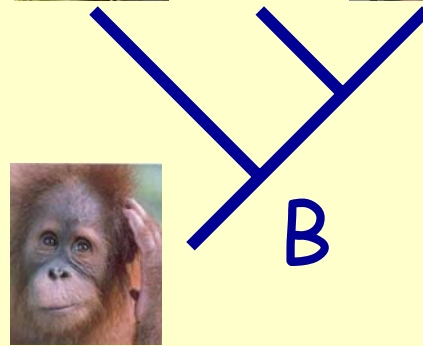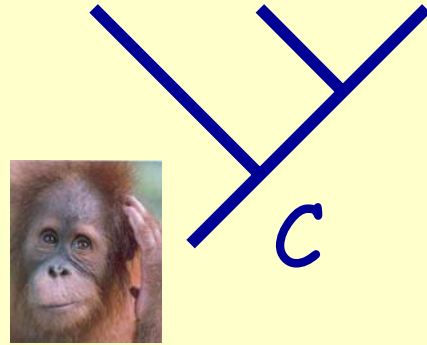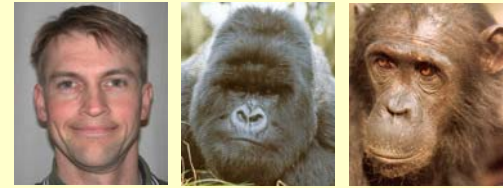# Infer relationships among three species:
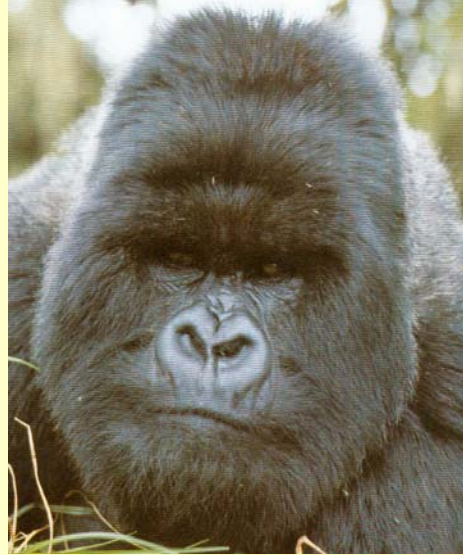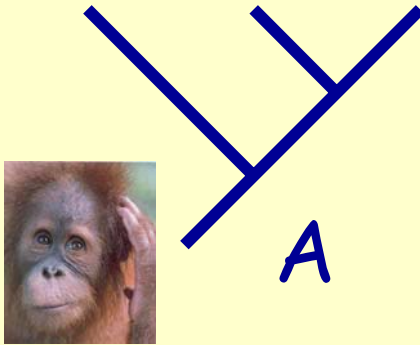
Outgroup:

Three possible trees (topologies):

A

B

C

Model

probability

1.0

Prior distribution

Data (observations)

probability

1.0
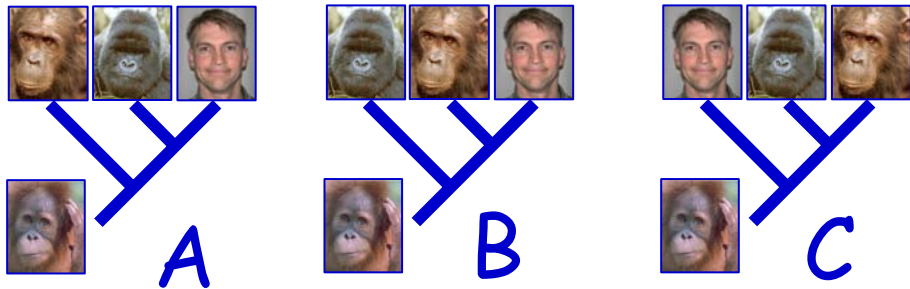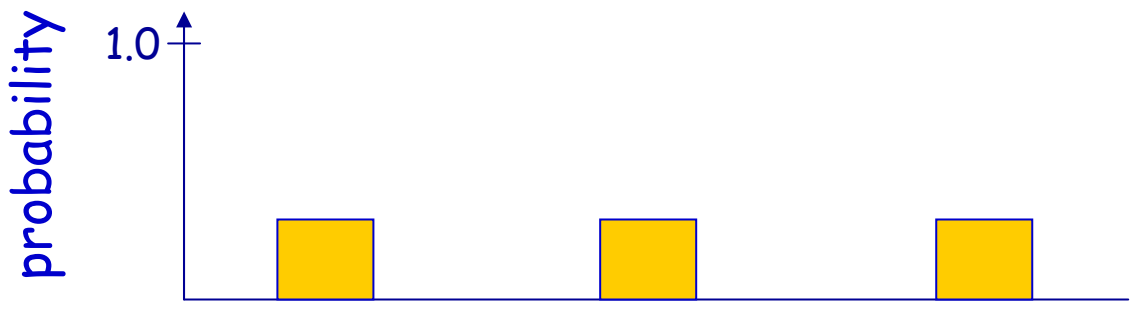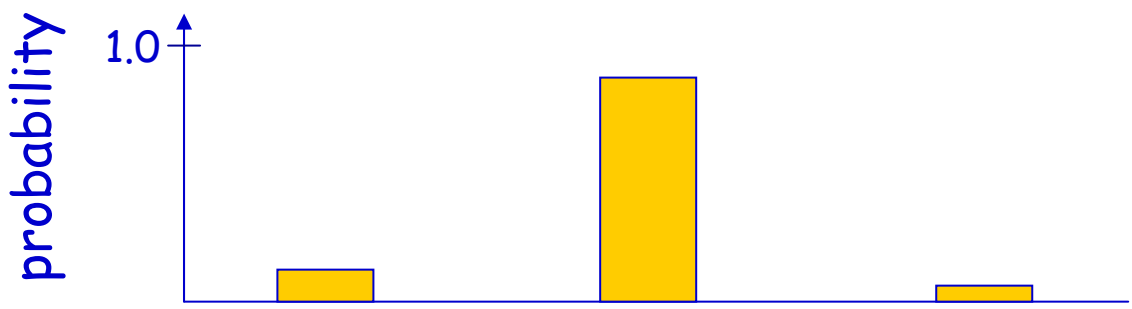
Posterior distribution

Bayes' theorem

Posterior distribution

Prior distribution

"Likelihood"

$$f(\theta \mid X) = \frac{f(\theta) f(X \mid \theta)}{\int f(\theta) f(X \mid \theta) d\theta}$$

Normalizing constant

Model: topology AND branch lengths

$$\theta \quad \text{Parameters}$$
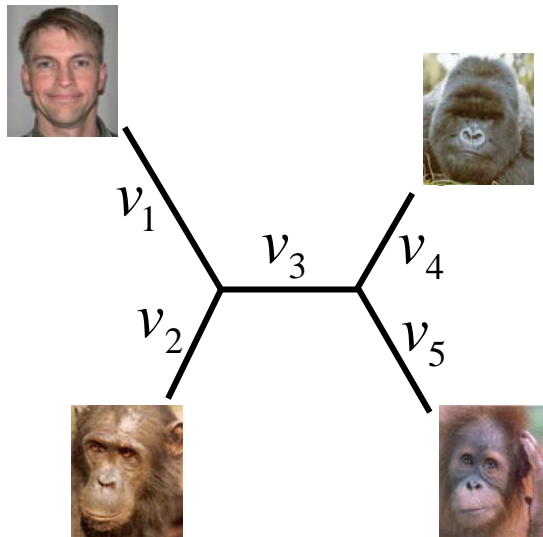
topology (τ)

branch lengths (v_i)
(expected amount of change)

$$\theta = (\tau, v)$$

Posterior probability distribution
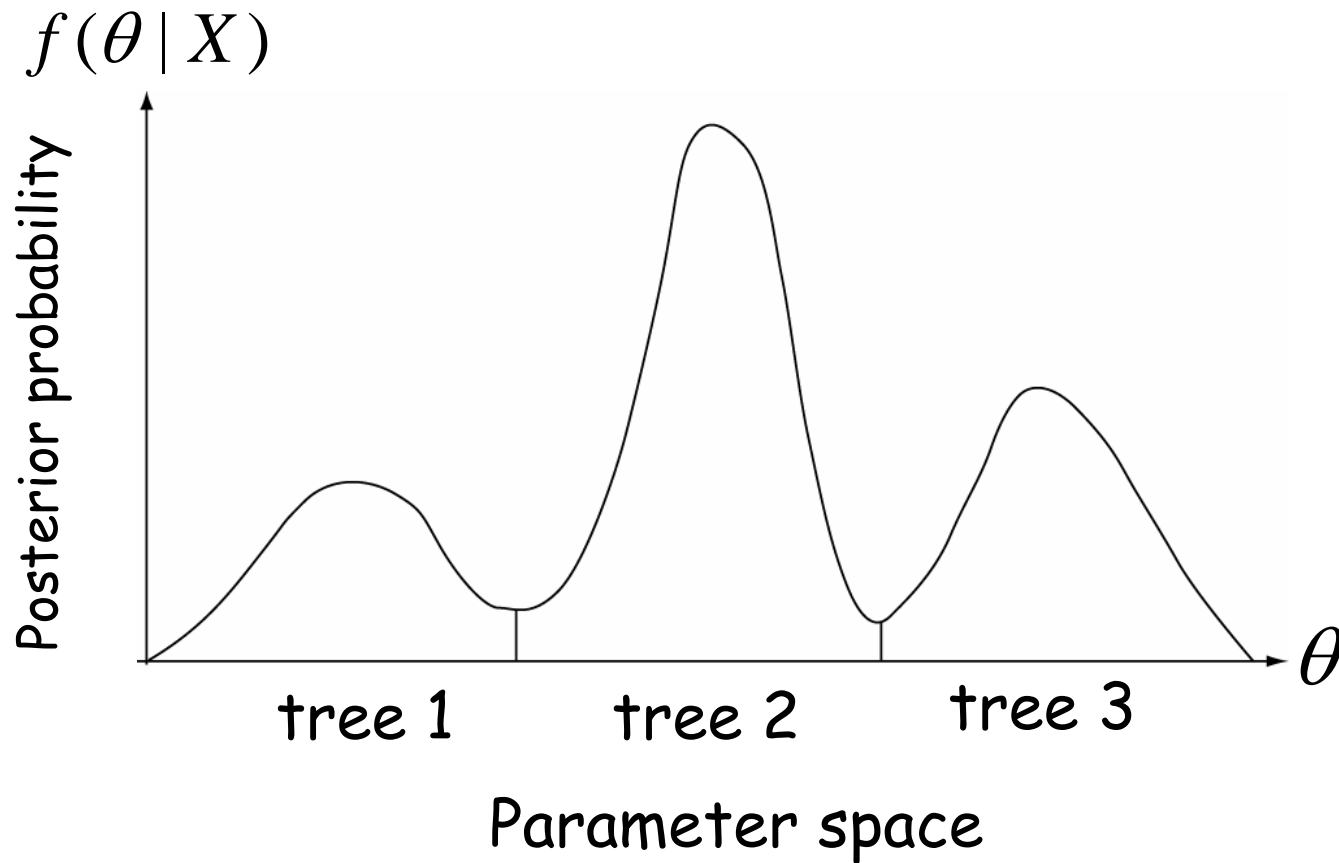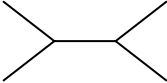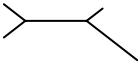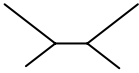
We can focus on any parameter of interest (there are no nuisance parameters) by marginalizing the posterior over the other parameters (integrating out the uncertainty in the other parameters)

20% 48% 32%

tree 1 tree 2 tree 3
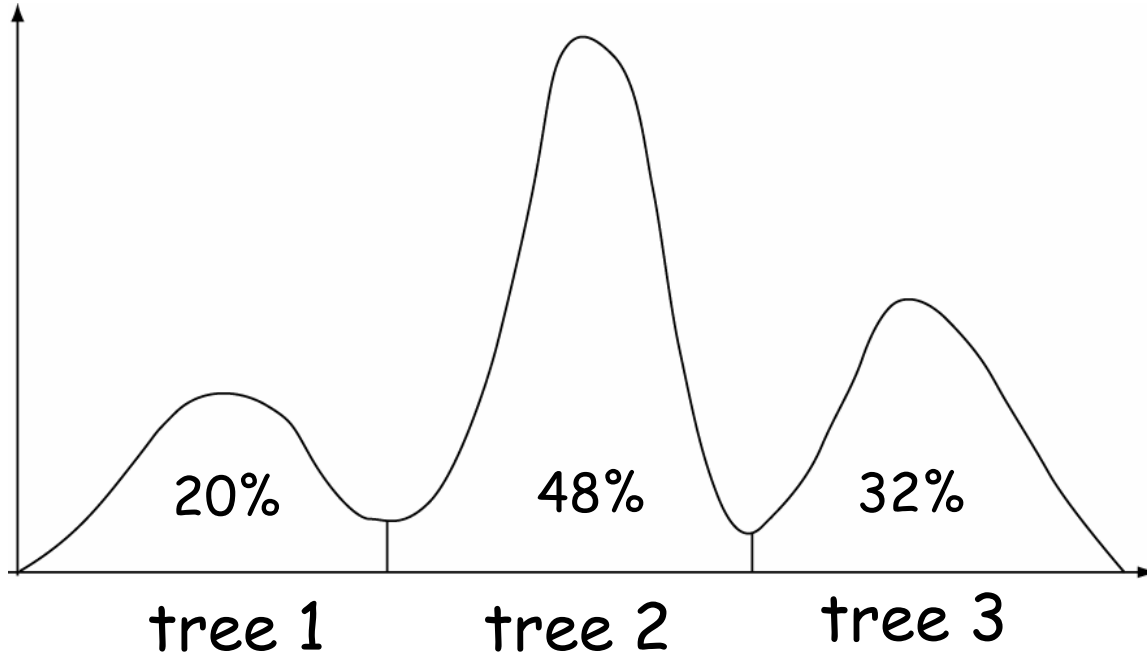
(Percentages denote marginal probability distribution on trees)

Why is it called marginalizing?

trees

joint probabilities

	τ_1	τ_2	τ_3	
ν^1	0.10	0.07	0.12	0.29
ν^2	0.05	0.22	0.06	0.33
ν^3	0.05	0.19	0.14	0.38
	0.20	0.48	0.32	

branch length vectors

marginal probabilities

Markov chain Monte Carlo

1. Start at an arbitrary point
2. Make a small random move
3. Calculate height ratio (r) of new state to old state:
    1. r > 1 -> new state accepted
    2. r < 1 -> new state accepted with probability r. If new state not accepted, stay in the old state
4. Go to step 2

always accept

2a

1

2b   accept sometimes

20 %   48 %   32 %

tree 1   tree 2   tree 3
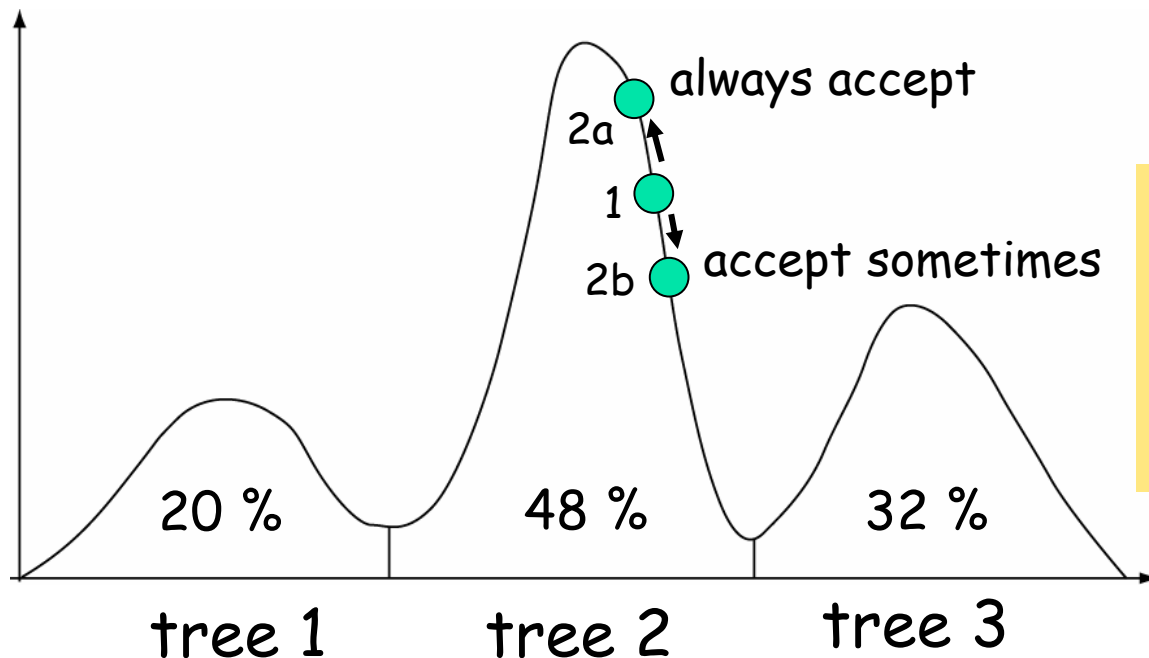
The proportion of time the MCMC procedure samples from a particular parameter region is an estimate of that region's posterior probability density

stationary phase sampled with thinning
(rapid mixing essential)

burn-in

InL

Generation
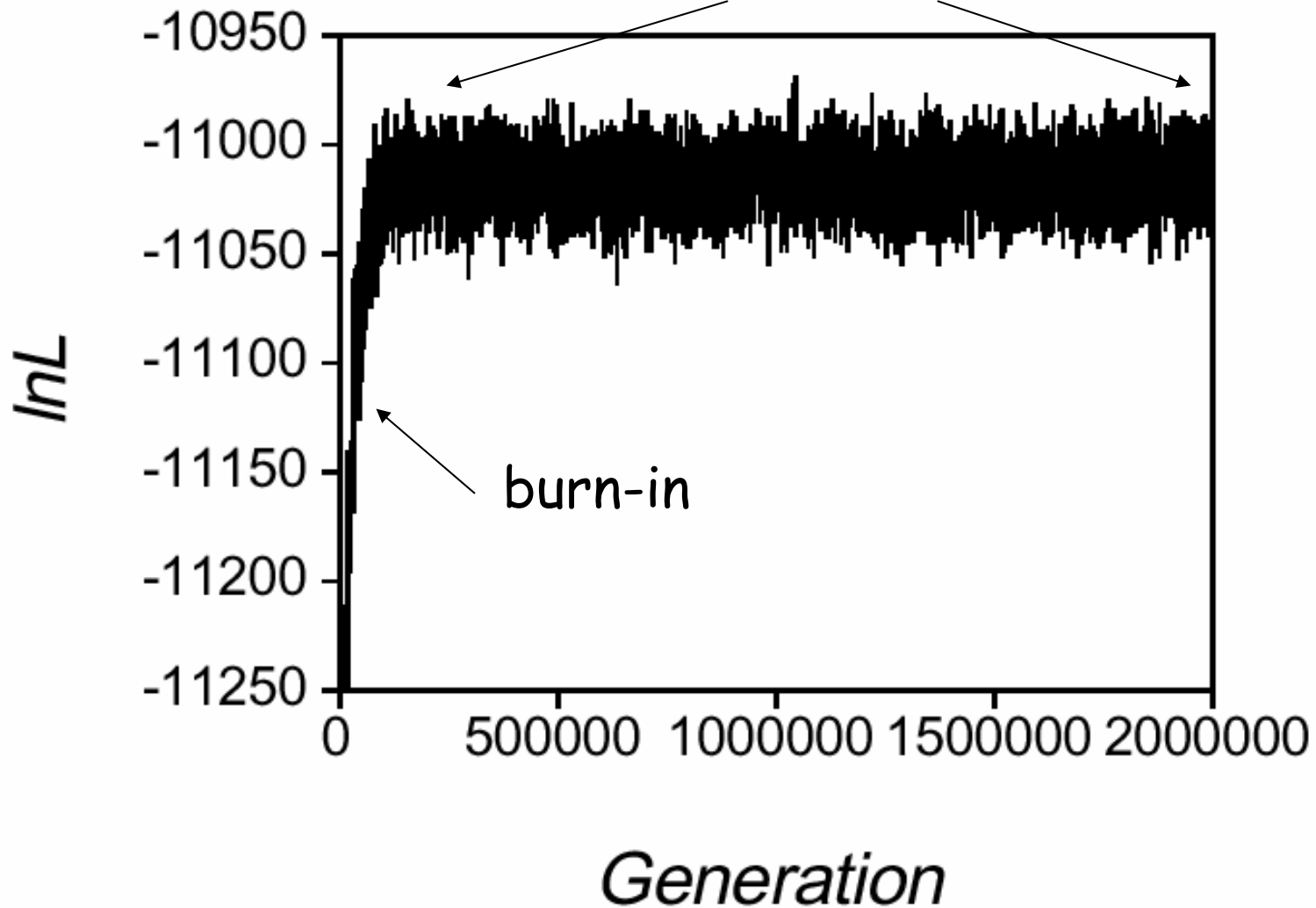
Metropolis-
coupled
Markov chain
Monte Carlo

a. k. a.

MCMCMC

a. k. a.

$(MC)^3$
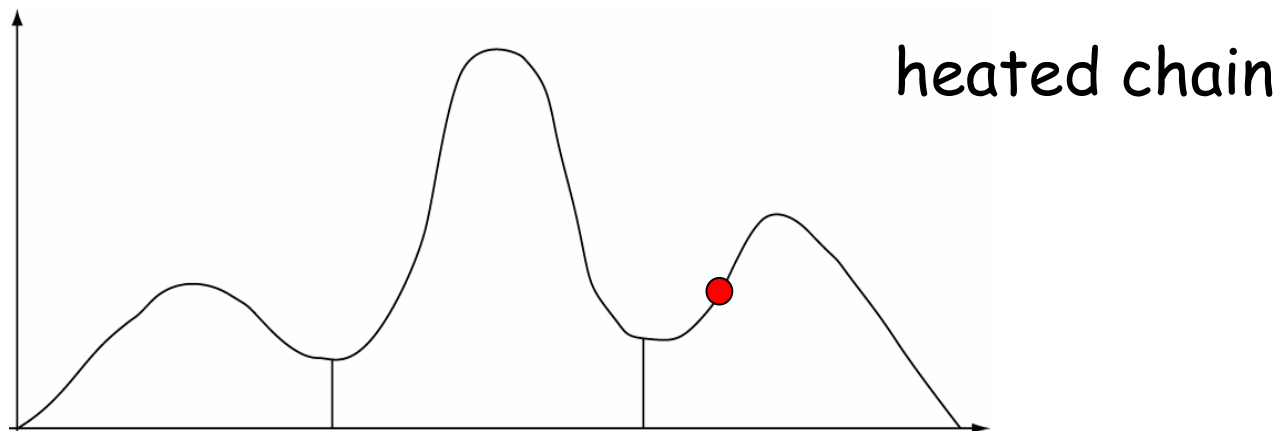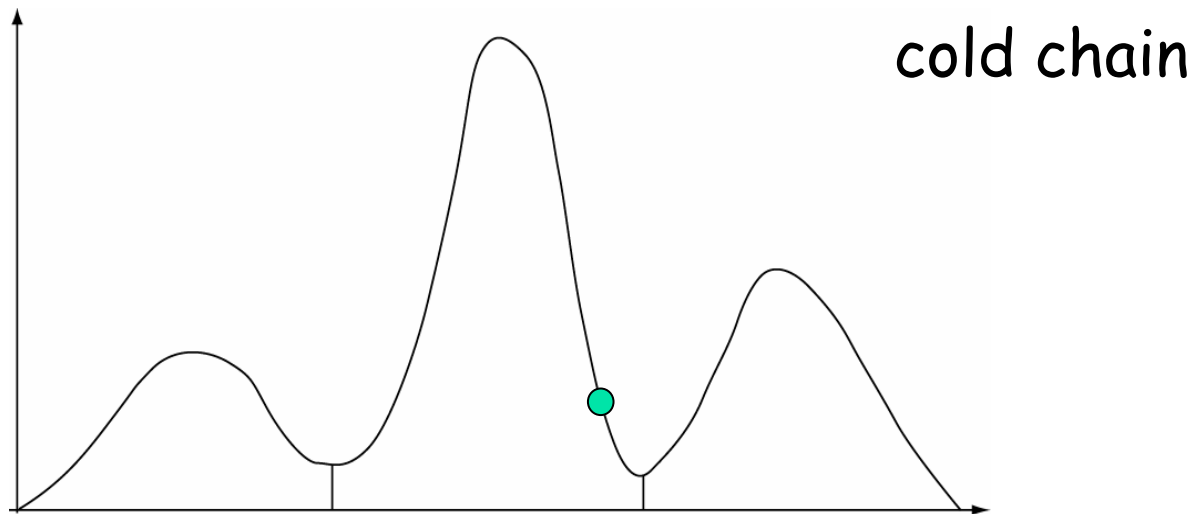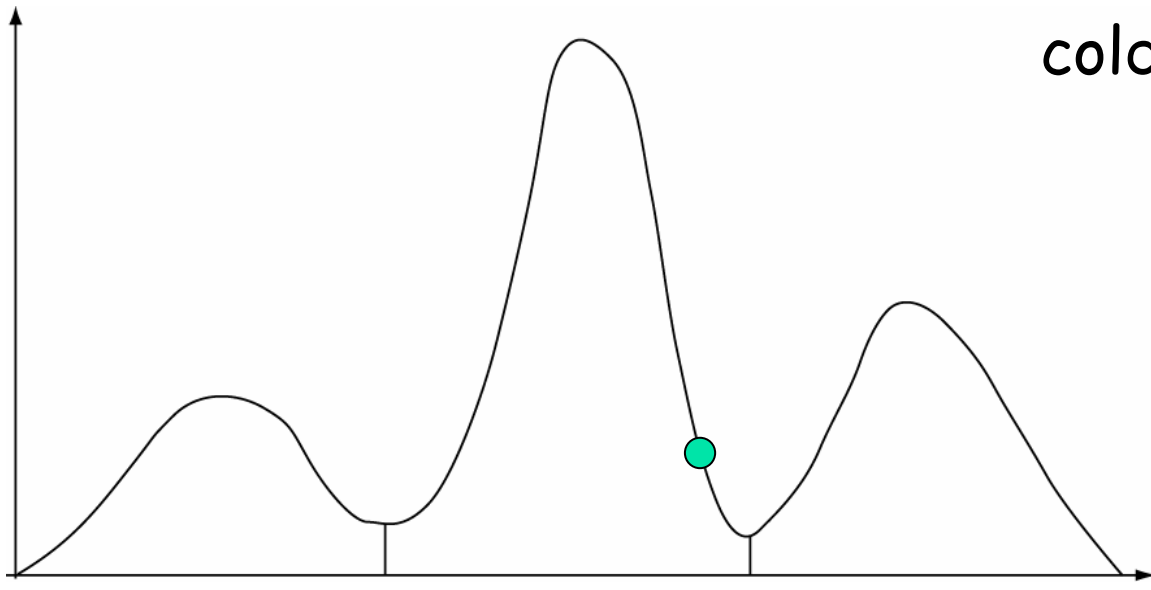
cold chain
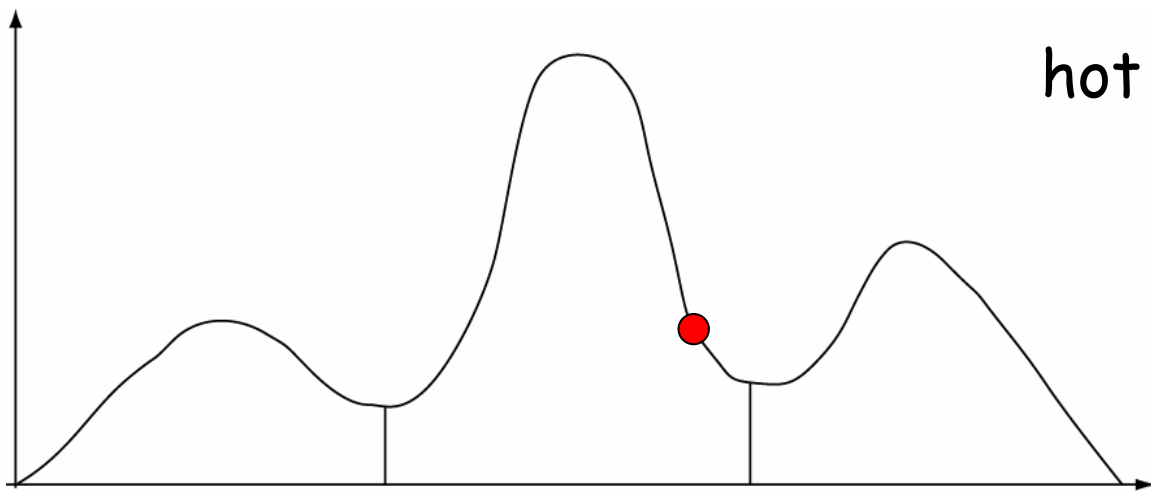
heated chain
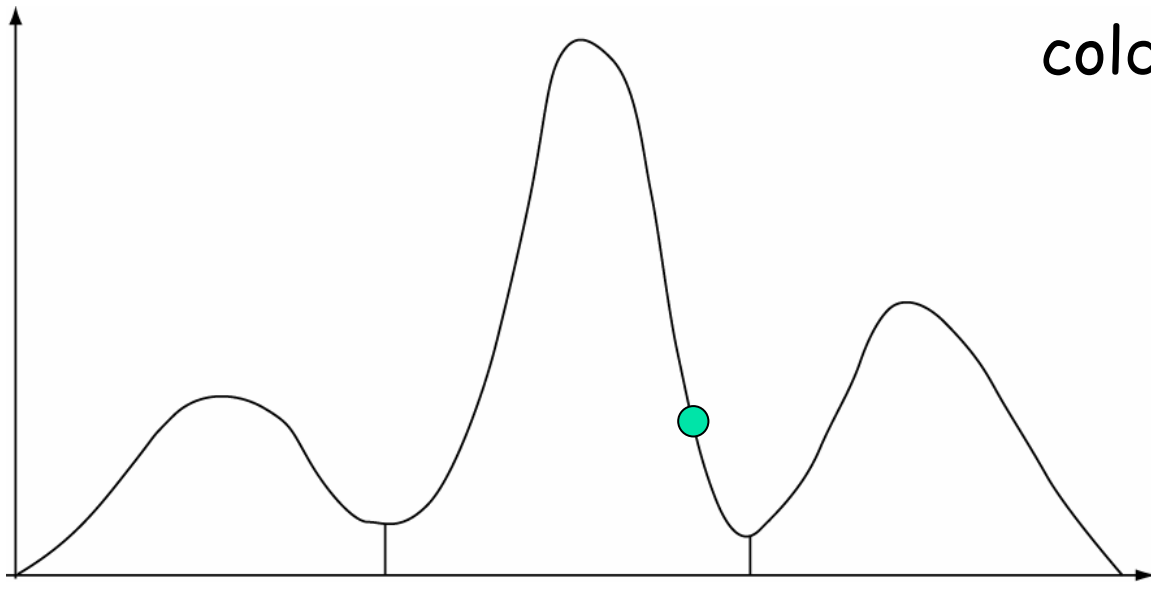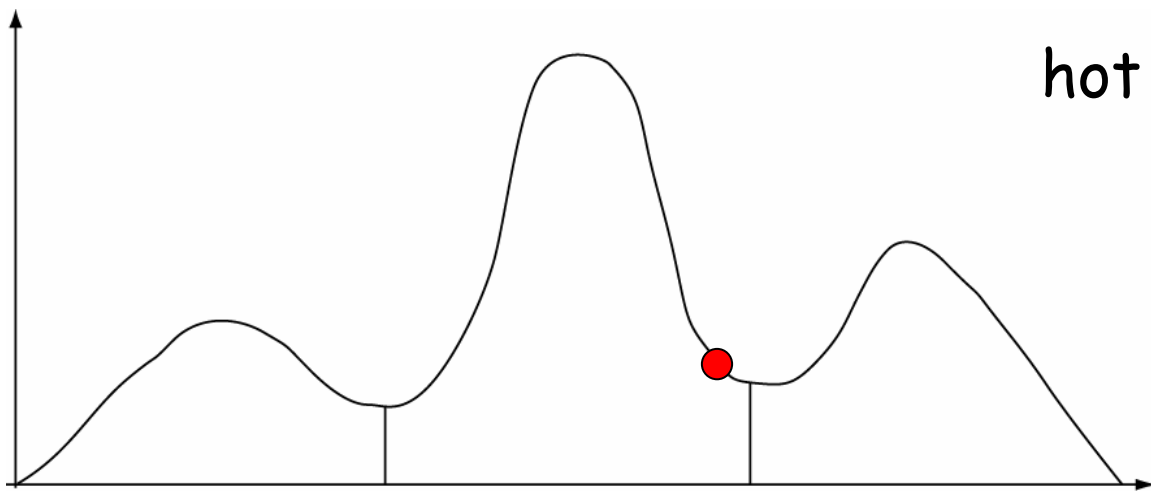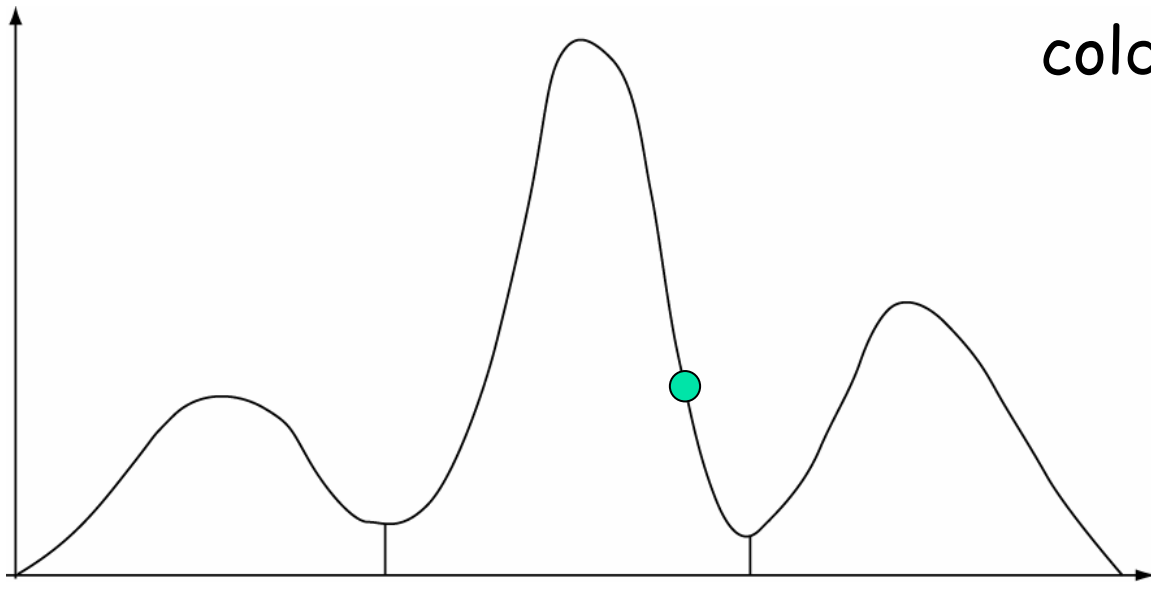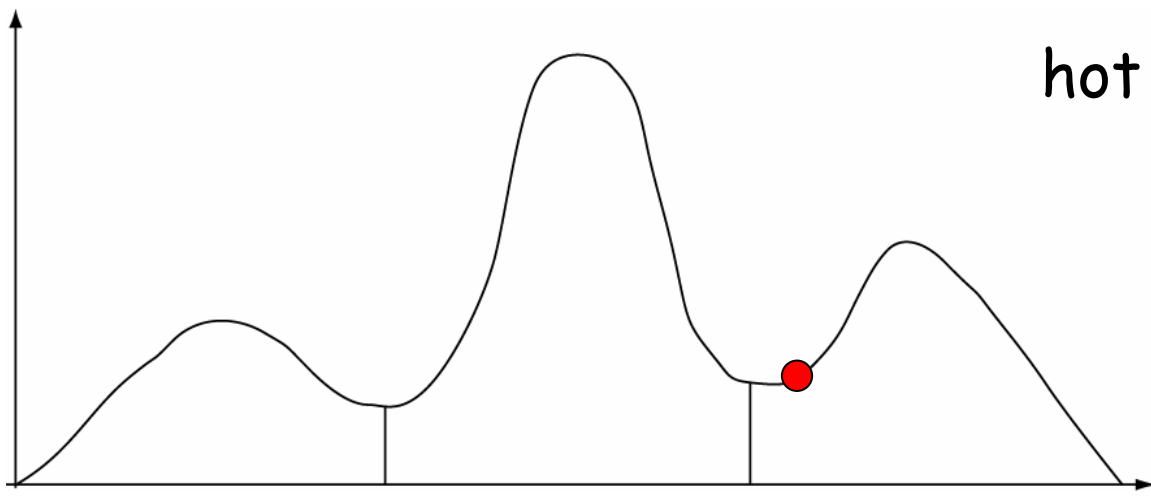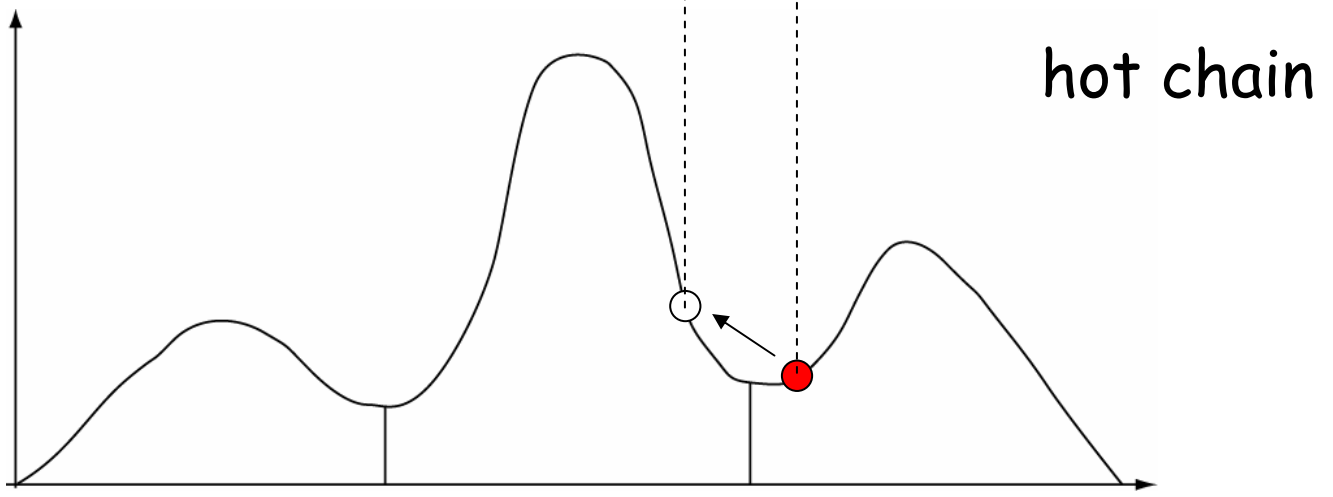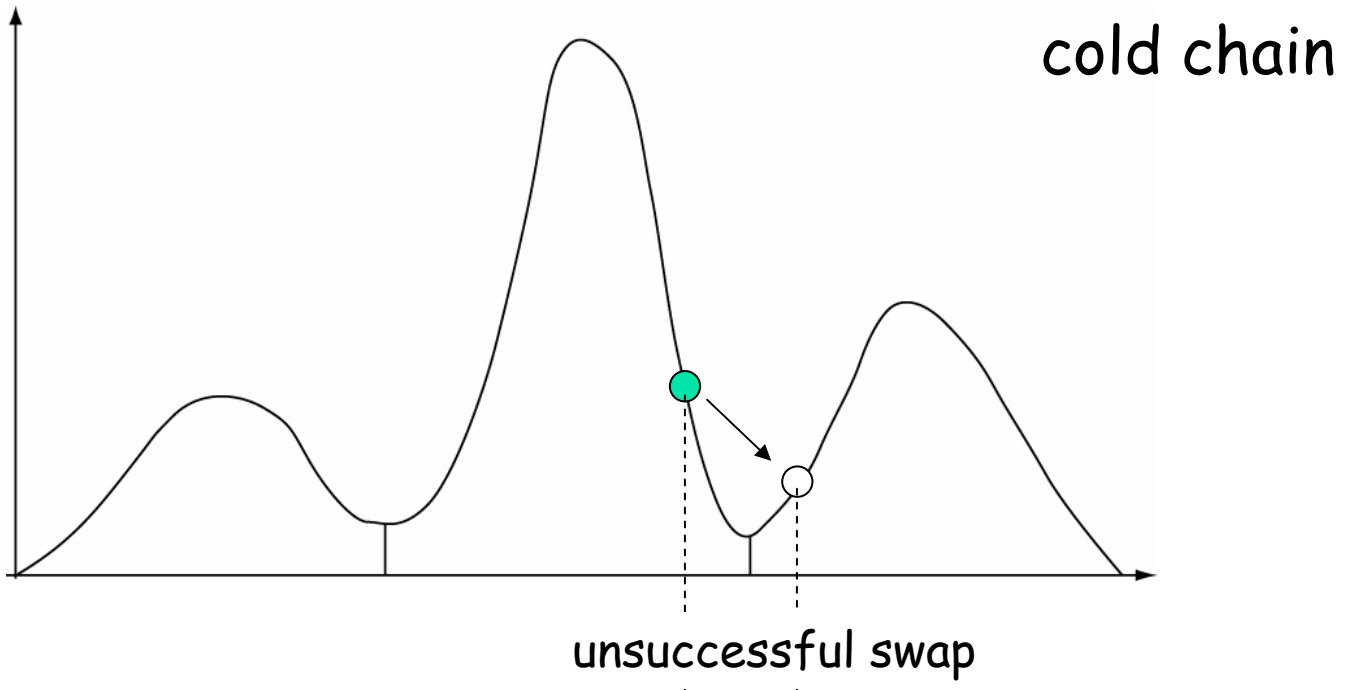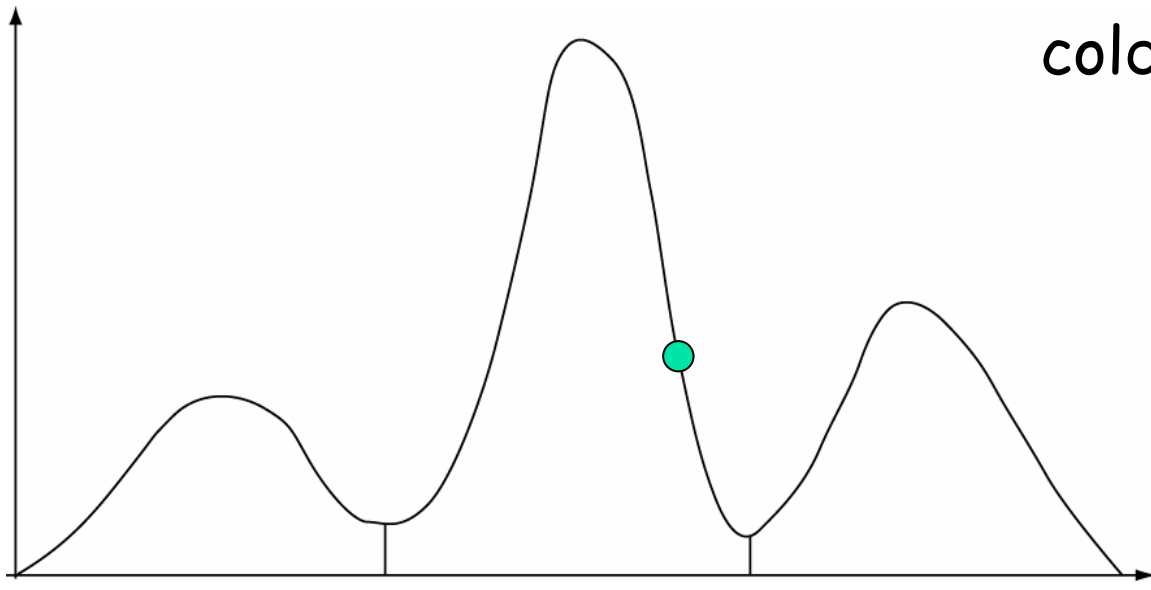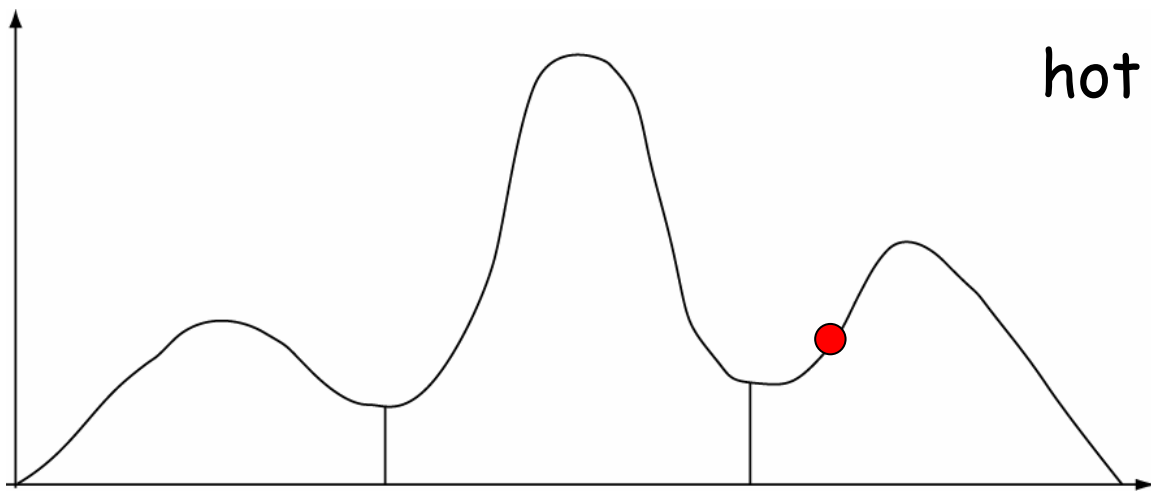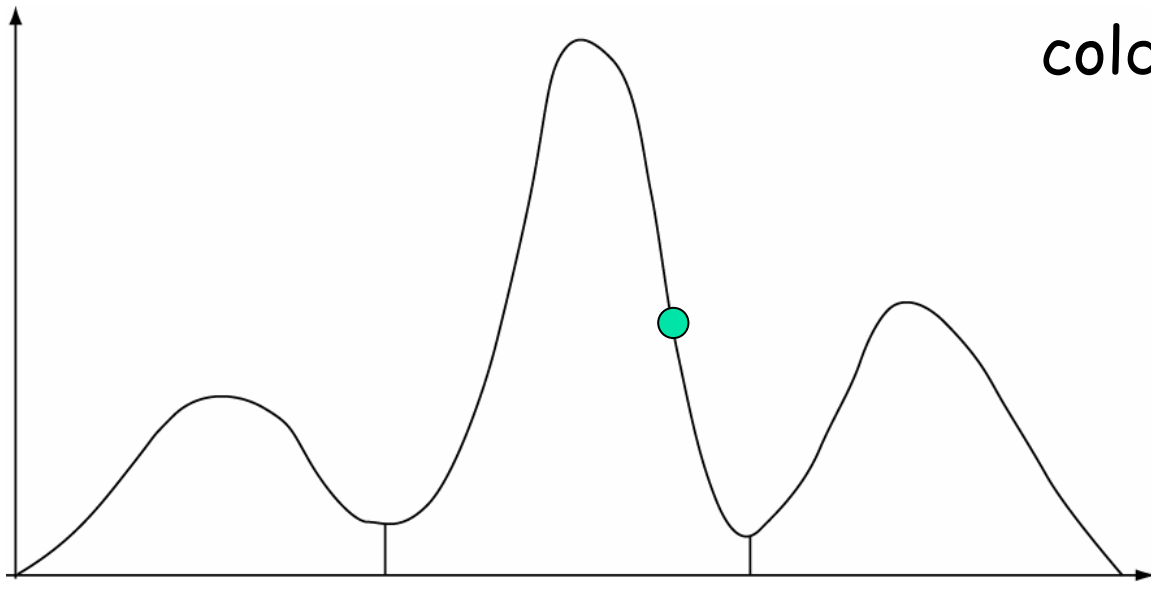
cold chain

hot chain

cold chain

hot chain

cold chain

hot chain

cold chain

unsuccessful swap

hot chain

cold chain

hot chain
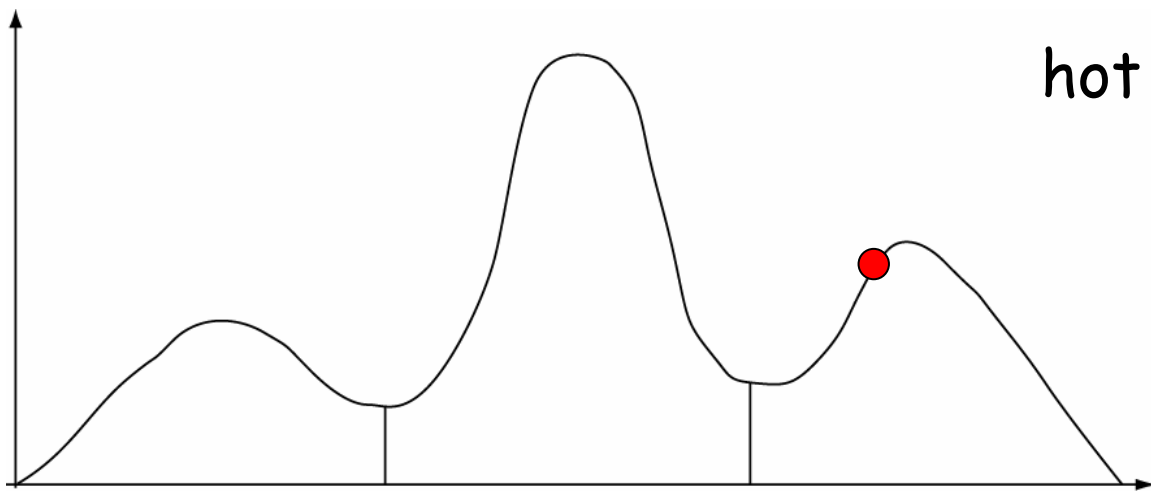
cold chain
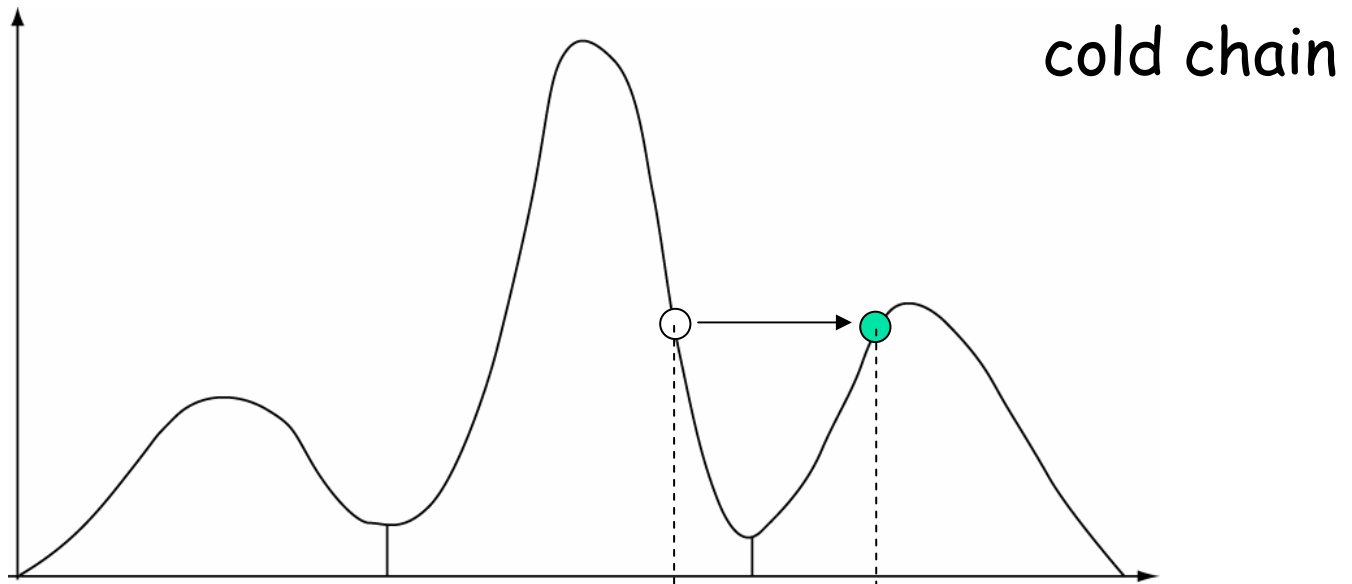
hot chain

cold chain
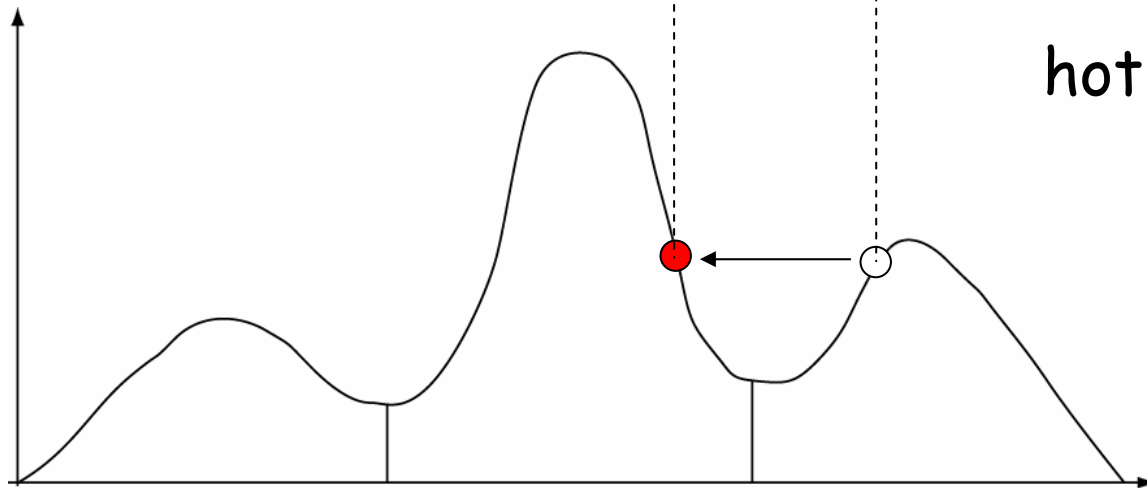
successful swap
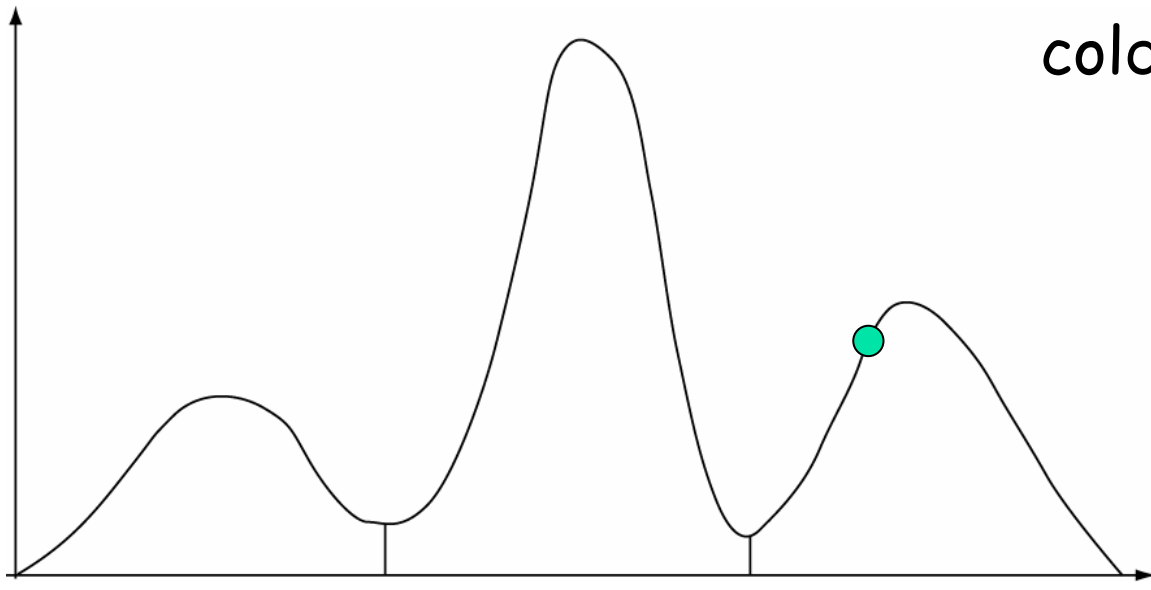
hot chain
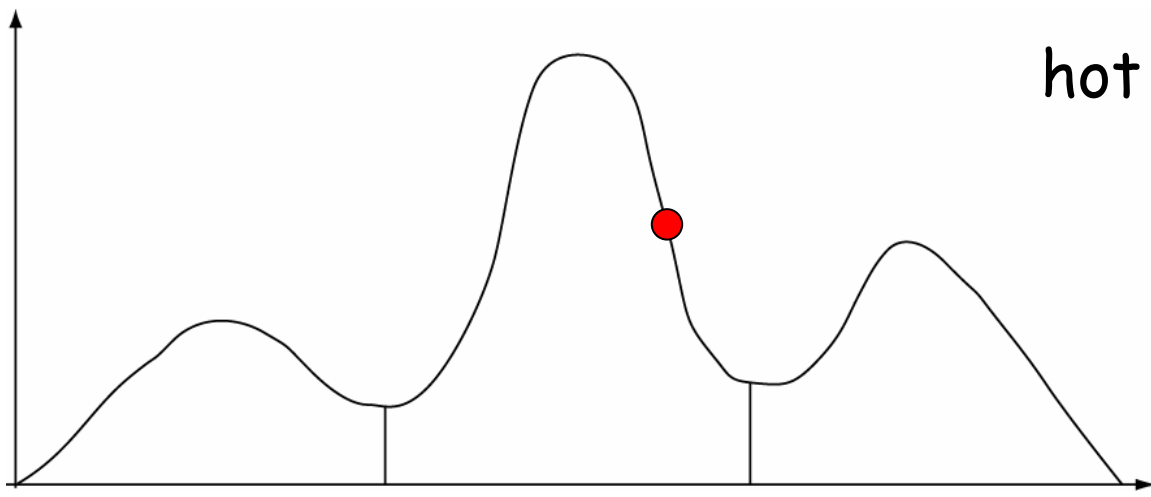
cold chain
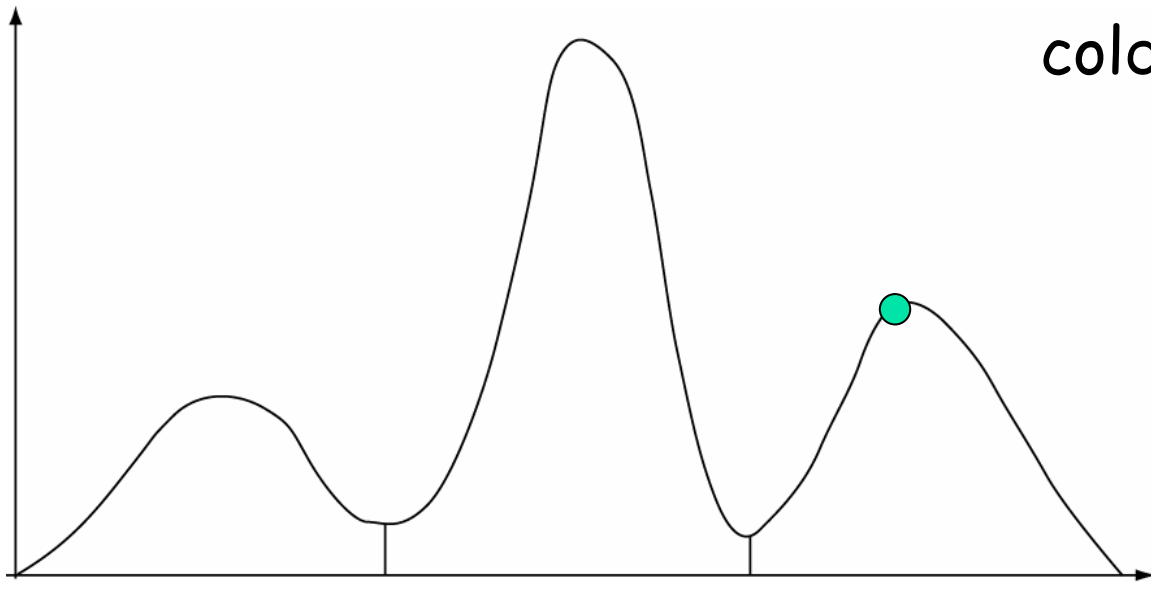
hot chain
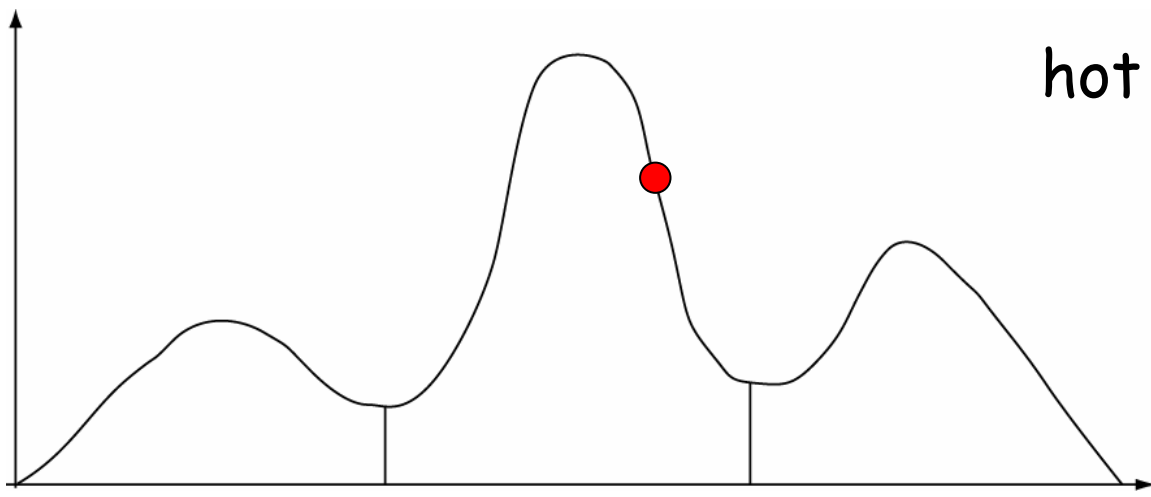
cold chain
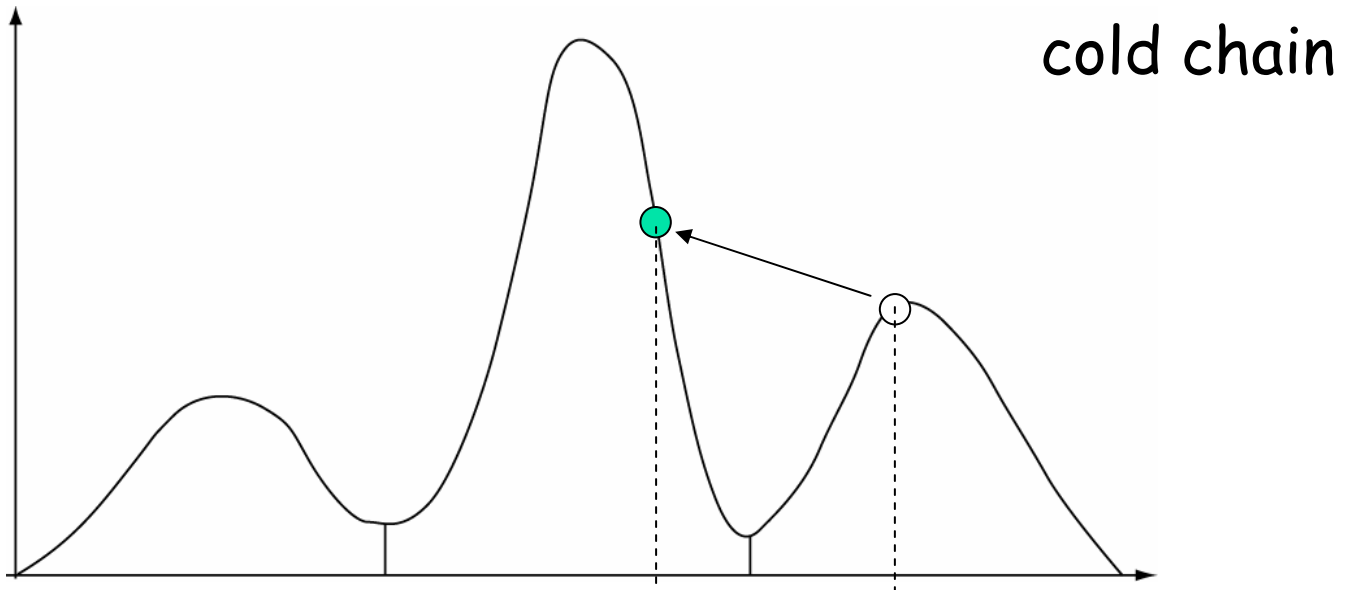
hot chain

cold chain

successful swap
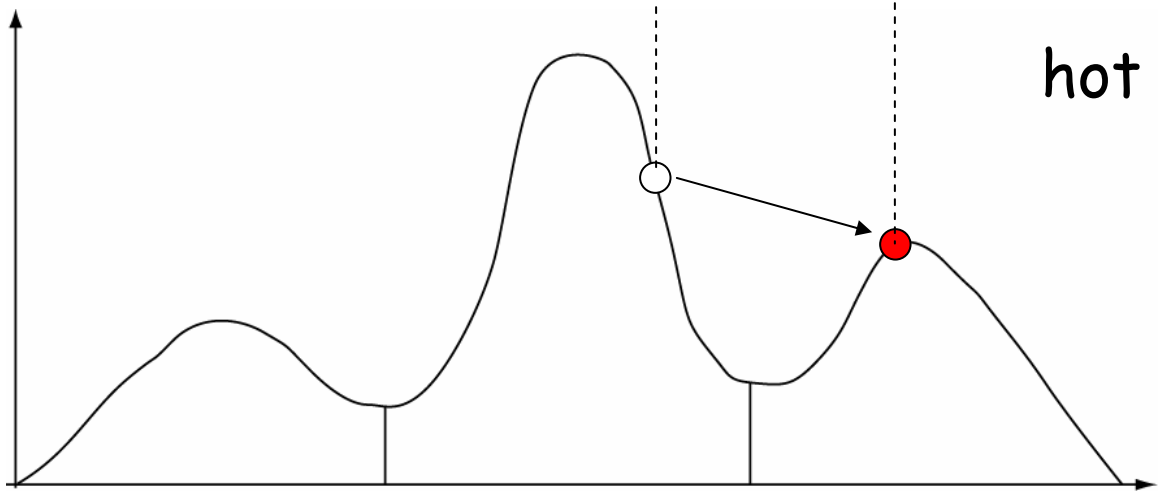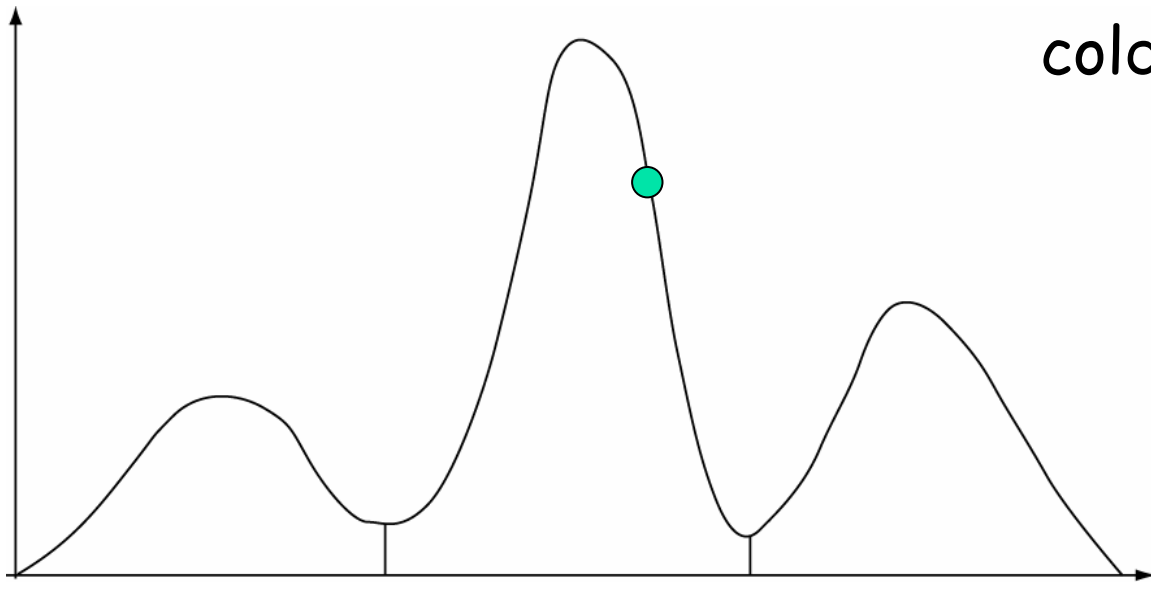
hot chain
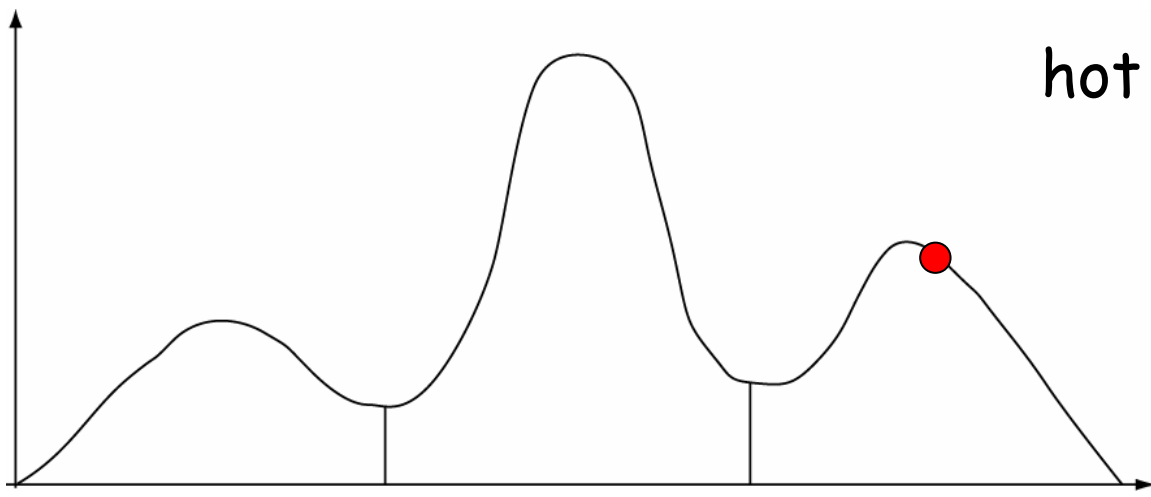
cold chain

hot chain

Incremental Heating

T is temperature, λ is heating coefficient

$$T = 1/(1 + \lambda i) \qquad i = \{0, 1, ..., n-1\}$$

Example for λ = 0.2:

i	T	Distr.	
0	1.00	$f(\theta \mid X)^{1.00}$	⟵ cold chain
1	0.83	$f(\theta \mid X)^{0.83}$	
2	0.71	$f(\theta \mid X)^{0.71}$	heated chains
3	0.62	$f(\theta \mid X)^{0.62}$	

Assessing Convergence

- Plateau in the trace plot (unreliable!!)
- Compare windows within the same run (better)
- Compare independent runs starting from different randomly chosen topologies (best)
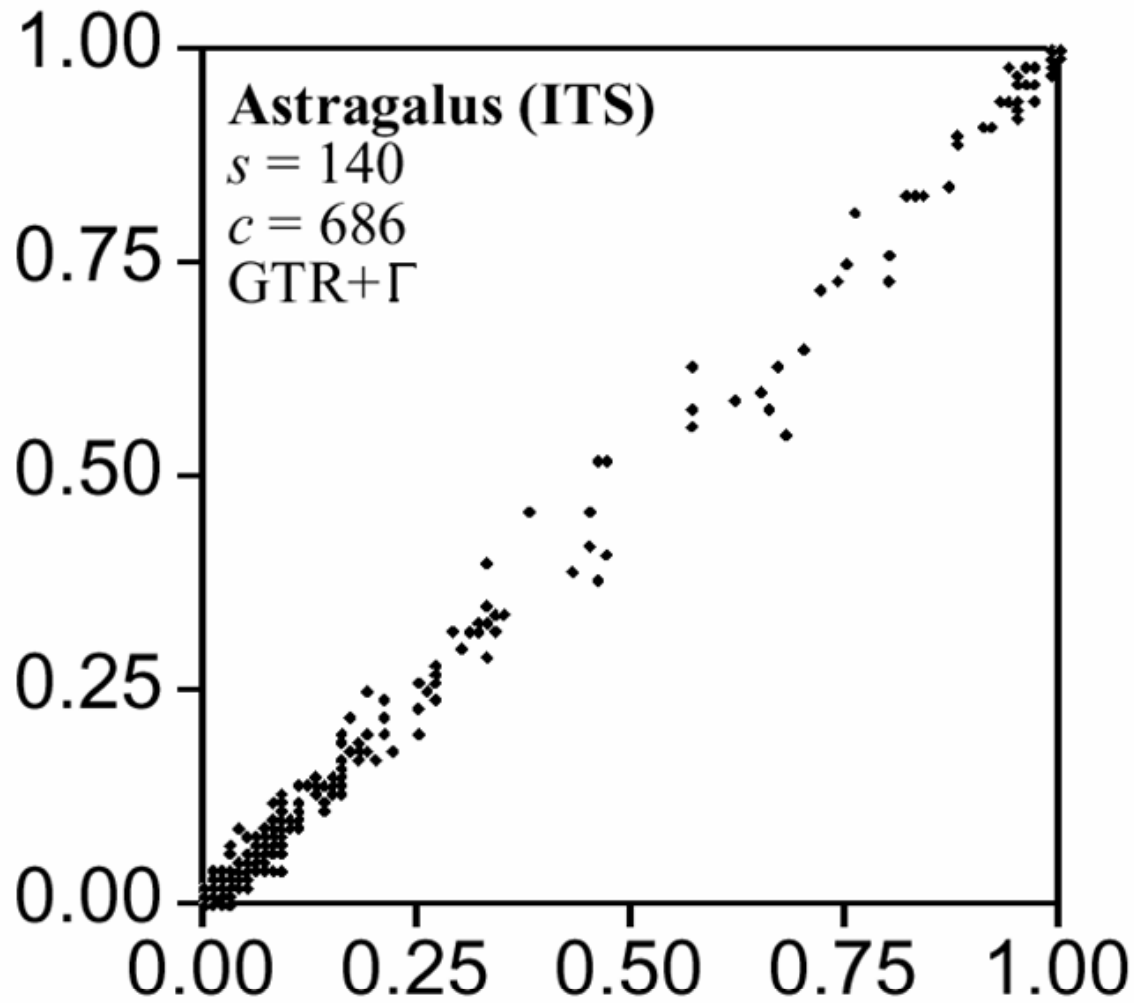
Axis labels:
- Y-axis: Clade probability in analysis 2
- X-axis: Clade probability in analysis 1

Plot annotation:
Astragalus (ITS)
$s = 140$
$c = 686$
GTR+Γ

Angiosperms (atpB)
$s = 357$
$c = 1497$
GTR+SS

X-axis: Clade probability in analysis 1
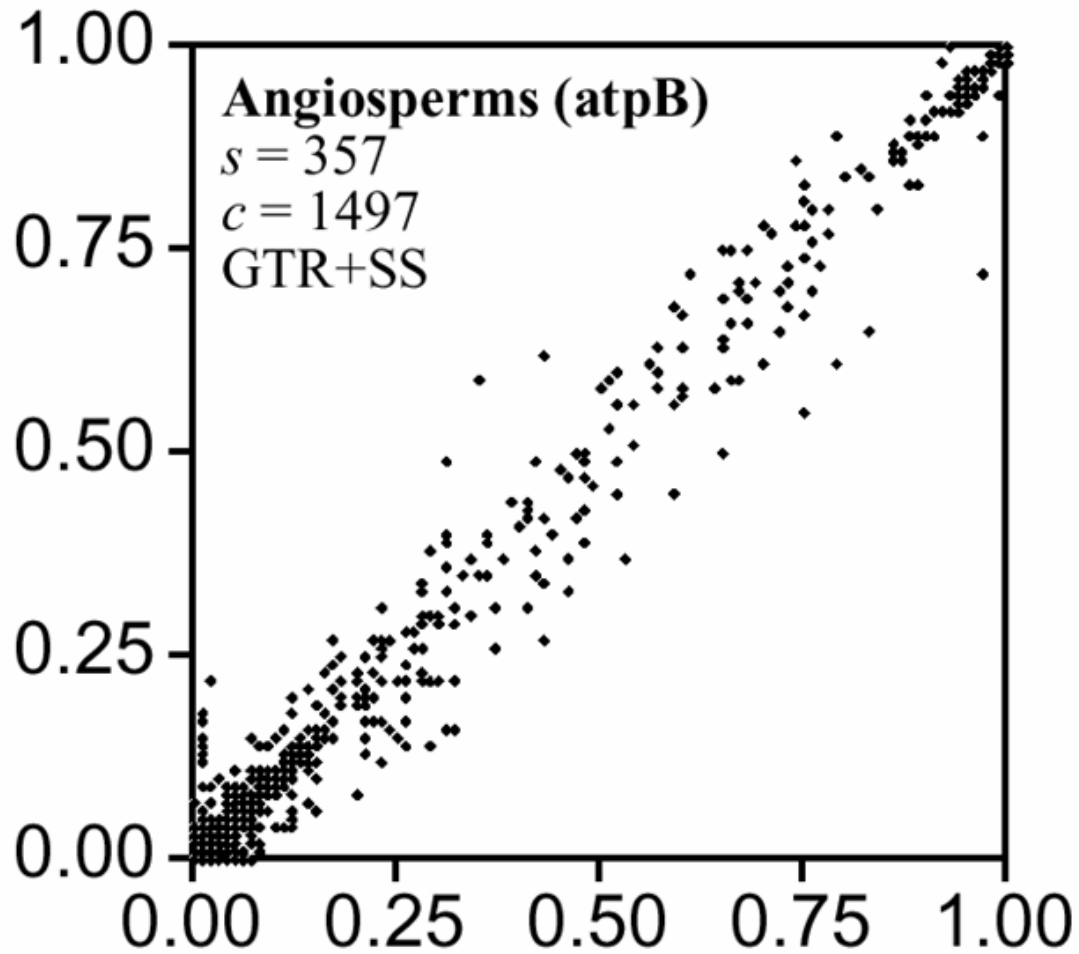
Y-axis: Clade probability in analysis 2

Improving Convergence

(Only if convergence diagnostics indicate problem!)

- Change tuning parameters of proposals to bring acceptance rate into the range 10 % to 70 %
- Propose changes to 'difficult' parameters more often
- Use different proposal mechanisms
- Change heating temperature to bring acceptance rate of swaps between adjacent chains into the range 10 % to 70 %.
- Run the chain longer
- Increase the number of heated chains
- Make the model more realistic
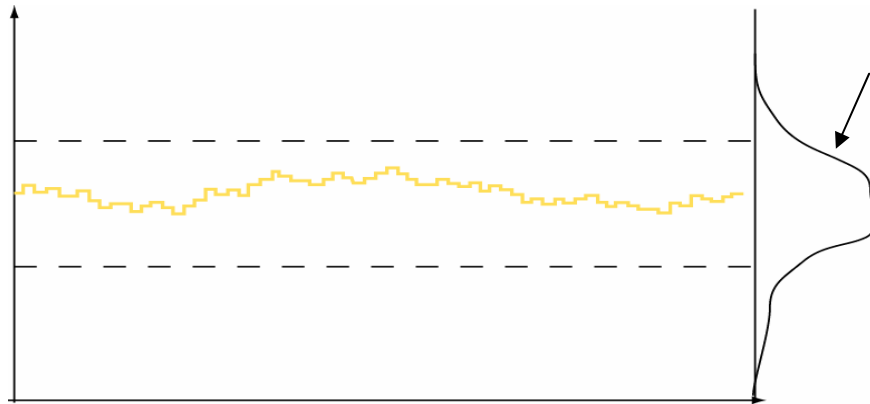
Target distribution

Too modest proposals
Acceptance rate too high
Poor mixing
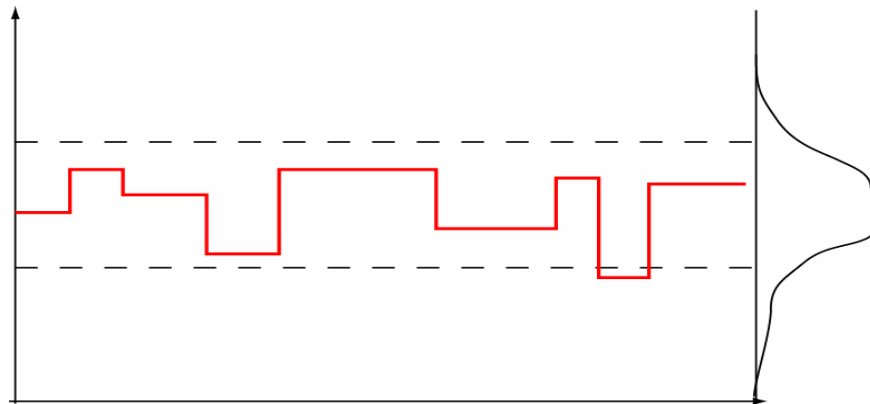
Too bold proposals
Acceptance rate too low
Poor mixing
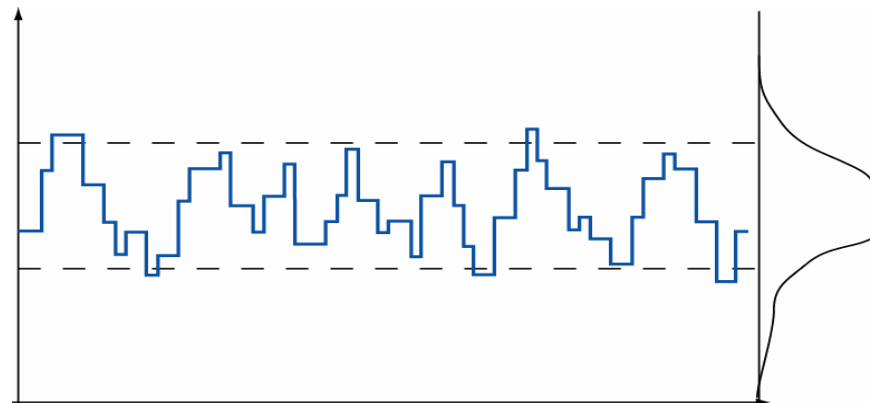
Moderately bold proposals
Acceptance rate intermediate
Good mixing

Sampled value

Running MrBayes

- Use **execute** to bring data in a Nexus file into MrBayes
- Set the model and priors using **lset** and **prset**
- Run the chain using **mcmc**
- Summarize the parameter samples using **sump**
- Summarize the tree samples using **sumt**
- Note that MrBayes 3.1 runs **two independent analyses** by default

Ver. 3.1 Convergence Diagnostics
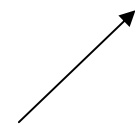
- By default performs two independent analyses starting from different random trees (**mcmc nruns=2**)
- Average standard deviation of clade frequencies calculated and presented during the run (**mcmc mcmcdiagn=yes diagnfreq=1000**) and written to file (.mcmc)
- Standard deviation of each clade frequency and potential scale reduction for branch lengths calculated with **sumt**
- Potential scale reduction calculated for all substitution model parameters with **sump**

Bayes' theorem

$$f(\theta \mid X) = \frac{f(\theta)f(X \mid \theta)}{\int f(\theta)f(X \mid \theta)d\theta} = \frac{f(\theta)f(X \mid \theta)}{f(X)}$$

Marginal likelihood (of the model)

We have implicitly conditioned on a model:

$$f(\theta \mid X, M) = \frac{f(\theta \mid M)f(X \mid \theta, M)}{f(X \mid M)}$$

Bayesian Model Choice

Posterior model odds:

$$\frac{f(M_1)f(X/M_1)}{f(M_0)f(X/M_0)}$$

Bayes factor:

$$B_{10} = \frac{f(X/M_1)}{f(X/M_0)}$$

Bayesian Model Choice

- The normalizing constant in Bayes' theorem, the marginal probability of the model, $f(X)$ or $f(X|M)$, can be used for model choice
- $f(X|M)$ can be estimated by taking the harmonic mean of the likelihood values from the MCMC run (MrBayes will do this automatically with 'sump')
- Any models can be compared: nested, non-nested, data-derived
- No correction for number of parameters
- Can prefer a simpler model over a more complex model
- Critical values in Kass and Raftery (1997)

Bayes Factor Comparisons

Interpretation of the Bayes factor

$2\ln(B_{10})$	B_{10}	Evidence against M_0
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
> 10	> 150	Very strong

Resources

- MrBayes web site (www.mrbayes.net)
- MrWiki manual (on the MrBayes web site)
- MrBayes 3.1 command reference
- MrBayes 3.1 manual (pdf) with two tutorials: a simple analysis (primates.nex) and an analysis of partitioned data (cynmix.nex)
- Last pages of the manual contain graphical summaries of the MrBayes 3 models and most common types of proposal mechanisms
- MrBayes 3 chapter: complex partitioned analysis (kim.nex)

Models supported by MrBayes 3 (simplified)

Data Type	State Frequencies (Substitution Rates)	Across-Site Rate Variation	Coding Bias	Misc.
Restriction 0 - 1	fixed/estimated (Dirichlet) *prset statefreqpr*	equal/gamma *lset rates*	all/variable/ nopresencesites/noabsencesites *lset coding*	
Standard 0 - 9	equal/estimated (SymmDir) *prset symdirihyperpr*	equal/gamma *lset rates*	all/variable/informative *lset coding*	unordered/ordered *ctype*

Data Type	Model Type	State Frequencies	Substitution Rates	Across-Site Rate Variation	Across-Tree Rate Variation
DNA A C G T	4by4 *lset nucmodel*	fixed/est. (Dirichlet) *prset statefreqpr*	F81/HKY/GTR *lset nst=1/2/6*	equal/gamma/ propinv/invgamma/ adgamma *lset rates*	yes/no *lset covarion*
	doublet *lset nucmodel*	fixed/est. (Dirichlet) (over 16 states) *prset statefreqpr*	F81/HKY/GTR *lset nst=1/2/6*	equal/gamma/ propinv/invgamma *lset rates*	

Across-Site Omega Variation

Data Type	Model Type	State Frequencies	Substitution Rates	Across-Site Omega Variation
	codon *lset nucmodel*	fixed/est. (Dirichlet) (over 61 states) *prset statefreqpr*	F81/HKY/GTR *lset nst=1/2/6*	equal/Ny98/M3 *lset omegavar*

Data Type	**Model Type**	**State Frequencies**	**Substitution Rates**	**Across-Site Rate Variation**	**Across-Tree Rate Variation**
Protein A - Y	GTR *prset aamodelpr*	fixed/est. (Dirichlet) *prset statefreqpr*	fixed/est. (Dirichlet) *prset statefreqpr*	equal/gamma/ propinv/invgamma/ adgamma *lset rates*	yes/no *lset covarion*
	Equalin *prset aamodelpr*	fixed/est. (Dirichlet) *prset statefreqpr*	fixed to equal	equal/gamma/ propinv/invgamma/ adgamma *lset rates*	yes/no *lset covarion*
	Poisson/Jones/ Dayhoff/Mtrev/ Mtmam/Wag/ Rtrev/Cprev/ Vt/Blossum/ mixed *prset aamodelpr*	fixed/mixed	fixed/mixed	equal/gamma/ propinv/invgamma/ adgamma *lset rates*	yes/no *lset covarion*

Tree Type	**Brlens Type**	**Brlens Prior**		
Non-clock	Unconstrained *prset brlenspr*	Exponential/Uniform *prset brlenspr*		
Clock	Clock *prset brlenspr*	Uniform *prset brlenspr*	Treeheight	
		Coalescence *prset brlenspr*	Theta, Ploidy Growth	
		Birth-Death *prset brlenspr*	Speciation Extinction Sampleprob	

Additional parameters
*see prset
(lset for ploidy)*

Brlens Variation Across Partitions

equal/proportional
prset ratepr

unlinked
unlink brlens

Topology Variation Across Partitions

same/unlinked
*link topology
unlink topology*

Sliding Window Proposal

New values are picked uniformly from a sliding window
of size δ centered on x.
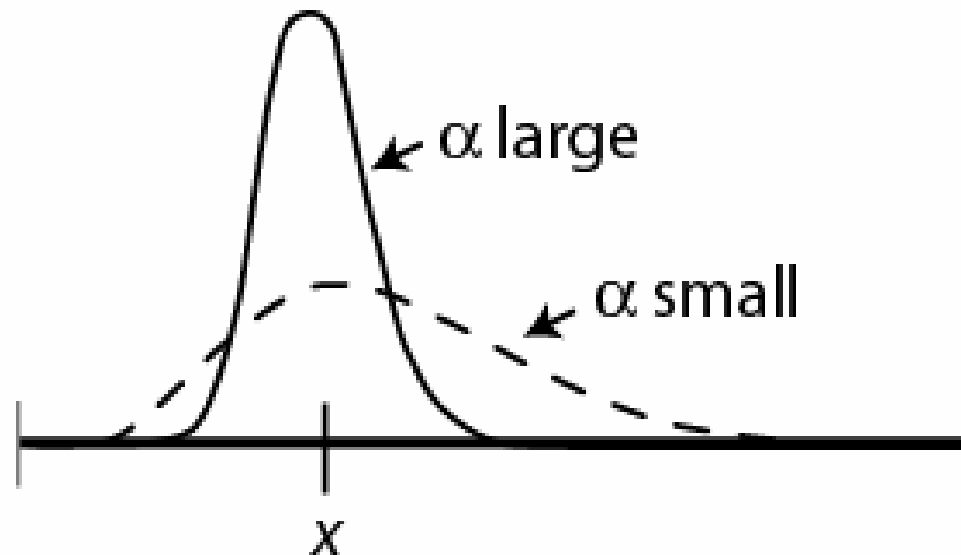
Tuning parameter: δ

Bolder proposals: increase δ

More modest proposals: decrease δ

*Works best when the effect on the probability of the
data is similar throughout the parameter range*

Dirichlet proposal

New values are picked from a Dirichlet (or Beta) distribution
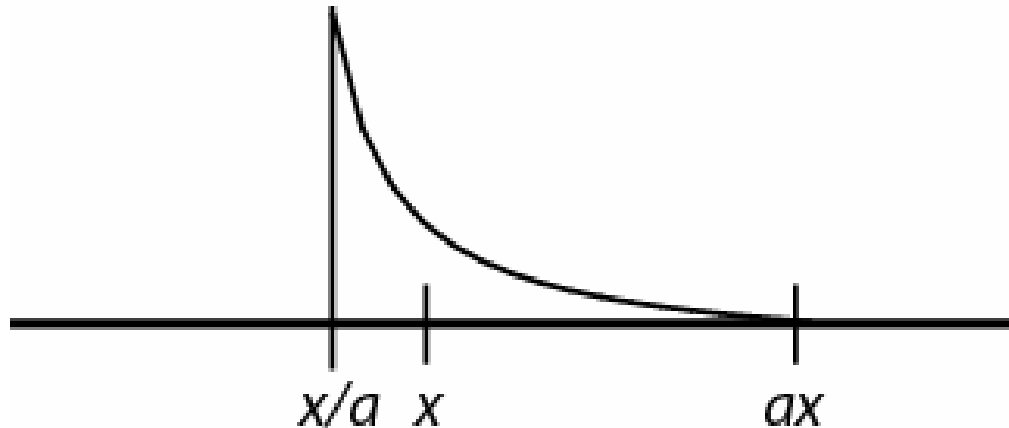   centered on x.

Tuning parameter: α

Bolder proposals: decrease α

More modest proposals: increase α

Works well for proportions, such as revmat and statefreqs.

Multiplier Proposal

New values are picked from the equivalent of a
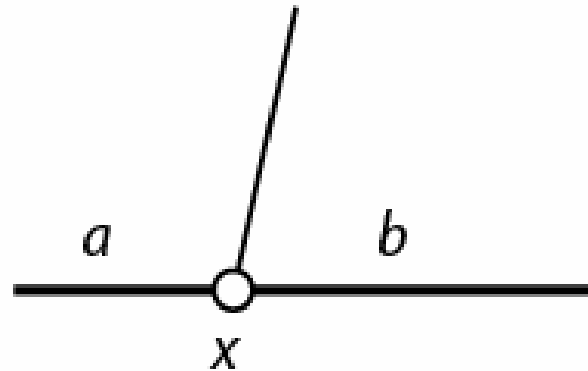 sliding window on the log-transformed x axis.
Tuning parameter: $\lambda = 2 \ln a$
Bolder proposals: increase λ
More modest proposals: decrease λ

*Works well when changes in small values of x have
a larger effect on the probability of data than
changes in large values of x. Example: branch lengths.*

Node Slider

Two adjacent branches a and b are chosen at random
The length of $a + b$ is changed using a multiplier with tuning paremeter λ
The node x is randomly inserted on $a + b$ according to a uniform distribution

Bolder proposals: increase λ
More modest proposals: decrease λ

The boldness of the proposal depends heavily on the uniform reinsertion of x, so changing λ may have limited effect

LOCAL
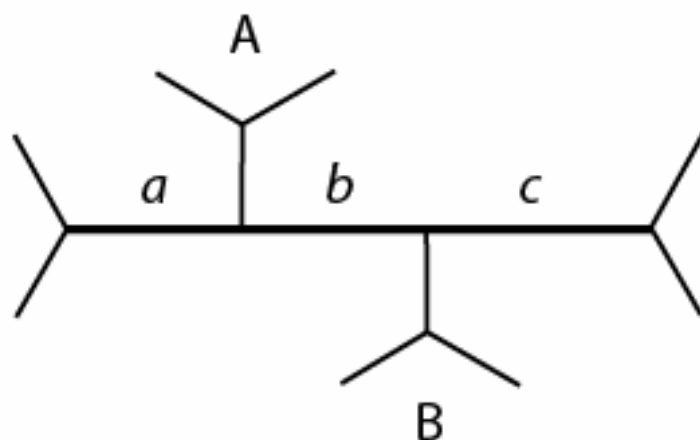
Three internal branches - a, b, and c - are chosen at random.
Their total length is changed using a multiplier with tuning
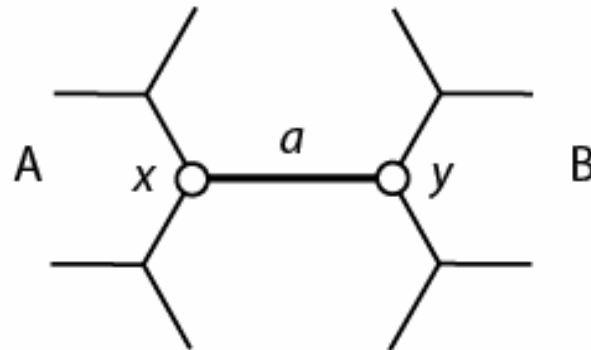   paremeter λ.
One of the subtrees A or B is picked at random.
It is randomly reinserted on $a + b + c$ according to a uni-
   form distribution

Bolder proposals: increase λ
More modest proposals: decrease λ

Extending TBR

An internal branch *a* is chosen at random

The length of *a* is changed using a multiplier with tuning paremeter λ

The node *x* is moved, with one of the adjacent branches, in subtree A, one node at a time, each time the probability of moving one more branch is *p* (the extension probability).

The node *y* is moved similarly in subtree B.

Bolder proposals: increase *p*

More modest proposals: decrease *p*

Changing λ has little effect on the boldness of the proposal.

Why Bayesian?

- Priors: accumulate scientific knowledge
- Easy to deal with complex models, and current models need to be improved
- Computational efficiency: today hundreds of taxa, can probably be extended considerably
- Convergence diagnostics to detect problems with convergence
- Model testing with Bayes factors
- Model averaging using reversible-jump MCMC sampling