

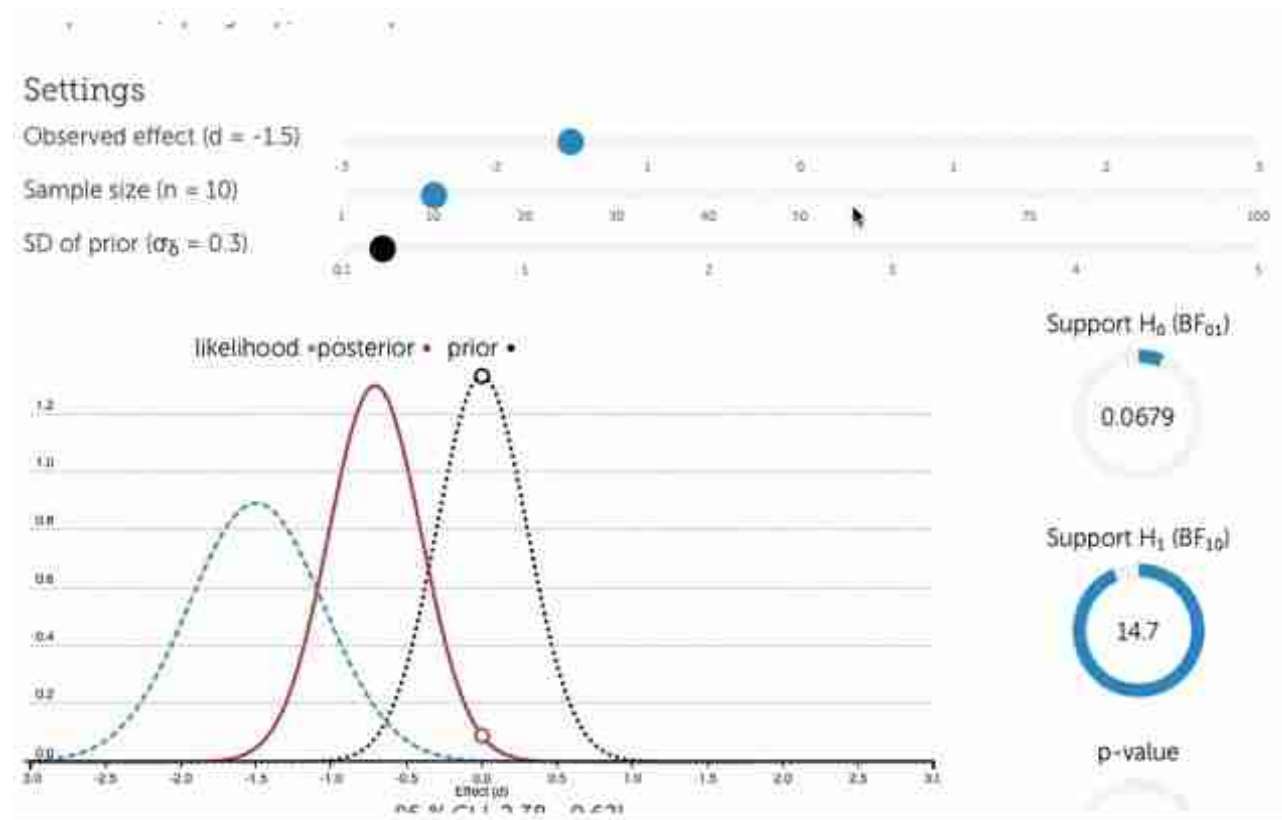
# Introduction to Bioinformatics for Computer Scientists

## Lecture 13

# Plan for next lectures

- Today
  - Convergence of MCMC chains
  - Escaping local optima
  - advanced MCMC
- Next time: population genetics

# What does Bayesian Inference do?

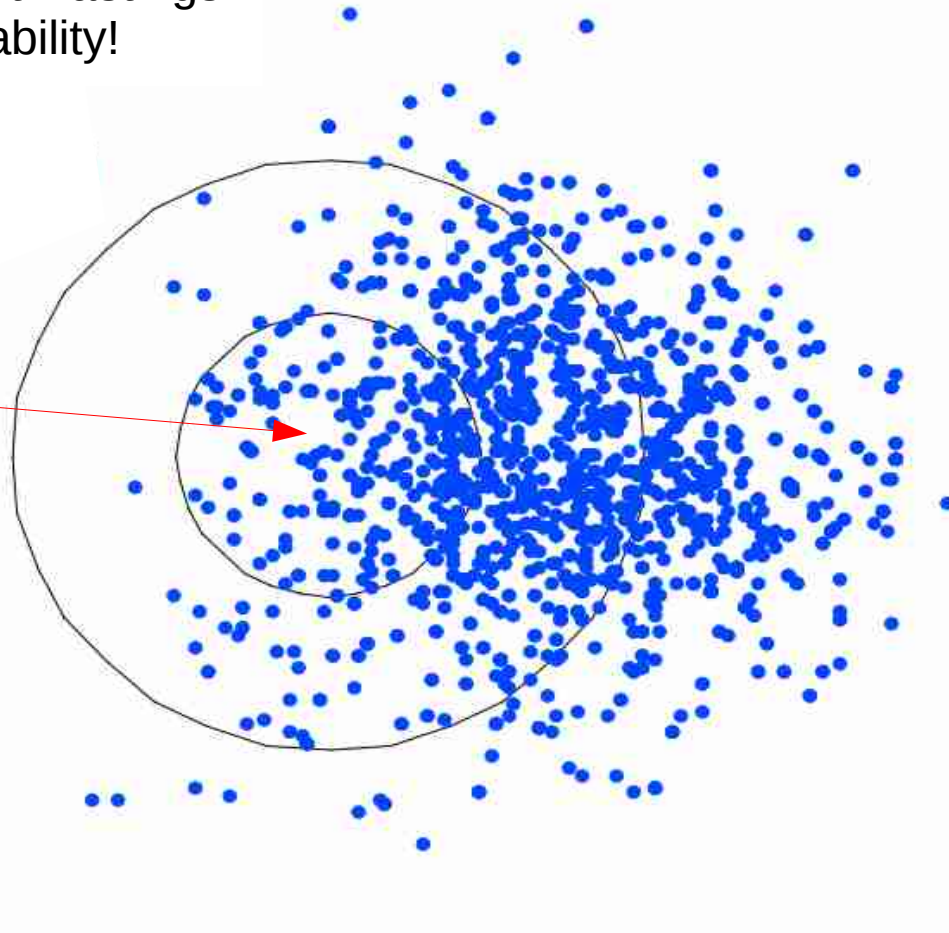


Live demo: <http://rpsychologist.com/d3/bayes/>

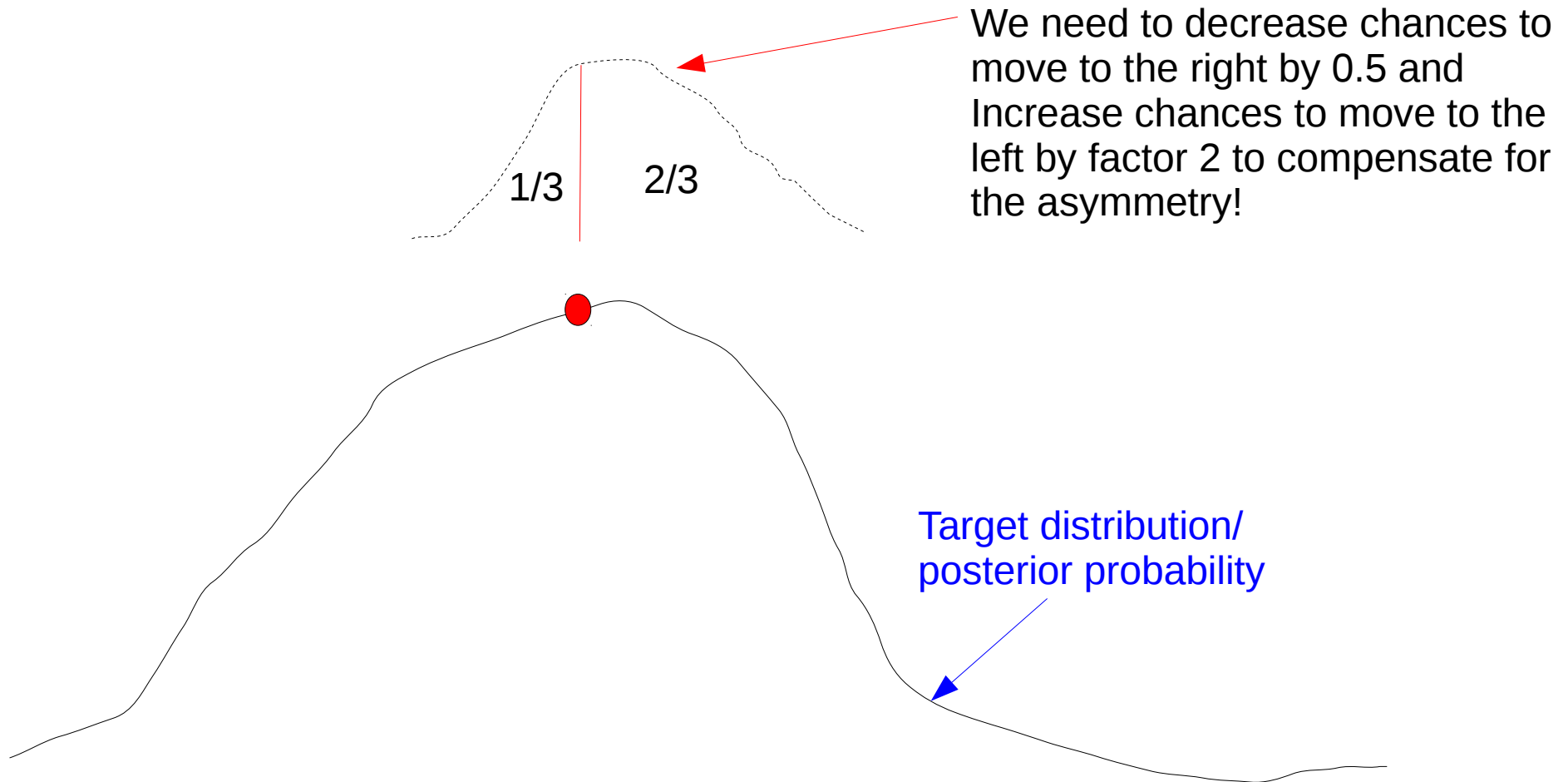
# Uncorrected Proposal Distribution A Robot in 3D

Example: MCMC proposed moves to the right 80% of the time without Hastings correction for acceptance probability!

Peak area



# Hastings Correction



# Hastings Correction

$$R = \left( \frac{\Pr(\text{point2})}{\Pr(\text{point1})} \right) * \left( \frac{\Pr(\text{data}|\text{point2})}{\Pr(\text{data}|\text{point1})} \right) * \left( \frac{Q(\text{point1}|\text{point2})}{Q(\text{point2}|\text{point1})} \right)$$

Prior ratio: for uniform priors this is 1 !

Likelihood ratio

Hastings ratio: if  $Q$  is symmetric  
 $Q(\text{point1}|\text{point2}) = Q(\text{point2}|\text{point1})$  and  
the hastings ratio is 1 → we obtain the  
normal Metropolis algorithm

# Hastings Correction more formally

$$R = \left( \frac{f(\theta^*)}{f(\theta_i)} \right) * \left( \frac{f(\text{data}|\theta^*)}{f(\text{data}|\theta_i)} \right) * \left( \frac{Q(\theta_i|\theta^*)}{Q(\theta^*|\theta_i)} \right)$$

Prior ratio

Likelihood ratio

Hastings ratio

# Hastings Correction is not trivial

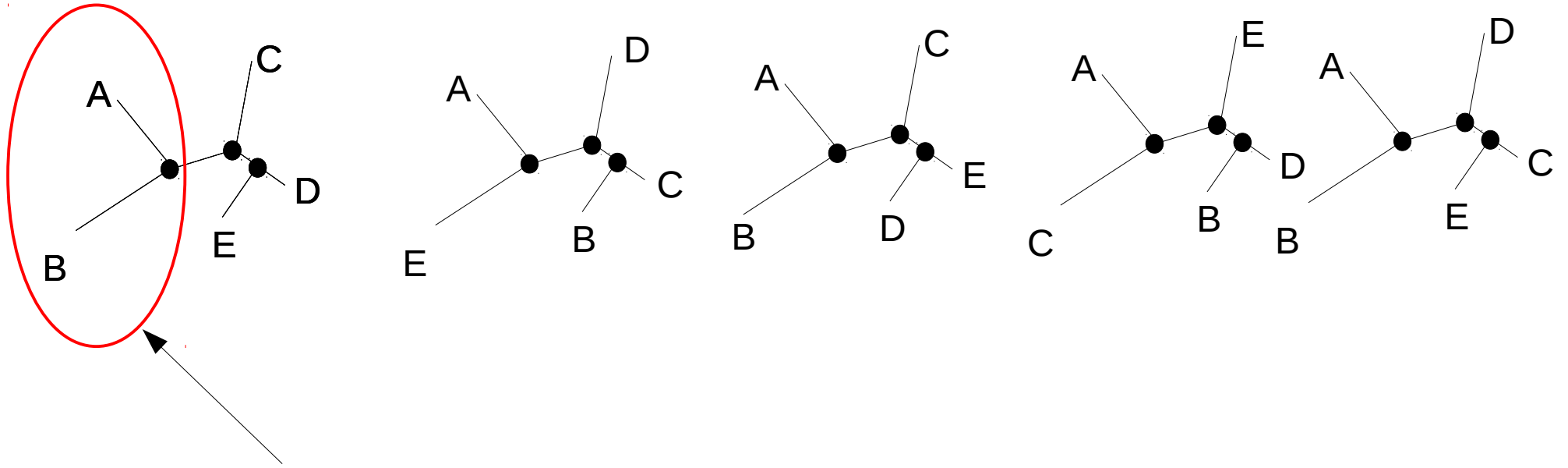
- Problem with the equation for the hastings correction
- M. Holder, P. Lewis, D. Swofford, B. Larget. 2005.  
**Hastings Ratio of the LOCAL Proposal Used in Bayesian Phylogenetics.** *Systematic Biology*. 54:961-965.  
<http://sysbio.oxfordjournals.org/content/54/6/961.full>

*“As part of another study, we estimated the marginal likelihoods of trees using different proposal algorithms and discovered repeatable discrepancies that implied that the published Hastings ratio for a proposal mechanism used in many Bayesian phylogenetic analyses is incorrect.”*

- Incorrect Hastings ratio used from 1999-2005

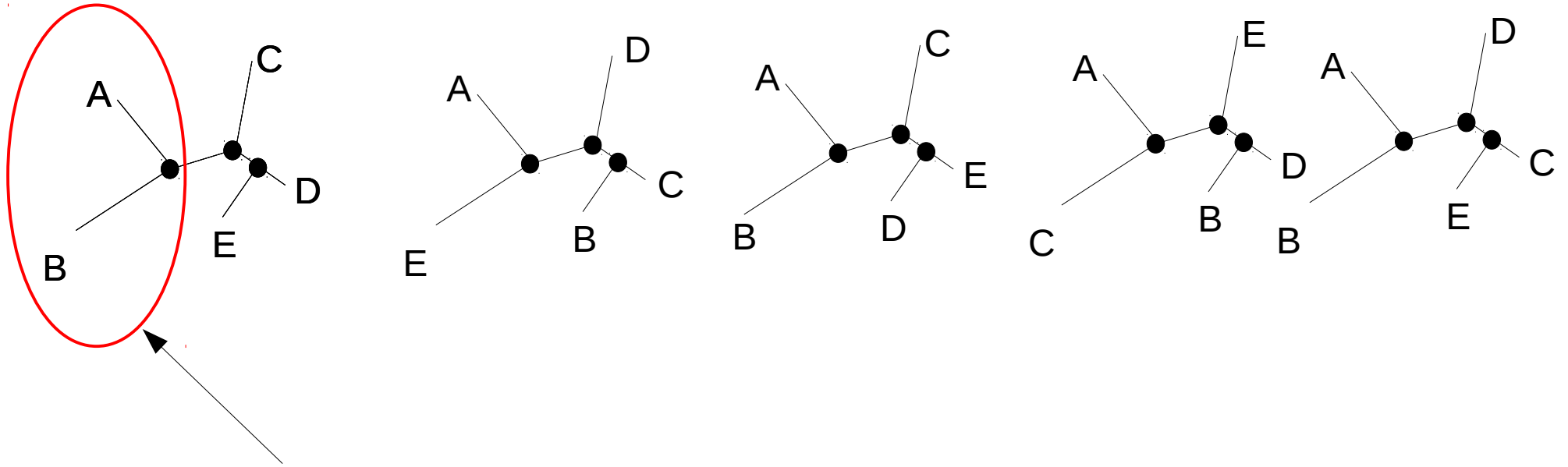


# Back to Phylogenetics



What's the posterior probability of bipartition  $AB|CDE$ ?

# Back to Phylogenetics

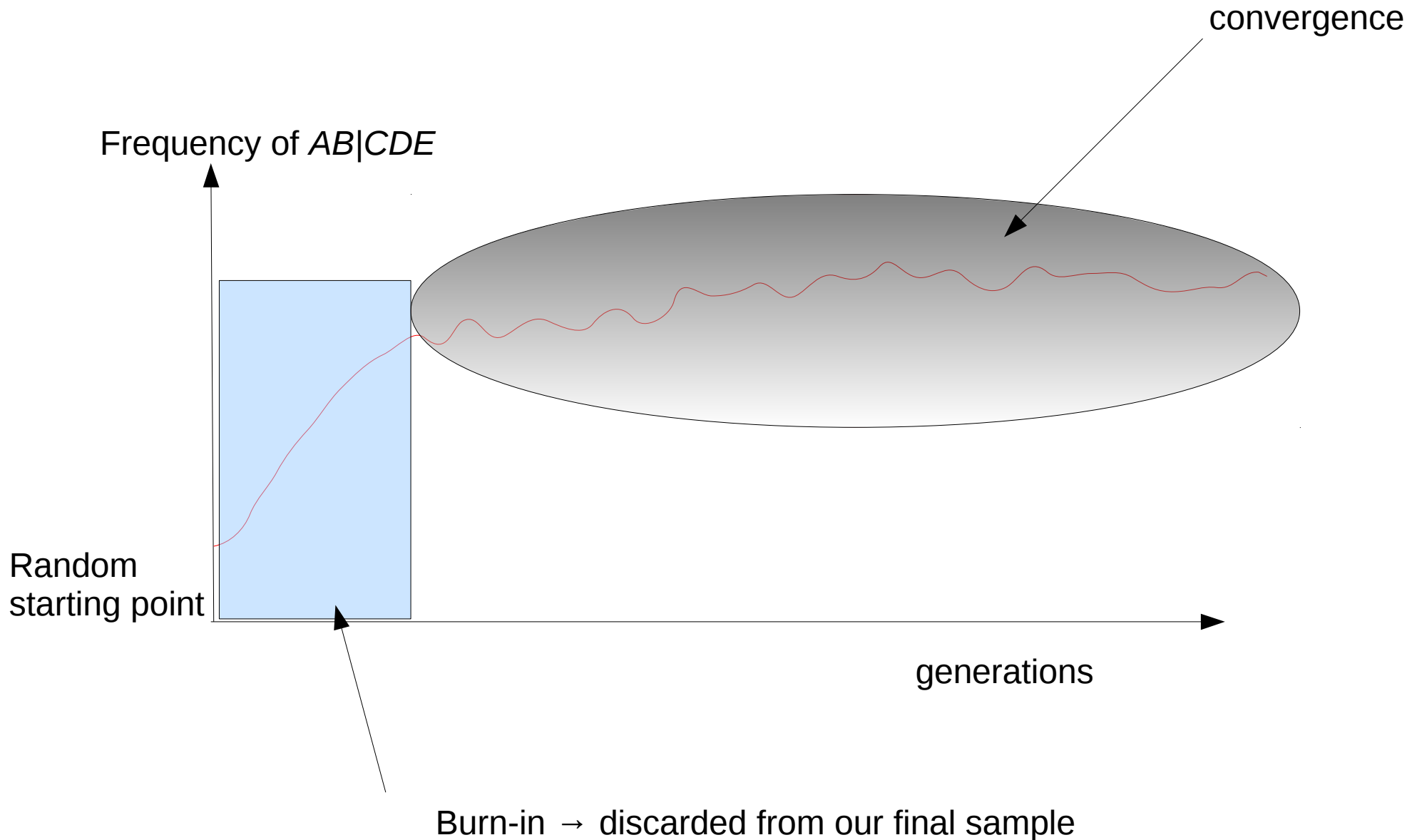


What's the posterior probability of bipartition  $AB|CDE$ ?

We just count from the sample generated by MCMC, here it's  $3/5 \rightarrow 0.6$

This approximates the true proportion (posterior probability) of bipartition  $AB|CDE$  **if** we have run the chain long enough and **if** it has converged

# MCMC in practice

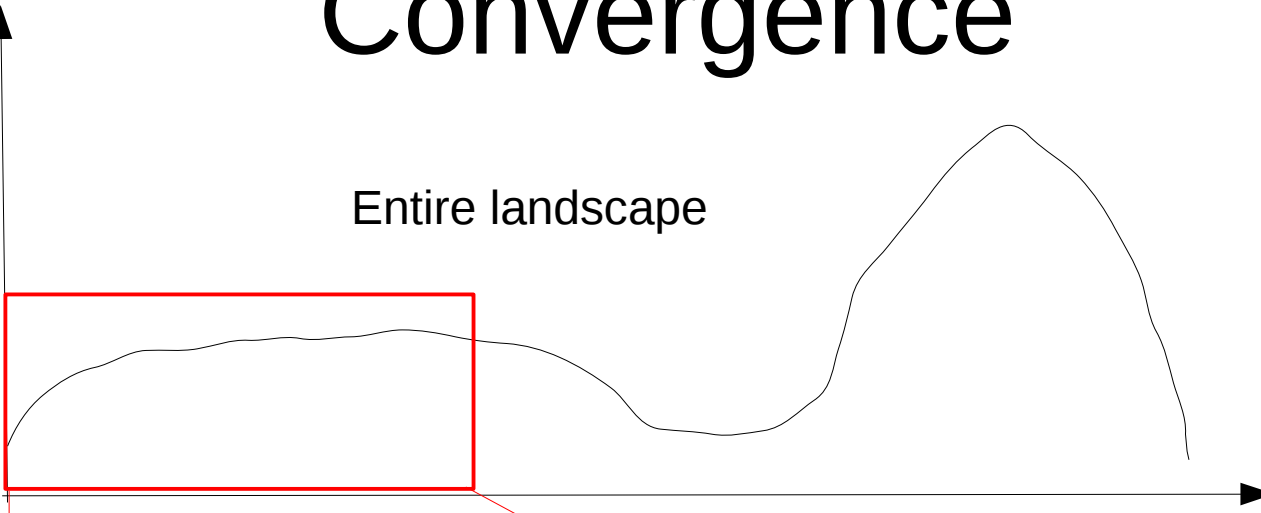


# Convergence

- How many samples do we need to draw to obtain an accurate approximation?
- When can we stop drawing samples?
- Methods for convergence diagnosis
  - we can **never** say that a MCMC-chain has converged
  - we can only diagnose that it has not converged
  - a plethora of tools for convergence diagnostics for phylogenetic MCMC exists

# Convergence

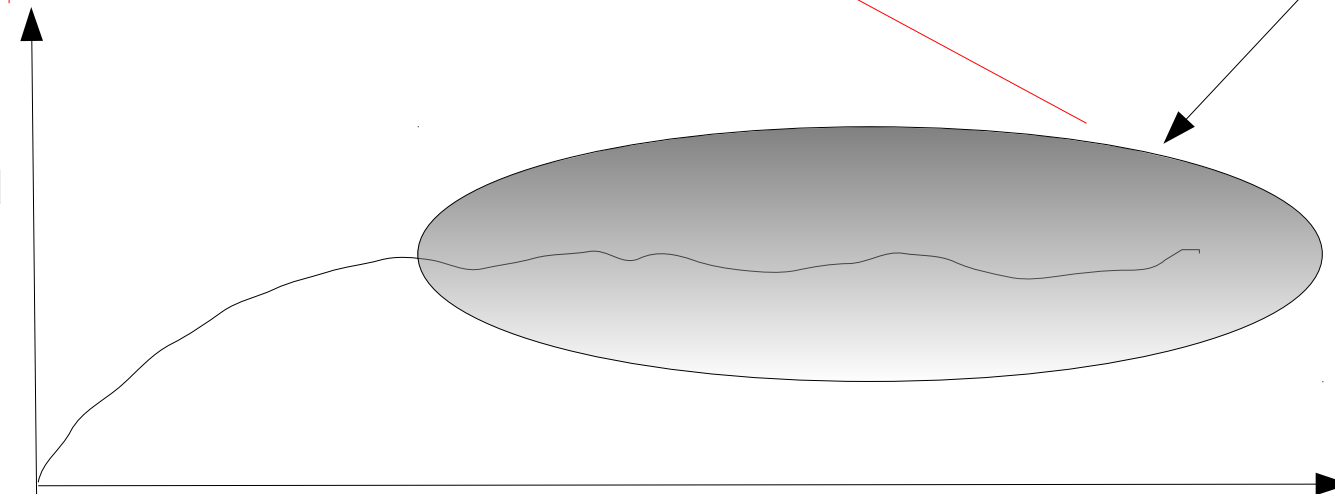
Likelihood score



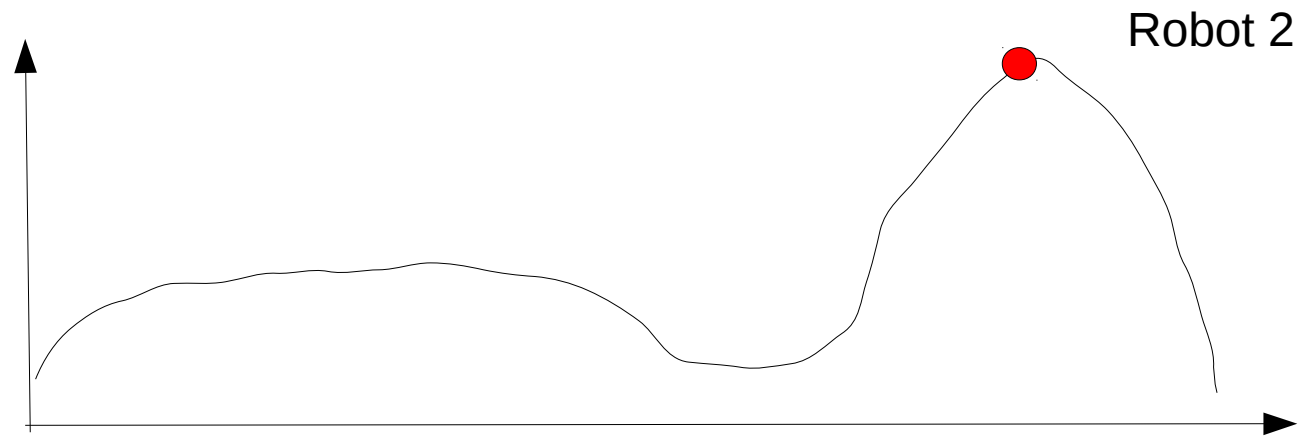
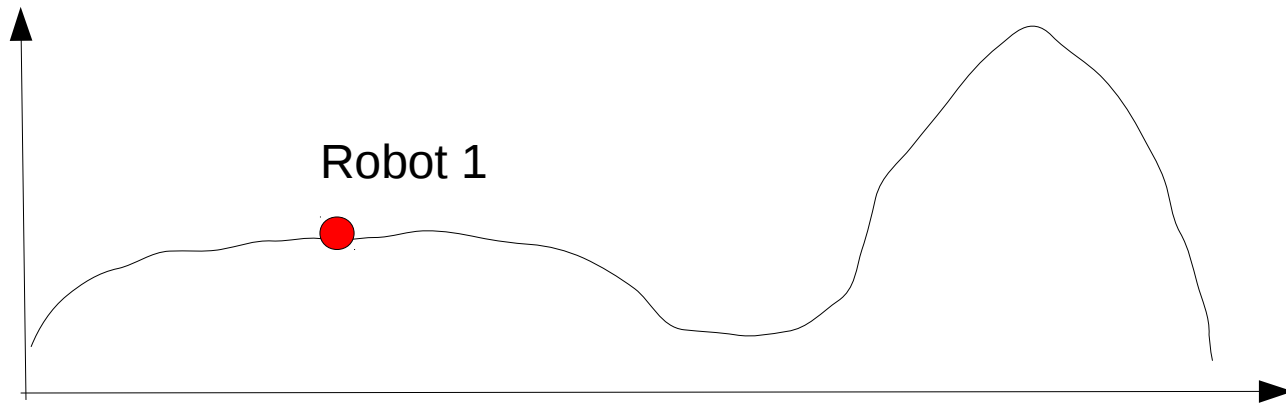
Area of apparent convergence

Zoom in

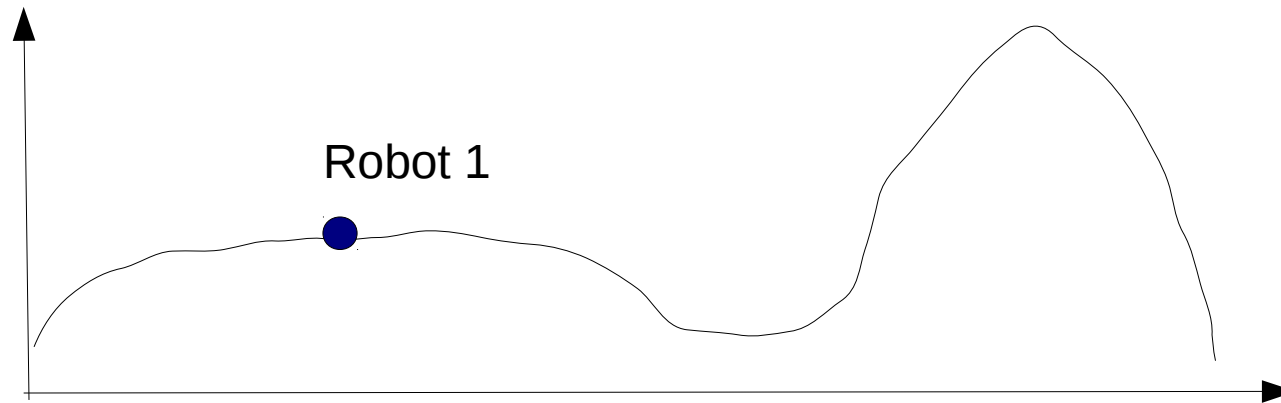
Likelihood Score output MCMC method



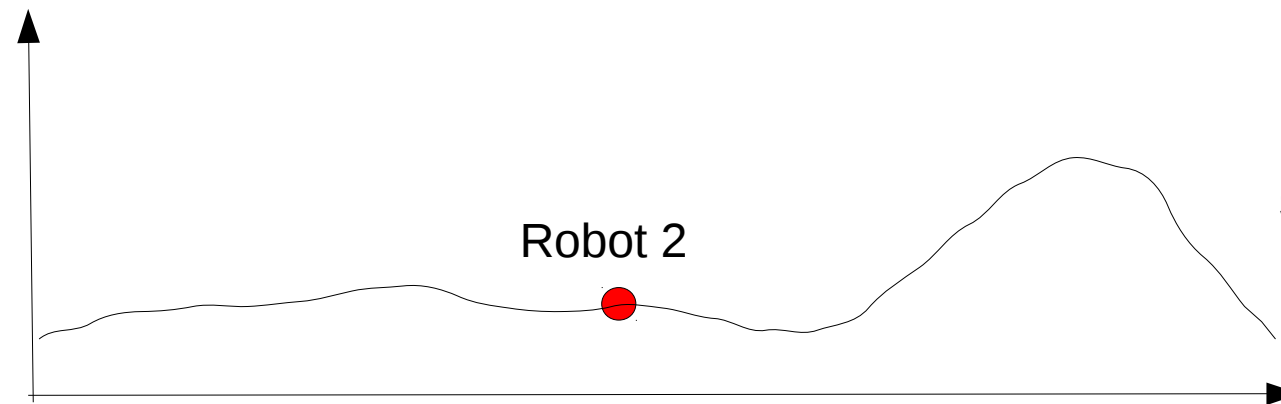
# Solution: Run Multiple Chains



# Heated versus Cold Chains



Cold chain: sees landscape as is



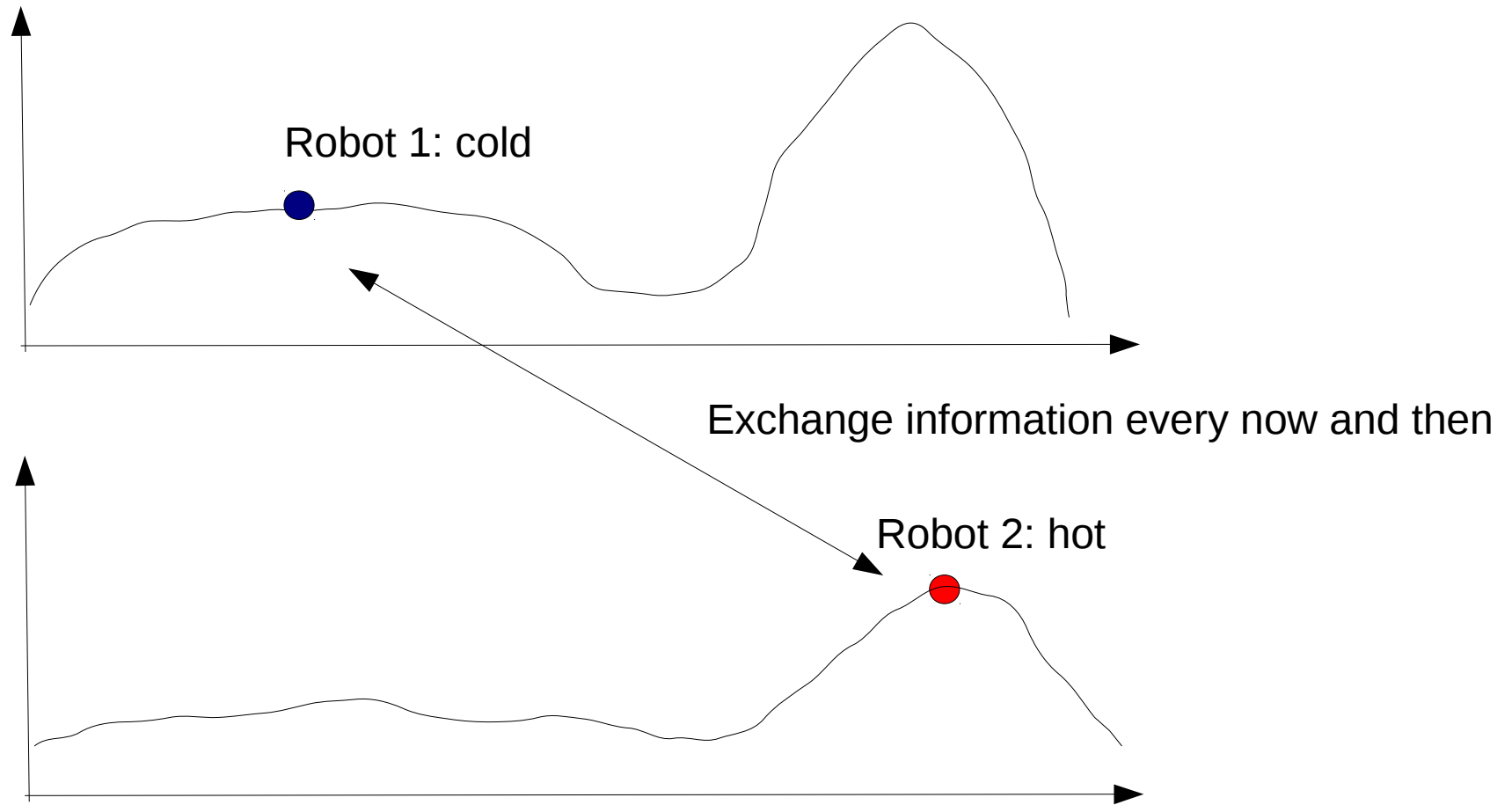
Hot chain: sees a flatter version of the same landscape → Moves more easily between peaks

# Known as MCMCMC

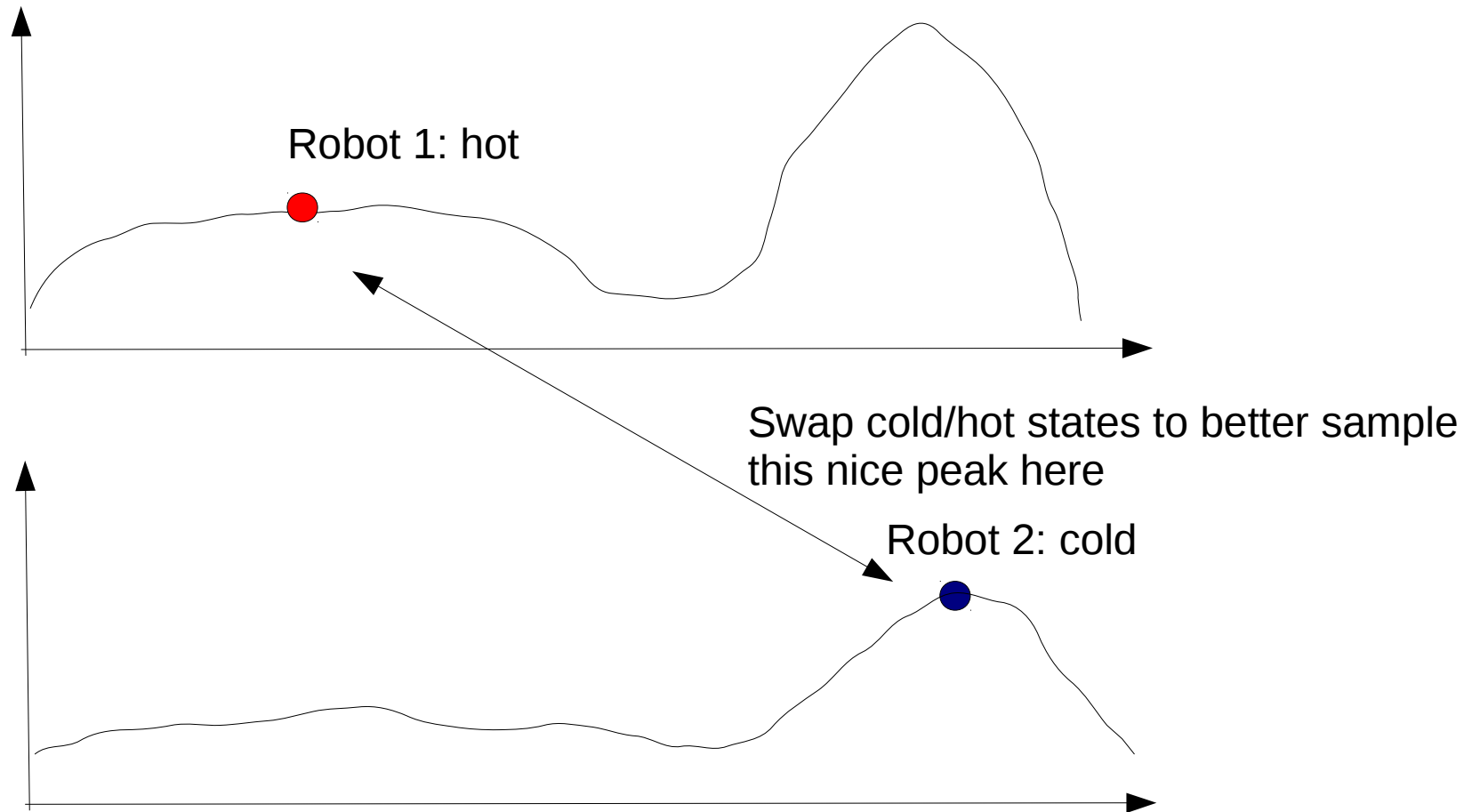
- Metropolis-Coupled Markov-Chain Monte Carlo
- Run several chains simultaneously
  - 1 **cold** chain (the one that samples)
  - Several **heated** chains (the ones that explore)
- Heated chain robots explore the parameter space in larger steps
- To flatten the landscape the acceptance ratio  $R$  is modified as follows:  $R^{1/1+H}$  where  $H$  is the so-called temperature
  - For the cold chain  $H := 0.0$
  - Setting the temperature for the hot chains is a bit of woo-do



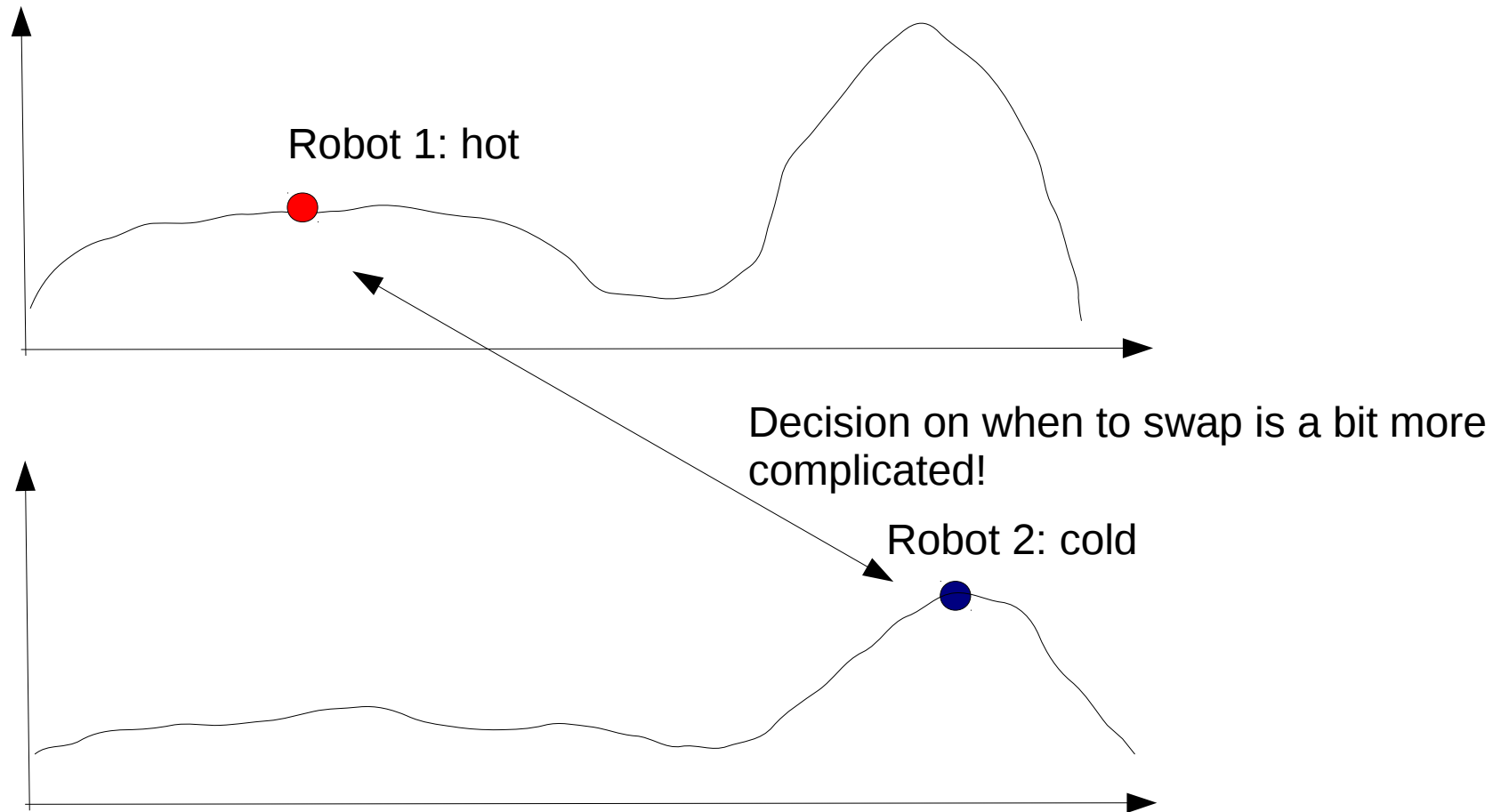
# Heated versus Cold Chains



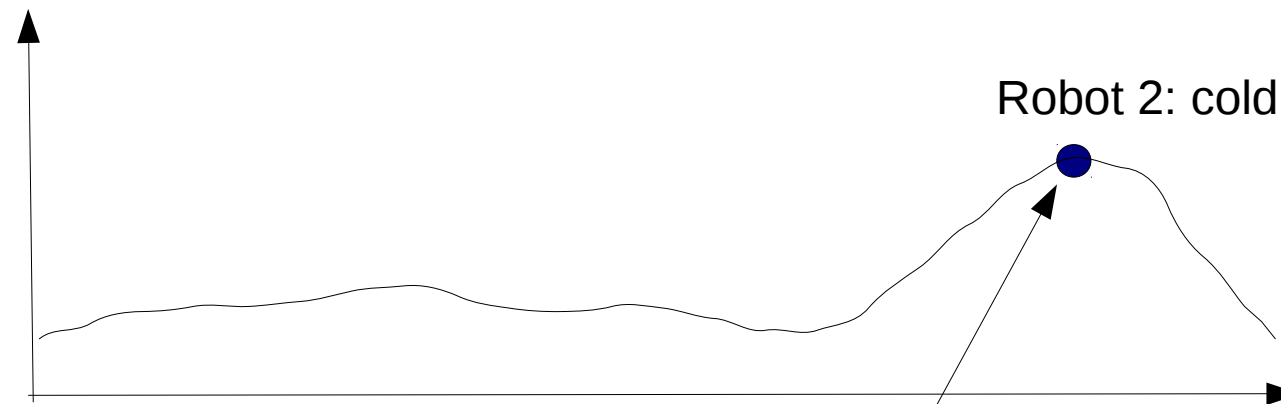
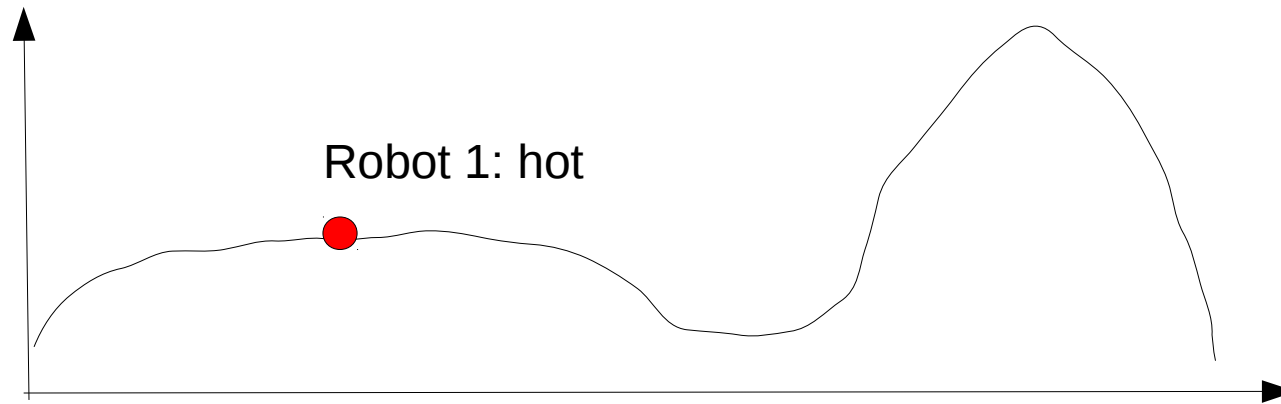
# Heated versus Cold Chains



# Heated versus Cold Chains



# Heated versus Cold Chains



Only the cold robot actually emits states (writes samples to file)

# A few words about priors

- Prior probabilities convey the scientist's beliefs, before having seen the data
- Using uninformative prior probability distributions (e.g., uniform priors, also called flat priors)
  - differences between prior and posterior distribution are attributable to likelihood differences
- Priors can bias an analysis
- For instance, we could chose an arbitrary prior distribution for branch lengths in the range [1.0,20.0]
  - what happens if branch lengths are much shorter?

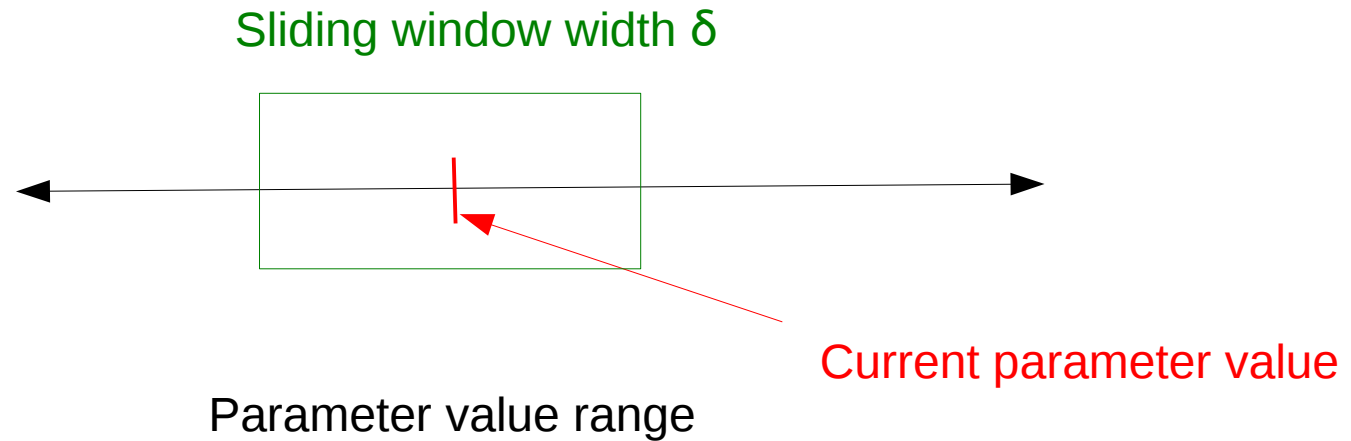
# Outline

- Some phylogenetic MCMC proposals
- Reversible jump MCMC

# Some Phylogenetic Proposal Mechanisms

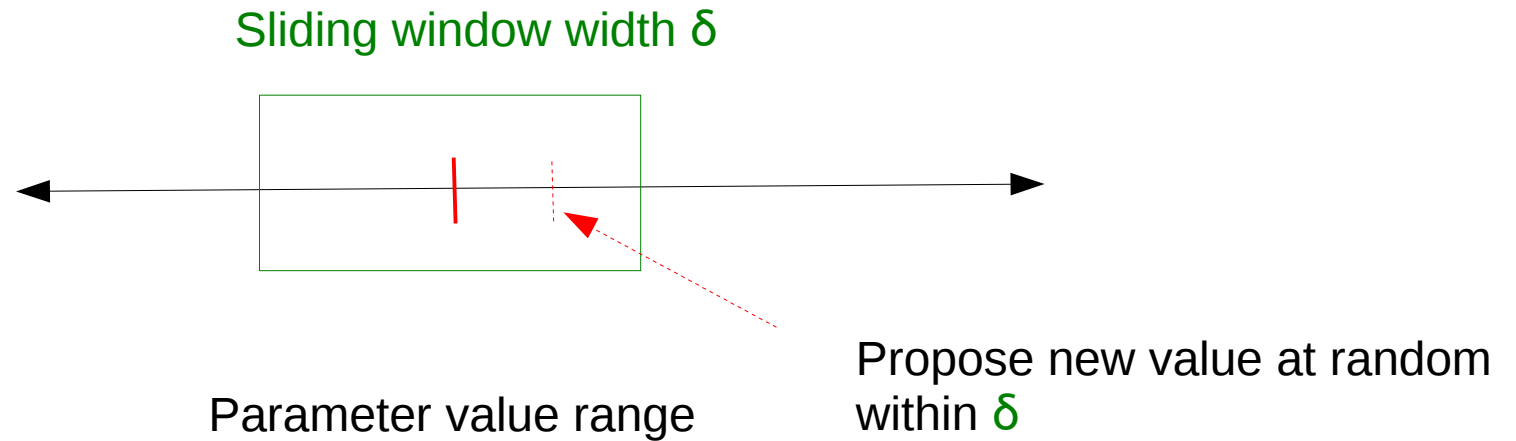
- Univariate parameters & branch lengths
  - Sliding Window Proposal
- Branch lengths
  - Node slider proposal
- Topologies
  - Local Proposal (the one with the bug in the Hastings ratio!)
- Remember: We need to design proposals for which
  - We either *don't need to* calculate the Hastings ratio
  - Or for which we *can* calculate it
  - That have an appropriate acceptance rate
    - all sorts of tricks being used, e.g., parsimony-biased topological proposals
    - acceptance rate should be around 25% (empirical observation)
    - for sampling from a multivariate normal distribution it has been formally shown that an acceptance rate of 23.4% is optimal

# Sliding Window Proposal

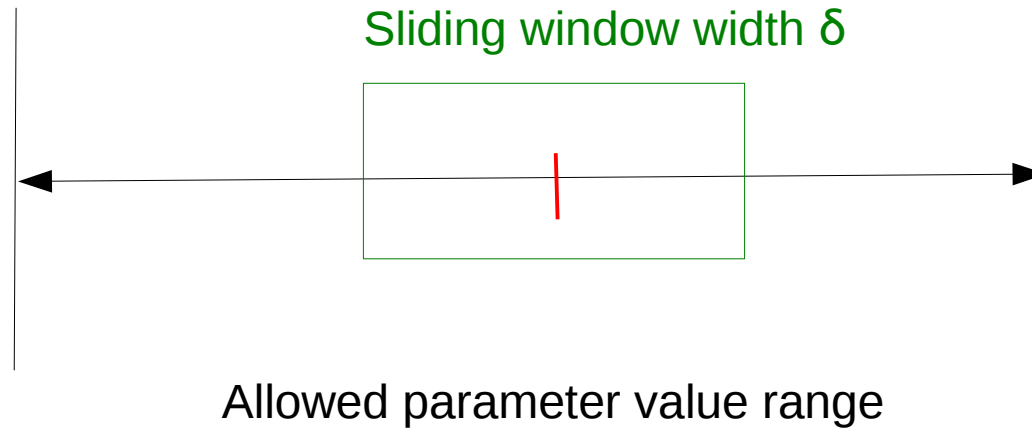




# Sliding Window Proposal



# Sliding Window Proposal

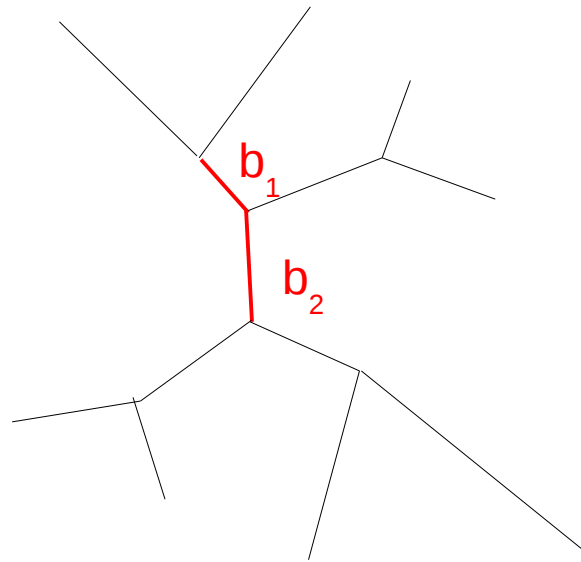


Notes:

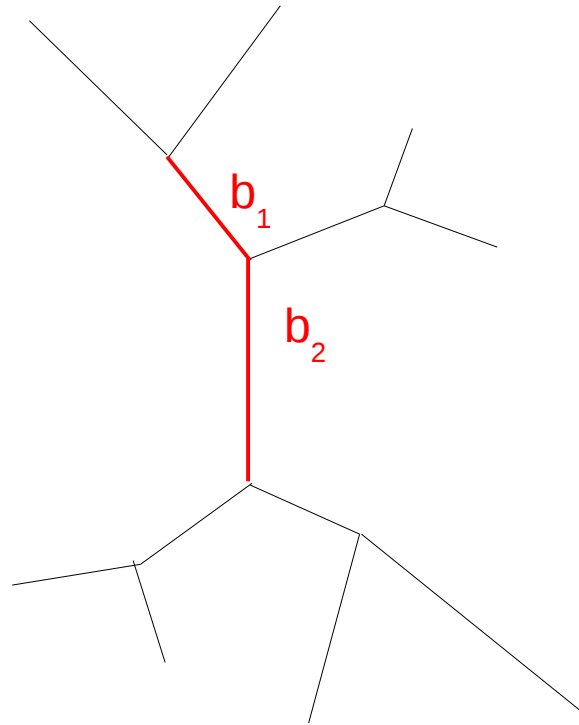
1. The Hastings ratio of this move is  $1$
2. The edge cases can be handled by back-projection
3. The window size  $\delta$  can be tuned itself (auto-tuning) to obtain an acceptance rate of  $\approx \frac{1}{4}$
4. This proposal can be used, e.g., for the  $\alpha$ -shape parameter of the  $\Gamma$  function in rate heterogeneity models

# The Node Slider Proposal

1. Pick **2 contiguous branches** at random

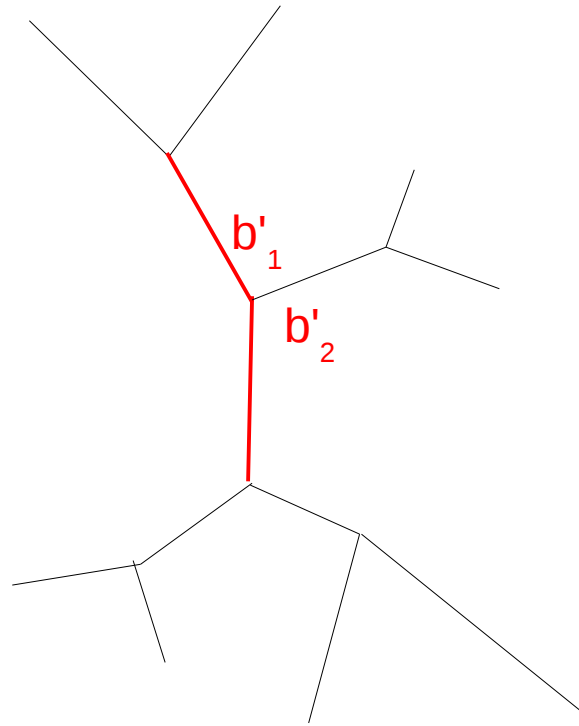


# The Node Slider Proposal



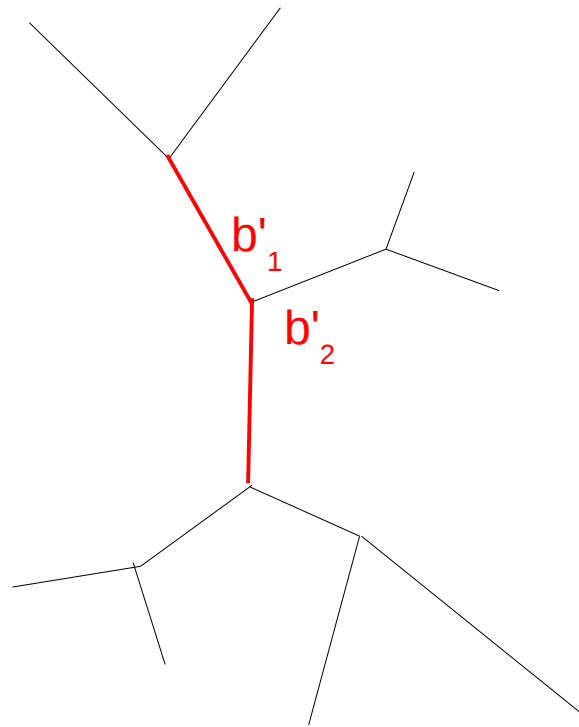
1. Pick **2 contiguous branches** at random
2. Multiply the **2 branches** by the same random number

# The Node Slider Proposal



1. Pick **2 contiguous branches** at random
2. Multiply the **2 branches** by the same random number
3. Propose a new branch ratio  $b'_1/b'_2$  at random

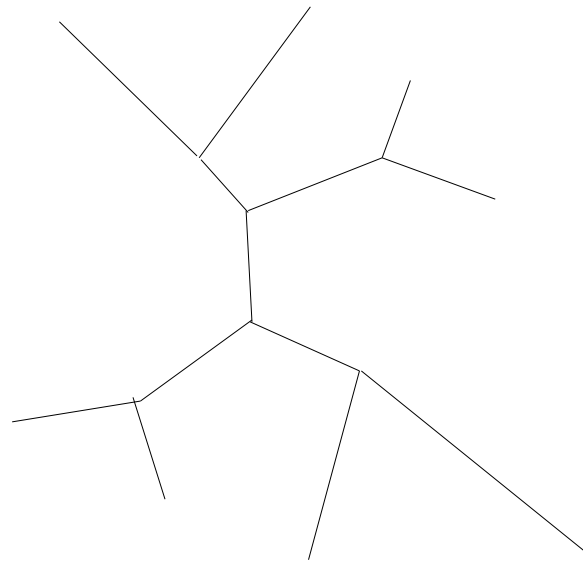
# The Node Slider Proposal



1. Pick **2 contiguous branches** at random
2. Multiply the **2 branches** by the same random number
3. Propose a new branch ratio  $b'_1/b'_2$  at random

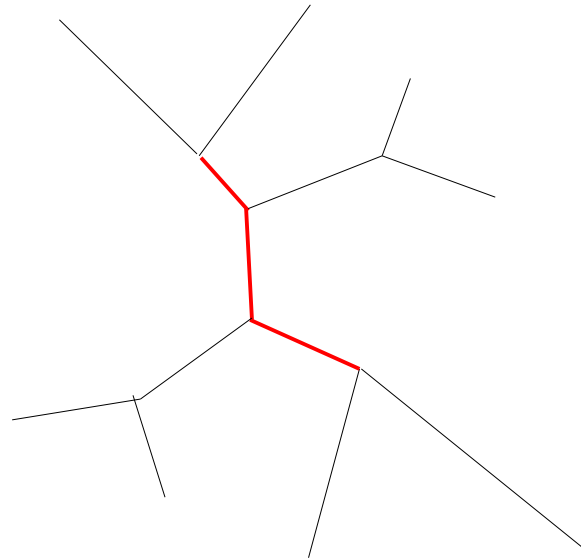
The Hastings ratio of this move is not 1!

# Moving through Tree Space



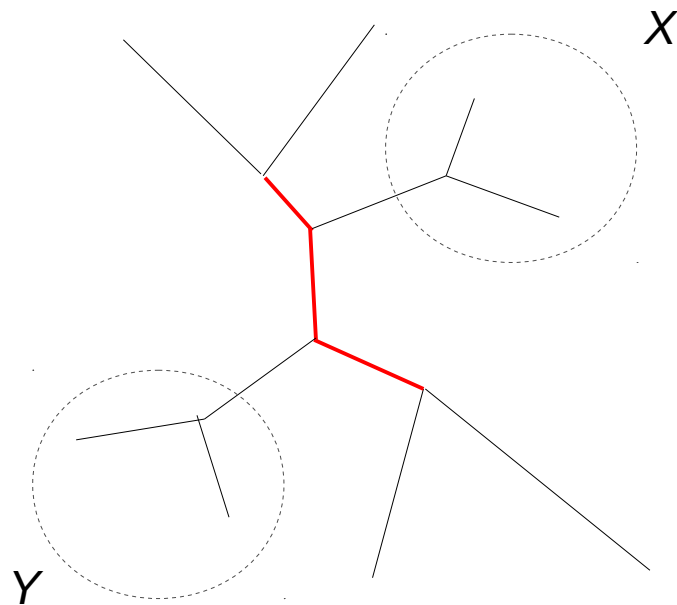
# Moving through Tree Space

1. Pick **3 contiguous branches** at random



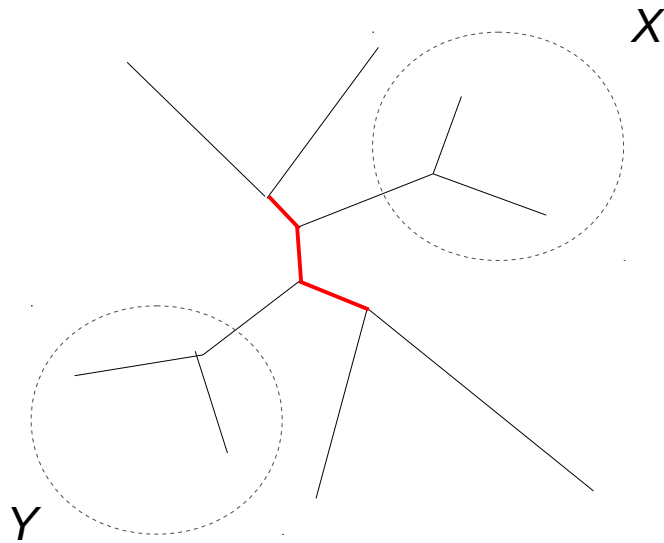


# Moving through Tree Space



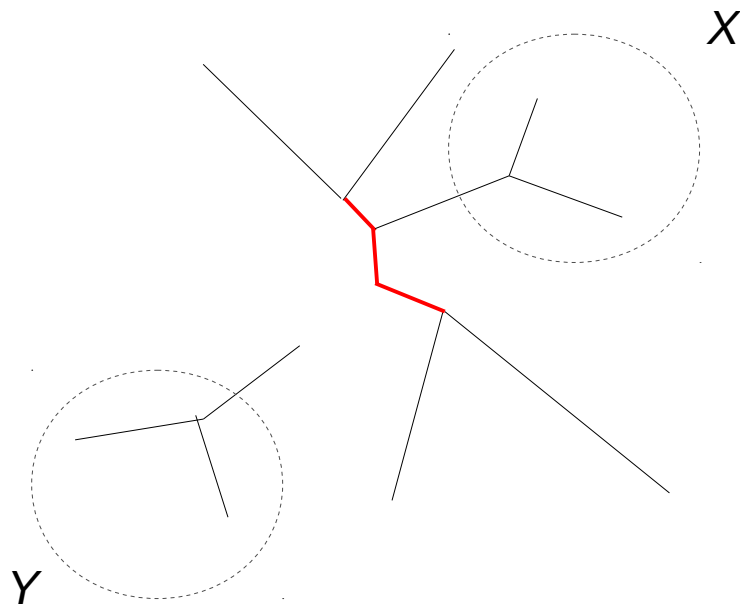
1. Pick **3 contiguous branches** at random that define 2 Subtrees X and Y

# Moving through Tree Space



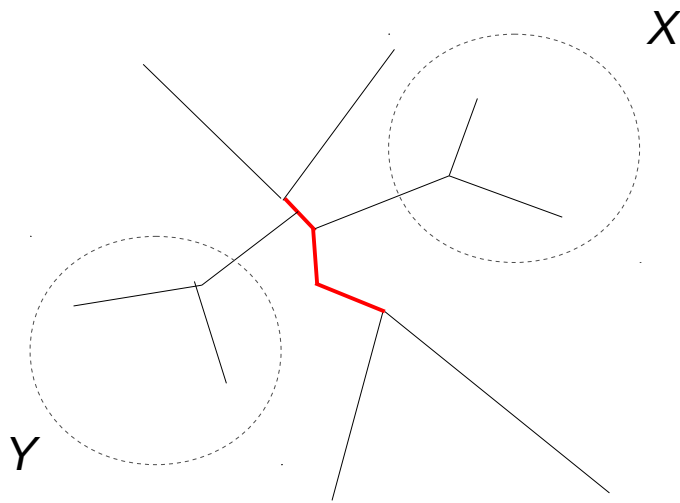
1. Pick **3 contiguous branches** at random that define 2 Subtrees X and Y
2. shrink or grow selected **3 branch segment** by a random amount

# Moving through Tree Space



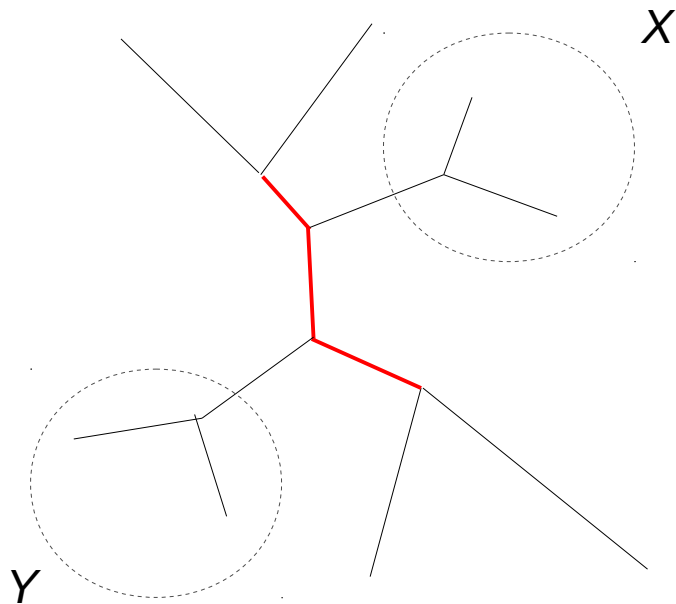
1. Pick **3 contiguous branches** at random that define 2 Subtrees X and Y
2. shrink or grow selected **3 branch segment** by a random Amount
3. Chose either X or Y at random and prune it from the tree

# Moving through Tree Space

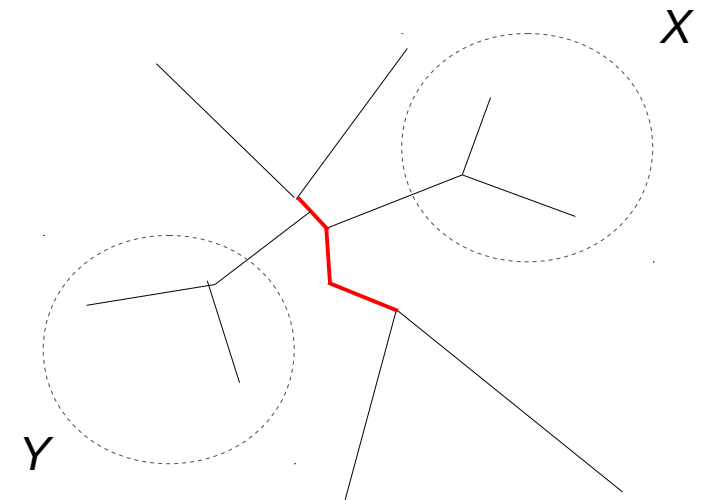


1. Pick **3 contiguous branches** at random that define 2 Subtrees X and Y
2. shrink or grow selected **3 branch segment** by a random Amount
3. Chose either X or Y at random And prune it from the tree
4. Re-insert Y at random into The **3 branch segment**

# Moving through Tree Space



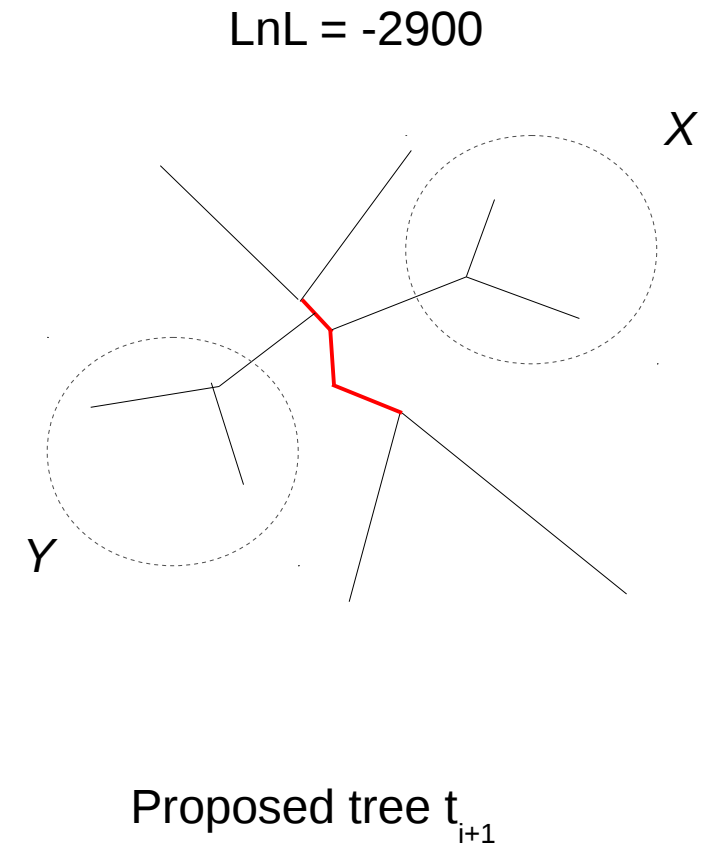
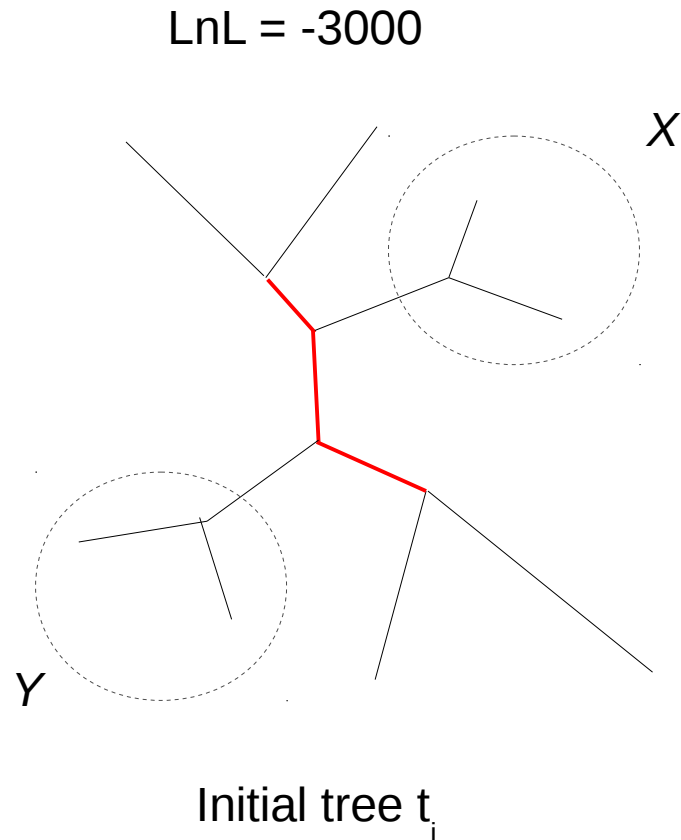
Initial tree  $t_i$



Proposed tree  $t_{i+1}$

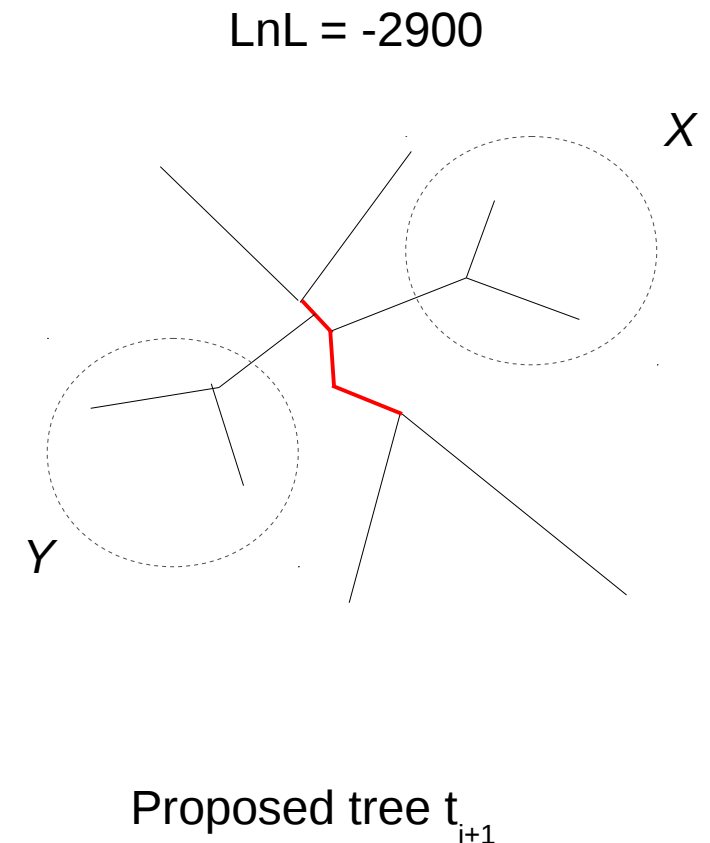
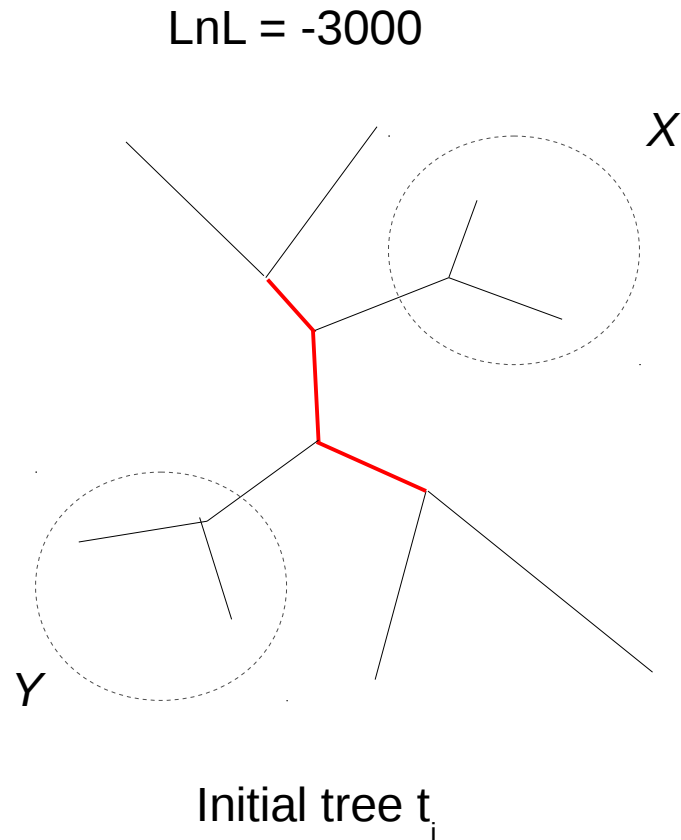
Proposed tree: 3 branch lengths changed and one NNI (Nearest Neighbor Interchange) move applied

# Moving through Tree Space



The proposed tree has a better likelihood!  
Will the proposed tree always be accepted?

# Moving through Tree Space



The proposed tree has a better likelihood!  
Will the proposed tree always be accepted?  
→ think about Priors and Hastings ratio!

# Outline for today

- Some phylogenetic MCMC proposals
- Reversible jump MCMC



# How do we select models using MCMC?

- **Example:** Consider all possible time-reversible nucleotide substitution models ranging from Jukes Cantor (JC, 1 rate) to the General Time Reversible Model (GTR, 6 rates)
- We will denote rate configurations by strings, e.g.,
  - 111111 is the JC model
  - ...
  - 123456 is the GTR model
- Do I need to explain this further?

# How many time-reversible DNA models are there?

- Number of ways a set with  $n$  objects can be partitioned into disjoint non-empty sets

- **Example:** the set  $\{a,b,c\}$  can be partitioned as follows:

$\{ \{a\}, \{b\}, \{c\} \}$

$\{ \{a\}, \{b, c\} \}$

$\{ \{b\}, \{a, c\} \}$

$\{ \{c\}, \{a, b\} \}$

$\{ \{a, b, c\} \}$

- The number of combinations for  $n$  (3 in our example) is given by the so-called *Bell* number, for details see

[https://en.wikipedia.org/wiki/Bell\\_number](https://en.wikipedia.org/wiki/Bell_number)

# The Bell Numbers

- $n:= 1 \rightarrow 1$
- $n:= 2 \rightarrow 2$
- $n:=3 \rightarrow 5$
- $n:= 4 \rightarrow 15$
- $n:= 5 \rightarrow 52$
- $n:= 6 \rightarrow 203$
- $n:= 7 \rightarrow 877$
- etc...

# What do we need?

- Apart from our usual suspect parameters (tree topology, branch lengths, stationary frequencies, substitution rates,  $a$ ), we also want to integrate over different models now ...
- What are the problems we need to solve?

# What do we need?

- Apart from our usual suspect parameters (tree topology, branch lengths, stationary frequencies, substitution rates,  $\alpha$ ), we also want to integrate over different models now ...
- What are the problems we need to solve?
  - Problem #1: we need to design proposals for moving between different models
  - Problem #2: those models have different numbers of parameters, we can not directly compare likelihoods
- Here we use MCMC to not only sample model parameters, **but also** models

# Problem #1

## Model Proposals

- Any ideas?

# Problem #1

## Model Proposals

- Split move

Chose a set of substitution rates with  $> 1$  member at random

111222 (two-parameter model)

and split it randomly into two rates

111223 (three-parameter model)

- Merge move

Chose two substitution rate sets at random

111223

and merge them into one substitution rate set

111222

# Problem #1

## Model Proposals

- Split move

Chose a set of substitution rates with  $> 1$  member at random

111222 (two-parameter model)

and split it randomly in

111223 (three-param

Clear to everyone what the respective rate matrix looks like?

- Merge move

Chose two substitution rate sets at random

111223

and merge them into one substitution rate set

111222



# Problem #2

## Sampling Different Models

- Use reversible jump MCMC (rjMCMC) to jump between models (posterior probability distributions) with different number of parameters (posterior distributions with different dimensions)
- The model proposal moves we designed are reversible jump moves!
- Evidently, we need to somehow modify our proposal ratio calculation ...
- In general terms, the acceptance ratio is calculated as:

$r = \text{likelihood ratio} * \text{prior ratio} * \text{proposal ratio} * \text{Jacobian}$

A Jacobian defines a linear map from  $R^n \rightarrow R^m$  at point  $x$ , if function  $f(x)$  is differentiable at  $x$

# Problem #2

## Sampling Different Models

- Use reversible jump MCMC (rjMCMC) to jump between models (posterior probability distributions) with different number of parameters (posterior distributions with different dimensions)
- The model proposal moves we designed are reversible jump moves!
- Evidently, we need to somehow modify our proposal ratio calculation ...
- In general terms, the acceptance ratio is calculated as:

$r = \text{likelihood ratio} * \text{prior ratio} * \text{proposal ratio} * \text{Jacobian}$

I will not provide further Details; see work by Peter Green (1995, 2003) who developed the rjMCMC methods

# rjMCMC - summary

- Need to design moves that can jump back and forth between models of different dimensions (parameter counts)
- Need to extend acceptance ratio calculation to account for jumps between different models
- The posterior probability of a specific model (e.g., *JC* or *GTR*) is calculated as the fraction of time (fraction of samples) the MCMC chain visited/spent time/generations sampling within that model ...