
Lecture 2.2

Bayesian Phylogenetics

The Bayesian framework

Bayesian phylogenetic analysis

- Bayesian phylogenetic analysis was developed in the mid 1990s
- Now one of the most widely used methods
- Bayes's theorem (1763)
- Reverend Thomas Bayes



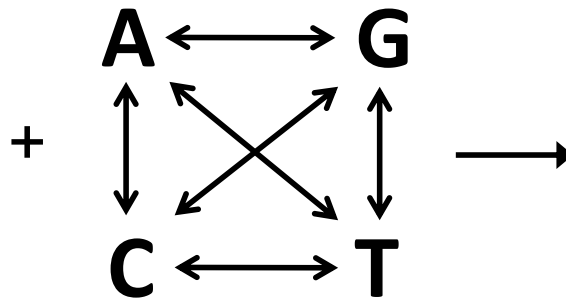
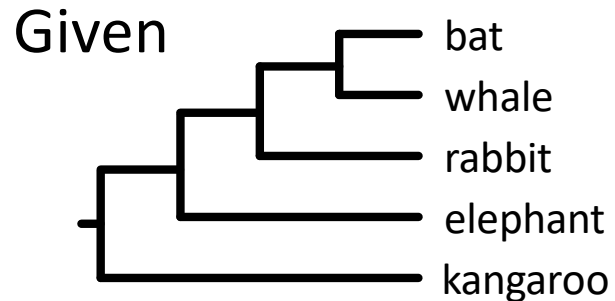
*Image probably
not of Thomas Bayes*

In phylogenetics

$$\Pr(\theta \mid D) \propto \Pr(\theta) \Pr(D \mid \theta)$$

Bayesian phylogenetic analysis

Maximum likelihood

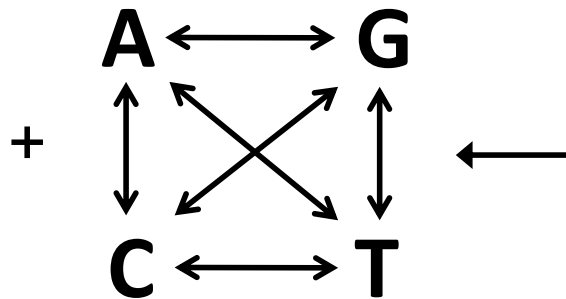
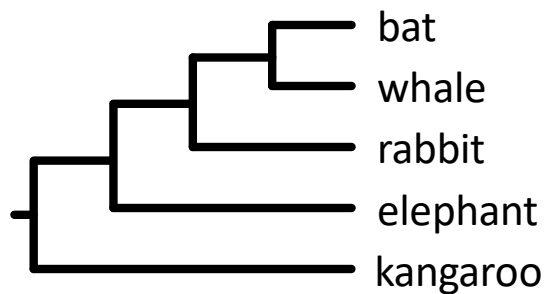


Probability of?

| | |
|----------|--------------|
| bat | CCGTTAGTAACT |
| whale | CCGTTAGTAACT |
| rabbit | CCGATAGTTACT |
| elephant | TCGTTAGTTACC |
| kangaroo | TCATTGGTTACT |

Bayesian inference

Probability of?



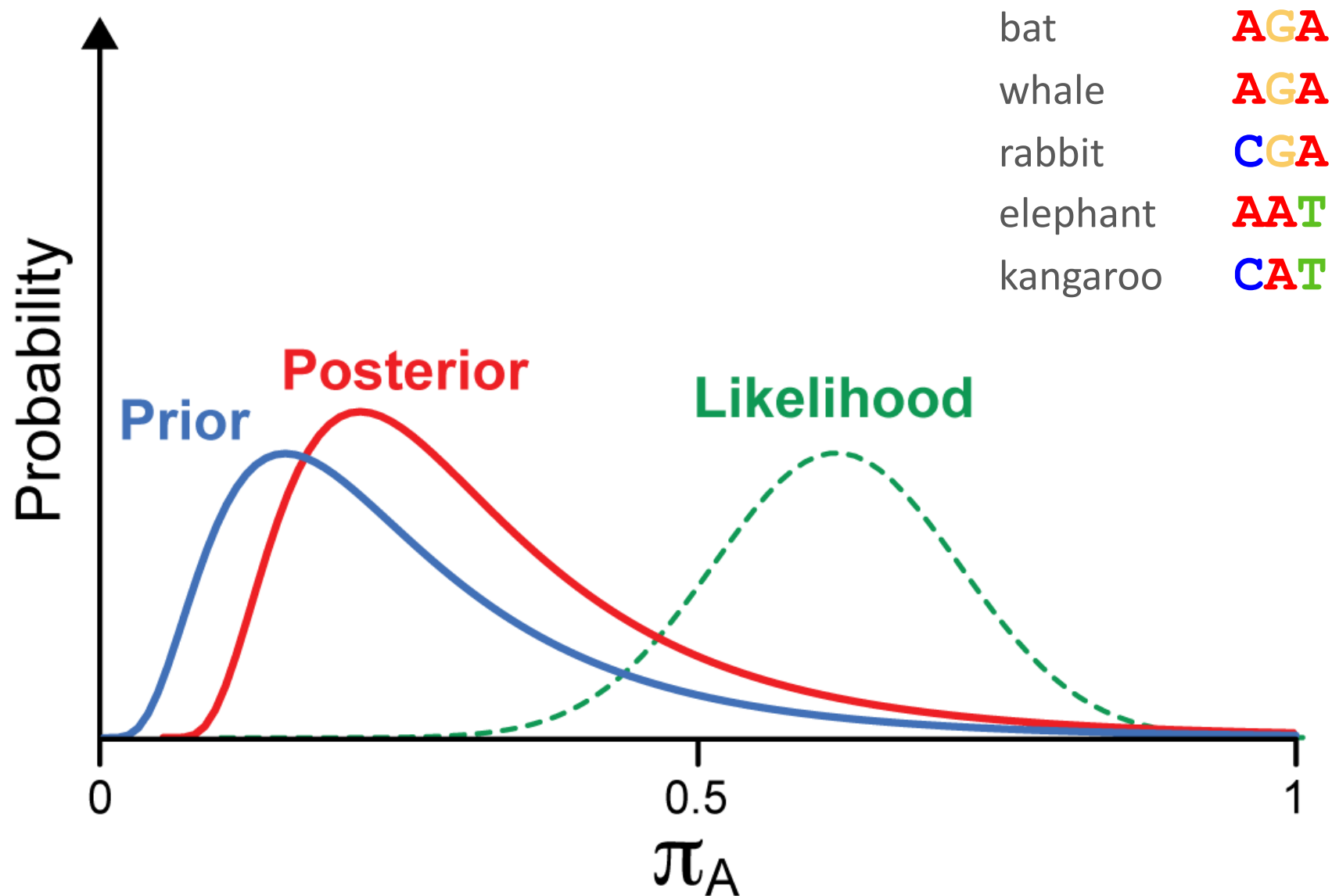
Given

| | |
|----------|--------------|
| bat | CCGTTAGTAACT |
| whale | CCGTTAGTAACT |
| rabbit | CCGATAGTTACT |
| elephant | TCGTTAGTTACC |
| kangaroo | TCATTGGTTACT |

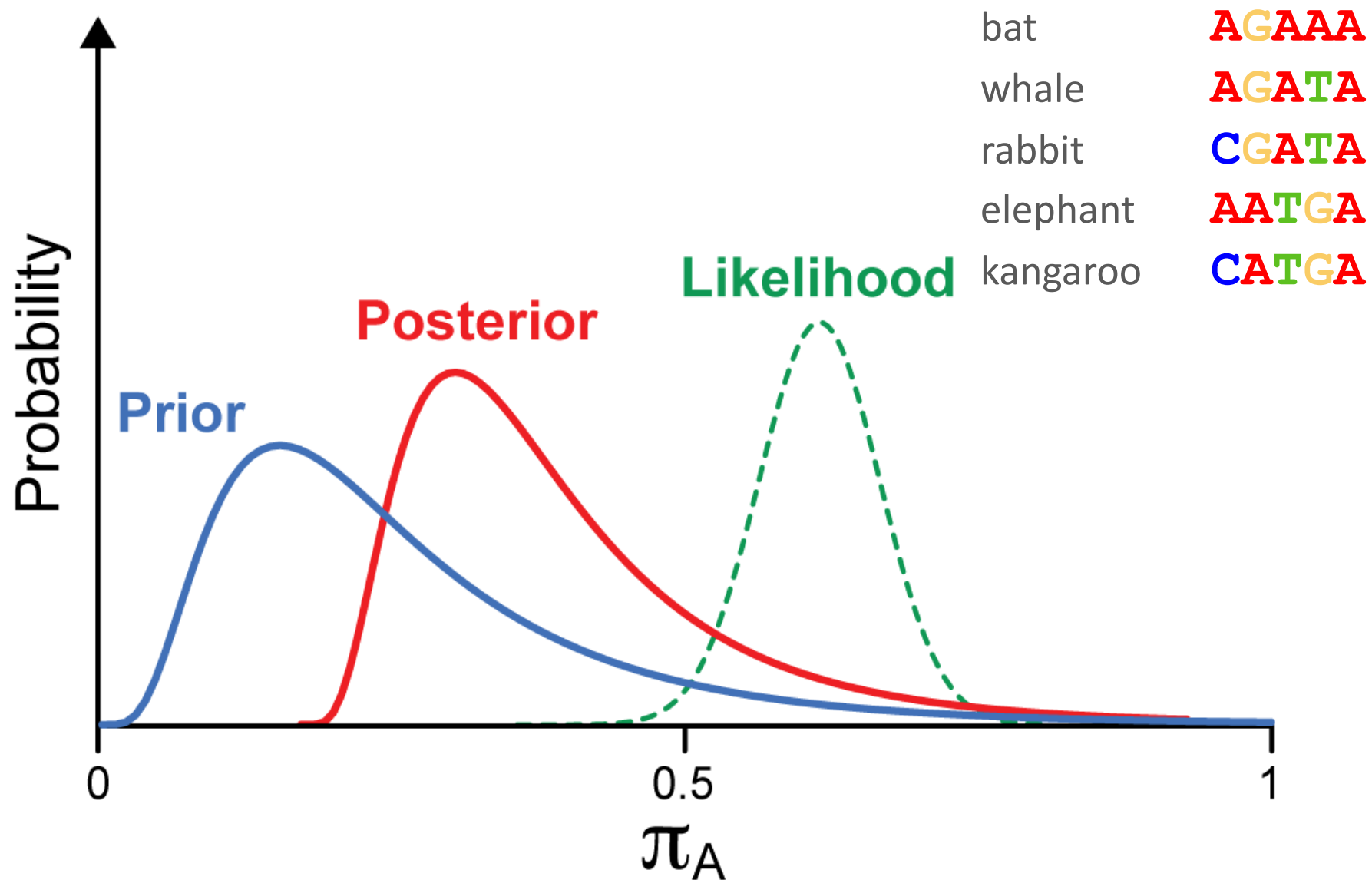
The Bayesian paradigm

- Contrast with frequentist statistics (likelihood)
- Parameters have **distributions**
- Before the data are observed, each parameter has a **prior distribution**
- The **likelihood** of the data is computed
- The prior distribution is combined (updated) with the likelihood to yield the **posterior distribution**

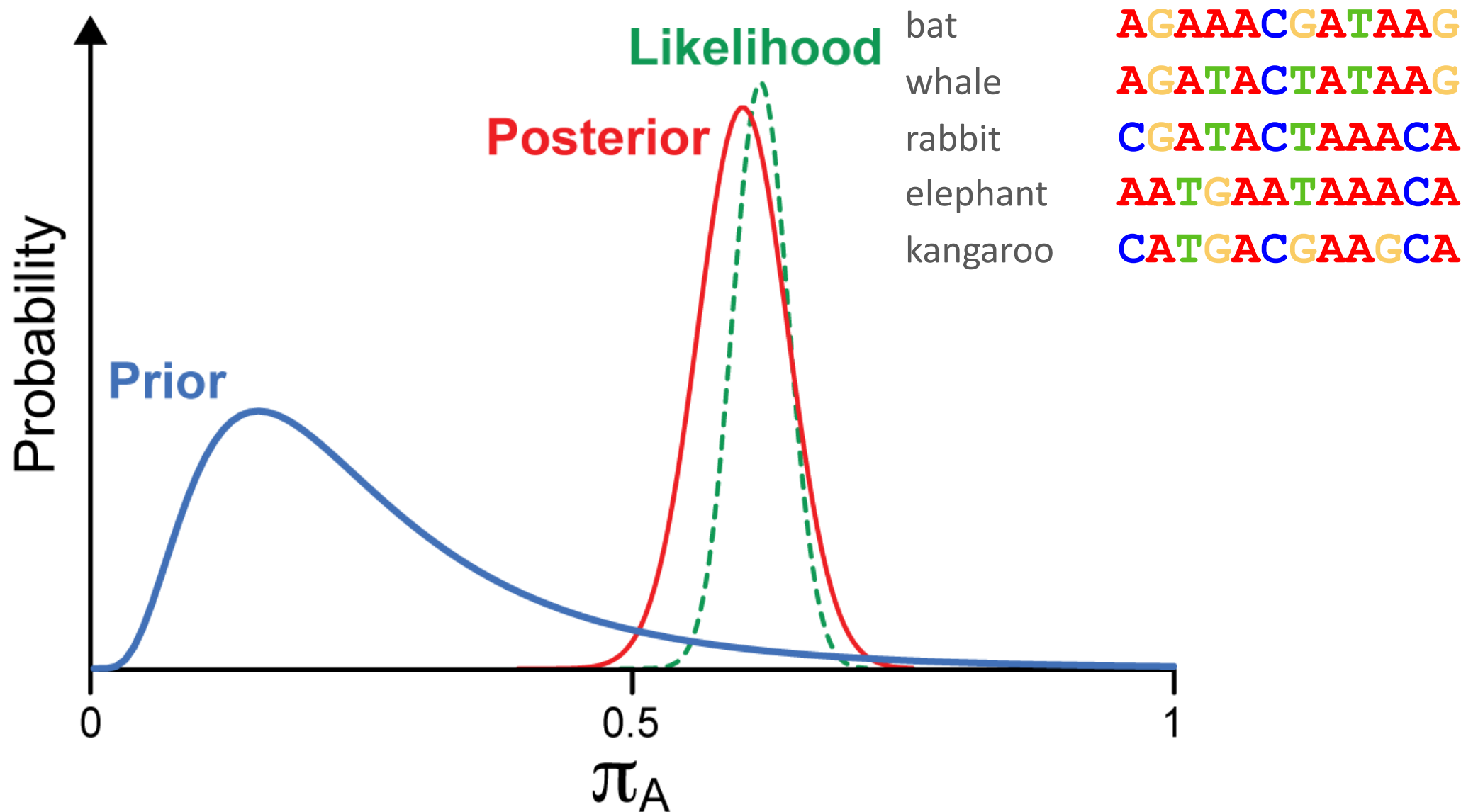
Simple example



Simple example



Simple example



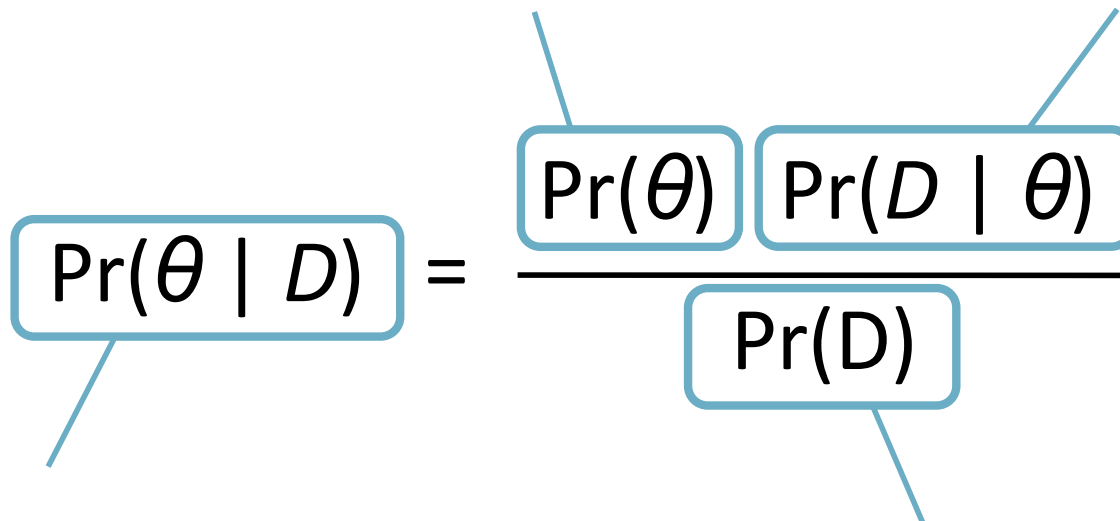
Bayesian inference

Prior

Specified by user,
independent of data

Likelihood

Calculated from data



The diagram illustrates the Bayesian inference formula:
$$\Pr(\theta \mid D) = \frac{\Pr(\theta) \Pr(D \mid \theta)}{\Pr(D)}$$
 Each term in the formula is enclosed in a light blue rounded rectangle. A line connects the 'Prior' label to the $\Pr(\theta)$ term. Another line connects the 'Likelihood' label to the $\Pr(D \mid \theta)$ term. A third line connects the 'Posterior' label to the $\Pr(\theta \mid D)$ term. A fourth line connects the 'normalising constant' label to the $\Pr(D)$ term in the denominator.

Posterior

This is what we
want to estimate

normalising constant
marginal likelihood of the data
model likelihood

Bayesian inference

Prior prob of tree

Topology
Branch lengths

Prior prob of substitution model parameters

Rate parameters
Base frequencies

$$\boxed{\Pr(\tau, M \mid D)} = \frac{\boxed{\Pr(\tau)} \boxed{\Pr(M)} \boxed{\Pr(D \mid \tau, M)}}{\Pr(D)}$$

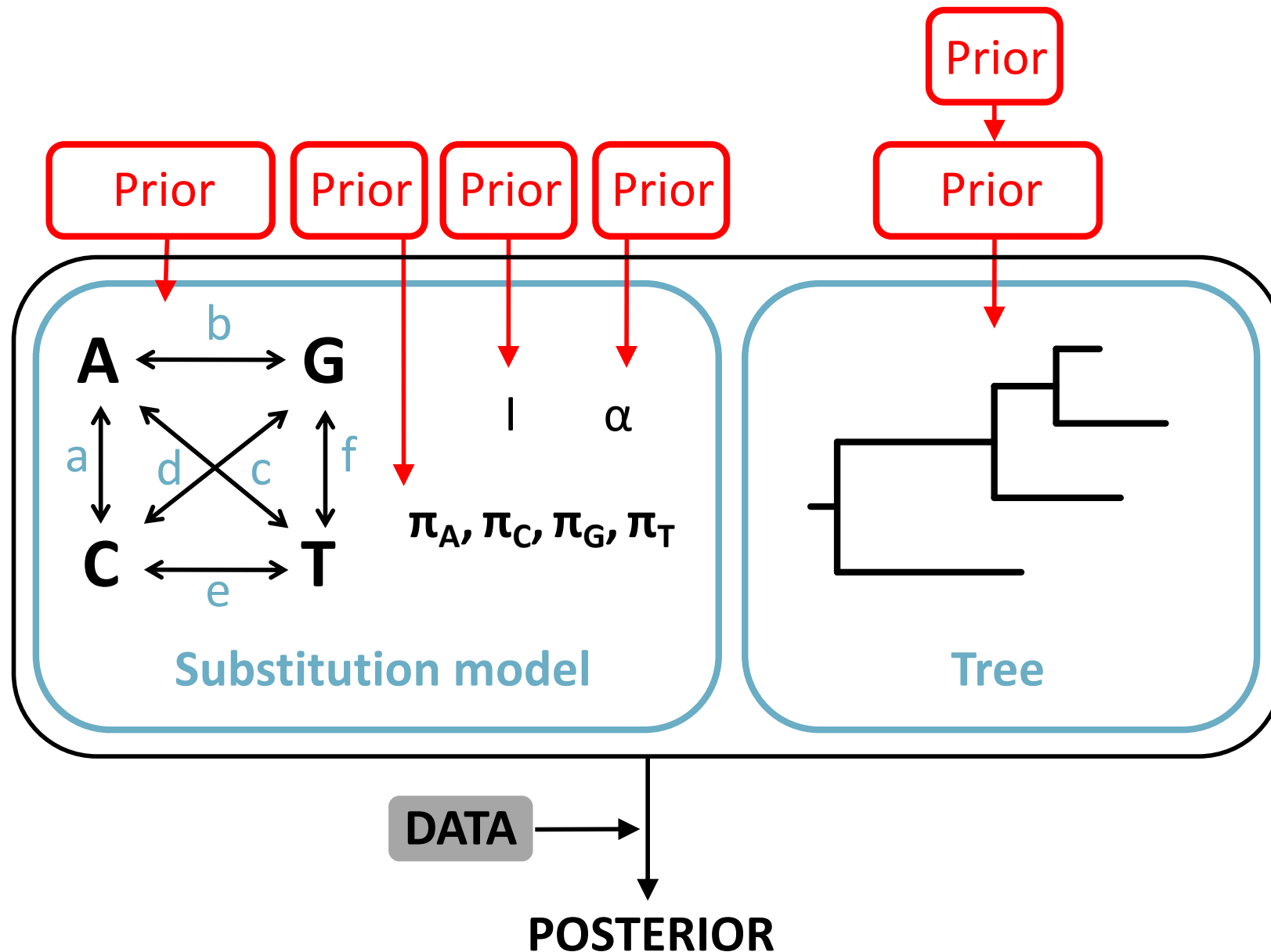
Posterior

This is what we
want to estimate

Likelihood

Calculated from data

Bayesian hierarchical model



Priors

Priors

- Priors are chosen in the form of probability distributions
- Reflect our prior expectations (and uncertainty) about values of parameters (without knowledge of the data)
 - Past observations
 - Personal beliefs
 - Use of a biological model

Continuous distributions

- Uniform
- Normal

Used to specify prior distributions of various continuous parameters

- Exponential
- Lognormal
- Gamma

Used to specify prior distributions of continuous parameters that cannot take negative values

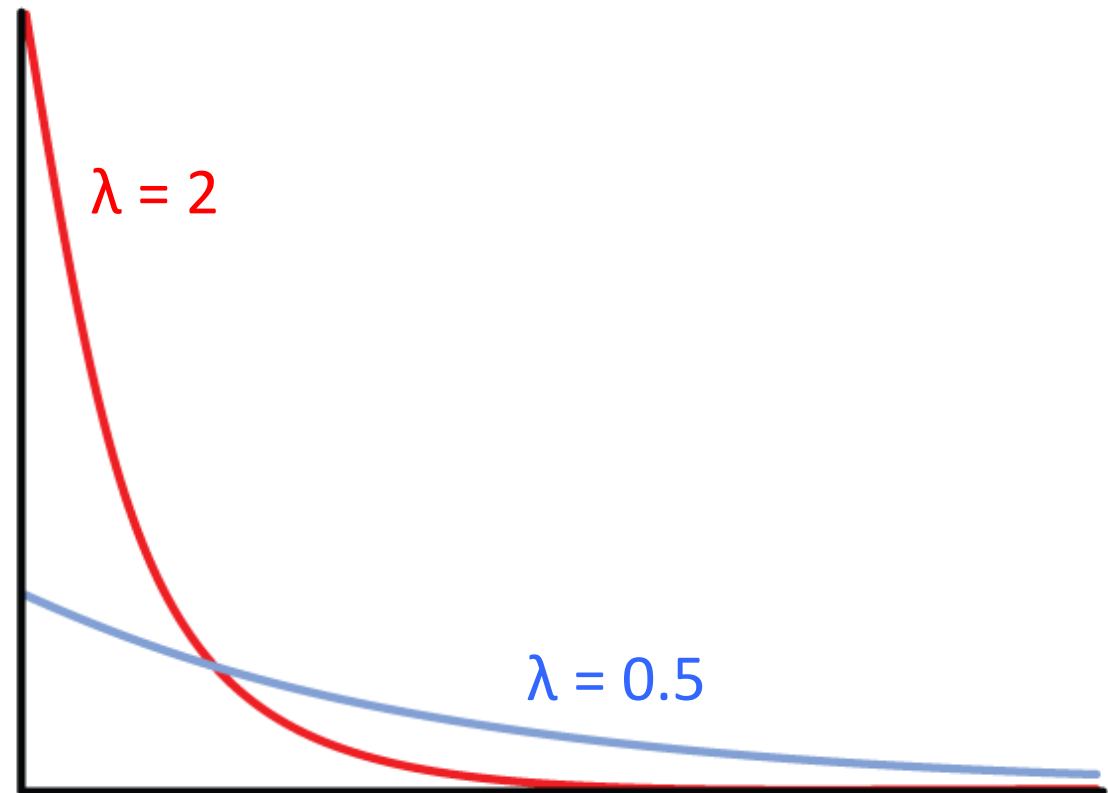
- Beta
- Dirichlet

Continuous distributions

- Uniform
- Normal
- **Exponential**
- Lognormal
- Gamma
- Beta
- Dirichlet

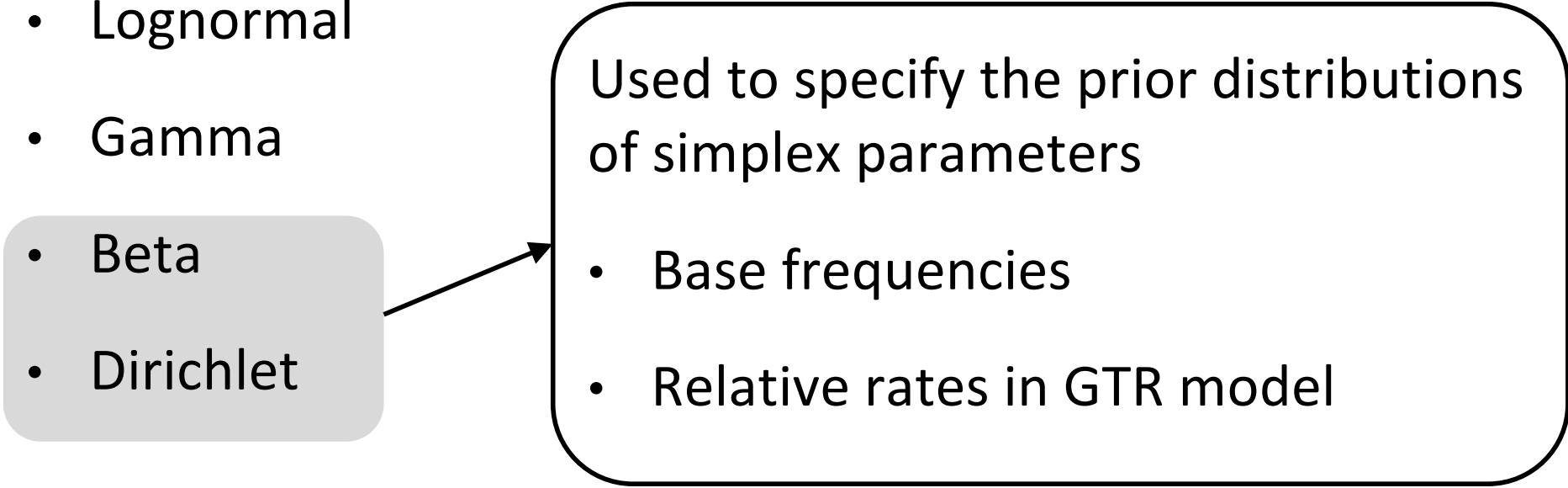
Parameters

- λ = rate of decay



Continuous distributions

- Uniform
- Normal
- Exponential
- Lognormal
- Gamma
- Beta
- Dirichlet



Used to specify the prior distributions of simplex parameters

The diagram consists of a light gray rounded rectangle on the left containing the list items 'Beta' and 'Dirichlet'. An arrow points from the right side of this rectangle to a larger rounded rectangle on the right. This right rectangle contains the text 'Used to specify the prior distributions of simplex parameters' followed by a bulleted list: 'Base frequencies' and 'Relative rates in GTR model'.

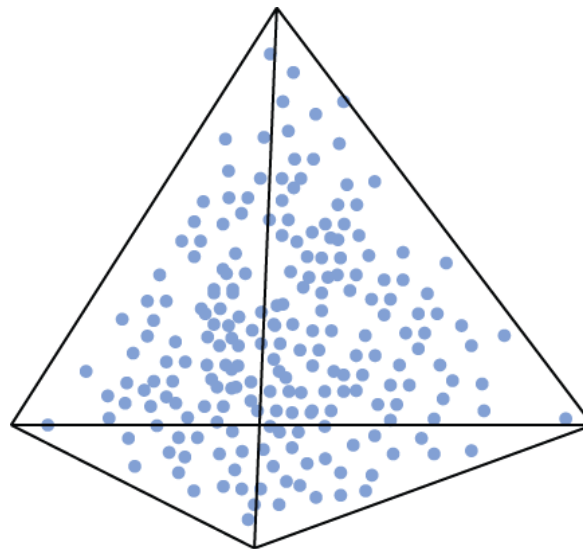
- Base frequencies
- Relative rates in GTR model

Continuous distributions

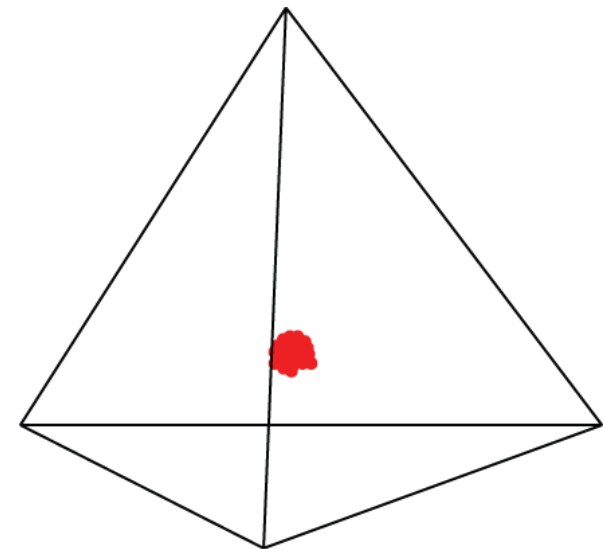
- Uniform
- Normal
- Exponential
- Lognormal
- Gamma
- Beta
- **Dirichlet**

Parameters

- $\alpha_1, \alpha_2, \dots$ = shape parameters



$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$$



$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 300$$

Discrete distributions

- Bernoulli distribution
- Binomial
- Multinomial
- Poisson

Default priors

| | <i>BEAST2</i> | <i>MrBayes</i> |
|--|----------------------|------------------------|
| Rate matrix parameters | Gamma(0.05,10) | Dirichlet(1,1,1,1,1,1) |
| Base frequencies | Uniform(0,1) | Dirichlet(1,1,1,1) |
| Shape parameter (α) | Exponential(1) | Exponential(2) |
| Proportion invariable | Uniform(0,1) | Uniform(0,1) |

Can specify uninformative priors where appropriate

Tree Prior

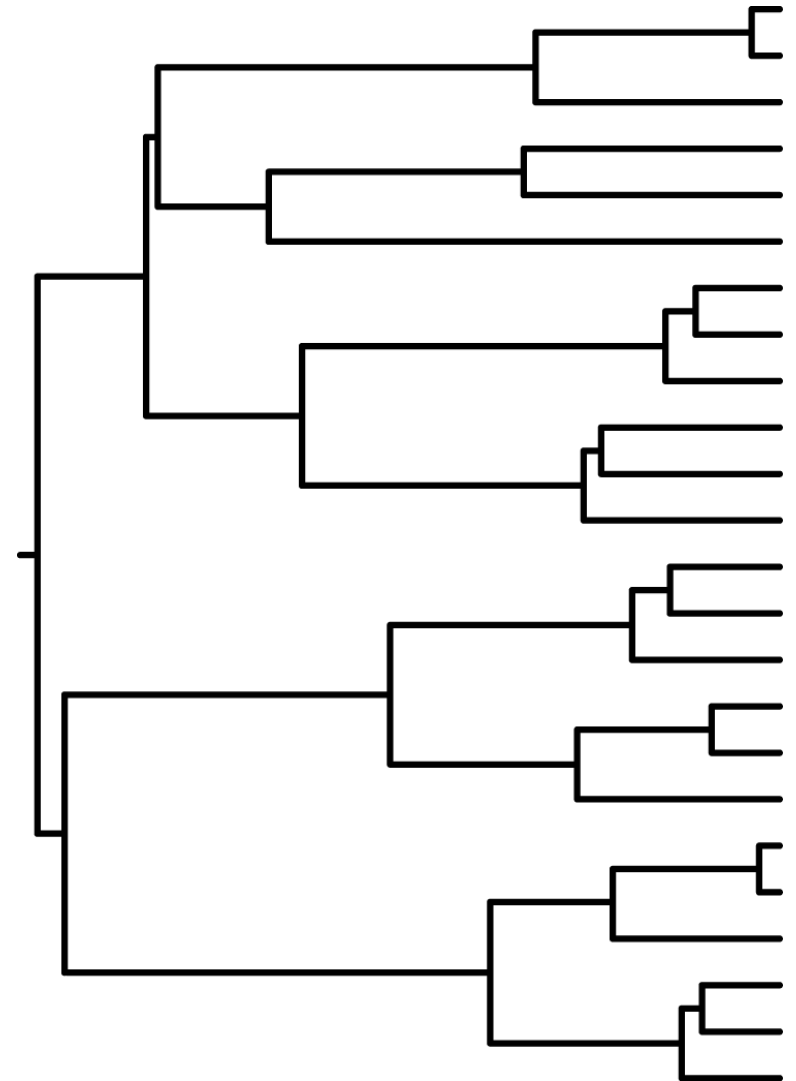
Tree prior

1. Use a **flat prior** (*MrBayes*)
 - All trees have equal probability
 - Also need a prior for branch lengths or node times

 2. Use a **biological model** (*BEAST* and *MrBayes*)
 - Among species: speciation model
 - Within species: coalescent model
- } Priors on rooted trees

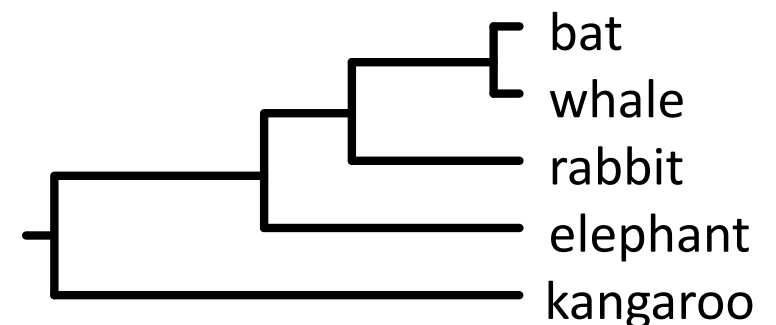
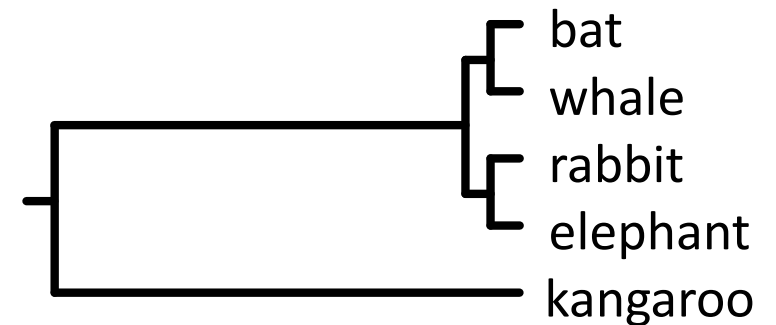
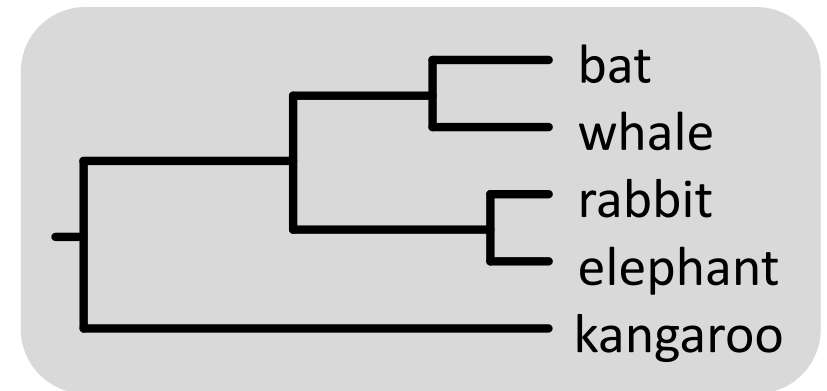
Speciation model

- Tree shape described by a stochastic branching process
- **Yule process**
 - The root lineage splits into two
 - Lineages split at a constant rate
 - Simulates speciation process
- **Birth-death process**
 - Allow lineages to go extinct



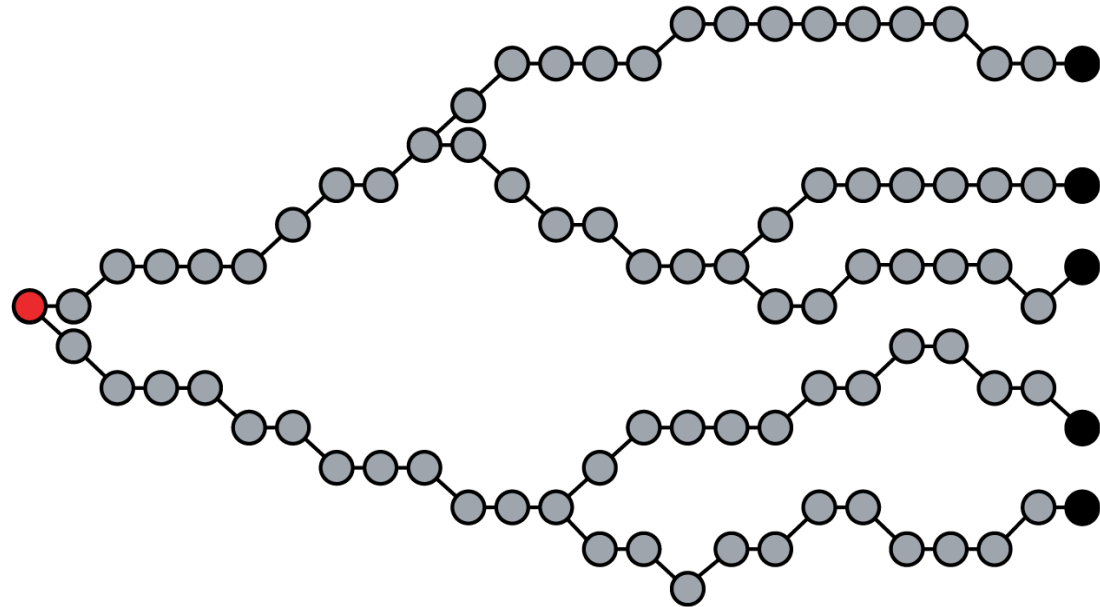
Speciation model

- Tree shape described by a stochastic branching process
- **Yule process**
 - The root lineage splits into two
 - Lineages split at a constant rate
 - Simulates speciation process
- **Birth-death process**
 - Allow lineages to go extinct

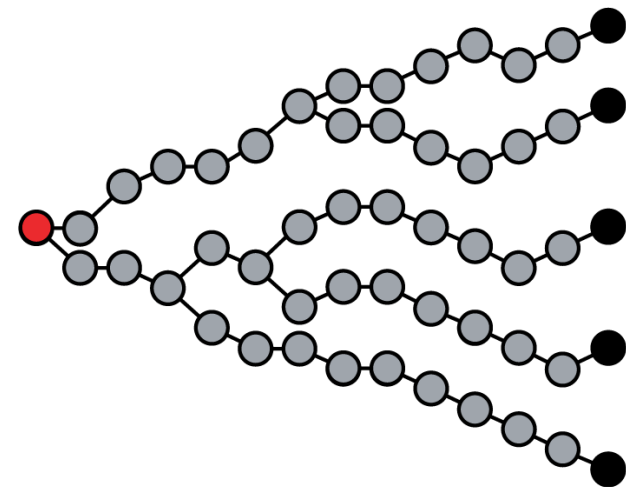


Coalescent model

Constant size



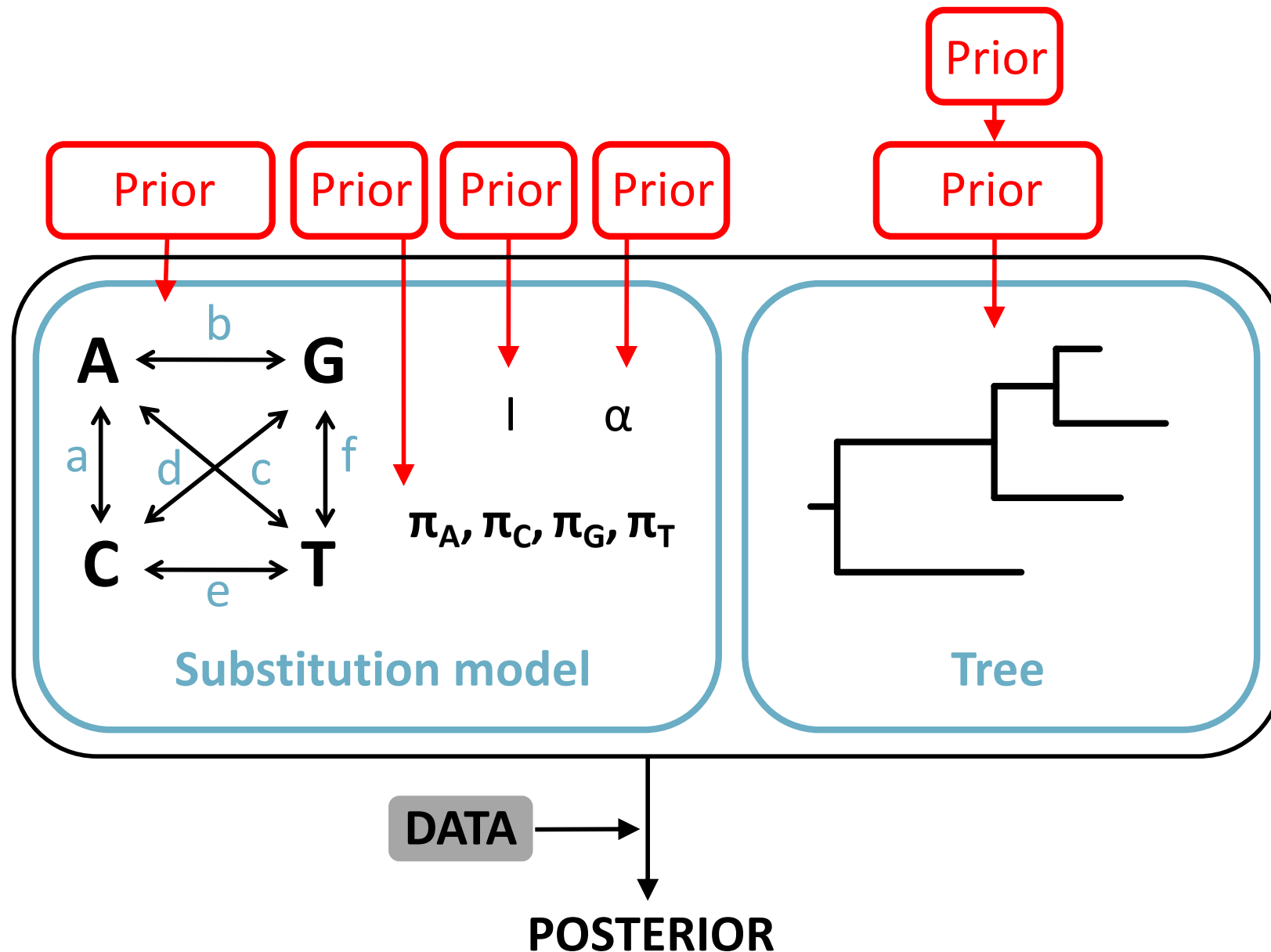
Exponential growth



Choosing a tree prior

- Test whether inferences are robust to the choice of tree prior
- Mixed data sets: multiple sequences from each species
 - Birth-death prior generally works well
- Compare tree priors using Bayesian model selection

Bayesian hierarchical model

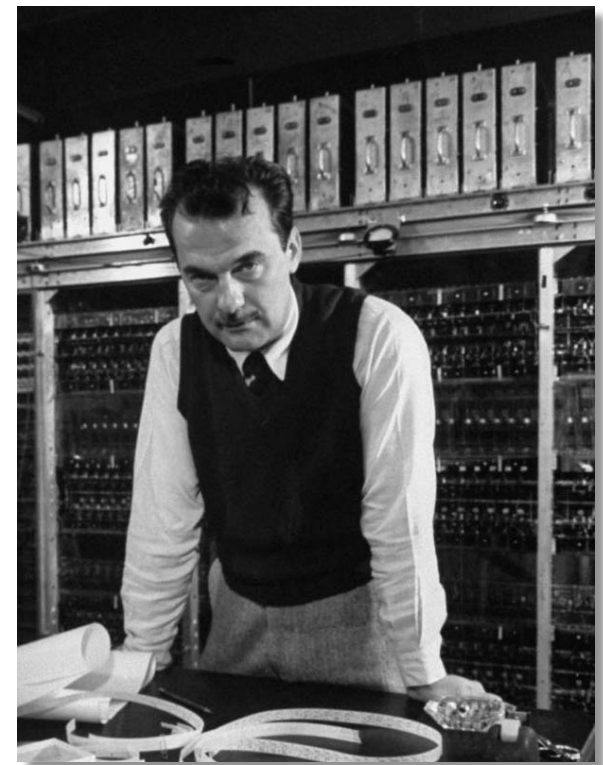


Posterior Distribution

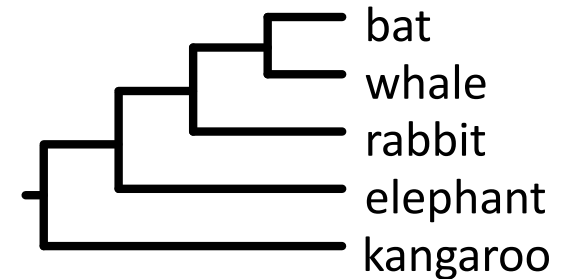
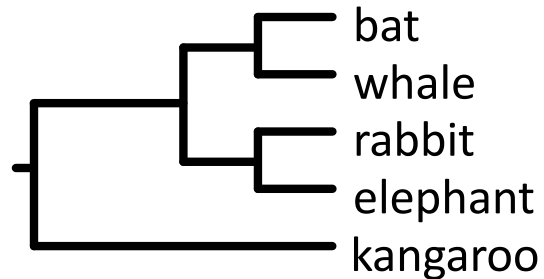
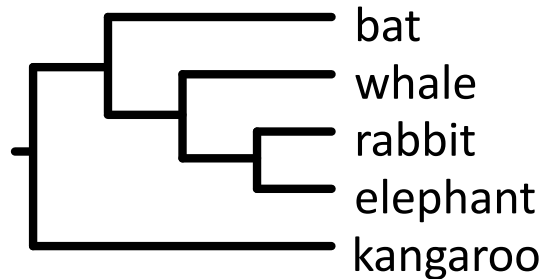
Estimating the posterior

- Impossible to obtain the posterior directly
- Instead, the posterior can be estimated using **Markov chain Monte Carlo simulation**
- This is usually done using the **Metropolis-Hastings algorithm**

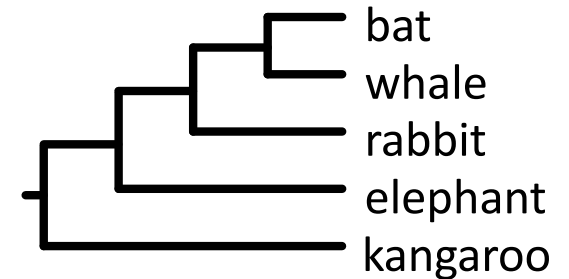
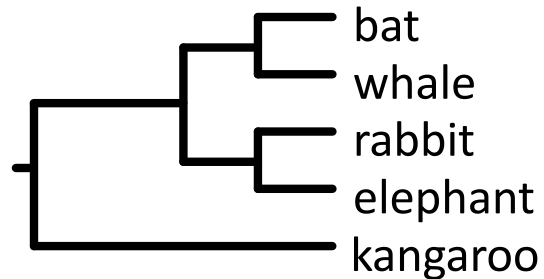
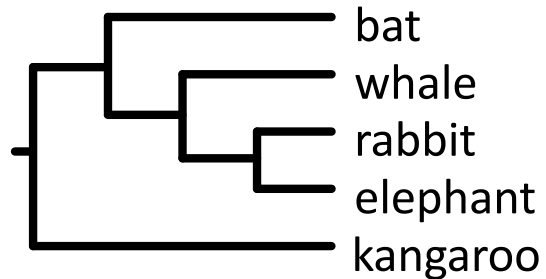
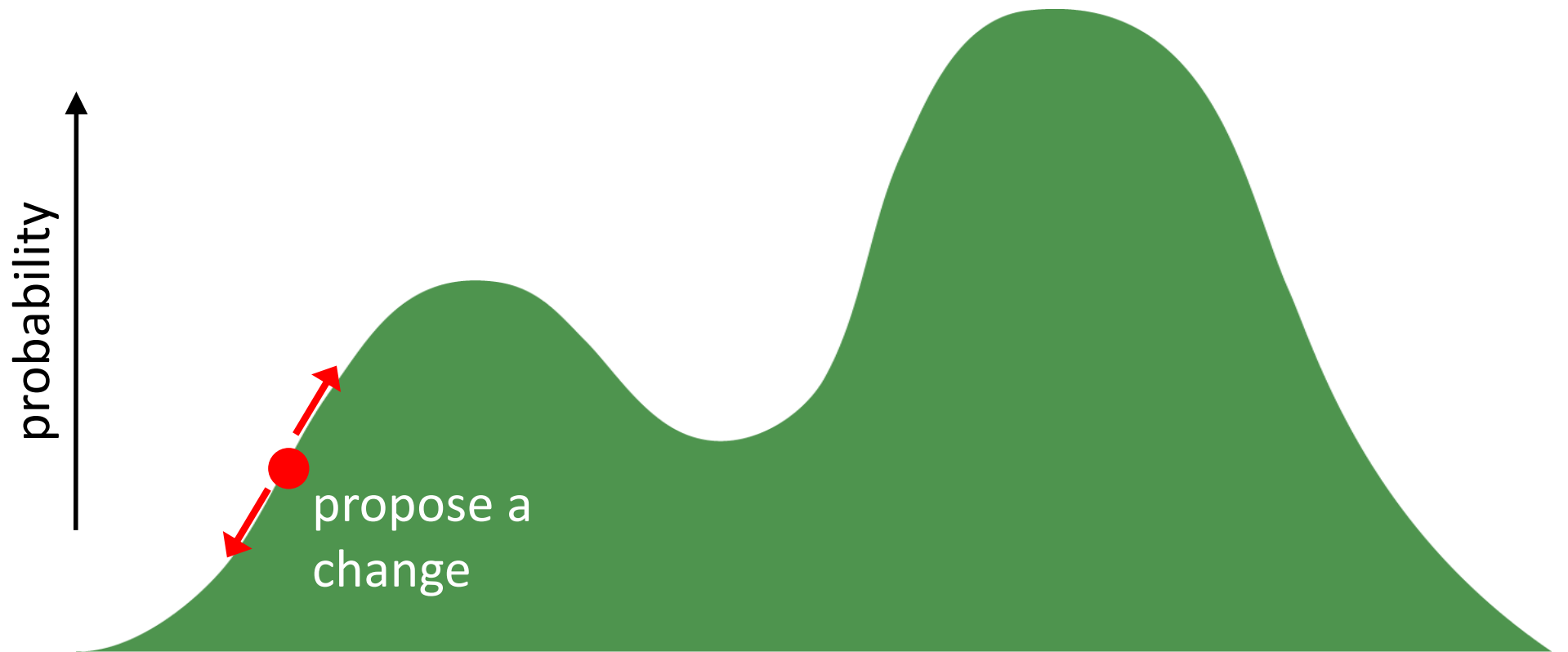
Nicholas Metropolis
Los Alamos, 1953



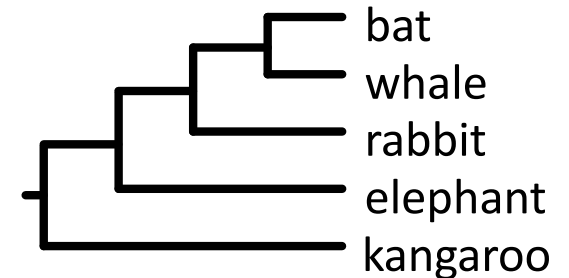
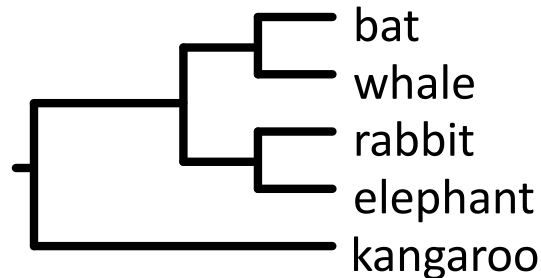
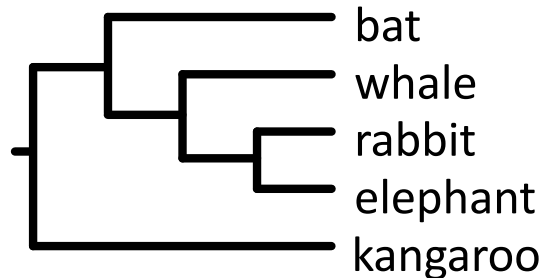
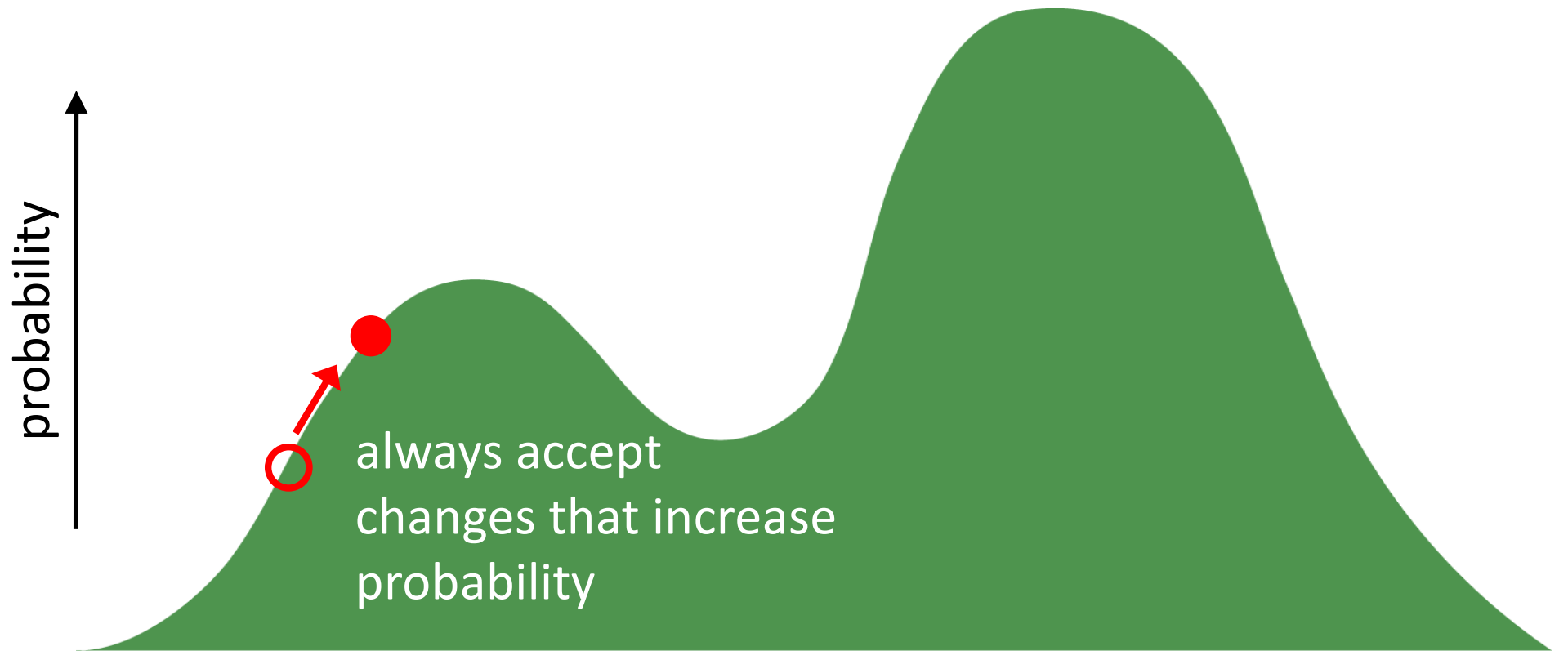
MCMC simulation



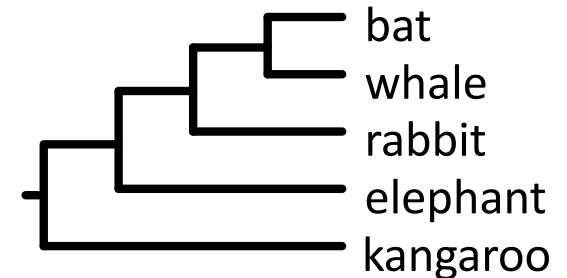
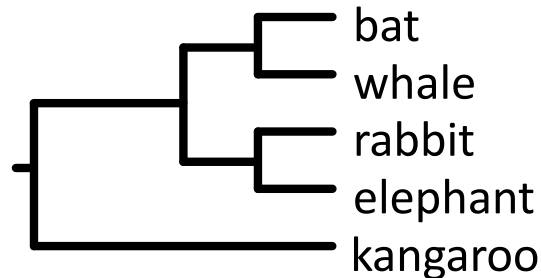
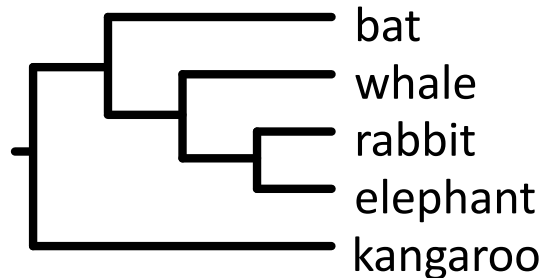
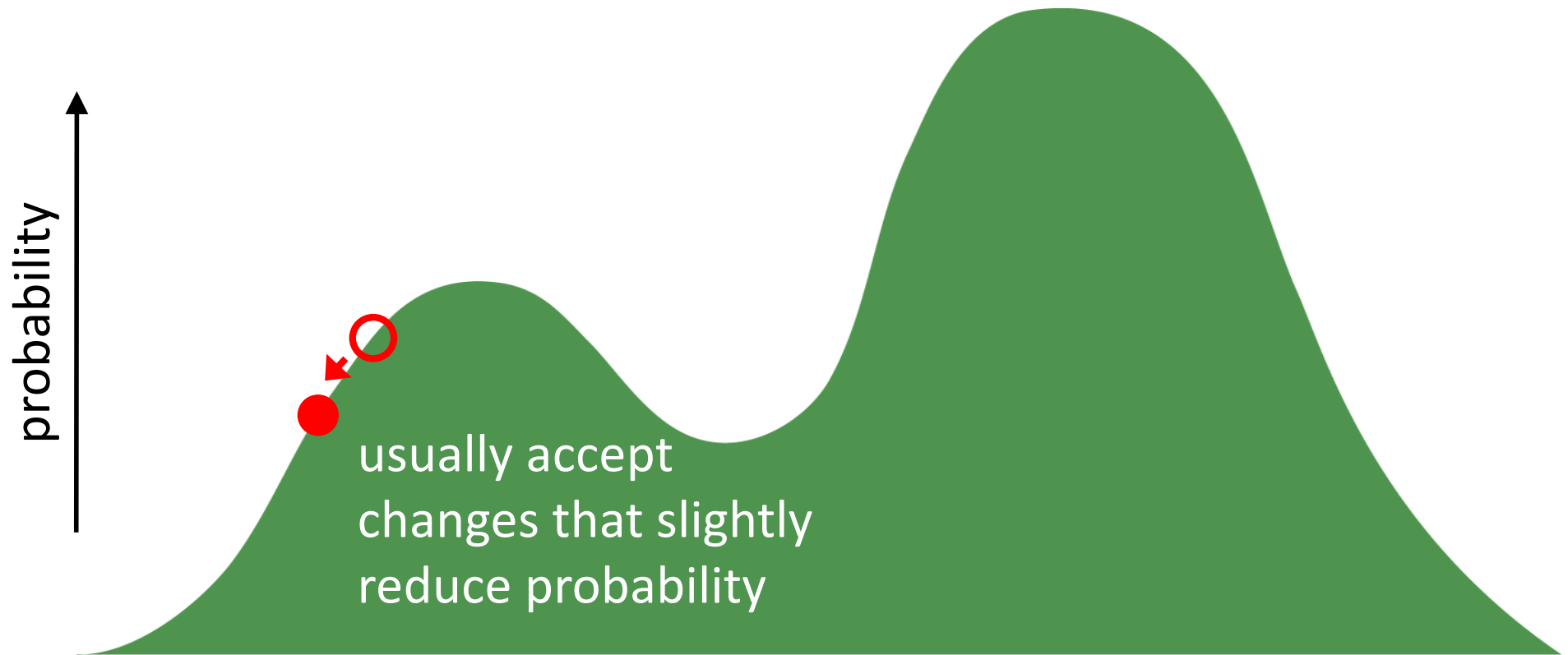
MCMC simulation



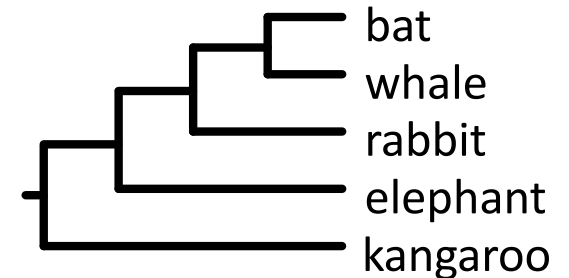
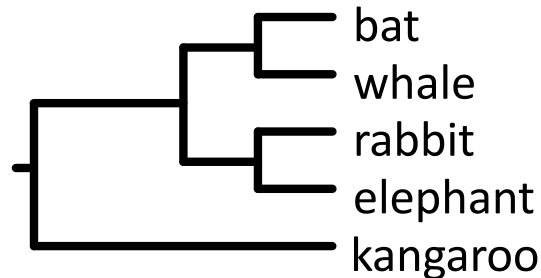
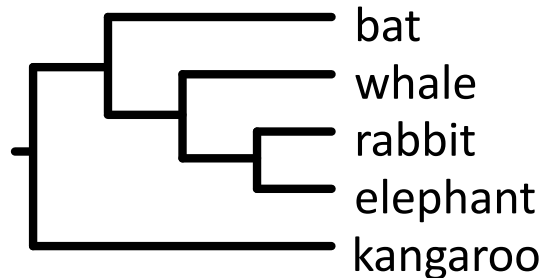
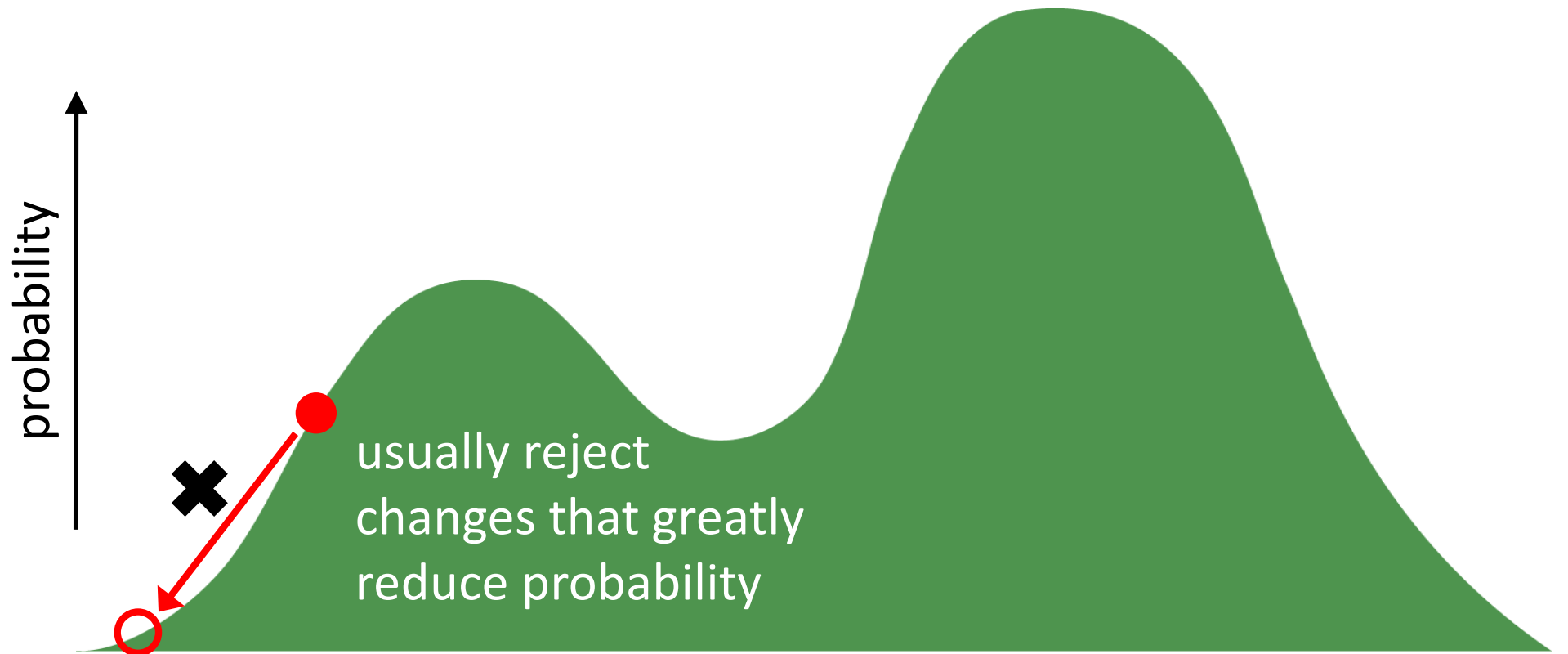
MCMC simulation



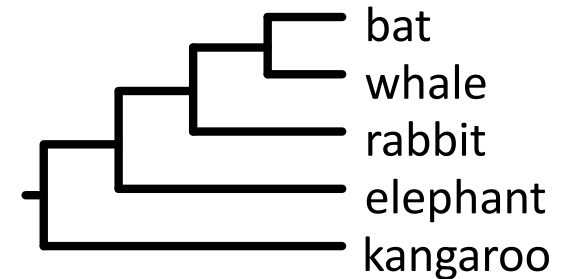
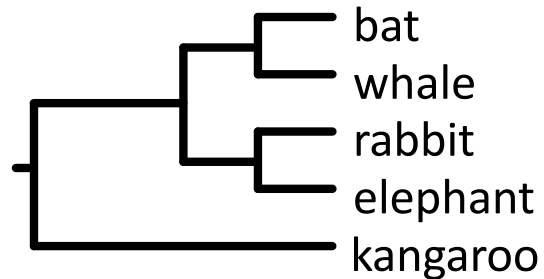
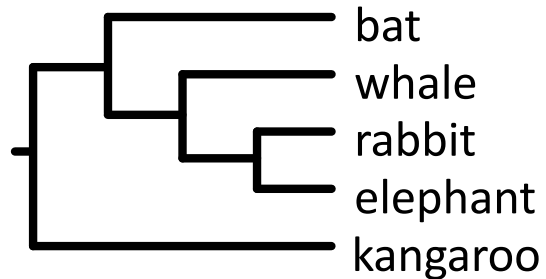
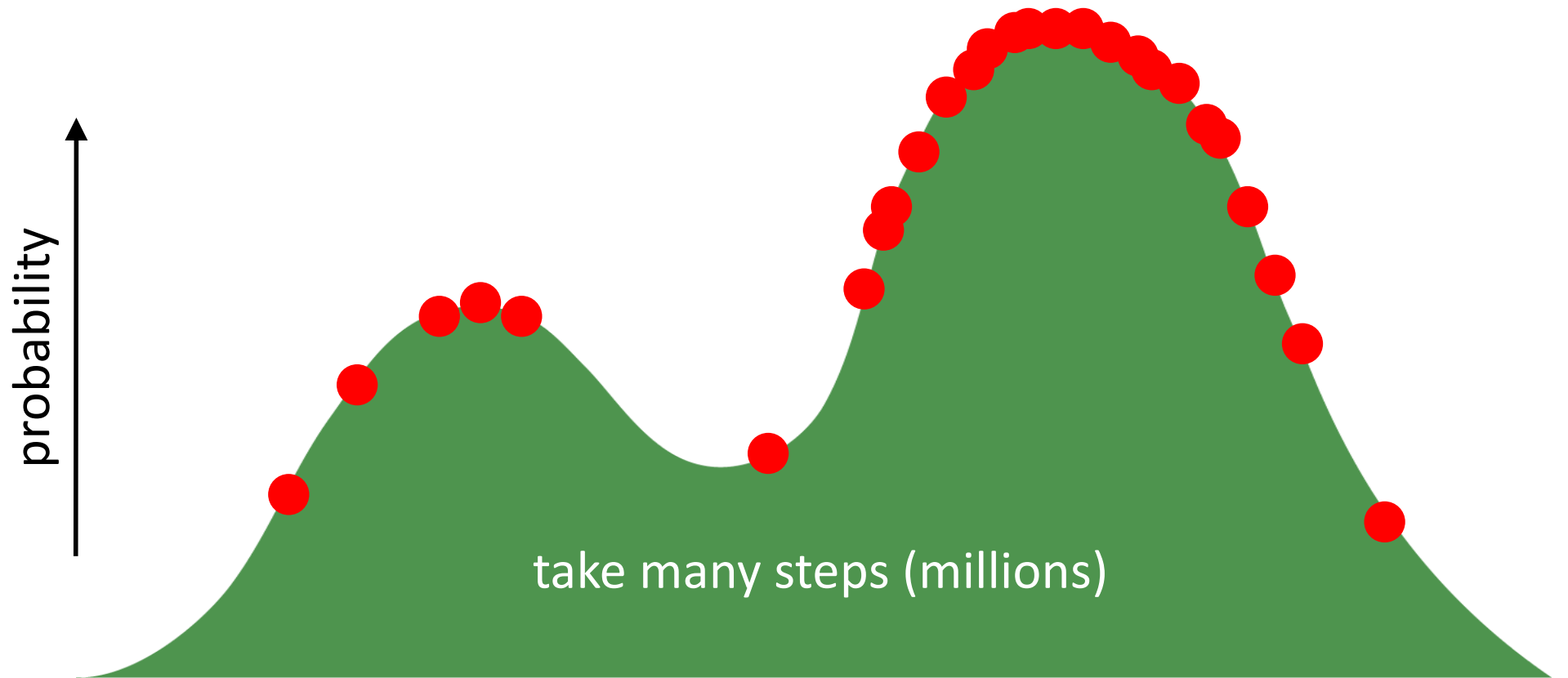
MCMC simulation



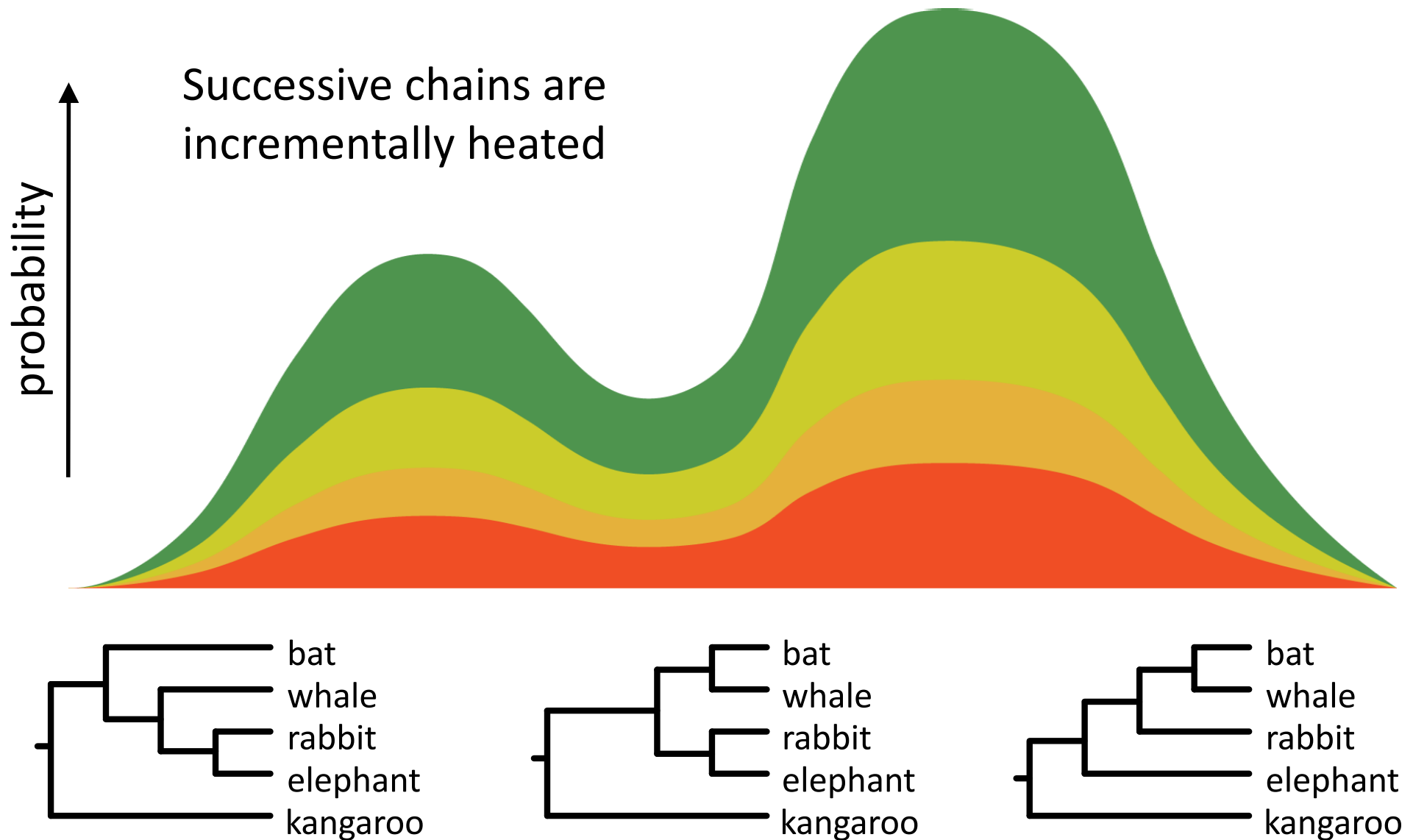
MCMC simulation



MCMC simulation



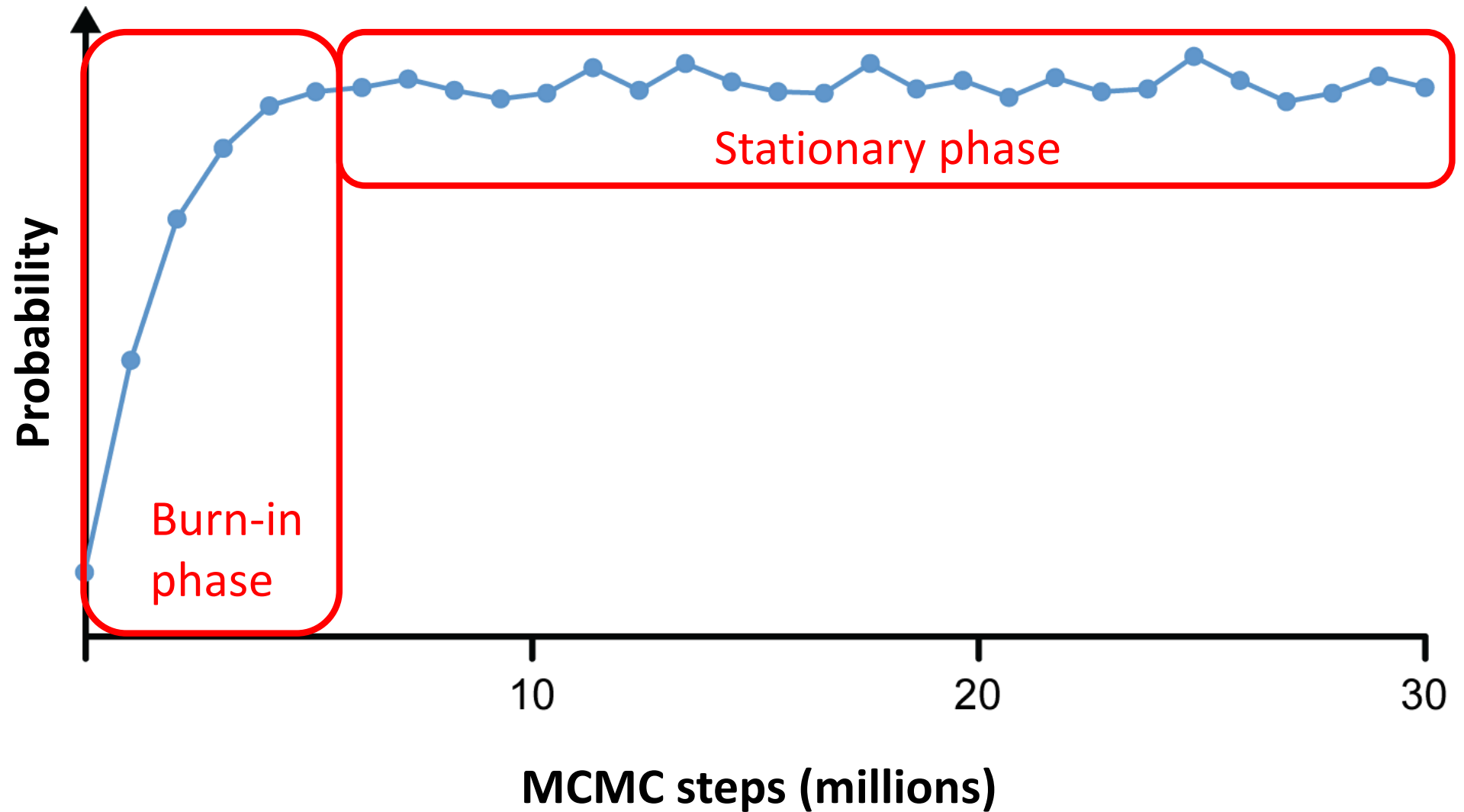
Metropolis-coupled MCMC



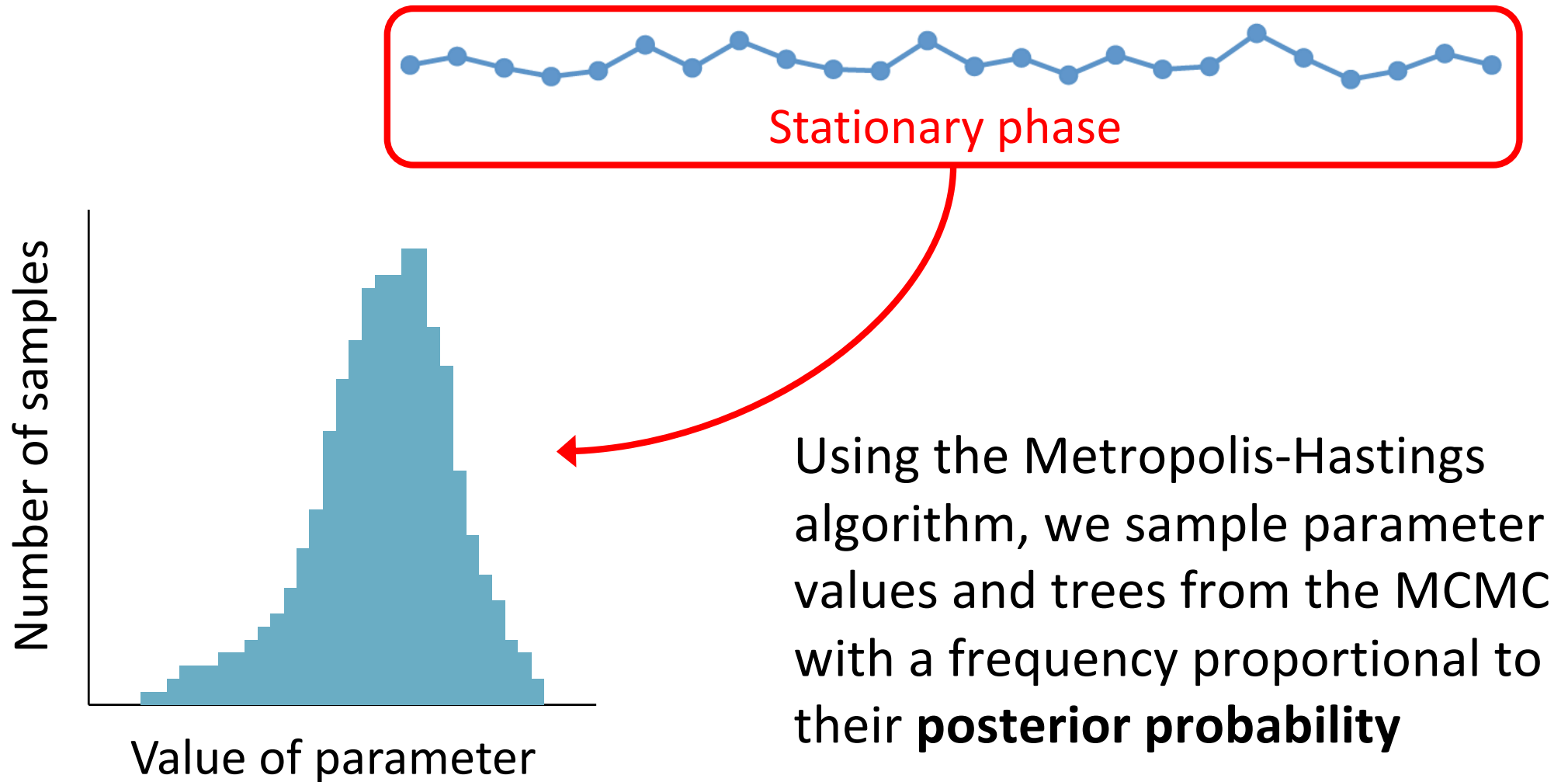
Samples from the MCMC

- Output from a Bayesian phylogenetic analysis:
 - A list of the **parameter values** visited by the Markov chain
(.p file in *MrBayes*, .log file in *BEAST*)
 - A list of the **trees** visited by the Markov chain
(.t file in *MrBayes*, .trees file in *BEAST*)

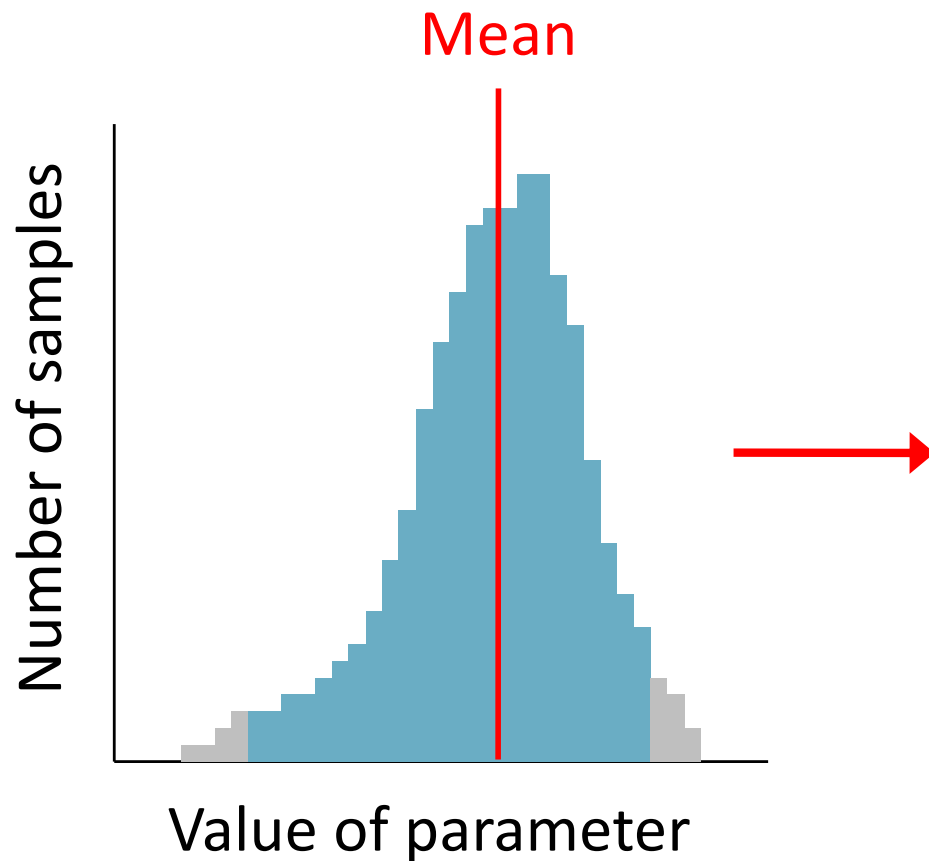
Samples from the MCMC



Samples from the MCMC



Samples from the MCMC



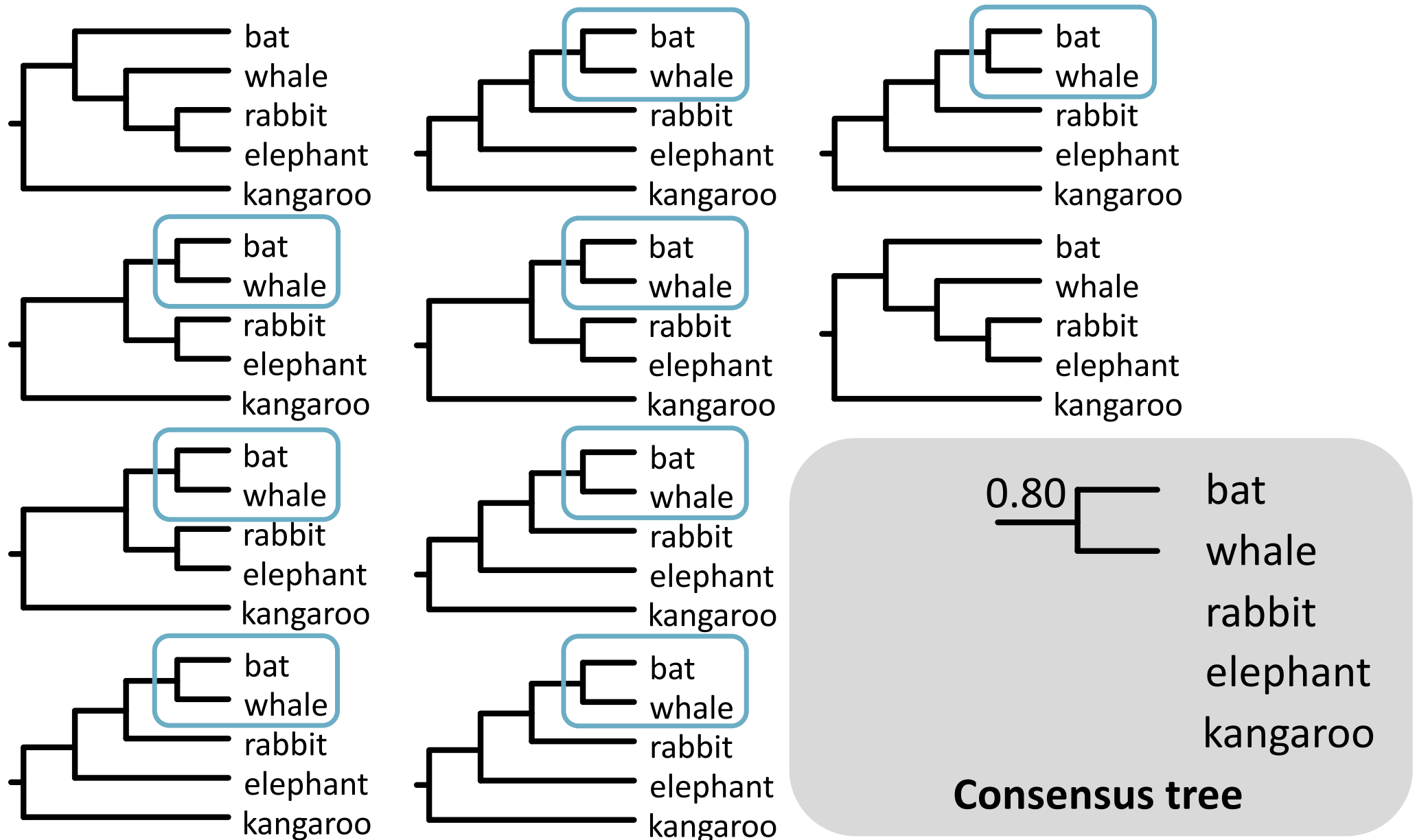
- Take the mean of the sampled values

Mean posterior estimate

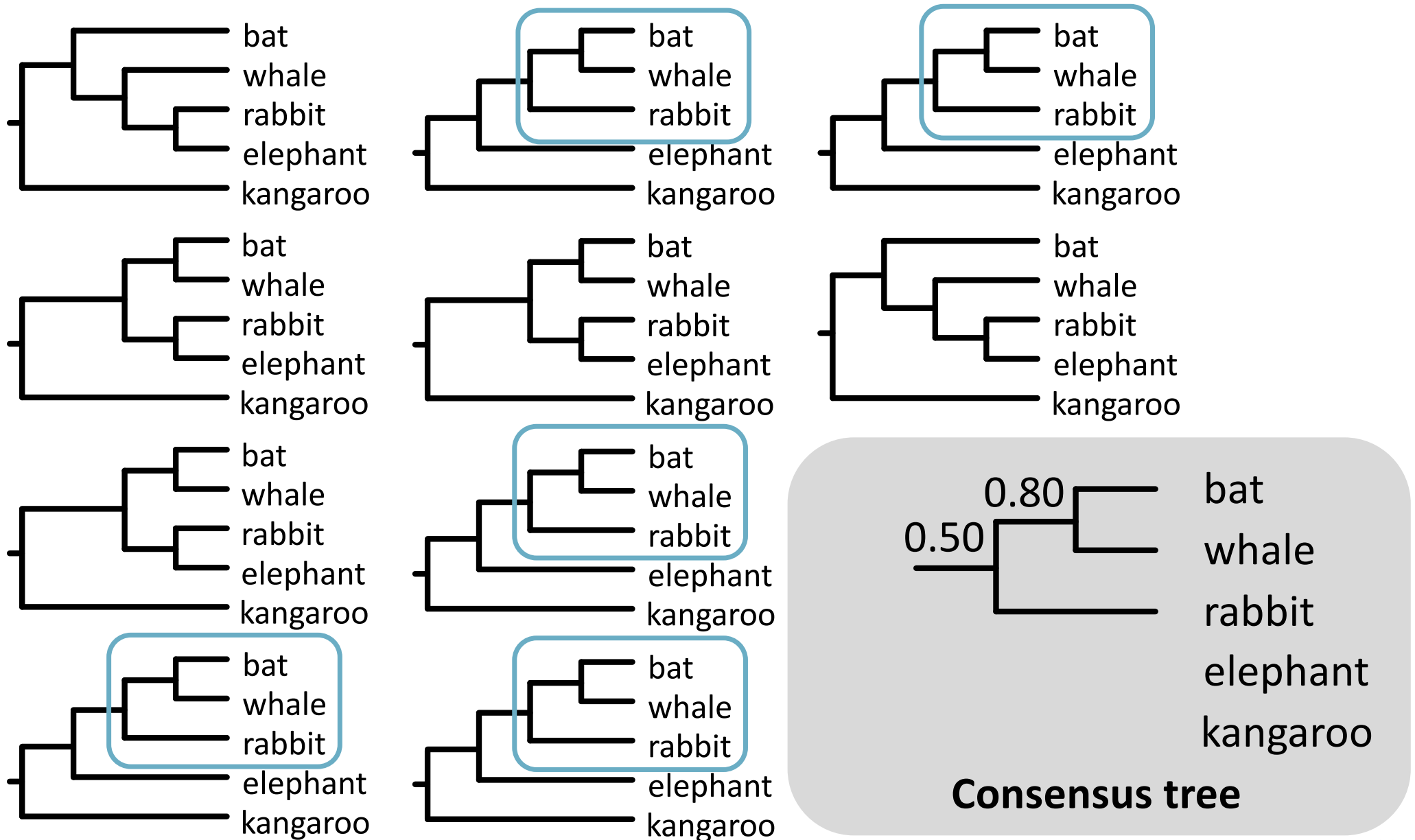
- Take the 'central' 95% of the sampled values

95% credibility interval

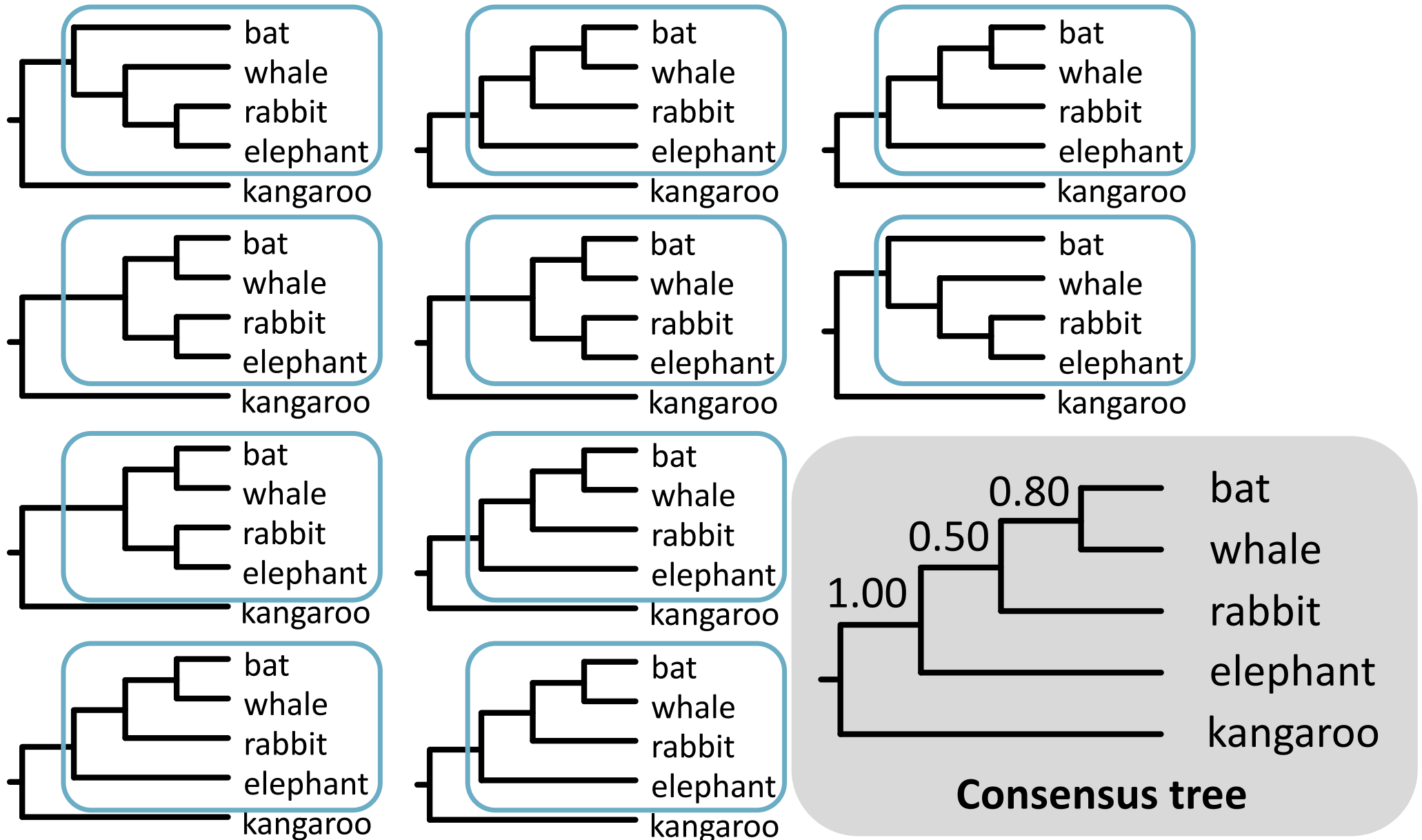
Samples from the MCMC



Samples from the MCMC



Samples from the MCMC



Samples from the MCMC

- **Majority-rule consensus tree (*MrBayes*)**
Shows all nodes with posterior probability >0.50
- **Maximum a posteriori (MAP) tree**
Sampled tree with highest posterior probability
- **Maximum clade credibility (MCC) tree (*BEAST/TreeAnnotator*)**
Sampled tree with highest sum or product of posterior node probabilities

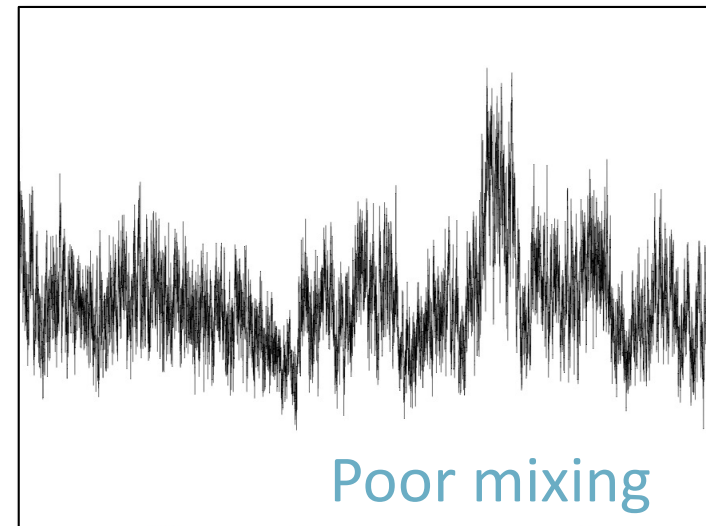
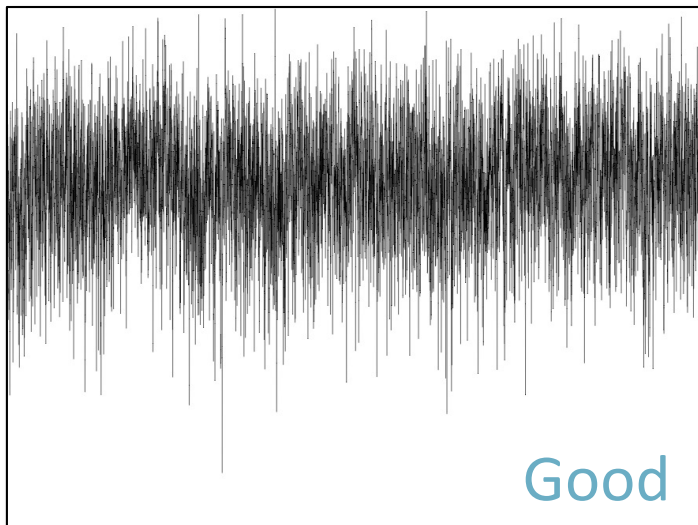
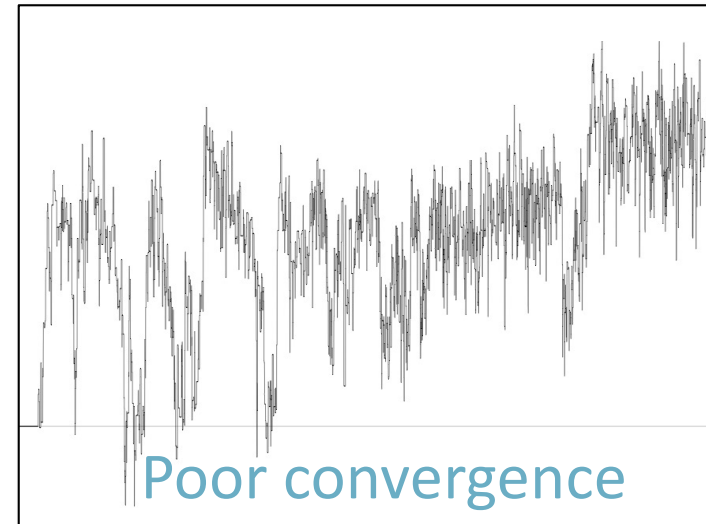
Diagnostics

1. Convergence

Are we drawing samples from the stationary distribution?

2. Sufficient sampling

Have we drawn enough samples to allow a reliable estimate of the posterior distribution?



Convergence

- Run at least 2 (preferably more) independent chains
- Likelihoods should be similar
- Estimates of model parameters should be similar

Sufficient sampling

- **Effective sample size (ESS)**

Have we drawn enough independent samples to produce a reliable estimate of the posterior distribution?

- ESS is preferably **>200** for each parameter
- ESS can be increased by:
 - Increasing the length of the MCMC
(and decreasing the frequency of sampling)
 - Modifying the MCMC proposals

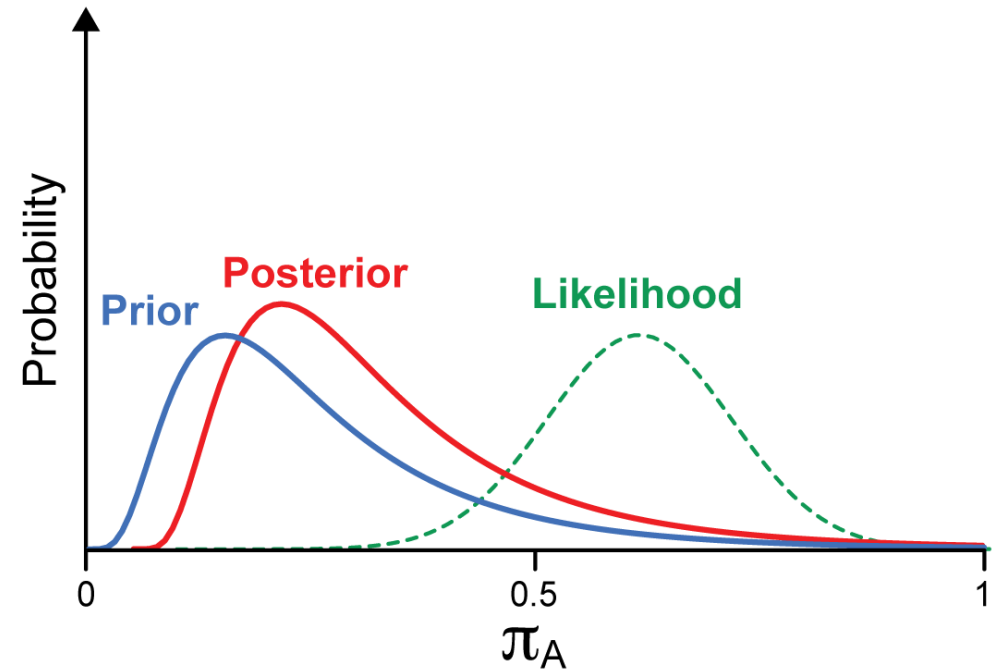
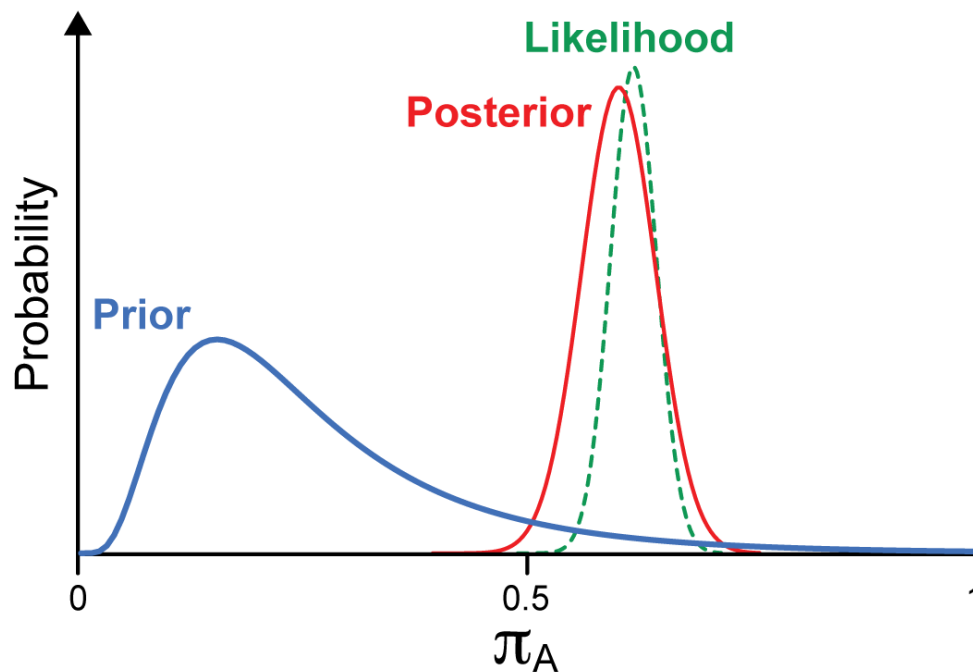
Advantages and Problems

Advantages

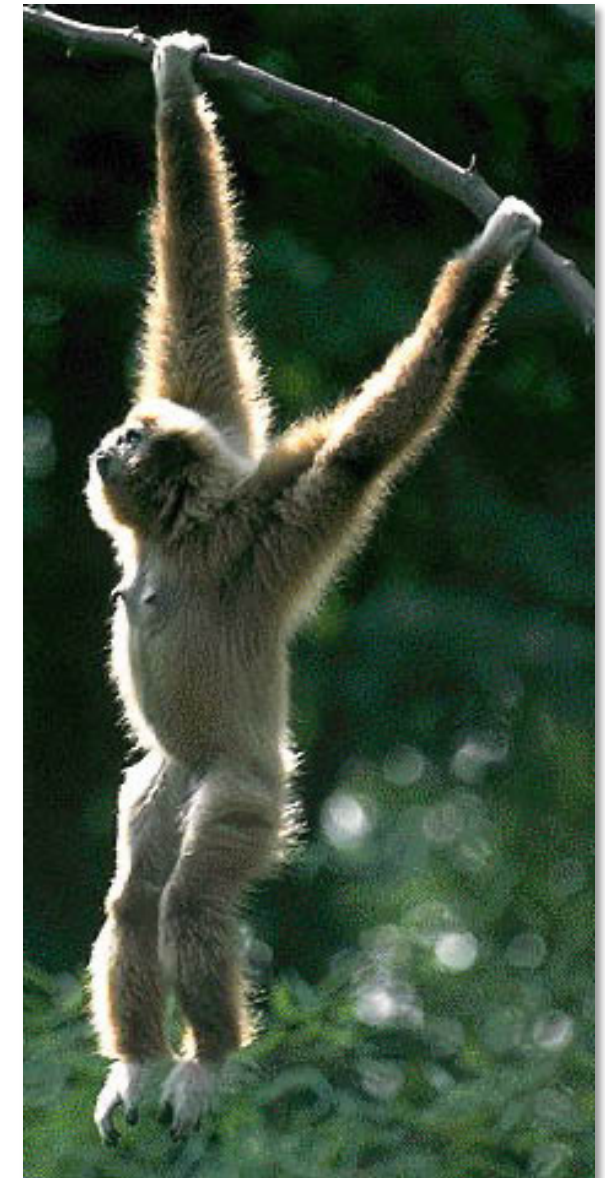
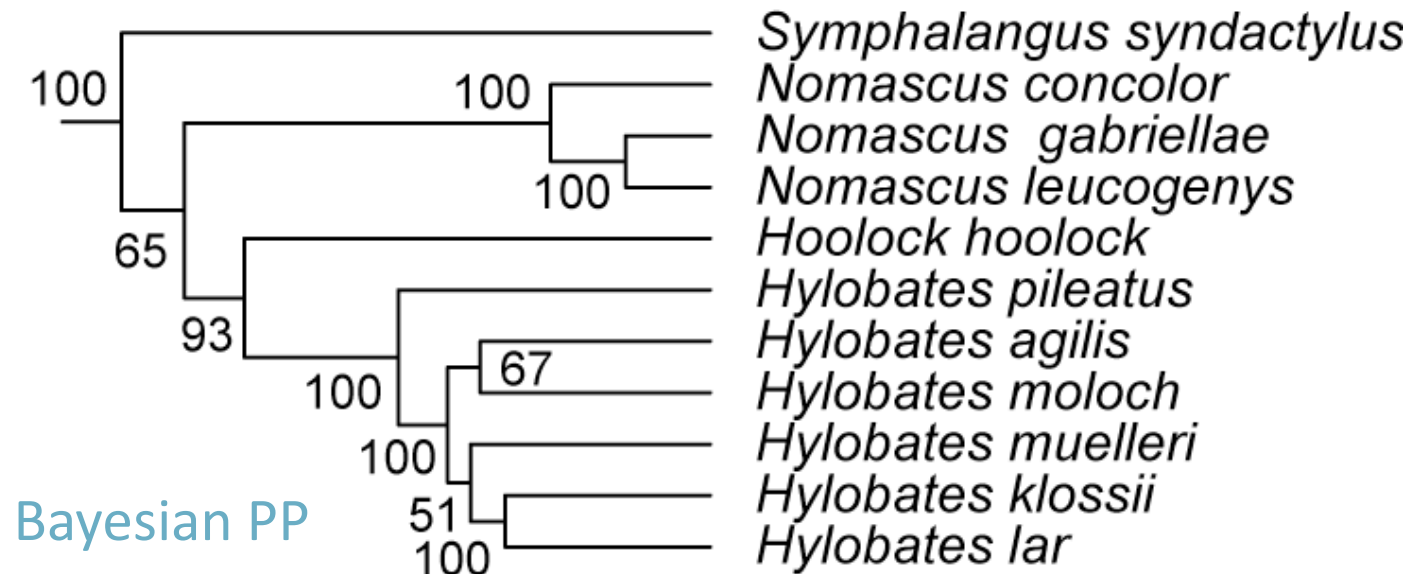
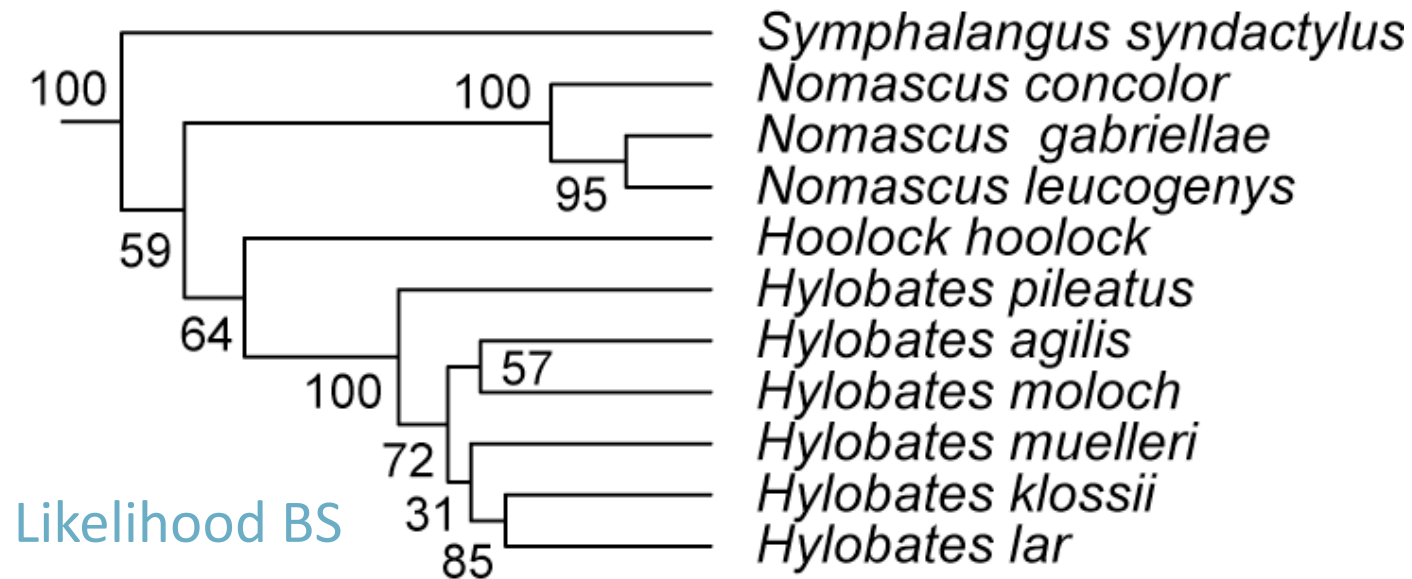
- Able to implement complex (highly parameterised) models
- Estimating node support is straightforward
- Posterior probabilities have an intuitive interpretation
- Can incorporate independent information (in the prior)
- Integrate over ‘nuisance’ parameters

Influence of priors

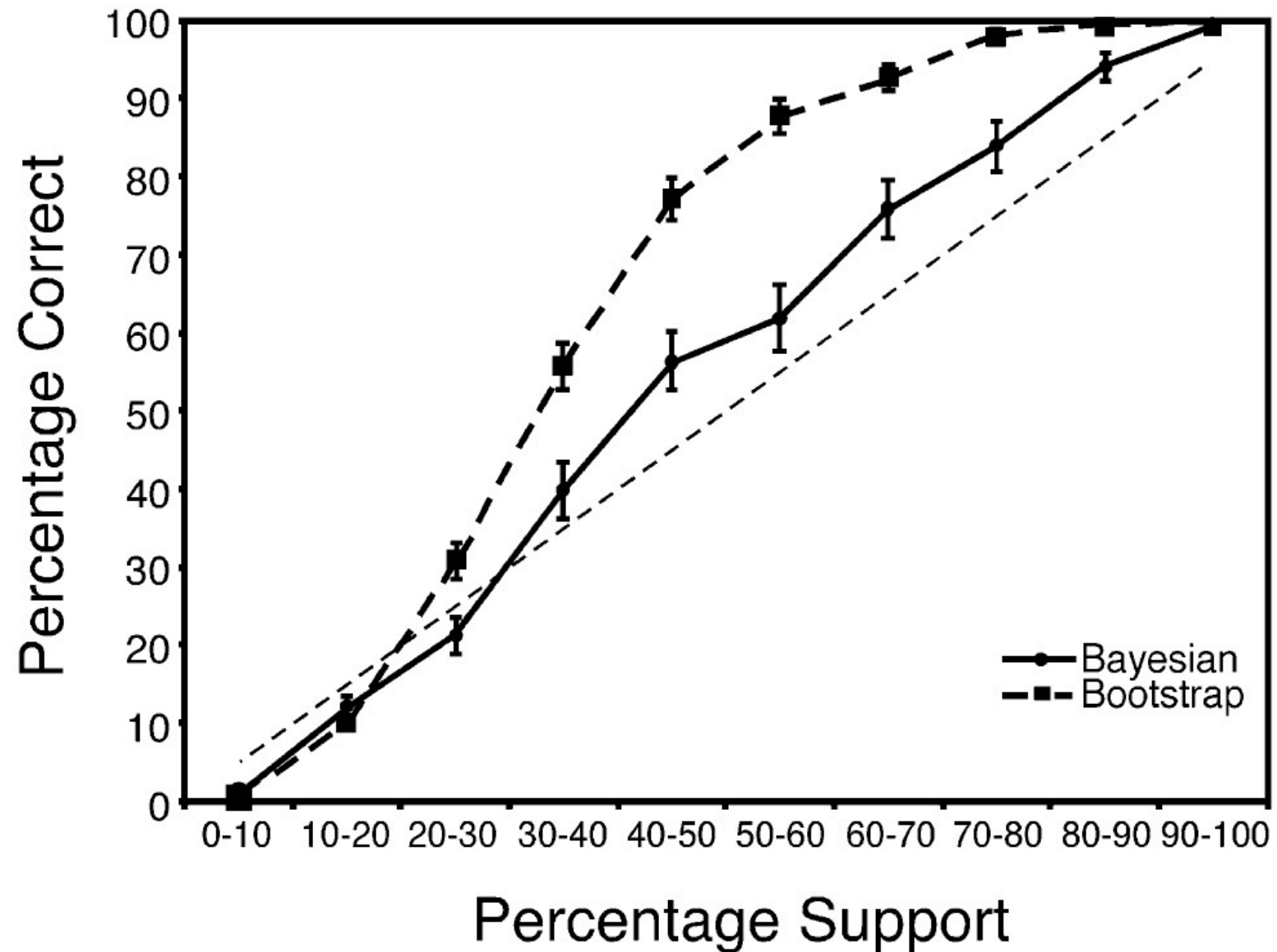
- Sensitivity of the posterior to the prior
- This problem can occur if the data are uninformative, the prior is strong, or both



Node support



Node support



Software for Bayesian phylogenetics

MrBayes



BEAST 1

RevBayes



BEAST 2

Taming the BEAST
workshops



Useful references

