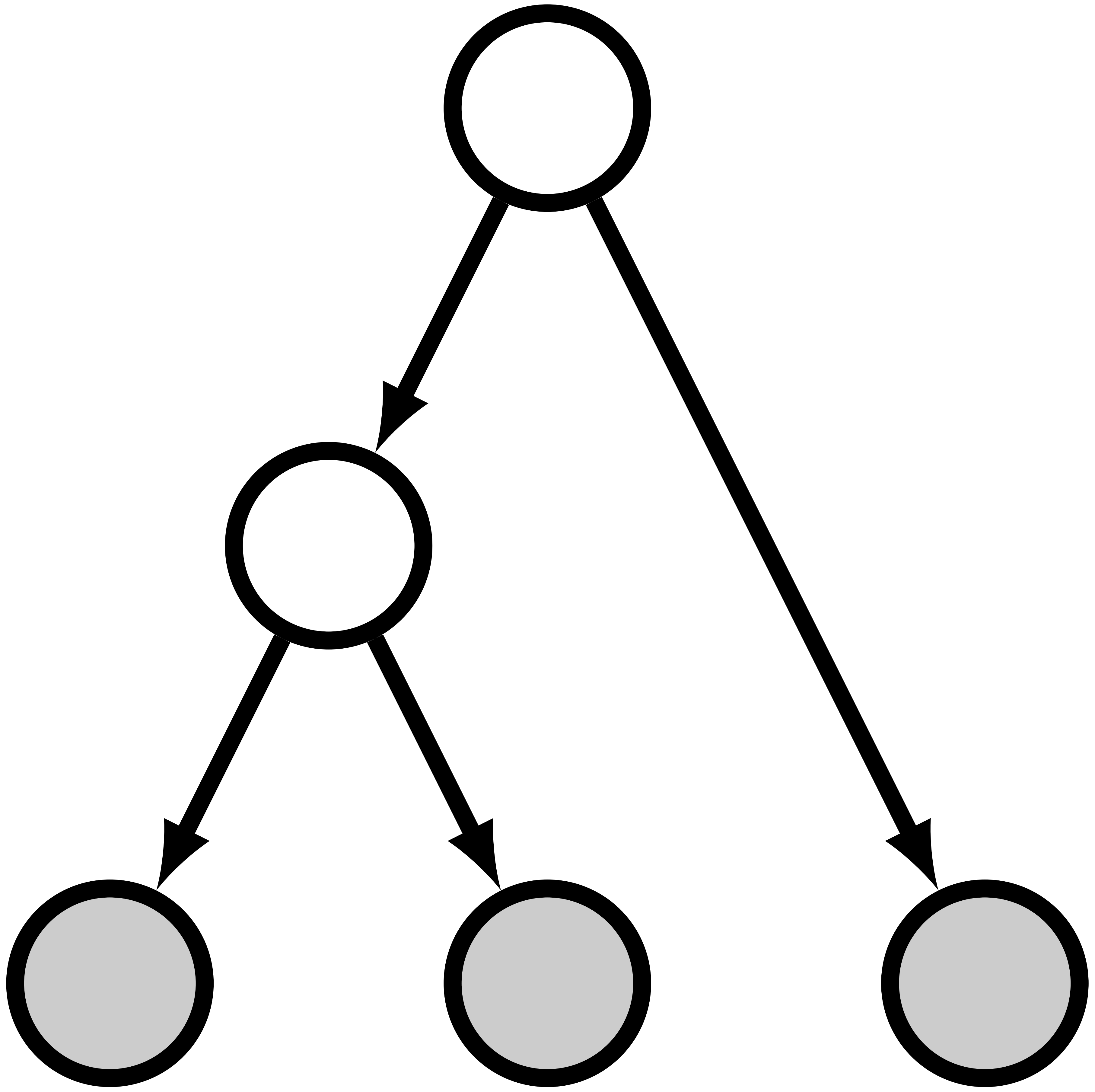
Intro to Bayesian inference using RevBayes

Analytical Paleobiology workshop

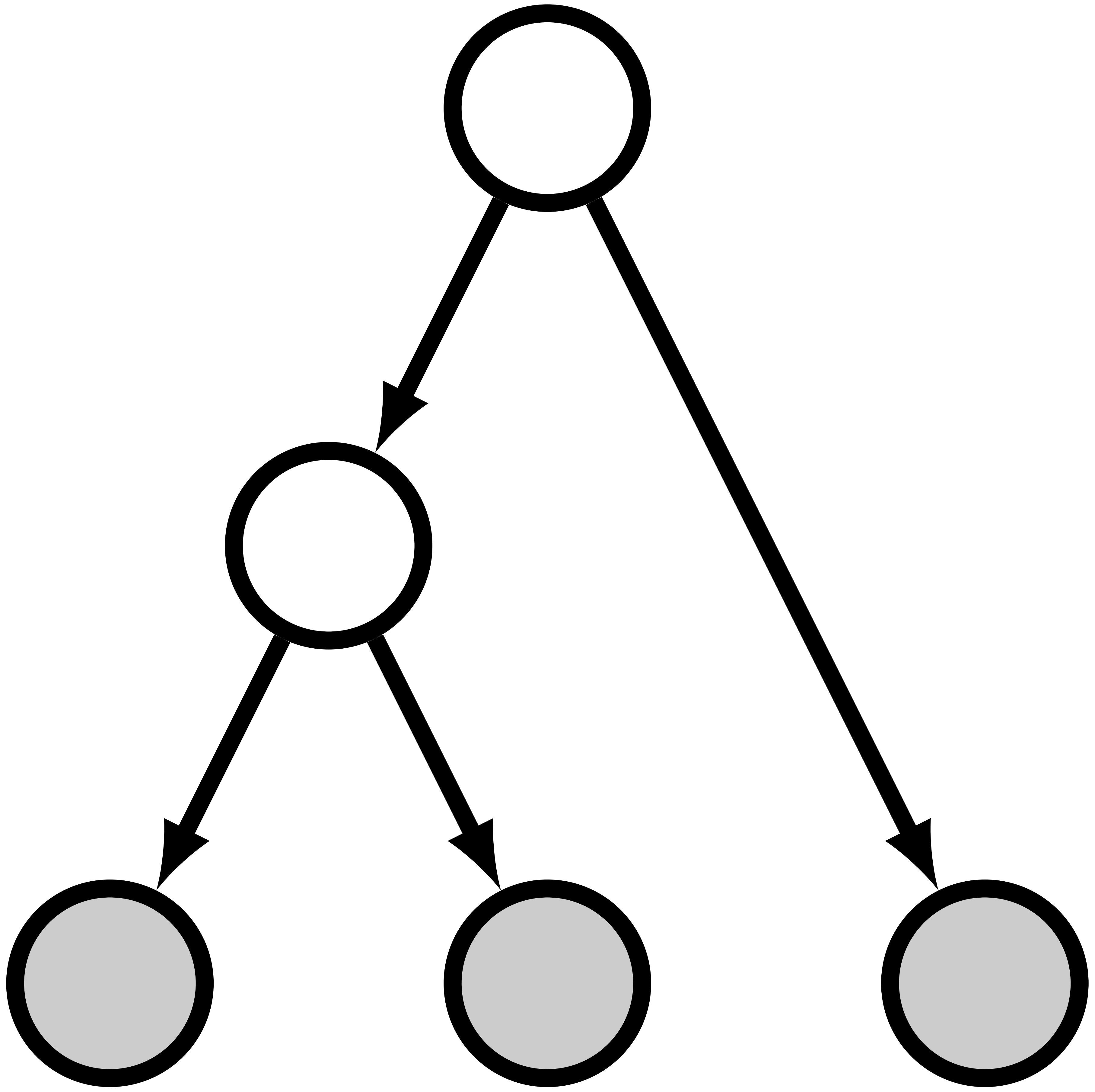
Rachel Warnock

28.08.25



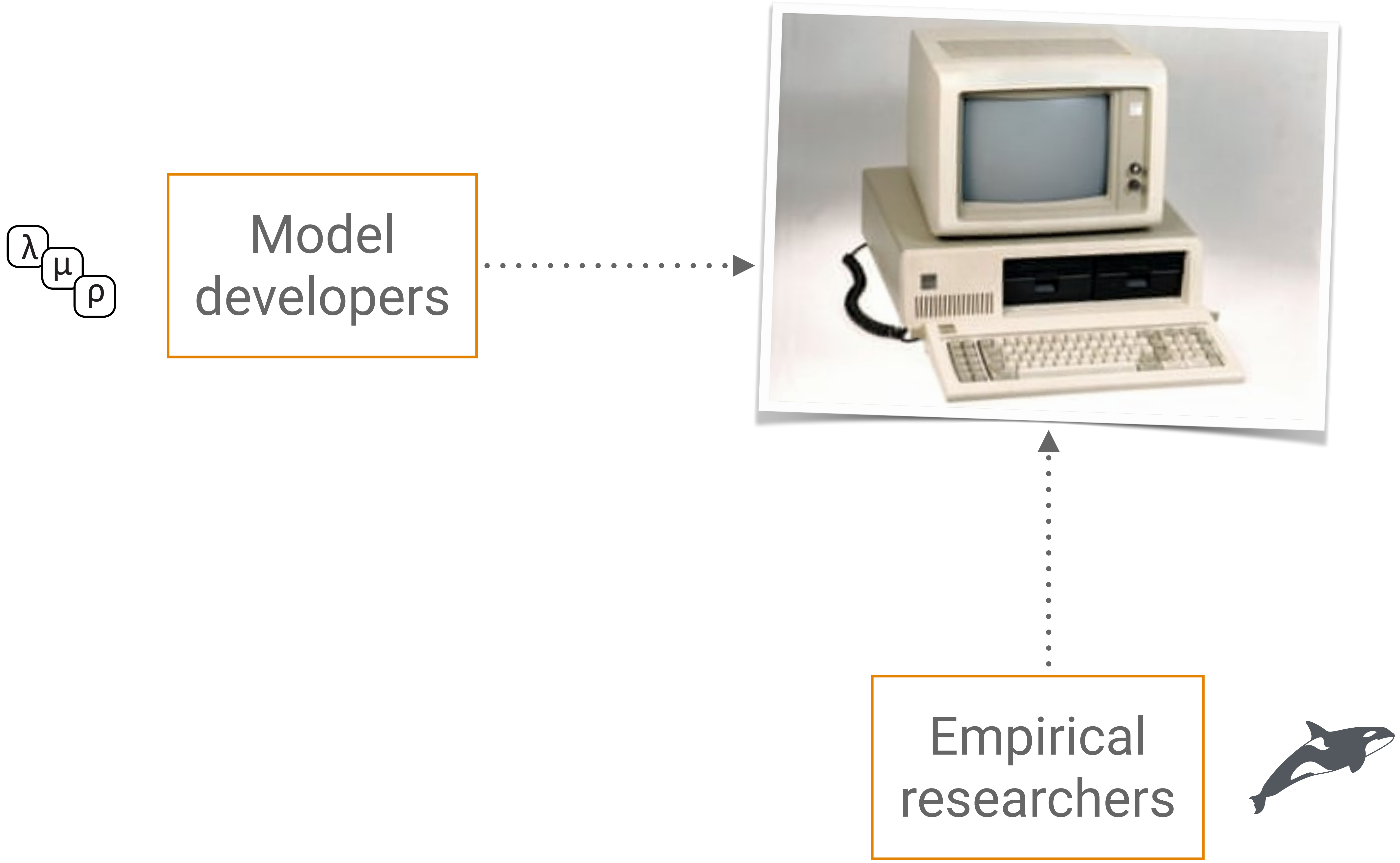
Objectives

- Graphical models
- RevBayes
- Recap: Bayesian inference
- Recap: MCMC



RevBayes

Phylogenetic inference — the old way

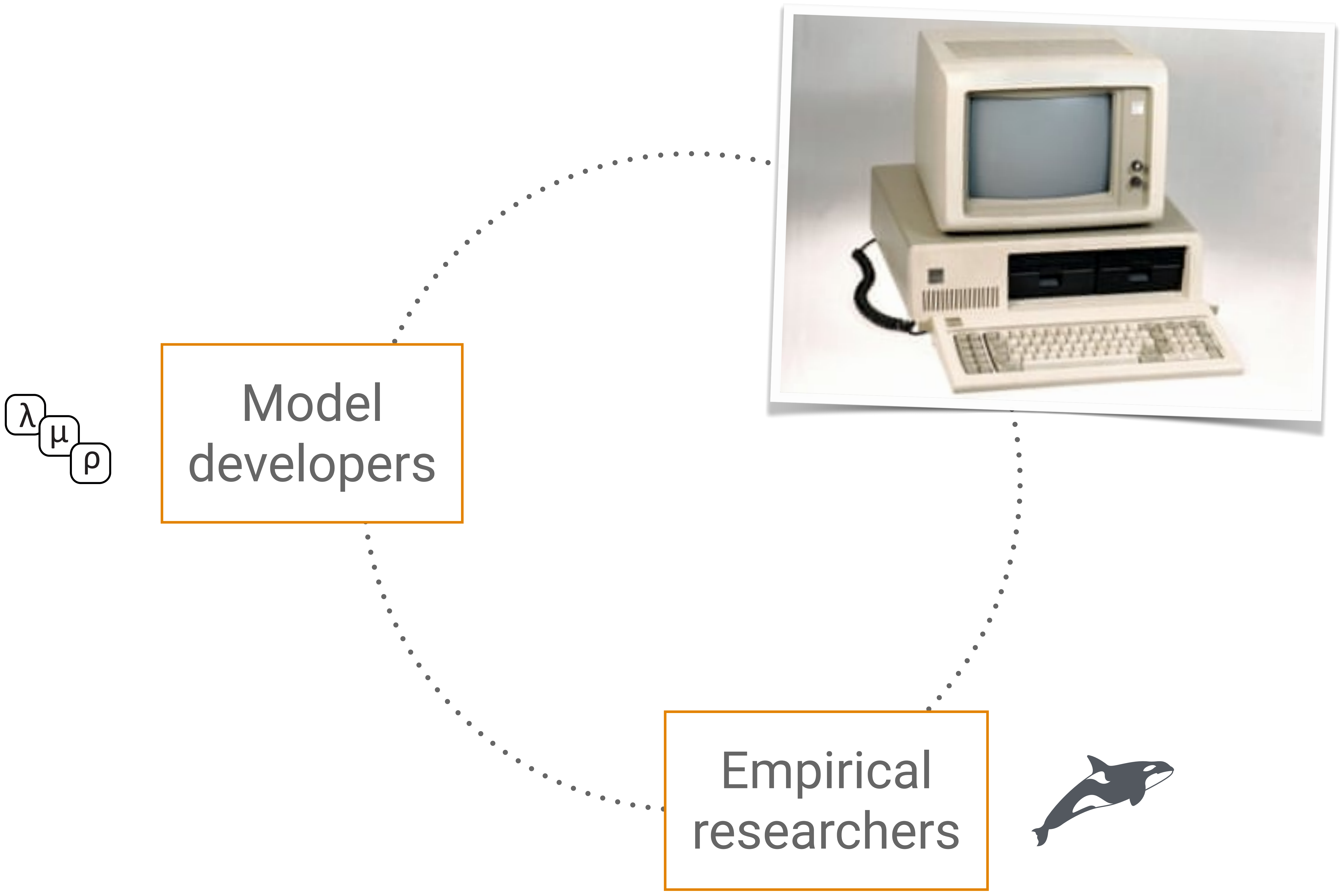


Model developers

Empirical researchers

What we might call a
“black box” approach

Phylogenetic inference — a better way?



Model developers

Empirical researchers

The goal is to bring researchers with different expertise together, increase transparency, and do better research

RevBayes

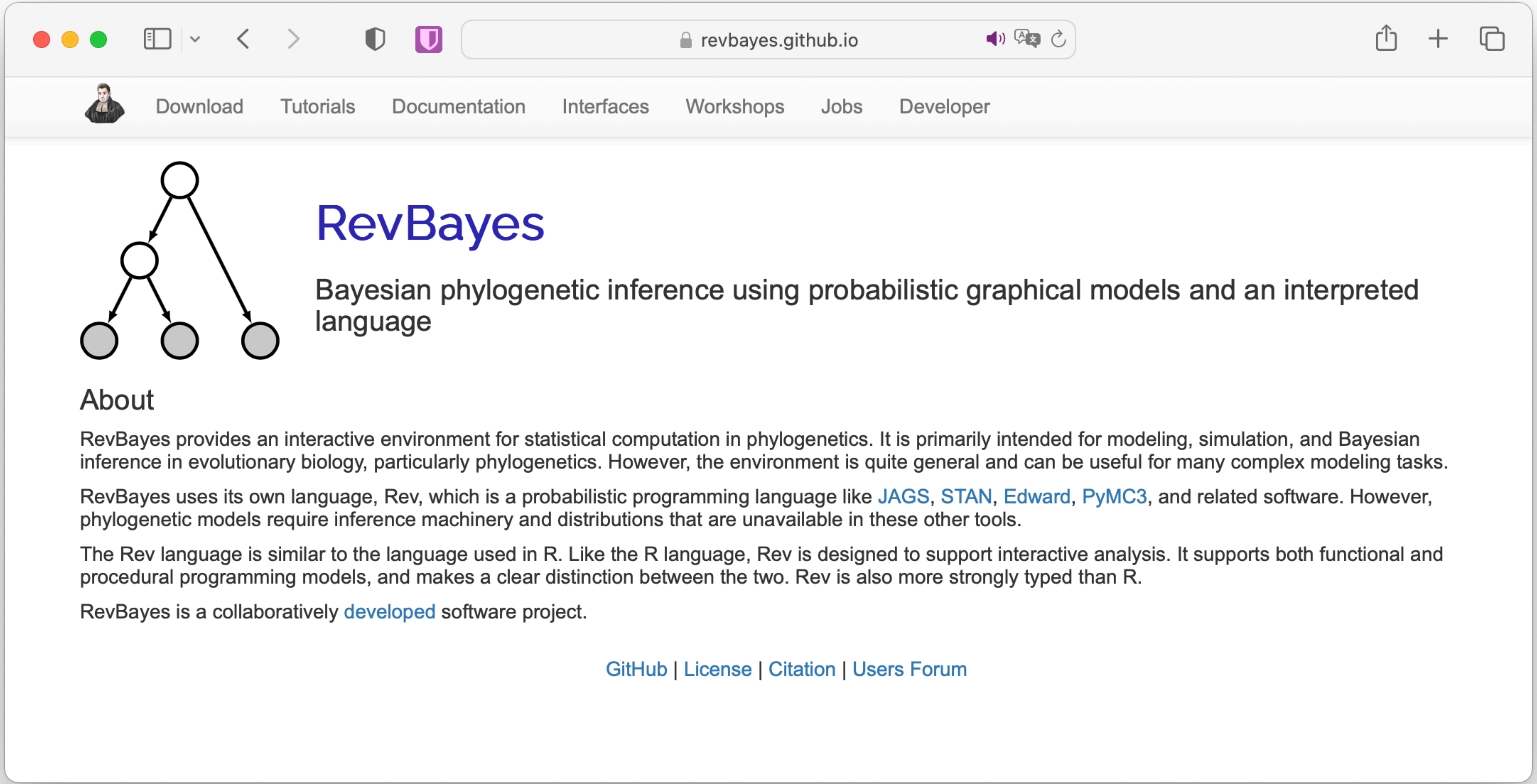
Named after Reverend Bayes, “descended” from the software MrBayes

Designed with extendability and flexibility in mind

Rev language, similar to R, and uses a graphical modelling framework

Developed and supported by a large international team of developers

revbayes.github.io



RevBayes

Bayesian phylogenetic inference using probabilistic graphical models and an interpreted language

About

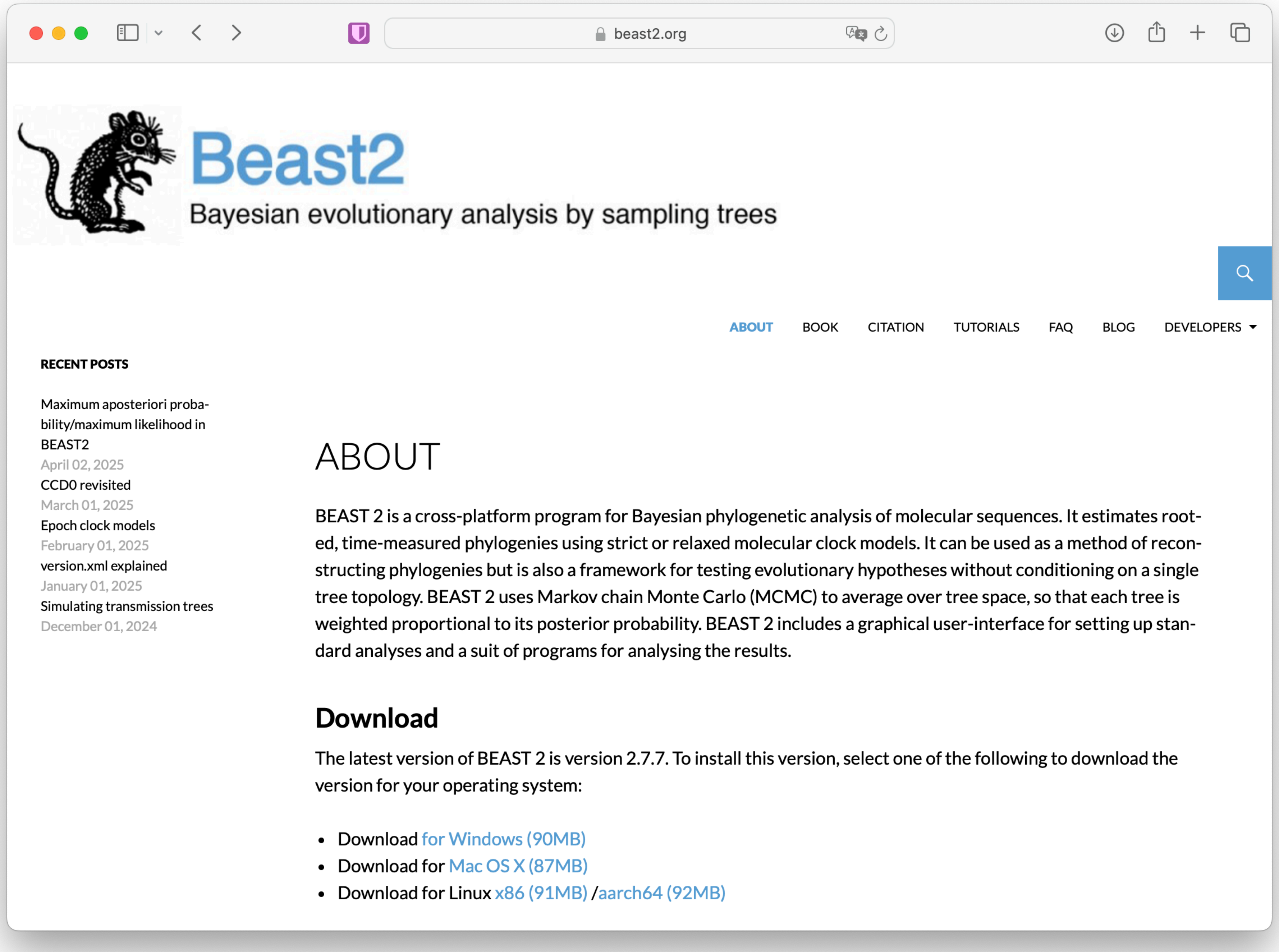
RevBayes provides an interactive environment for statistical computation in phylogenetics. It is primarily intended for modeling, simulation, and Bayesian inference in evolutionary biology, particularly phylogenetics. However, the environment is quite general and can be useful for many complex modeling tasks.

RevBayes uses its own language, Rev, which is a probabilistic programming language like JAGS, STAN, Edward, PyMC3, and related software. However, phylogenetic models require inference machinery and distributions that are unavailable in these other tools.

The Rev language is similar to the language used in R. Like the R language, Rev is designed to support interactive analysis. It supports both functional and procedural programming models, and makes a clear distinction between the two. Rev is also more strongly typed than R.

RevBayes is a collaboratively developed software project.

GitHub | License | Citation | Users Forum



Beast2

Bayesian evolutionary analysis by sampling trees

RECENT POSTS

Maximum aposteriori probability/maximum likelihood in BEAST2
April 02, 2025
CCD0 revisited
March 01, 2025
Epoch clock models
February 01, 2025
version.xml explained
January 01, 2025
Simulating transmission trees
December 01, 2024

ABOUT

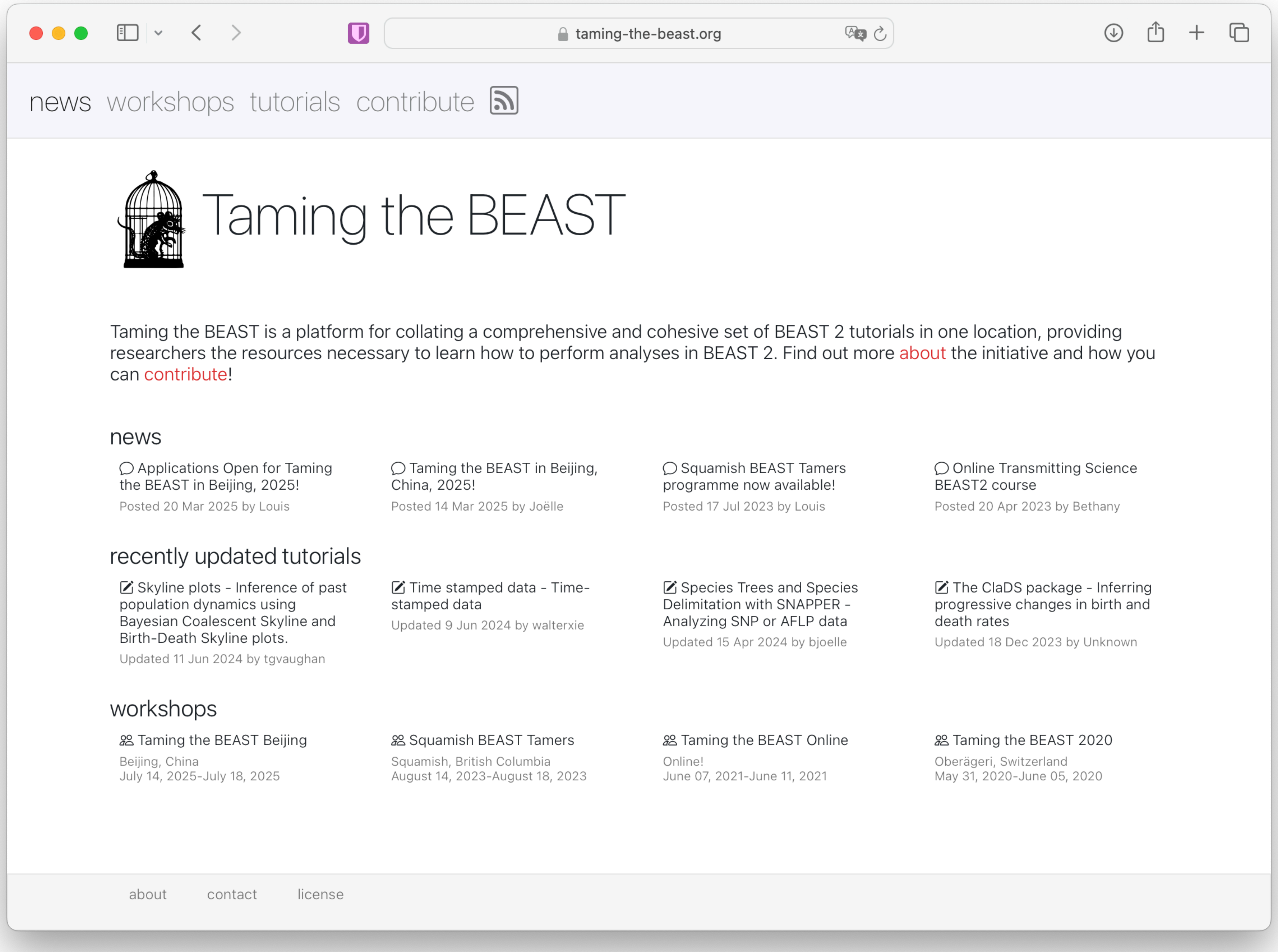
BEAST 2 is a cross-platform program for Bayesian phylogenetic analysis of molecular sequences. It estimates rooted, time-measured phylogenies using strict or relaxed molecular clock models. It can be used as a method of reconstructing phylogenies but is also a framework for testing evolutionary hypotheses without conditioning on a single tree topology. BEAST 2 uses Markov chain Monte Carlo (MCMC) to average over tree space, so that each tree is weighted proportional to its posterior probability. BEAST 2 includes a graphical user-interface for setting up standard analyses and a suit of programs for analysing the results.

Download

The latest version of BEAST 2 is version 2.7.7. To install this version, select one of the following to download the version for your operating system:

- Download for Windows (90MB)
- Download for Mac OS X (87MB)
- Download for Linux x86 (91MB) /aarch64 (92MB)

www.beast2.org



Taming the BEAST

Taming the BEAST is a platform for collating a comprehensive and cohesive set of BEAST 2 tutorials in one location, providing researchers the resources necessary to learn how to perform analyses in BEAST 2. Find out more about the initiative and how you can contribute!

news

- Applications Open for Taming the BEAST in Beijing, 2025!
  Posted 20 Mar 2025 by Louis
- Taming the BEAST in Beijing, China, 2025!
  Posted 14 Mar 2025 by Joëlle
- Squamish BEAST Tamers programme now available!
  Posted 17 Jul 2023 by Louis
- Online Transmitting Science BEAST2 course
  Posted 20 Apr 2023 by Bethany

recently updated tutorials

- Skyline plots - Inference of past population dynamics using Bayesian Coalescent Skyline and Birth-Death Skyline plots.
  Updated 11 Jun 2024 by tgvaughan
- Time stamped data - Time-stamped data
  Updated 9 Jun 2024 by walterxie
- Species Trees and Species Delimitation with SNAPPER - Analyzing SNP or AFLP data
  Updated 15 Apr 2024 by bjoelle
- The ClaDS package - Inferring progressive changes in birth and death rates
  Updated 18 Dec 2023 by Unknown

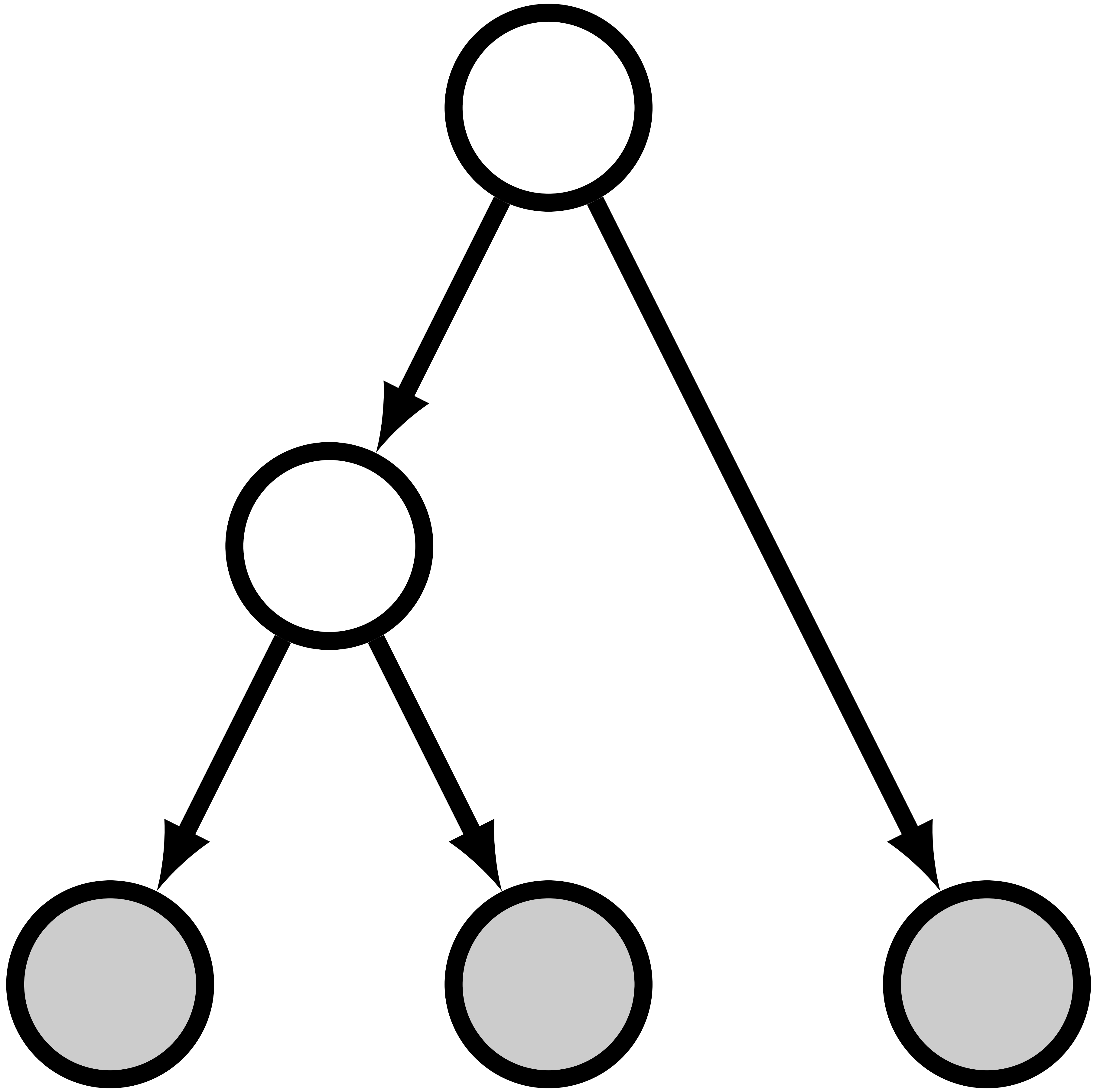
workshops

- Taming the BEAST Beijing
  Beijing, China
  July 14, 2025-July 18, 2025
- Squamish BEAST Tamers
  Squamish, British Columbia
  August 14, 2023-August 18, 2023
- Taming the BEAST Online
  Online!
  June 07, 2021-June 11, 2021
- Taming the BEAST 2020
  Oberägeri, Switzerland
  May 31, 2020-June 05, 2020

taming-the-beast.org

Graphical models

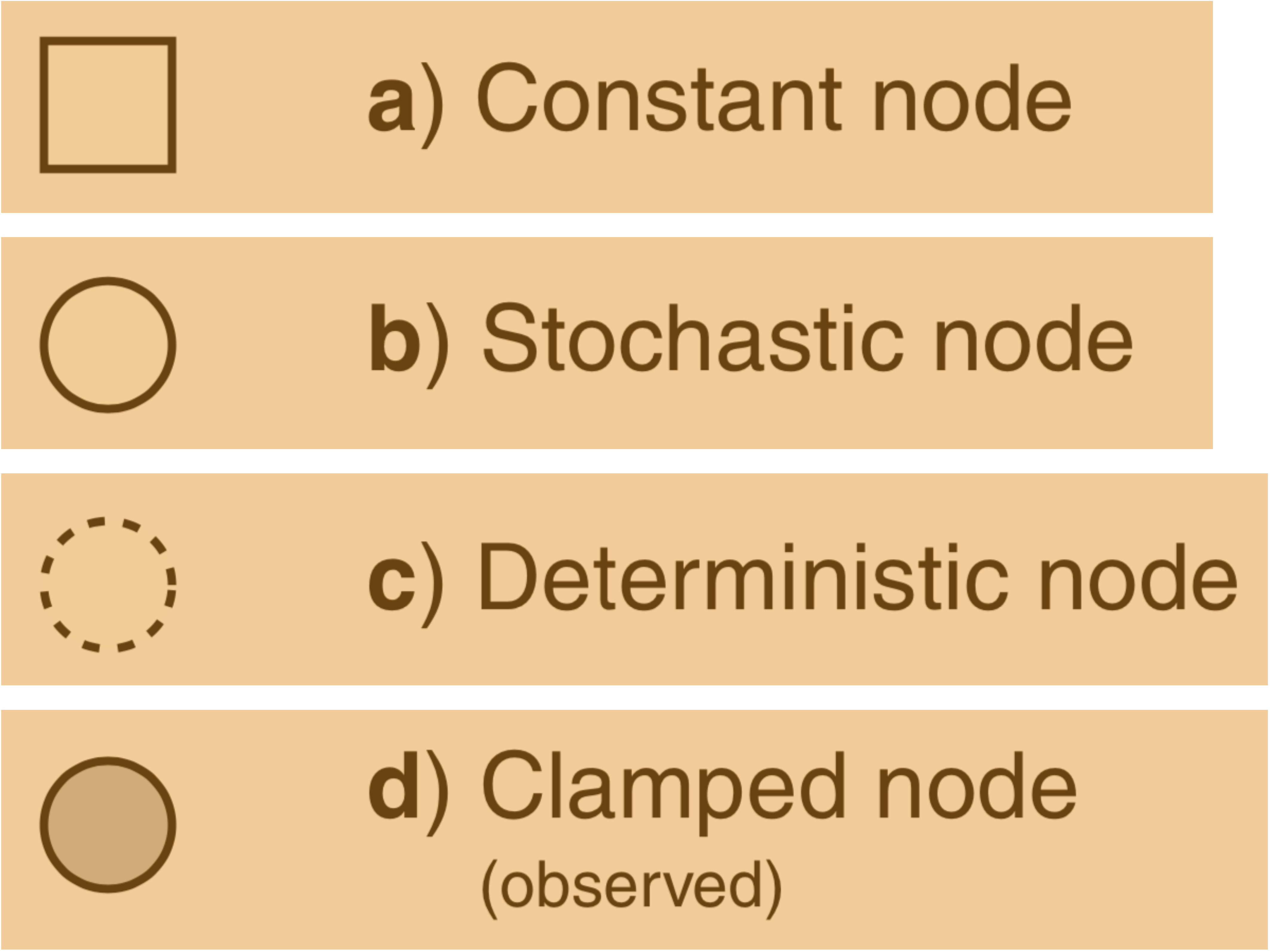
Graphical models



Provide tools for visually and computationally representing complex, parameter-rich models

Depict the conditional dependence structure of parameters and other random variables

Types of variables (nodes)

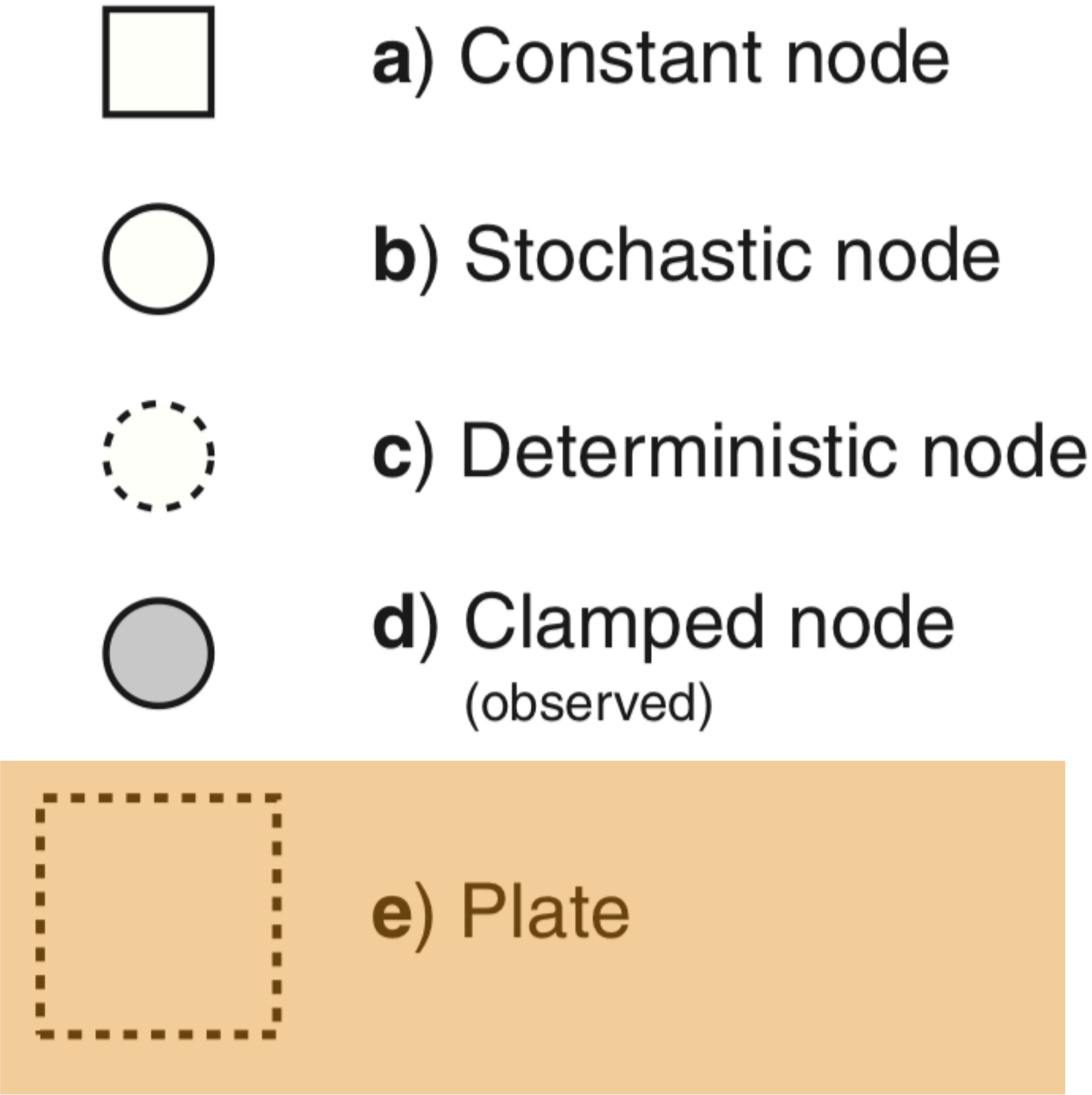


a. fixed value variables

b. random variables that depend on other variables

c. variables determined by a function applied other variables (transformations)

d. observed stochastic variables (data)



a. fixed value variables

b. random variables that depend on other variables

c. variables determined by a function applied other variables (transformations)

d. observed stochastic variables (data)

e. repetition over multiple variables (equivalent to a loop)

Specifying graphical models using the Rev syntax

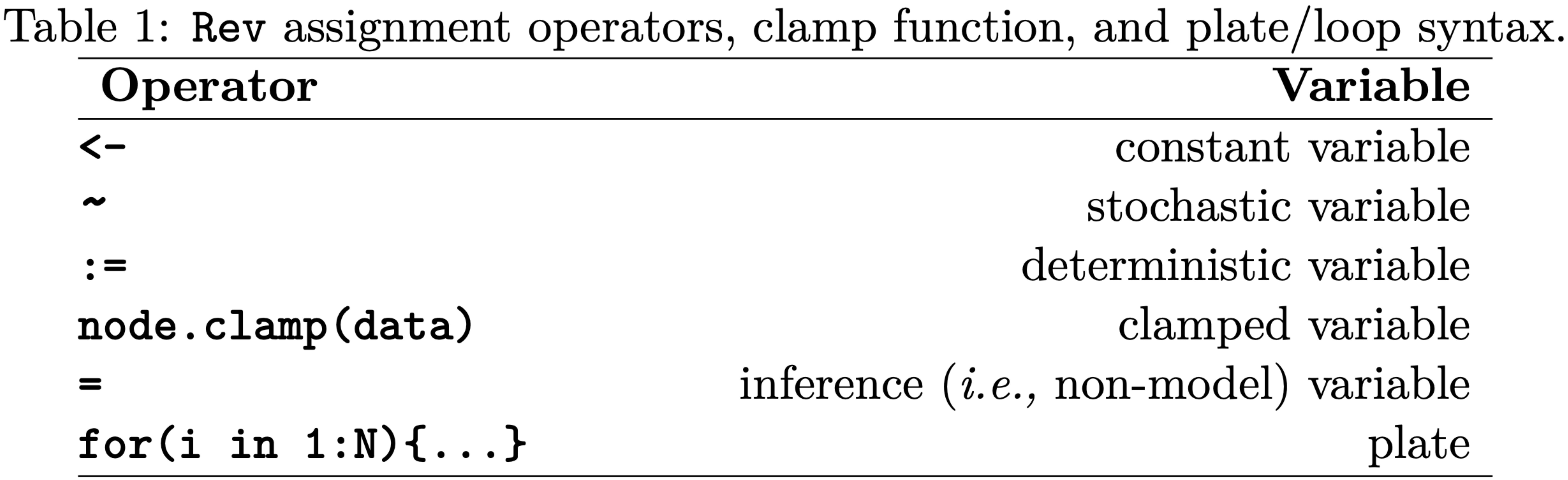
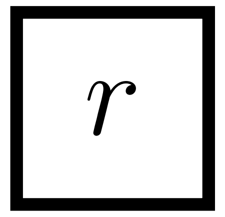


Table 1: `Rev` assignment operators, clamp function, and plate/loop syntax.

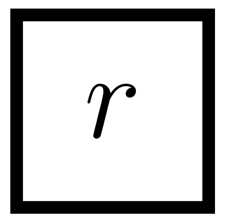
Operator	**Variable**
`<-`	constant variable
`~`	stochastic variable
`:=`	deterministic variable
`node.clamp(data)`	clamped variable
`=`	inference (*i.e.*, non-model) variable
`for(i in 1:N){...}`	plate

a)

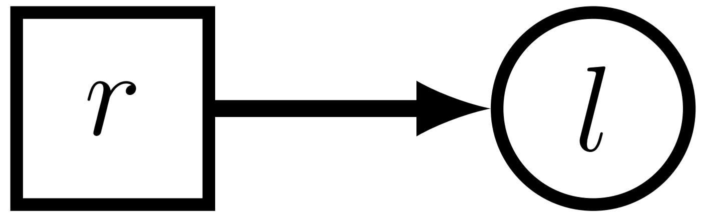


```
# constant node
r <- 10
```

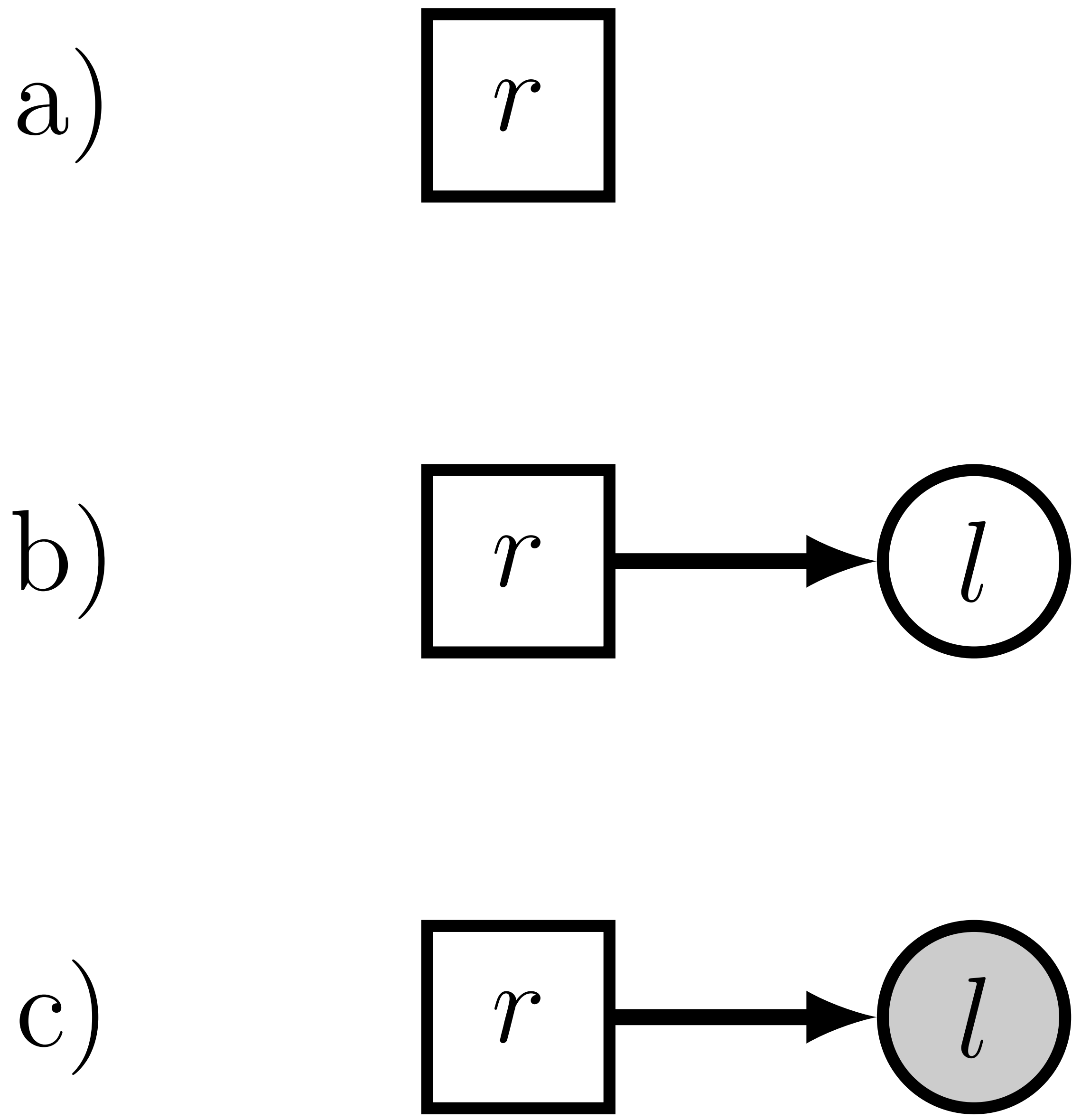
a)



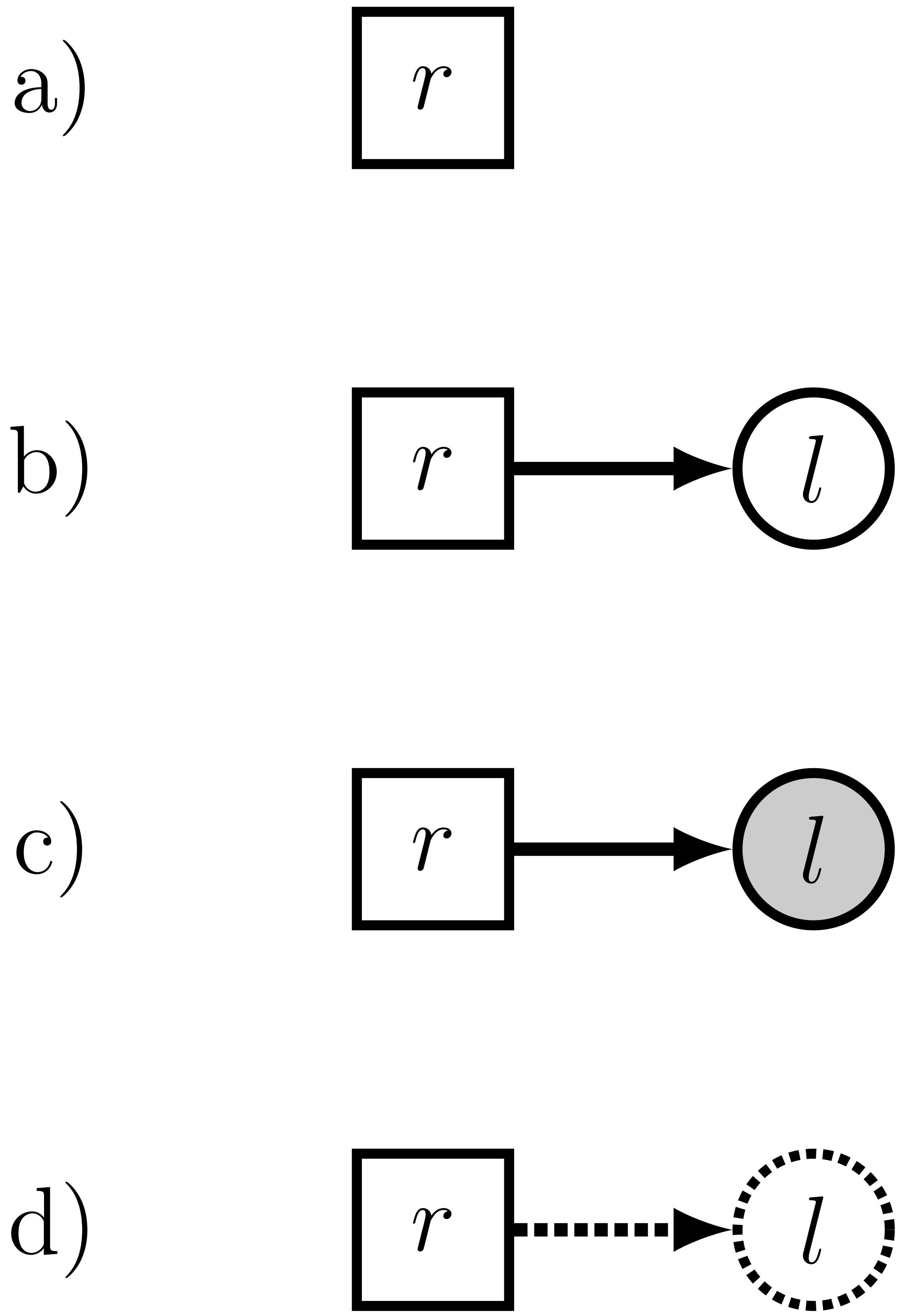
b)



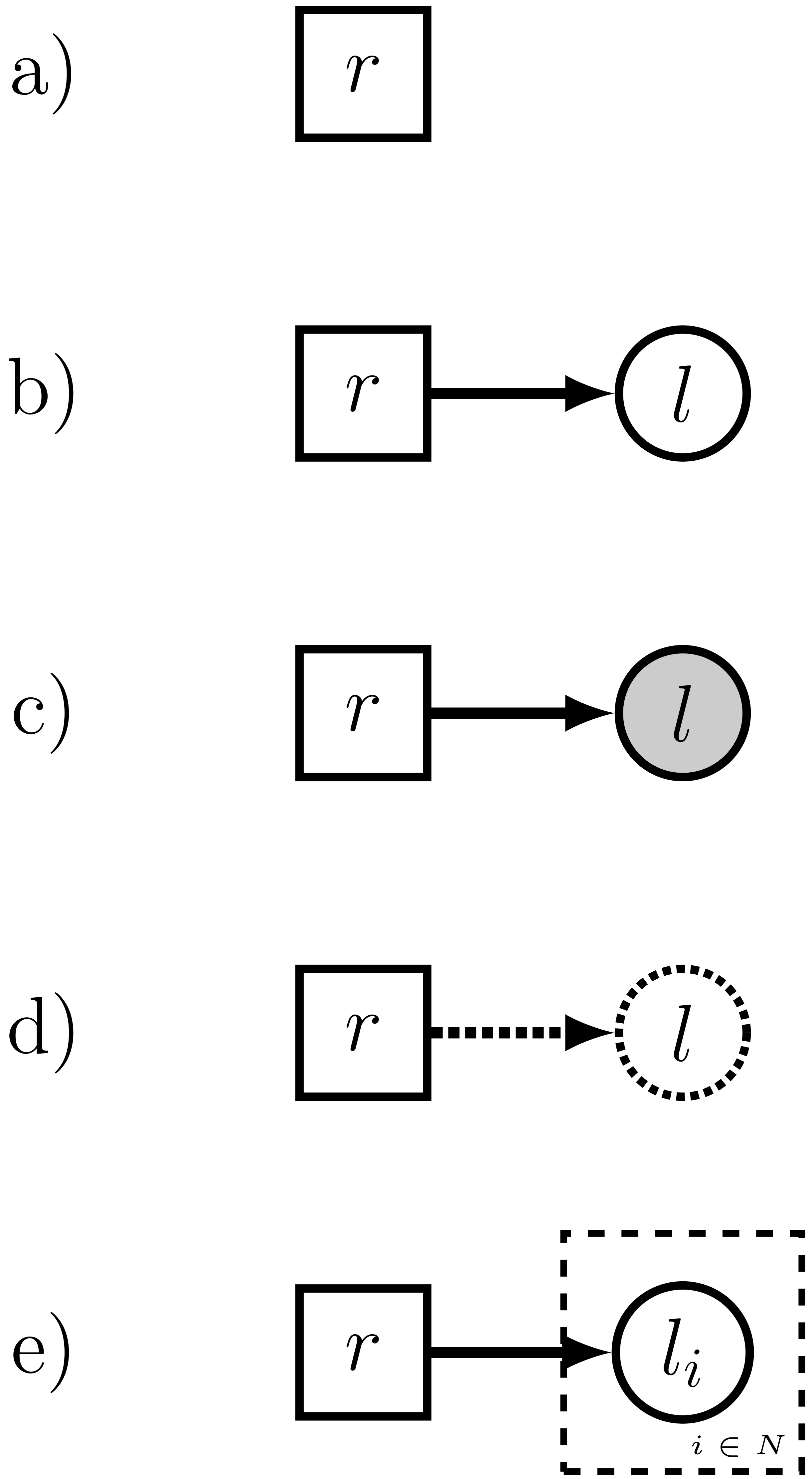
```
# constant node
r <- 10


# stochastic node
l ~ dnExp(r)
```



```
# constant node
r <- 10


# stochastic node
l ~ dnExp(r)


# stochastic node (observed)
l.clamp(0.1)
```



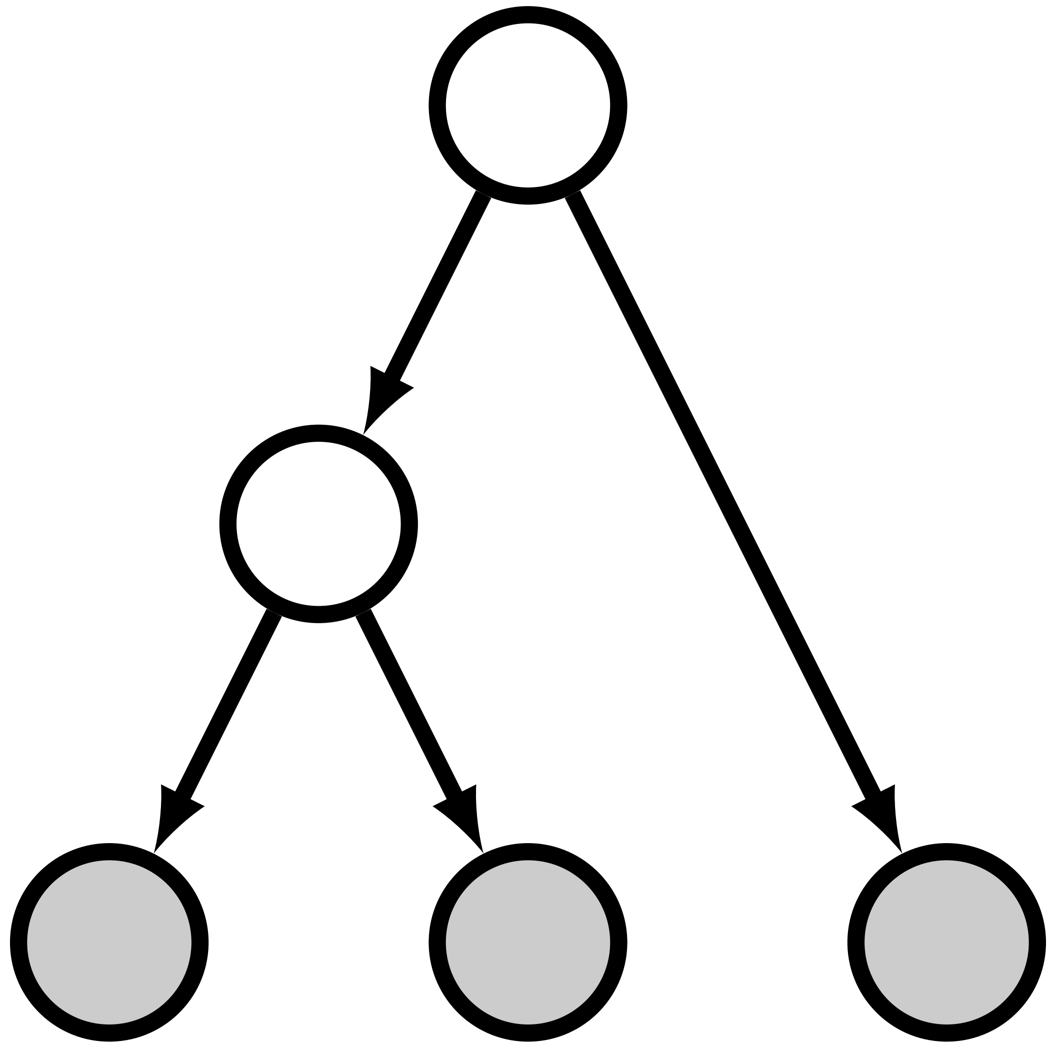
```
# constant node
r <- 10


# stochastic node
l ~ dnExp(r)


# stochastic node (observed)
l.clamp(0.1)


# deterministic node
l := exp(r)
```



```
# constant node
r <- 10


# stochastic node
l ~ dnExp(r)


# stochastic node (observed)
l.clamp(0.1)


# deterministic node
l := exp(r)


# stochastic nodes (iid)
for (i in 1:N) {
    l[i] ~ dnExp(r)
}
```

Running RevBayes demo

Exercise



Recap: Bayesian tree inference

Bayes' theorem

$$\Pr(\text{model} \mid \text{data}) = \frac{\Pr(\text{data} \mid \text{model})\,\Pr(\text{model})}{\Pr(\text{data})}$$

Bayes’ theorem

The probability of the data given the model assumptions and parameter values

Likelihood

$$\Pr(\text{data} \mid \text{model}) \Pr(\text{model})$$

$$\Pr(\text{model} \mid \text{data}) = \frac{\Pr(\text{data} \mid \text{model}) \Pr(\text{model})}{\Pr(\text{data})}$$

Bayes' theorem

This represents our prior knowledge of the model parameters

Priors

$$\Pr(\text{model} \mid \text{data}) = \frac{\Pr(\text{data} \mid \text{model}) \, \Pr(\text{model})}{\Pr(\text{data})}$$

Bayes' theorem

$$\Pr(\text{model} \mid \text{data}) = \frac{\Pr(\text{data} \mid \text{model}) \Pr(\text{model})}{\Pr(\text{data})}$$

Marginal probability

The probability of the data, given all possible parameter values. Can be thought of as a normalising constant

Bayes' theorem

Reflects our combined knowledge based on the likelihood and the priors

posterior

$$\Pr(\text{model} \mid \text{data}) = \frac{\Pr(\text{data} \mid \text{model}) \Pr(\text{model})}{\Pr(\text{data})}$$

Bayes' theorem

$$\Pr(\textbf{model} \mid \textbf{data}) \propto \Pr(\textbf{data} \mid \textbf{model}) \Pr(\textbf{model})$$

The posterior is proportional to the product of the prior and the likelihood

How do we find the ‘best’ tree?

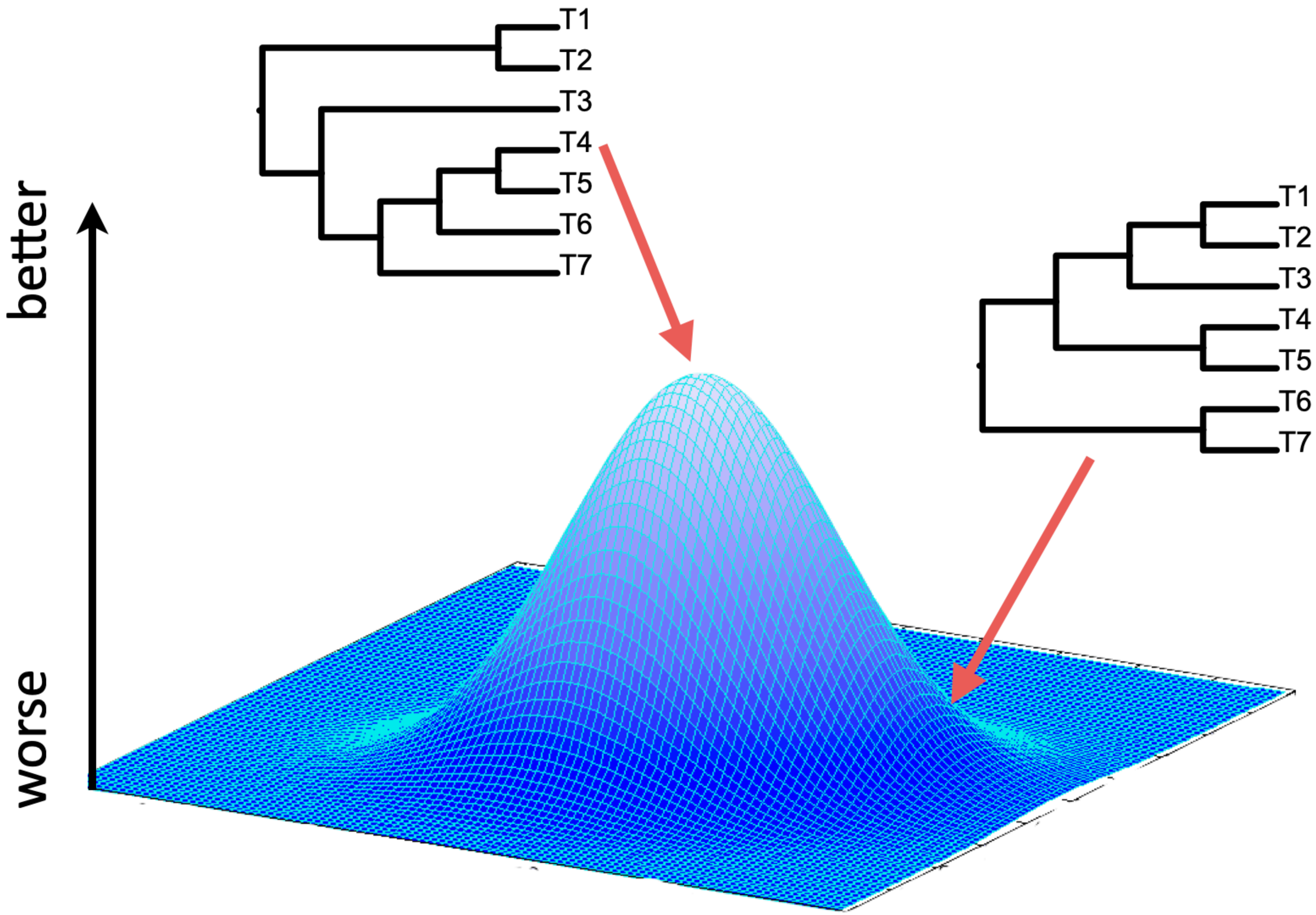


Image source Tracy Heath

It depends how you measure ‘best’

Method	Criterion (tree score)
Maximum parsimony	Minimum number of changes
Maximum likelihood	Likelihood score (probability), optimised over branch lengths and model parameters
Bayesian inference	Posterior probability, integrating over branch lengths and model parameters

Both maximum likelihood and Bayesian inference are model-based approaches

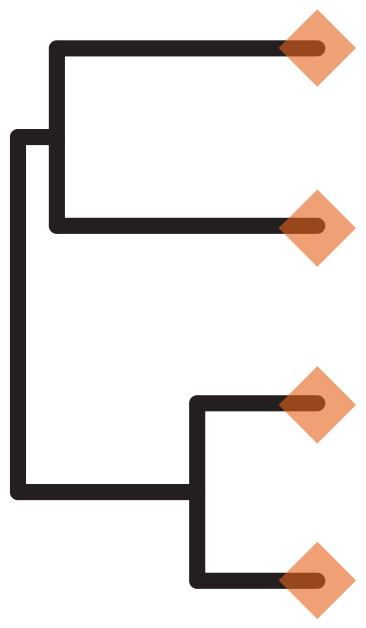
Note these are not the only approaches to tree-building but they are the most widely used

Components used to infer trees

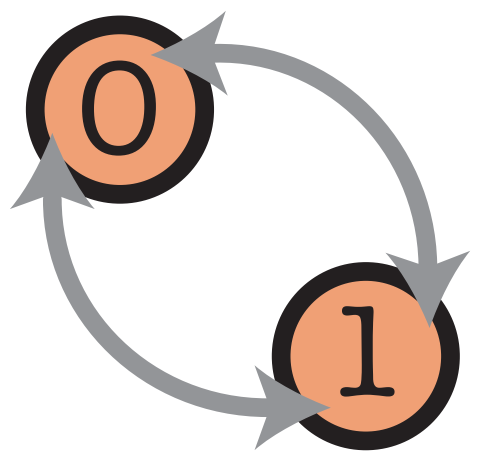
without considering time

0101...
1101...
0100...

data
sequences or
characters

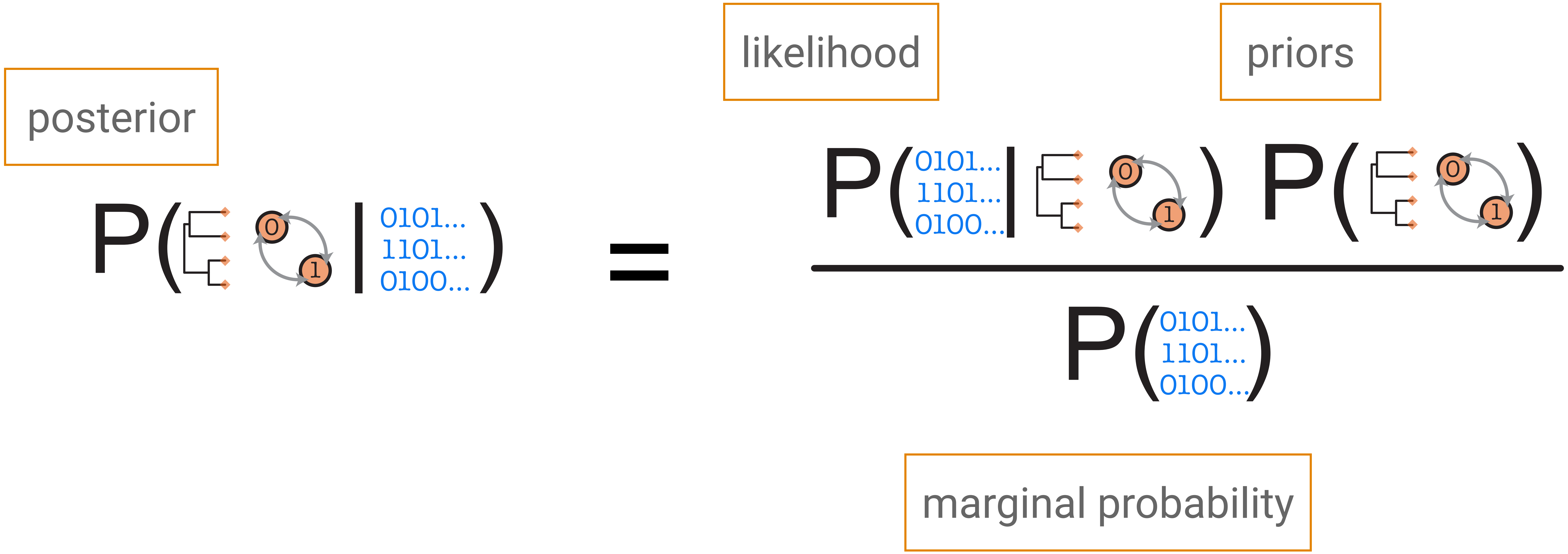


tree
topology and
branch lengths



substitution
model

Bayesian tree inference



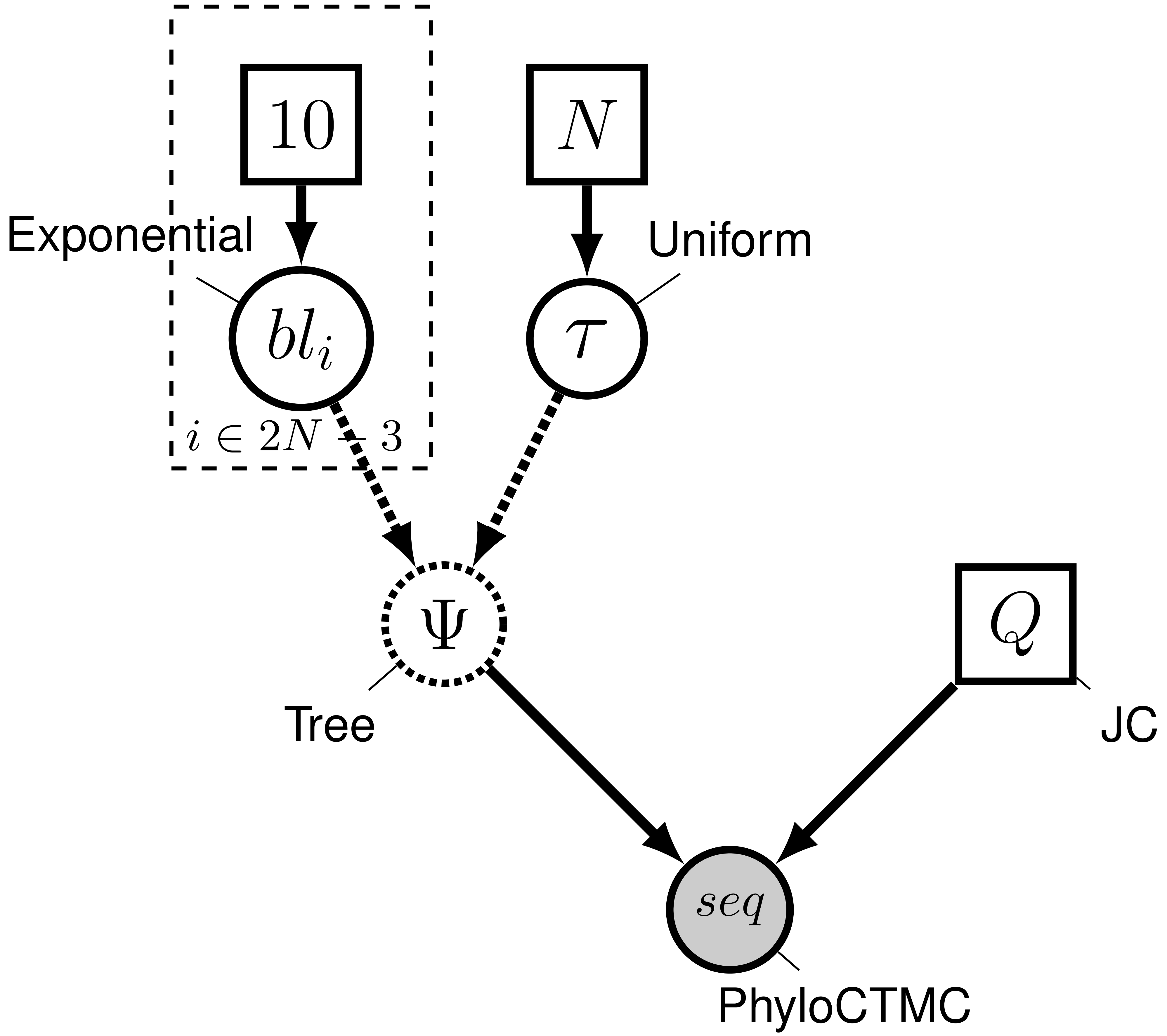
posterior

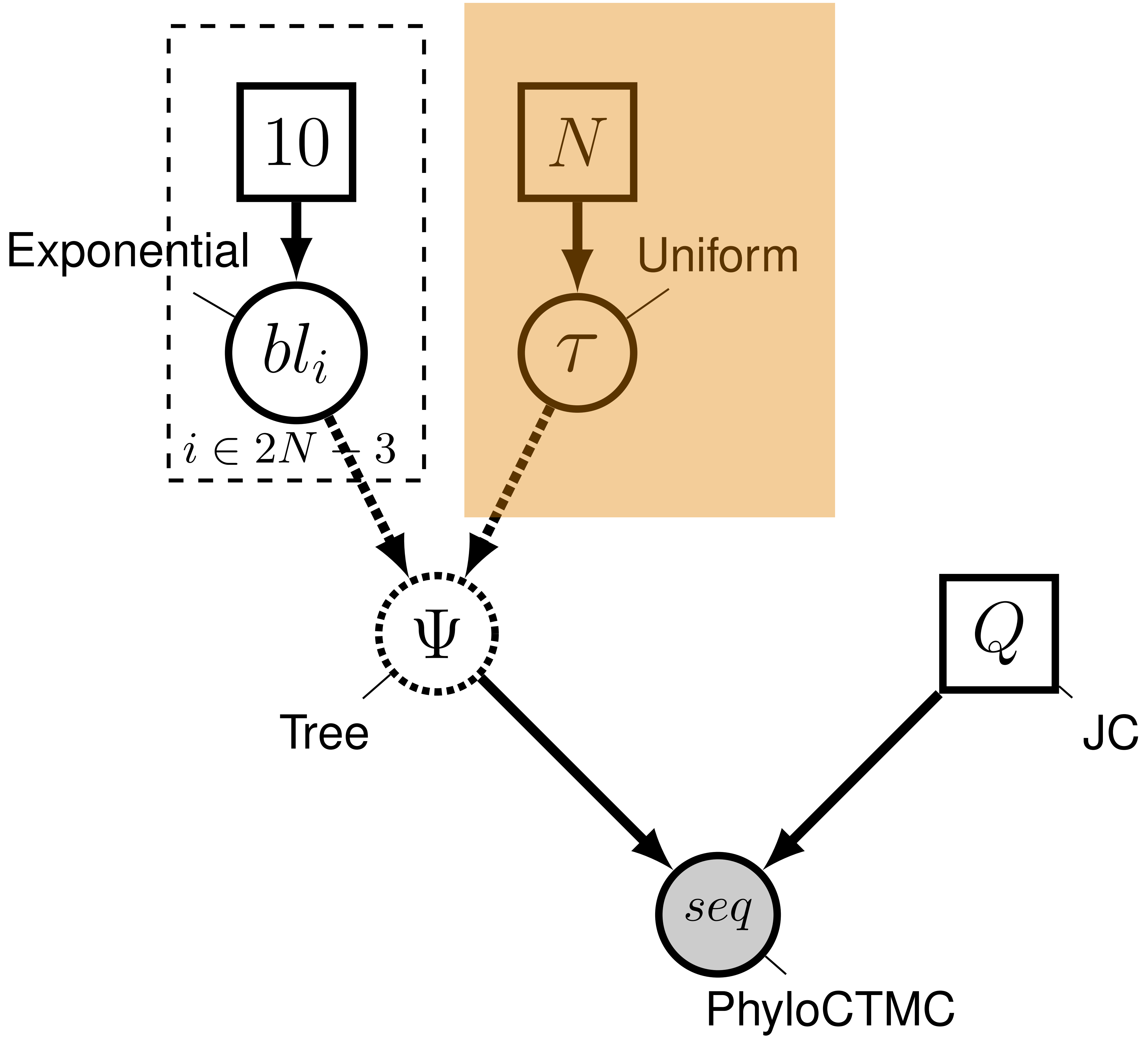
likelihood

priors

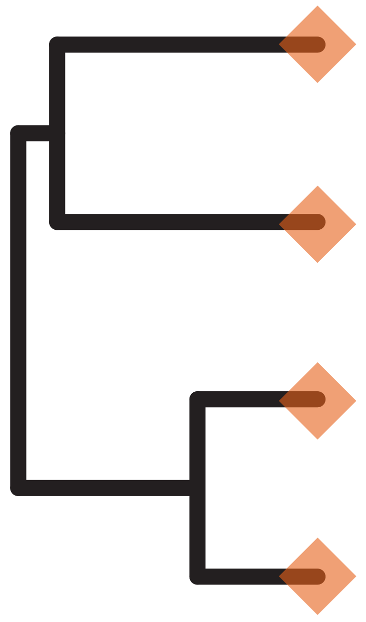
marginal probability

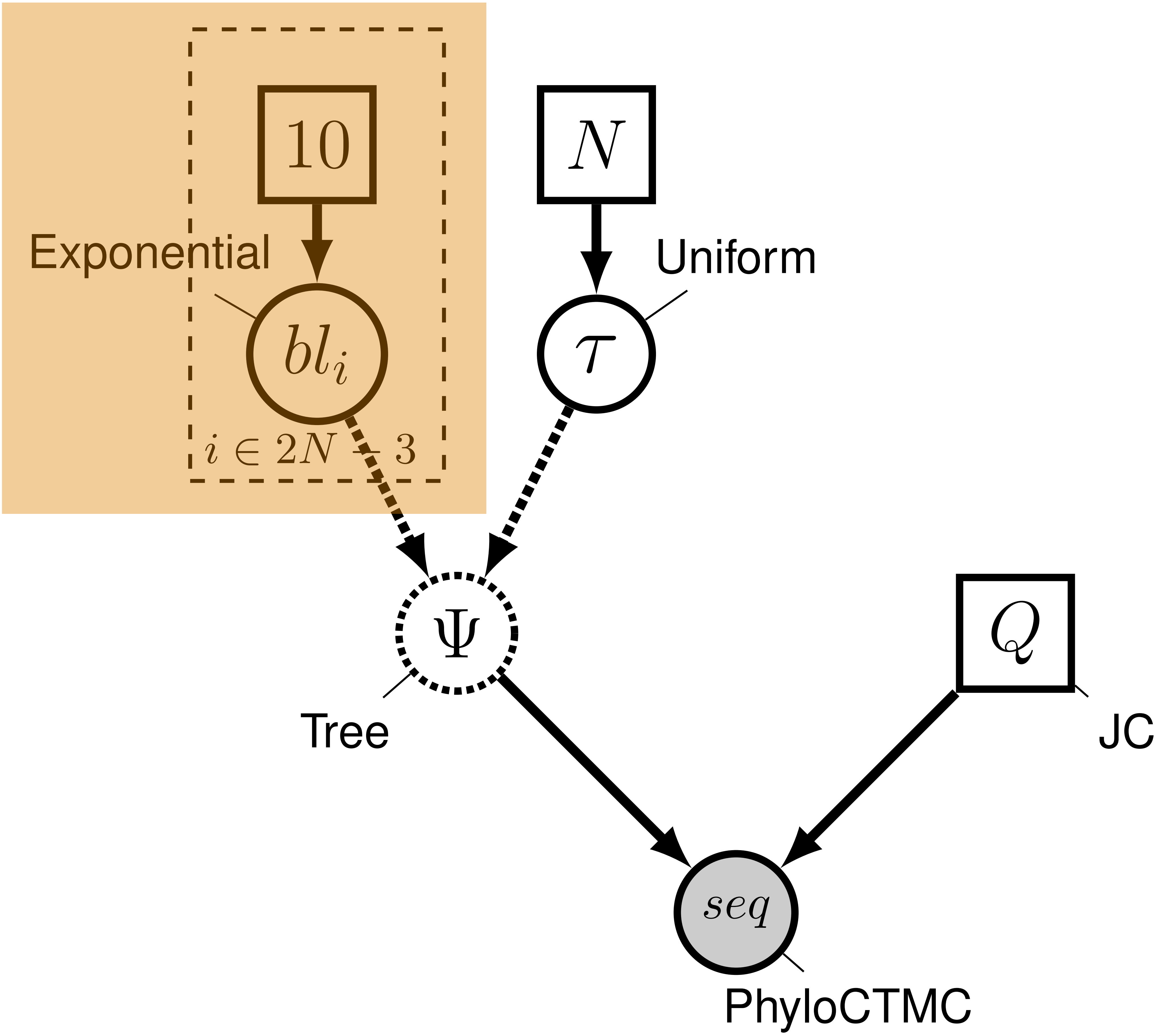
$


10
N
Exponential
Uniform
bl_i
τ
$i \in 2N - 3$
Ψ
Q
Tree
JC
seq
PhyloCTMC

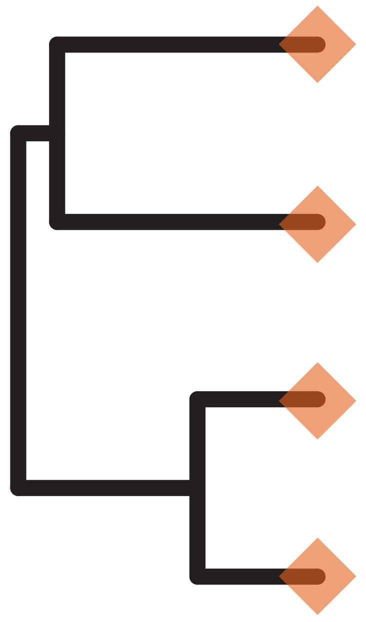


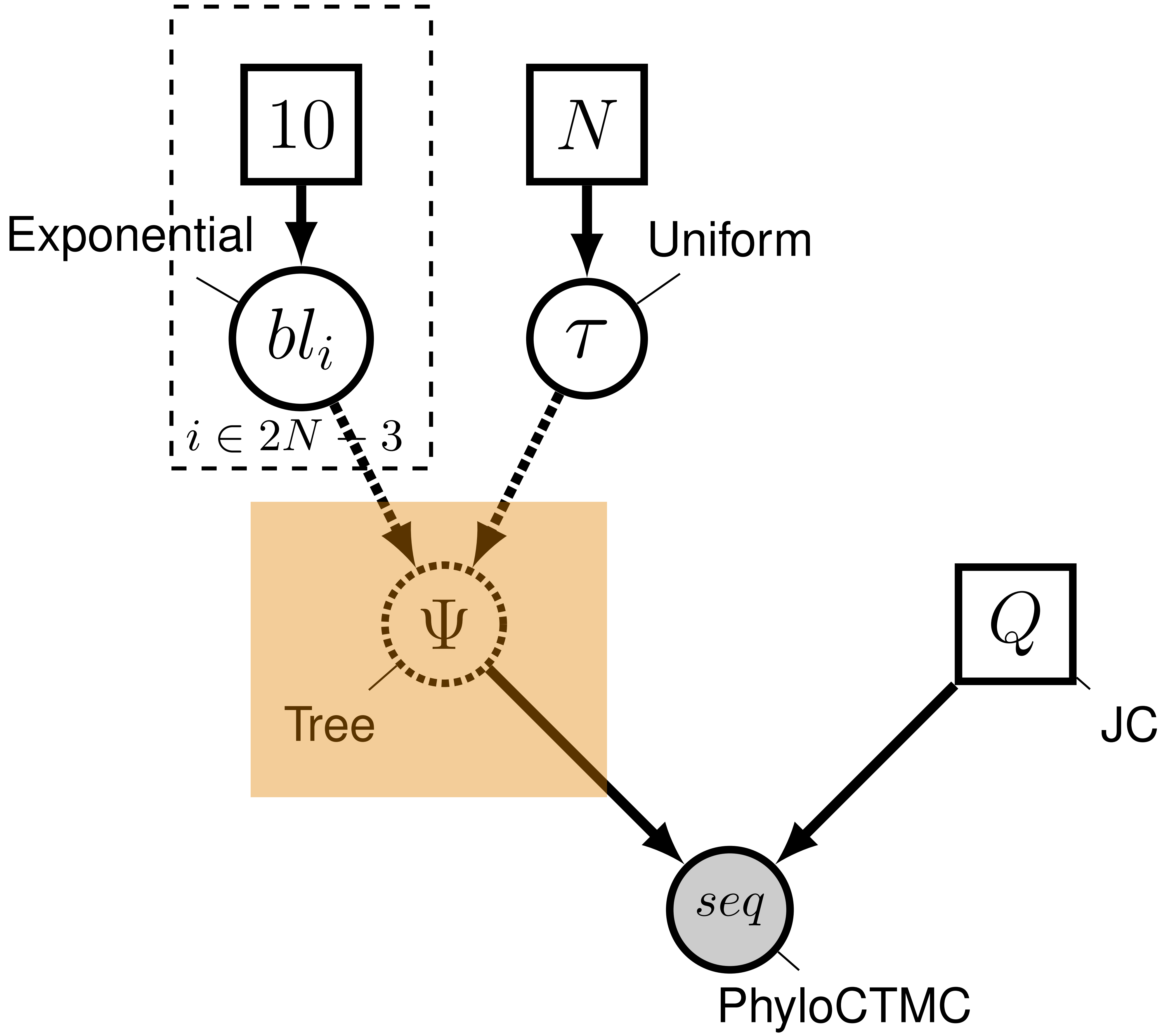
prior on the tree topology



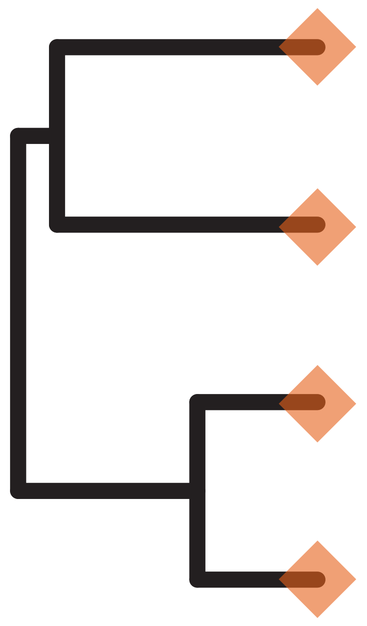


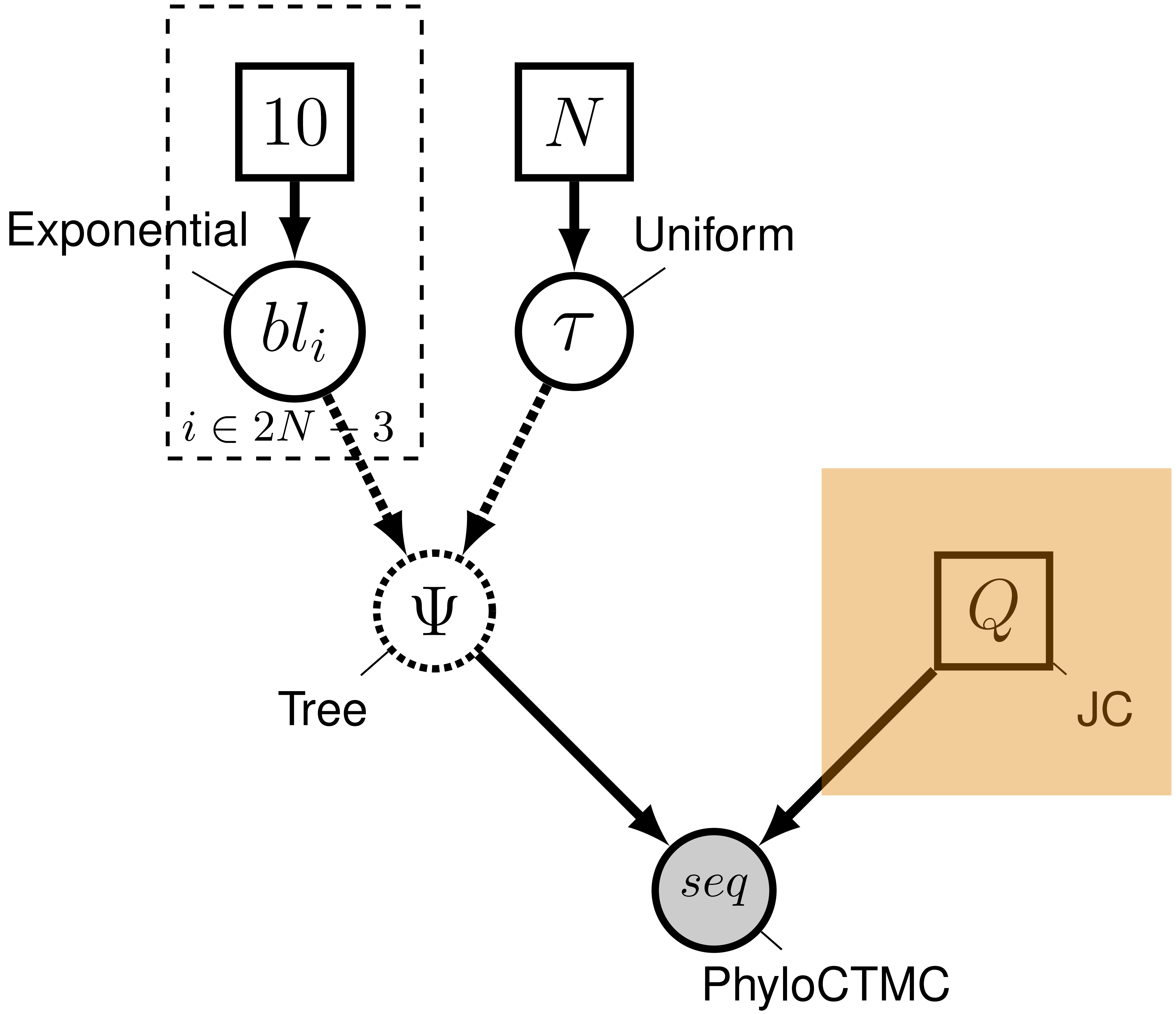
prior on the branch lengths



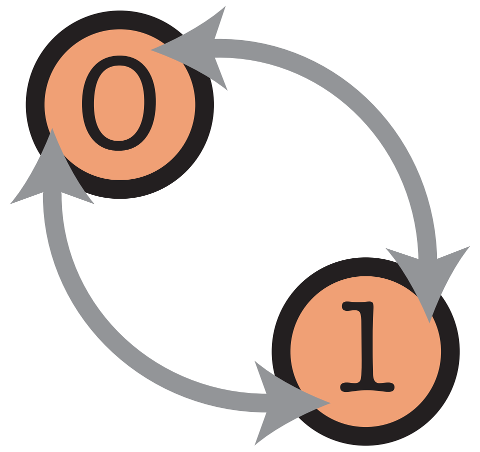


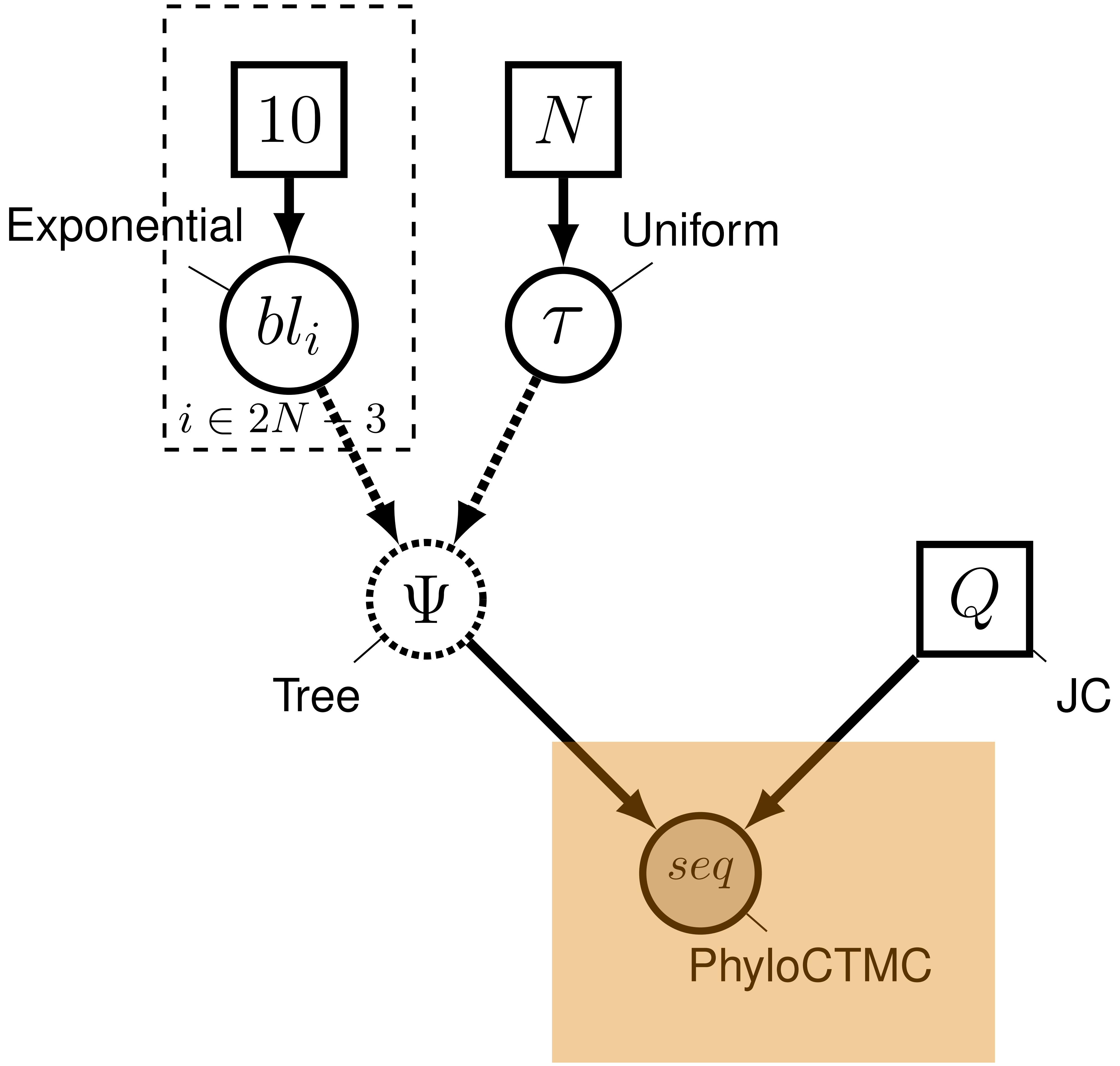
we can combine the topology and branch lengths





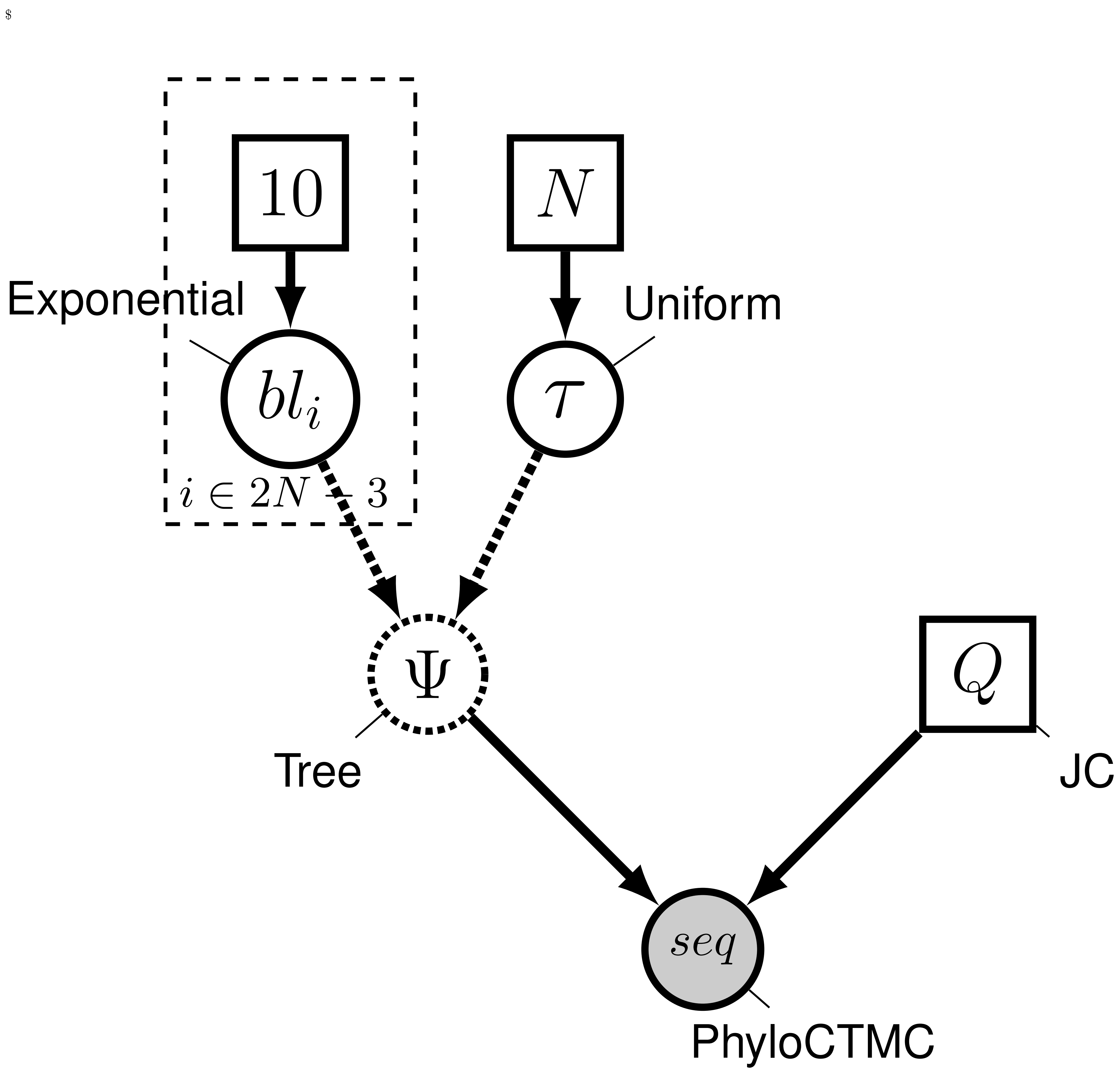
Substitution model





Observed data

0101...
1101...
0100...



```
for (i in 1:n_brances){
    bl[i] ~ dnExponential(10.0)
}
topology ~ dnUniformToplogy(taxa)
tree := treeAssembly(topology, bl)

Q <- fnJC(4)

seq ~ dnPhyloCTMC(tree = tree,
                        Q = Q,
                     type = "DNA")

seq.clamp(data)
```

Recap: MCMC

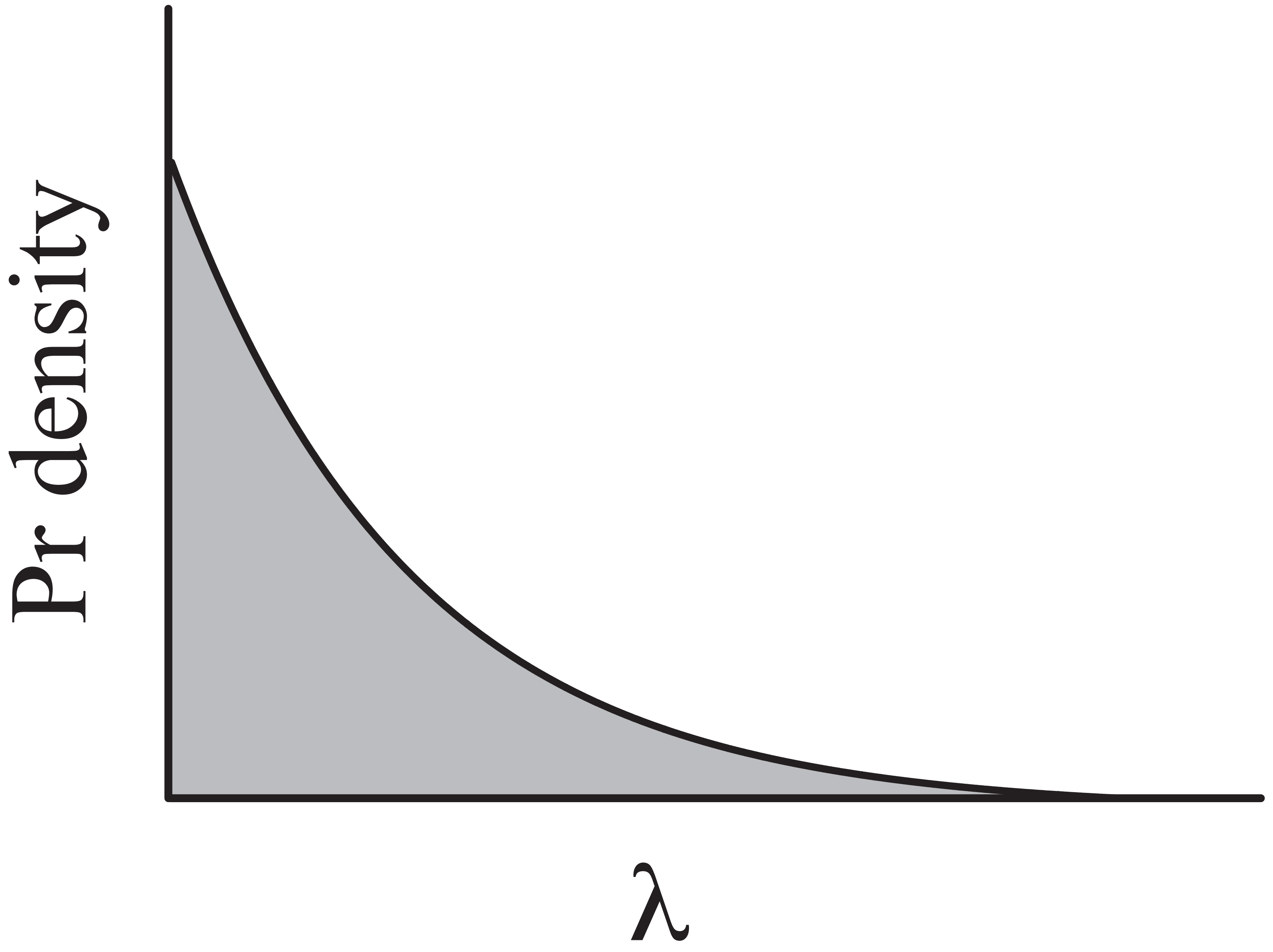
Probabilities vs. probability densities

In phylogenetics, probabilities are not normally discrete (i.e., represented by a single value)

We’re often dealing with a lot of uncertainty and typically work with probability densities

Probability densities introduce some complexity

Probabilities vs. probability densities



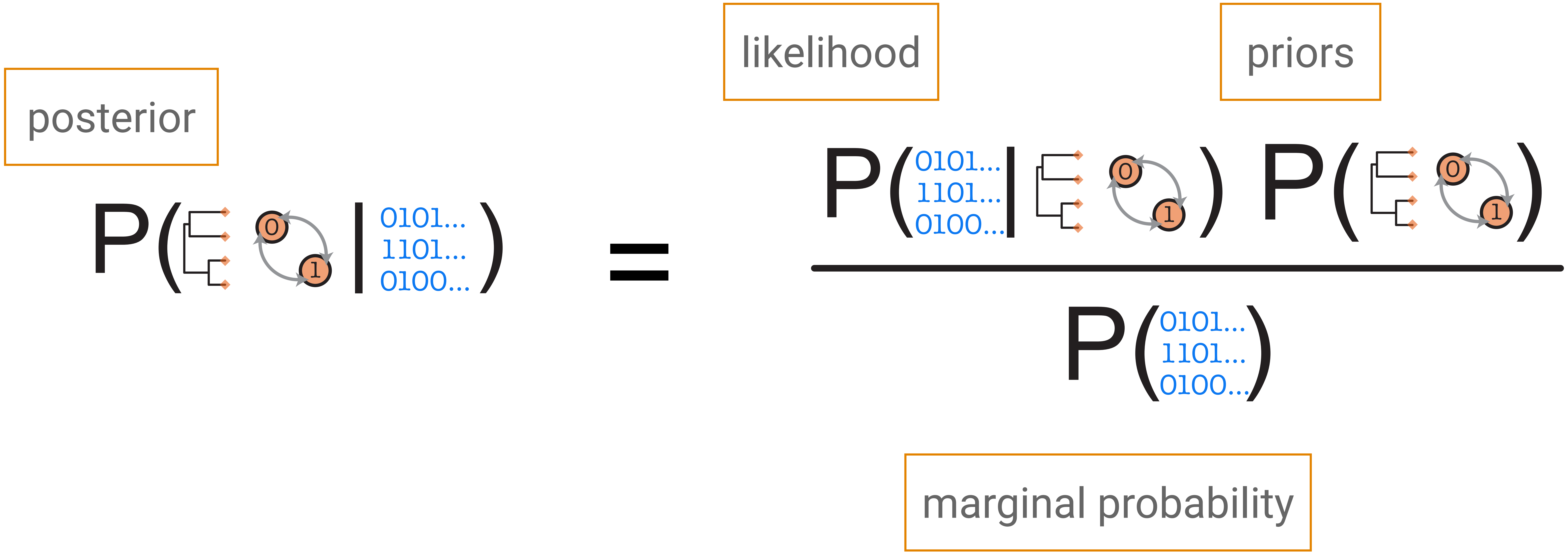
λ is drawn from an exponential distribution with mean δ

The x-axis represents the value of our parameter λ

The y-axis does have a value but it is not so easily interpretable

The distribution height reflects the relative probability of a given range of values

Bayesian tree inference



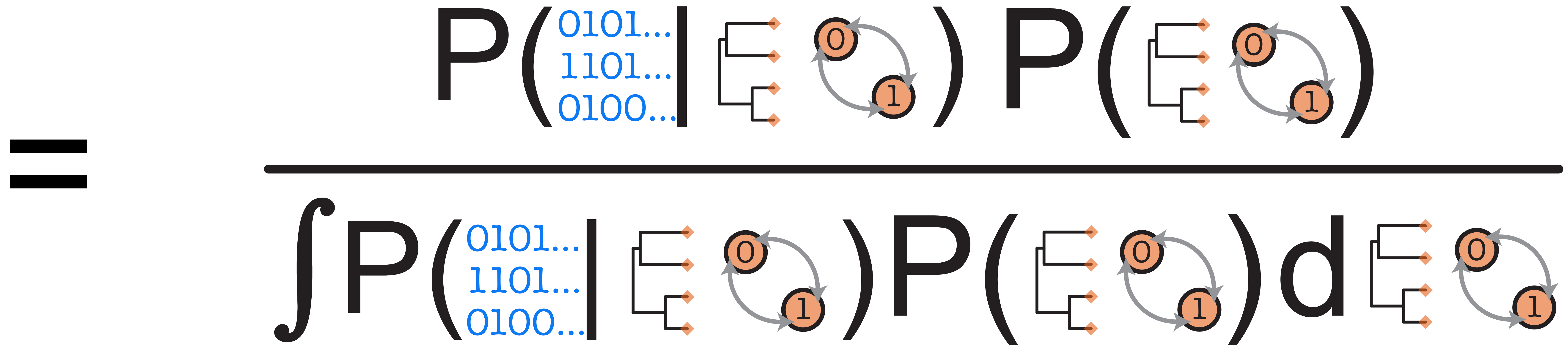
posterior

likelihood

priors

marginal probability

Bayesian tree inference



this part is incredibly difficult to calculate!

What is Markov chain Monte Carlo (MCMC)?

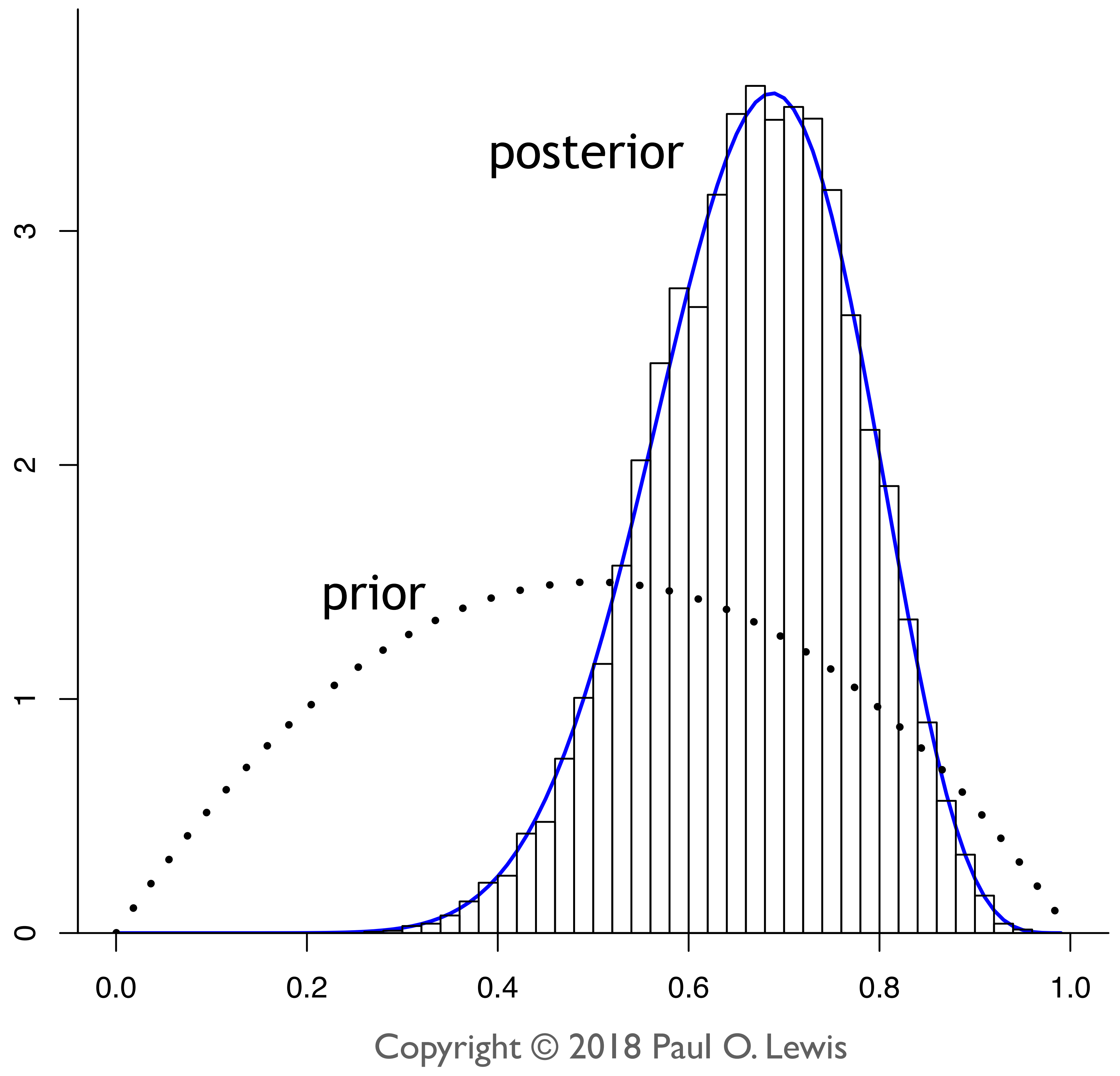
A group of algorithms for approximating the posterior distribution (also known as samplers)

Markov chain means the progress of the algorithm doesn’t depend on its past

Monte Carlo (named for the casino in Monaco) methods estimate a distribution via random sampling

We use this algorithm to visit different regions the parameter space. The number of times a given region is visited will be in proportion to its posterior probability

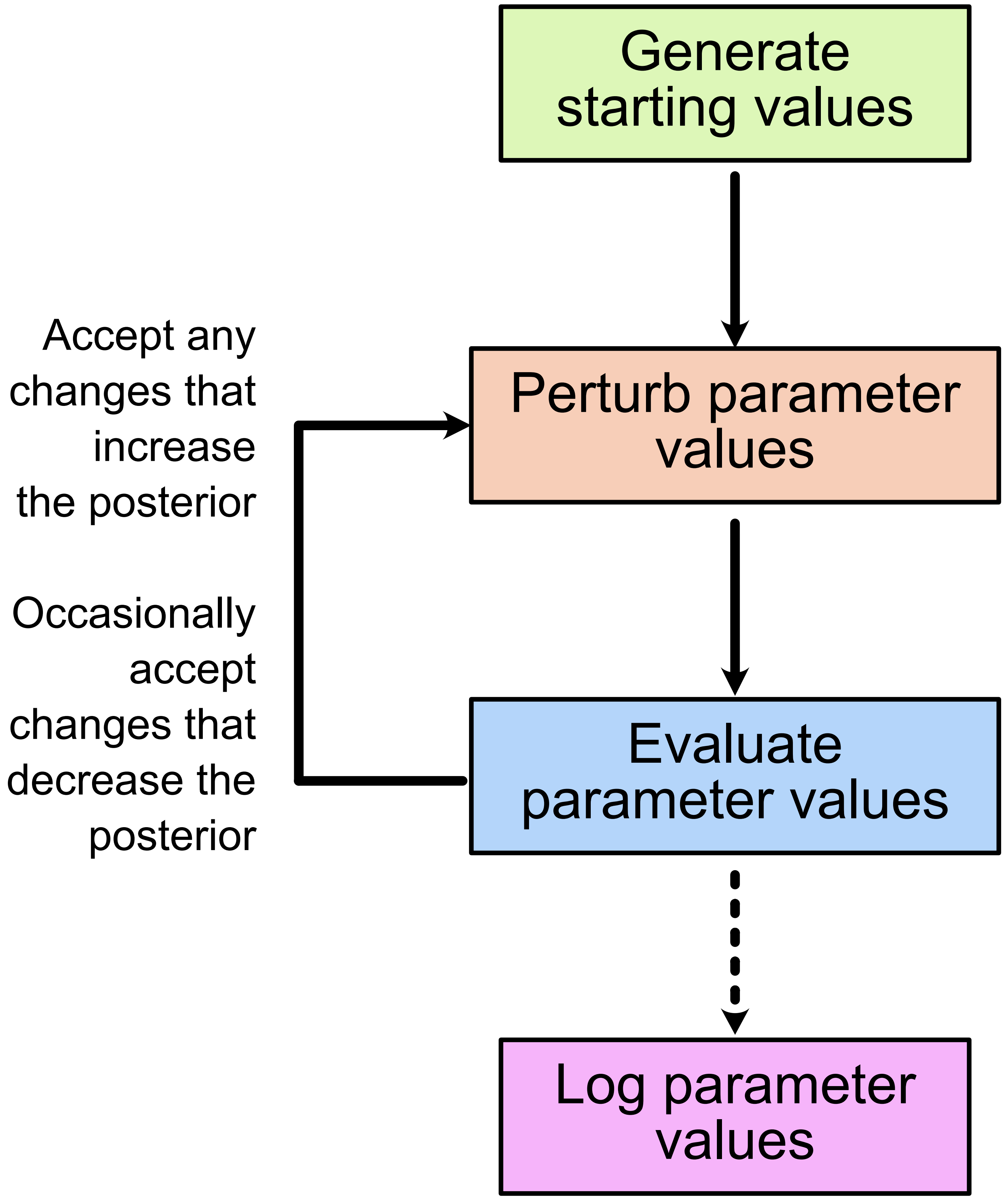
What is Markov chain Monte Carlo (MCMC)?



The aim is to produce a histogram that provides a good approximation of the posterior

The Metropolis-Hastings algorithm

Flowchart

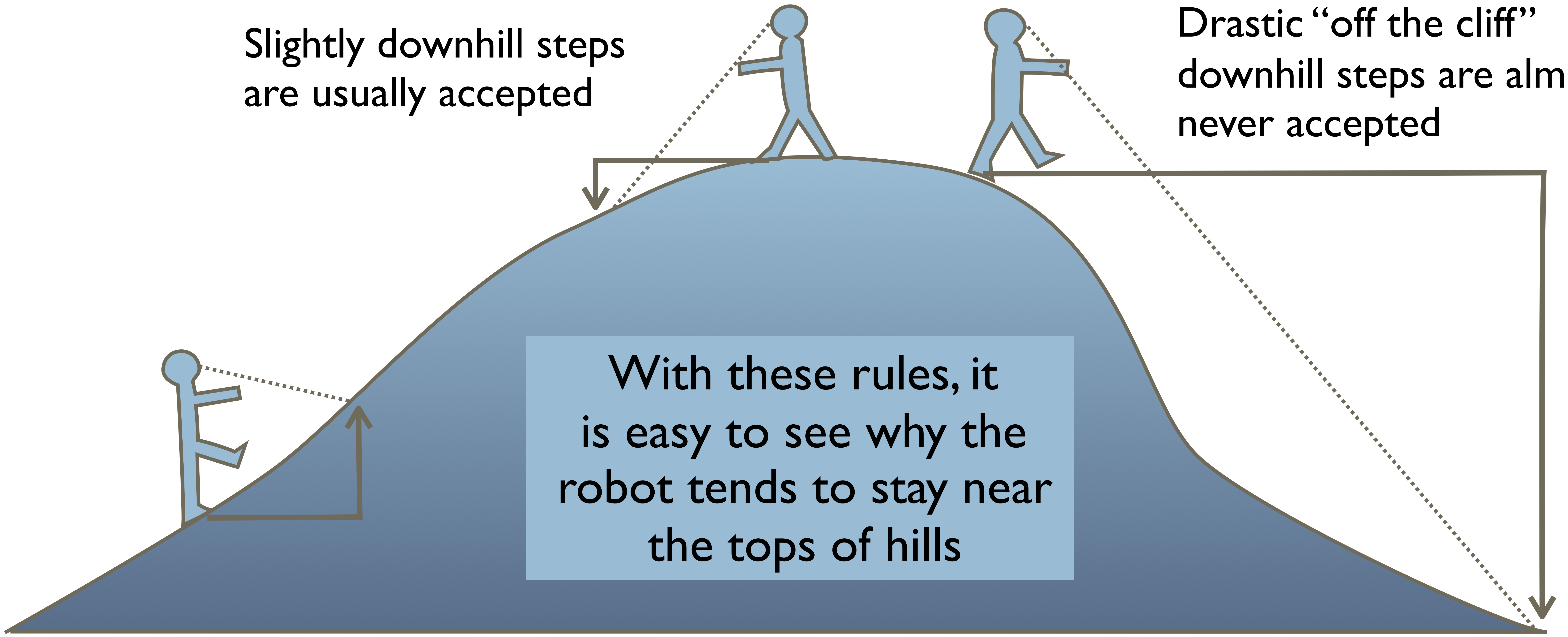


Pseudocode

```
initialize starting values;

for i in mcmc steps
do
    propose new parameter values;
    calculate the Hastings ratio R;

    if( R > 1 )
        accept the new values;
    else
        accept the new values with Pr = R;

    store the values with frequency j;
done
```

Barido-Sottani et al. (2024)

MCMC robot's rules

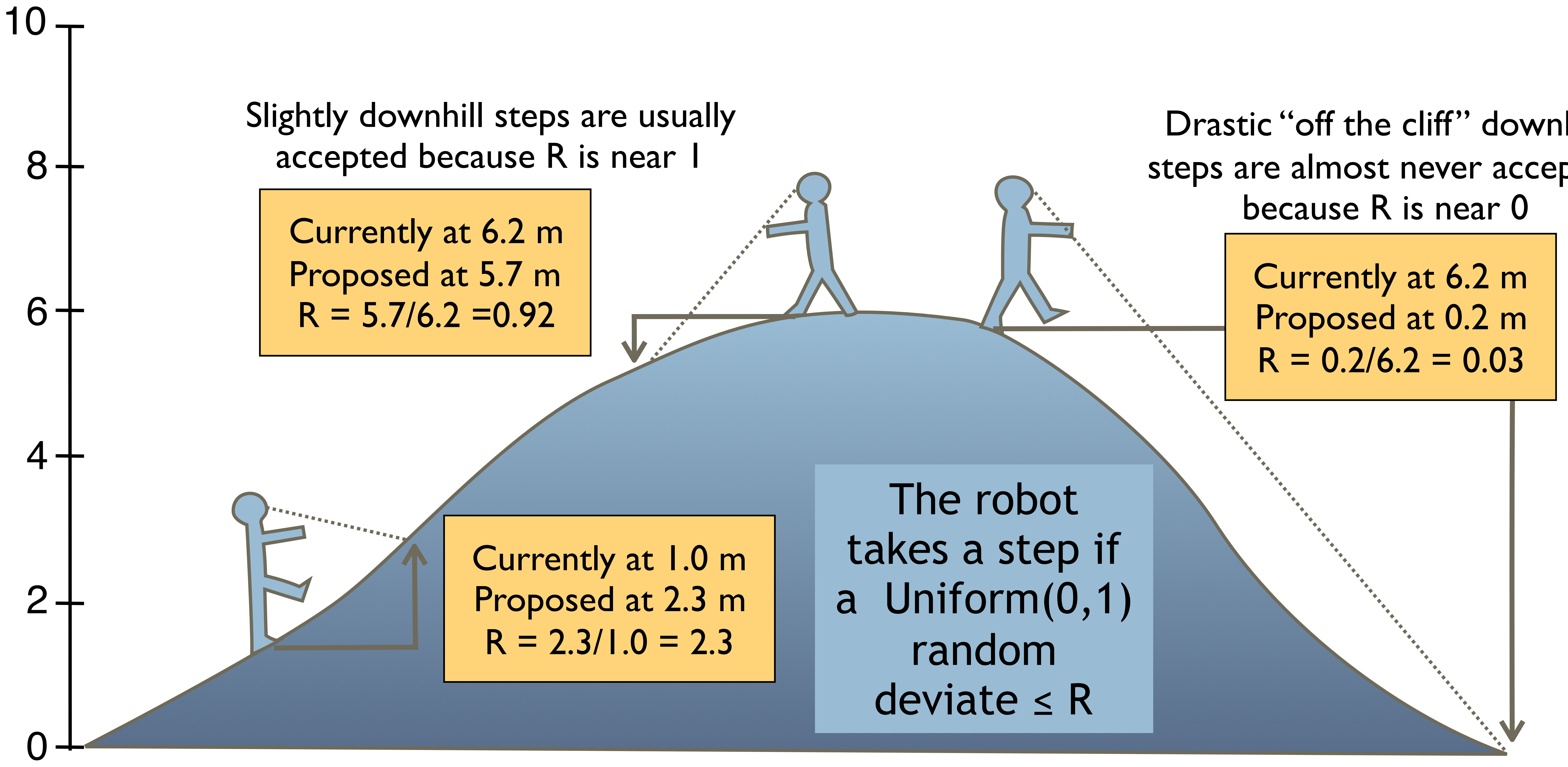


Slightly downhill steps are usually accepted

Drastic "off the cliff" downhill steps are almost never accepted

With these rules, it is easy to see why the robot tends to stay near the tops of hills

Uphill steps are always accepted



Slightly downhill steps are usually accepted because R is near 1

Currently at 6.2 m
Proposed at 5.7 m
R = 5.7/6.2 =0.92

Drastic "off the cliff" downhill steps are almost never accepted because R is near 0

Currently at 6.2 m
Proposed at 0.2 m
R = 0.2/6.2 = 0.03

Currently at 1.0 m
Proposed at 2.3 m
R = 2.3/1.0 = 2.3

The robot takes a step if a Uniform(0,1) random deviate ≤ R

Uphill steps are always accepted because R > 1

Metropolis et al. 1953. Equation of state calculations by fast computing machines. *J. Chem. Physics* 21(6):1087-1092.

When calculating the ratio (*R*) of posterior densities, the marginal probability of the data cancels.

$$\frac{p(\theta^* \,|\, D)}{p(\theta \,|\, D)} = \frac{\dfrac{p(D \,|\, \theta^*)\; p(\theta^*)}{\cancel{p(D)}}}{\dfrac{p(D \,|\, \theta)\; p(\theta)}{\cancel{p(D)}}} = \frac{p(D \,|\, \theta^*)\; p(\theta^*)}{p(D \,|\, \theta)\; p(\theta)}$$

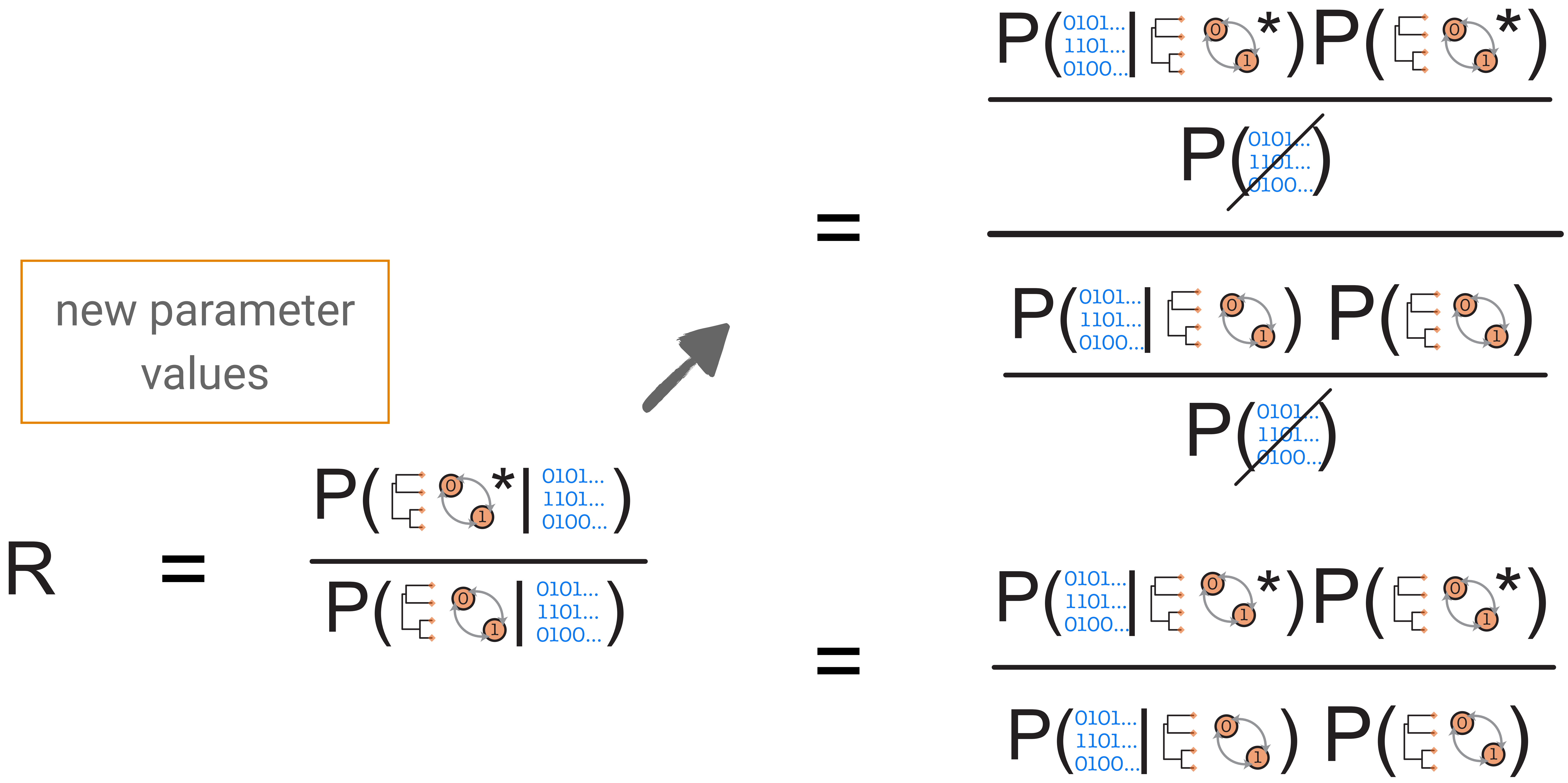
Posterior odds

Apply Bayes' rule to both top and bottom

Likelihood ratio

Prior odds

Hastings ratio



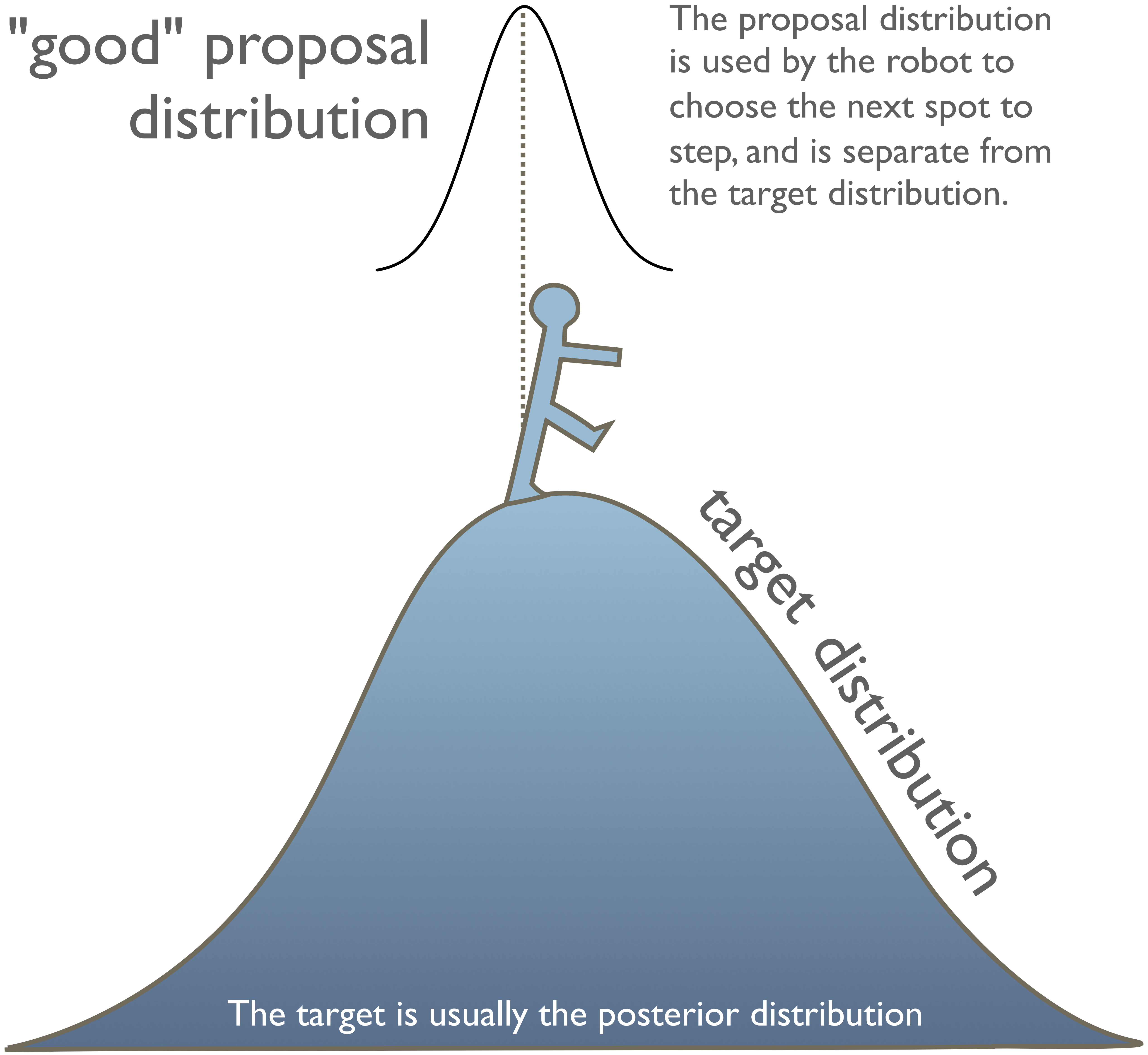
The marginal probability of the data cancels out

All we're left to calculate is the likelihood ratio and the prior odds ratio

Moves

Known as moves, proposals or operators — *how* we move across the parameter space

i.e., how do we select the next set of parameters to evaluate

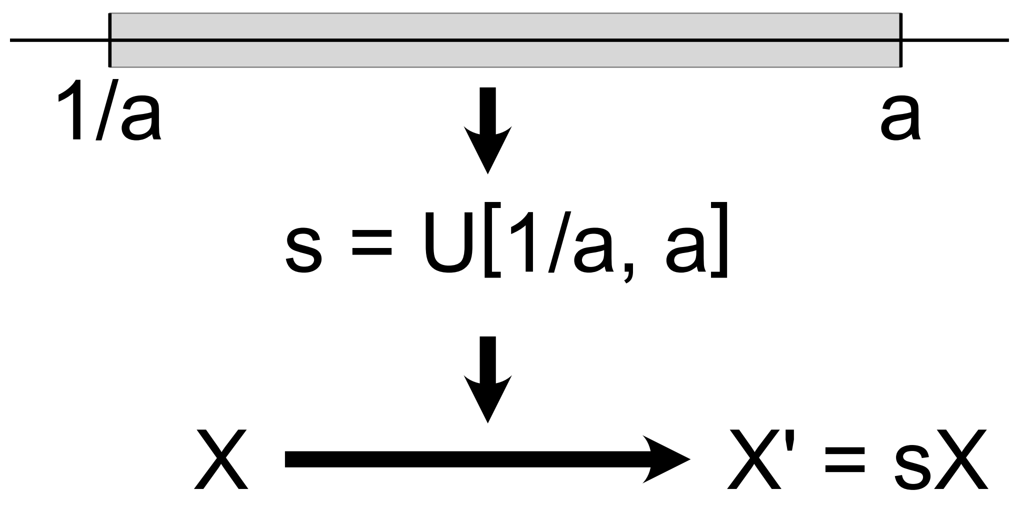


The proposal distribution is used by the robot to choose the next spot to step, and is separate from the target distribution.

The target is usually the posterior distribution

Moves

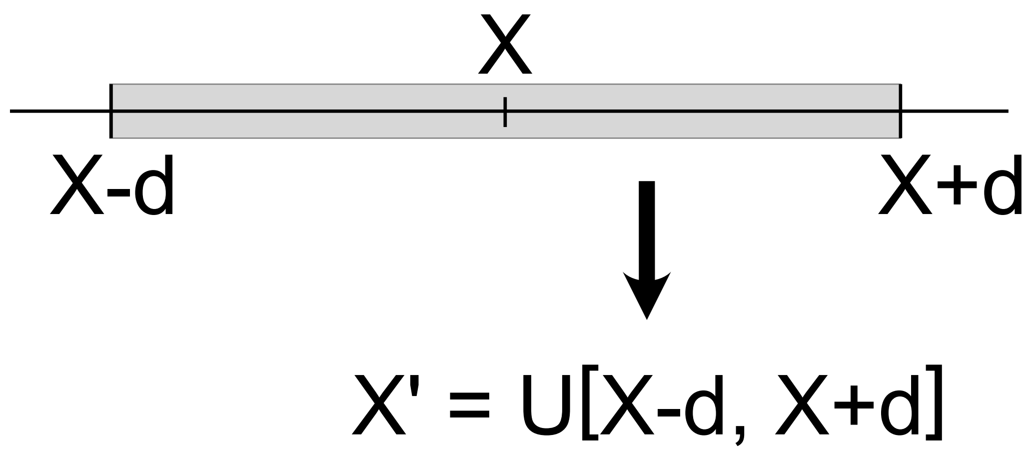
Scaling move



s = U[1/a, a]

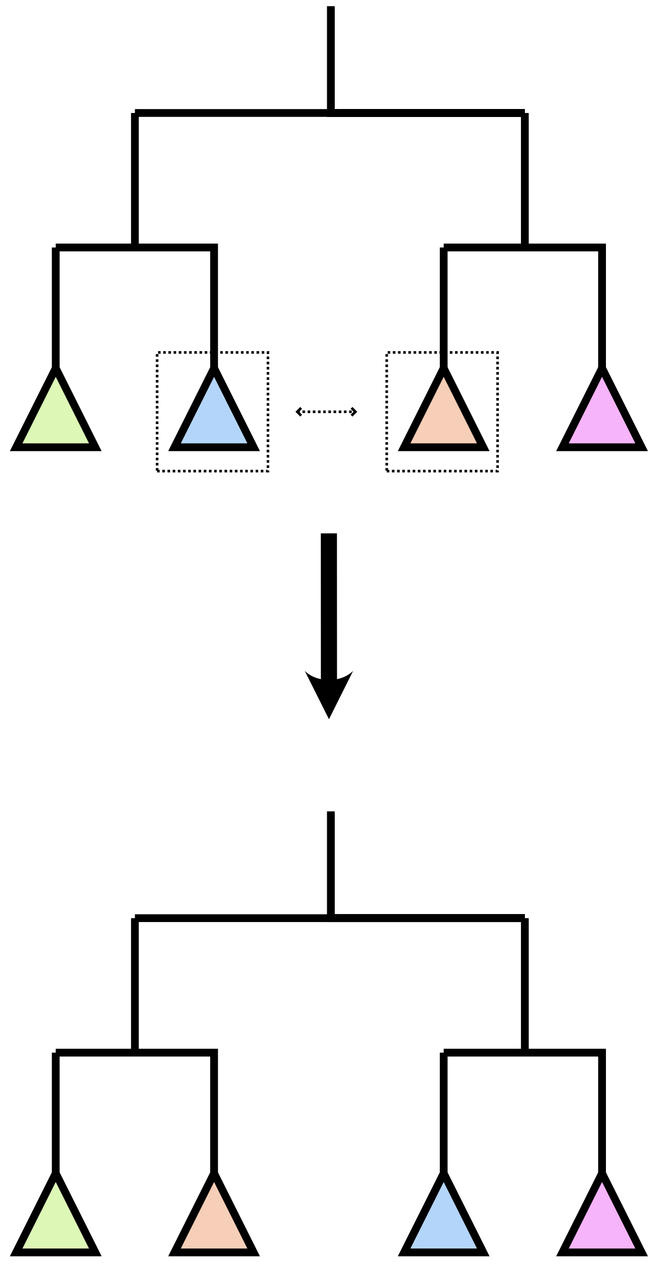
X → X' = sX

Sliding move

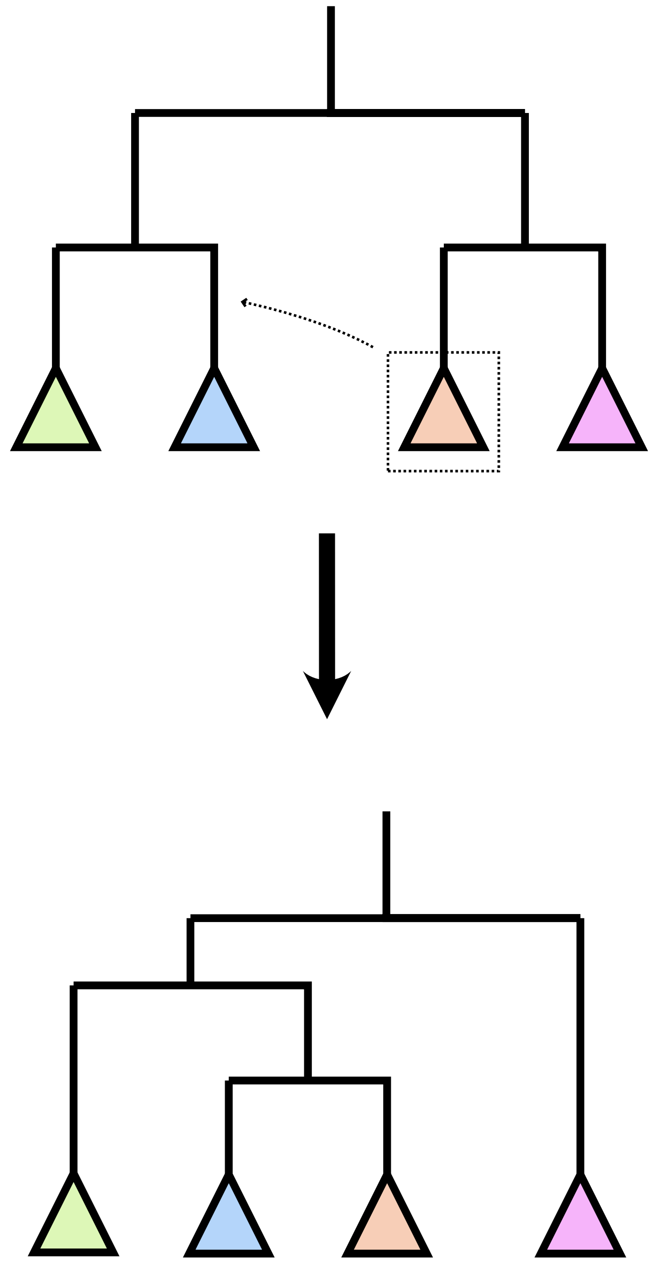


X' = U[X-d, X+d]

SubTree eXchange (STX) move



Subtree Pruning and Regrafting (SPR) move



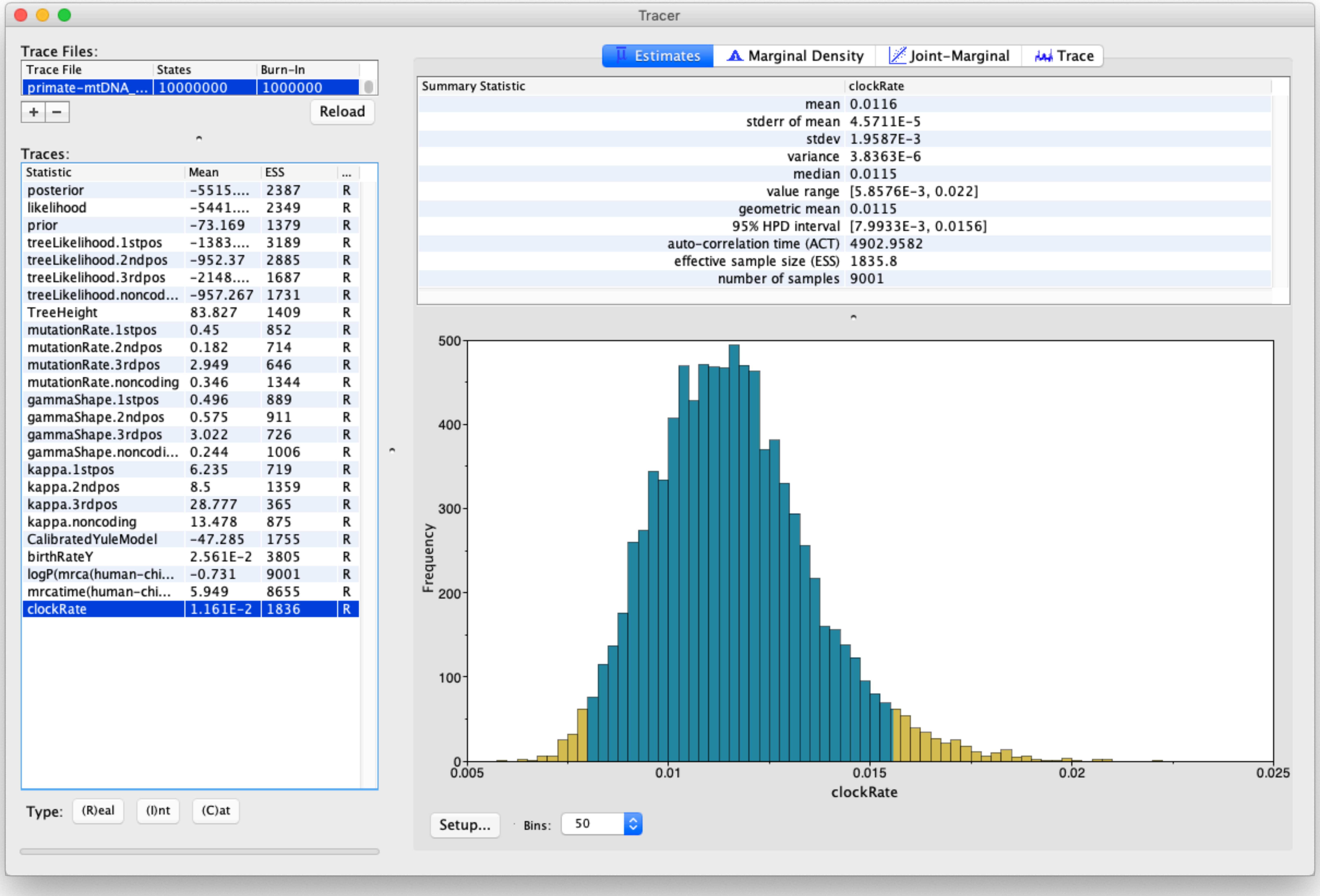
Barido-Sottani et al. (2024)

MCMC Robot demo (Paul Lewis)

MCMC Demo (Chi Feng)

Summarising the posterior

Tracer is an amazing program for exploring MCMC output



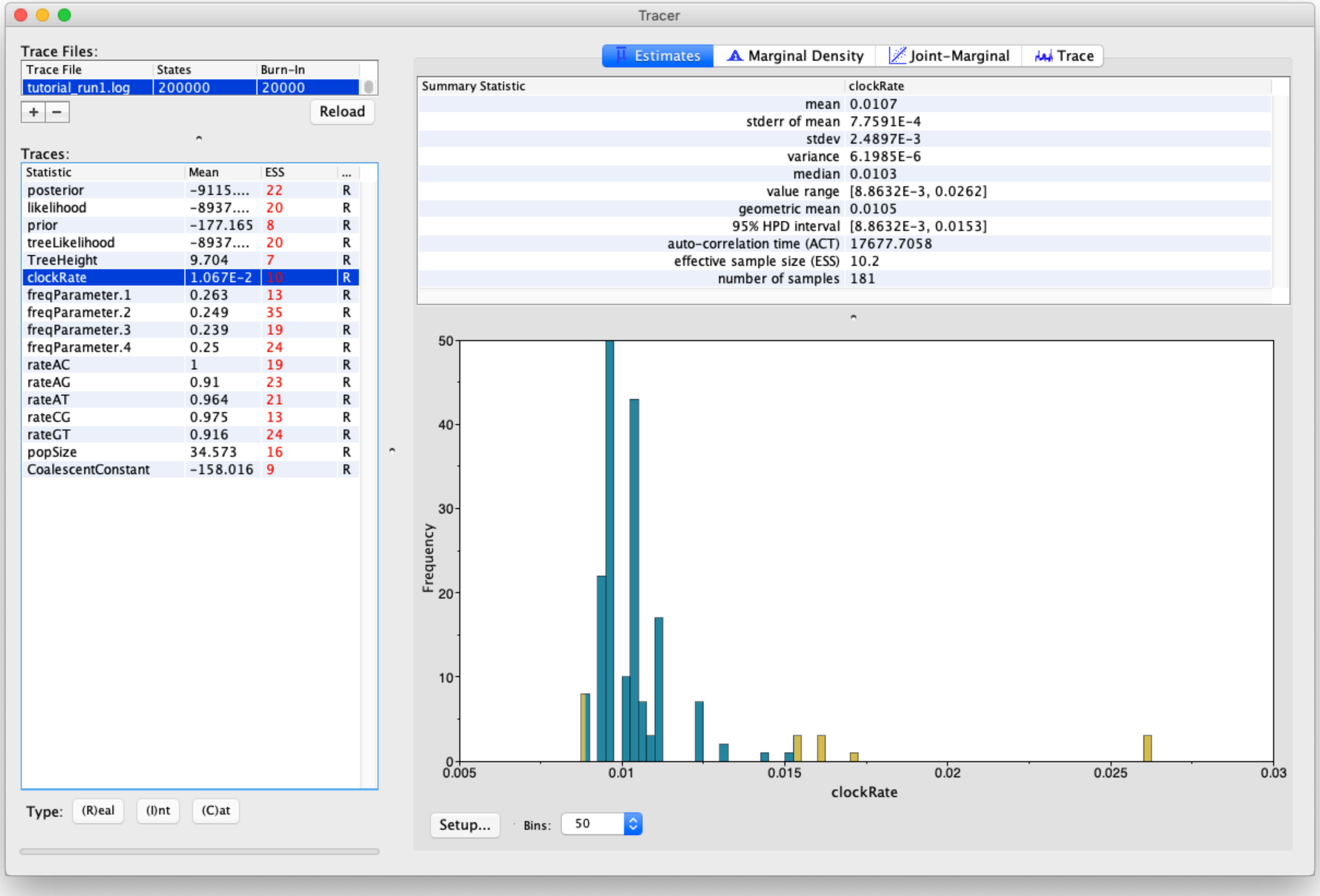
Trace File	States	Burn-In
primate-mtDNA_...	10000000	1000000

Statistic	Mean	ESS	...
posterior	-5515....	2387	R
likelihood	-5441....	2349	R
prior	-73.169	1379	R
treeLikelihood.1stpos	-1383....	3189	R
treeLikelihood.2ndpos	-952.37	2885	R
treeLikelihood.3rdpos	-2148....	1687	R
treeLikelihood.noncod...	-957.267	1731	R
TreeHeight	83.827	1409	R
mutationRate.1stpos	0.45	852	R
mutationRate.2ndpos	0.182	714	R
mutationRate.3rdpos	2.949	646	R
mutationRate.noncoding	0.346	1344	R
gammaShape.1stpos	0.496	889	R
gammaShape.2ndpos	0.575	911	R
gammaShape.3rdpos	3.022	726	R
gammaShape.noncodi...	0.244	1006	R
kappa.1stpos	6.235	719	R
kappa.2ndpos	8.5	1359	R
kappa.3rdpos	28.777	365	R
kappa.noncoding	13.478	875	R
CalibratedYuleModel	-47.285	1755	R
birthRateY	2.561E-2	3805	R
logP(mrca(human-chi...	-0.731	9001	R
mrcatime(human-chi...	5.949	8655	R
clockRate	1.161E-2	1836	R

Summary Statistic	clockRate
mean	0.0116
stderr of mean	4.5711E-5
stdev	1.9587E-3
variance	3.8363E-6
median	0.0115
value range	[5.8576E-3, 0.022]
geometric mean	0.0115
95% HPD interval	[7.9933E-3, 0.0156]
auto-correlation time (ACT)	4902.9582
effective sample size (ESS)	1835.8
number of samples	9001

Summarising the posterior

Tracer is an amazing program for exploring MCMC output



Trace Files:

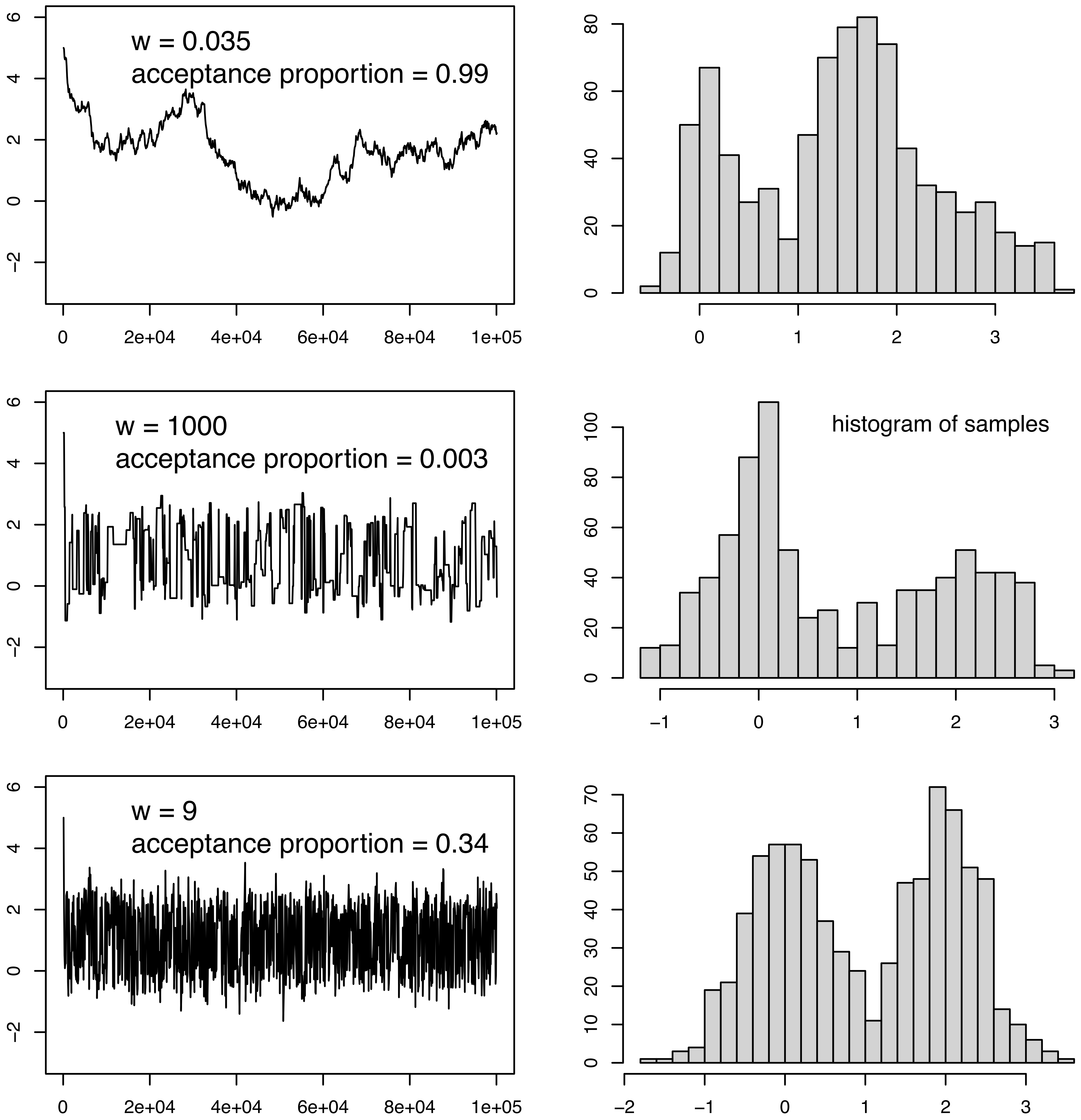
Trace File	States	Burn-In
tutorial_run1.log	200000	20000

Traces:

Statistic	Mean	ESS	...
posterior	-9115....	22	R
likelihood	-8937....	20	R
prior	-177.165	8	R
treeLikelihood	-8937....	20	R
TreeHeight	9.704	7	R
clockRate	1.067E-2	10	R
freqParameter.1	0.263	13	R
freqParameter.2	0.249	35	R
freqParameter.3	0.239	19	R
freqParameter.4	0.25	24	R
rateAC	1	19	R
rateAG	0.91	23	R
rateAT	0.964	21	R
rateCG	0.975	13	R
rateGT	0.916	24	R
popSize	34.573	16	R
CoalescentConstant	-158.016	9	R

Summary Statistic	clockRate
mean	0.0107
stderr of mean	7.7591E-4
stdev	2.4897E-3
variance	6.1985E-6
median	0.0103
value range	[8.8632E-3, 0.0262]
geometric mean	0.0105
95% HPD interval	[8.8632E-3, 0.0153]
auto-correlation time (ACT)	17677.7058
effective sample size (ESS)	10.2
number of samples	181

Better!

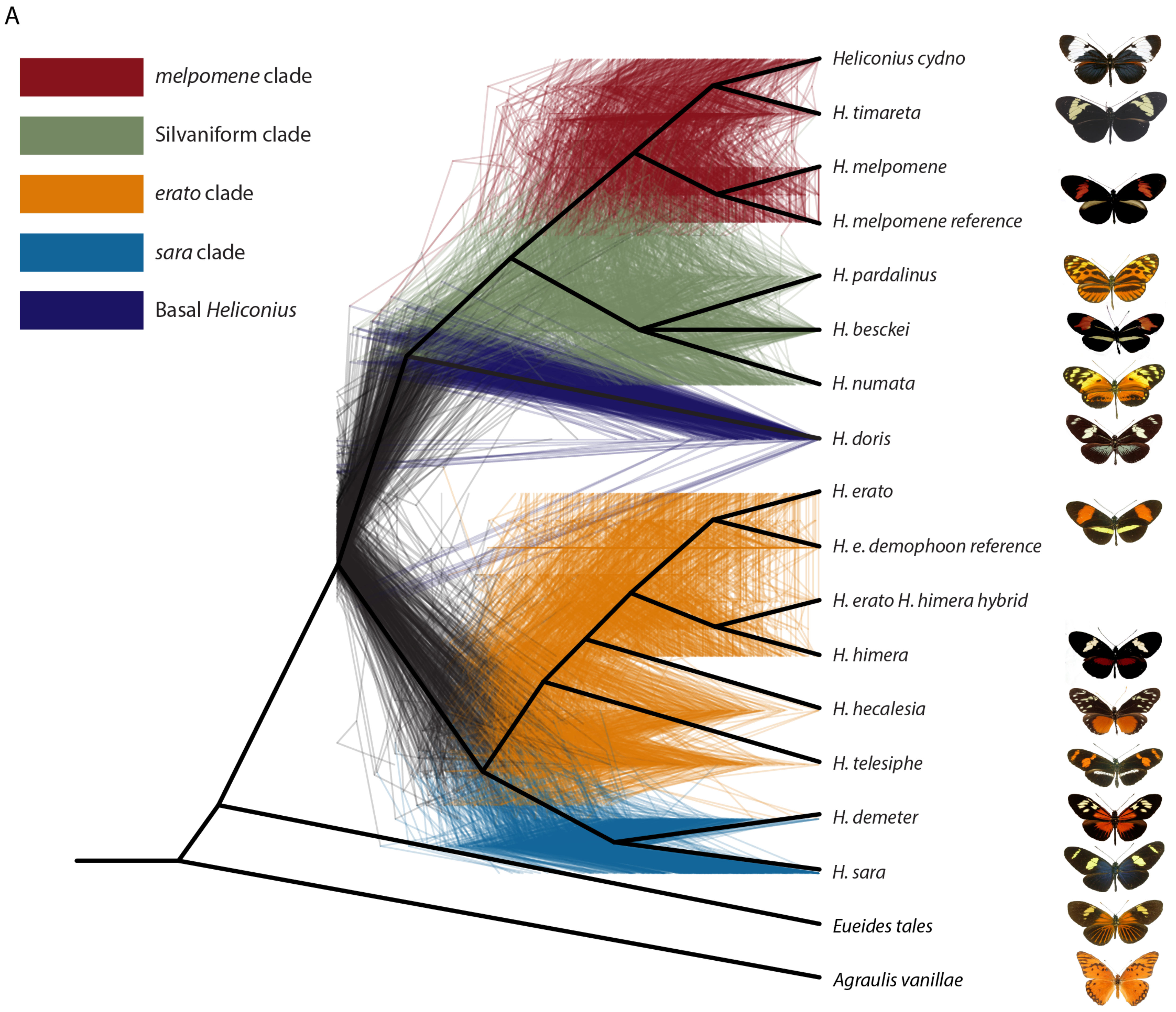


Barido-Sottani et al. ([2024](https://))

Summarising the posterior

Summarising trees is much more challenging

Presenting a single summary tree can be misleading



Summarising the posterior

Maximum clade credibility (MCC) tree — the tree in the posterior sample that has the highest posterior probability (i.e., clade support) across all nodes

The 95% highest posterior density (HPD) — the shortest interval that contains 95% of the posterior probability. The Bayesian equivalent of the 95% confidence interval

Marginal posterior density — the probability of a parameter regardless of the value of the others, represented by the histogram

Exercise (demo)