

**Molecular Phylogenetics Course**

## **Statistical frameworks for modelling in phylogenetics**

Jadranka Rota

Some slides from Paul Lewis (University of Connecticut, USA)

# **Maximum likelihood**

## Maximum Likelihood Estimation (MLE)

- Statistical method for estimating parameters of a model (e.g. mean and variance of a normal distribution)
- Originally developed by R. A. Fisher in the 1920s
- Adapted for phylogenetics by Joe Felsenstein

Joe Felsenstein



Joe Felsenstein

Born Joseph Felsenstein  
May 9, 1942 (age 79)

Alma mater University of Chicago

Known for PHYLIP  
Felsenstein's tree-pruning algorithm

[Wikipedia](#)

EVOLUTIONARY TREES FROM DNA-SEQUENCES - A MAXIMUM-LIKELIHOOD APPROACH

[FELSENSTEIN, J.](#)

[1981 | JOURNAL OF MOLECULAR EVOLUTION](#) 17 (6), pp.368-376

10,183

Citations

26

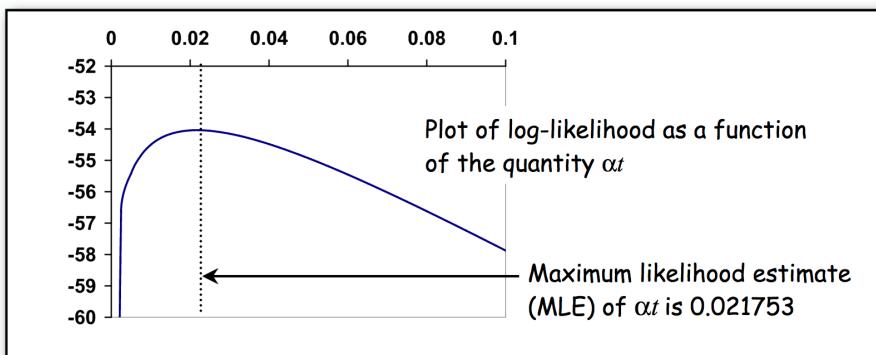
References

[Web of Science 2021-11-03](#)

## Likelihood of a hypothesis

- Likelihood ( $L$ ) is proportional to the probability ( $P$ ) of observing the data ( $D$ ) given a model ( $M$ ) – *conditional probability*
$$- L(M) = \Pr(D | M)$$
- We can examine this likelihood function to find where it is highest and identify the parameters of the model at this point -> Maximum Likelihood Estimates

## Likelihood function



Paul O. Lewis (2014 Woods Hole Workshop in Molecular Evolution)

22

## Likelihood of a hypothesis

- Likelihood (L) is proportional to the probability (P) of observing the data (D) given a model (M) – *conditional probability*
  - $L(M) = \Pr(D | M)$
- We can examine this likelihood function to find where it is highest and identify the parameters of the model at this point -> Maximum Likelihood Estimates
- In molecular phylogenetics, likelihood is the **probability of observing the sequences given our model** (e.g. GTR+G and our tree topology)

## Maximum Likelihood

- For reconstructing phylogenies

Model

– which tree topology ( $\tau$ ), branch lengths, and parameters of DNA evolution model ( $\theta$ ) (e.g. transition/transversion ratio, base frequencies, ...) are maximizing the probability of observing the sequences at hand?

Data

$$L(\tau, \theta) = \Pr(\text{Data} \mid \tau, \theta)$$

## ML analysis in short

- Tree topology is obtained
- Branch lengths and parameters of the DNA substitution model are optimized
- Different topologies (with branch lengths and DNA substitution model parameters optimized) are compared based on their likelihood as the optimality criterion
- The topology with the highest likelihood needs to be found

## Likelihood of a single sequence

First 32 nucleotides of the  $\psi\eta$ -globin gene of gorilla:

**GAAGTCCTTGAGAAATAACTGCACACACTGG**

$$L = \pi_G \pi_A \pi_A \pi_G \pi_T \pi_C \pi_T \pi_G \pi_A \pi_A \pi_A \pi_T \pi_A \pi_A \pi_C \pi_T \pi_G \pi_C \pi_A \pi_C \pi_A \pi_T \pi_G \pi_G \\ = \pi_A^{12} \pi_C^7 \pi_G^7 \pi_T^6$$

$$\ln L = 12 \ln(\pi_A) + 7 \ln(\pi_C) + 7 \ln(\pi_G) + 6 \ln(\pi_T)$$

We can already see by eye-balling this that the F81 model will fit better than the JC69 model because there are about twice as many As as there are Cs, Gs and Ts.

Paul O. Lewis 2005

## Likelihood ratio test

Find lnL under F81 model:

$$\begin{aligned} \ln L &= 12 \ln(\pi_A) + 7 \ln(\pi_C) + 7 \ln(\pi_G) + 6 \ln(\pi_T) \\ &= 12 \ln(0.375) + 7 \ln(0.21875) + 7 \ln(0.21875) + 6 \ln(0.1875) \\ &= -43.1 \end{aligned}$$

Find lnL under JC69 model:

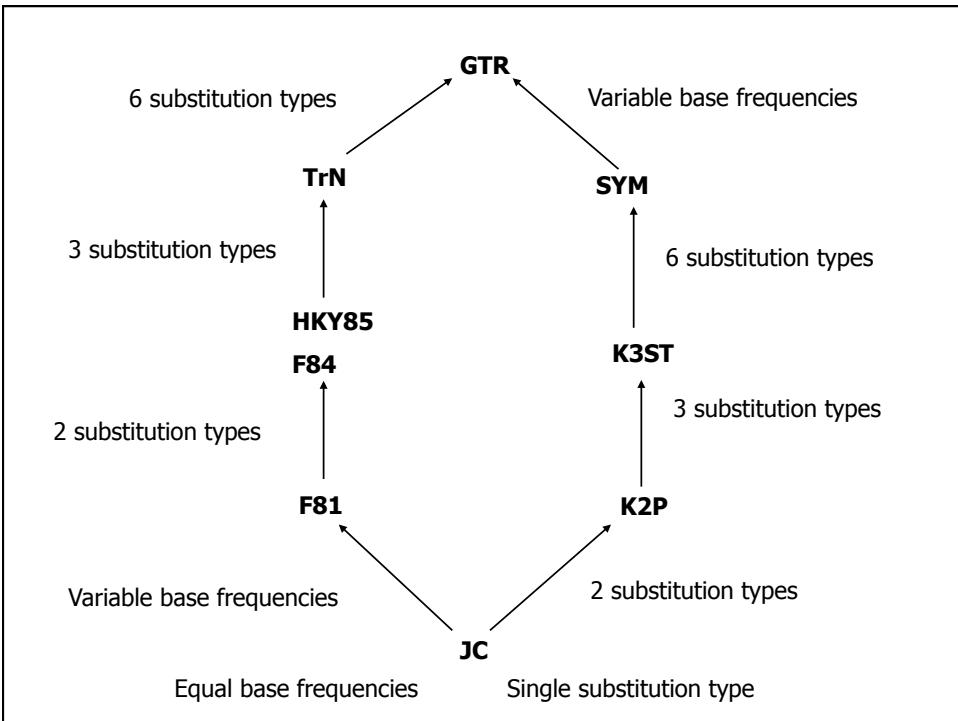
$$\begin{aligned} \ln L &= 12 \ln(\pi_A) + 7 \ln(\pi_C) + 7 \ln(\pi_G) + 6 \ln(\pi_T) \\ &= 12 \ln(0.25) + 7 \ln(0.25) + 7 \ln(0.25) + 6 \ln(0.25) \\ &= -44.4 \end{aligned}$$

F81 does fit better ( $-43.1 > -44.4$ ), but not significantly better ( $P = 0.457$ , chi-squared with 3 d.f.\*)

Find likelihood ratio test statistic:

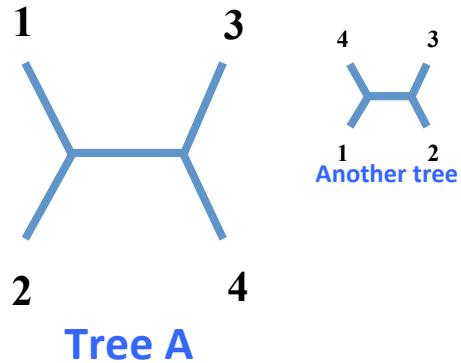
$$LR = -2(\ln L_{JC69} - \ln L_{F81}) = -2[-44.4 - (-43.1)] = 2.6$$

\*The number of degrees of freedom equals the difference between the two models in the number of parameters. In this case, F81 has 3 parameters and JC69 has 0, so d.f. = 3 - 0 = 3



## Maximum Likelihood tree reconstruction

**1** CGAGAC  
**2** AGCGAC  
**3** AGATTC  
**4** GGATAG



What is the likelihood that **Tree A** (rather than another tree) could have generated the sequence alignment?

# Likelihood of the simplest tree

sequence 1 ————— sequence 2

To keep things simple, assume that the sequences are only 2 nucleotides long:

$$\begin{aligned}
 L &= L_1 L_2 \\
 &= \left[ \left( \frac{1}{4} \right) \left( \frac{1}{4} + \frac{3}{4} e^{-4\alpha t} \right) \right] \left[ \left( \frac{1}{4} \right) \left( \frac{1}{4} - \frac{1}{4} e^{-4\alpha t} \right) \right]
 \end{aligned}$$

$\Pr(G)$      $\Pr(G|G, \alpha t)$      $\Pr(A)$      $\Pr(G|A, \alpha t)$

Note that we are NOT assuming independence here

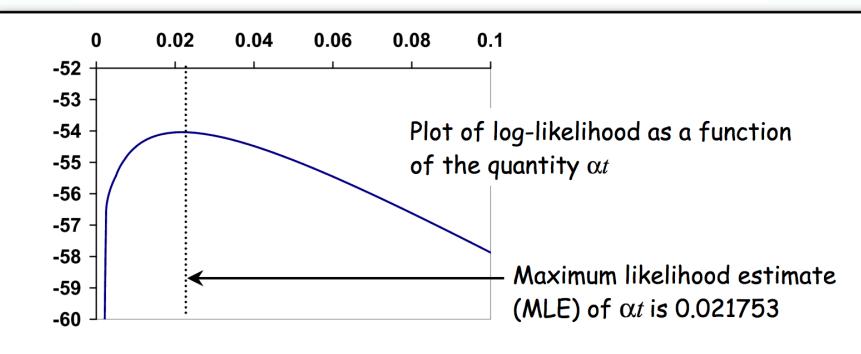
21

# Maximum likelihood estimation

First 32 nucleotides of the  $\psi\eta$ -globin gene of gorilla and orangutan:

gorilla    GAAGTCCTTGAGAAATAAACTGCACACACTGG  
orangutan    GGACTCCTTGAGAAATAAACTGCACACACTGG

$$L = \left[ \left( \frac{1}{4} \right) \left( \frac{1}{4} + \frac{3}{4} e^{-4\alpha t} \right) \right]^{30} \left[ \left( \frac{1}{4} \right) \left( \frac{1}{4} - \frac{1}{4} e^{-4\alpha t} \right) \right]^2$$

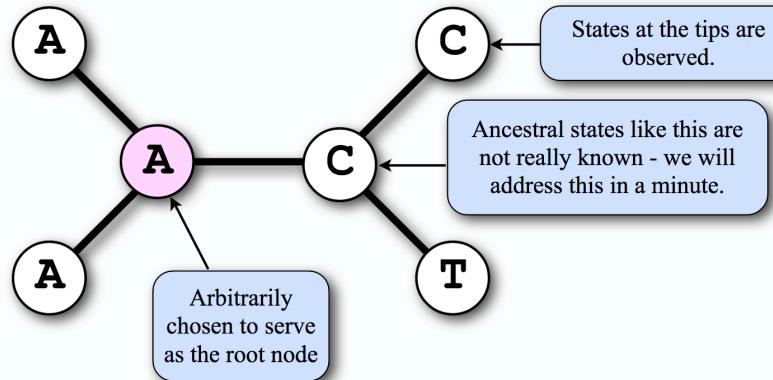


Paul O. Lewis (2014 Woods Hole Workshop in Molecular Evolution)

22

## Likelihood of an unrooted tree

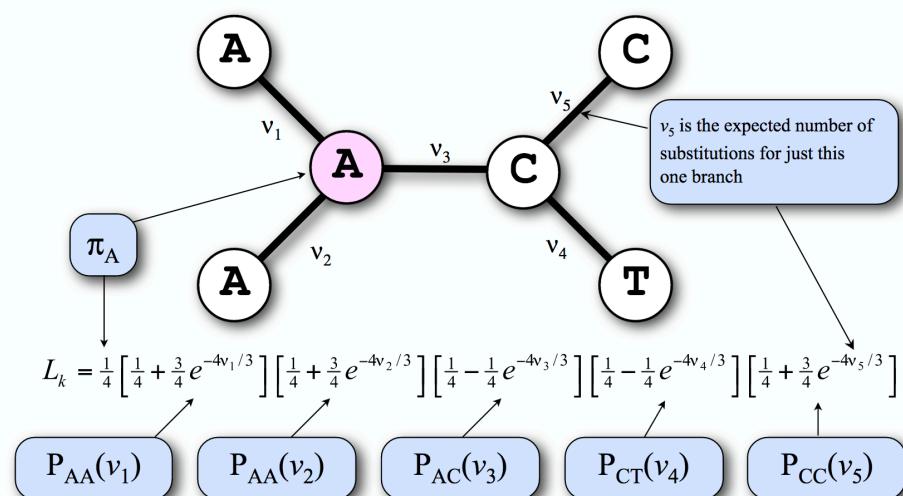
(data shown for only one site)



Paul O. Lewis (2014 Woods Hole Workshop in Molecular Evolution)

27

## Likelihood for site $k$

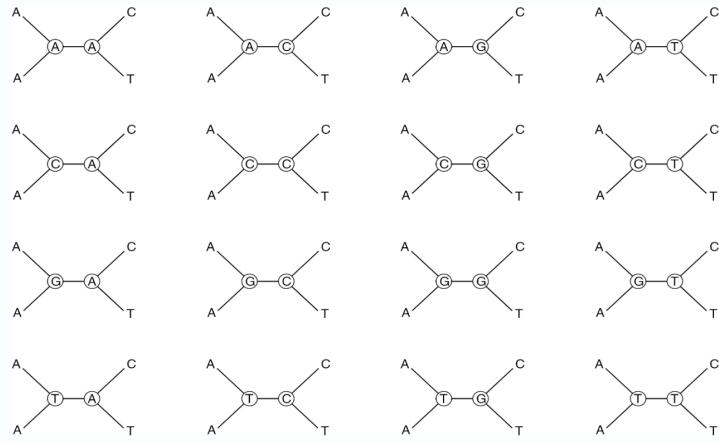


Paul O. Lewis (2014 Woods Hole Workshop in Molecular Evolution)

Note use of the AND probability rule

28

Brute force approach would be to calculate  $L_k$  for all 16 combinations of ancestral states and sum them

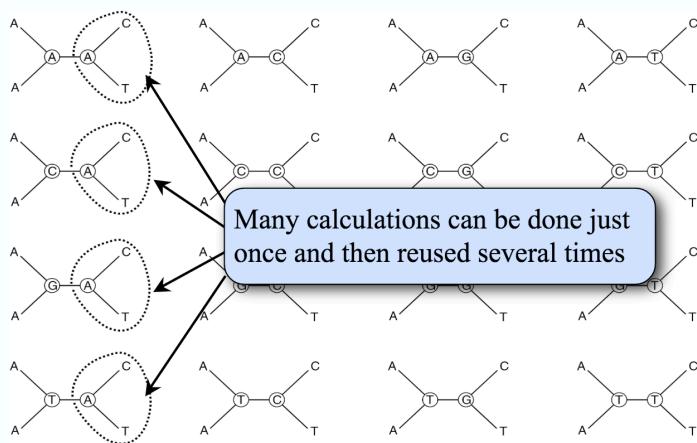


Note use of the OR probability rule

Paul O. Lewis (2014 Woods Hole Workshop in Molecular Evolution)

29

## Pruning algorithm (same result, less time)



Felsenstein, J. 1981. Evolutionary trees from DNA sequences:  
a maximum likelihood approach. *Journal of Molecular Evolution* 17:368-376

Paul O. Lewis (2014 Woods Hole Workshop in Molecular Evolution)

30

## ML analysis in summary

- The likelihood for each site in the alignment for a given topology, branch lengths and the DNA substitution model is calculated
- The probability of the single site is the sum of probabilities of each scenario, taking into account all of the possible nucleotides that may have existed as states at the internal nodes
- The likelihood for a given tree topology for the whole alignment is the product of the likelihoods for each site

## Finding the maximum likelihood of a tree

- Problem: the number of possible trees (e.g. for 10 taxa, 2 million unrooted trees possible; for 60 taxa, more possible trees than atoms in the universe!)

- for each tree topology we need to identify the maximum likelihood estimate for evolutionary parameters and branch lengths
  - then compare the likelihood among all the trees
    - This is simply computationally not feasible

► Solution:

- Currently no method guarantees finding the best tree
  - Starting tree made usually using MP or NJ
  - Heuristic approaches are used:
    - e.g. NNI = Nearest Neighbour Interchange, SPR = subtree pruning and regrafting, TBR = tree-bisection and reconnection

## Typical assumptions of ML substitution models

- The probability of any change is independent of the prior history of the site ([a Markov Model](#))
- Relative frequencies of A, G, C, and T are at equilibrium ([stationarity](#)) - S
- Change is [time reversible](#) e.g. the rate of change of A to T is the same as T to A - R
- Substitution probabilities do not change with time or over the tree ([a homogeneous Markov process](#)) – H
- [SRH](#) – we assume that sequence evolution is stationary, reversible, and homogeneous or SRH

## Advantages of ML

- Appropriate for DNA sequences: can be reasonably modelled by stochastic processes  
    ⇒statistical description of the stochastic processes
- statistically well understood
- estimation method least affected by sampling error
- can evaluate different tree topologies (vs. NJ)
- uses all the sequence information (vs. Distance)

## Disadvantages of ML (?)

- Very computationally intensive (less of an issue nowadays)
- Potentially problematic when missing data not randomly distributed (*Simmons, M.P., 2011. Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. Cladistics. 27:1-15*)
- Questionably applicable to complex data like morphology
- Philosophically less well established compared to parsimony
- Compared to Bayesian, there is only one tree per tree search and no direct test for robustness
  - > bootstrap!

### How often are the SRH assumptions broken?

The Prevalence and Impact of Model Violations in Phylogenetic Analysis 

Suha Naser-Khdour , Bui Quang Minh, Wenqi Zhang, Eric A Stone, Robert Lanfear

*Genome Biology and Evolution*, Volume 11, Issue 12, December 2019, Pages 3341–3352, <https://doi.org/10.1093/gbe/evz193>

Published: 19 September 2019 Article history 

- Review of 35 published phylogenetic datasets
- 23% of them reject the SRH assumptions
- The authors partitioned the data into its various subsets (e.g. 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> codon positions, nuclear, mitochondrial, introns, exons, etc.)
- SRH tests done on each partition
- In 25% of datasets, tree topology different between the partitions that **do** and **do not violate** these assumptions

The Prevalence and Impact of Model Violations in Phylogenetic Analysis 

Suha Naser-Khdour , Bui Quang Minh, Wenqi Zhang, Eric A Stone,  
Robert Lanfear

*Genome Biology and Evolution*, Volume 11, Issue 12, December 2019, Pages 3341–3352, <https://doi.org/10.1093/gbe/evz193>

Published: 19 September 2019 Article history ▾

**Table 2**

The Proportion of Partitions That Failed At Least One of the Three Tests—MaxSymTest, MaxSymTest<sub>mar</sub>, and MaxSymTest<sub>int</sub>

Type/Genome	Nuclear	Mitochondrial	Plastid	Virus
First codon positions	20.2%	27.6%	33.3%	25.0%
Second codon positions	21.0%	7.4%	0.0%	25.0%
Third codon positions	76.6%	44.8%	0.0%	75.0%
Other (e.g., intron)	27.8%	100.0%	0.0%	
rRNA	30.0%	25.0%		
UCE	22.5%			
tRNA		0.0%		

- Take-home message: test your partitions for model violations!

Maximum Likelihood should be seen as a tree estimation procedure instead of a tree reconstruction

“we are making a *best estimate* of an evolutionary history based on incomplete information” Swofford, 1990

## **Some ML programs**

- IQ-Tree
- RAxML
- PhyML

## **A Bayesian Approach to Phylogenetics**

## A Bayesian approach compared to ML

- The likelihood is the probability of observing the data given a hypothesis
  - $L = \Pr(D | \theta)$ .
- In ML we search for the parameter values of the model that maximize the likelihood function
- In a Bayesian analysis, we get the probability of a hypothesis given the data (probability of the tree given the sequences)
  - We combine the likelihood of a given hypothesis with a prior expectation for this hypothesis to obtain a posterior probability of the hypothesis

## Bayes' rule in statistics

$$\Pr(\theta|D) = \frac{\Pr(D|\theta) \Pr(\theta)}{\sum_{\theta} \Pr(D|\theta) \Pr(\theta)}$$

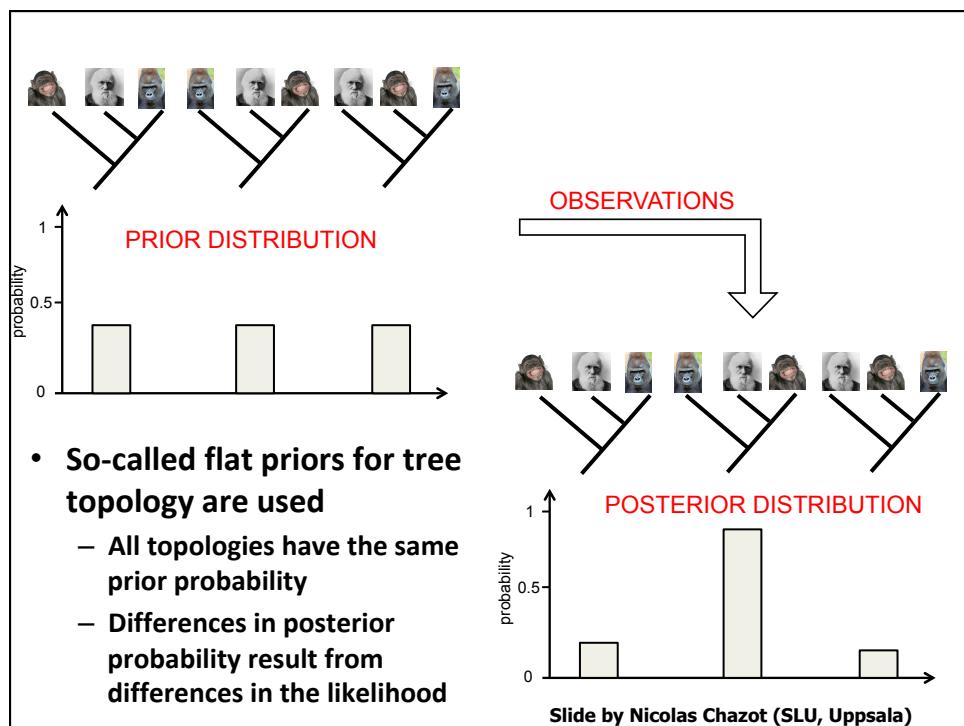
Diagram illustrating Bayes' rule:

- Likelihood of hypothesis  $\theta$  (blue box)
- Prior probability of hypothesis  $\theta$  (orange box)
- Posterior probability of hypothesis  $\theta$  (purple box)
- Marginal probability of the data (marginalizing over hypotheses) (green box)

Arrows point from the Likelihood and Prior boxes to the numerator of the equation. Arrows point from the Posterior and Marginal boxes to the denominator of the equation.

# Bayesian inference in general

- D stands for data
- $\Theta$  (Gr. theta) means any one of a number of things:
  - a discrete hypothesis
  - a distinct model (e.g. JC, HKY, GTR, etc.)
  - a tree topology
  - one of an infinite number of continuous model parameter values (e.g. ts:tv rate ratio)
- Prior vs. posterior probability
- Posterior probability can be calculated by multiplying the prior probability of a tree (and model parameters) and the likelihood of the observed data (given a tree and its parameters) divided by a normalizing constant



## Major difference between ML and BI

- In **ML joint estimation of parameters** – likelihood for all parameters estimated at once
  - Likelihood of each parameter depends on likelihood estimation of every other parameter
- In **BI marginal estimation** – posterior probability of any one parameter is calculated independently
- So even if using flat priors and the same model of DNA evolution, **ML and BI could infer different trees** because of differences between joint and marginal likelihood estimation

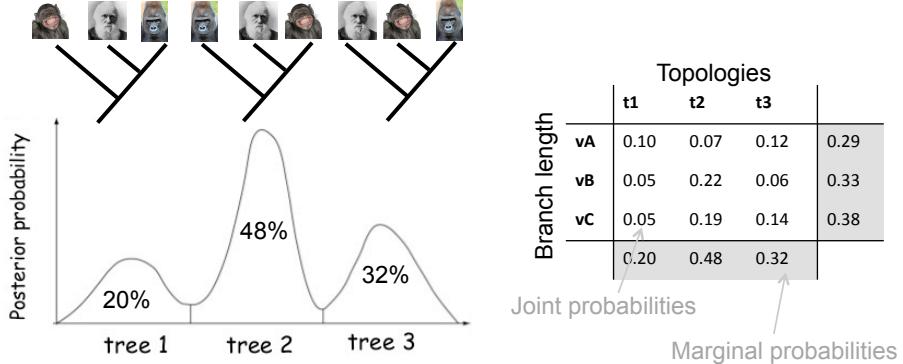
## Bayes' rule: continuous case

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{\int f(D|\theta)f(\theta)d\theta}$$

Diagram illustrating Bayes' rule for the continuous case:

- Likelihood:  $f(D|\theta)$  (blue box)
- Prior probability density:  $f(\theta)$  (orange box)
- Posterior probability density:  $f(\theta|D)$  (purple box)
- Marginal probability of the data:  $\int f(D|\theta)f(\theta)d\theta$  (green box)

Arrows point from Likelihood and Prior probability density to the numerator of the equation. Arrows point from Posterior probability density and Marginal probability of the data to the denominator of the equation.



**Problem:** it is impossible, in most cases,  
to derive the posterior probability analytically

or even estimate it by drawing random samples from it

We want something that will “walk” across this parameter space  
and actively search for the highest point in the parameter  
“landscape”

Slide by Nicolas Chazot (SLU, Uppsala)

## Markov chain Monte Carlo (MCMC)

Larget & Simon 1999

## How does MCMC work? 1/2

A simple summary ☺

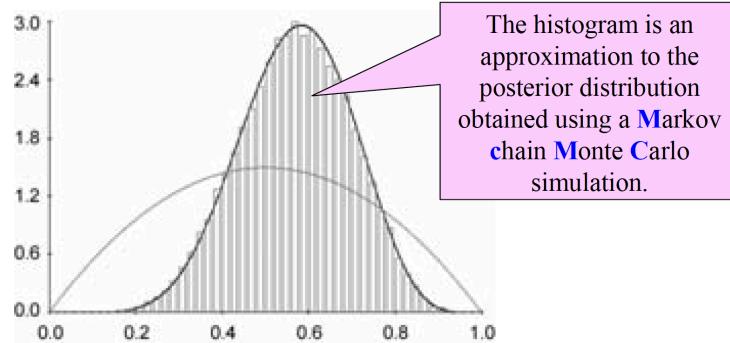
- A Markov chain generates a series of random variables
- The probability distribution of future states depends only on the current state (Markov property)
- Phylogenetic inference starts with a randomly generated tree with branch lengths
- Next step is to generate a new tree, based on the previous tree e.g. using tree rearrangements (NNI, SPR, TBR) or changing branch lengths -> this is a **proposal**
- The **proposal is accepted or rejected** given a probability based on the Metropolis-Hastings algorithm
  - In practice, it's accepted if it has a better likelihood
- If it is accepted, it becomes the new current state and a new proposal is made

## How does MCMC work? 2/2

A simple summary ☺

- Running a Markov chain relatively quickly finds better trees
- After a while no better trees can be found and all sampled trees are close to the optimum – “stationary distribution”
- The number of times the tree is visited by the chain – interpreted as posterior probability of that tree

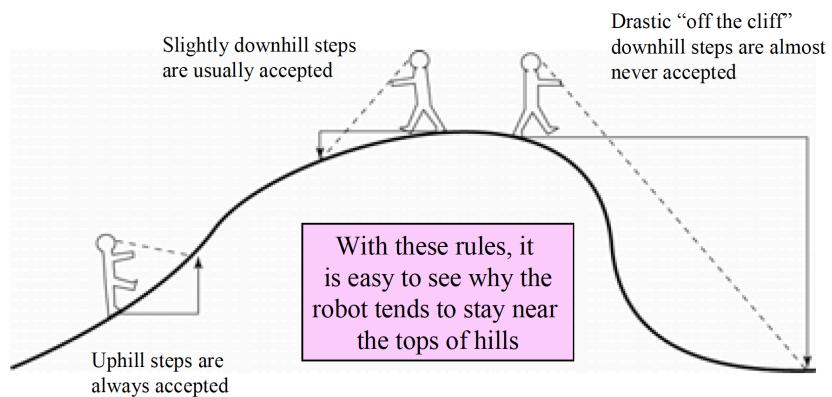
## Markov chain Monte Carlo (MCMC)



For more complex problems, we might settle for a  
**good approximation**  
to the posterior distribution

Copyright © 2005 by Paul O. Lewis

## MCMC robot's rules



Copyright © 2005 by Paul O. Lewis

## How to decide?

$\theta$  = initial position in the parameter space  
 $\theta^*$  = new position proposed randomly

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{\int f(D|\theta)f(\theta)d\theta}$$

- **Ratio of posterior probabilities R:**

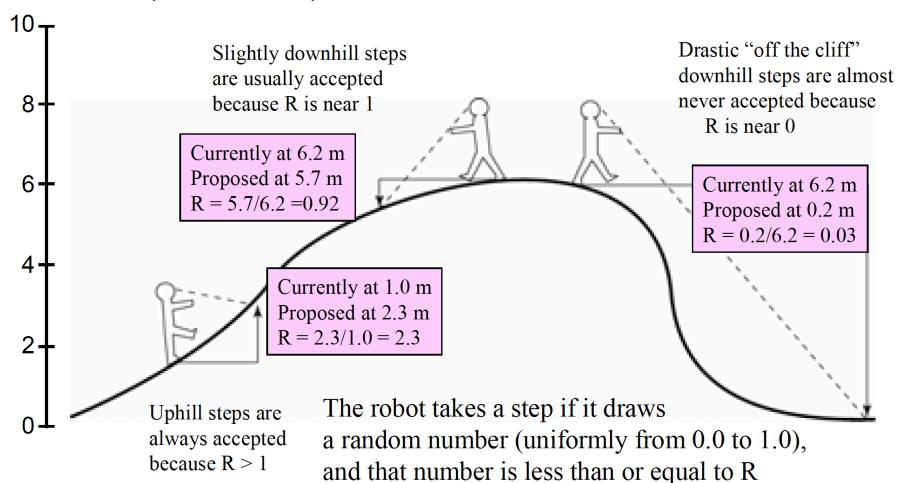
$$\frac{f(\theta^*|D)}{f(\theta|D)} = \frac{\frac{f(D|\theta^*)f(\theta^*)}{f(D)}}{\frac{f(D|\theta)f(\theta)}{f(D)}} = \frac{f(D|\theta^*)f(\theta^*)}{f(D|\theta)f(\theta)} = R$$

- **Random number between [0,1]:**  $n$

**if  $n \leq R$**  => new position accepted

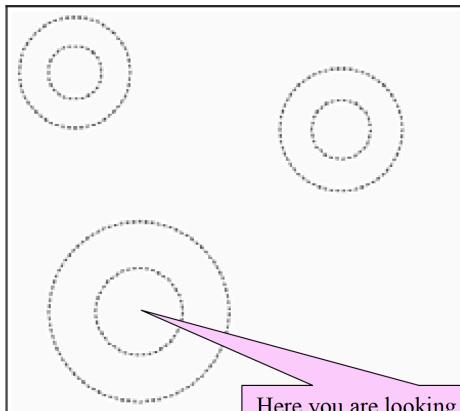
**if  $n > R$**  => new position rejected

## (Actual) MCMC robot rules



Copyright © 2005 by Paul O. Lewis

## What MCRobot can teach us about Markov chain Monte Carlo



### Posterior distribution:

- equal mixture of 3 bivariate normal “hills”
- inner contours: 50%
- outer contours: 95%

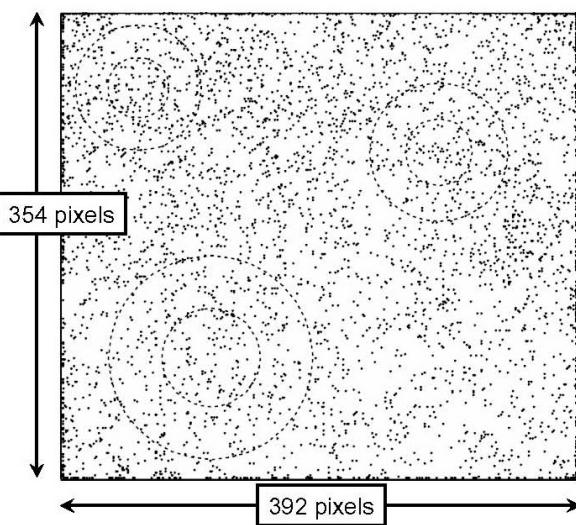
### Proposal scheme:

- random direction
- gamma-distributed step length
- reflection at edges

Here you are looking down from above at  
one of the three bivariate normal hills

Copyright © 2005 by Paul O. Lewis

## Pure random walk



### Proposal scheme:

- random direction
- gamma-distributed step length (mean 45 pixels, s.d. 40 pixels)
- reflection at edges

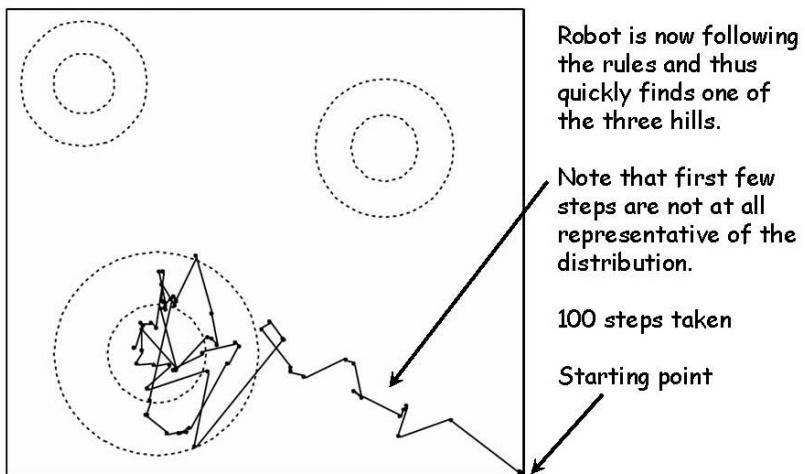
### Target distribution:

- equal mixture of 3 bivariate normal “hills”
- inner contours: 50%
- outer contours: 95%

In this case, the robot  
is accepting every step

5000 steps shown

## Burn-in



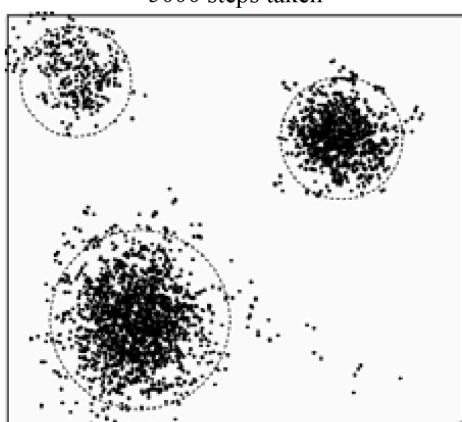
© 2005 Paul O. Lewis

Bayesian Phylogenetics

28

## Target distribution approximation

5000 steps taken



How good is the MCMC approximation?

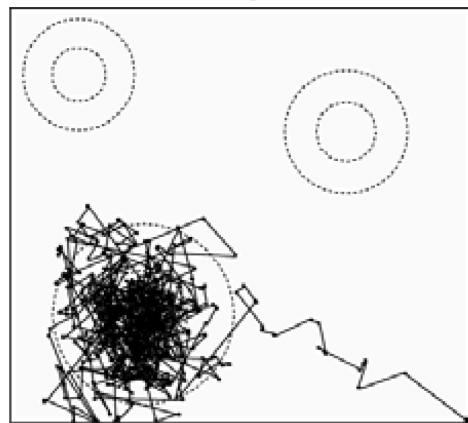
- 51.2% of points are inside inner contours (cf. 50% actual)
- 93.6% of points are inside outer contours (cf. 95% actual)

Approximation gets better the longer the chain is allowed to run.

Copyright © 2005 by Paul O. Lewis

## Just how long is a long run?

1000 steps taken

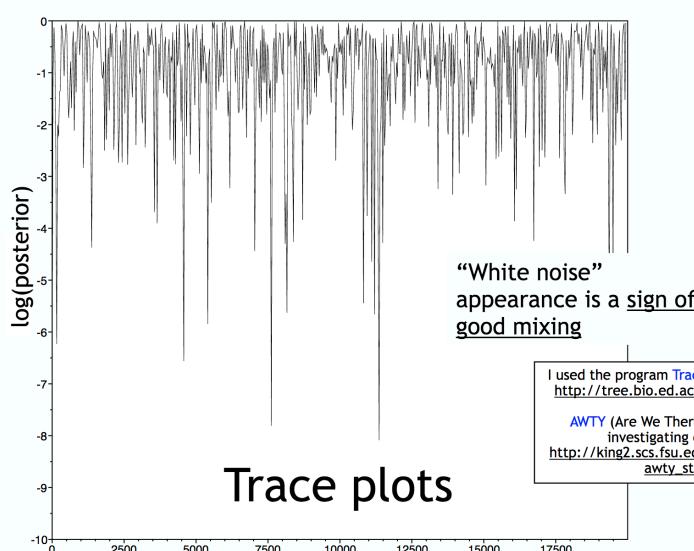


What would you conclude about the target distribution had you stopped the robot at this point?

The way to avoid this mistake is to perform **several runs**, each one beginning from a different randomly-chosen starting point.

Results different among runs? Probably none of them were long enough!

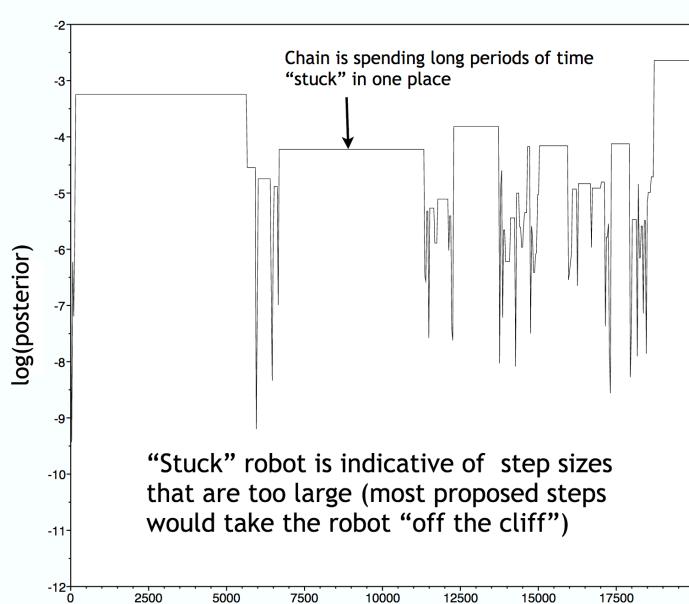
Copyright © 2005 by Paul O. Lewis





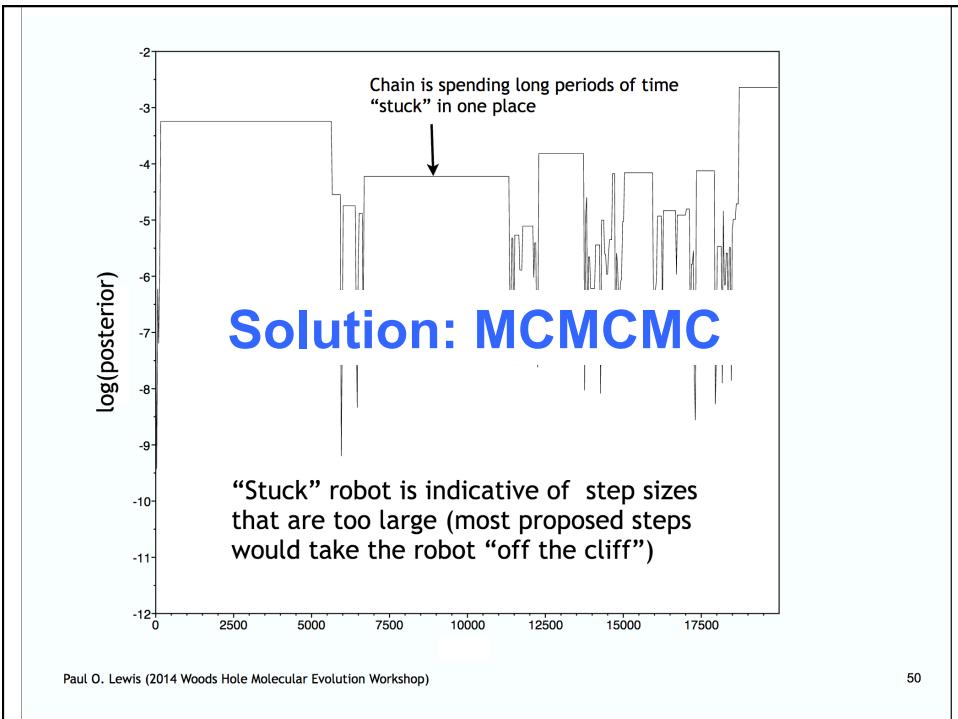
Paul O. Lewis (2014 Woods Hole Molecular Evolution Workshop)

48



Paul O. Lewis (2014 Woods Hole Molecular Evolution Workshop)

50



## Metropolis-coupled Markov chain Monte Carlo (MCMCMC, or MC<sup>3</sup>)



- MC<sup>3</sup> involves running several chains simultaneously (one “cold” and several “heated”)
- The cold chain is the one that counts, the heated chains are “scouts”
- Chain is heated by raising densities to a power less than 1.0 (values closer to 0.0 are warmer)

## Cold vs. heated landscapes



Cold landscape: note peaks separated by deep valleys



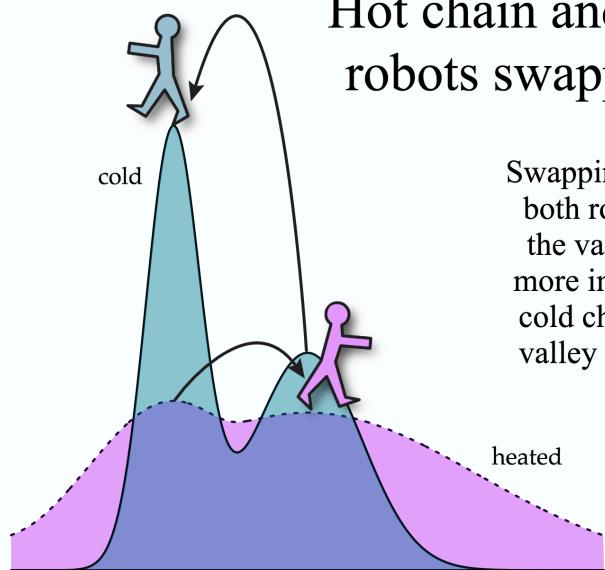
Heated landscape: note shallow (easy to cross) valleys

© 2005 Paul O. Lewis

Bayesian Phylogenetics

36

## Hot chain and cold chain robots swapping places

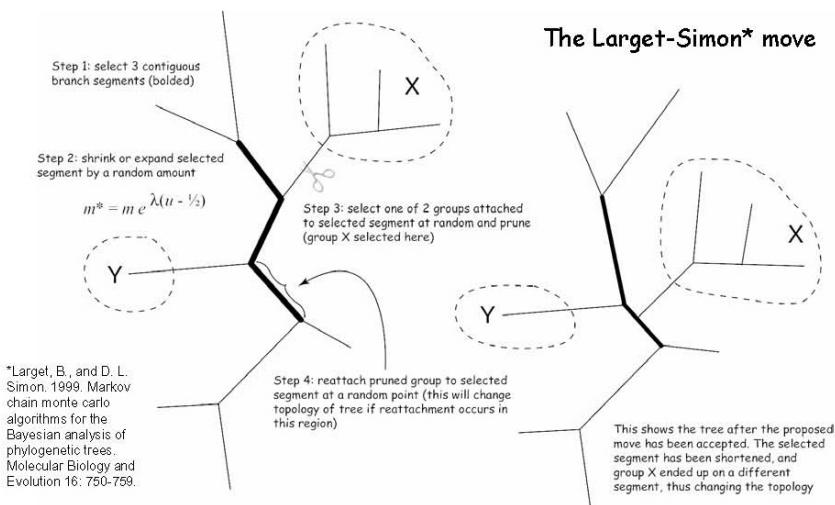


Swapping places means both robots can cross the valley, but this is more important for the cold chain because its valley is much deeper

Paul O. Lewis (2014 Woods Hole Molecular Evolution Workshop)

55

## Moving through treespace



© 2005 Paul O. Lewis

Bayesian Phylogenetics

40

## MCMC, in short:

- Start with **random tree** and arbitrary **initial values for branch lengths and model parameters**
- Each generation consists of one of these (chosen at random):
  - **Propose a new tree** (e.g. Larget-Simon move) and either accept or reject the move
  - **Propose** (and either accept or reject) a new model parameter value
- Every k generations save tree topology, branch lengths and all model parameters (i.e. sample the chain)
- After n generations, summarize sample using histograms, means, credible intervals, etc.

## How does MCMC work? 1/2

A simple summary 😊

- A Markov chain generates a series of random variables
- The probability distribution of future states depends only on the current state (Markov property)
- Phylogenetic inference starts with a randomly generated tree with branch lengths
- Next step is to generate a new tree, based on the previous tree e.g. using tree rearrangements (NNI, SPR, TBR) or changing branch lengths -> this is a **proposal**
- The **proposal is accepted or rejected** given a probability based on the Metropolis-Hastings algorithm
  - In practice, it's accepted if it has a better likelihood
- If it is accepted, it becomes the new current state and a new proposal is made

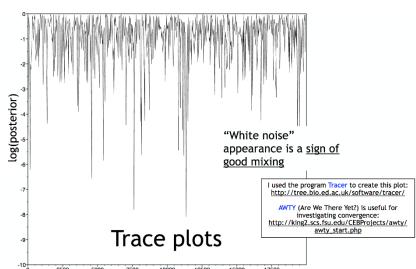
## How does MCMC work? 2/2

A simple summary 😊

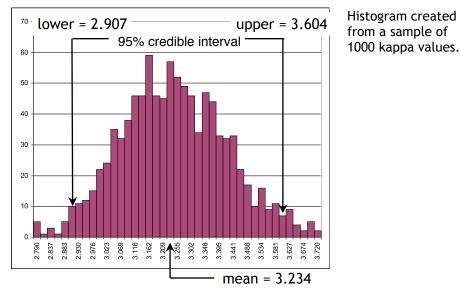
- Running a Markov chain relatively quickly finds better trees
- After a while no better trees can be found and all sampled trees are close to the optimum – “stationary distribution”
- The number of times the tree is visited by the chain – interpreted as posterior probability of that tree

## Looking at the results

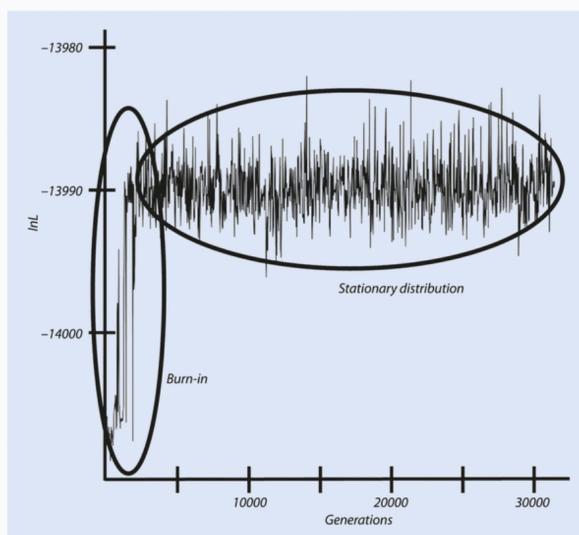
### - Graphically



### Marginal Posterior Distribution of $\kappa$



- **Statistically, by computing Effective Sample Size (ESS, minimum should be 200 or more)**
- **Checking convergence of each run, and all runs together**



**Fig. 8.12** Likelihood scores of a MCMC run plotted against generations. Once stationarity is achieved, trees from this distribution are sampled by discarding all other trees as burn-in. A majority-rule consensus of sampled trees will provide posterior probabilities for every node

## Summarizing the results

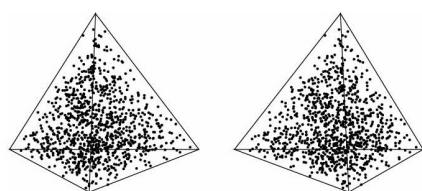
- ▶ **Autocorrelation:** between values that are sampled one after another, so need to sample values at a lower frequency – e.g. every 1000 steps
  - ▶ MCMC easily runs over millions of states
- => **Synthesize/summarize the parameters we are interested in**
- By computing the marginal posterior distribution of these parameters
- mean, median or variance
  - 95% credibility interval
- By identifying one or more “best” topologies
- e.g. the splits most frequently identified
- The number of times a clade in the tree is accepted during the MCMC defines the posterior probability of the clade, and therefore indicates the support for the node

## Prior distributions

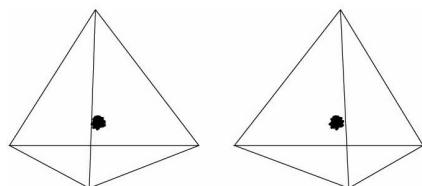
## Prior Distributions

- For topologies: discrete Uniform distribution
- For proportions: Beta( $a,b$ ) distribution
  - flat when  $a=b$
  - peaked above 0.5 if  $a=b$  and both are greater than 1
- For base frequencies: Dirichlet( $a,b,c,d$ ) distribution
  - flat when  $a=b=c=d$
  - all base frequencies close to 0.25 if  $v=a=b=c=d$  and  $v$  large (e.g. 300)
- For GTR model relative rates: Dirichlet( $a,b,c,d,e,f$ ) distribution

### 4-parameter Dirichlet( $a,b,c,d$ )



Flat prior:  
 $a = b = c = d = 1$



Informative prior:  
 $a = b = c = d = 300$

(stereo pairs)

## Prior Distributions

- For other model parameters and branch lengths: **Gamma(a,b) distribution**
  - Exponential( $\lambda$ ) equals Gamma(1,  $\lambda$ -1) distribution
  - Mean of Gamma(a,b) is ab (so mean of an Exponential(10) distribution is 0.1)
  - Variance of a Gamma(a,b) distribution is ab<sup>2</sup> (so variance of an Exponential(10) distribution is 0.01)

**10 important considerations**

## Top 10 List (of important considerations)

1. Beware of arbitrarily truncated priors
2. Branch length priors particularly important
3. Beware of high posteriors for very short branch lengths
4. Partition with care (prefer fewer subsets) and run MCMC for longer
5. MCMC run length should depend on number of parameters
6. Pay attention to parameter estimates
7. Pay attention to the behavior of the MCMC for ALL parameters
8. Run without data to explore prior
9. Run long
10. Run several times and compare runs

## 3. Branch length priors

*Syst. Biol.* 59(1):108–117, 2010  
© The Author(s) 2009. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.  
For Permissions, please email: journals.permissions@oxfordjournals.org  
DOI:10.1093/sysbio/syp080  
Advance Access publication on November 17, 2009

### Cryptic Failure of Partitioned Bayesian Phylogenetic Analyses: Lost in the Land of Long Trees

DAVID C. MARSHALL\*

Department of Ecology and Evolutionary Biology, University of Connecticut, 75 N. Eagleville Road, U-3043, Storrs, CT 06269, USA;  
\*Correspondence to be sent to: Department of Ecology and Evolutionary Biology, University of Connecticut, 75 N. Eagleville Road, U-3043,  
Storrs, CT 06269, USA; E-mail: david.marshall@uconn.edu.

*Syst. Biol.* 59(2):145–161, 2010  
© The Author(s) 2009. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.  
For Permissions, please email: journals.permissions@oxfordjournals.org  
DOI:10.1093/sysbio/syp081  
Advance Access publication on December 10, 2009

### When Trees Grow Too Long: Investigating the Causes of Highly Inaccurate Bayesian Branch-Length Estimates

JEREMY M. BROWN<sup>1,2,\*</sup>, SHANNON M. HEDTKE<sup>1</sup>, ALAN R. LEMMON<sup>3,4</sup>, AND EMILY MORIARTY LEMMON<sup>3,5</sup>

## Recommended reading on priors in Bayesian Inference

### Priors and Posteriors in Bayesian Timing of Divergence Analyses: The Age of Butterflies Revisited

Nicolas Chazot , Niklas Wahlberg, André Victor Lucci Freitas, Charles Mitter, Conrad Labandeira, Jae-Cheon Sohn, Ranjit Kumar Sahoo, Noemy Seraphim, Rienk de Jong, Maria Heikkilä

*Systematic Biology*, Volume 68, Issue 5, September 2019, Pages 797–813,

<https://doi.org/10.1093/sysbio/syz002>

Published: 25 February 2019 Article history ▾

- 4: Partition with care (prefer fewer subsets)
- 5: MCMC run length should depend on number of parameters

#### 100 taxa, 2000 sites, F81 model

Unpartitioned		197 branch lengths, 3 base frequencies = <u>200 parameters</u> 10 sites/parameter
subset 1	subset 2	394 branch lengths, 6 base frequencies = <u>400 parameters</u> 5 sites/parameter
subset 1	subset 2	591 branch lengths, 9 base frequencies = <u>600 parameters</u> 3.3 sites/parameter

- Partitioning reduces information for estimating some model parameters.
- Might want to run 3-subset case 3 times longer than the unpartitioned case.

## **Some Bayesian programs**

- **MrBayes**
- **ExaBayes**
- **BEAST**

## **Maximum Likelihood and Bayesian methods: summary**

- Both methods are very popular in molecular systematics
- Maximum likelihood is the most important method in phylogenomics
- Bayesian methods are able to take into account uncertainty in parameter estimates
- Bayesian methods can relax the assumption of a homogenous Markov model for rates of change in a tree