

# **Lecture 2.2:**

# **Bayesian Phylogenetics**

# Phylogenetic methods

---

	<b>Algorithm-based</b>	<b>Optimality criterion</b>	<b>Other</b>
No explicit evolutionary model		Maximum parsimony	
Explicit evolutionary model	Distance-based methods	Maximum likelihood	Bayesian inference

# Maximum parsimony

---

- Identifies the tree that requires fewest evolutionary steps to explain the data set
- **Advantages**
  - Readily used for morphological and molecular data
  - Relatively quick
  - Does not require a lot of assumptions
- **Disadvantages**
  - Susceptible to long-branch attraction
  - Unable to estimate evolutionary rates and timescales

# Distance-based methods

---

- Uses a matrix of pairwise distances to infer the phylogeny
- **Advantages**
  - Can use various types of distance data
  - Extremely quick
  - Uses model to account for multiple changes
- **Disadvantages**
  - Information is lost in pairwise comparisons
  - Unable to implement complex evolutionary models

# Maximum likelihood

---

- Identifies the tree and model parameters that yield the highest likelihood score
- **Advantages**
  - Statistically rigorous
  - Uses model to account for multiple changes
- **Disadvantages**
  - Relatively slow (although new methods can be very quick)
  - Unable to implement complex, parameter-rich models
  - Bootstrapping can be slow

# Bayesian phylogenetic analysis

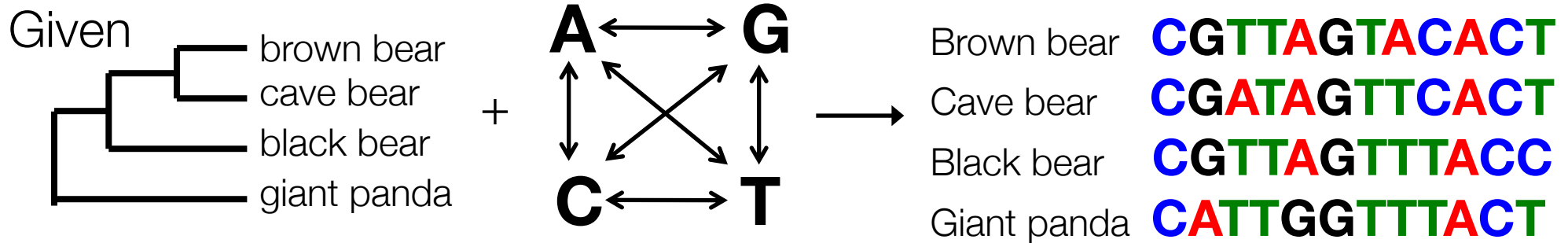
---

- Bayesian phylogenetic analysis was developed in the mid 1990s
- Now one of the most widely used methods
- Software:
  - MrBayes
  - BEAST

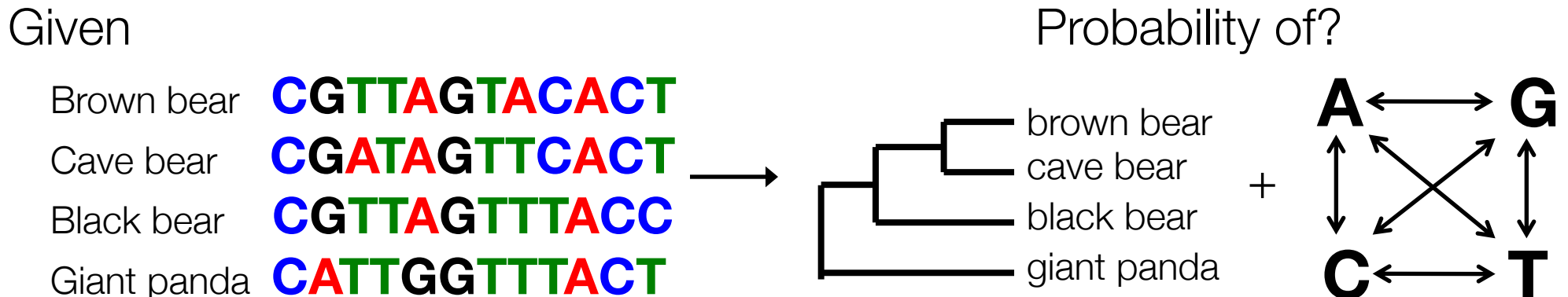


# Bayesian *versus* likelihood

## Maximum likelihood



## Bayesian inference



# The Bayesian paradigm

---

- 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 with the likelihood to yield the **posterior distribution**



# Bayesian inference

---

$$\boxed{\text{Posterior}} \propto \boxed{\text{Prior}} \times \boxed{\text{Likelihood}}$$

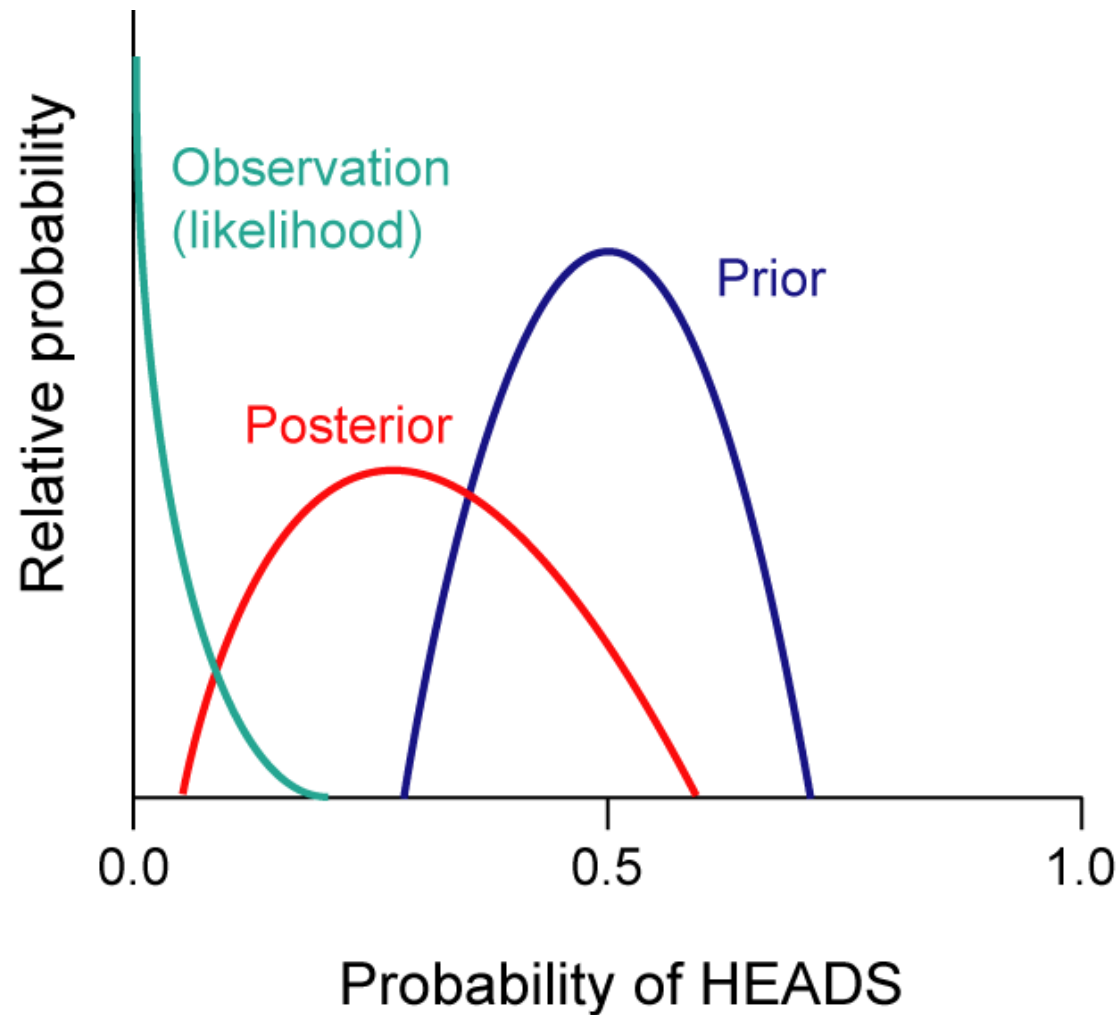
This is what we want  
to estimate

Specified by user,  
Independent of data

Calculated from data

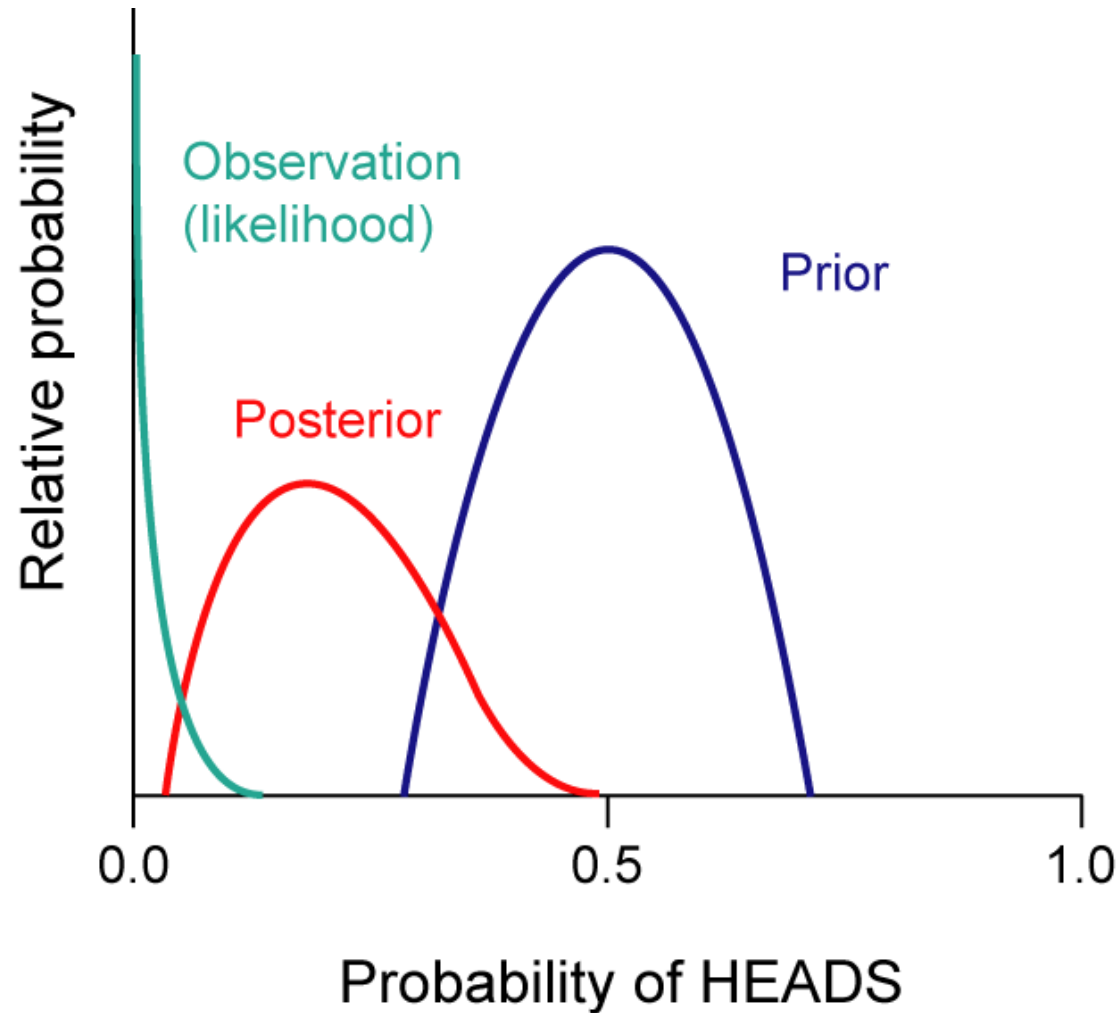
# Coin toss example

---



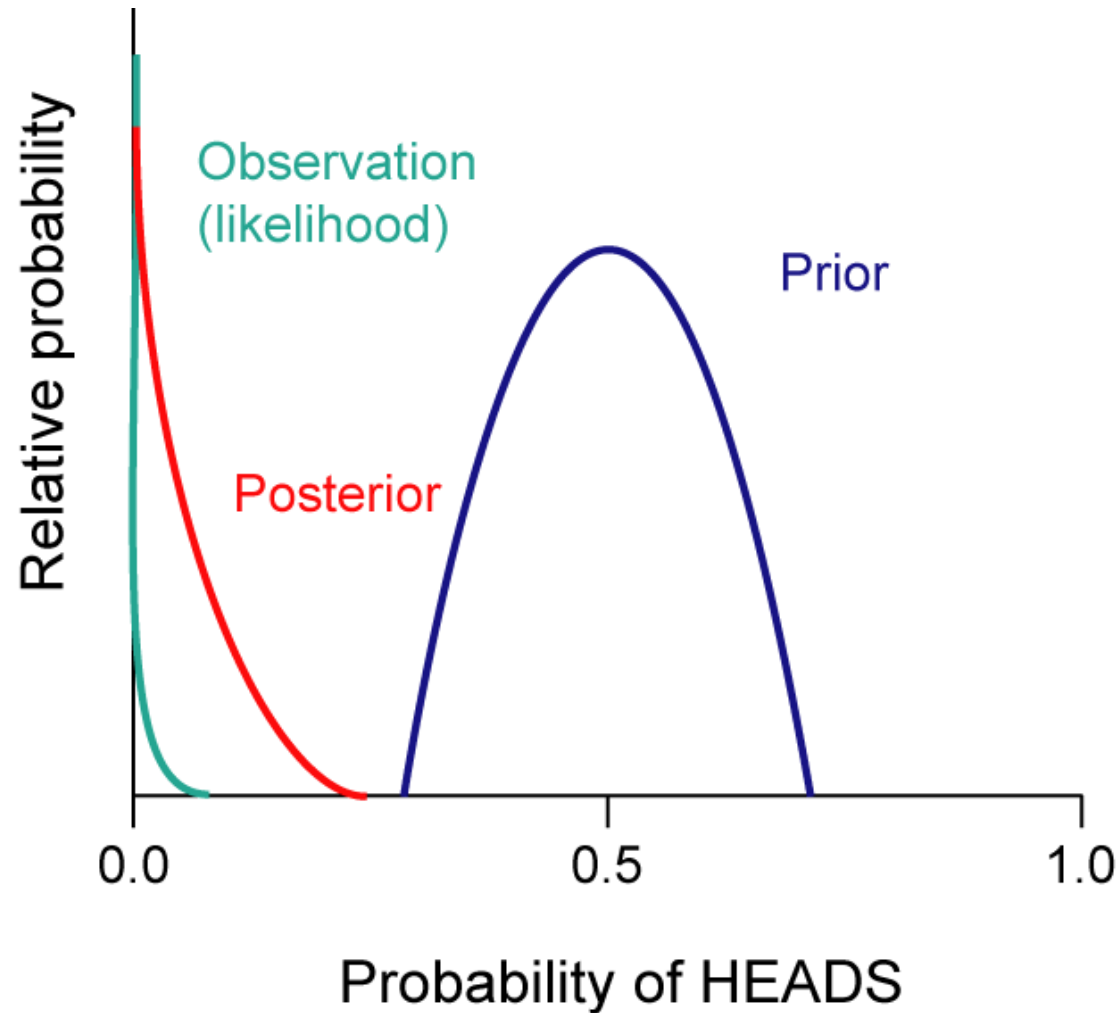
# Coin toss example

---



# Coin toss example

---



# Questions?

---

# Priors

---

- Reflect our prior expectations (and uncertainty) about values of parameters (without knowledge of the data)
  1. Past observations
  2. Personal beliefs
  3. Use of a biologically realistic model

# Priors

---

- Priors are chosen in the form of probability distributions
  - Example: substitution rate  
Probably around  $3.2 \times 10^{-8}$   $\rightarrow$   $Normal(3.2 \times 10^{-8}, \sigma)$
- If unsure, use **uninformative/flat prior**

# Prior for the tree

---

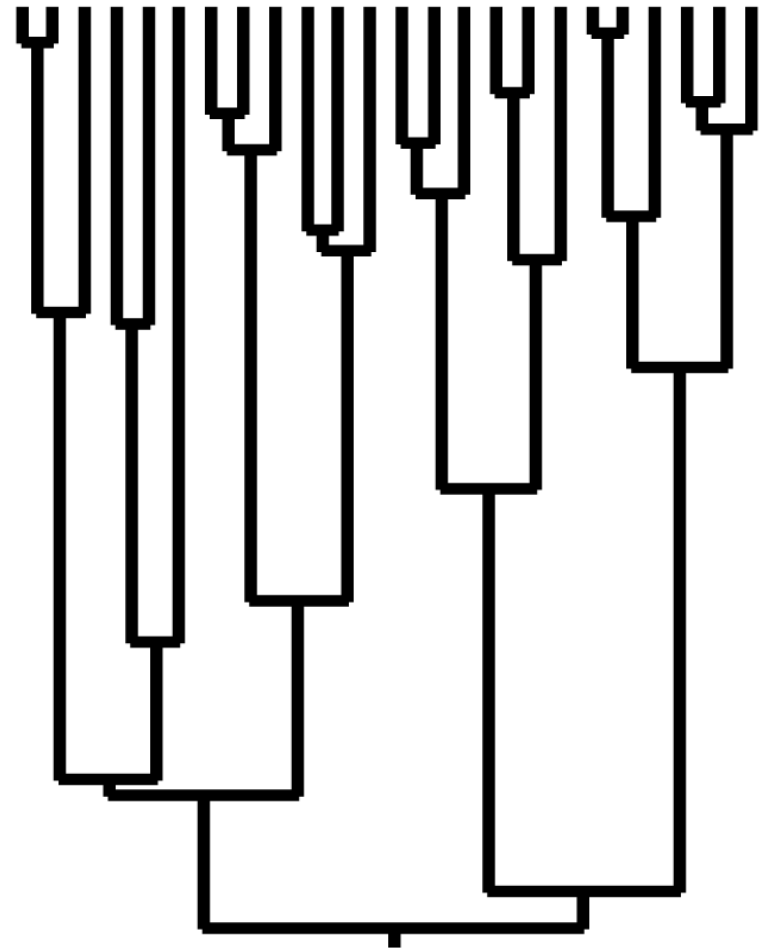
1. Use a flat prior for tree topology (*MrBayes*)
  - All trees have equal probability
  - Also need a prior for branch lengths or node times
2. Use a biological model to generate prior distribution (*BEAST*)
  - Among species: speciation model
  - Within species: coalescent model



# Among species

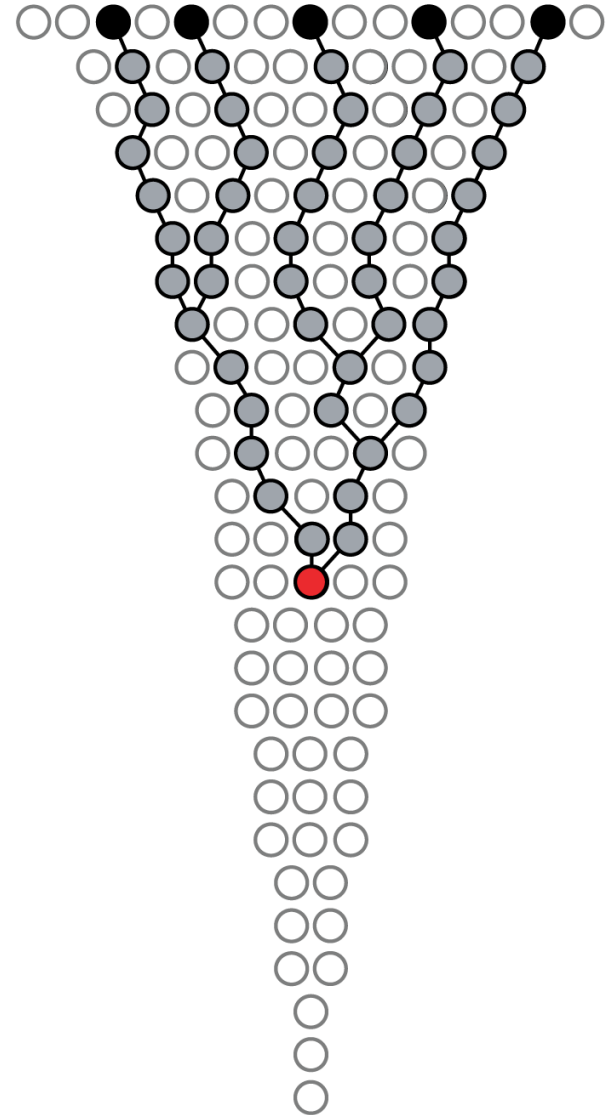
---

- Tree shape described by a stochastic branching process, such as:
- Yule process:
  - The root lineage splits into two
  - Lineages split at a constant rate
  - Simulates speciation process
- Birth-death process



# Within species

- Coalescent model used to put a prior on the tree
- Time between coalescent events depends on population size
- Different demographic models:
  - Constant population
  - Exponential growth
  - Logistic growth



# Questions?

---

# Markov Chain Monte Carlo Sampling

# Estimating the posterior

---

- Impossible to obtain the posterior analytically
- 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*



# Estimating the posterior

---

Likelihood      Prior

$$f(\tau_i | X) = \frac{f(X | \tau_i) f(\tau_i)}{\sum_{j=1}^{B(s)} f(X | \tau_i) f(\tau_i)}$$

Marginal likelihood / normalising constant

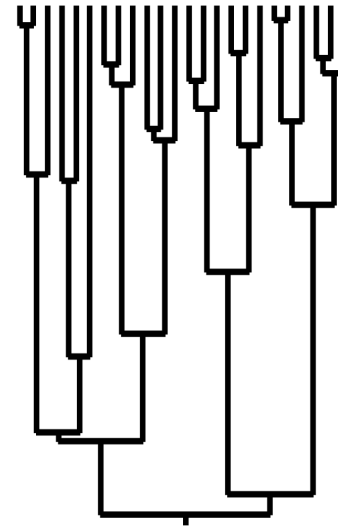
$$f(X | \tau_i) = \int_{v_i} \int_{\theta} f(X | \tau_i, v_i, \theta) f(v_i) f(\theta) dv_i d\theta$$

# Estimating the posterior

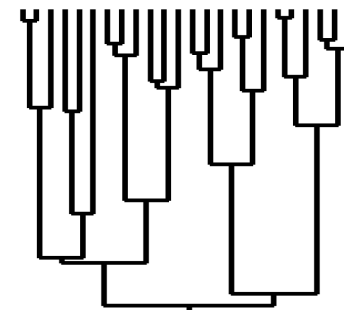
$$R = \frac{Pr(\text{Point}_2|\text{Data})}{Pr(\text{Point}_1|\text{Data})}$$

$$R = \frac{\frac{Pr(\text{Point}_2)L(\text{Point}_2)}{Pr(\text{Data})}}{\frac{Pr(\text{Point}_1)L(\text{Point}_1)}{Pr(\text{Data})}}$$

$$R = \frac{Pr(\text{Point}_2)L(\text{Point}_2)}{Pr(\text{Point}_1)L(\text{Point}_1)}$$



Point 1



Point 2

# Questions?

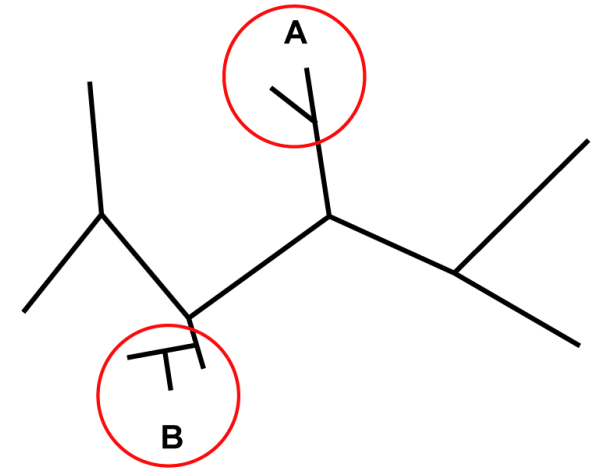
---



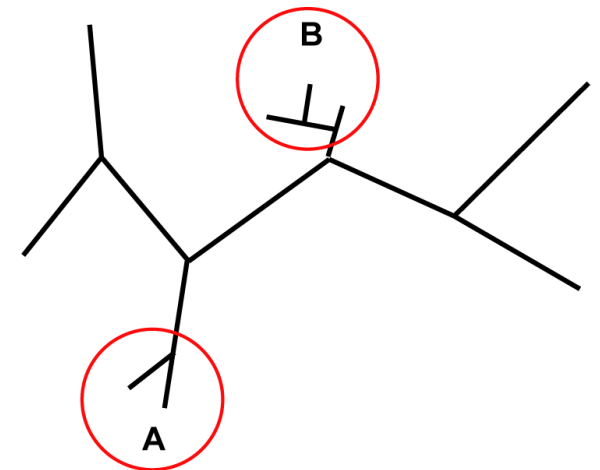
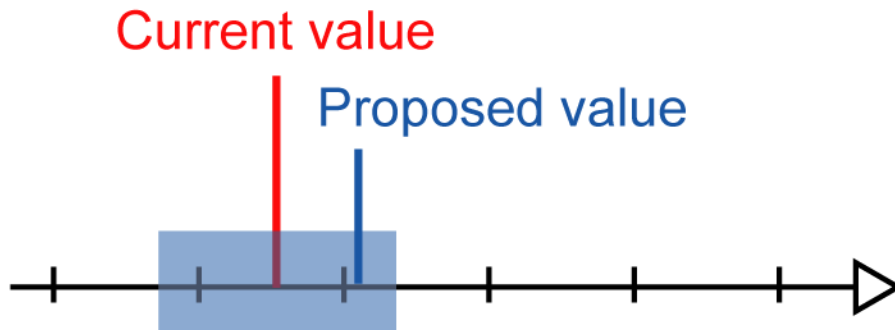
# Metropolis-Hastings algorithm

1. Choose a starting tree and parameter values
2. Calculate prior x likelihood
3. Propose a change to a parameter or to the tree

Tree



Continuous parameters



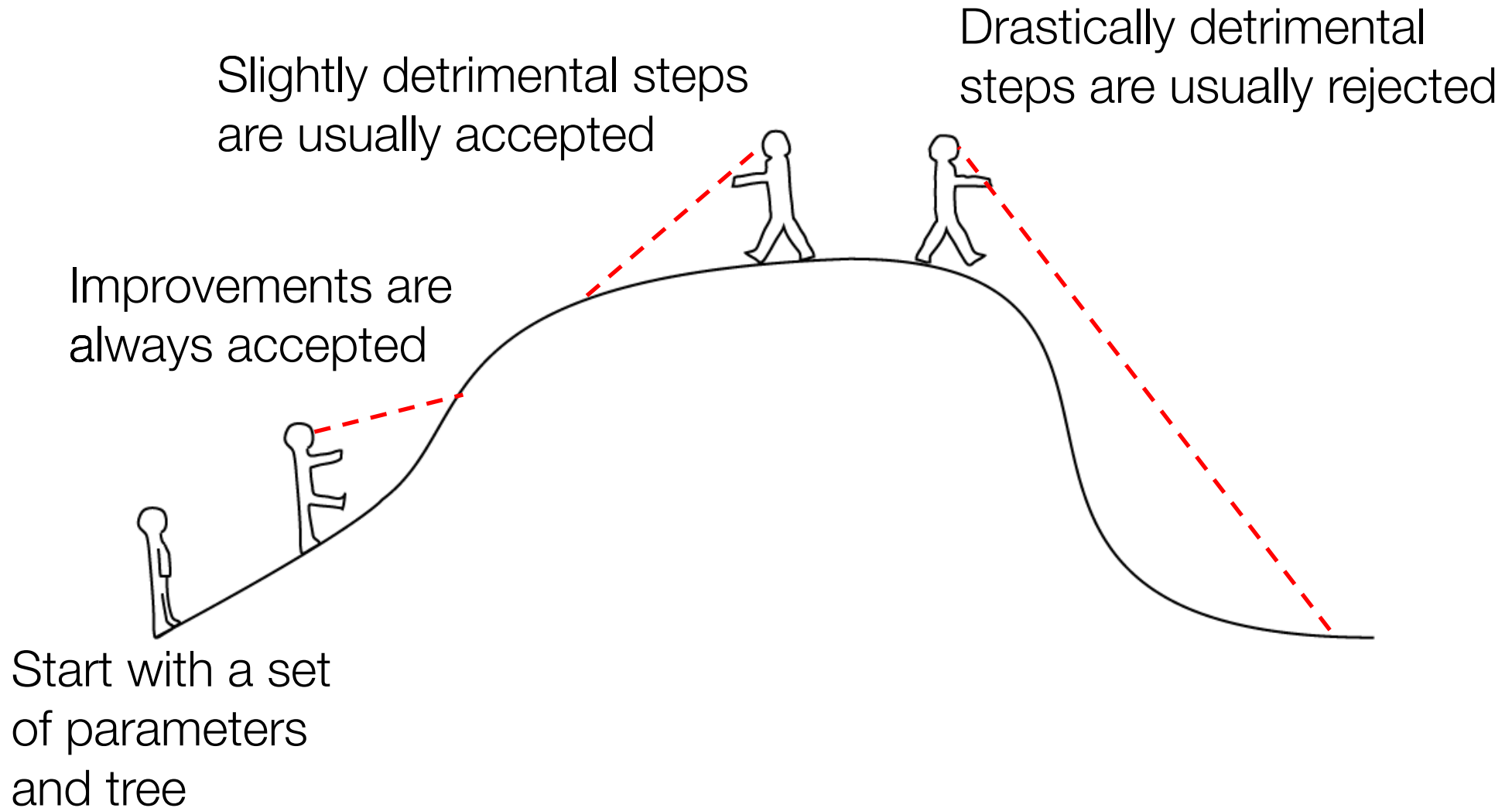
# Metropolis-Hastings algorithm

---

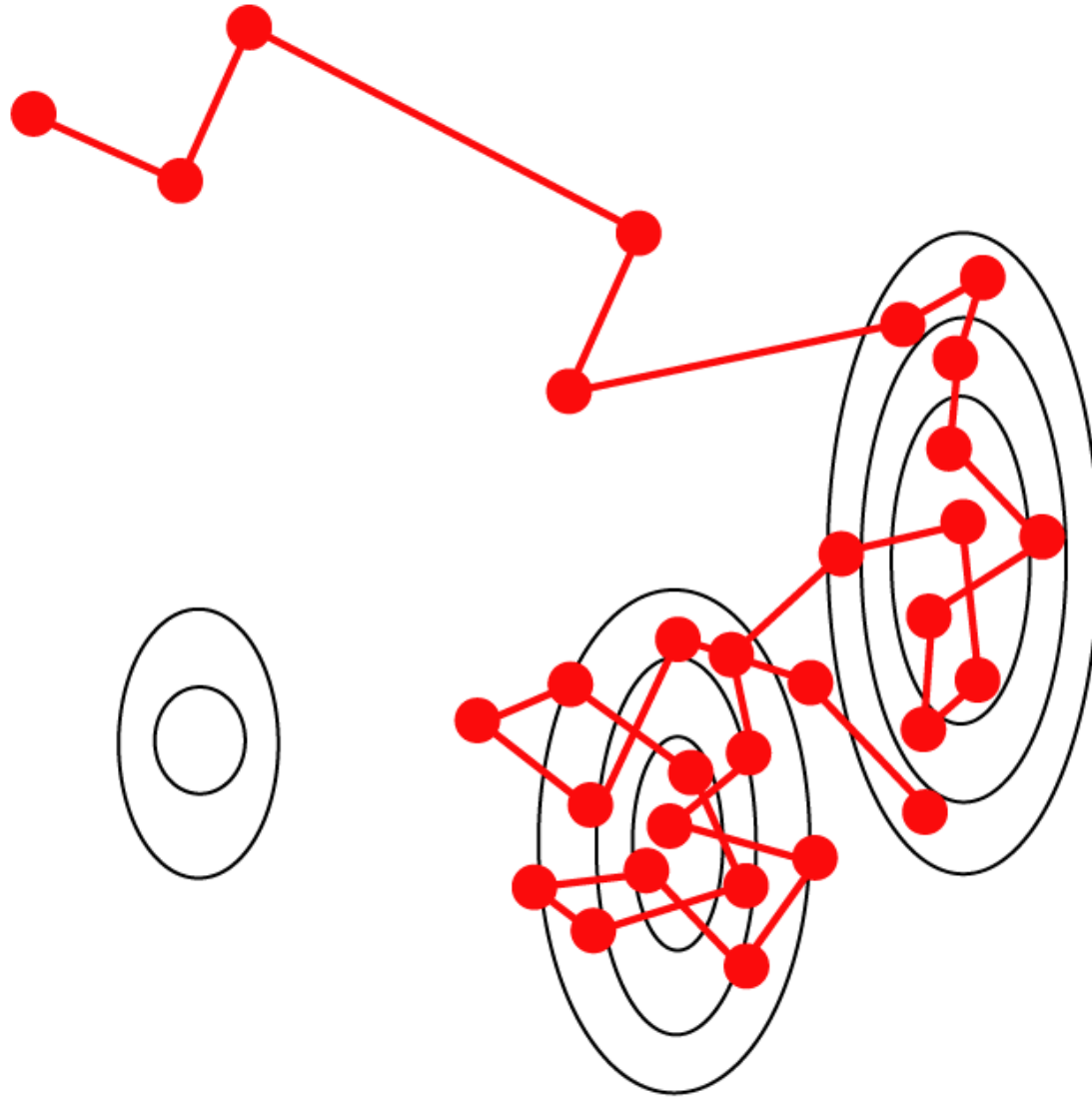
1. Choose a starting tree and parameter values
2. Calculate prior  $\times$  likelihood
3. Propose a change to a parameter or to the tree
4. Two situations:
  2. If proposed location is better, move to the new location
  3. If proposed location is worse, move to the new location with probability equal to ratio of new to old location (“Hastings ratio”)
5. Record the tree and parameter values at each step

# Metropolis-Hastings algorithm

---

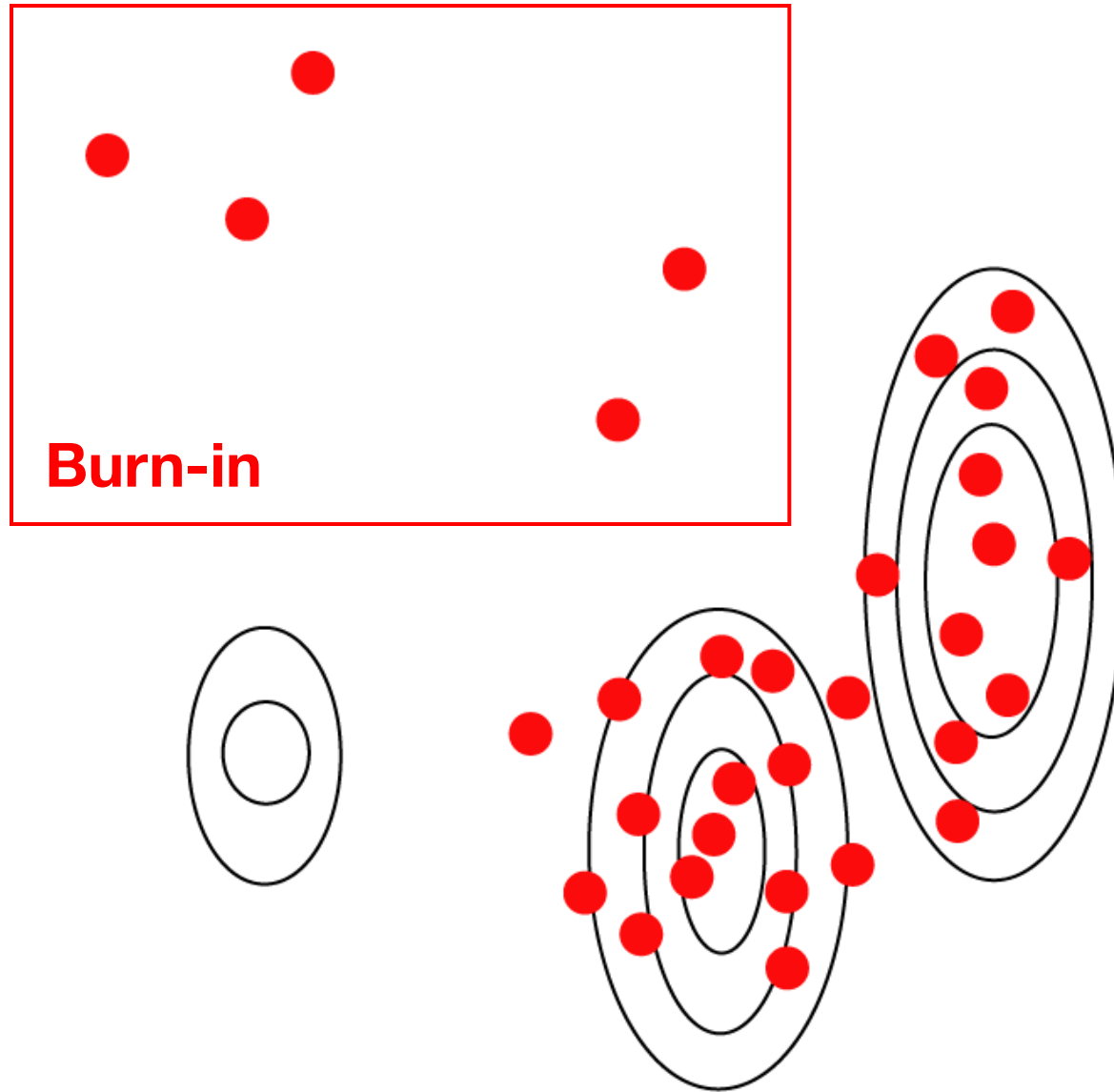


# Metropolis-Hastings algorithm



# Metropolis-Hastings algorithm

---



# Questions?

---

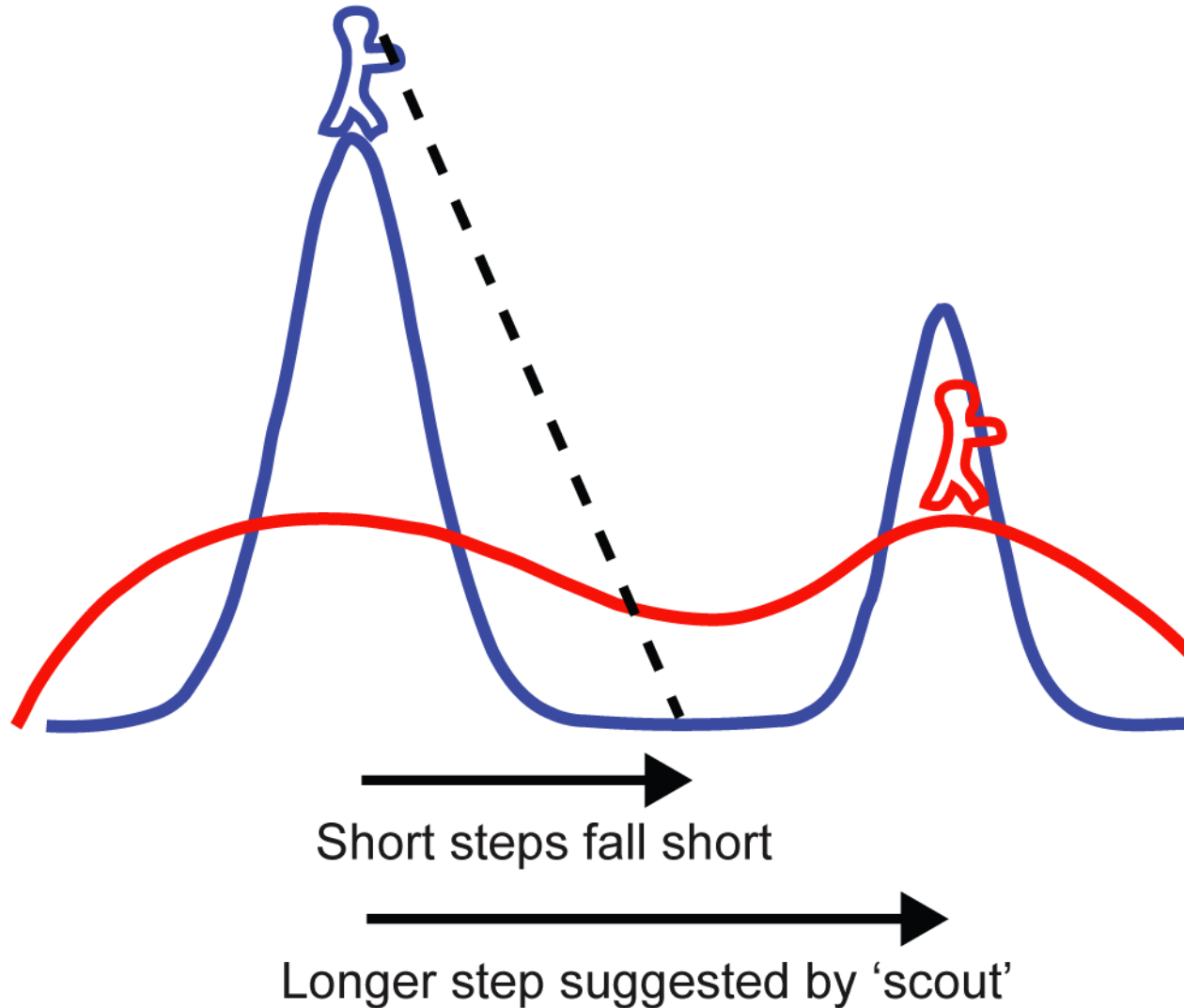
# Metropolis coupling

---

- Sometimes the Metropolis-Hastings algorithm is unable to traverse parameter space efficiently
- This can be addressed by **Metropolis-coupled Markov chain Monte Carlo** (MCMCMC)
- Heated chains are added to the original cold chain
- The heated chains have a higher probability of accepting downhill steps
- Heated chains act as ‘scouts’ for the cold chain

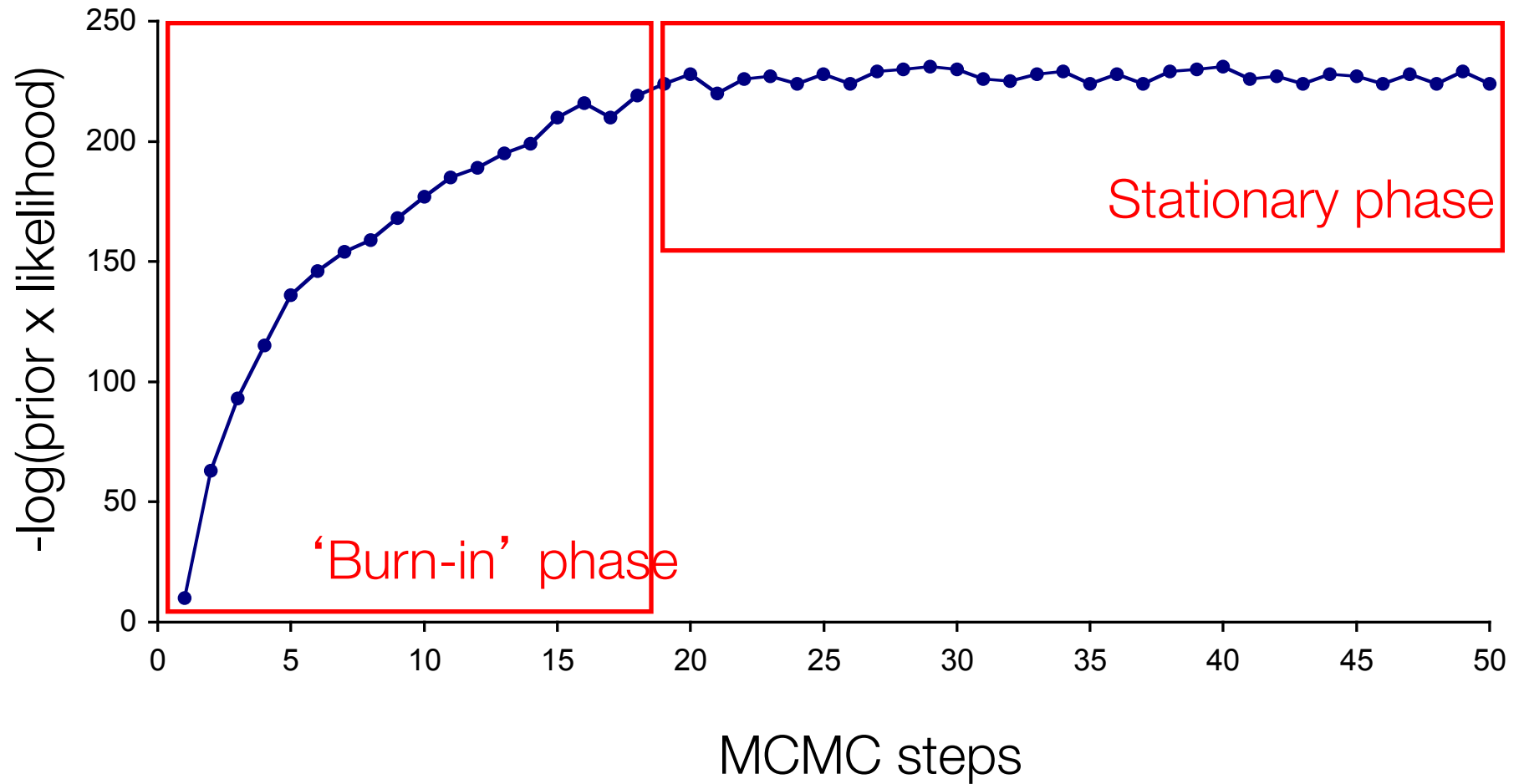
# Metropolis coupling

---

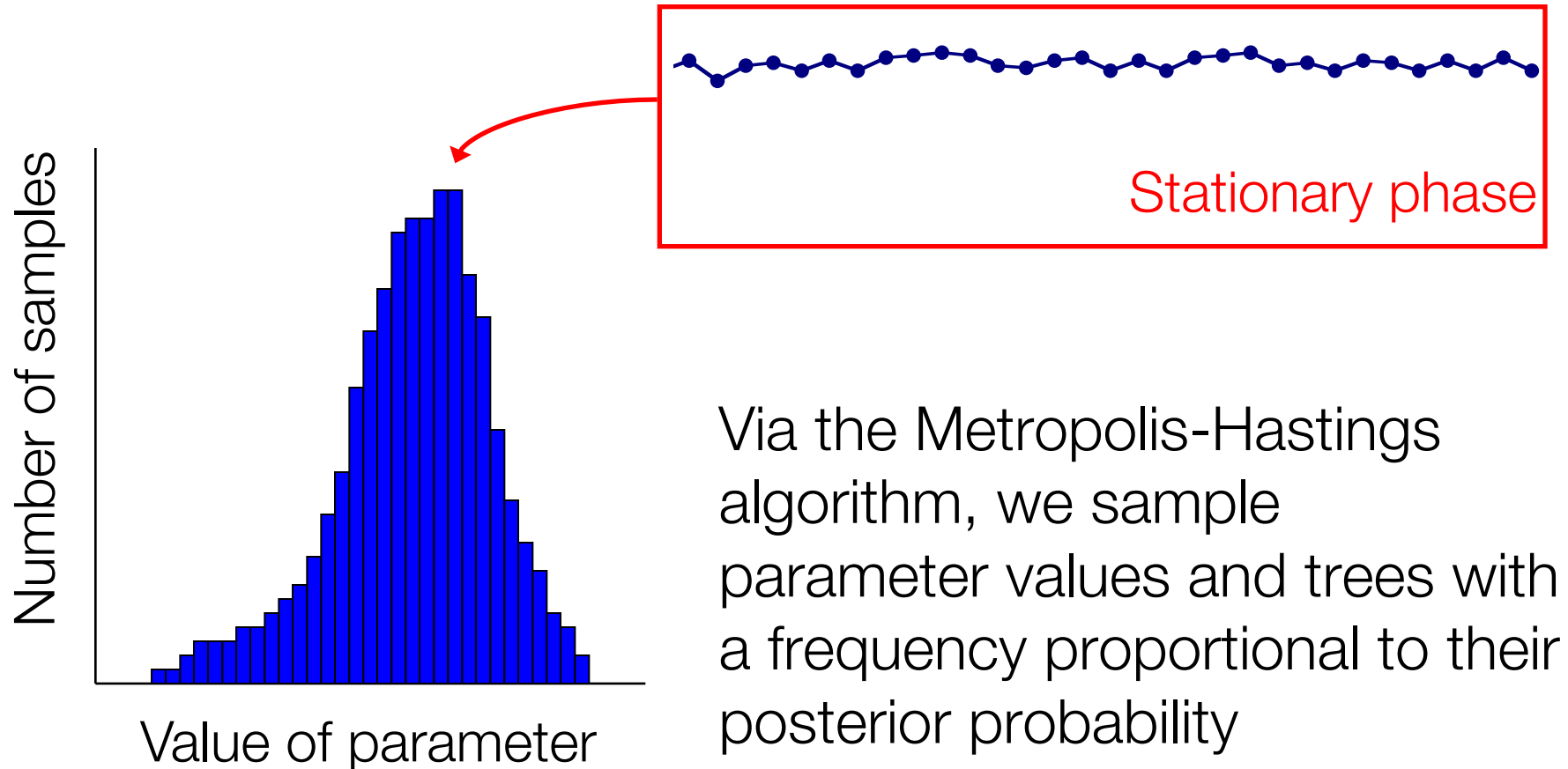




# Posterior distribution



# Posterior distribution

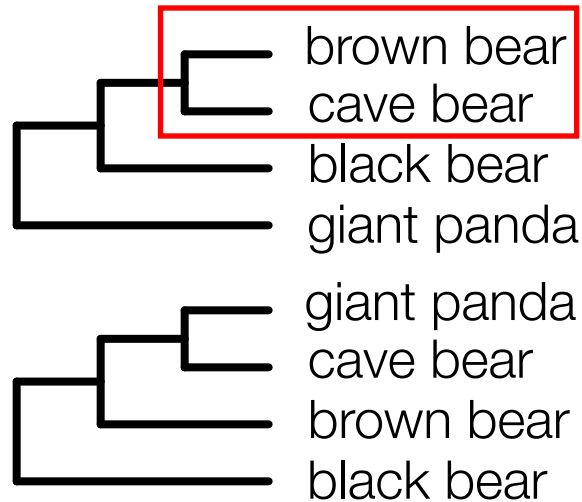
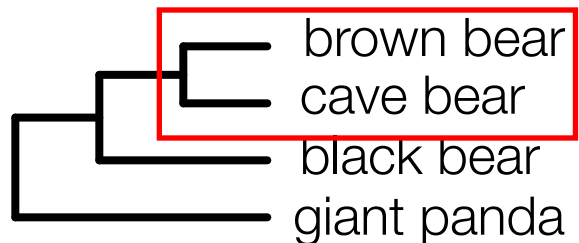
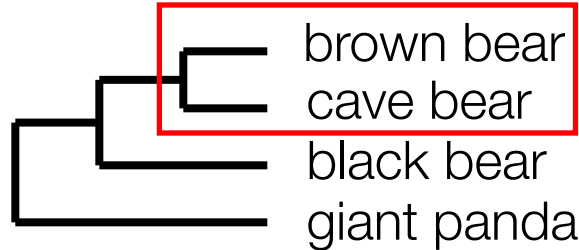
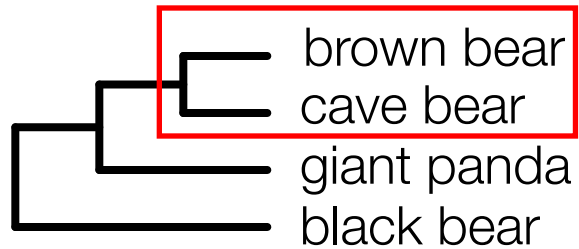


# Summarising the posterior

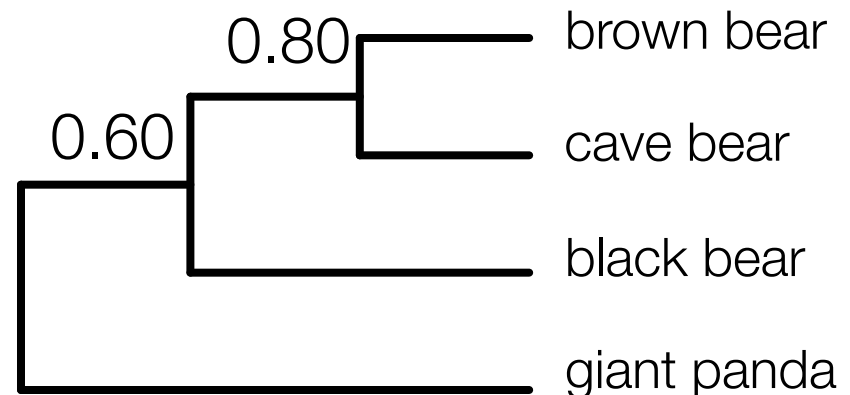
---

- 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*)
- Take the mean of the sampled values  
**Mean posterior estimate**
- Take the central 95% of the sampled values  
**95% credibility interval**

# Summarising the trees



**Compare this to  
bootstrapping in  
ML**



# Summarising the trees

---

- For each observed node, calculate the proportion of sampled trees in which the node is present
- For each node, this proportion is the ‘posterior probability’ of that node
- Finding a single representative tree
  - **Maximum *a posteriori* (MAP) tree**  
Sampled tree with highest posterior probability
  - **Maximum clade credibility (MCC) tree**  
Sampled tree with highest product of posterior node probabilities

# Questions?

---

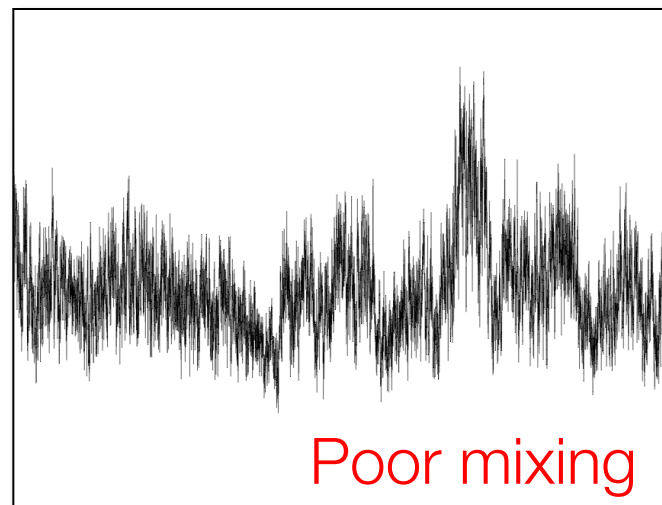
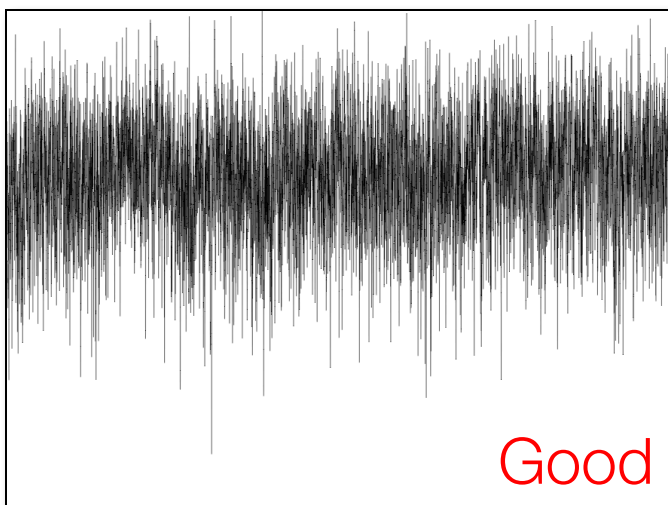
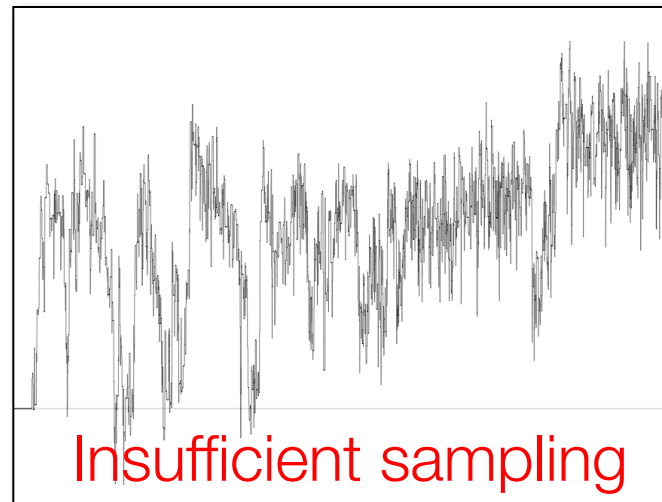
# Diagnostics

## 1. **Convergence/sufficient sampling**

Has the MCMC reached the stationary distribution?

## 2. **Mixing**

Are the samples from the stationary distribution reasonably independent of each other?



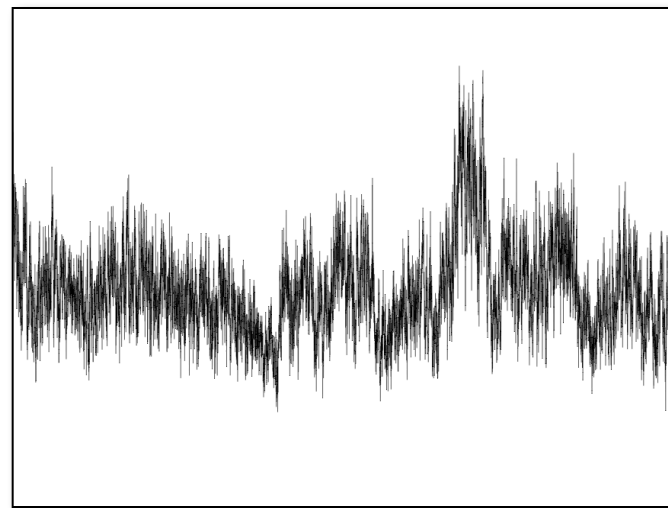
# Diagnostics

---

- **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
  - Decreasing the frequency of sampling from the MCMC





# Advantages and Disadvantages

# Advantages

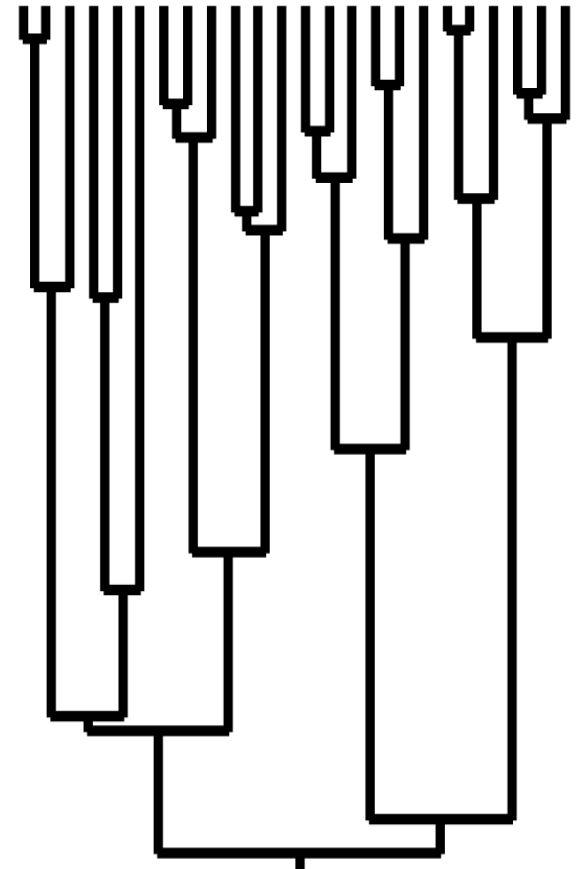
---

- Able to **implement highly parameterised models**
- **Estimating tree uncertainty** is straightforward
  - Can only do this indirectly in likelihood (via bootstrapping)
- **Posterior probabilities** have an intuitive interpretation
- Can incorporate **independent information** (in the prior)
- Can easily **integrate over ‘nuisance’ parameters** (i.e., those that are not of immediate interest)

# Advantages: Nuisance parameters

---

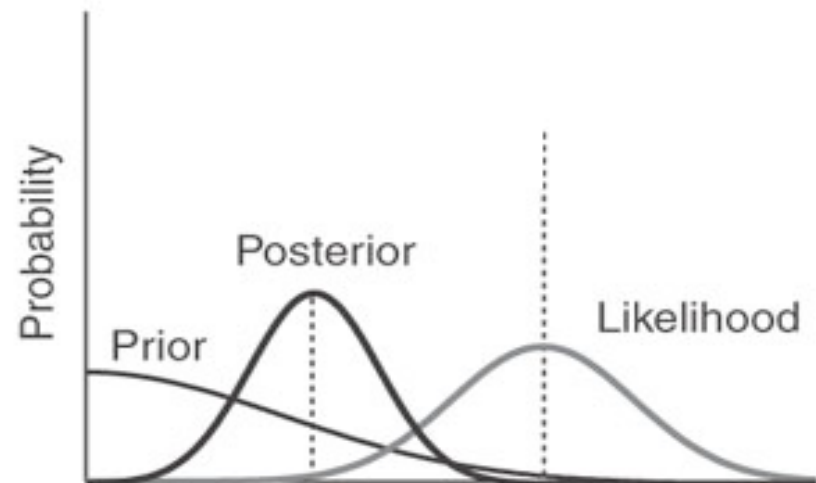
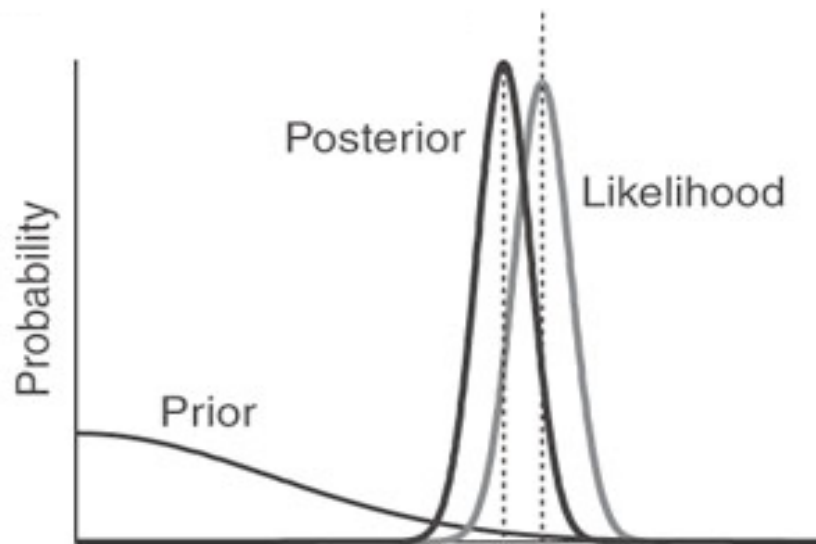
- Marginal likelihood of a parameter of interest
- Integrate over ‘nuisance’ parameters
- Example
  - Interested in evolutionary rate but not in the phylogenetic tree



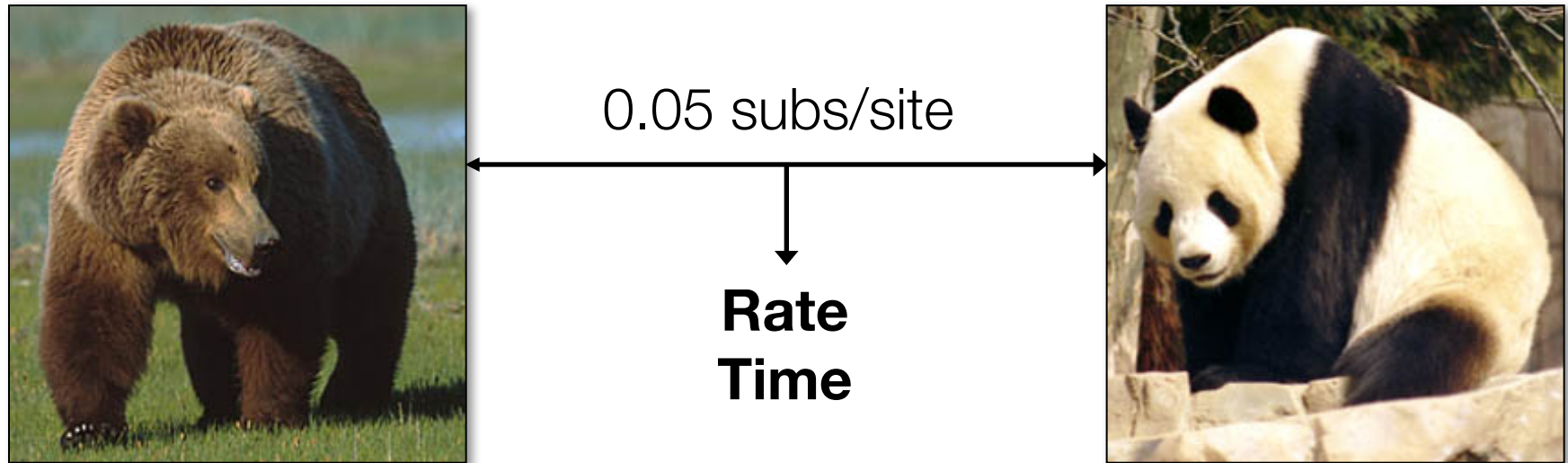
# Problems: Influence of priors

- Sensitivity of the posterior to the prior
  - This problem can arise if the data are uninformative

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$



# Problems: Overparameterisation

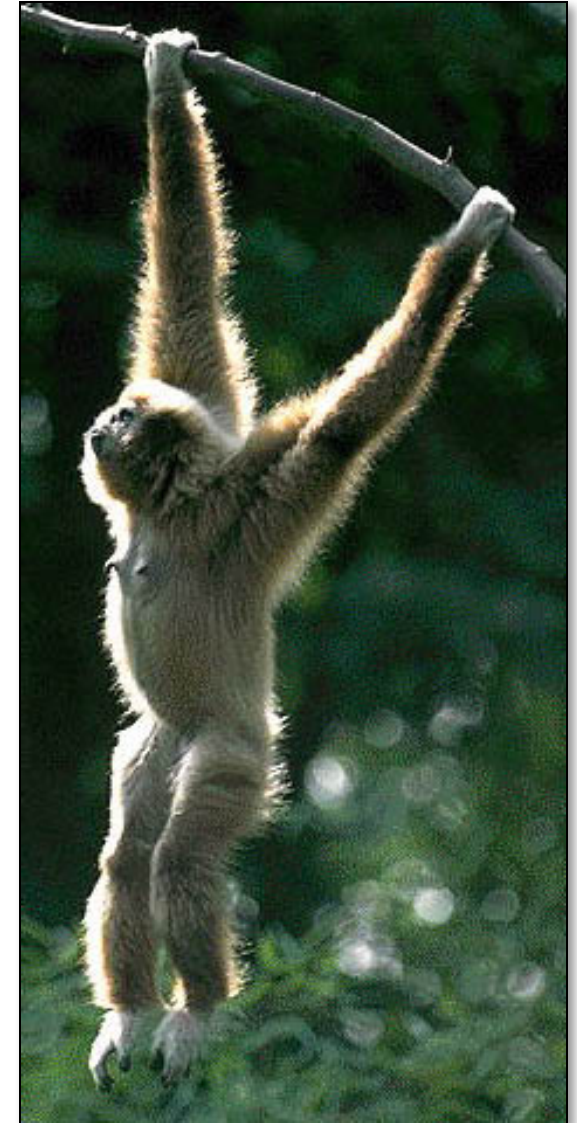
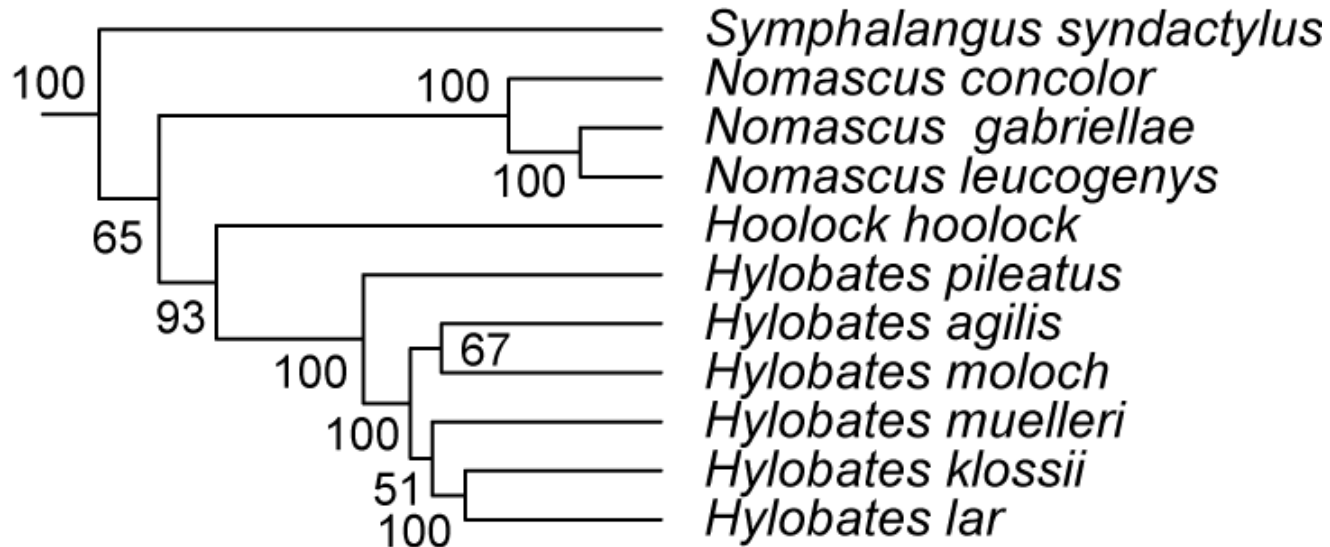
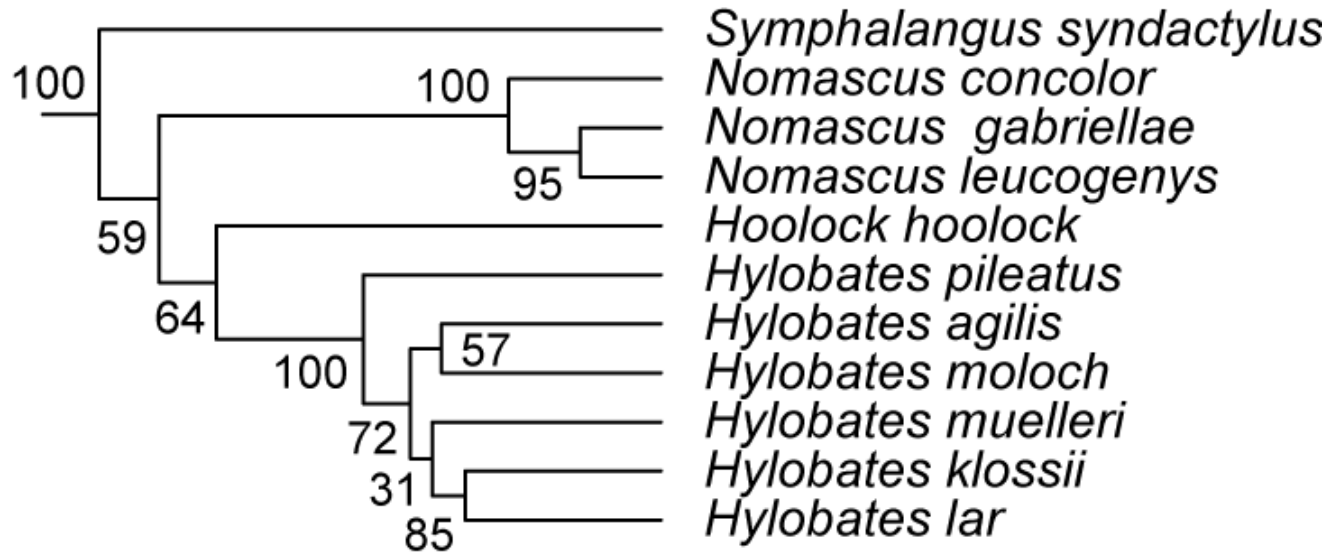


Parameter(s) are non-identifiable

This problem is not always obvious in the analysis

Bayesian methods are often robust to moderate overparameterisation

# Problems: Inflated support values?



# Methods in practice

- Maximum parsimony commonly used for morphological data, but rarely used otherwise
- Distance-based methods rarely used
- Maximum likelihood and Bayesian inference most commonly used for past 15 years or so
- Fast ML methods

