

# Workshop on Advanced Bayesian Phylogenetics

## Tuesday 18 November 2014

# About Trees

**Simon Tierney**

Australian Centre for Evolutionary Biology & Biodiversity

School of Earth & Environmental Sciences

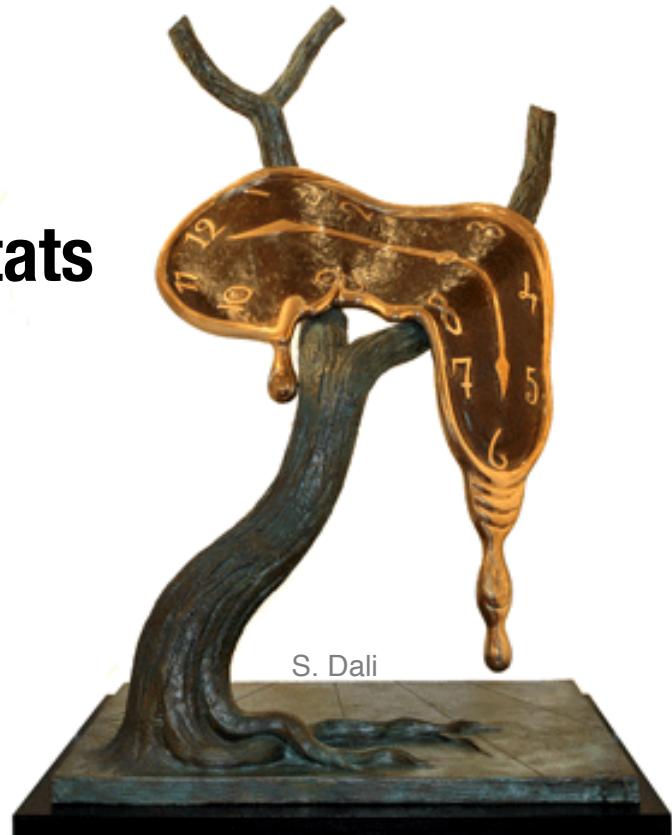
[simon.tierney@adelaide.edu.au](mailto:simon.tierney@adelaide.edu.au)



# Overview

## About Phylogenetic trees:

- The tree space problem
- Tree proposals (moves)
- Summarising trees
- Tree-specific convergence stats

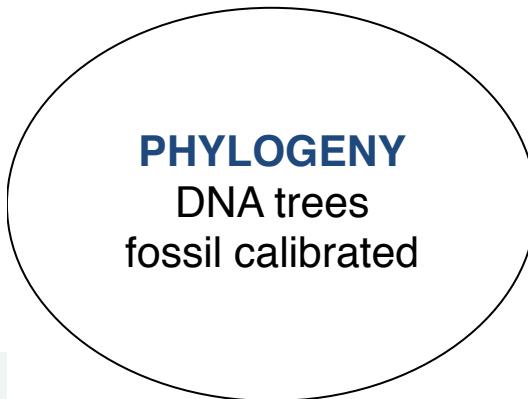
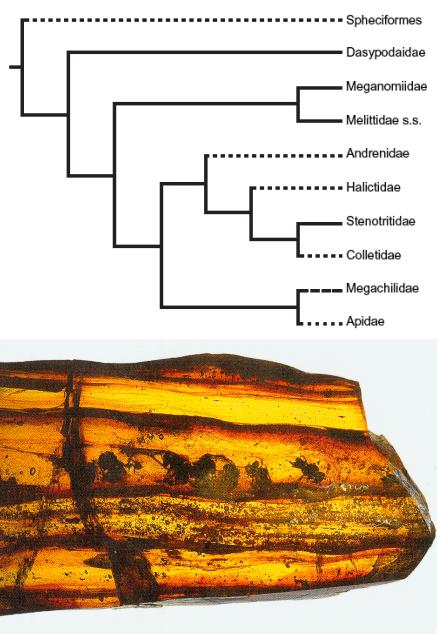


# Research Interests

- 1. Evolution of Social Behaviour**
- 2. Evolution of Nocturnal Foraging**
- 3. Regressive Evolution of Eye Genes**

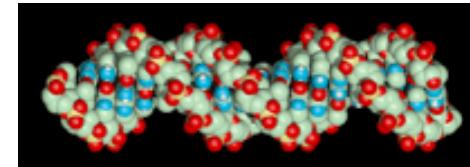
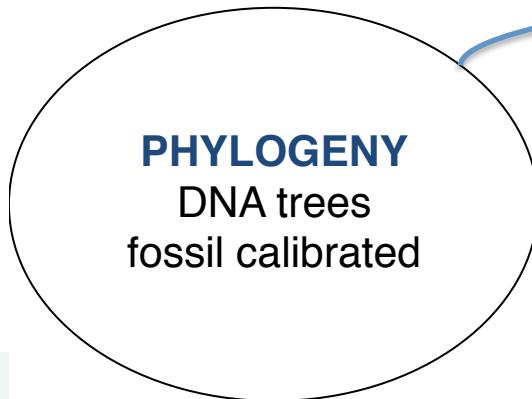
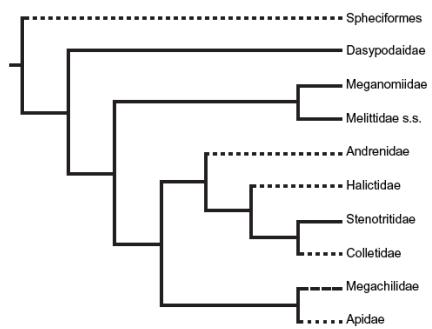
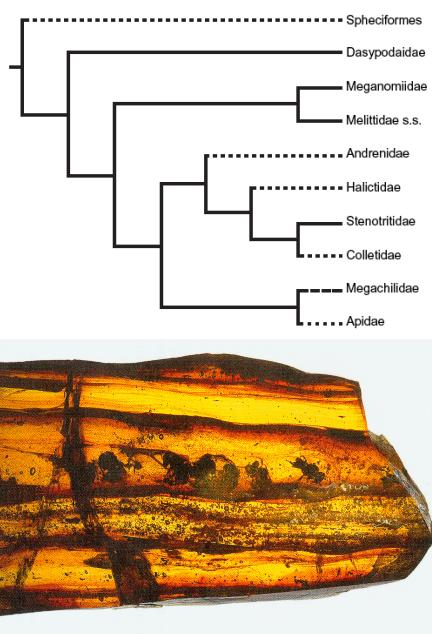
# Research Interests

1. Evolution of Social Behaviour
2. Evolution of Nocturnal Foraging
3. Regressive Evolution of Eye Genes



# Research Interests

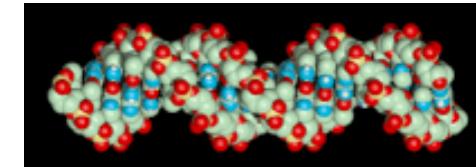
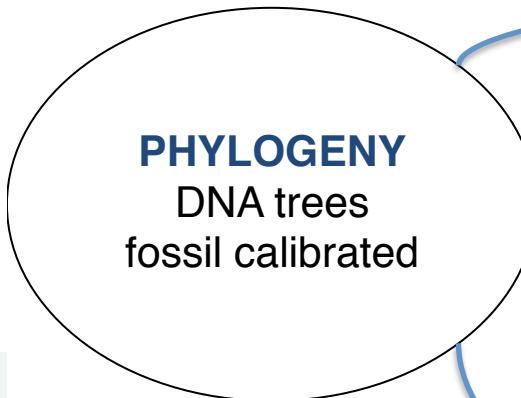
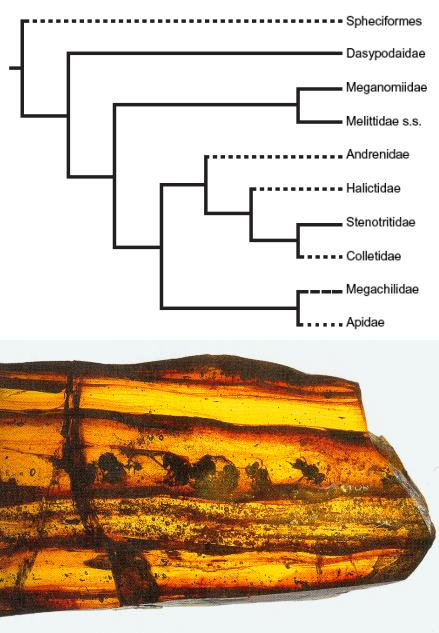
1. Evolution of Social Behaviour
2. Evolution of Nocturnal Foraging
3. Regressive Evolution of Eye Genes



**MOLECULAR EVOLUTION**  
Functional proteins  
Gene evolution

# Research Interests

1. Evolution of Social Behaviour
2. Evolution of Nocturnal Foraging
3. Regressive Evolution of Eye Genes



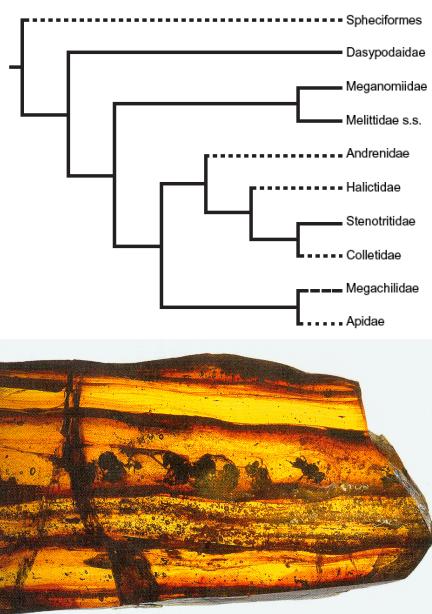
**MOLECULAR EVOLUTION**  
Functional proteins  
Gene evolution



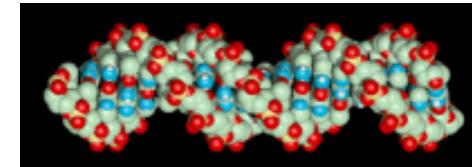
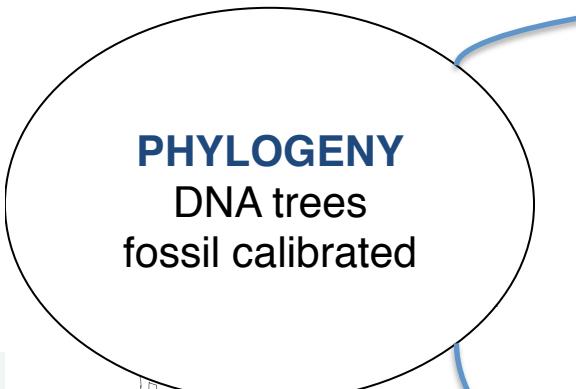
**ETHOLOGY**  
Evolution of worker castes  
Origins of social parasitism  
Photic niche shifts

# Research Interests

1. Evolution of Social Behaviour
2. Evolution of Nocturnal Foraging
3. Regressive Evolution of Eye Genes



**MACROEVOLUTION**  
Biogeography  
Speciation



**MOLECULAR EVOLUTION**  
Functional proteins  
Gene evolution



**ETHOLOGY**  
Evolution of worker castes  
Origins of social parasitism  
Photic niche shifts

# Evolutionary Ecology

**A brief history in science...**

Old World  
 **facultatively social bees**



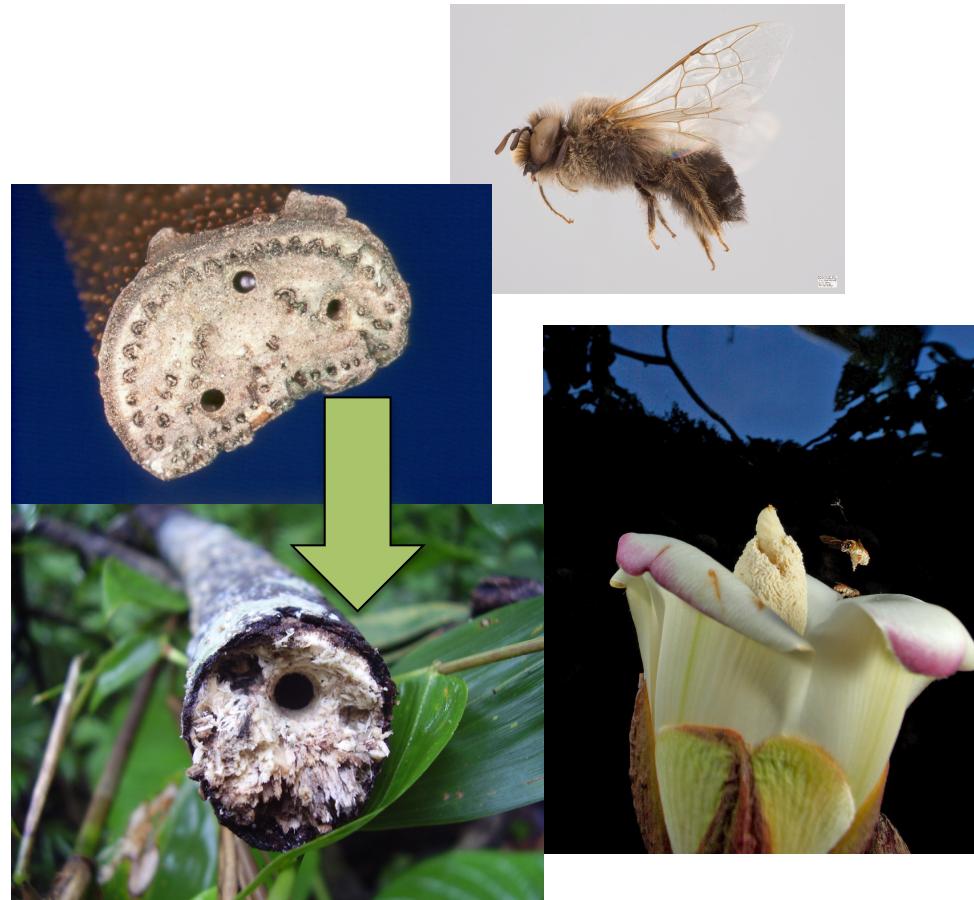
# Evolutionary Ecology

## A brief history in science...

Old World  
 **facultatively social bees**



Neotropical  
 **facultatively social bees**  
dim-light foraging



# Evolutionary Ecology

## A brief history in science...

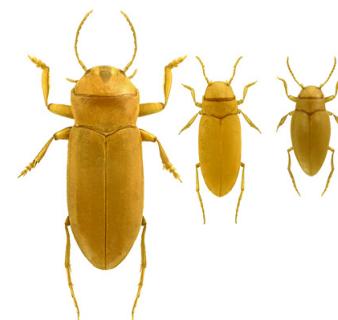
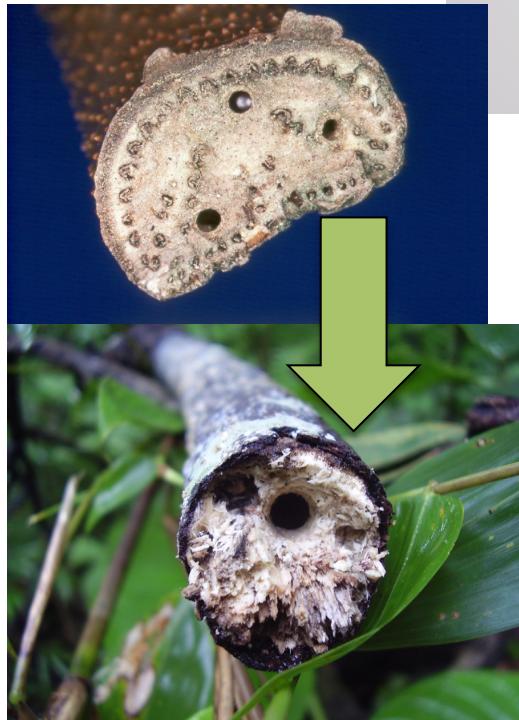
Old World  
 **facultatively social bees**



Neotropical  
 **facultatively social bees**  
dim-light foraging



Australian  
 **groundwater-dwelling**  
blind predatory water beetles



*C. Zeigler*



C. Zeigler





# Tree Space Problem

## Problem:

Sometimes the informative posterior probability is located in a very small area of parameter space



# Tree Space Problem

## Problem:

Sometimes the informative posterior probability is located in a very small area of parameter space

**Exemplified by Phylogenetics**  
especially in relation to tree topology



# Tree Space Problem

## Problem:

Sometimes the informative posterior probability is located in a very small area of parameter space

**Exemplified by Phylogenetics**  
especially in relation to tree topology

**9 taxa > 135k trees:**

odds of being hit by lightning	(3:100,000)	0.00003
randomly choosing best tree	(1:135,135)	0.000007



# Tree Space Problem

## Problem:

Sometimes the informative posterior probability is located in a very small area of parameter space

### Exemplified by Phylogenetics

especially in relation to tree topology

#### 9 taxa > 135k trees:

odds of being hit by lightning	(3:100,000)	0.00003
randomly choosing best tree	(1:135,135)	0.000007



#### 50 taxa = 2.83806E+74 trees:

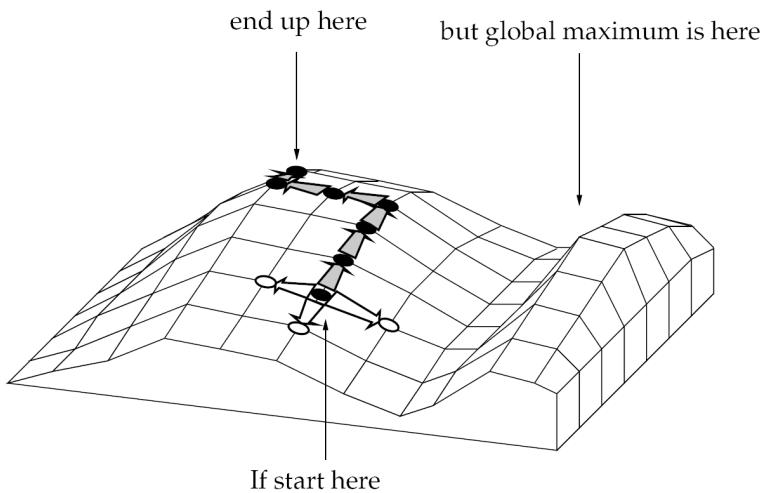
More possible tree topologies than atoms in the known universe  $\sim 10^{80}$

# Tree Space Problem

## Problem:

Sometimes the informative posterior probability is located in a very small area of parameter space

## Space above a 2D plane

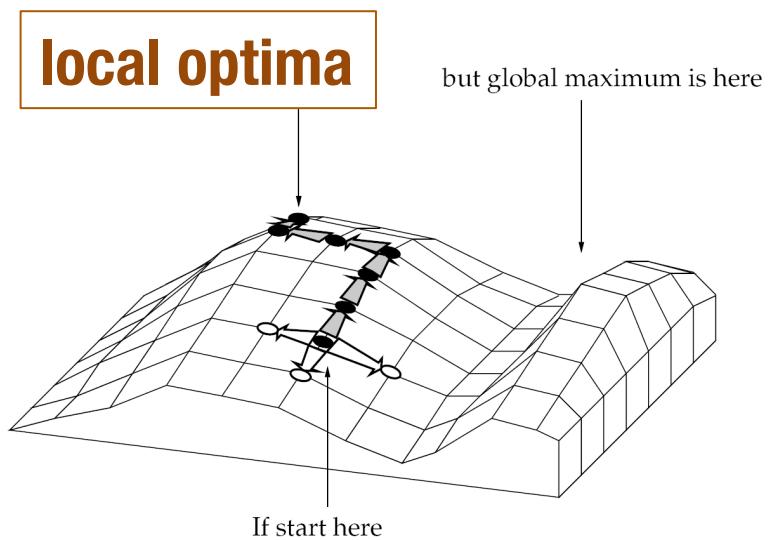


# Tree Space Problem

## Problem:

Sometimes the informative posterior probability is located in a very small area of parameter space

## Space above a 2D plane

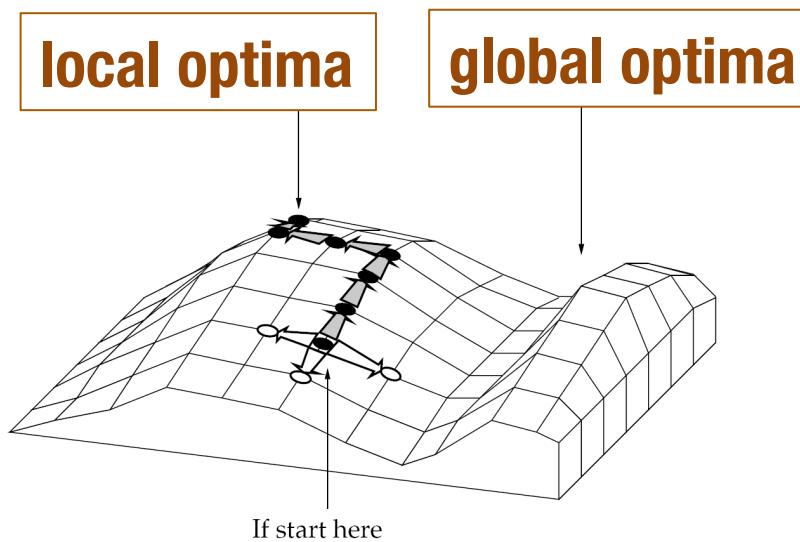


# Tree Space Problem

## Problem:

Sometimes the informative posterior probability is located in a very small area of parameter space

## Space above a 2D plane



# Tree Space Problem

## **Problem:**

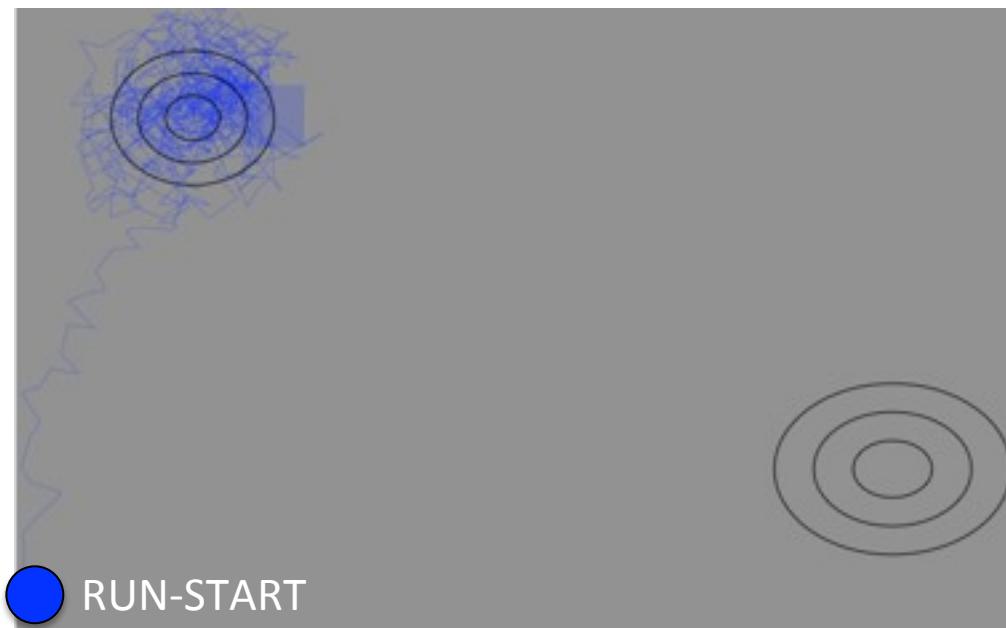
Susceptible to becoming trapped

# Tree Space Problem

## Problem:

Susceptible to becoming trapped

Plan view



# Tree Space Problem

## Problem:

Susceptible to becoming trapped

Plan view

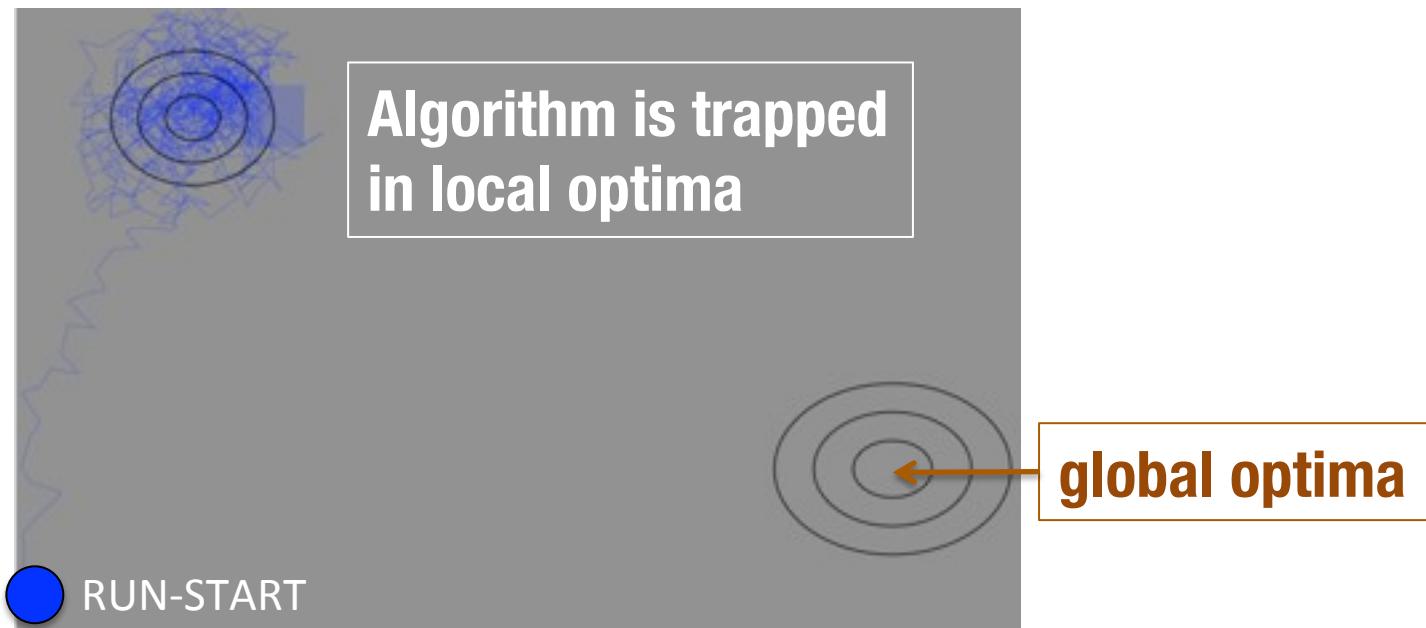


# Tree Space Problem

## Problem:

Susceptible to becoming trapped

Plan view



# Obvious Solution

1) Use a big needle  
and a small haystack



# Obvious Solution

1) Use a big needle  
and a small haystack



2) Look at the haystack

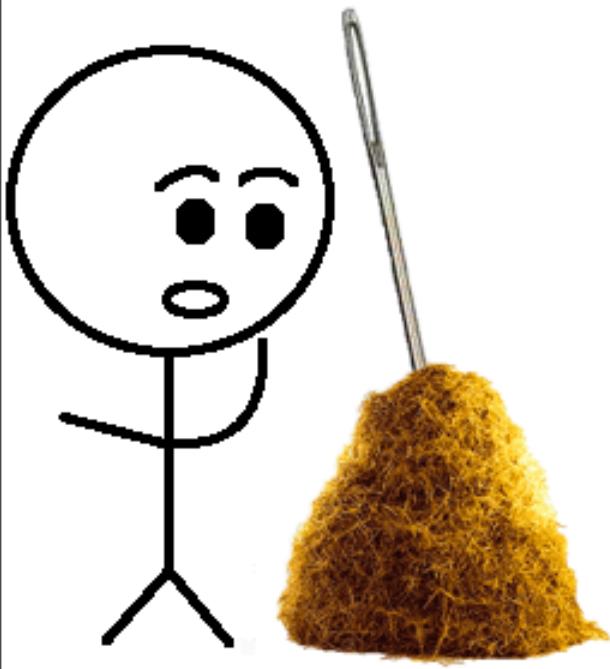


# Obvious Solution

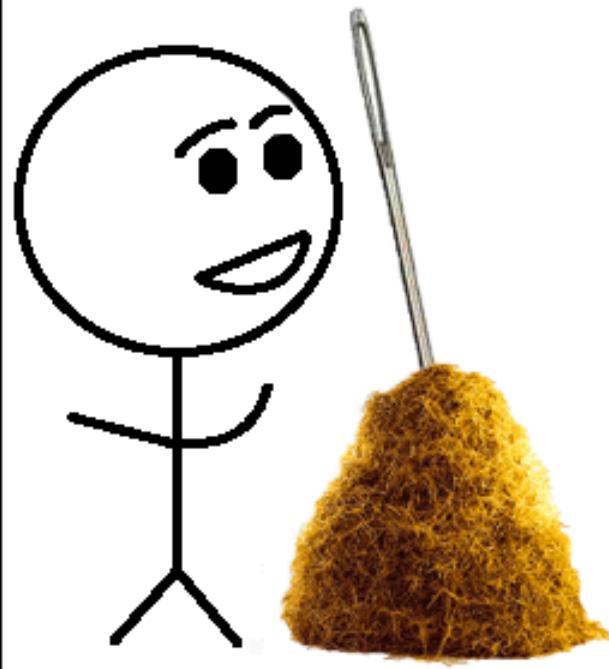
1) Use a big needle  
and a small haystack



2) Look at the haystack



3) Look at the needle



# MCMC redux

## Markov Chain

Memory-less random process

Succeeding state is dependent only on the current state

**no historical context**



# MCMC redux

## Markov Chain

Memory-less random process

Succeeding state is dependent only on the current state

**no historical context**

## Monte Carlo Simulation

Repeated random sampling



# MCMC redux

## Markov Chain

Memory-less random process

Succeeding state is dependent only on the current state

**no historical context**

## Monte Carlo Simulation

Repeated random sampling

## Metropolis-Hastings Sampler

Undertake small random changes to the parameter of interest and then accept/reject the change based on probability ratio:  $r$



# MCMC redux

## **MCMC Steps:**

*Simplified version*

1. Start at a random point in the parameter space

# MCMC redux

## **MCMC Steps:**

*Simplified version*

1. Start at a random point in the parameter space
2. Make a small random move to a new state\* in that landscape

# MCMC redux

## **MCMC Steps:**

*Simplified version*

1. Start at a random point in the parameter space
2. Make a small random move to a new state\* in that landscape
3. Calculate the **ratio of posterior probability**  
between the existing state and the new state\*

# MCMC redux

## **MCMC Steps:**

*Simplified version*

1. Start at a random point in the parameter space
2. Make a small random move to a new state\* in that landscape
3. Calculate the **ratio of posterior probability**  
between the existing state and the new state\*
  - (a)  $r > 1$       new state\* is always accepted

# MCMC redux

## MCMC Steps:

*Simplified version*

1. Start at a random point in the parameter space
2. Make a small random move to a new state\* in that landscape
3. Calculate the **ratio of posterior probability**  
between the existing state and the new state\*
  - (a)  $r > 1$       new state\* is always accepted
  - (b)  $r < 1$       new state\* sometimes accepted  
or reject and maintain existing state

# MCMC redux

## MCMC Steps:

*Simplified version*

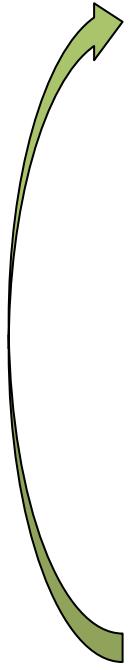
1. Start at a random point in the parameter space
2. Make a small random move to a new state\* in that landscape
3. Calculate the **ratio of posterior probability**  
between the existing state and the new state\*
  - (a)  $r > 1$       new state\* is always accepted
  - (b)  $r < 1$       new state\* sometimes accepted  
or reject and maintain existing state

**this constitutes one iteration of the process**

# MCMC redux

## MCMC Steps:

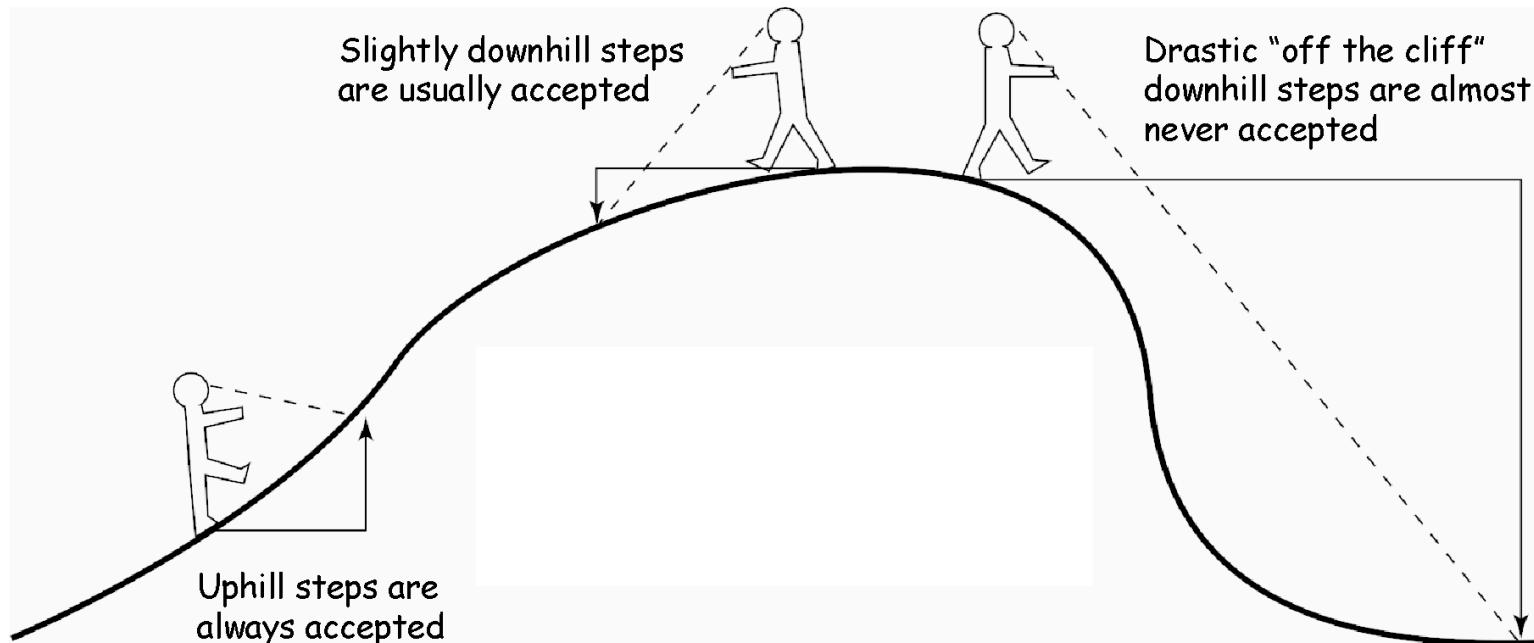
*Simplified version*

- 
1. Start at a random point in the parameter space
  2. Make a small random move to a new state\* in that landscape
  3. Calculate the **ratio of posterior probability**  
between the existing state and the new state\*
    - (a)  $r > 1$       new state\* is always accepted
    - (b)  $r < 1$       new state\* sometimes accepted  
or reject and maintain existing state
  4. Go to step 2 and repeat

**this constitutes one iteration of the process**

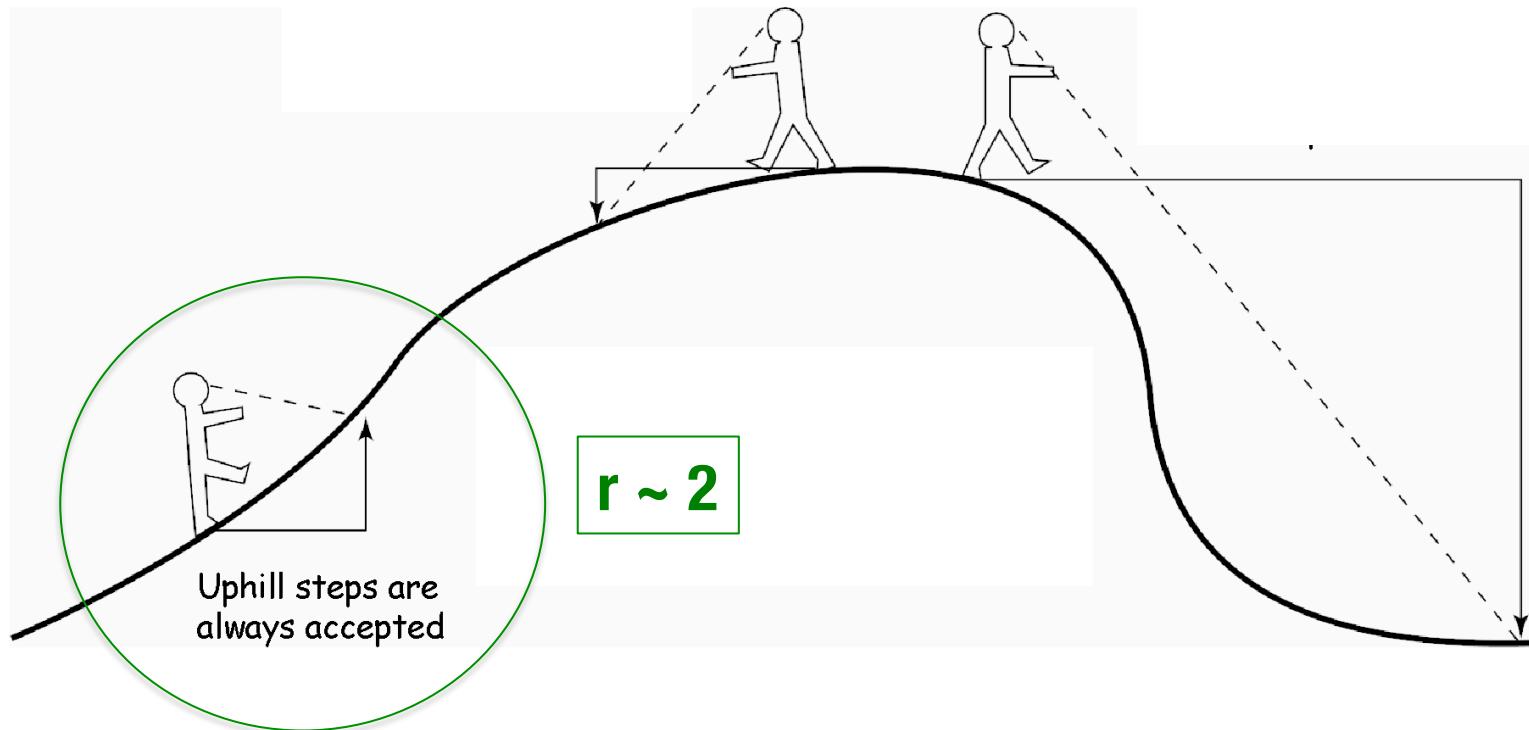
# Maximum Likelihood Zombie

Zombie mindlessly aims for the top of the hill



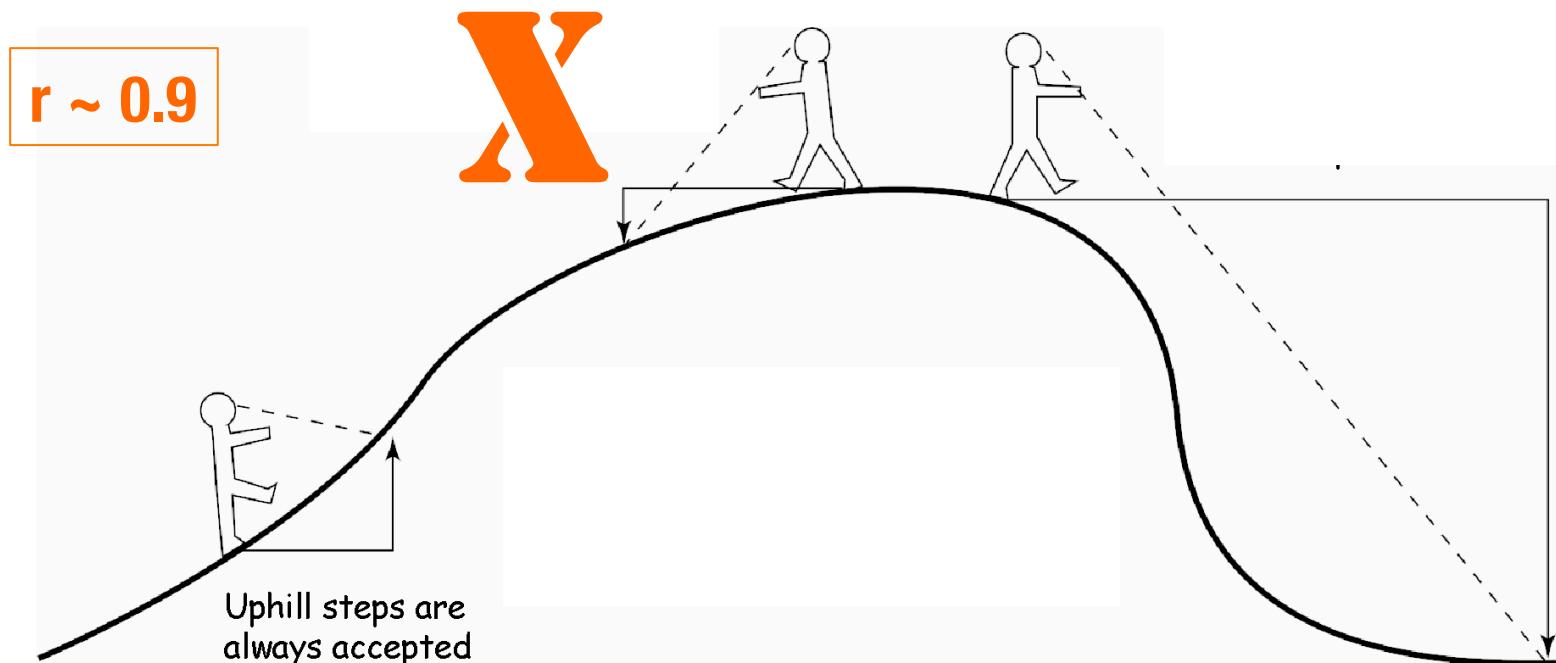
# Maximum Likelihood Zombie

Zombie mindlessly aims for the top of the hill



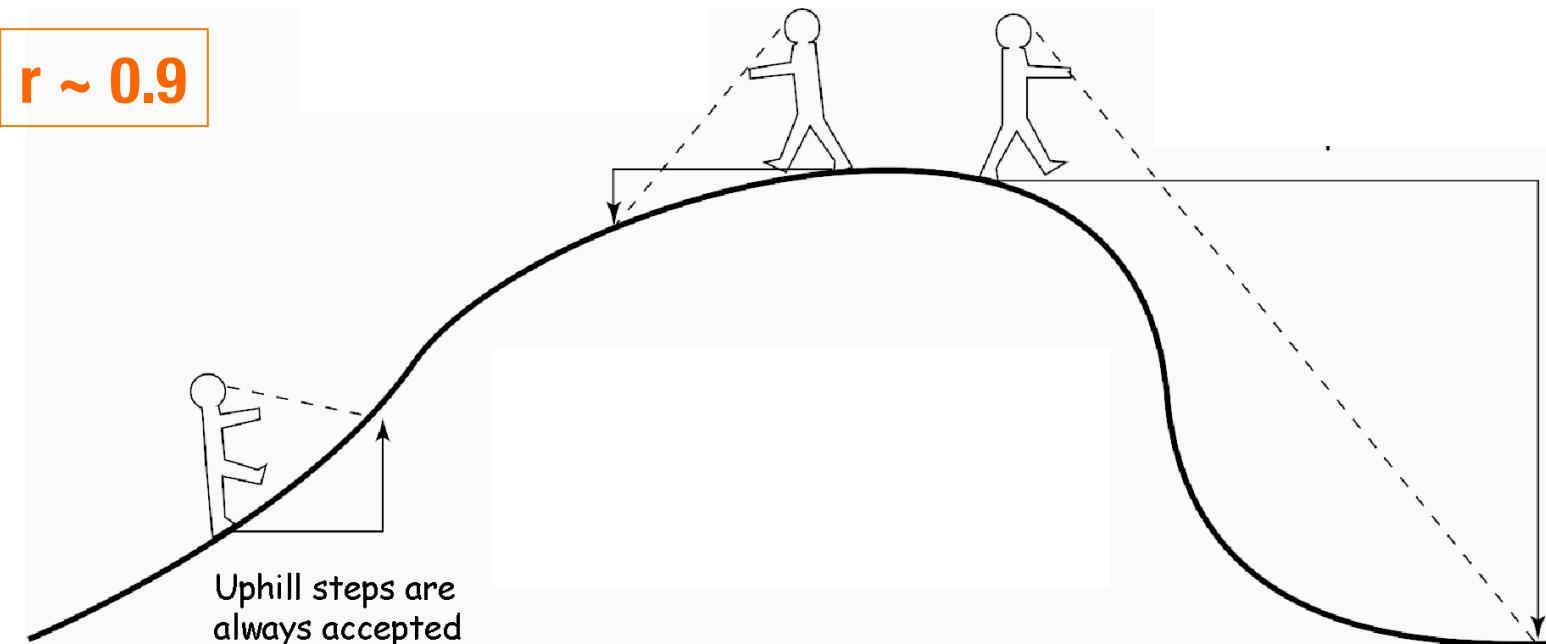
# Maximum Likelihood Zombie

Zombie would not accept a move of lower probability



# MCMC Diligent Robot

Will consider joint posterior probability (of likelihood & prior)

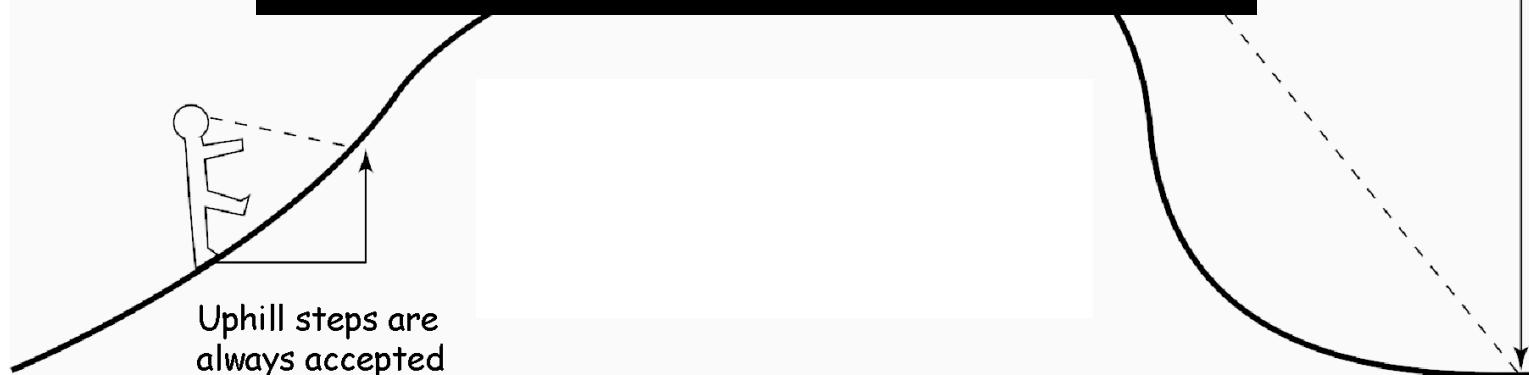


# MCMC Diligent Robot

**Will consider joint posterior probability (of likelihood & prior)**

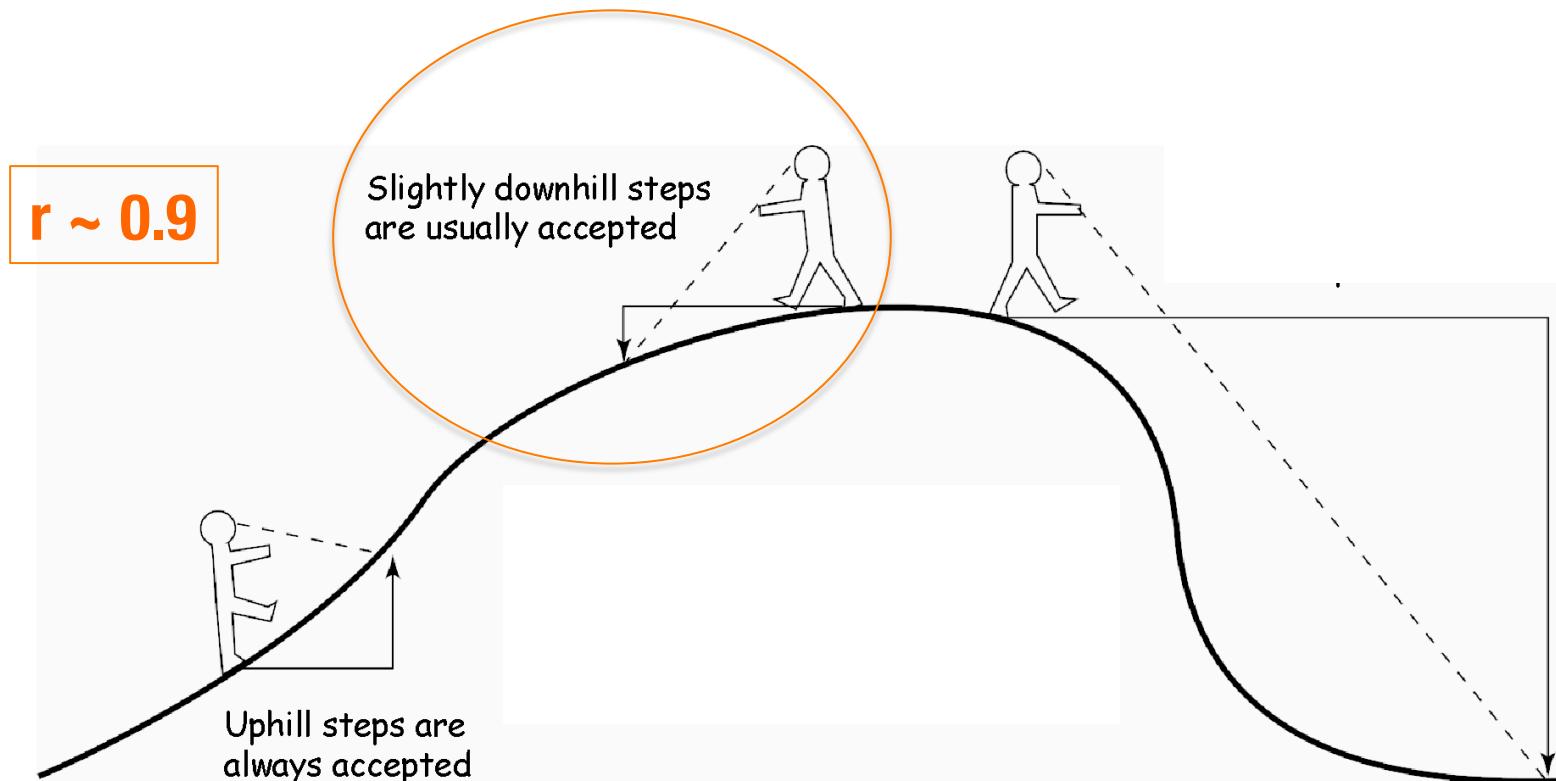
**r ~ 0.9**

**If the Hastings Proposal prior = 1  
Decision to move is totally dependent  
on the joint probabilities of  
the likelihood & the prior**



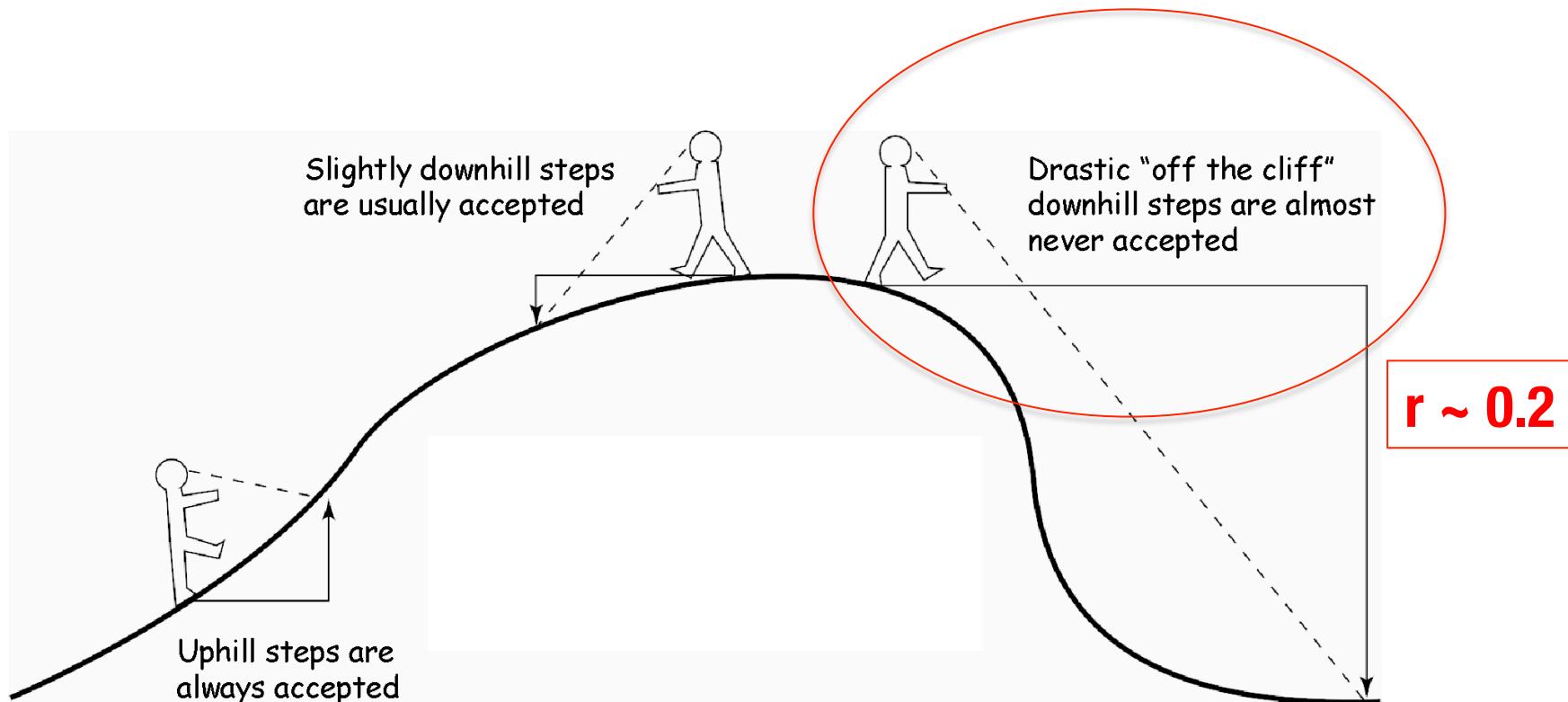
# MCMC Diligent Robot

Will consider joint posterior probability (of likelihood & prior)



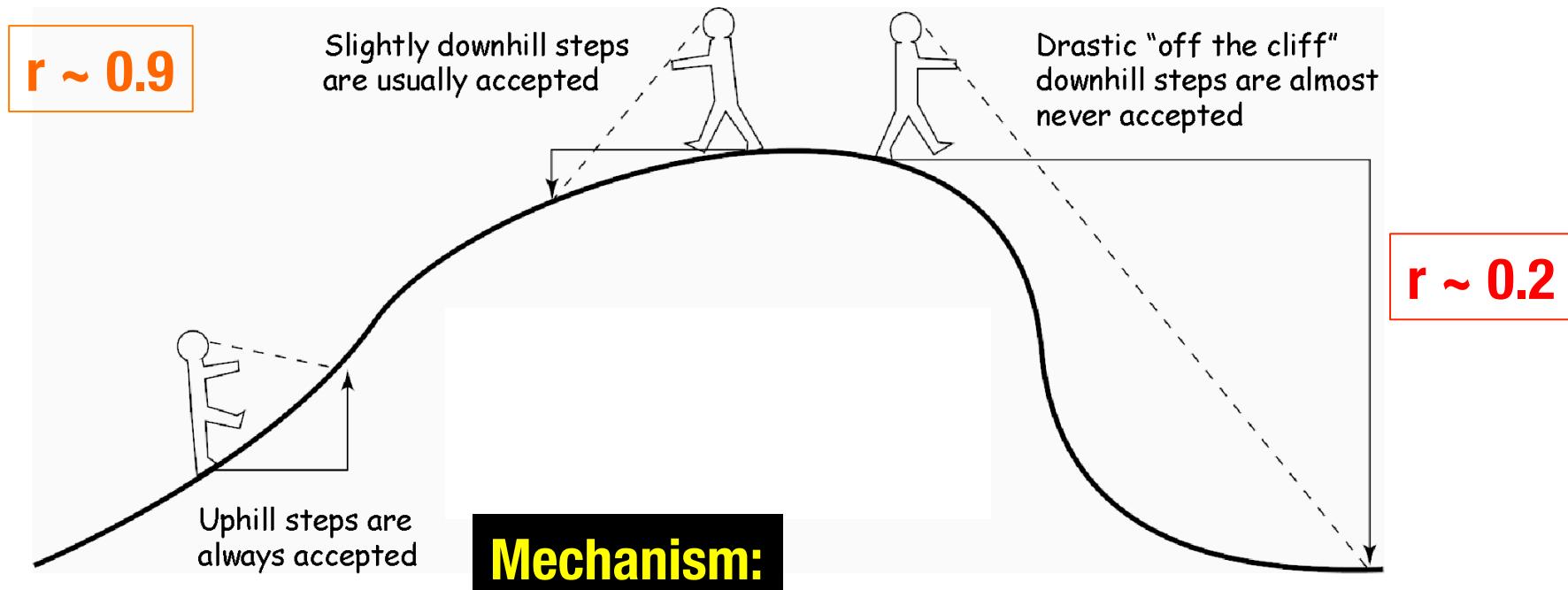
# MCMC Diligent Robot

Will consider joint posterior probability (of likelihood & prior)



# MCMC Diligent Robot

Will consider joint posterior probability (of likelihood & prior)



If  $r$  is  $< 1$

Decision to accept/reject move is dependent upon  $U$

$U$  = uniform number randomly drawn (0, 1)

If  $r > U$  move is accepted

# Tree Space Problem

## Solution 1:

MCMC permits a much more thorough exploration of parameter space.

MCMC inference is averaged out over the joint posterior probability of all parameters (including trees & branch lengths).

Therefore much more robust than ML inference.



# Probability ratio

## **Calculation of r:**

*Simplified version*

# Probability ratio

## **Calculation of r:**

*Simplified version*

X data

# Probability ratio

## **Calculation of r:**

*Simplified version*

X data

Θ existing parameter state

# Probability ratio

## **Calculation of r:**

*Simplified version*

$X$  data

$\Theta$  existing parameter state

$\Theta^*$  new parameter state

# Probability ratio

## Calculation of r:

*Simplified version*

**X** data

**Θ** existing parameter state

**Θ \*** new parameter state

**P** probability

# Probability ratio

## Calculation of r:

*Simplified version*

**X** data

**Θ** existing parameter state

**Θ \*** new parameter state

**P** probability

**r** ratio of existing and new\* parameter state probabilities

# Probability ratio

## Calculation of r:

*Simplified version*

X data

$\Theta$  existing parameter state

$\Theta^*$  new parameter state

P probability

r ratio of existing and new\* parameter state probabilities

prior ratio

$$r = \frac{P(\Theta^*)}{P(\Theta)}$$

# Probability ratio

## Calculation of r:

*Simplified version*

X data

$\Theta$  existing parameter state

$\Theta^*$  new parameter state

P probability

r ratio of existing and new\* parameter state probabilities

**prior ratio**

**likelihood ratio**

$$r = \frac{P(\Theta^*)}{P(\Theta)} \times \frac{P(X | \Theta^*)}{P(X | \Theta)}$$

# Probability ratio

## Calculation of r:

*Simplified version*

X data

$\Theta$  existing parameter state

$\Theta^*$  new parameter state

P probability

r ratio of existing and new\* parameter state probabilities

**prior ratio**

**likelihood ratio**

$$r = \frac{P(\Theta^*)}{P(\Theta)} \times \frac{P(X | \Theta^*)}{P(X | \Theta)}$$

**joint posterior probability**

# Probability ratio

## Calculation of $r$ :

*Simplified version*

$X$  data

$\Theta$  existing parameter state

$\Theta^*$  new parameter state

$P$  probability

$r$  ratio of existing and new\* parameter state probabilities

Hastings

prior ratio

likelihood ratio

proposal ratio

$$r = \frac{P(\Theta^*)}{P(\Theta)} \times \frac{P(X | \Theta^*)}{P(X | \Theta)} \times \frac{P(\Theta | \Theta^*)}{P(\Theta^* | \Theta)}$$

joint posterior probability

# Tree Topology Parameter

**For phylogenetics**

Our data is a character state matrix

# Tree Topology Parameter

## For phylogenetics

Our data is a character state matrix  
(morphological or gene sequence alignment)



# Tree Topology Parameter

## For phylogenetics

Our data is a character state matrix  
(morphological or gene sequence alignment)

T	A	C
T	A	C

Our parameter is the tree topology



# Tree Topology Parameter

## For phylogenetics

Our data is a character state matrix  
(morphological or gene sequence alignment)

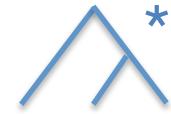
T	A	C
T	A	C

Our parameter is the tree topology



## Metropolis-Hastings sampler

Proposes a new topology



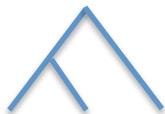
# Tree Topology Parameter

## For phylogenetics

Our data is a character state matrix  
(morphological or gene sequence alignment)

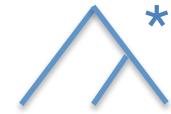
T	A	C
T	A	C

Our parameter is the tree topology



## Metropolis-Hastings sampler

Proposes a new topology



$$\frac{P(\text{ } \wedge \text{ } | \text{ } \wedge^*)}{P(\wedge^* | \wedge)}$$

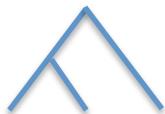
# Tree Topology Parameter

## For phylogenetics

Our data is a character state matrix  
(morphological or gene sequence alignment)

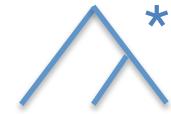
T	A	C
T	A	C

Our parameter is the tree topology



## Metropolis-Hastings sampler

Proposes a new topology



$$\frac{P(\text{[TAC TAC]} | \text{ } \wedge^*)}{P(\text{[TAC TAC]} | \text{ } \wedge)} \times \frac{P(\text{ } \wedge | \text{ } \wedge^*)}{P(\text{ } \wedge^* | \text{ } \wedge)}$$

# Tree Topology Parameter

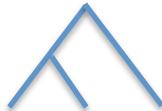
## For phylogenetics

Our data is a character state matrix  
(morphological or gene sequence alignment)

T	A	C
T	A	C

Our parameter is the tree topology

## Metropolis-Hastings sampler



Proposes a new topology



$$r = \frac{P(\text{ } \nearrow^*)}{P(\text{ } \nearrow)} \times \frac{P(\text{ } \boxed{\text{TAC}} \text{ } | \text{ } \nearrow^*)}{P(\text{ } \boxed{\text{TAC}} \text{ } | \text{ } \nearrow)} \times \frac{P(\text{ } \nearrow \text{ } | \text{ } \nearrow^*)}{P(\text{ } \nearrow^* \text{ } | \text{ } \nearrow)}$$

# Metropolis sampler



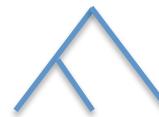
**Each sampler targets one (or a related set) parameter of the model**

# Metropolis sampler



**Each sampler targets one (or a related set) parameter of the model**

Ratio of the rates at which the Markov chain switches between



&



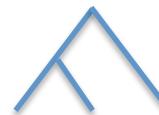
is proportional to the posterior probability **at stationarity**

# Metropolis sampler



**Each sampler targets one (or a related set) parameter of the model**

Ratio of the rates at which the Markov chain switches between



&



is proportional to the posterior probability **at stationarity**

*Simplistically*

Chain spends 77% of the time at parameter state  $\Theta^*$

# Metropolis sampler



**Each sampler targets one (or a related set) parameter of the model**

Ratio of the rates at which the Markov chain switches between



is proportional to the posterior probability **at stationarity**

*Simplistically*

Chain spends 77% of the time at parameter state  $\Theta^*$

Thus:

Node support for topology is 0.77 posterior probability

# Sampler types

**A large parameter-rich model requires multiple different samplers**

In order to move to a new parameter state  $\Theta^*$

**Four basic modes to propose a new parameter state:**

1. Sliding Window proposal
2. Normal proposal
3. Multiplier proposal
4. Beta & Dirichlet proposal



Canvassed by Mike Lee

**But these are not appropriate for moving to a new topology**

This is the subject of the next section

# Tree Space Problem

**MCMC is a more robust search engine**  
but still susceptible to getting trapped in local optima

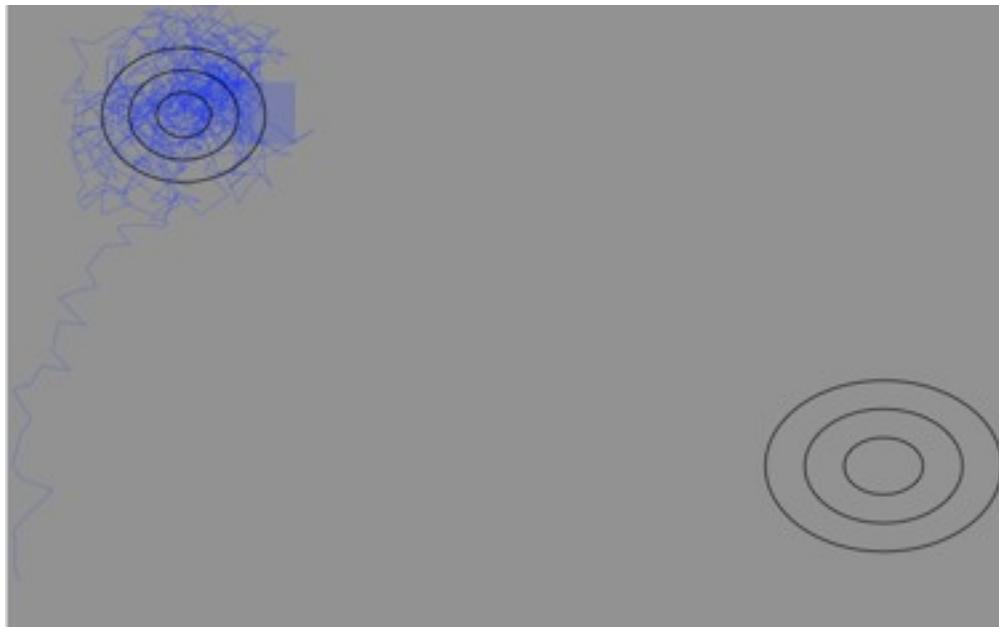


# Tree Space Problem

## Solution 2:

Undertake parallel runs that start at independent points in parameter space

Plan view

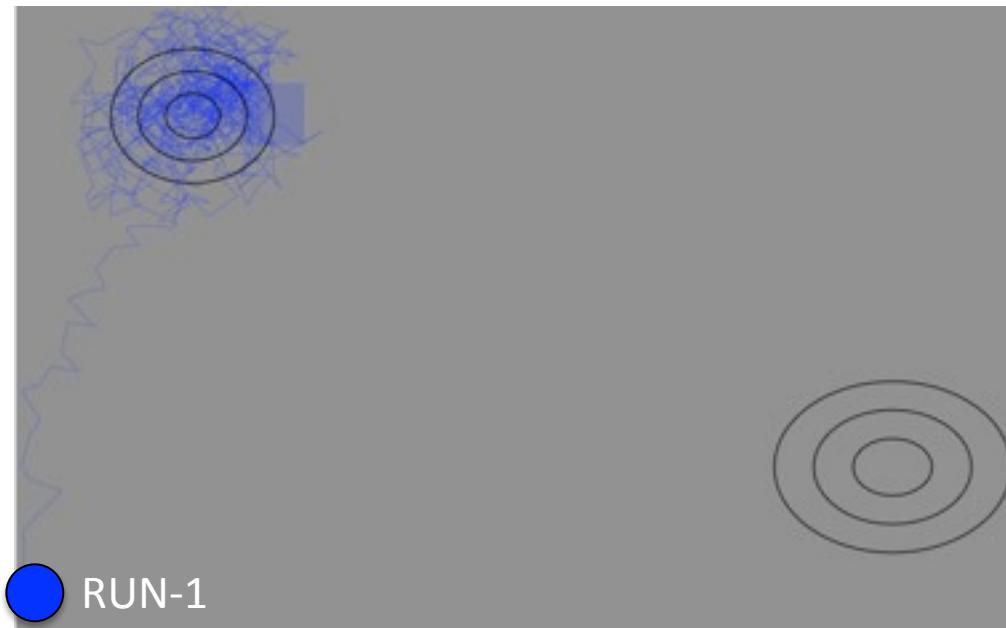


# Tree Space Problem

## Solution 2:

Undertake parallel runs that start at independent points in parameter space

Plan view

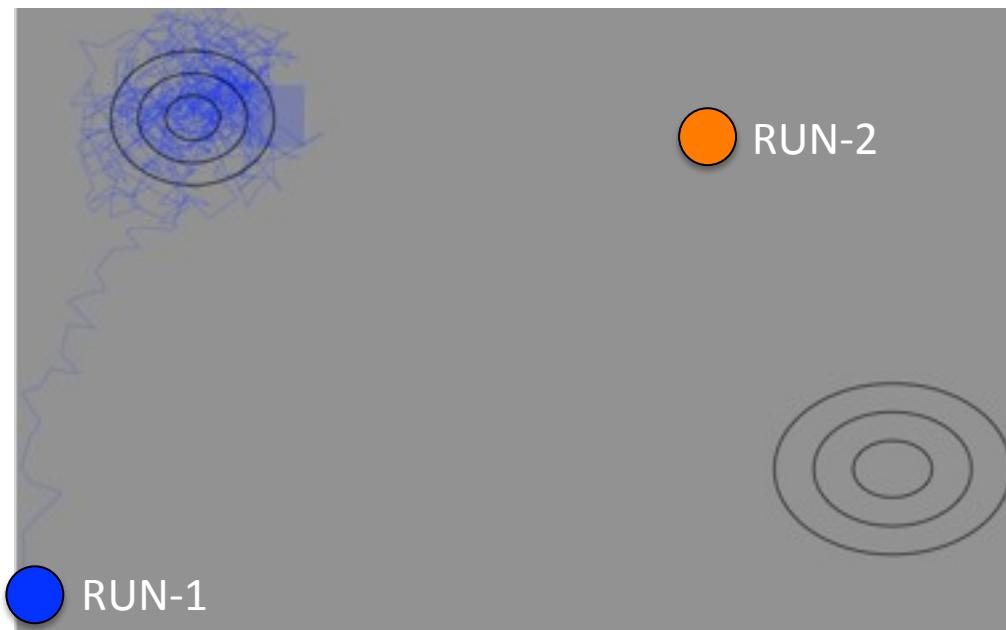


# Tree Space Problem

## Solution 2:

Undertake parallel runs that start at independent points in parameter space

Plan view

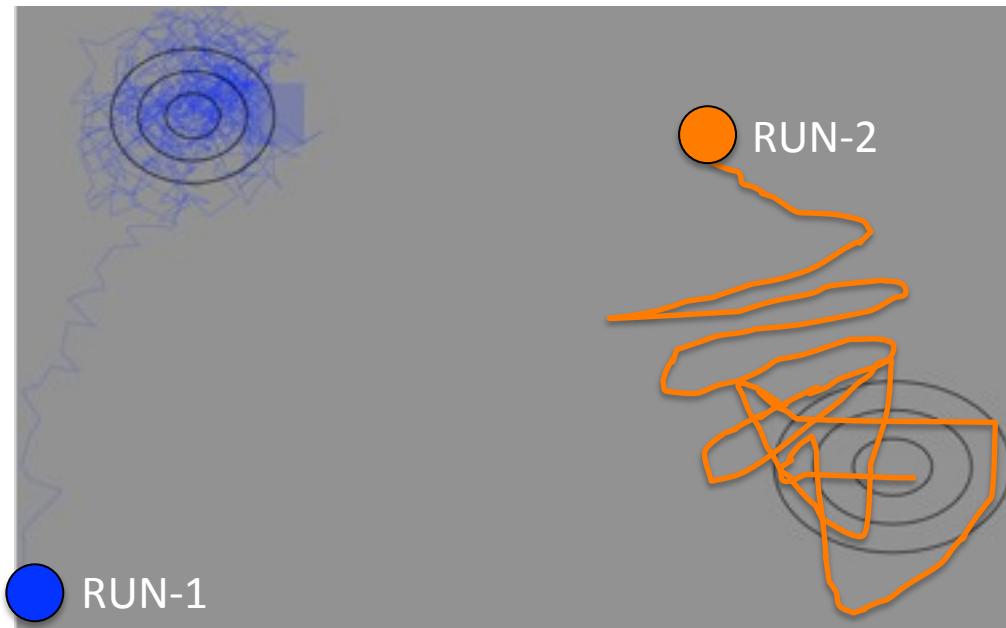


# Tree Space Problem

## Solution 2:

Undertake parallel runs that start at independent points in parameter space

Plan view



Adapted from Paul Lewis

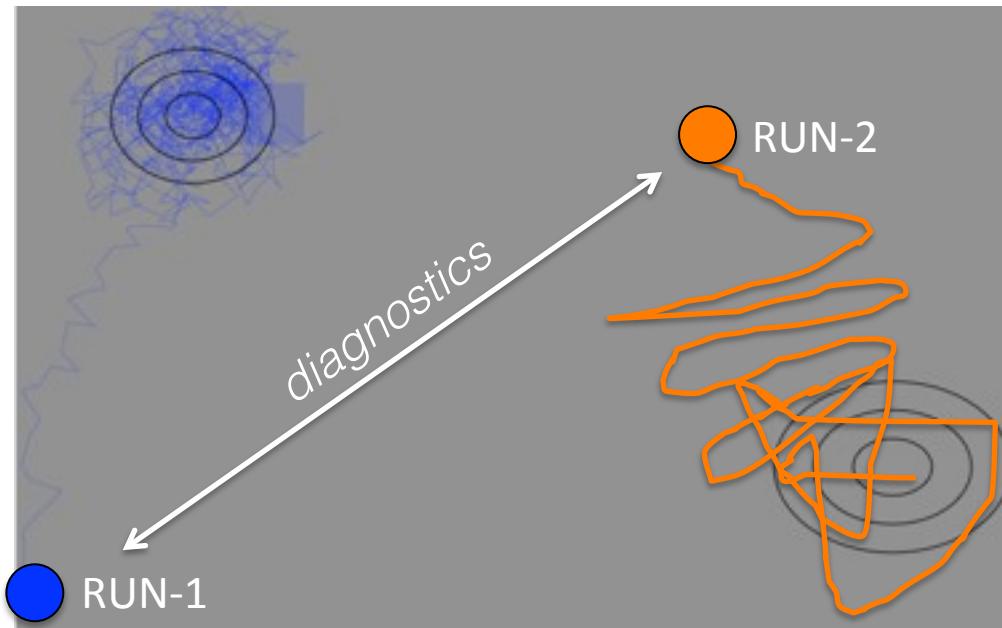
# Tree Space Problem

## Solution 2:

Undertake parallel runs that start at independent points in parameter space

Examine the convergence between RUN-1 & RUN-2

Plan view

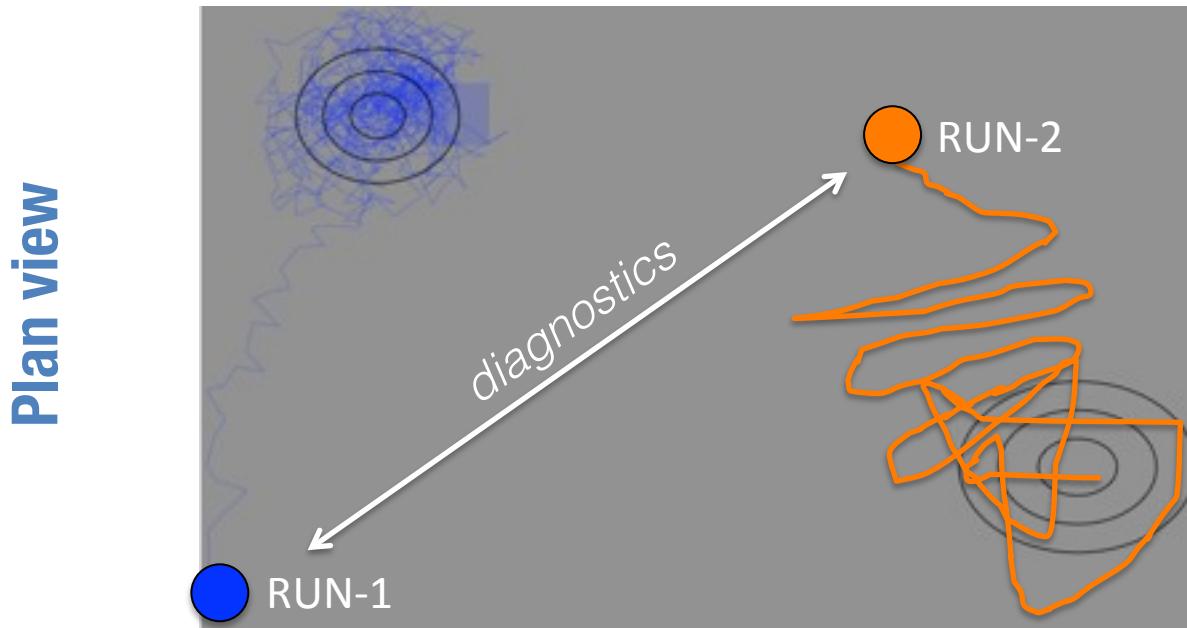


# Tree Space Problem

## Solution 2:

Undertake parallel runs that start at independent points in parameter space

Examine the convergence between RUN-1 & RUN-2



Thus avoid becoming trapped in a local optima

Adapted from Paul Lewis

# Sampling space

**Should now understand how MCMC navigates through parameter space in order to arrive at new parameter states**





# Tree Proposals (moves)

## Heuristic Searches

Attempt to find the best trees without looking at all trees.

“Greedy” searches = good solution in a reasonable time.

Indication of which trees are “promising”.

# Tree Proposals (moves)

## Heuristic Searches

Attempt to find the best trees without looking at all trees.

“Greedy” searches = good solution in a reasonable time.

Indication of which trees are “promising”.

**Tree topology** - Tau

**Branch length** - V

# Tree Proposals (moves)

## Heuristic Searches

Attempt to find the best trees without looking at all trees.

“Greedy” searches = good solution in a reasonable time.

Indication of which trees are “promising”.

**Tree topology** - Tau

**Branch length** - V

### Simplistic

Nearest Neighbour Interchanges (NNI)

### Stochastic branch rearrangement

Subtree Pruning & Refgrafting (SPR)

Tree Bissection & Reconnection (TBR)

# Tree Proposals (moves)

## Heuristic Searches

Attempt to find the best trees without looking at all trees.

“Greedy” searches = good solution in a reasonable time.

Indication of which trees are “promising”.

**Tree topology** - Tau

**Branch length** - V

### Simplistic

Nearest Neighbour Interchanges      (NNI)

### Stochastic branch rearrangement

Subtree Pruning & Refactoring      (SPR)

Tree Bissection & Reconnection      (TBR)



Increasing complexity

# Tree Proposals (moves)

## Default proposal probabilities (MrBayes)

MCMC sampler uses the following moves:

With prob.	Chain will use move
1.00 %	Dirichlet (Repmat)
1.00 %	Slider (Repmat)
1.00 %	Dirichlet (Pi)
1.00 %	Slider (Pi)
2.00 %	Multiplier (Alpha)
10.00 %	ExtSPR (Tau,V)
10.00 %	ExtTBR (Tau,V)
10.00 %	ExtTBR (Tau,V)
10.00 %	ParsSPR (Tau,V)
40.00 %	Multiplier (V)
10.00 %	Nodeslider (V)
4.00 %	TLMultiplier (V)

# Tree Proposals (moves)

## Default proposal probabilities (MrBayes)

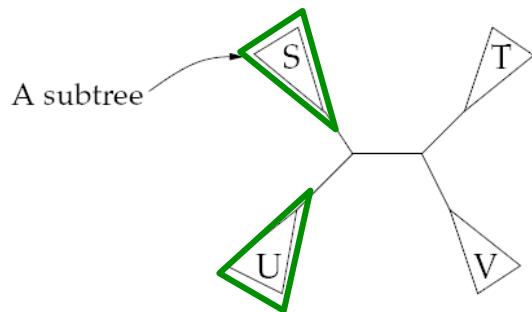
MCMC sampler uses the following moves:

With prob.	Chain will use move
1.00 %	Dirichlet (Revmat)
1.00 %	Slider (Revmat)
1.00 %	Dirichlet (Pi)
1.00 %	Slider (Pi)
2.00 %	Multiplier (Alpha)
10.00 %	ExtSPR (Tau,V)
10.00 %	ExtTBR (Tau,V)
10.00 %	ExtTBR (Tau,V)
10.00 %	ParsSPR (Tau,V)
40.00 %	Multiplier (V)
10.00 %	Nodeslider (V)
4.00 %	TLMultiplier (V)

MCMC will allocate 40%  
of search effort to changing  
tree topology alone



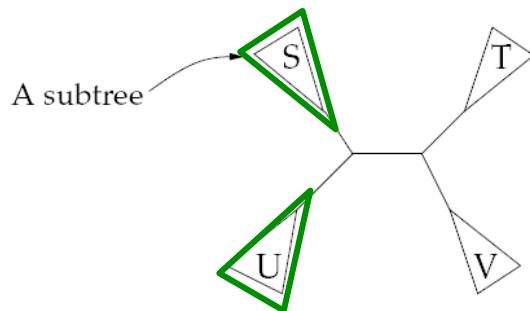
# NN-Nearest Neighbour Interchange



**Simplistic**  
swaps 2-adjacent branches

is rearranged by dissolving the connections to an interior branch

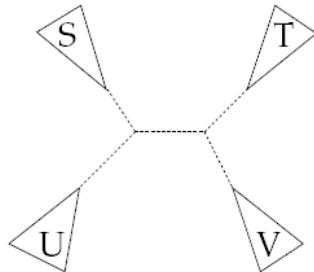
# NN-Nearest Neighbour Interchange



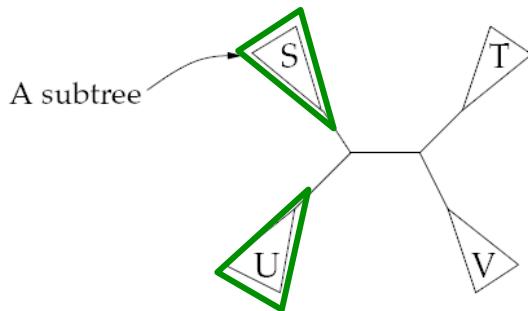
**Simplistic**  
swaps 2-adjacent branches

5-branches dissolved  
to isolate 4-subtrees

is rearranged by dissolving the connections to an interior branch



# NN-Nearest Neighbour Interchange

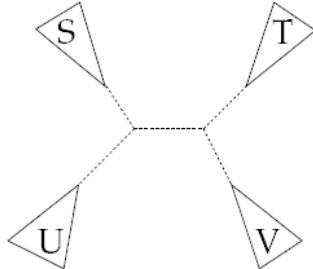


is rearranged by dissolving the connections to an interior branch

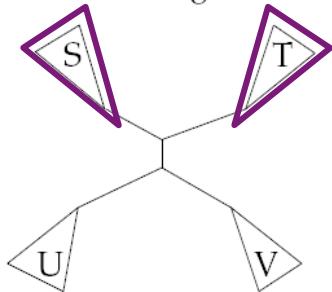
**Simplistic**  
swaps 2-adjacent branches

5-branches dissolved  
to isolate 4-subtrees

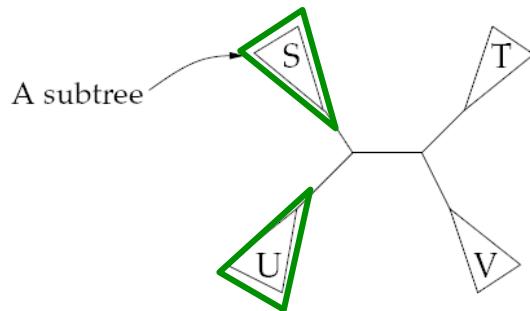
In this instance there are 3-options



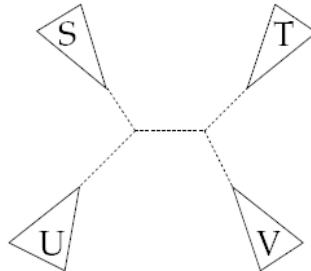
and reforming them in one of the two possible alternative ways:



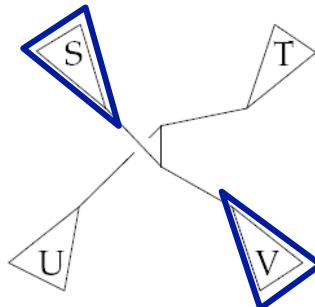
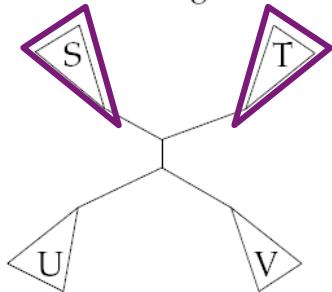
# NN-Nearest Neighbour Interchange



is rearranged by dissolving the connections to an interior branch



and reforming them in one of the two possible alternative ways:

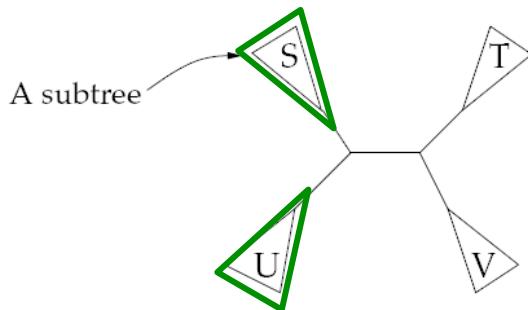


**Simplistic**  
swaps 2-adjacent branches

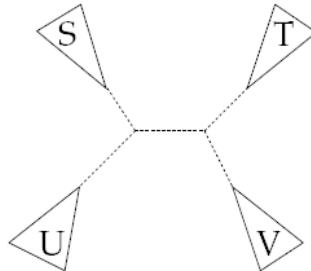
5-branches dissolved  
to isolate 4-subtrees

In this instance there are 3-options

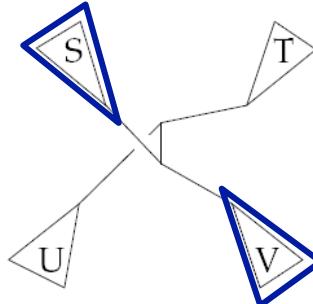
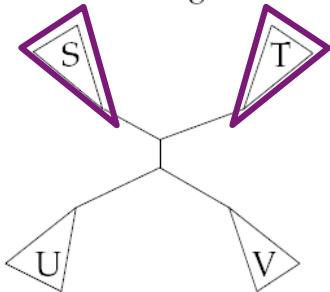
# NN-Nearest Neighbour Interchange



is rearranged by dissolving the connections to an interior branch



and reforming them in one of the two possible alternative ways:



Subtree

**Simplistic**

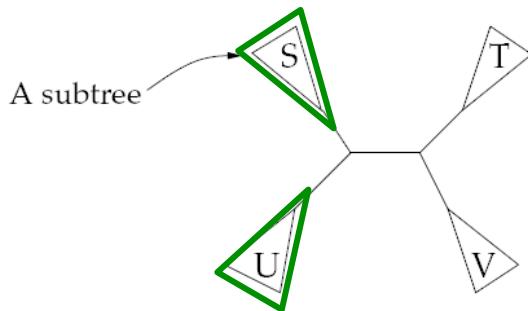
swaps 2-adjacent branches

5-branches dissolved  
to isolate 4-subtrees

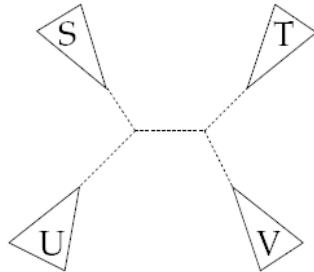
In this instance there are 3-options

But a tree with 20-terminals has  
34 neighbours under NNI...

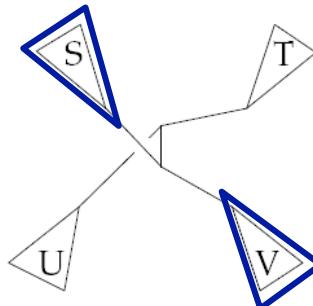
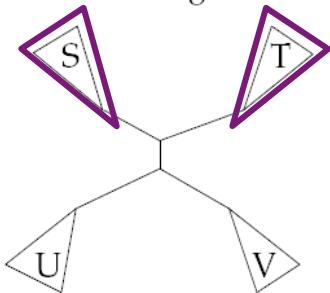
# NN-Nearest Neighbour Interchange



is rearranged by dissolving the connections to an interior branch



and reforming them in one of the two possible alternative ways:



Felsenstein 2004

## Simplistic

swaps 2-adjacent branches

5-branches dissolved  
to isolate 4-subtrees

In this instance there are 3-options

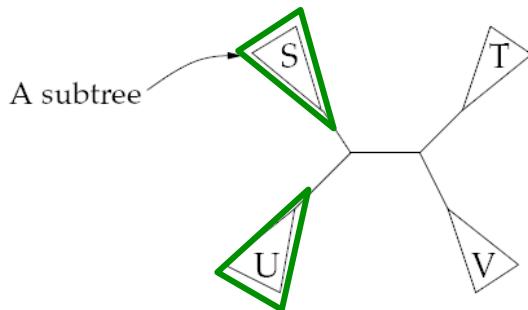
But a tree with 20-terminals has  
34 neighbours under NNI...

## Question of Greed:

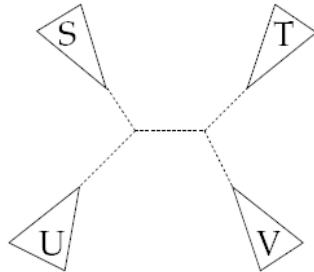
Choose 1<sup>st</sup> neighbour with improved  
 $\Theta^*$  that is encountered?

Or explore all neighbouring trees  
with tied  $\Theta^*$  values?

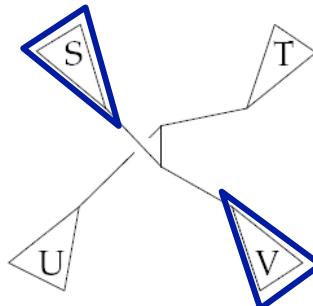
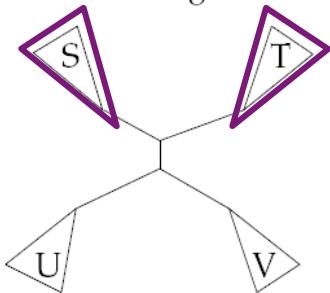
# NN-Nearest Neighbour Interchange



is rearranged by dissolving the connections to an interior branch



and reforming them in one of the two possible alternative ways:



Felsenstein 2004

## Simplistic

swaps 2-adjacent branches

5-branches dissolved  
to isolate 4-subtrees

In this instance there are 3-options

But a tree with 20-terminals has  
34 neighbours under NNI...

## Question of Greed:

Choose 1<sup>st</sup> neighbour with improved  
 $\Theta^*$  that is encountered?

FASTER

Or explore all neighbouring trees  
with tied  $\Theta^*$  values?

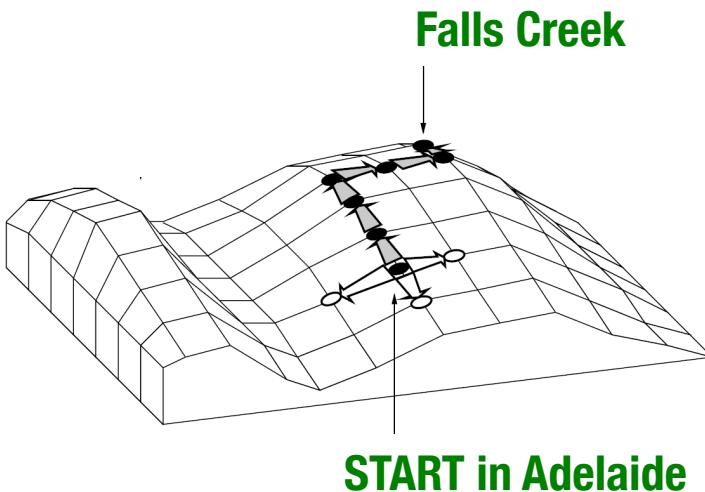
COMPREHENSIVE

# No Chairlift to Everest



# No Chairlift to Everest

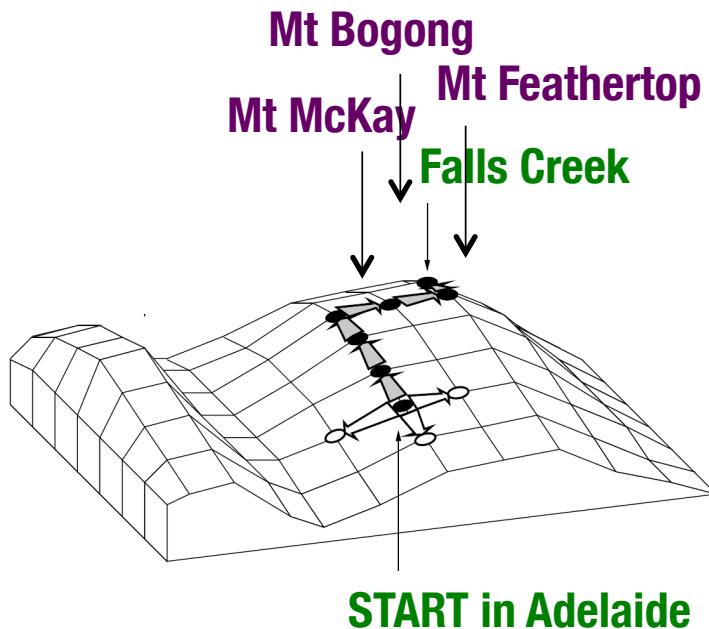
**Most parsimonious NNI:** Ski-lift gets you to Fall Creek Summit (easiest).



# No Chairlift to Everest

**Most parsimonious NNI:** Ski-lift gets you to Fall Creek Summit (easiest).

**Tied-rank NNI:** might get you to surrounding peaks of greater height.

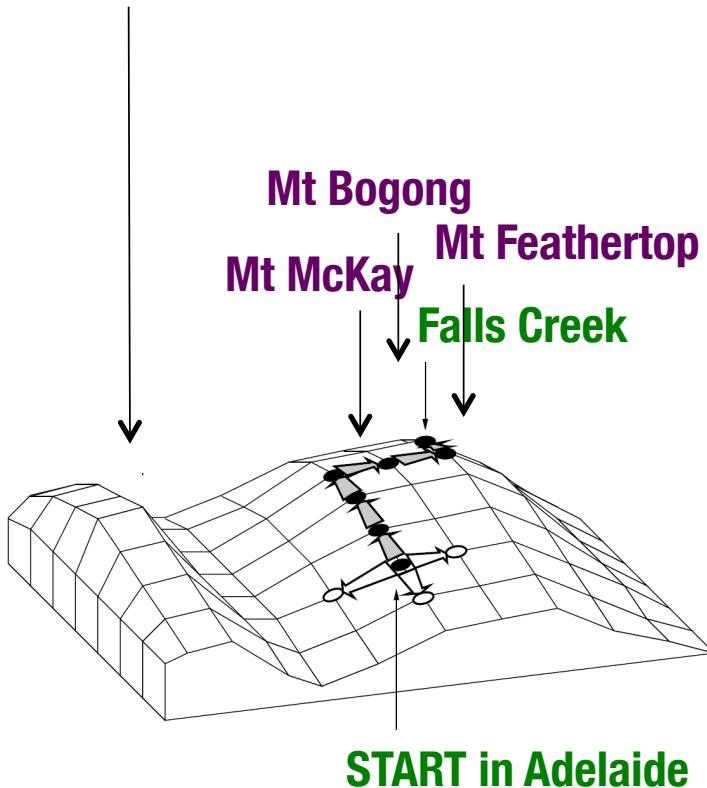


# No Chairlift to Everest

**Most parsimonious NNI:** Ski-lift gets you to Fall Creek Summit (easiest).

**Tied-rank NNI:** might get you to surrounding peaks of greater height.

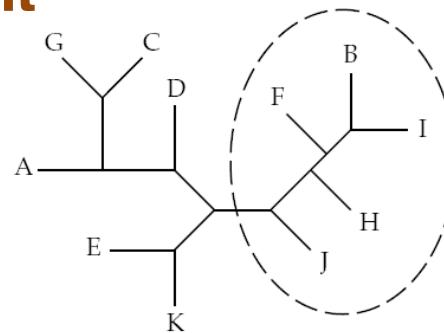
**Unlikely to get you to Everest**



# SPR - Subtree Pruning & Regrafting

## Stochastic branch rearrangement

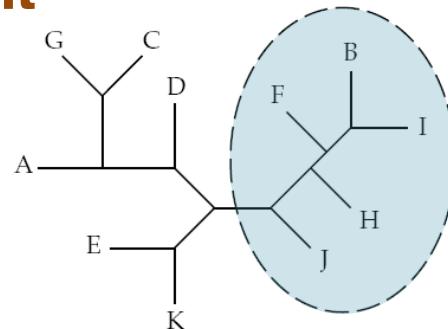
Remove a branch with a subtree  
(interior or exterior)



# SPR - Subtree Pruning & Regrafting

## Stochastic branch rearrangement

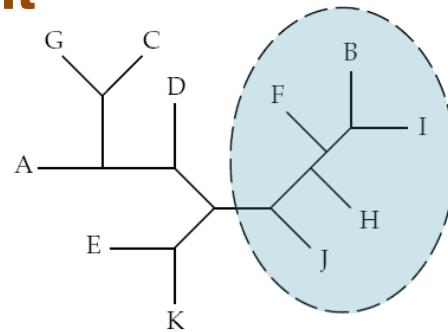
Remove a branch with a subtree  
(interior or exterior)



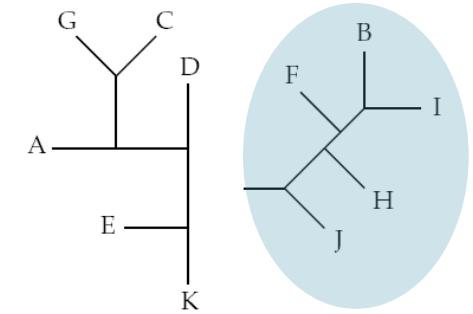
# SPR - Subtree Pruning & Regrafting

## Stochastic branch rearrangement

Remove a branch with a subtree  
(interior or exterior)



Break a branch, remove a subtree

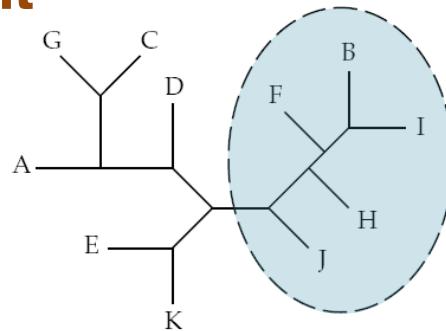


# SPR - Subtree Pruning & Regrafting

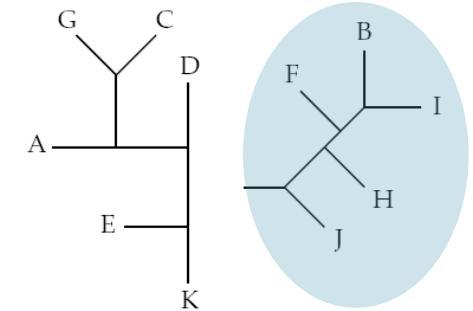
## Stochastic branch rearrangement

Remove a branch with a subtree  
(interior or exterior)

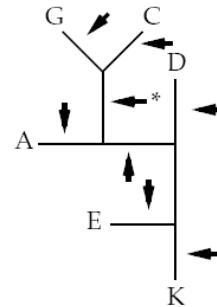
& re-insert it in all other places  
(8-other options)



Break a branch, remove a subtree



Add it in, attaching it to one (\*)  
of the other branches

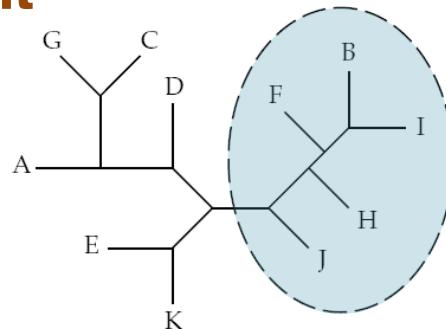


# SPR - Subtree Pruning & Regrafting

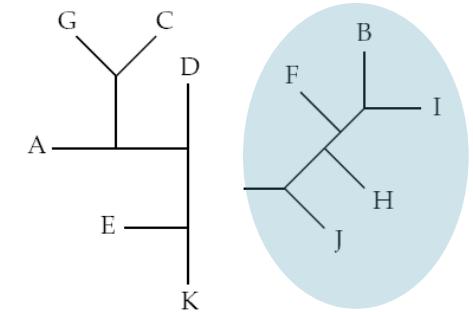
## Stochastic branch rearrangement

Remove a branch with a subtree  
(interior or exterior)

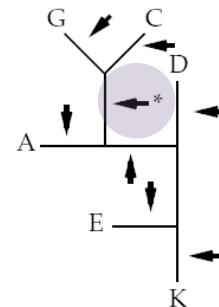
& re-insert it in all other places  
(8-other options)



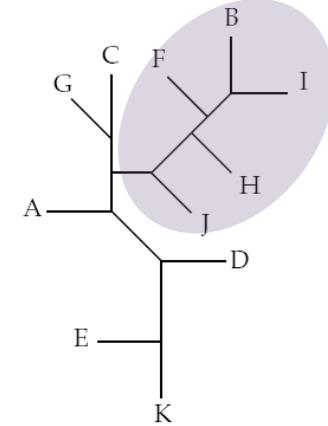
Break a branch, remove a subtree



Add it in, attaching it to one (\*)  
of the other branches



Here is the result:



Felsenstein 2004

# SPR - Subtree Pruning & Regrafting

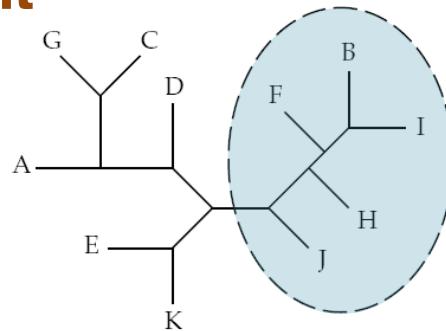
## Stochastic branch rearrangement

Remove a branch with a subtree  
(interior or exterior)

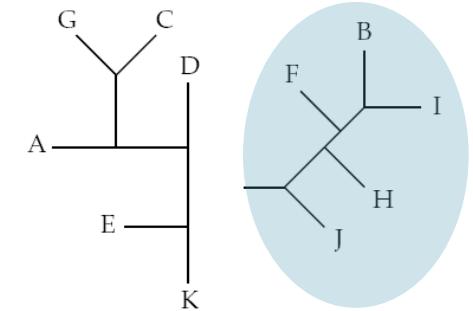
& re-insert it in all other places  
(8-other options)

**SPR examines 288 neighbours**

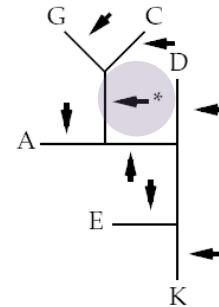
**NNI would examine 16 of these**



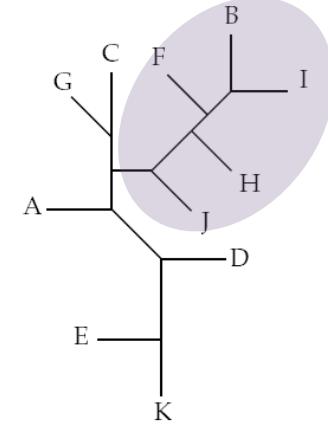
Break a branch, remove a subtree



Add it in, attaching it to one (\*)  
of the other branches

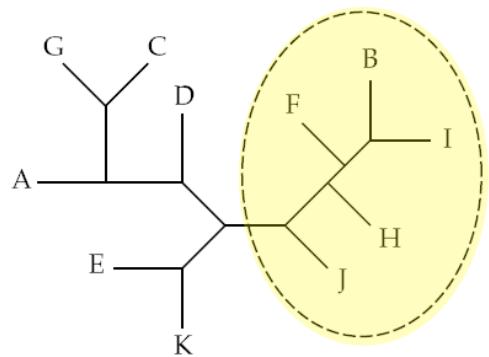


Here is the result:



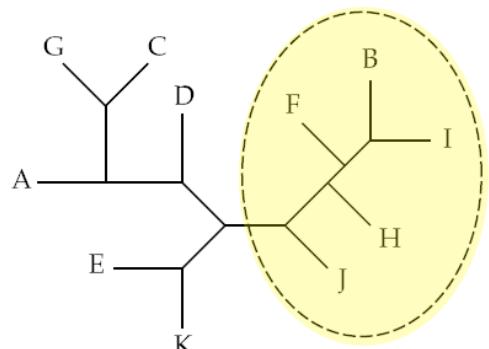
Felsenstein 2004

# TBR - Tree Bisection & Reconnection

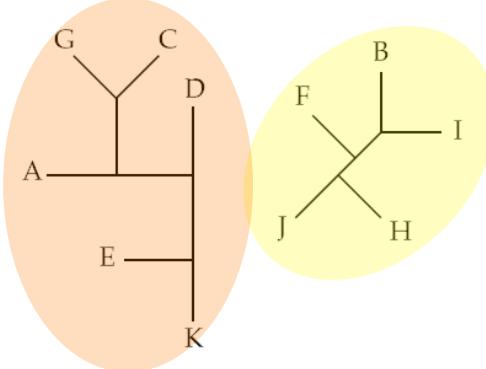


**Stochastic branch rearrangement 2°**  
Break interior branch  
into 2-subtrees

# TBR - Tree Bisection & Reconnection



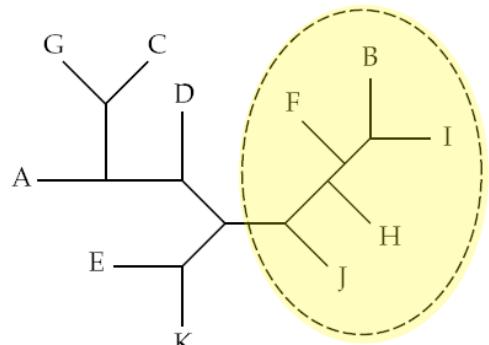
Break a branch, separate the subtrees



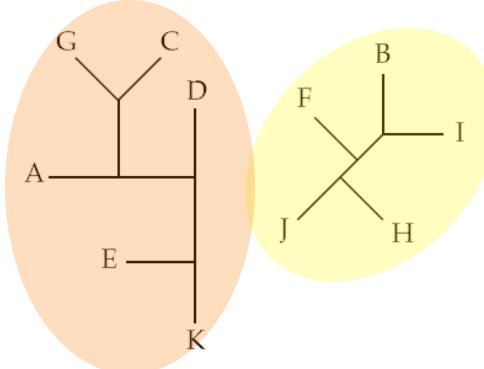
**Stochastic branch rearrangement 2°**

Break interior branch  
into 2-subtrees

# TBR - Tree Bisection & Reconnection



Break a branch, separate the subtrees

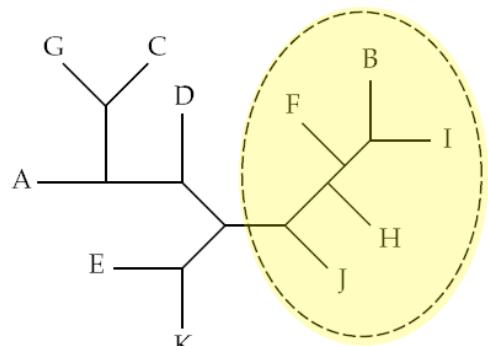


**Stochastic branch rearrangement 2°**

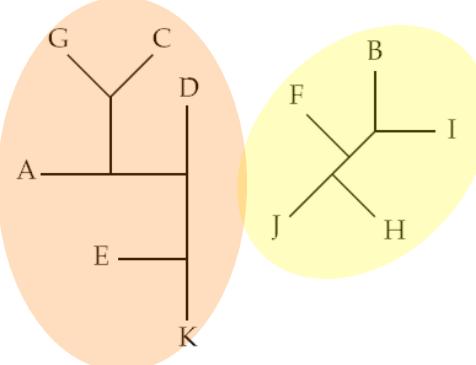
Break interior branch  
into 2-subtrees

2-fragments are reconnected using  
all possible branch insertion  
re-arrangements

# TBR - Tree Bisection & Reconnection



Break a branch, separate the subtrees

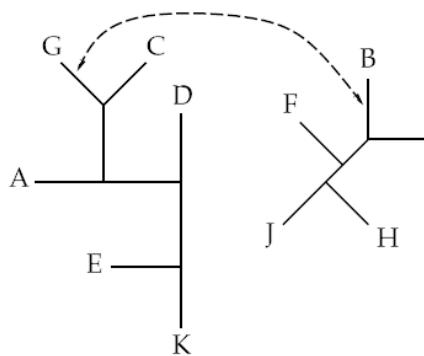


**Stochastic branch rearrangement 2°**

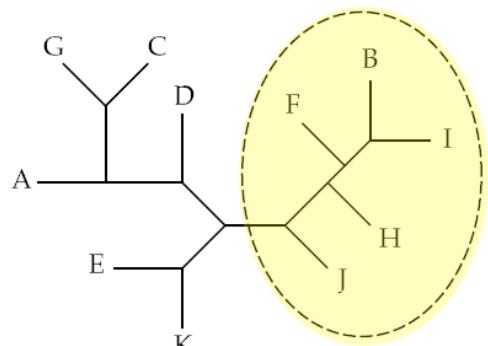
Break interior branch  
into 2-subtrees

2-fragments are reconnected using  
all possible branch insertion  
re-arrangements

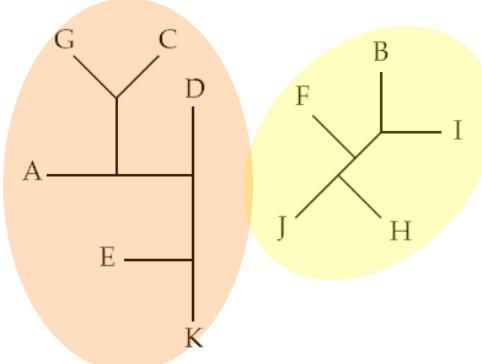
Connect a branch of one  
to a branch of the other



# TBR - Tree Bisection & Reconnection



Break a branch, separate the subtrees

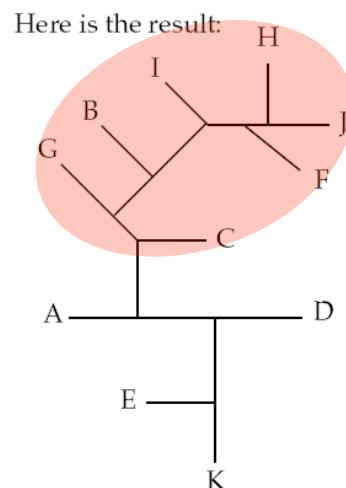
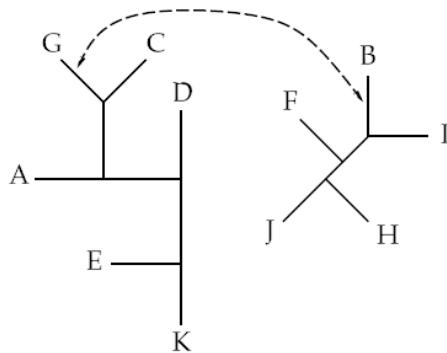


**Stochastic branch rearrangement 2°**

Break interior branch  
into 2-subtrees

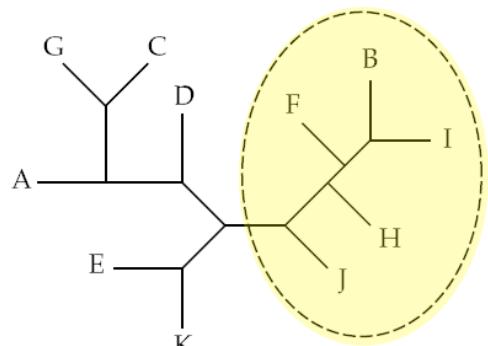
2-fragments are reconnected using  
all possible branch insertion  
re-arrangements

Connect a branch of one  
to a branch of the other

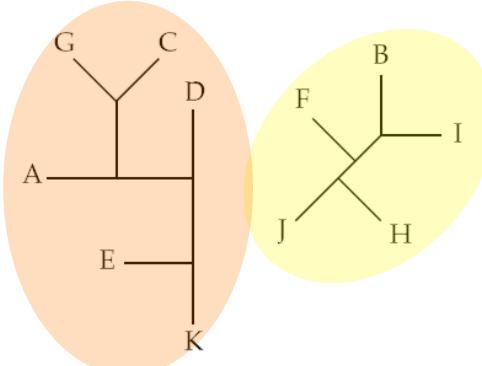


**TBR examines 296 internal branch neighbours**

# TBR - Tree Bisection & Reconnection



Break a branch, separate the subtrees

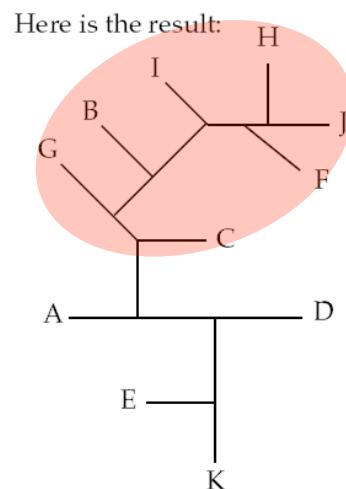
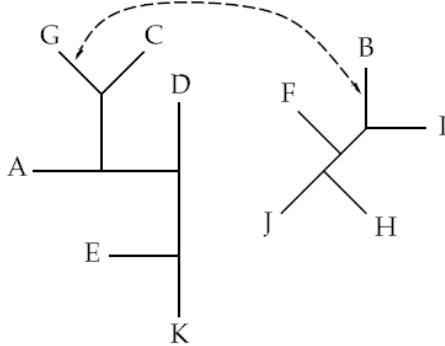


Stochastic branch rearrangement 2°

Break interior branch  
into 2-subtrees

2-fragments are reconnected using  
all possible branch insertion  
re-arrangements

Connect a branch of one  
to a branch of the other



TBR examines 296 internal branch neighbours

Allen & Steel (2001)  
Subtree transfer operations  
*Annals of Combinatorics* 5: 1-15

Algorithm determines efficient number of  
combinations needed for SPR & TBR  
to cover all possibilities

# Neighbourhood Exploration

## Method

Nearest Neighbour Interchange (NNI)

Subtree Pruning & Regrafting (SPR)

Tree Bisection & Reconnection (TBR)

## Exploration [n = tips]

$2(n - 3)$

$4(n - 3) (n - 2)$

$(2n_1 - 3) (2n_2 - 3)$

# Neighbourhood Exploration

## Method

Nearest Neighbour Interchange (NNI)

Subtree Pruning & Regrafting (SPR)

Tree Bisection & Reconnection (TBR)

## Exploration [n = tips]

$2(n - 3)$

$4(n - 3) (n - 2)$

$(2n_1 - 3) (2n_2 - 3)$

## Many other options:

Tree-fusing

Genetic algorithms

2-optimal trees

evolutionary simulation with fitness function

# Neighbourhood Exploration

## Method

Nearest Neighbour Interchange (NNI)

Subtree Pruning & Regrafting (SPR)

Tree Bisection & Reconnection (TBR)

## Exploration [n = tips]

$2(n - 3)$

$4(n - 3) (n - 2)$

$(2n_1 - 3) (2n_2 - 3)$

## Many other options:

Tree-fusing

Genetic algorithms

Sequential adding

Star decomposition

2-optimal trees

evolutionary simulation with fitness function

+ n & keep best tree, + n...

all n unrooted (star), group pairs of n to resolve

# Neighbourhood Exploration

## Method

Nearest Neighbour Interchange (NNI)

Subtree Pruning & Regrafting (SPR)

Tree Bisection & Reconnection (TBR)

## Exploration [n = tips]

$2(n - 3)$

$4(n - 3) (n - 2)$

$(2n_1 - 3) (2n_2 - 3)$

## Many other options:

Tree-fusing

Genetic algorithms

Sequential adding

Star decomposition

Simulated annealing

2-optimal trees

evolutionary simulation with fitness function

+ n & keep best tree, + n...

all n unrooted (star), group pairs of n to resolve  
temperature dependent Metropolis algorithm

# Neighbourhood Exploration

## Method

Nearest Neighbour Interchange (NNI)

Subtree Pruning & Regrafting (SPR)

Tree Bisection & Reconnection (TBR)

## Exploration [n = tips]

$2(n - 3)$

$4(n - 3) (n - 2)$

$(2n_1 - 3) (2n_2 - 3)$

## Many other options:

Tree-fusing

Genetic algorithms

Sequential adding

Star decomposition

Simulated annealing

2-optimal trees

evolutionary simulation with fitness function

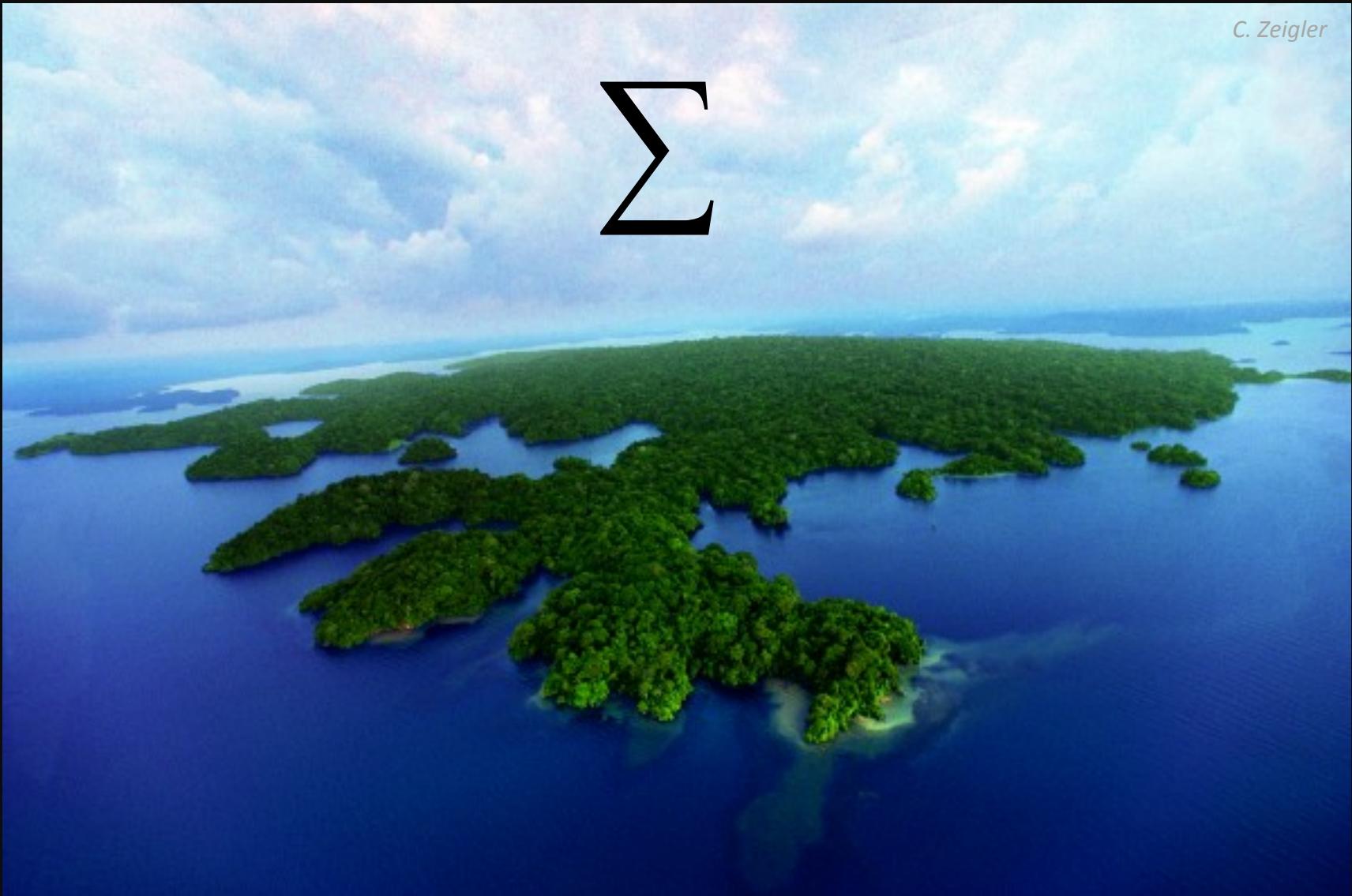
+ n & keep best tree, + n...

all n unrooted (star), group pairs of n to resolve  
temperature dependent Metropolis algorithm

**At this stage – we will largely consider the first three methods**

*C. Zeigler*

$$\sum$$



# Summarising Trees

## Continuous variables

Most model parameters are continuous  
summarised using standard stats: mean, variance, 95% credibility intervals

# Summarising Trees

## Continuous variables

Most model parameters are continuous  
summarised using standard stats: mean, variance, 95% credibility intervals

## Tree parameters

Tree topology parameters problematic  
many different possible branching patterns with high probability

# Summarising Trees

## Continuous variables

Most model parameters are continuous  
summarised using standard stats: mean, variance, 95% credibility intervals

## Tree parameters

Tree topology parameters problematic  
many different possible branching patterns with high probability

### Credible sets:

List topologies in decreasing order of probability

5-best trees:	probability
1	0.35
2	0.25
3	0.20
4	0.15
5	0.03

# Summarising Trees

## Continuous variables

Most model parameters are continuous  
summarised using standard stats: mean, variance, 95% credibility intervals

## Tree parameters

Tree topology parameters problematic  
many different possible branching patterns with high probability

### Credible sets:

List topologies in decreasing order of probability

Calculate number of trees that fall within a certain **cumulative** percentage

<b>5-best trees:</b>	<b>probability</b>	<b>cumulative probability</b>
1	0.35	0.35
2	0.25	0.60
3	0.20	0.80
4	0.15	0.95
5	0.03	0.98

# Summarising Trees

## Continuous variables

Most model parameters are continuous  
summarised using standard stats: mean, variance, 95% credibility intervals

## Tree parameters

Tree topology parameters problematic  
many different possible branching patterns with high probability

### Credible sets:

List topologies in decreasing order of probability

Calculate number of trees that fall within a certain **cumulative** percentage

	<b>5-best trees:</b>	<b>probability</b>	<b>cumulative probability</b>	
<b>2-trees</b>	1	0.35	0.35	<b>50% credibility set</b>
	2	0.25	0.60	
	3	0.20	0.80	
	4	0.15	0.95	
	5	0.03	0.98	

# Summarising Trees

## Continuous variables

Most model parameters are continuous  
summarised using standard stats: mean, variance, 95% credibility intervals

## Tree parameters

Tree topology parameters problematic  
many different possible branching patterns with high probability

### Credible sets:

List topologies in decreasing order of probability

Calculate number of trees that fall within a certain **cumulative** percentage

	<b>5-best trees:</b>	<b>probability</b>	<b>cumulative probability</b>	
<b>4-trees</b>	1	0.35	0.35	<b>95% credibility set</b>
	2	0.25	0.60	
	3	0.20	0.80	
	4	0.15	0.95	
	5	0.03	0.98	

# Summarising Trees

## Continuous variables

Most model parameters are continuous summarised using standard stats: mean, variance, 95% credibility intervals

## Tree parameters

Tree topology parameters problematic many different possible branching patterns

### Credible sets:

List topologies in order of probability

Calculate number of topologies

ONLY PRACTICAL IF THE MAJORITY OF PROBABILITY FALLS ON A SMALL NUMBER OF TOPOLOGIES

	5-best trees:	probability	cumulative probability	95% credibility set
4-trees	1	0.35	0.35	95% credibility set
	2	0.25	0.60	
	3	0.20	0.80	
	4	0.15	0.95	
	5	0.03	0.98	

# Summarising Trees

## Continuous variables

Most model parameters are continuous  
summarised using standard stats: mean, variance, 95% credibility intervals

## Tree parameters

Tree topology parameters problematic  
many different possible branching patterns with high probability

**Split (Clade) Frequencies:**

# Summarising Trees

## Continuous variables

Most model parameters are continuous  
summarised using standard stats: mean, variance, 95% credibility intervals

## Tree parameters

Tree topology parameters problematic  
many different possible branching patterns with high probability

### Split (Clade) Frequencies:

more common method because # splits < # topologies

# Summarising Trees

## Continuous variables

Most model parameters are continuous  
summarised using standard stats: mean, variance, 95% credibility intervals

## Tree parameters

Tree topology parameters problematic  
many different possible branching patterns with high probability

### Split (Clade) Frequencies:

more common method because # splits < # topologies

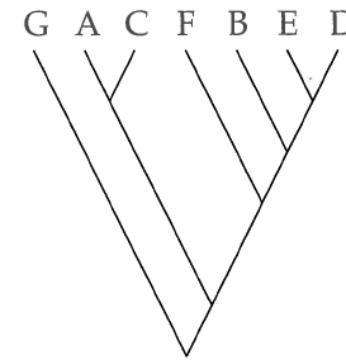
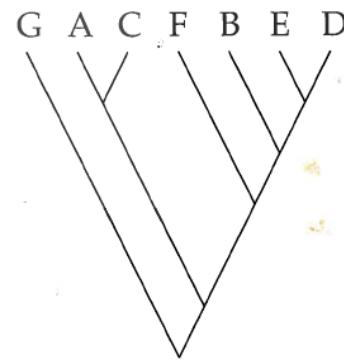
### 50% Majority Rule Tree

contains all topological splits in at least 50% of the trees sampled

convenient but is only ever an approximation of topological space

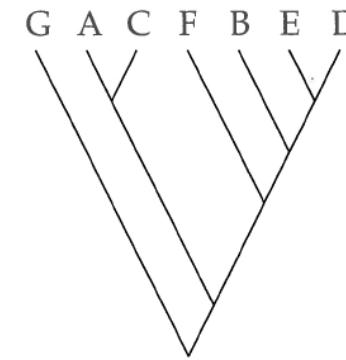
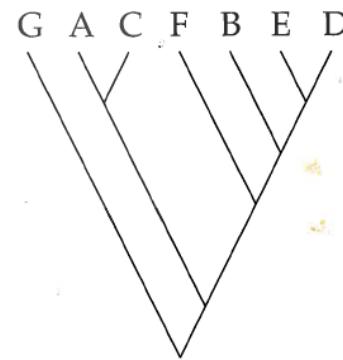
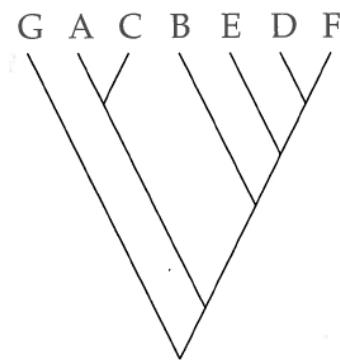
effectively a tree that summarises a series of trees

# Summarising Trees

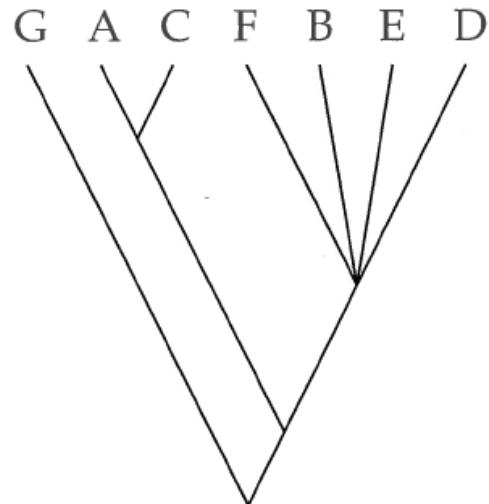


**3 Best trees = 80% credibility set**

# Summarising Trees

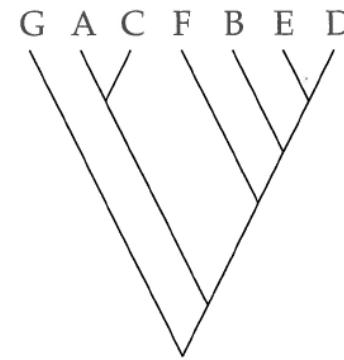


## Strict Consensus

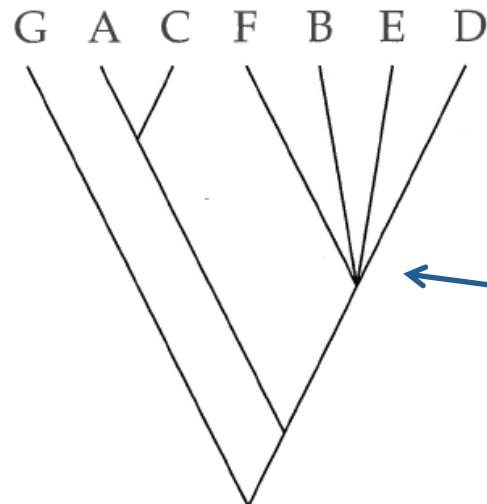


Felsenstein 2004

# Summarising Trees

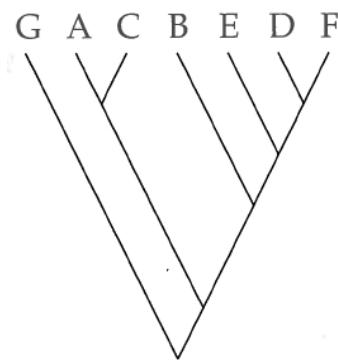


## Strict Consensus

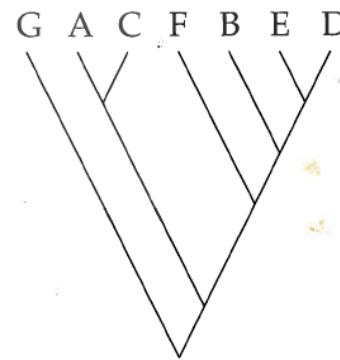


unresolved  
polytomy

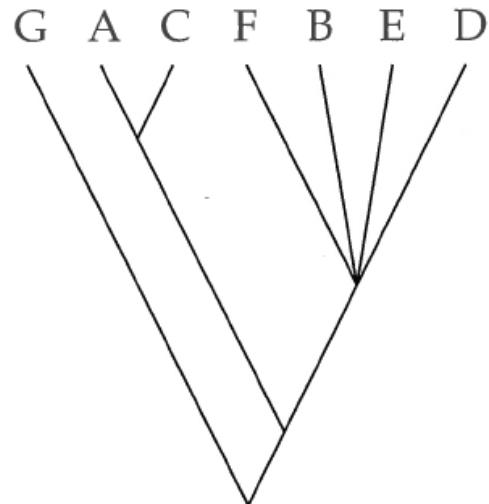
# Summarising Trees



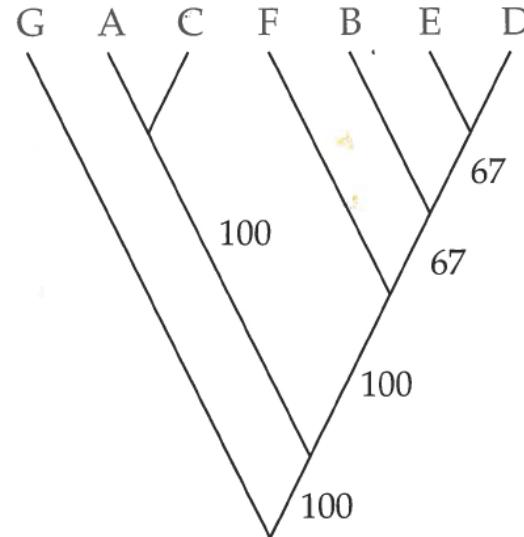
Strict Consensus



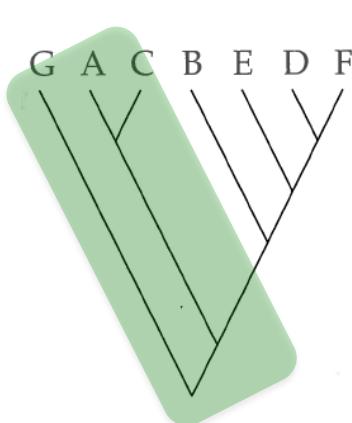
50% Majority Rule



Felsenstein 2004



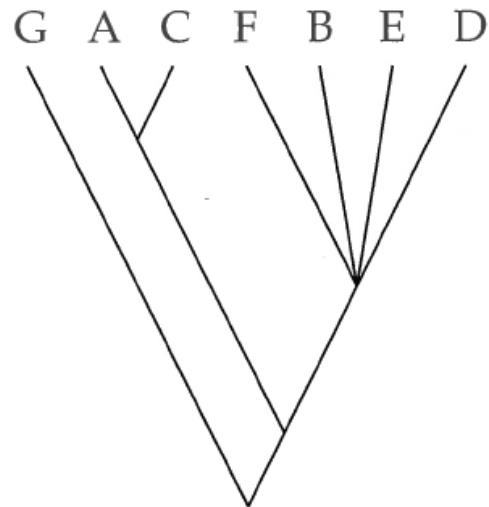
# Summarising Trees



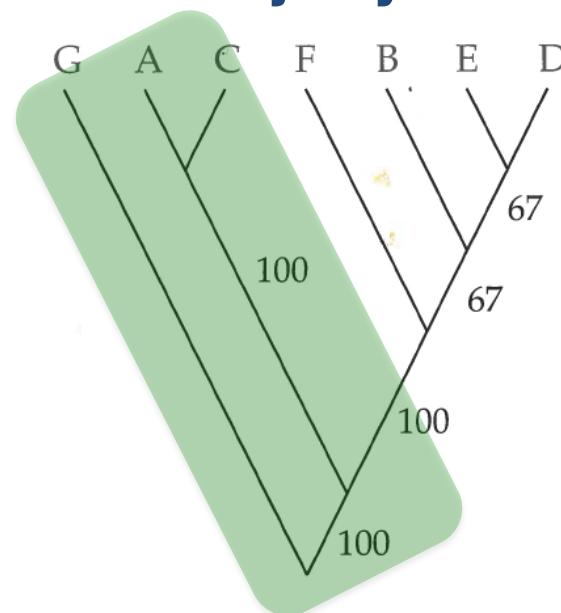
Strict Consensus



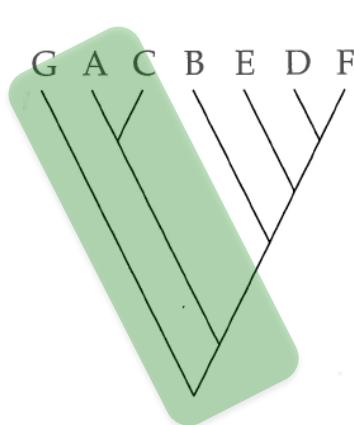
50% Majority Rule



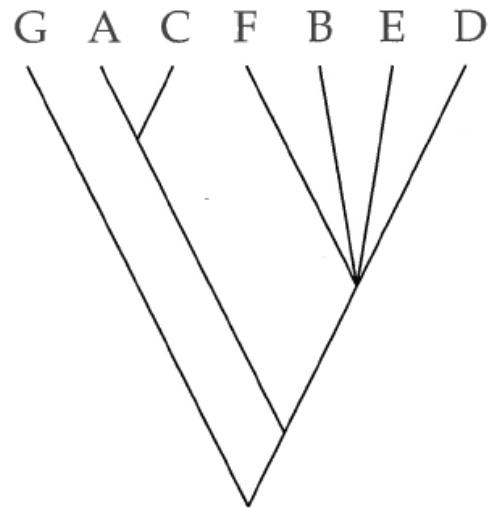
Felsenstein 2004



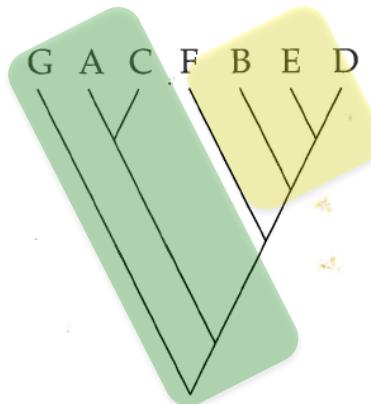
# Summarising Trees



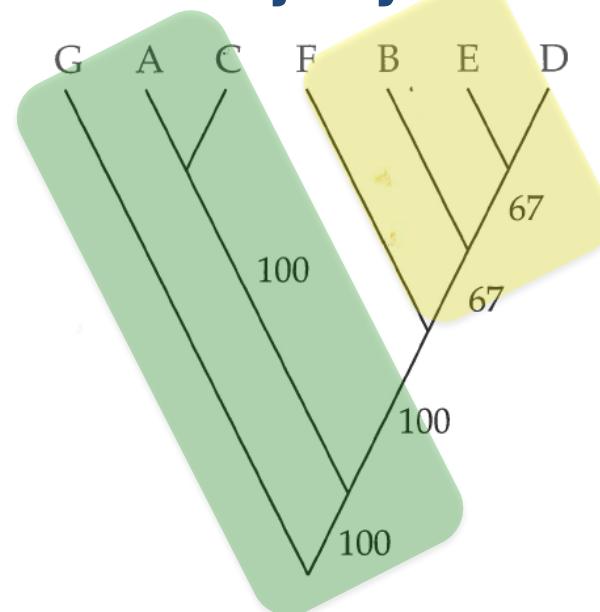
Strict Consensus



Felsenstein 2004



50% Majority Rule





# Tree Convergence

**Convergence why bother?**

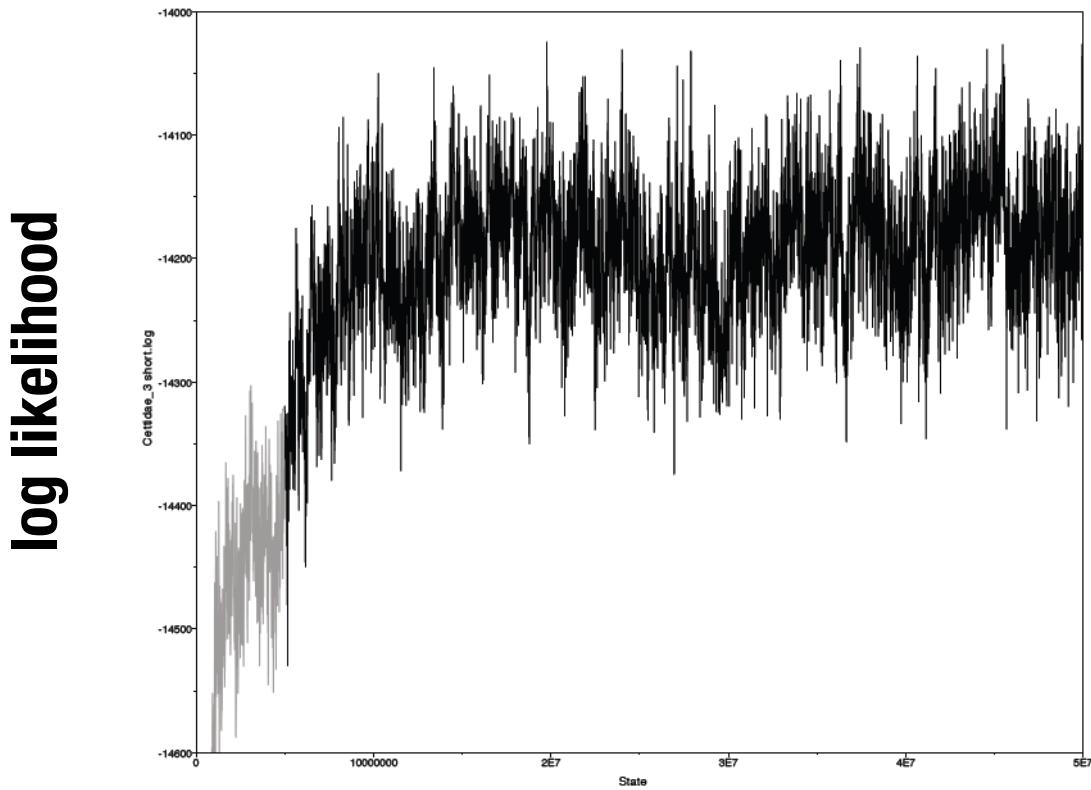
# Tree Convergence

Convergence why bother?



# Tree Convergence

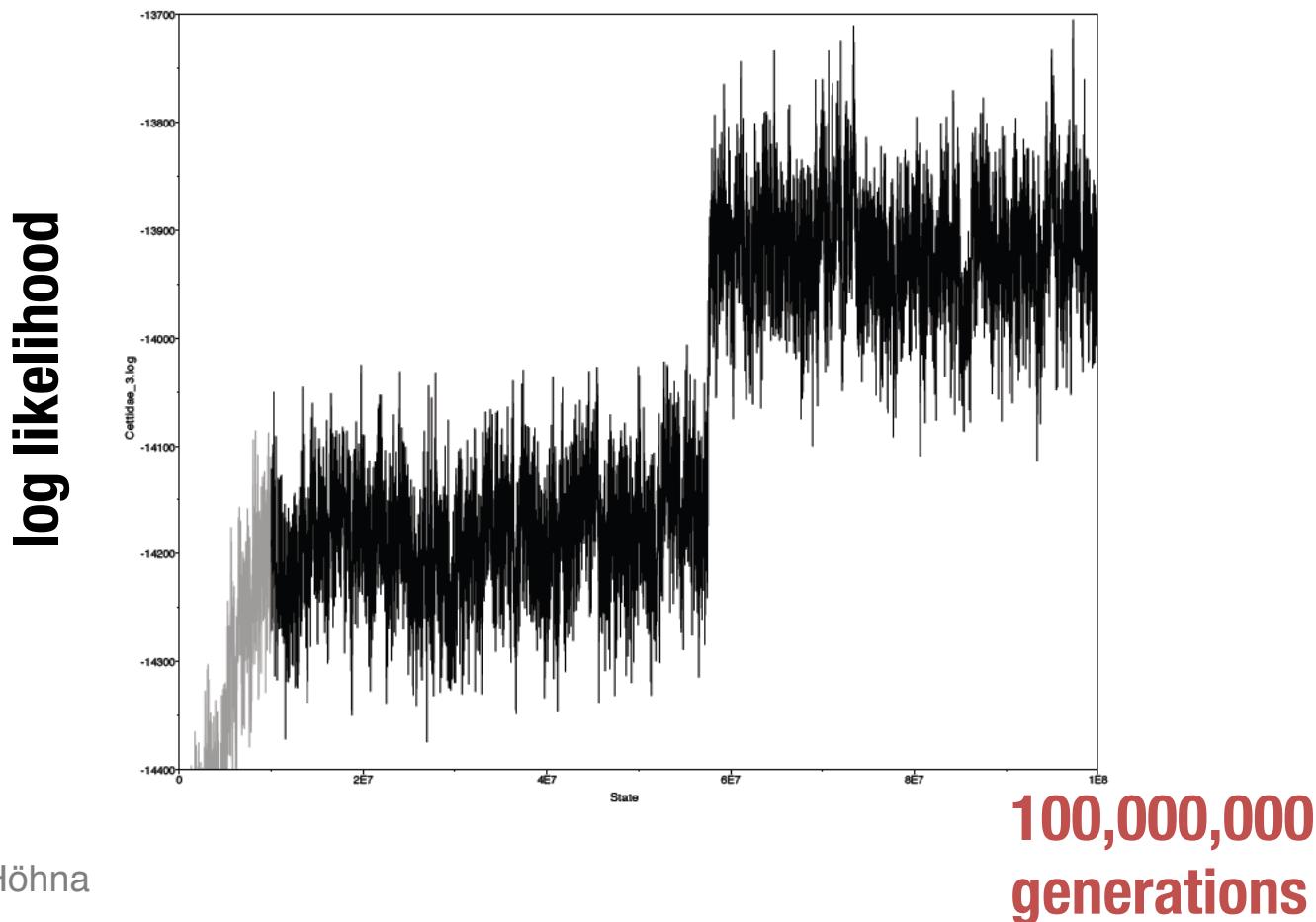
Convergence why bother?



50,000,000  
generations

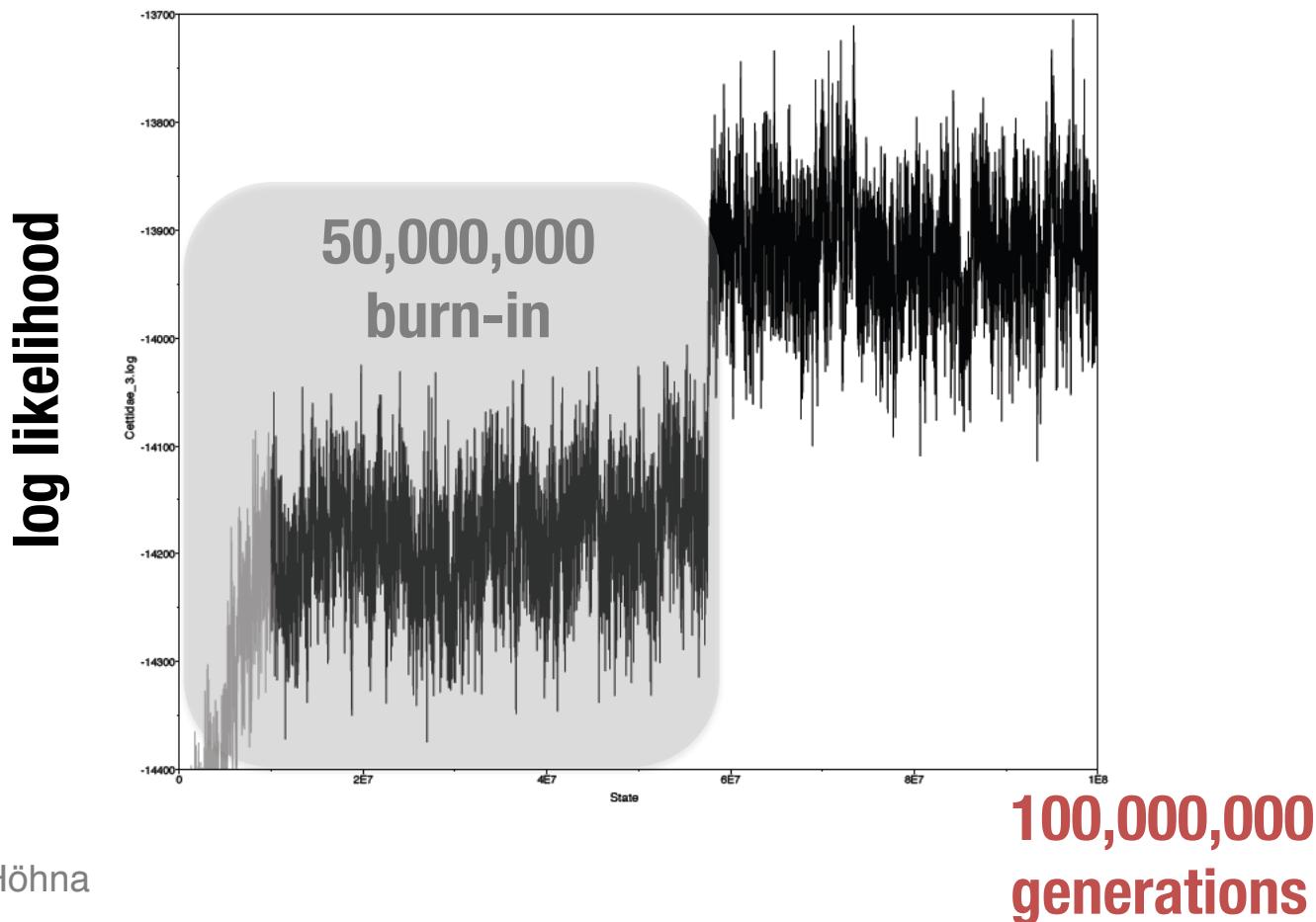
# Tree Convergence

Convergence why bother?



# Tree Convergence

Convergence why bother?



# Tree Convergence

## Convergence Assessment

Gives you an idea of how well the MCMC has sampled the posterior.

## General Diagnostics

For a single run:

Trace of log likelihoods

Effective Sample Sizes (ESS) of various parameters

**Tracer**

# Tree Convergence

## Convergence Assessment

Gives you an idea of how well the MCMC has sampled the posterior.

## General Diagnostics

For a single run:

Trace of log likelihoods

Effective Sample Sizes (ESS) of various parameters

**Tracer**

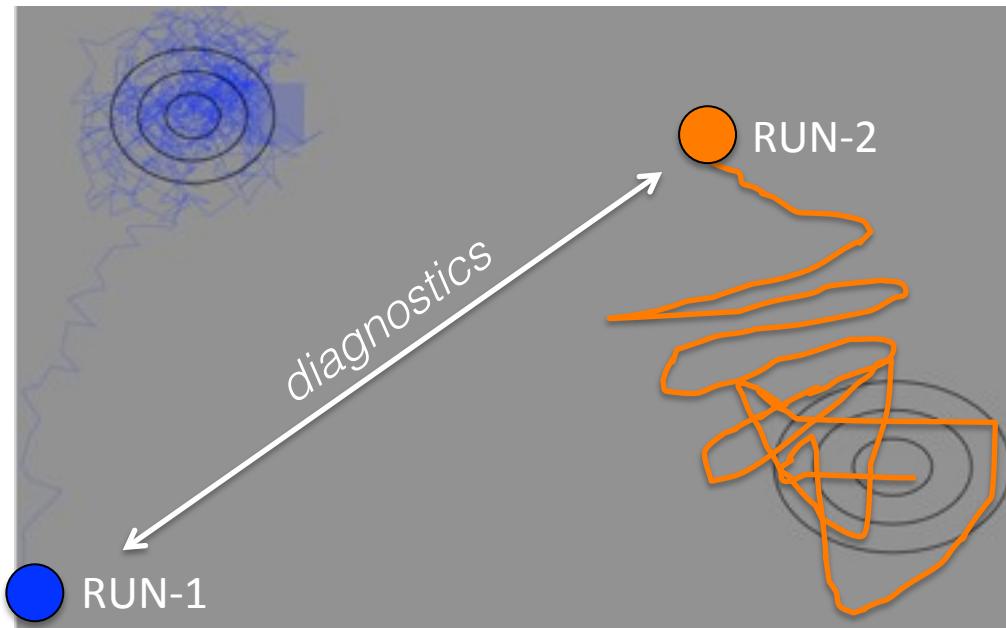
**Much more powerful when comparing multiple independent runs**

# Tree Convergence

## Solution 2:

Undertake parallel runs that start at independent points in parameter space

Examine the convergence between RUN-1 & RUN-2



# Tree Convergence

## Ideal outcome:

Verifiable convergence of RUN-1 & RUN-2 on the global optima



Adapted from Paul Lewis

# Tree Convergence

## Convergence Assessment

Gives you an idea of how well the MCMC has sampled the posterior.

## General Diagnostics

For a single run:

Trace of log likelihoods

Effective Sample Sizes (ESS) of various parameters

**Tracer**

**Much more powerful when comparing multiple independent runs**

MrBayes & RevBayes implements

**Potential Scale Reduction Factor**  
for continuous model parameters

# Tree Convergence

## Convergence Assessment

Gives you an idea of how well the MCMC has sampled the posterior.

## General Diagnostics

For a single run:

Trace of log likelihoods

Effective Sample Sizes (ESS) of various parameters

**Tracer**

**Much more powerful when comparing multiple independent runs**

MrBayes & RevBayes implements **Potential Scale Reduction Factor**  
for continuous model parameters

PSRF compares parameter variance (a) within and (b) between runs;  
so that when independent MCMC runs near convergence  
variances will be similar and...

PSRF will approach 1.0

# Tree Convergence

## Convergence Assessment

Gives you an idea of how well the MCMC has sampled the posterior.

## Tree-Specific Diagnostics

Comparison of tree parameter variance - as per PSRF?

# Tree Convergence

## Convergence Assessment

Gives you an idea of how well the MCMC has sampled the posterior.

## Tree-Specific Diagnostics

Comparison of tree parameter variance - as per PS

ONLY PRACTICAL IF THE MAJORITY  
OF PROBABILITY FALLS ON A SMALL  
NUMBER OF TOPOLOGIES

# Tree Convergence

## Convergence Assessment

Gives you an idea of how well the MCMC has sampled the posterior.

## Tree-Specific Diagnostics

Comparison of tree parameter variance - as per PS

Once again we can resort to split (clade) frequencies:

### Average Standard Deviation of Split Frequencies

ONLY PRACTICAL IF THE MAJORITY  
OF PROBABILITY FALLS ON A SMALL  
NUMBER OF TOPOLOGIES

# Tree Convergence

## Convergence Assessment

Gives you an idea of how well the MCMC has sampled the posterior.

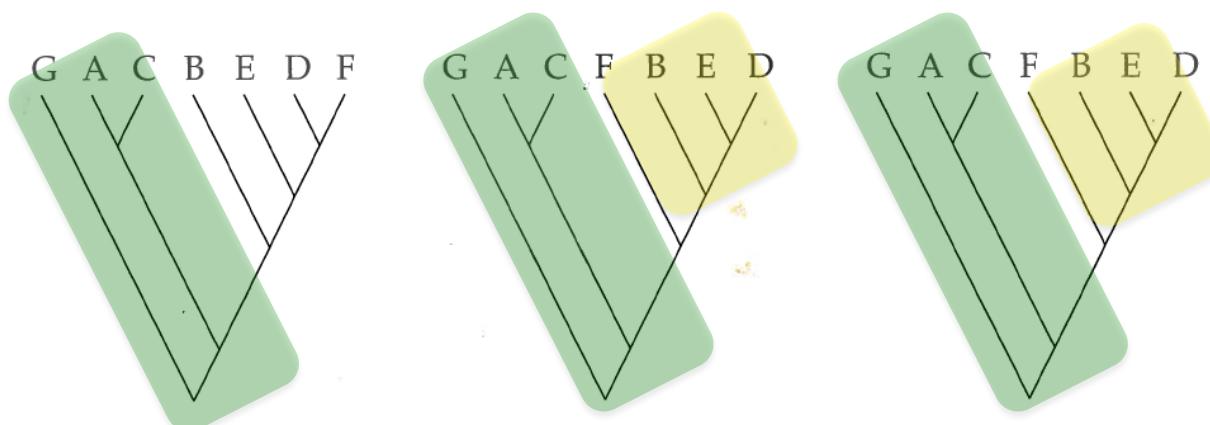
## Tree-Specific Diagnostics

Comparison of tree parameter variance - as per PS

Once again we can resort to split (clade) frequencies:

### Average Standard Deviation of Split Frequencies

Whereby, frequency of dominant clades represents the diagnostic



ONLY PRACTICAL IF THE MAJORITY OF PROBABILITY FALLS ON A SMALL NUMBER OF TOPOLOGIES

# Tree Convergence

## Convergence Assessment

Gives you an idea of how well the MCMC has sampled the posterior.

## Tree-Specific Diagnostics

Comparison of tree parameter variance - as per PS

Once again we can resort to split (clade) frequencies:

### Average Standard Deviation of Split Frequencies

Whereby, frequency of dominant clades represents the diagnostic

If two independent MCMC runs have converged on similar areas of tree-space they should exhibit similar clade frequencies

**Quite simply: calculation of the average std. dev. of  
clade frequencies between runs  
should approach zero.**

ONLY PRACTICAL IF THE MAJORITY  
OF PROBABILITY FALLS ON A SMALL  
NUMBER OF TOPOLOGIES

# Tree Convergence

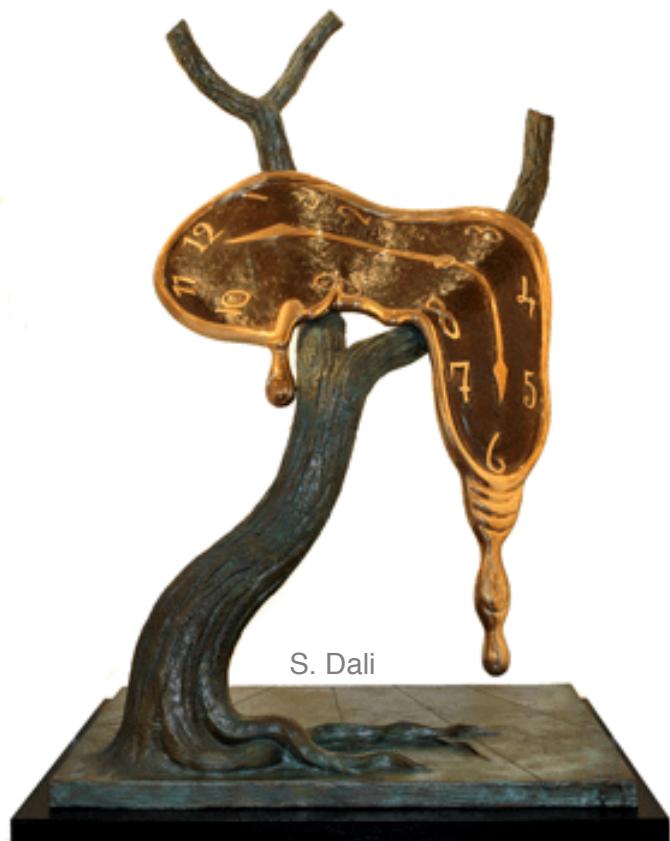
## Average Standard Deviation of Split Frequencies

Chain results:

```
1   -- (-7812.831) (-7523.685) [-7485.569] (-7700.309) * [-7832.045] (-7618.595) (-7776.608) (-7836.826)
100 -- (-6771.532) (-6857.529) (-6766.678) [-6682.527] * [-6506.277] (-6944.449) (-6784.126) (-6991.307) -- 0:01:39
200 -- (-6321.464) [-6179.561] (-6338.168) (-6272.242) * (-6339.400) (-6715.840) [-6265.329] (-6599.698) -- 0:01:38
300 -- (-6201.285) (-6084.899) (-6139.200) [-6049.061] * [-6073.056] (-6359.134) (-6106.834) (-6515.348) -- 0:01:37
400 -- (-6015.429) [-5924.177] (-6074.274) (-5967.236) * (-6002.941) (-6215.415) [-5980.096] (-6036.624) -- 0:01:12
500 -- (-5963.069) [-5851.415] (-6019.301) (-5948.768) * [-5925.423] (-6072.446) (-5925.549) (-5963.427) -- 0:01:16
600 -- (-5931.545) [-5802.472] (-5997.276) (-5927.204) * (-5899.590) (-6035.097) (-5904.327) [-5896.520] -- 0:01:18
700 -- (-5852.405) [-5782.934] (-5966.224) (-5880.975) * (-5890.220) (-6008.142) (-5876.956) [-5838.392] -- 0:01:19
800 -- (-5844.426) [-5754.416] (-5963.531) (-5860.182) * (-5883.474) (-5902.784) (-5790.765) [-5789.326] -- 0:01:20
900 -- (-5814.885) [-5741.101] (-5926.666) (-5855.916) * (-5870.806) (-5871.573) (-5783.012) [-5755.608] -- 0:01:20
1000 -- (-5811.949) [-5737.884] (-5888.234) (-5819.793) * (-5867.377) (-5851.693) (-5784.437) [-5749.264] -- 0:01:21
```

Average standard deviation of split frequencies: 0.073946

# What About Branch Lengths?



**Will discuss  
branch lengths & clock-analyses  
at a later date**

# What About Branch Lengths?

## Default proposal probabilities (MrBayes)

MCMC sampler uses the following moves:

With prob.	Chain will use move
1.00 %	Dirichlet (Revmat)
1.00 %	Slider (Revmat)
1.00 %	Dirichlet (Pi)
1.00 %	Slider (Pi)
2.00 %	Multiplier (Alpha)
10.00 %	ExtSPR (Tau,V)
10.00 %	ExtTBR (Tau,V)
10.00 %	ExtTBR (Tau,V)
10.00 %	ParsSPR (Tau,V)
40.00 %	Multiplier (V)
10.00 %	Nodeslider (V)
4.00 %	TLMultiplier (V)

MCMC will allocate 94%  
of search effort to changing  
tree branch lengths

# What About Branch Lengths?

## Default proposal probabilities (MrBayes)

MCMC sampler uses the following moves:

With prob.	Chain will use move
1.00 %	Dirichlet (Revmat)
1.00 %	Slider (Revmat)
1.00 %	Dirichlet (Pi)
1.00 %	Slider (Pi)
2.00 %	Multiplier (Alpha)
10.00 %	ExtSPR (Tau,V)
10.00 %	ExtTBR (Tau,V)
10.00 %	ExtTBR (Tau,V)
10.00 %	ParsSPR (Tau,V)
40.00 %	Multiplier (V)
10.00 %	Nodeslider (V)
4.00 %	TLMultiplier (V)

MCMC will allocate 94%  
of search effort to changing  
tree branch lengths

A relative indication of the importance of Tree Moves

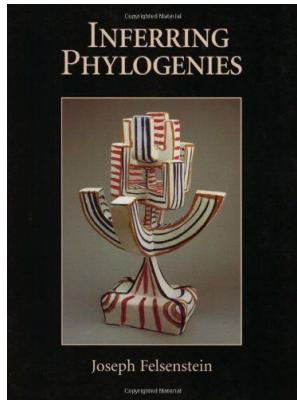
# MCMC Performance & Effort

**Performance of the MCMC algorithm used to asses joint posterior probability is the most important aspect of Bayesian Inference**

**MCMC algorithm expends large amount of search effort assessing tree parameter states**

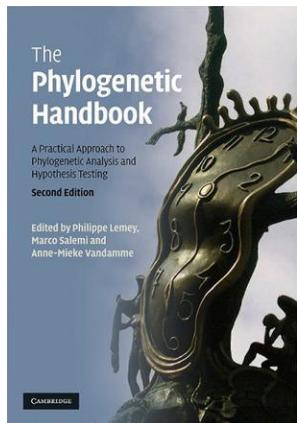


# References



Felsenstein J. 2004. *Inferring phylogenies*. Sinauer Associates: Sunderland MA.

Chapter 4



Ronquist F, van der Mark P & Huelsenbeck JP. 2009. Bayesian Phylogenetic analysis using MrBayes, pp. 210-266.

Chapter 7

In: Lemey P, Salemi M & Vandamme A-M (eds) *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press: Cambridge.