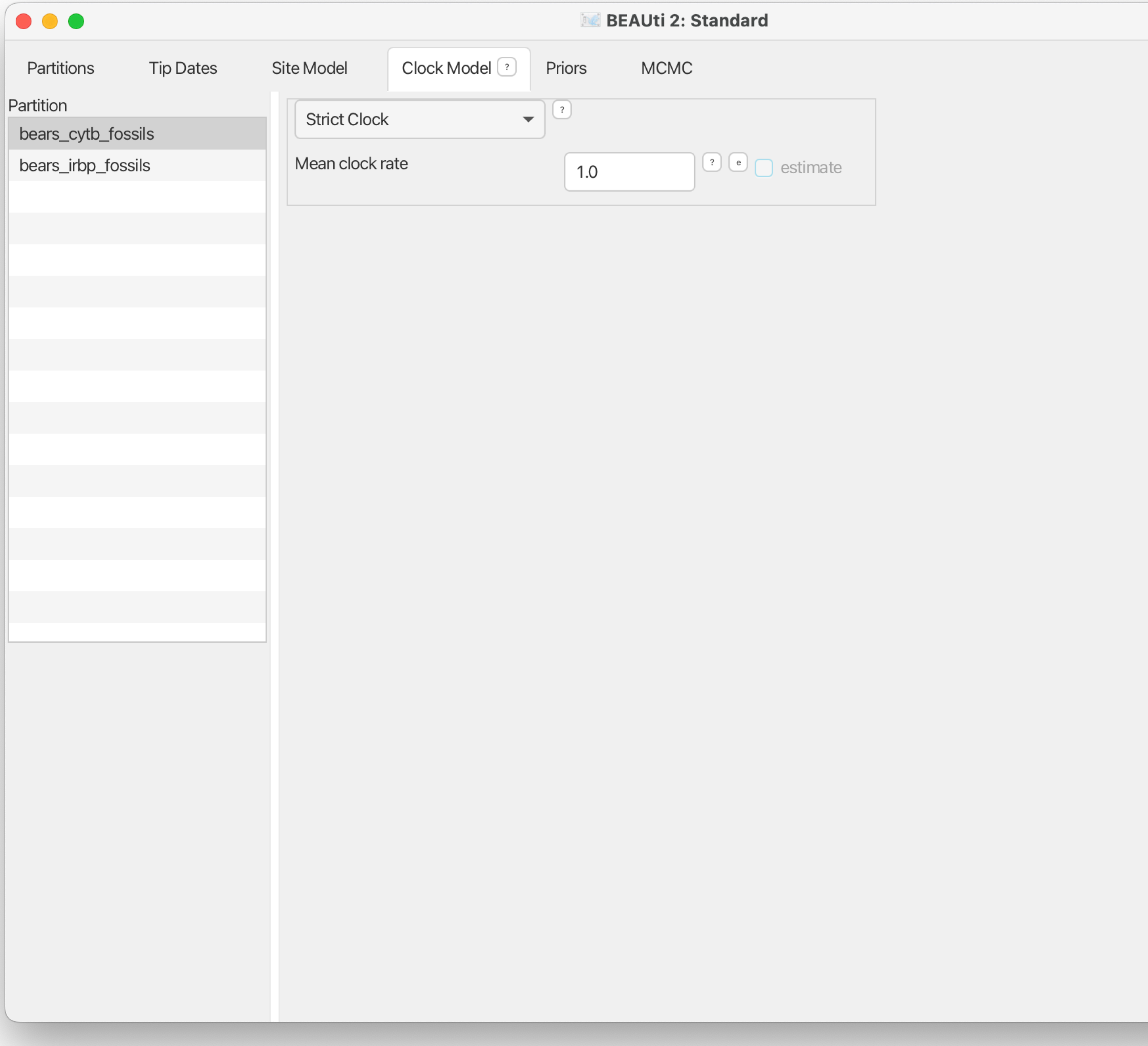
Substitution Rate: 1.0 ☐ estimate

Gamma Category Count: 0

Proportion Invariant: 0.0 ☐ estimate

Subst Model: JC69

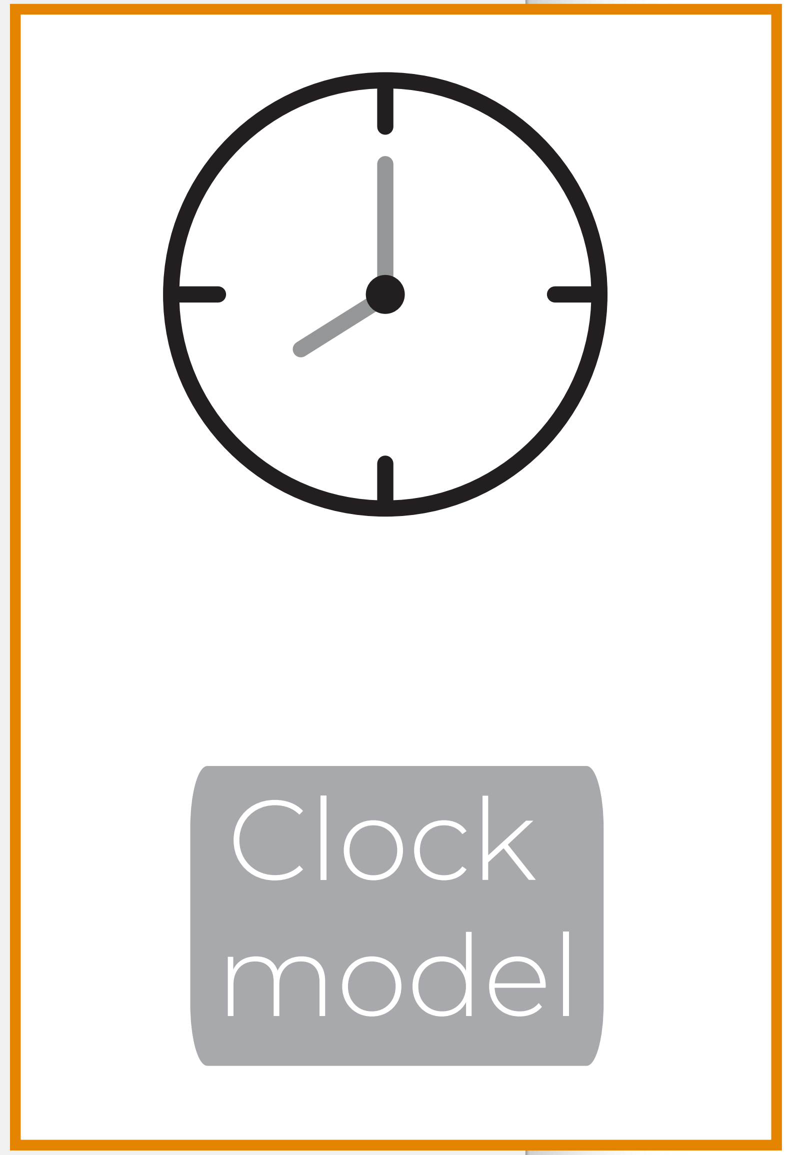
☐ Fix mean substitution rate

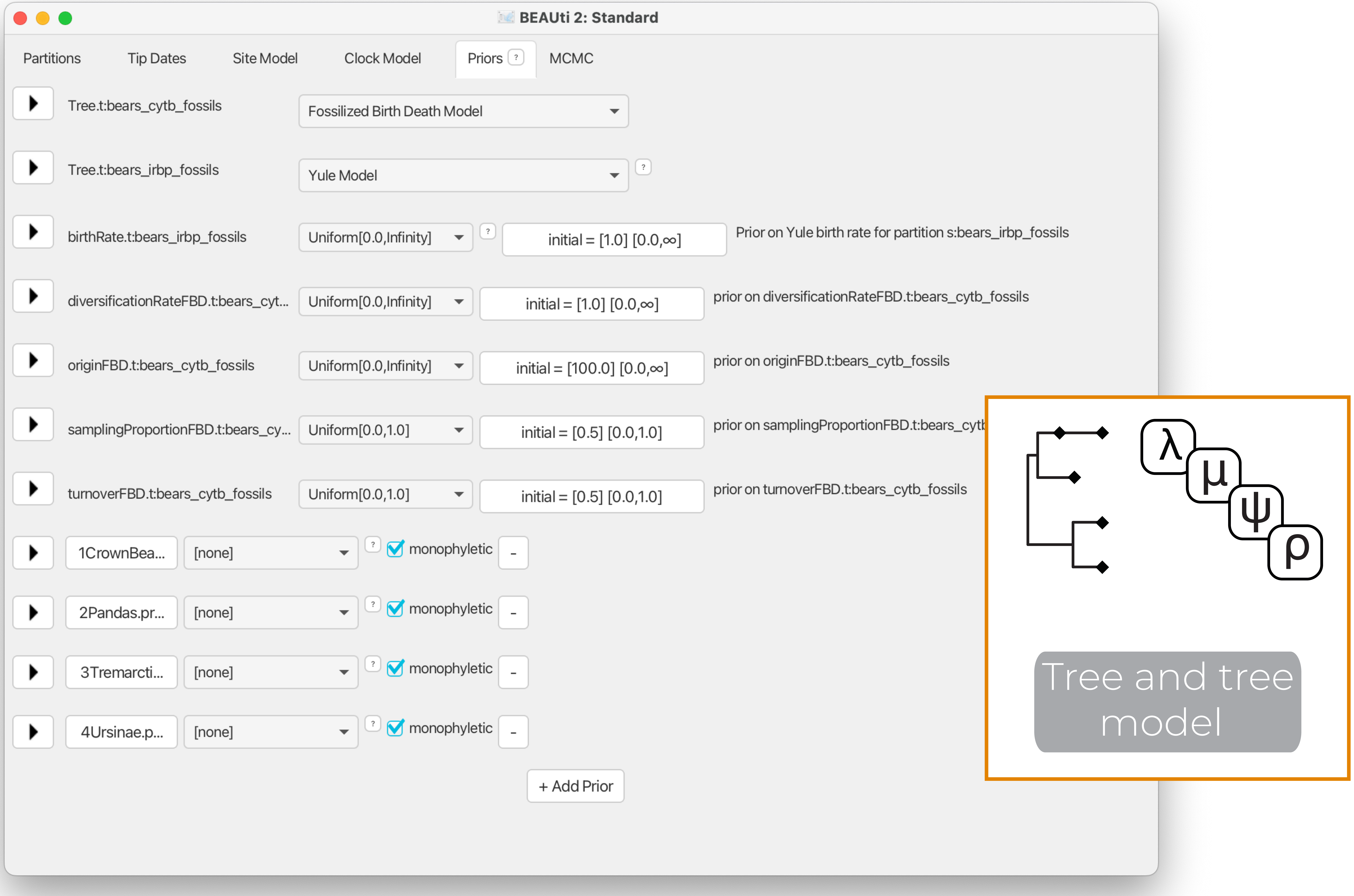


bears_cytb_fossils
bears_irbp_fossils

Strict Clock

Mean clock rate 1.0 ☐ estimate





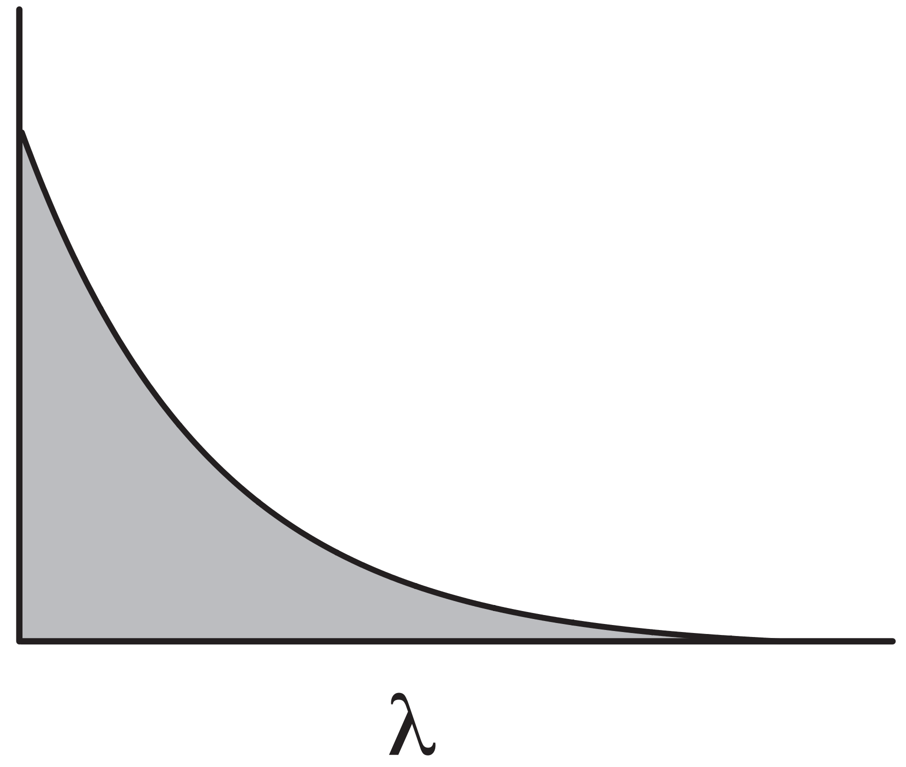
Tree.t:bears_cytb_fossils	Fossilized Birth Death Model		
Tree.t:bears_irbp_fossils	Yule Model		
birthRate.t:bears_irbp_fossils	Uniform[0.0,Infinity]	initial = [1.0] [0.0,∞]	Prior on Yule birth rate for partition s:bears_irbp_fossils
diversificationRateFBD.t:bears_cyt...	Uniform[0.0,Infinity]	initial = [1.0] [0.0,∞]	prior on diversificationRateFBD.t:bears_cytb_fossils
originFBD.t:bears_cytb_fossils	Uniform[0.0,Infinity]	initial = [100.0] [0.0,∞]	prior on originFBD.t:bears_cytb_fossils
samplingProportionFBD.t:bears_cy...	Uniform[0.0,1.0]	initial = [0.5] [0.0,1.0]	prior on samplingProportionFBD.t:bears_cyt
turnoverFBD.t:bears_cytb_fossils	Uniform[0.0,1.0]	initial = [0.5] [0.0,1.0]	prior on turnoverFBD.t:bears_cytb_fossils
1CrownBea...	[none]	☑ monophyletic	
2Pandas.pr...	[none]	☑ monophyletic	
3Tremarcti...	[none]	☑ monophyletic	
4Ursinae.p...	[none]	☑ monophyletic	

Tree and tree model

Key model parameters

- birth (speciation) rate – λ
- death (speciation) rate – μ
- fossil sampling rate – ψ
- extant sampling probability – *ρ*

birth, death and sampling are instantaneous rates and typically estimated during inference



extant species sampling is often fixed because this is the parameter we have the most information about

Parameterisation of the process

We can put priors on different combinations of parameters

Parameter	**Transformation**
Net diversification	$d = \lambda - \mu$
Turnover	$v = \mu/\lambda$
Sampling proportion	$s = \psi/(\mu + \psi)$
Speciation	$\lambda = d/(1 - v)$
Extinction	$\mu = (vd)(1 - v)$
Sampling	$\psi = (s/(1 - s))((vd)/(1 - v))$

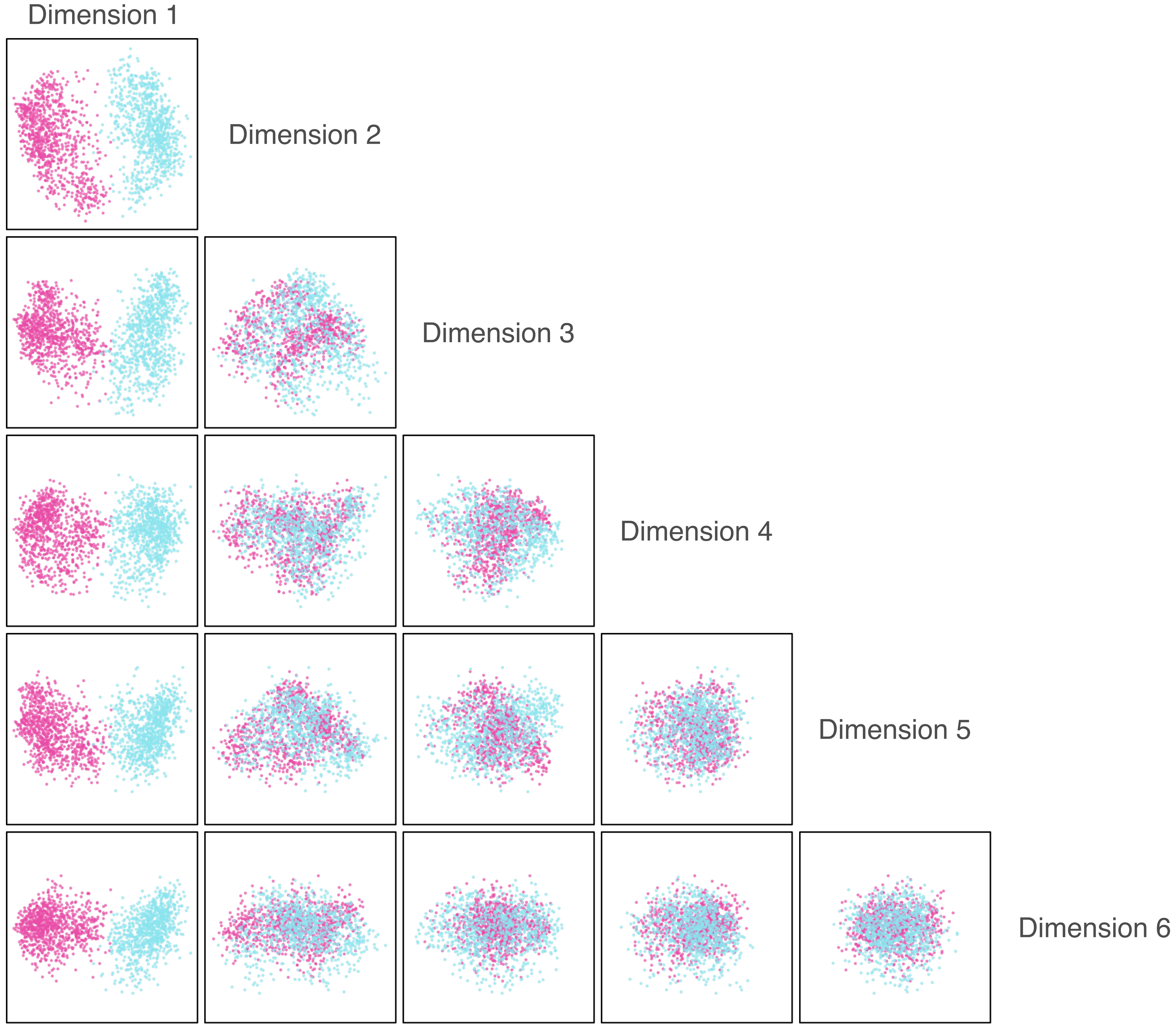
Take care when interpreting PP for trees with SAs

More options available via the pdf!

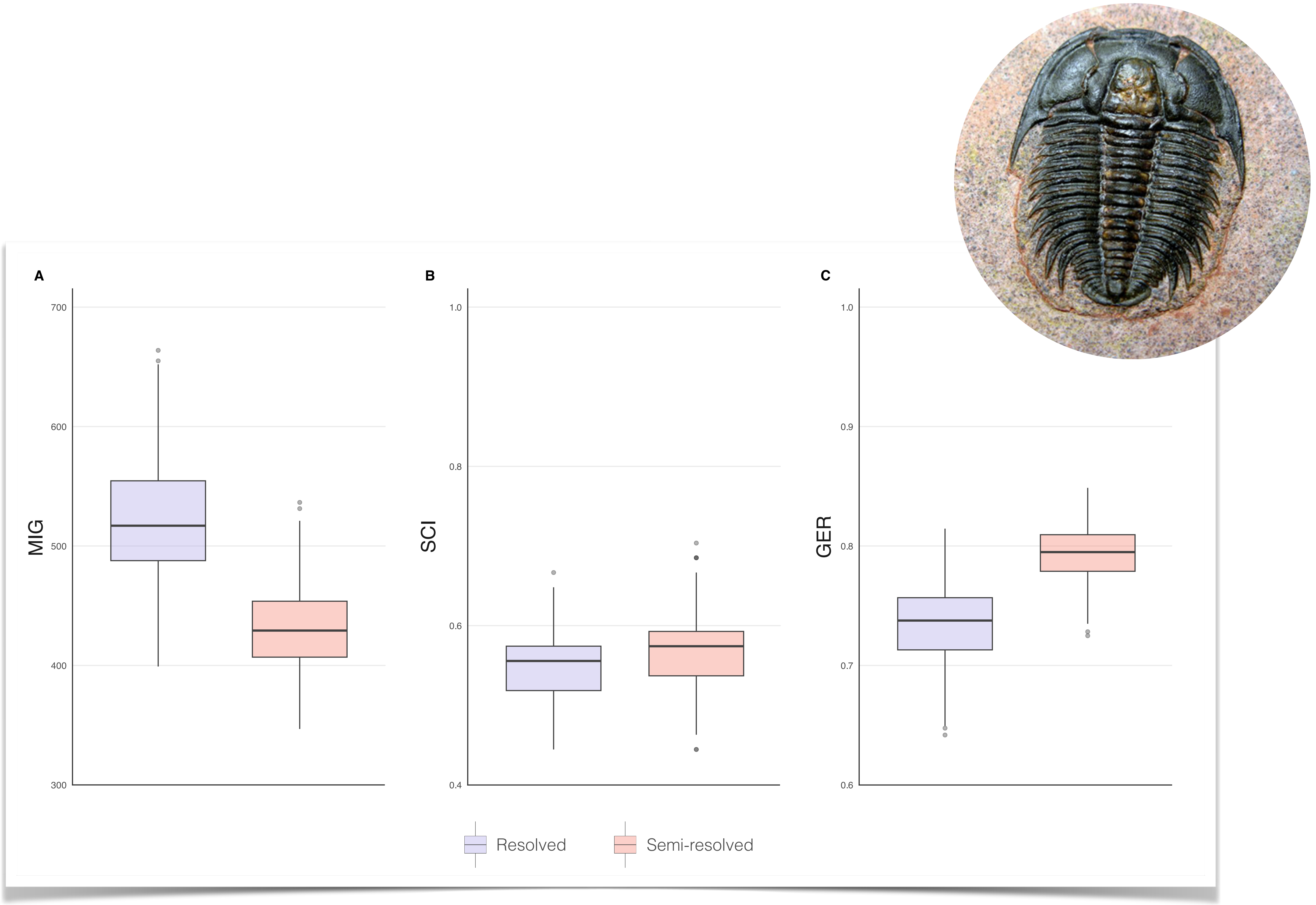
e.g., alternative parameterisations, fully extinct clades

Extra slides

Including samples with no character data improves inference



Trees occupy different regions of tree space

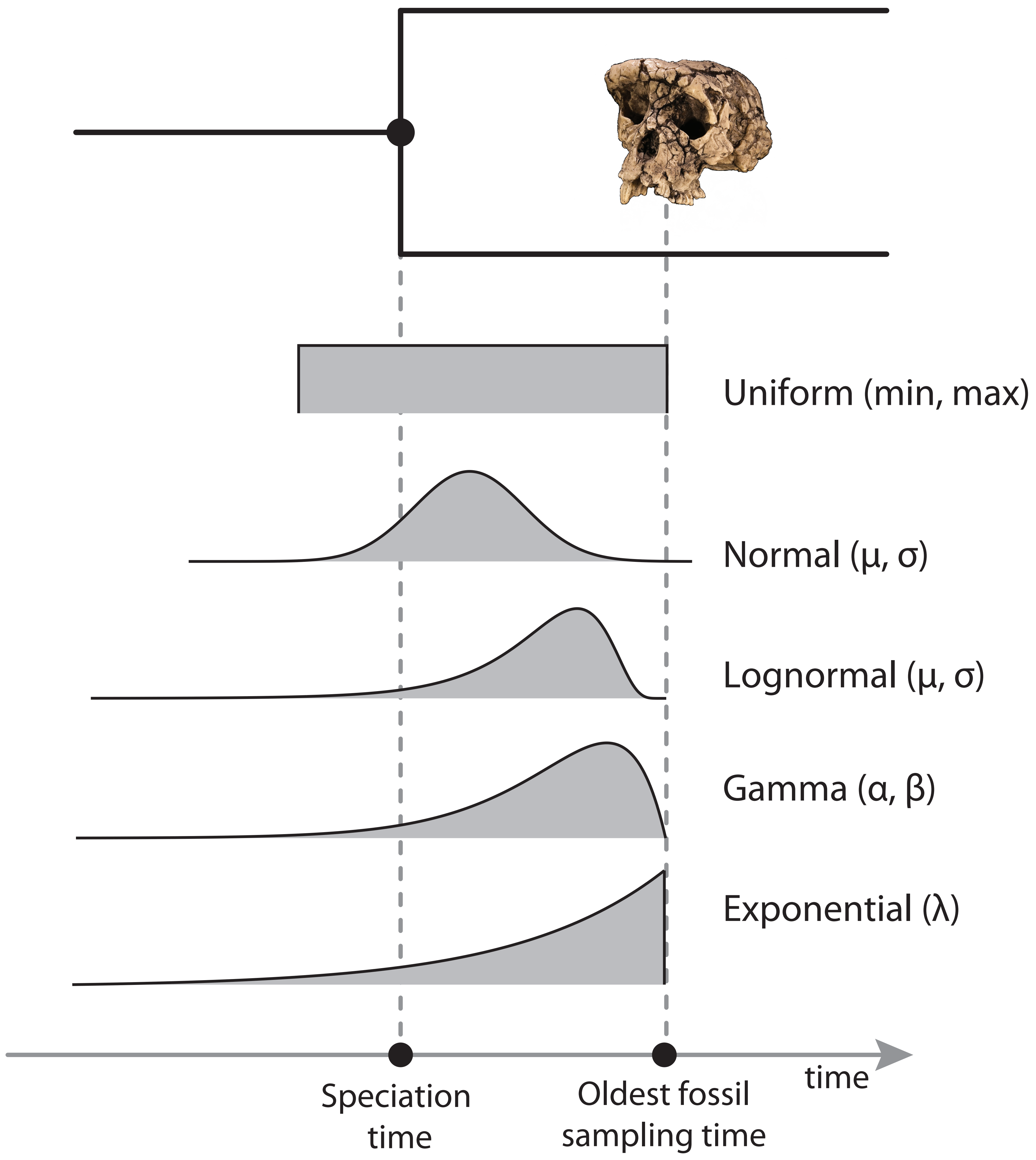


And show worse / better correspondence with the fossil record

Nikolic et al.. in review.

Scenario	**Software**
Node dating with large datasets	MCMCTree
If want (or have to) fix the tree topology	MCMCTree
If fossil sampling is sparse or complex	MCMCTree
If you have abundant fossil data, or are interested in the topological position of fossils	BEAST2, RevBayes
If you’re interested in the phylodynamic parameters	BEAST2, RevBayes
If you want to use a specific model	BEAST2, RevBayes, MCMCTree

Recap: Node dating



We can use a calibration density to constrain internal node ages

We typically use a birth-death process model to describe the tree generating process

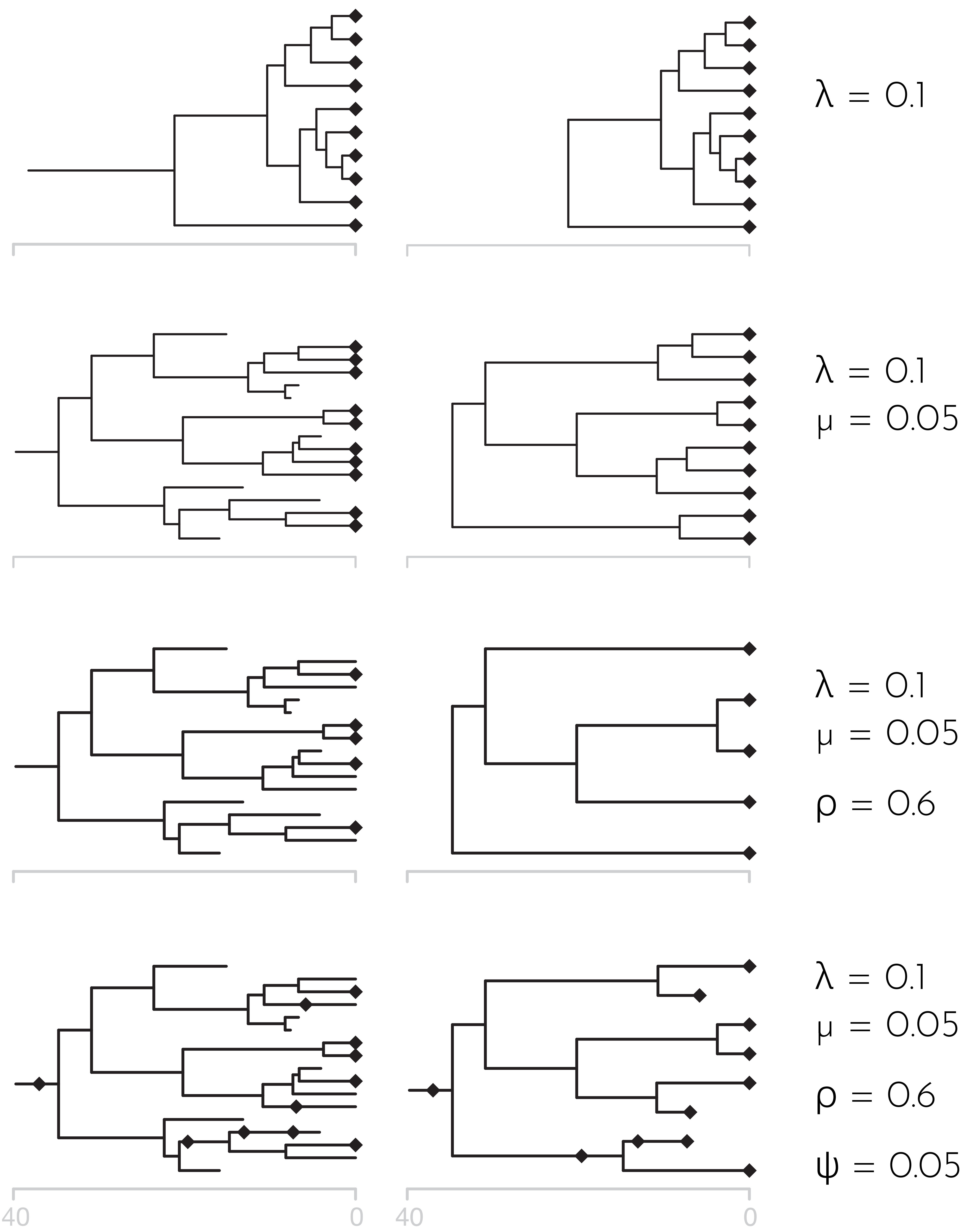
Adapted from *Heath* (2012). *Sys Bio*

Relationship to (some) other birth-death process models

We can also use *ρ* at *t > 0* to model serial sampling

For epidemiology: *Stadler et al.* (2012)
See also: *Stadler and Yang* (2013)
Review: *MacPherson et al.* (2022)
Image: *Warnock and Wright* (2020)

complete vs. reconstructed trees



Yang and Rannala (1997)
Stadler (2009)

Stadler (2010)

Table 2. Available fossilized birth–death (FBD) models and extensions

Model name	Description	Reference
FBD specimen process	Model assumes constant rates of diversification (λ, μ) and sampling through time (ψ) and assumes uniform sampling of extant taxa.	Stadler (2010); Heath et al. (2014)
FBD skyline	Model allows for diversification (λ_t, μ_t) and sampling rates (ψ_t) to vary through time across discrete intervals (i.e., in a piecewise-constant fashion).	Stadler et al (2012); Gavryushkina et al. (2014)
Diversified sampling	Model accounts for a non-uniform sampling strategy that aims to maximize the phylogenetic diversity represented in the extant tree. This is achieved by introducing the parameter x_{cut}, which represents the youngest node age, after which we have no more sampled nodes.	Höhna et al. (2011); Zhang et al. (2016)
Multi-type birth–death process	Model allows for variation in diversification and fossil sampling rates across different parts of the tree associated with different “types”. Types may correspond to any trait that effects differences in rates. Lineages can switch between types with rate *m.*	Kühnert et al. (2016); Barido-Sottani et al. (2020c)
Occurrence birth–death process	Model allows for the incorporation of data from fossil observations outside the tree, i.e., those for which we have no taxonomic information or morphological data, such as trace fossils. These observations are modeled using a separate sampling process, with rate parameter ω.	Manceau et al. (2021); Gupta et al. (2020); Andréoletti et al. (2022)
Episodic FBD process	Model allows for instantaneous speciation, extinction, or sampling events to affect the entire tree, corresponding to events such as mass extinctions or species radiations.	Magee and Höhna (2021)
FBD multispecies coalescent	Model allows separate genes to evolve independently under a coalescent model.	Ogilvie et al. (2022)
FBD range process	Model provides the framework to explicitly take stratigraphic range data into account, thereby associating multiple occurrences within a range of the same taxon.	Stadler et al. (2018)

Mulvey et al. (2025) Paleobiology 30

> *“It is, it must be admitted, a* ***humbling*** *task to infer ancient events, and the results in many cases are tenuous at best. Given the obvious limitations of working with extant species and few, if any, fossils,* ***it is necessary to integrate all of the available sources of evidence*** *if we hope to produce assuring answers.”*

Landis et al. (2023) *Systematic Biology*
Joint phylogenetic estimation of geographic movements and biome shifts

Further reading

More on FBD models and phylodynamics

From fossils to phylogenies: exploring the integration of paleontological data into Bayesian phylogenetic inference. *Mulvey et al*. (2025)

Integrating fossil observations into phylogenetics using the fossilized birth-death model. *Wright et al.* (2022)

Phylogenetic insights into diversification. *Morlon et al*. (2024)

Unifying birth-death models in epidemiology and macroevolution. *MacPherson et al.* (2022)

decodinggenomes.org (free pdf available) *Stadler et al.* (2024) Chap 9, Phylodynamics

Models of morphological evolution

A systematist’s guide to estimating Bayesian phylogenies from morphological data. *Wright* ([2019](2019))

Assessing the adequacy of morphological models using posterior predictive simulations. *Mulvey et al.* ([2024](2024))