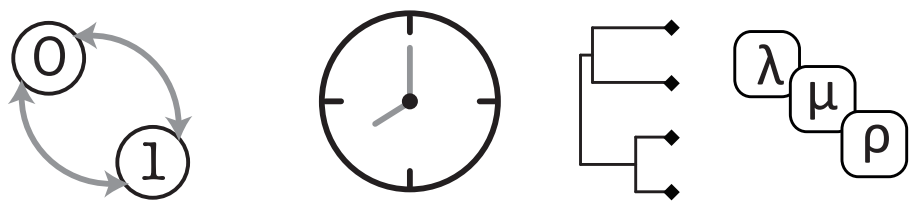
Introduction to MCMC

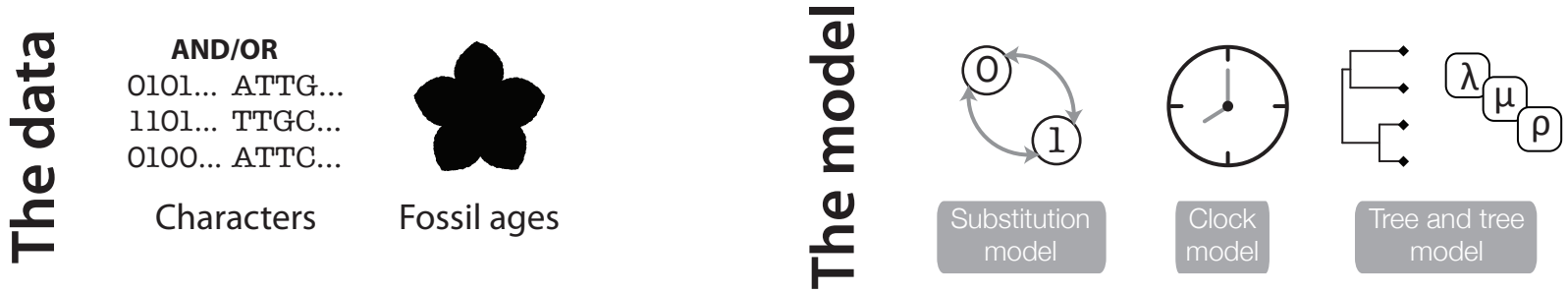
Recap: Bayesian phylogenetic dating requires three model components

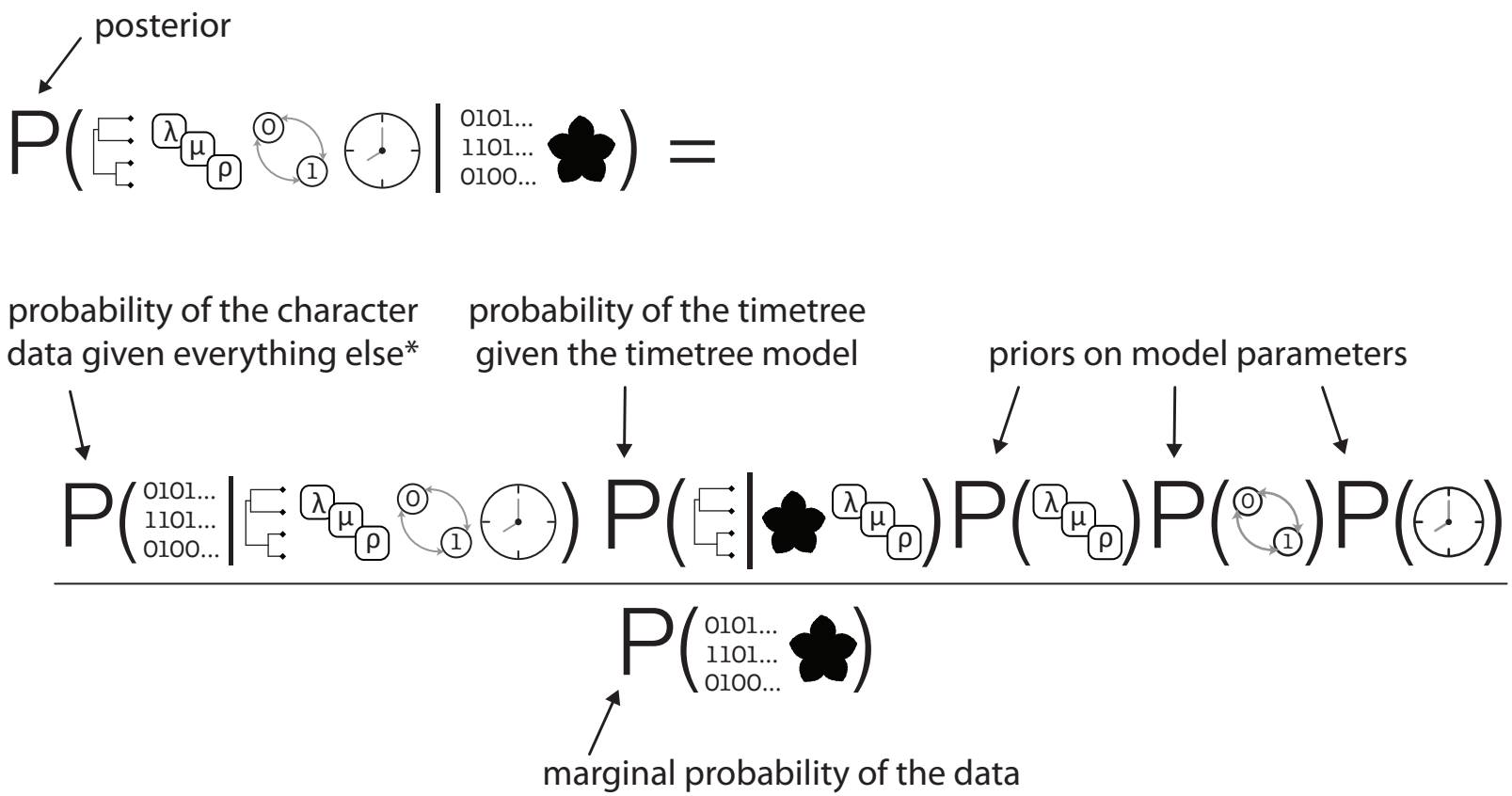
- The **substitution model** ← describes how sites evolve over time.
- The **clock model** ← describes how evolutionary rates vary across the tree.
- The **tree model** ← describes how trees grow over time. Temporal evidence is included here.



Substitution model | Clock model | Tree and tree model

Recap: Bayesian phylogenetic dating





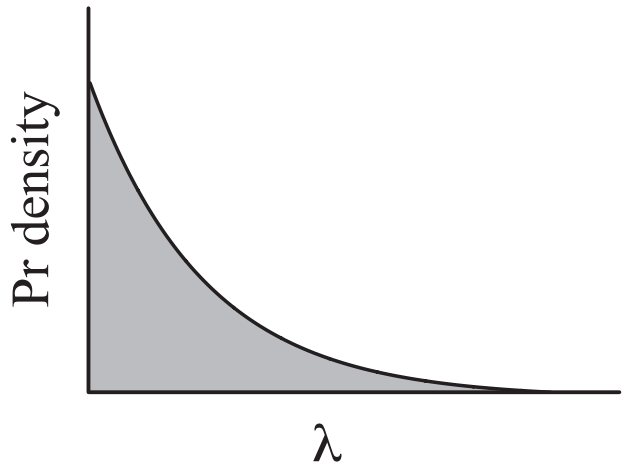
*the timetree, the parameters and the tripartite model

Probabilities versus probability densities

In phylogenetics, probabilities are not normally discrete (i.e. represented by a single value) and we're often dealing with a lot of uncertainty (esp. in the fossil record). Instead we typically work with **probability densities**.

See Paul Lewis's archery prior demo.

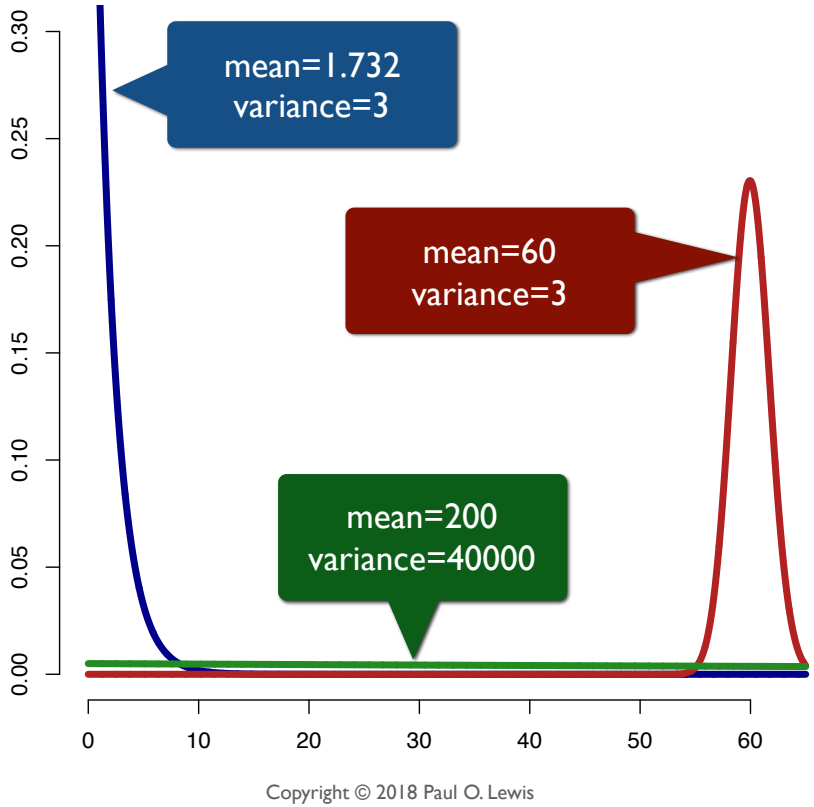
Probabilities versus probability densities



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 parameter values.

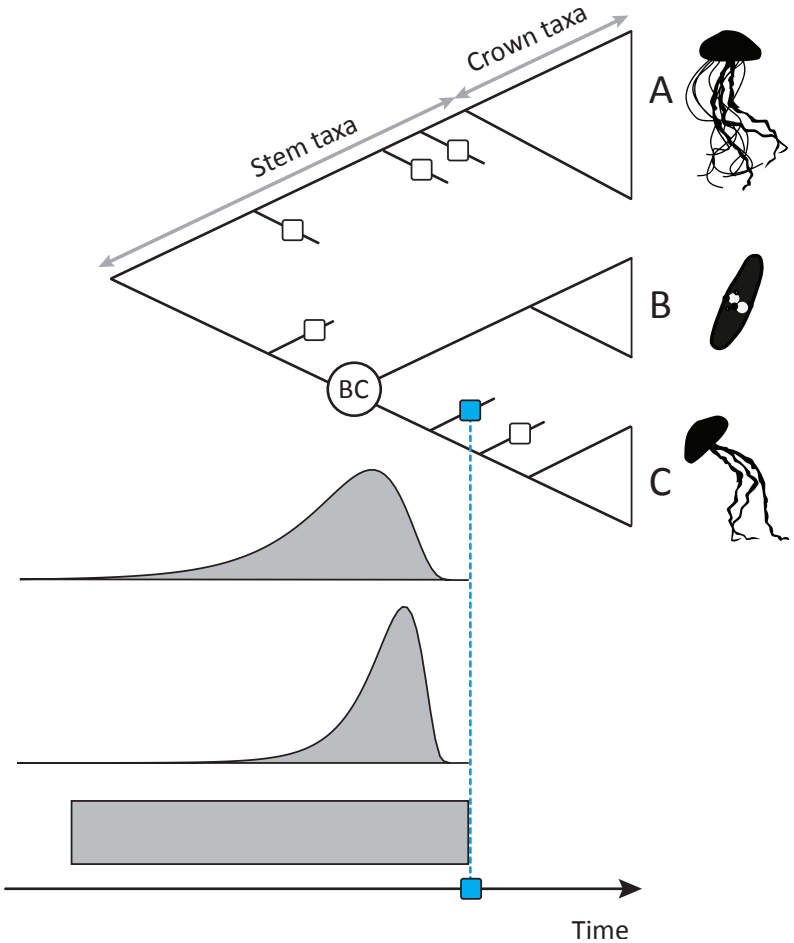
Probabilities versus probability densities



Variations of a gamma probability distribution.

- 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 parameter values.

Probabilities versus probability densities



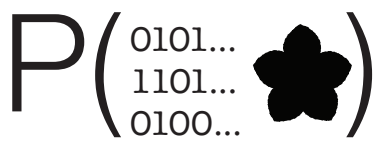
- The x-axis represents the value of our parameter λ.
- The y-axis does have a value but it is not so easily interpretable.
- The distribution height reflects the relative probability of a given range of parameter values.

Example calibration densities.

Why do we need Markov chain Monte Carlo?

Probability densities already introduce some complexity
$\rightarrow$ Remember the posterior is not usually a point estimate (i.e. a single value) but a range of values.

The marginal probability of the data is also very tricky to calculate.



Calculating this requires taking into account all possible alternative parameter combinations (e.g. all possible trees).

This makes it challenging to calculate the posterior analytically (i.e. exactly).

What is Markov chain Monte Carlo (MCMC)?

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

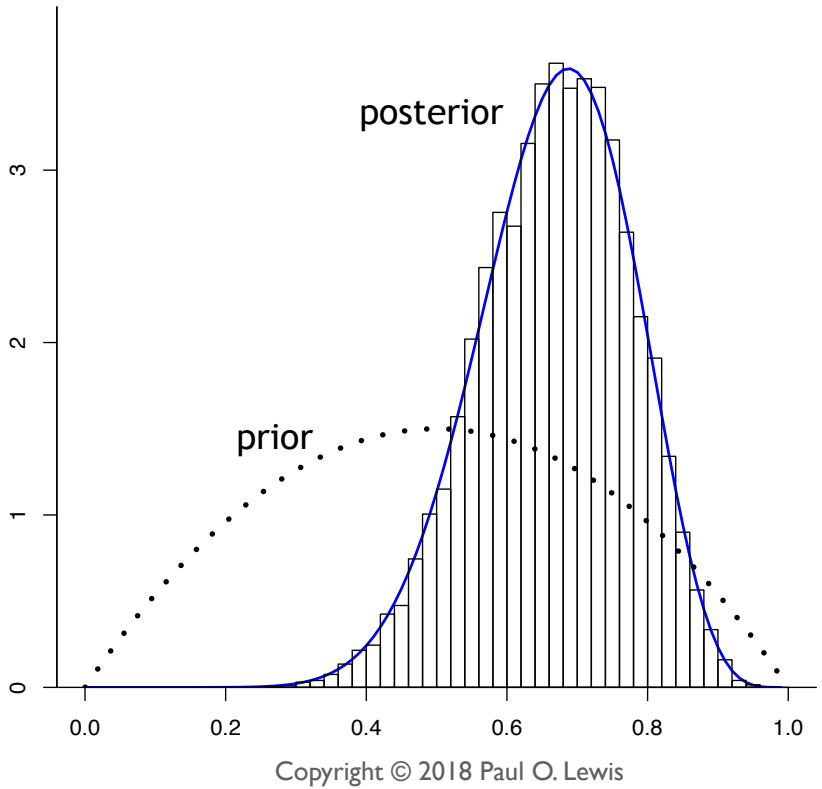
Markov chain means the progress of the algorithm doesn't depend on its past.

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.

Click here for a little bit of history.

What is Markov chain Monte Carlo (MCMC)?



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

Image source: Paul Lewis's phyloseminar lecture.

MCMC robot's rules

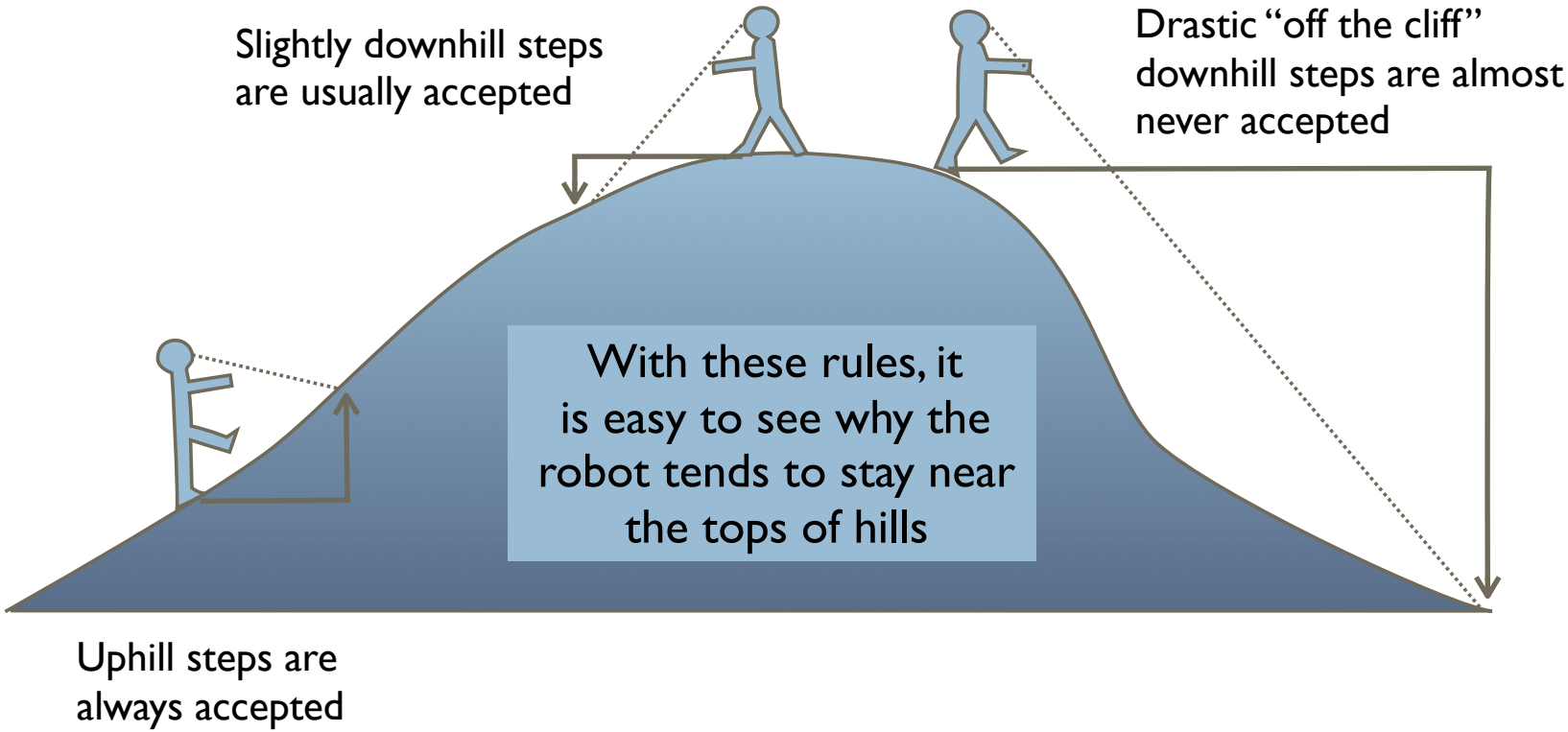


Image source: Paul Lewis's phyloseminar lecture.

Actual rules (Metropolis algorithm)

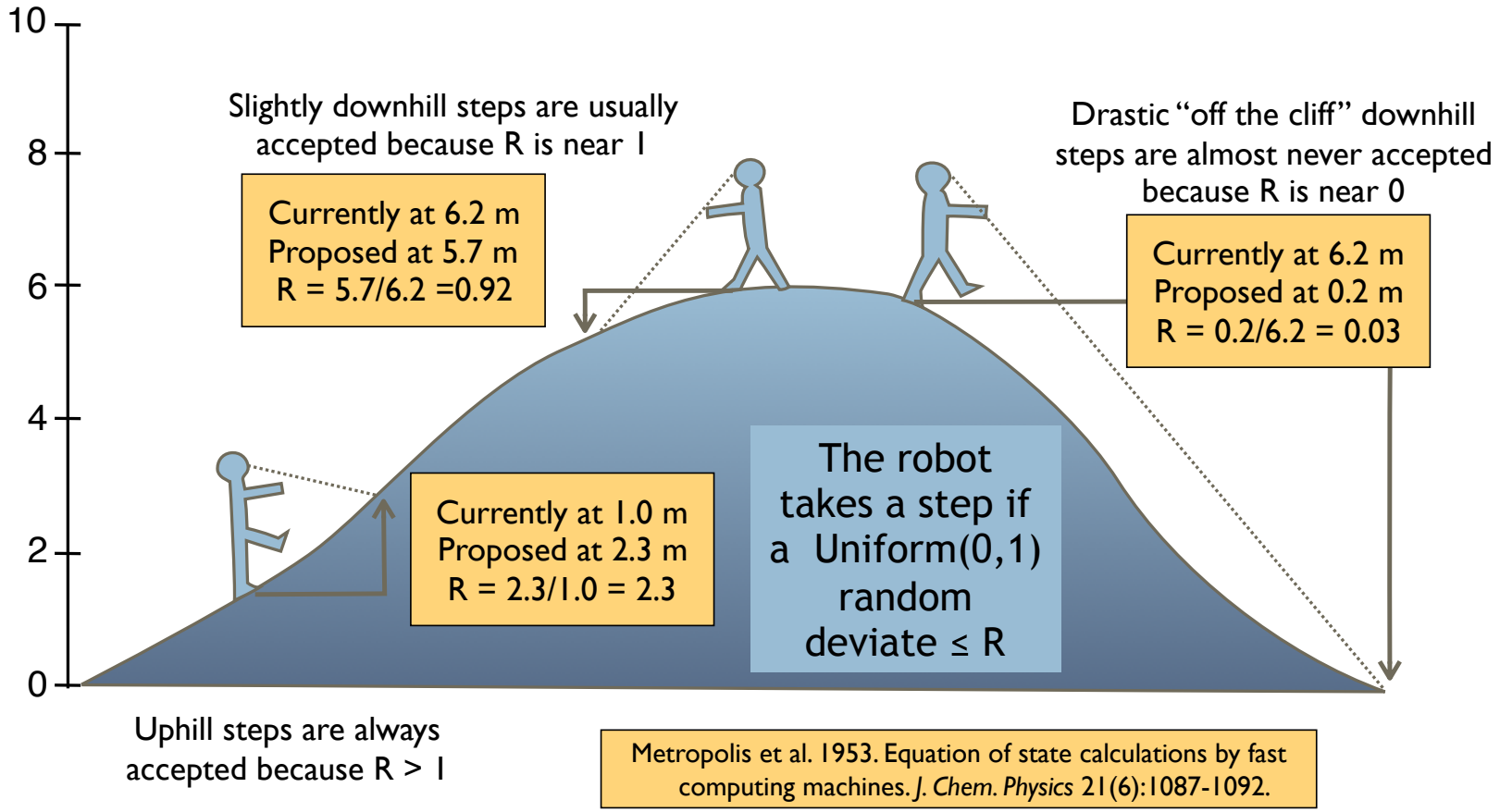


Image source: Paul Lewis's phyloseminar lecture.

The marginal likelihood is cancelled

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

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

Posterior odds

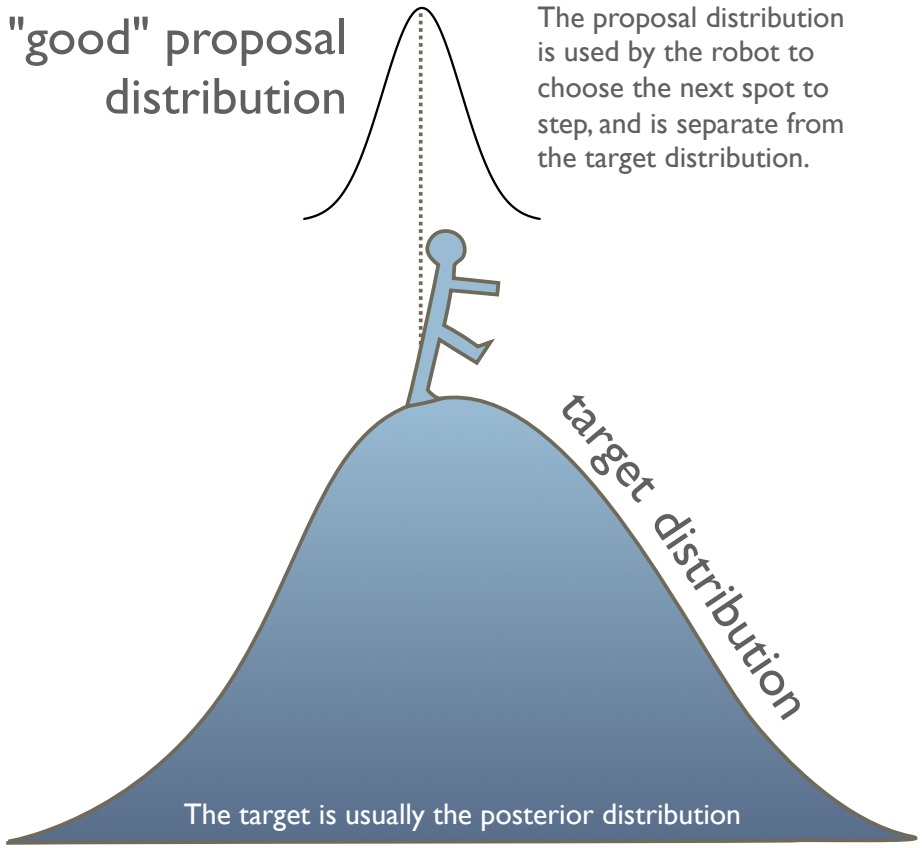
Apply Bayes' rule to both top and bottom

Likelihood ratio

Prior odds

Image source: Paul Lewis's phyloseminar lecture.

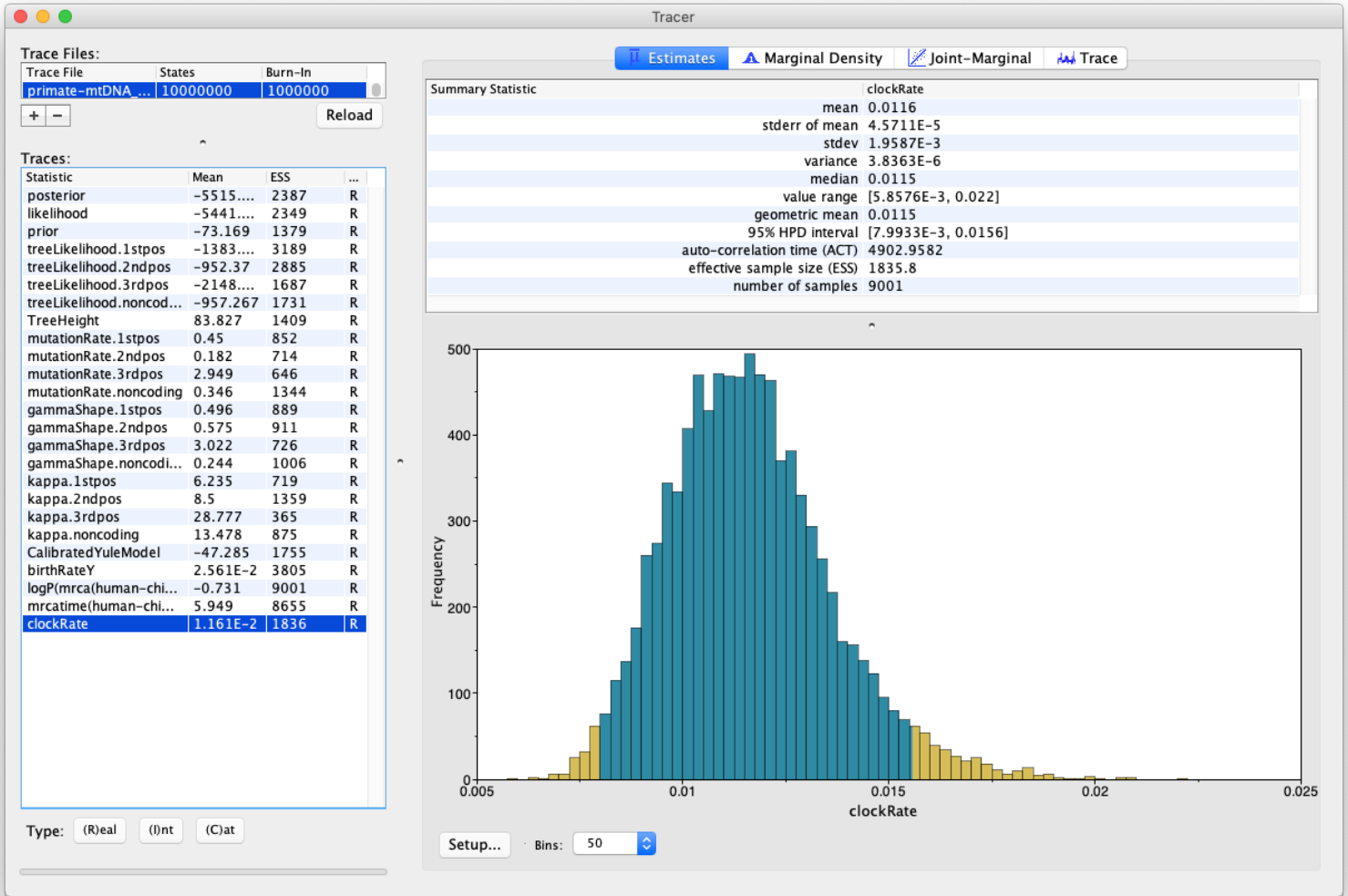
MCMC robot



See Paul Lewis's MCMC robot demo.

Summarising the posterior

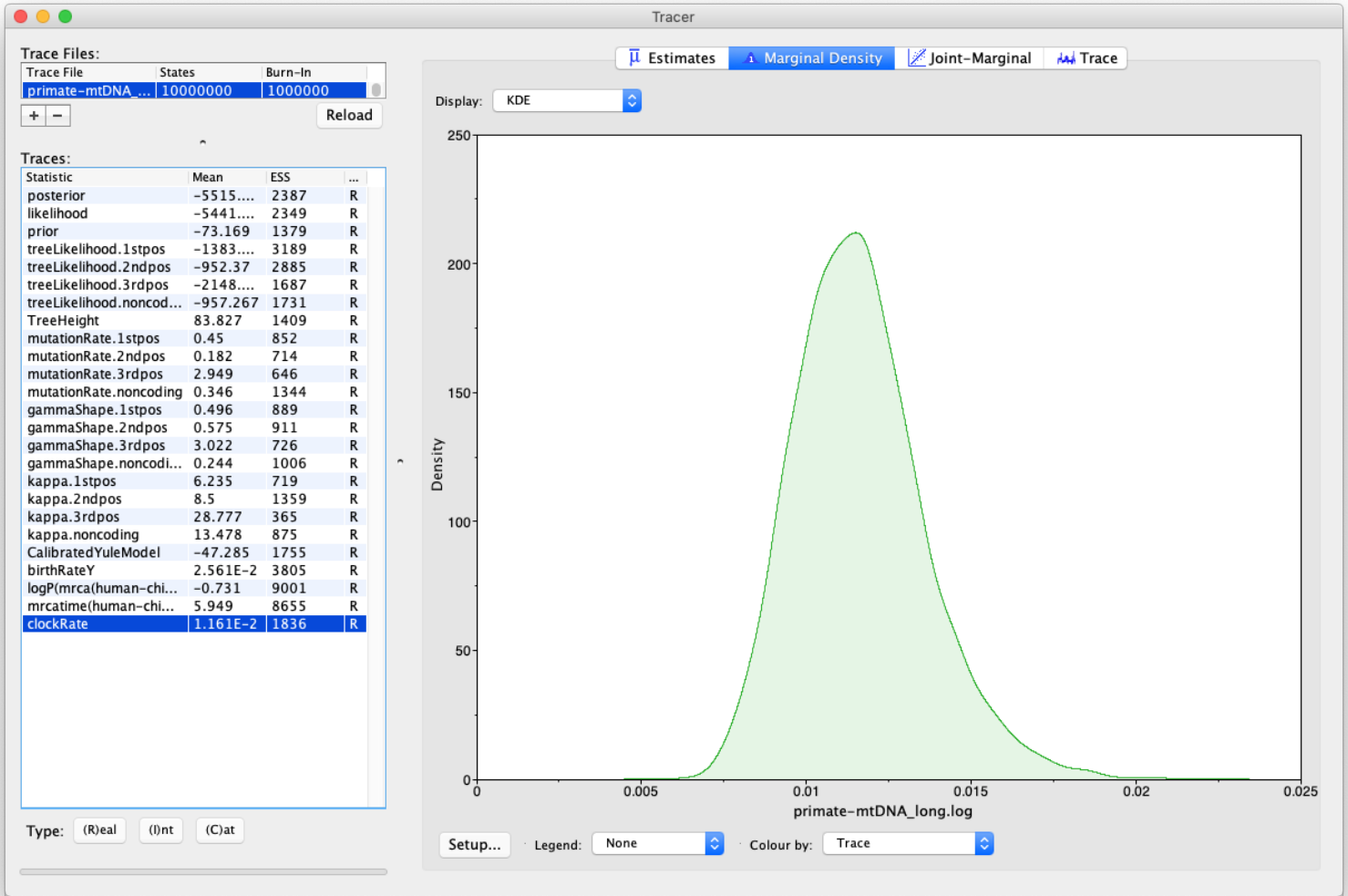
Tracer is an amazing program for exploring MCMC output.



Example MCMC output. Source: taming-the-beast.org.

Summarising the posterior

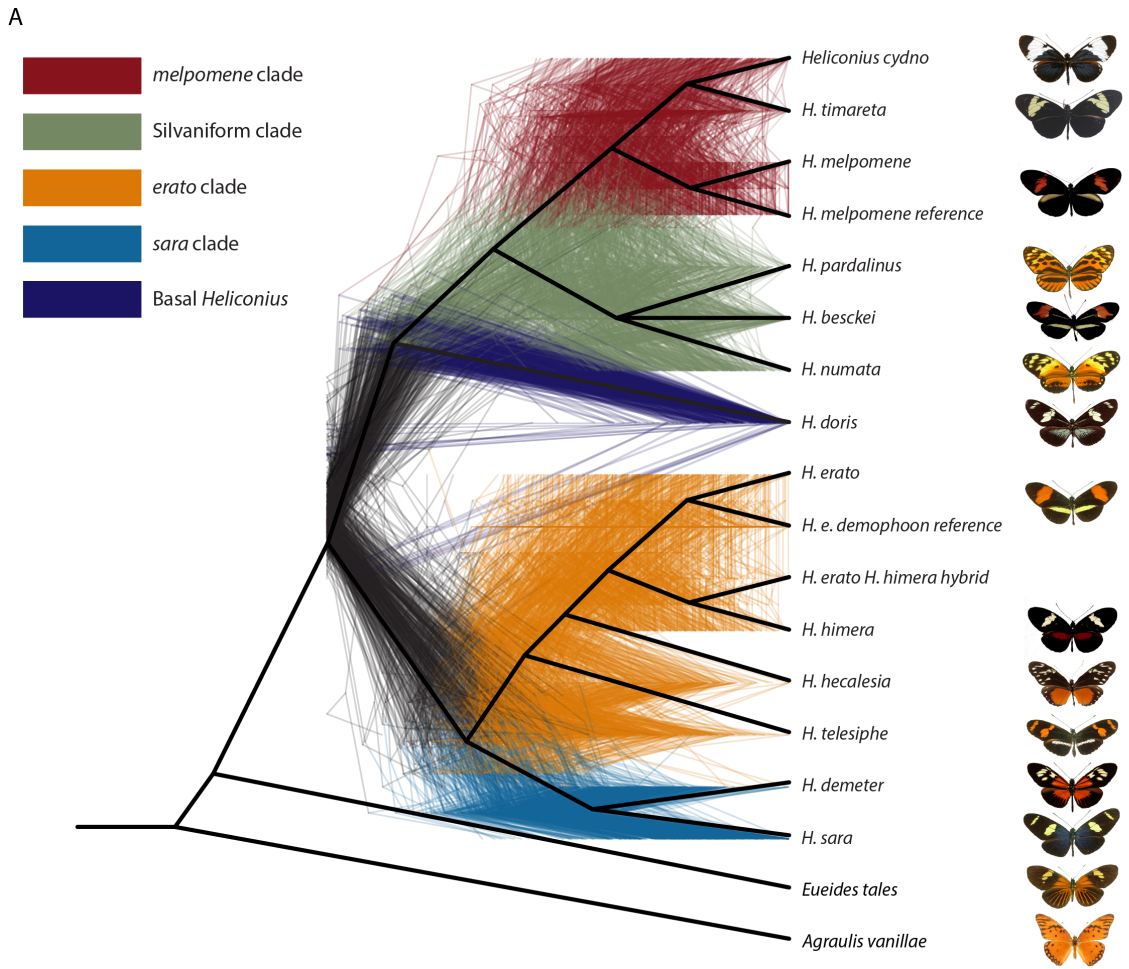
Tracer is an amazing program for exploring MCMC output.



Example MCMC output. Source: taming-the-beast.org.

Summarising the posterior

Summarising trees is much more challenging. Presenting a single summary tree can sometimes be misleading.



Edelman et al. (2019) *Science*.

Summarising the posterior

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.

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

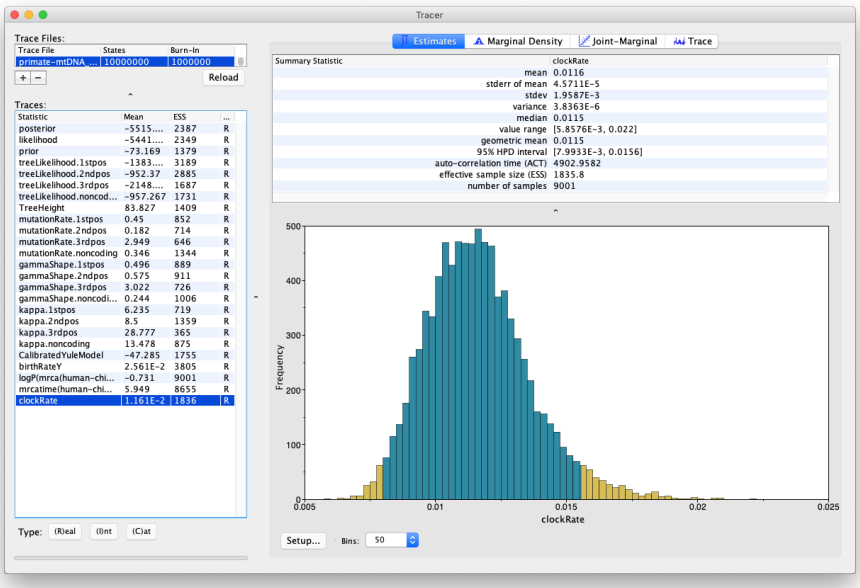
For more on issues associated with summary tree methods see O'Reilly & Donoghue (2018) *Sys Bio*.

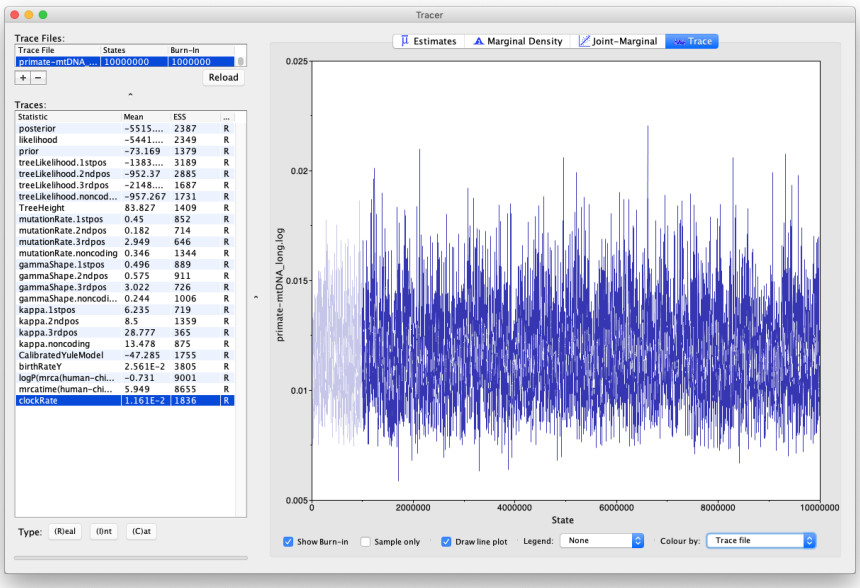
Convergence

How do you know if you've run the run the chain long enough? $\rightarrow$ You don't! But there are some clues.

Convergence

Good mixing.

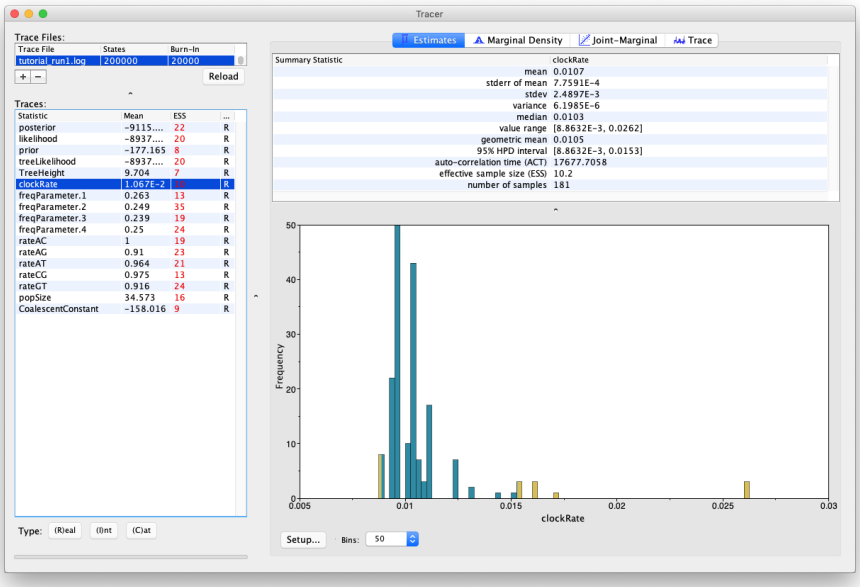


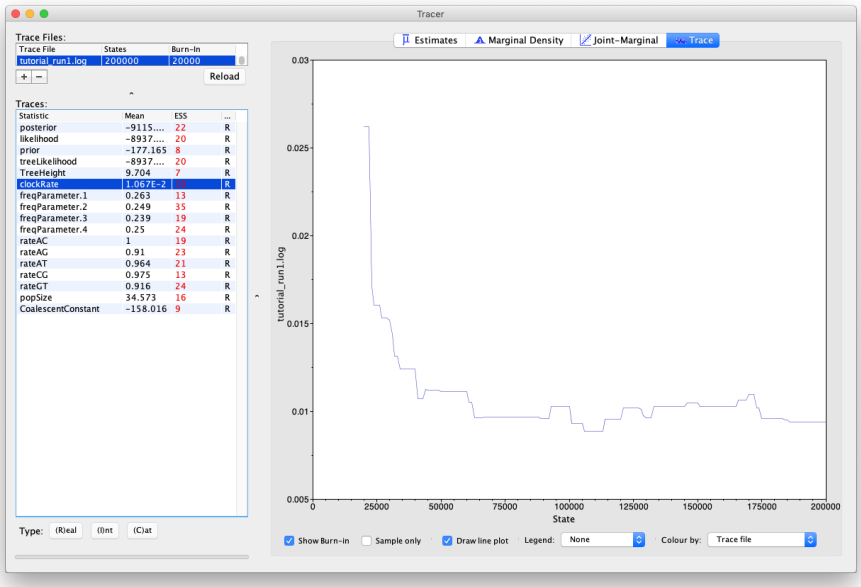


Example MCMC output. Source: taming-the-beast.org.

Convergence

Poor mixing.

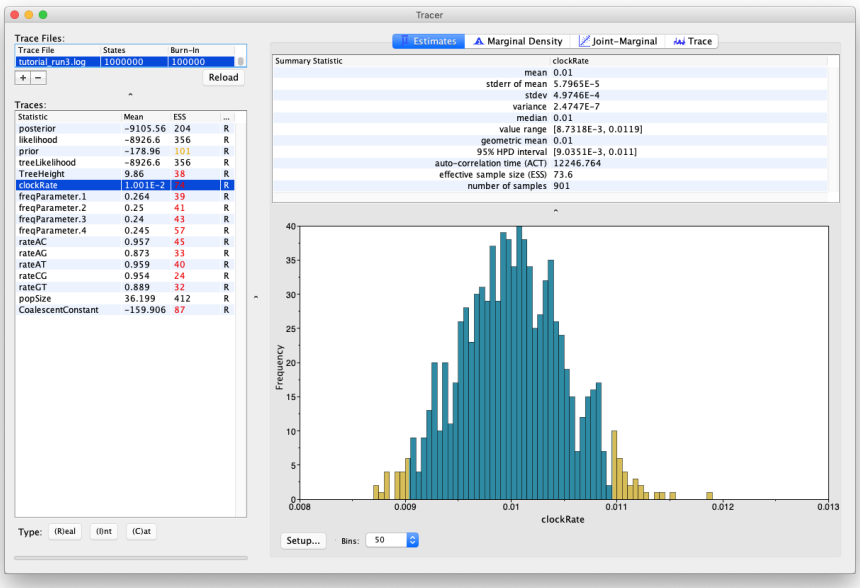


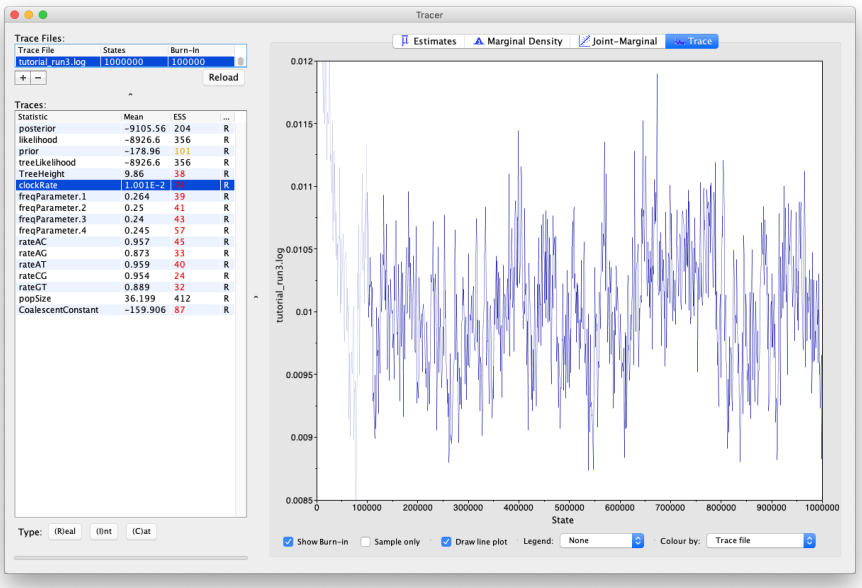


Example MCMC output. Source: taming-the-beast.org.

Convergence

Better mixing.





Revisit Paul Lewis's MCMC robot.

Take homes

MCMC is an elegant algorithm trick to infer the posterior distribution.

It samples values directly from posterior in proportion to how probable they are, resulting in a histogram, which provides a good approximation of the posterior.