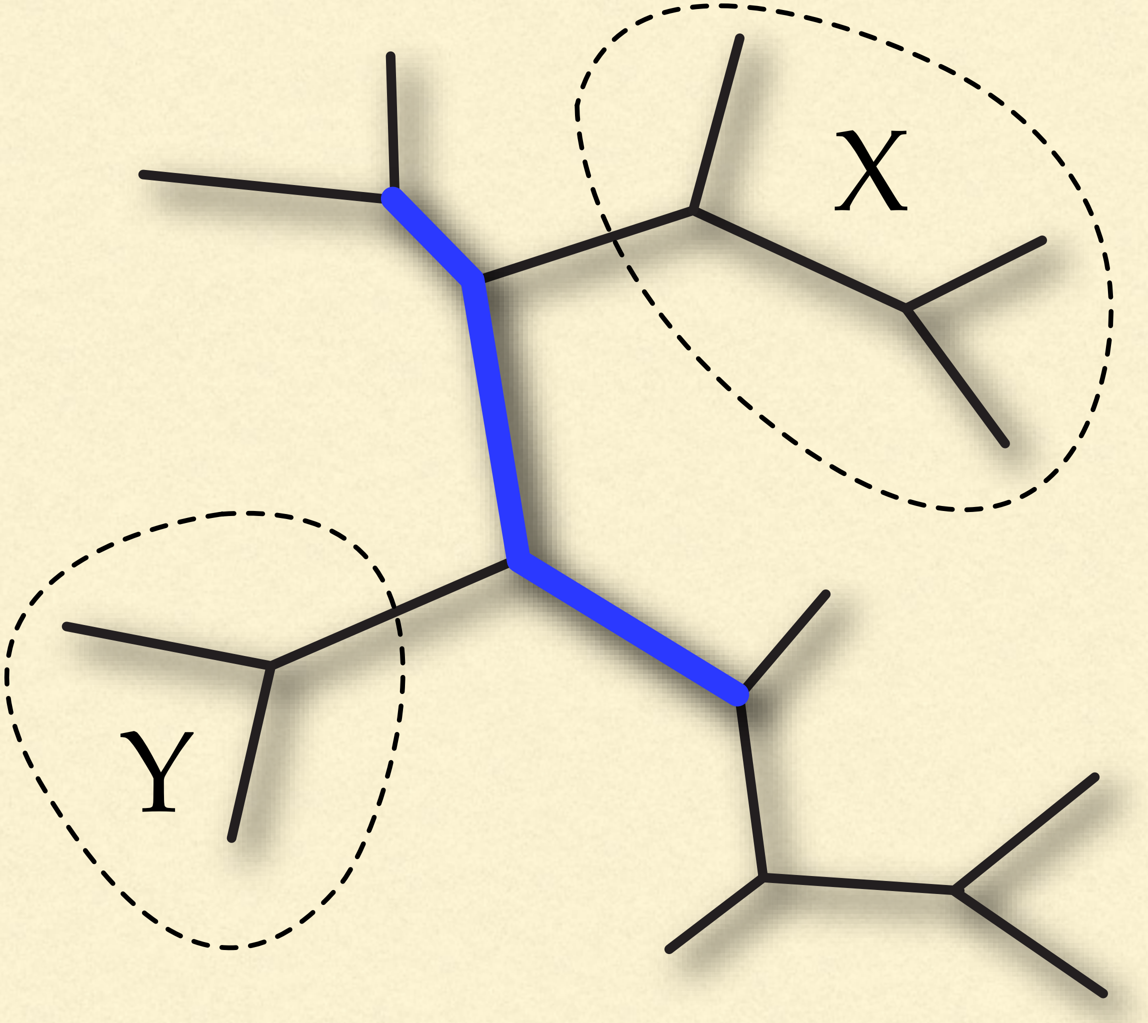
MCMC proposals

Moving through treespace

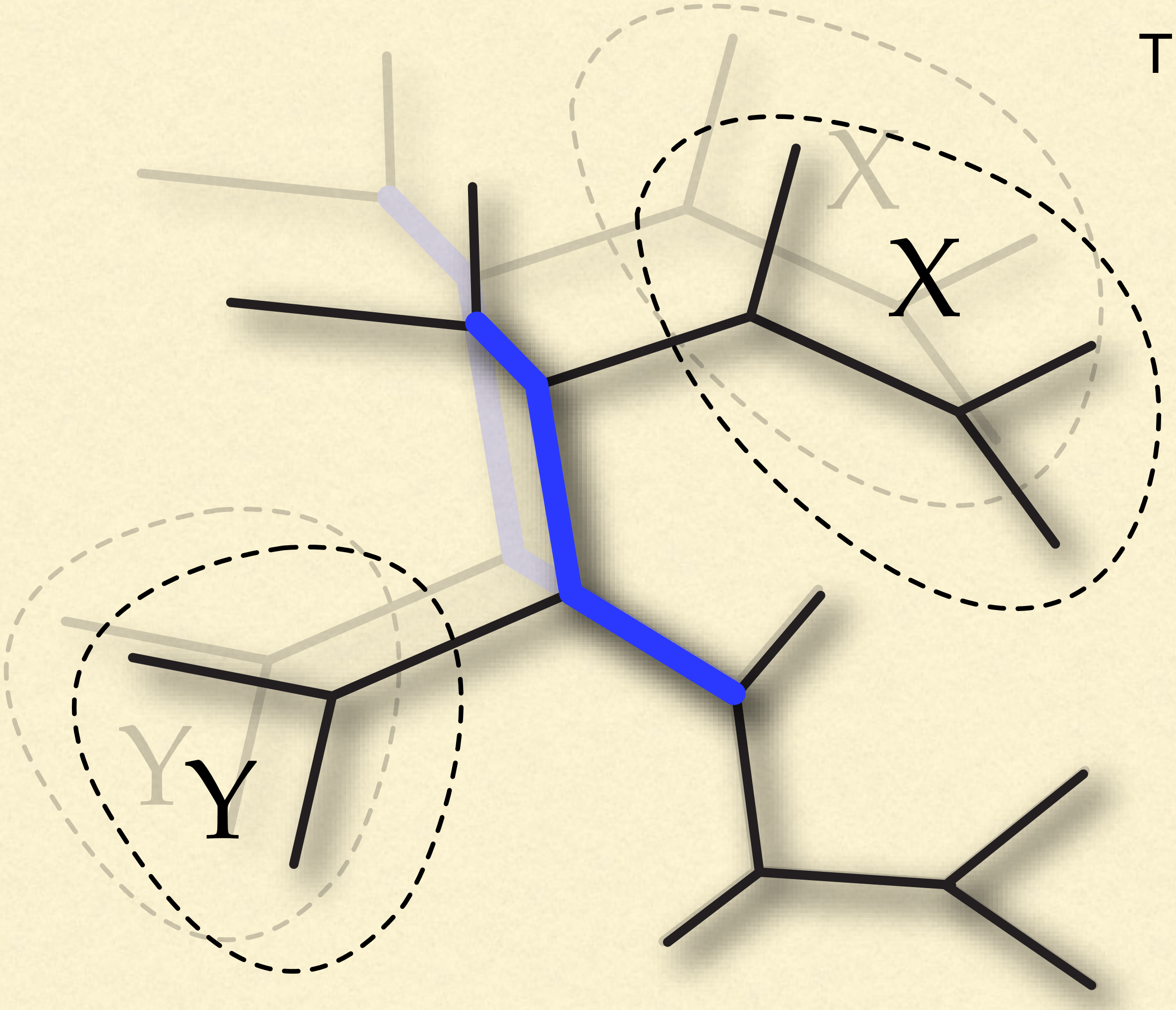


The Larget-Simon move

Step 1:
Pick 3 contiguous edges randomly, defining two subtrees, X and Y

Larget, B., and D. L. Simon. 1999. Mol. Biol. Evol. 16: 750-759.
See also: Holder et al. 2005. Syst. Biol. 54: 961-965.

Moving through treespace

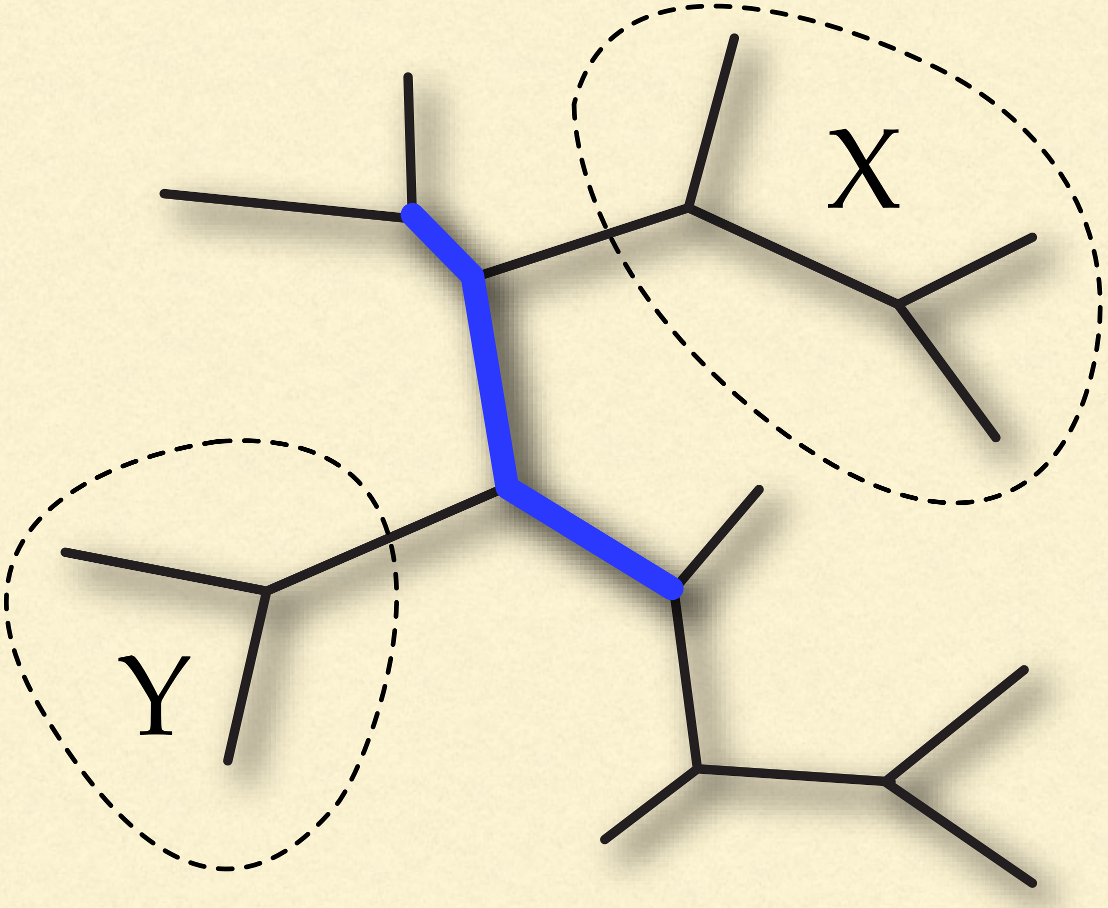


The Larget-Simon move

Step 1:
Pick 3 contiguous edges randomly, defining two subtrees, X and Y

Step 2:
Shrink or grow selected 3-edge segment by a random amount

Moving through treespace

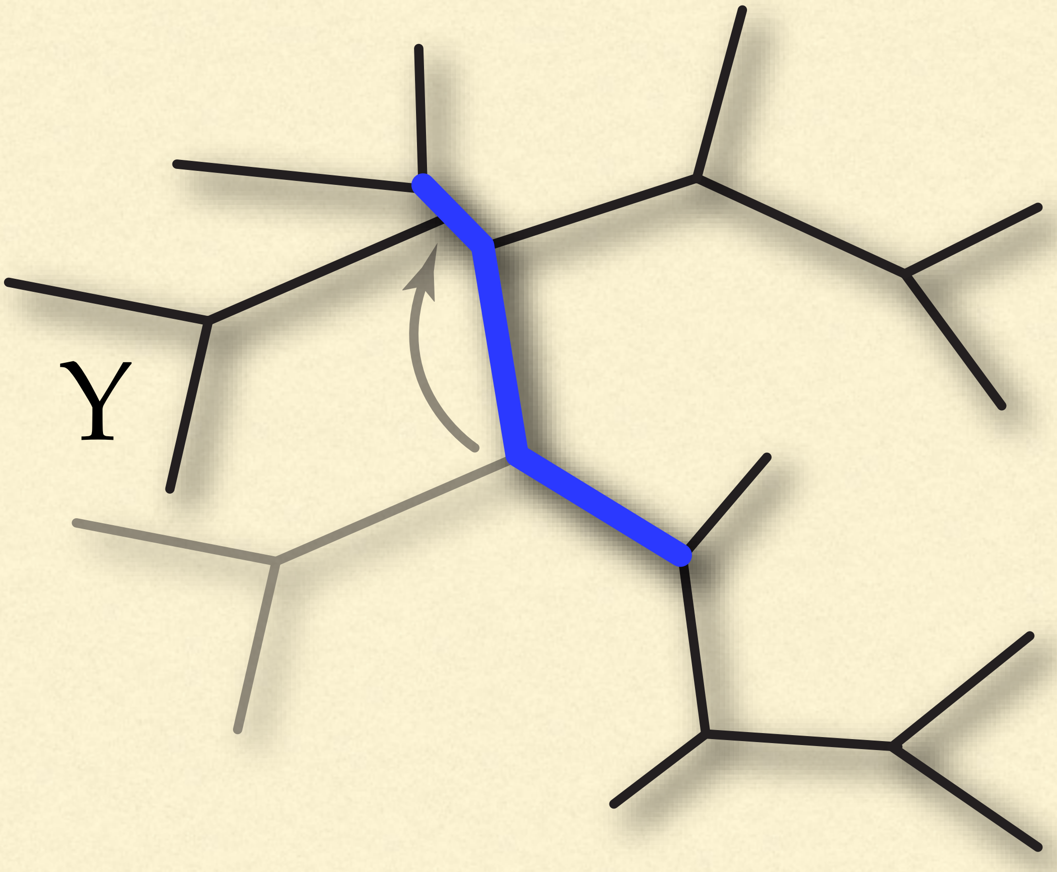


The Large-Simon move

Step 1:
Pick 3 contiguous edges randomly, defining two subtrees, X and Y

Step 2:
Shrink or grow selected 3-edge segment by a random amount

Moving through treespace



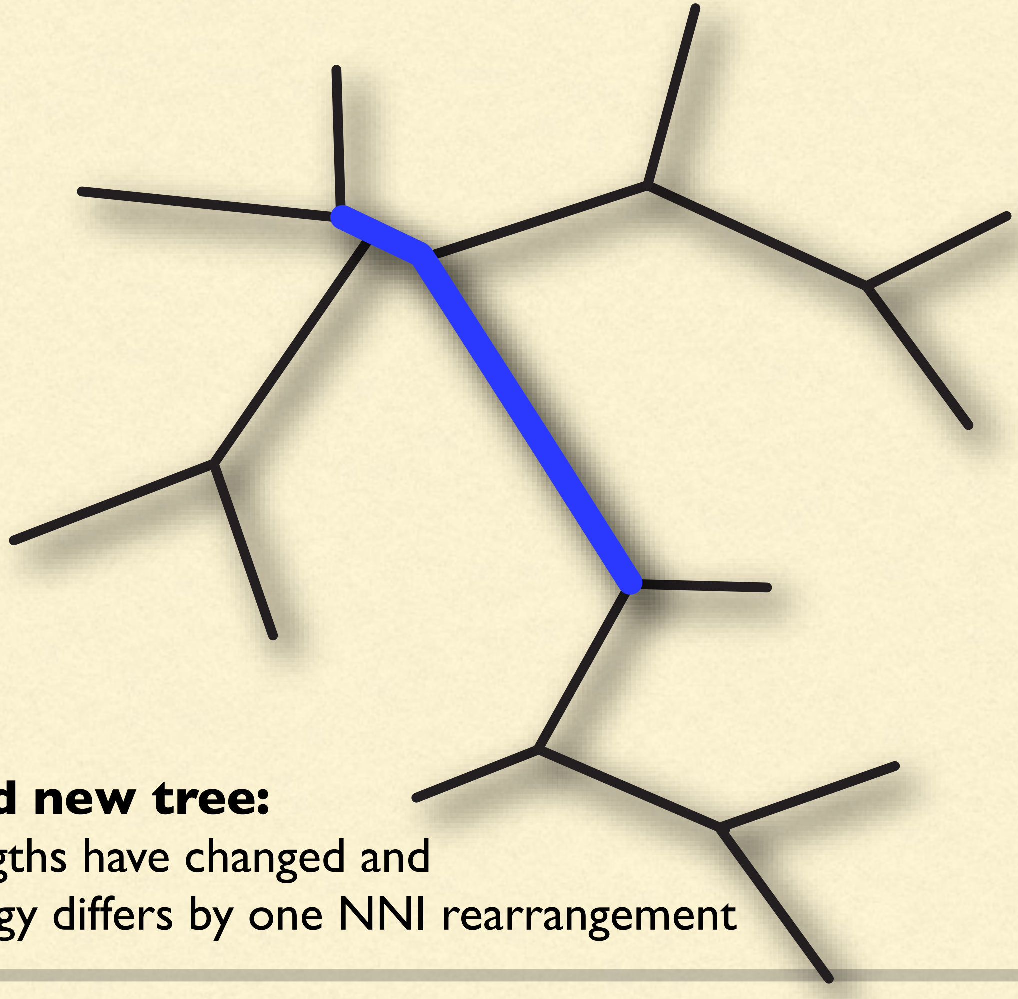
The Larget-Simon move

Step 1:
Pick 3 contiguous edges randomly, defining two subtrees, X and Y

Step 2:
Shrink or grow selected 3-edge segment by a random amount

Step 3:
Choose X or Y randomly, then reposition randomly

Moving through treespace



The Larget-Simon move

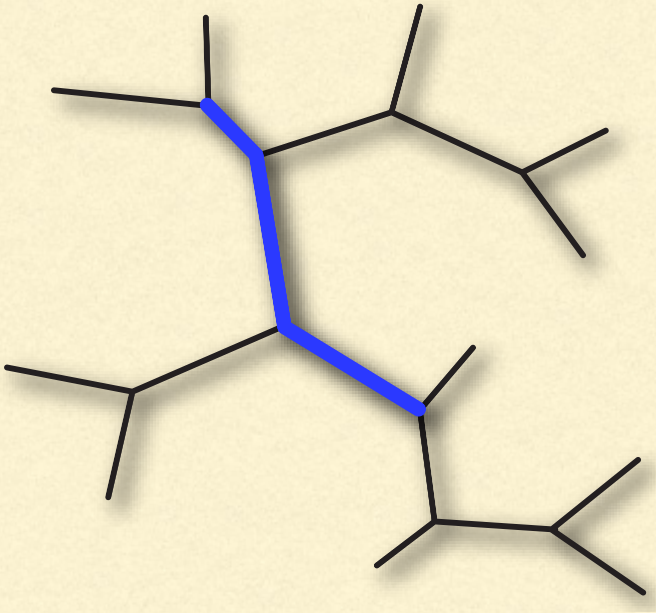
Step 1:
Pick 3 contiguous edges randomly, defining two subtrees, X and Y

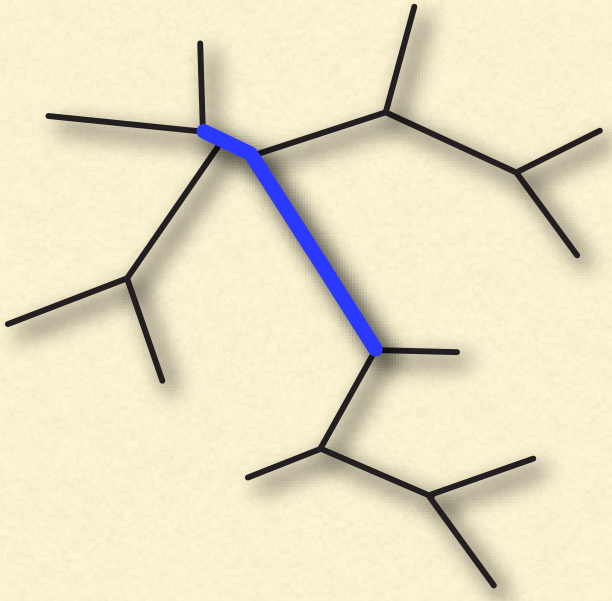
Step 2:
Shrink or grow selected 3-edge segment by a random amount

Step 3:
Choose X or Y randomly, then reposition randomly

Proposed new tree:
3 edge lengths have changed and
the topology differs by one NNI rearrangement

Moving through treespace





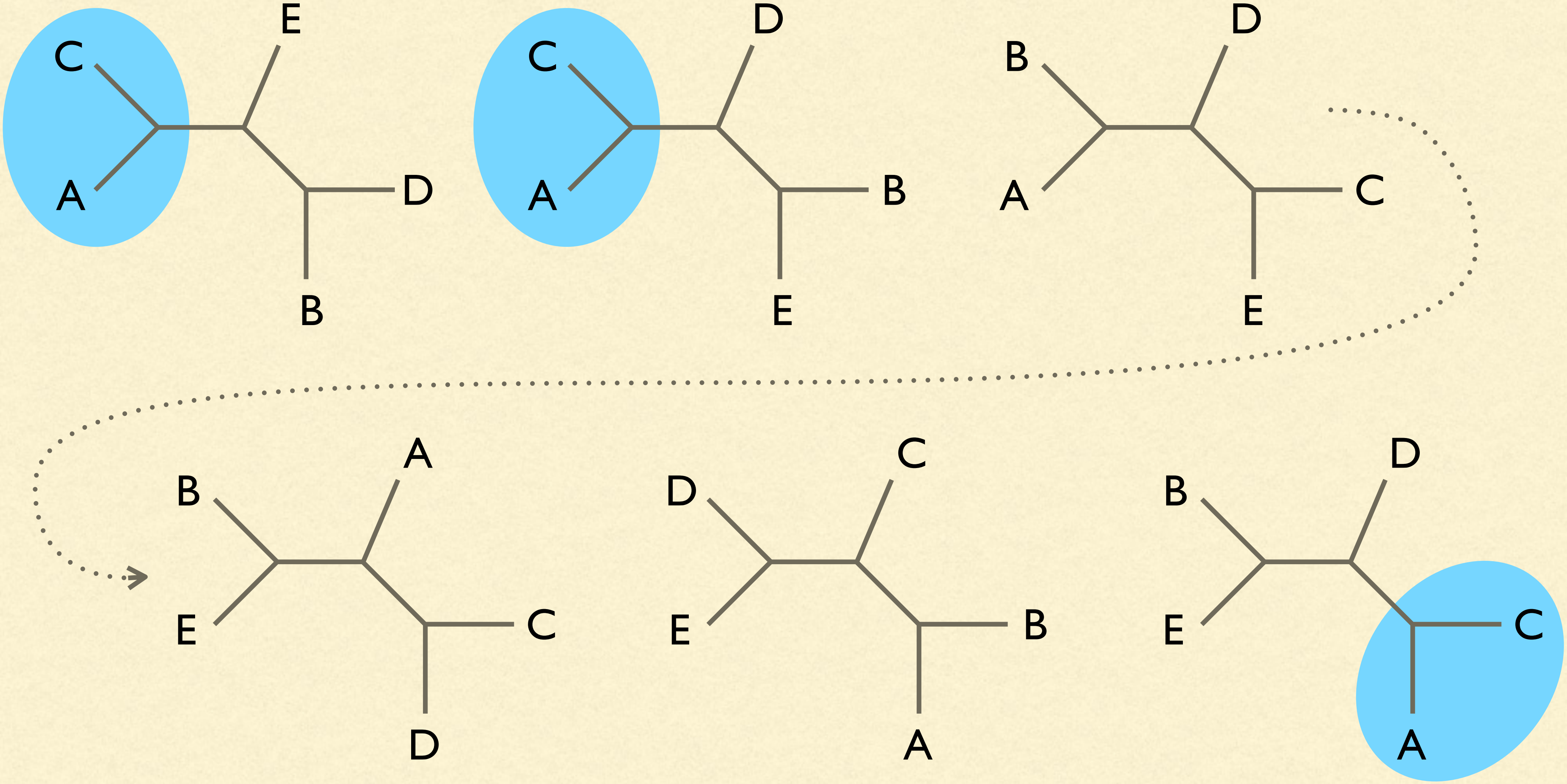
Current tree

log-posterior = -34256

Proposed tree

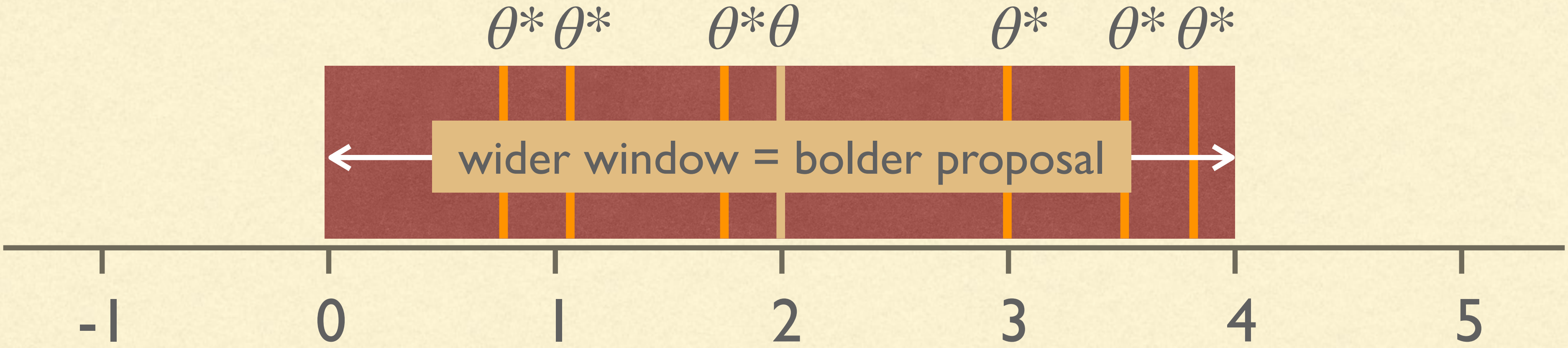
log-posterior = -32519
(better, so accept)

Marginal split posterior probabilities



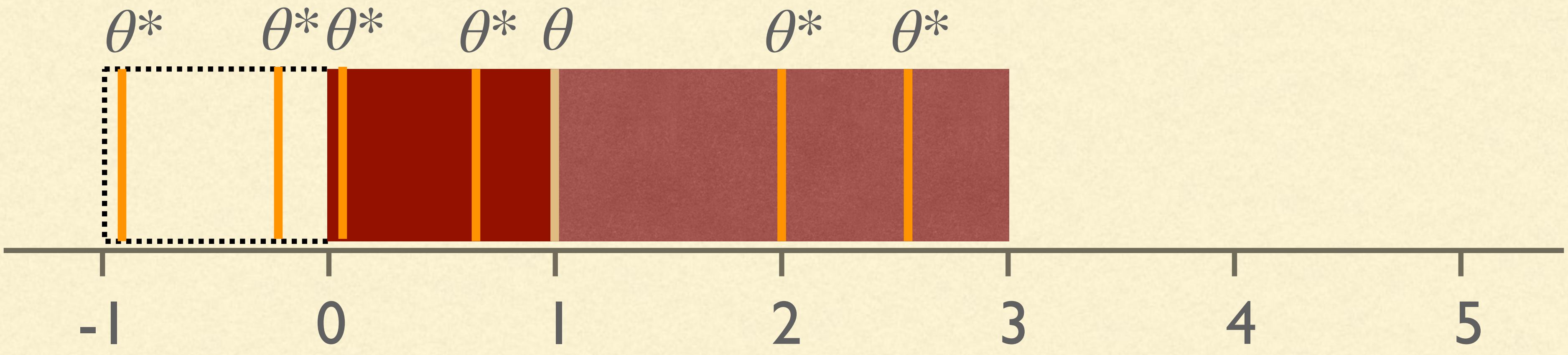
Sliding window proposal

propose new values (θ^*) uniformly
within a proposal window
centered at current value (θ)

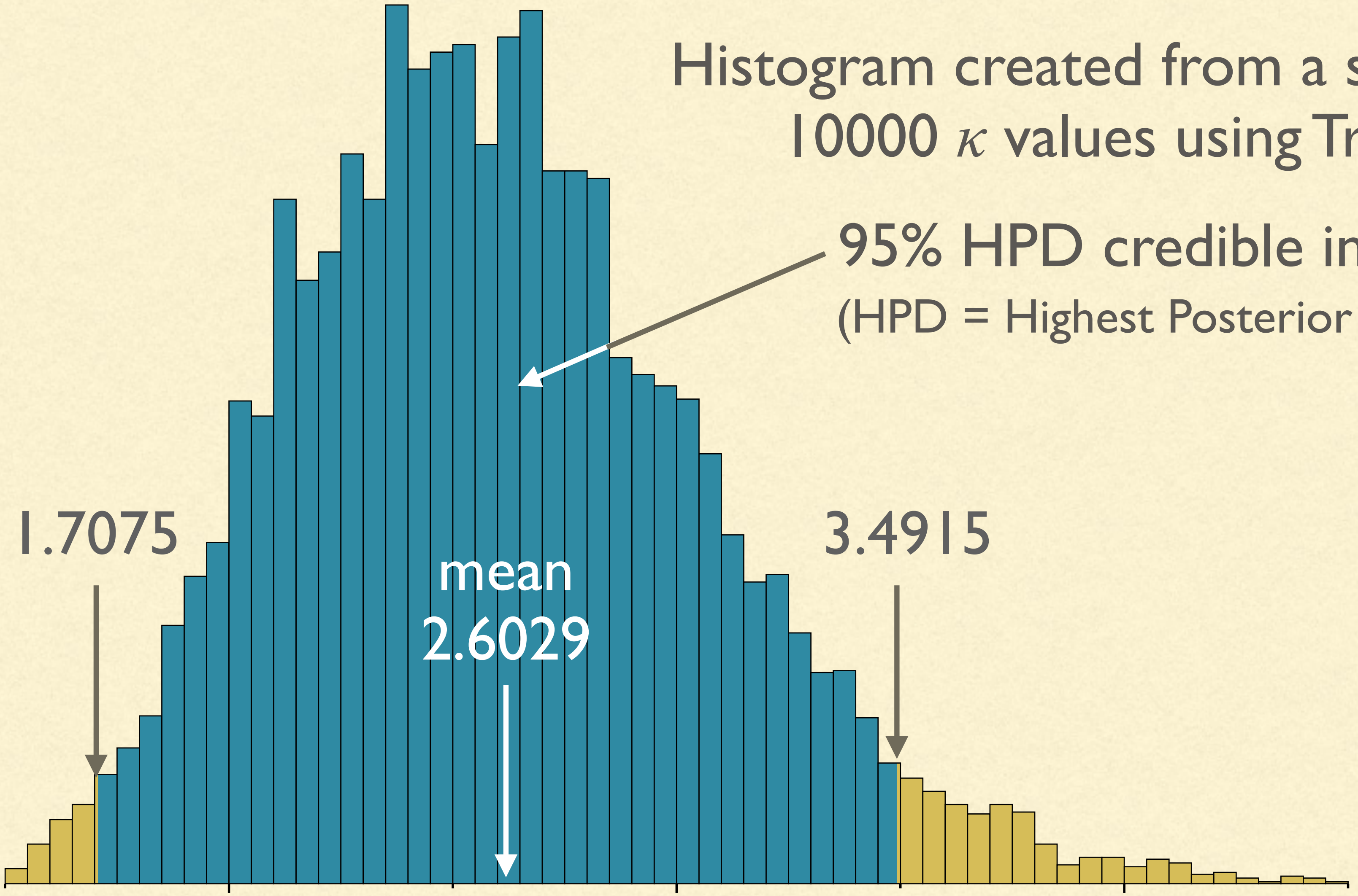


Sliding window proposal

If proposed value is out of bounds,
reflect it back in bounds



Marginal distributions and credible intervals



Histogram created from a sample of 10000 κ values using Tracer

95% HPD credible interval

(HPD = Highest Posterior Density)

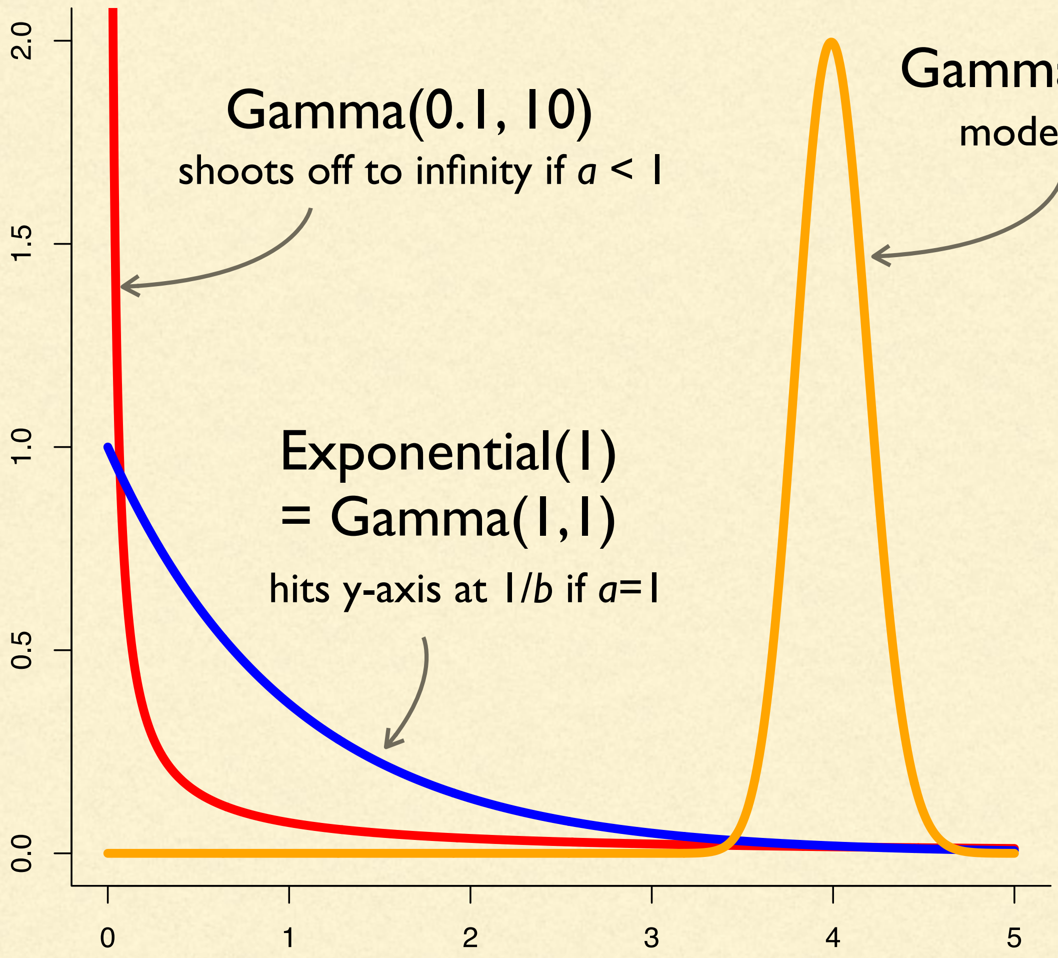
MOLE 2025 Paul O. Lewis

Tracer downloaded from https://github.com/beast-dev/tracer/releases/tag/v1.7.1

Prior distributions

$$p(\theta \mid D) = \frac{p(D \mid \theta)\, \boxed{p(\theta)}}{p(D)}$$

Gamma(a,b) distribution



Gamma(0.1, 10)
shoots off to infinity if $a < 1$

Exponential(1)
= Gamma(1,1)
hits y-axis at $1/b$ if $a=1$

Gamma(400, 0.01)
mode > 0 if $a>1$

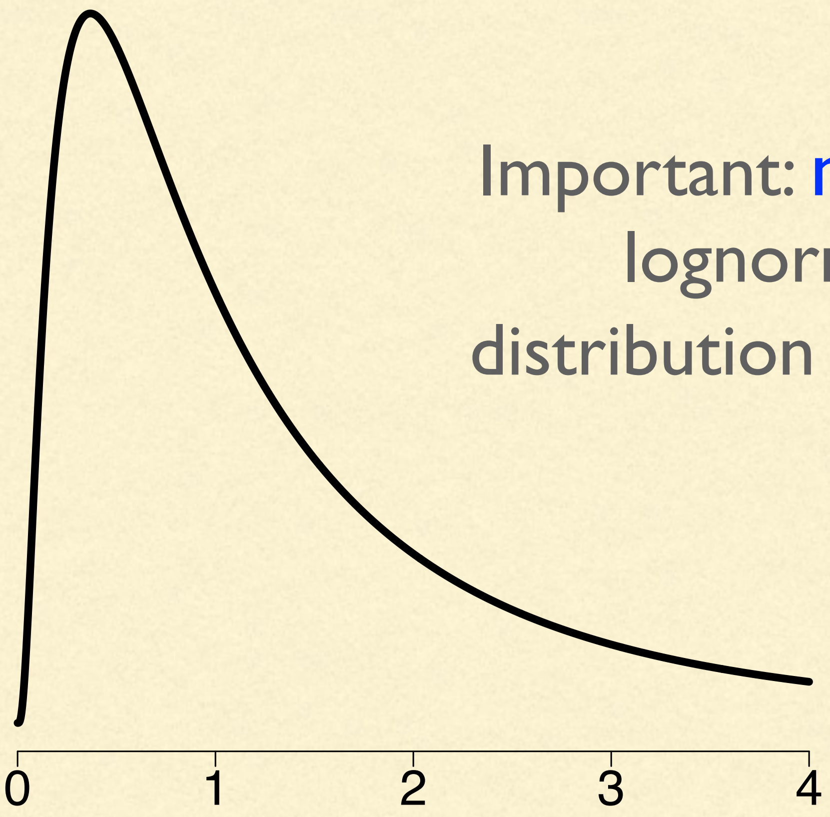
Gamma(a,b) distributions are appropriate for parameters that range from 0 to infinity (e.g. branch lengths)

a = shape
b = scale*
mean = ab
variance = ab^2

*Note: be aware that in many papers the Gamma distribution is defined such that the second parameter is the rate (*inverse* of the scale b used in this slide) rather than the scale! In this case, the mean and variance would be a/b and a/b^2, respectively.

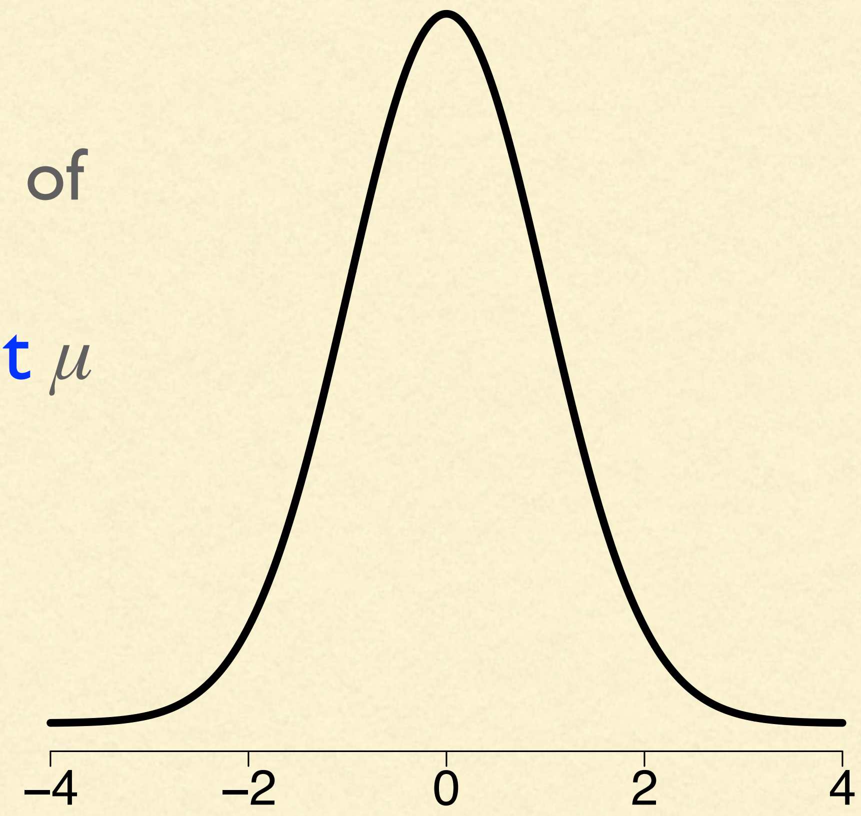
Lognormal(μ,σ) distribution

Lognormal($\mu = 0, \sigma = 1$)

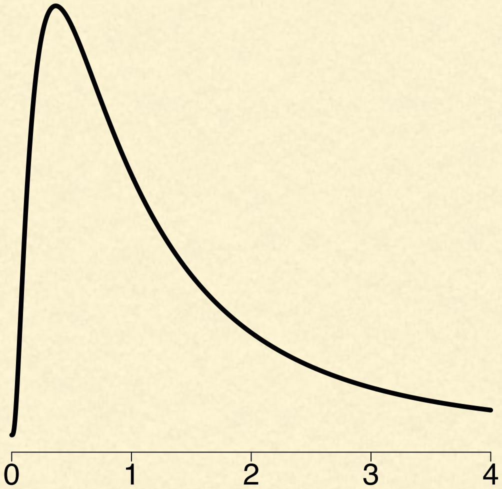


Important: **mean** of lognormal distribution is **not** μ

Normal($\mu = 0, \sigma = 1$)



Lognormal(μ,σ) distribution



$\mu = 0, \sigma = 1$

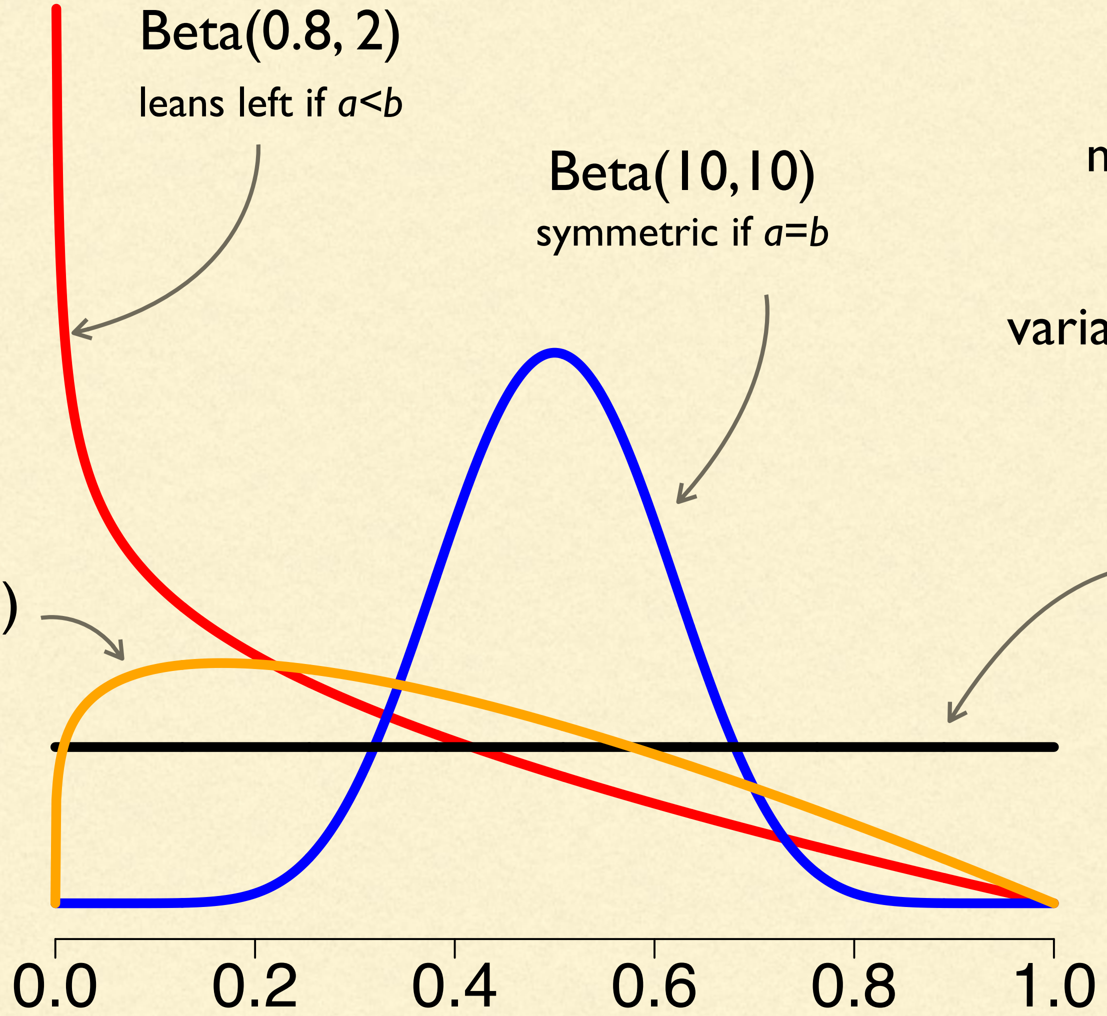
1.65	$\text{mean} = e^{\mu + \sigma^2/2}$
4.67	$\text{variance} = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$
0.37	$\text{mode} = e^{\mu - \sigma^2}$
1.0	$\text{median} = e^{\mu}$

To choose μ and σ to yield a particular mean (*m*) and variance (*v*) for a lognormal prior, use these formulas (log is natural logarithm):

$$\sigma^2 = \log\left(1 + \frac{v}{m^2}\right) \qquad \mu = \log(m) - \sigma^2/2$$

Beta(*a*,*b*) distribution

Beta(*a*,*b*) distributions are appropriate for proportions, which must lie between 0 and 1 (inclusive).



Beta(0.8, 2)
leans left if $a<b$

Beta(10,10)
symmetric if $a=b$

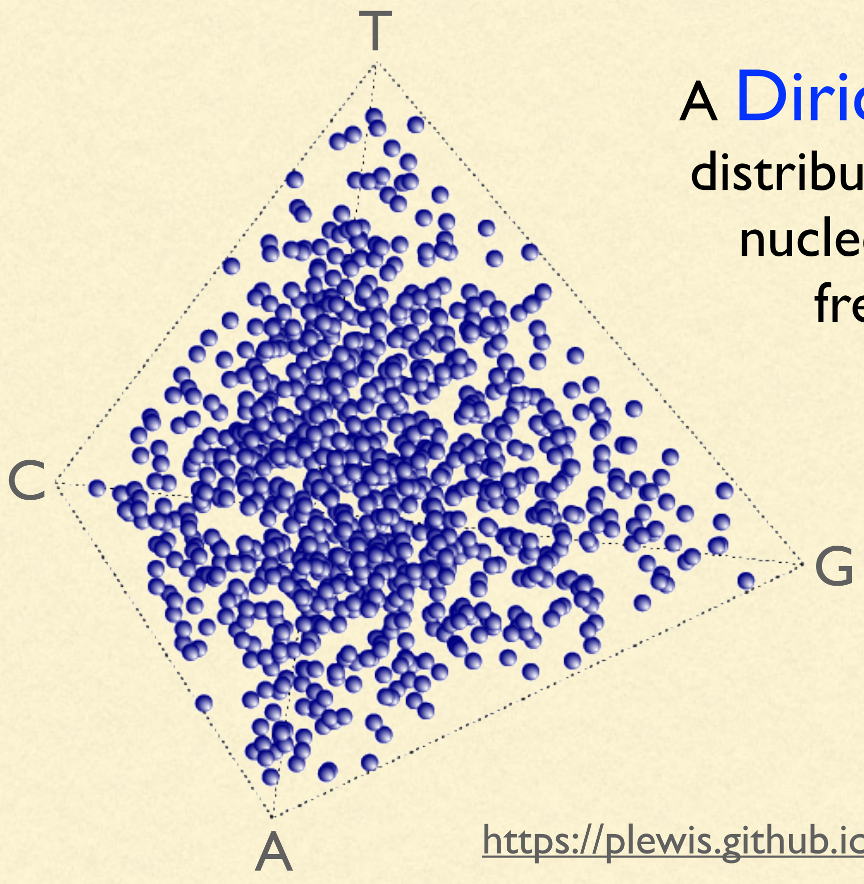
Beta(1.2,2)

Beta(1,1)
Uniform(0,1) if $a=b=1$

$$\text{mean} = \frac{a}{a+b}$$

$$\text{variance} = \frac{ab}{(a+b)^2(a+b+1)}$$

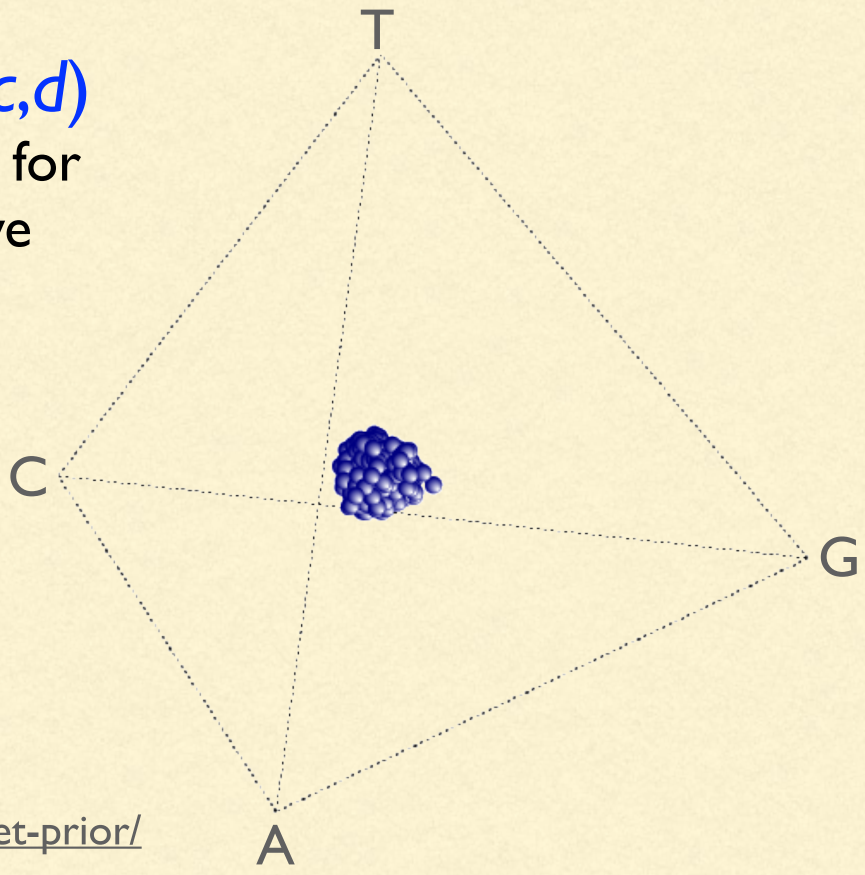
Dirichlet(*a*,*b*,*c*,*d*) distribution



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

(every combination equally probable)

A Dirichlet(*a*,*b*,*c*,*d*) distribution is ideal for nucleotide relative frequencies.

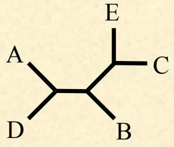


Informative: $a = b = c = d = 100$

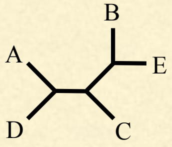
(frequencies tend to be nearly equal)

https://plewis.github.io/applets/dirichlet-prior/

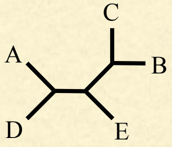
Topology: discrete uniform



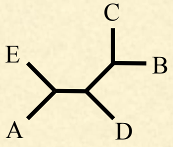
$\frac{1}{15}$



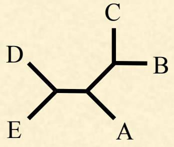
$\frac{1}{15}$



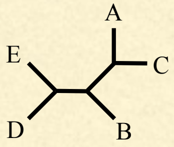
$\frac{1}{15}$



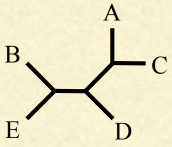
$\frac{1}{15}$



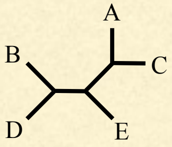
$\frac{1}{15}$



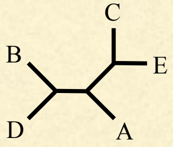
$\frac{1}{15}$



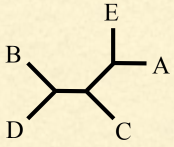
$\frac{1}{15}$



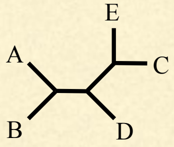
$\frac{1}{15}$



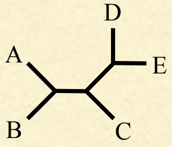
$\frac{1}{15}$



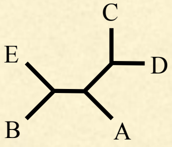
$\frac{1}{15}$



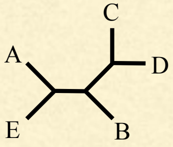
$\frac{1}{15}$



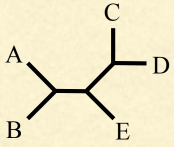
$\frac{1}{15}$



$\frac{1}{15}$

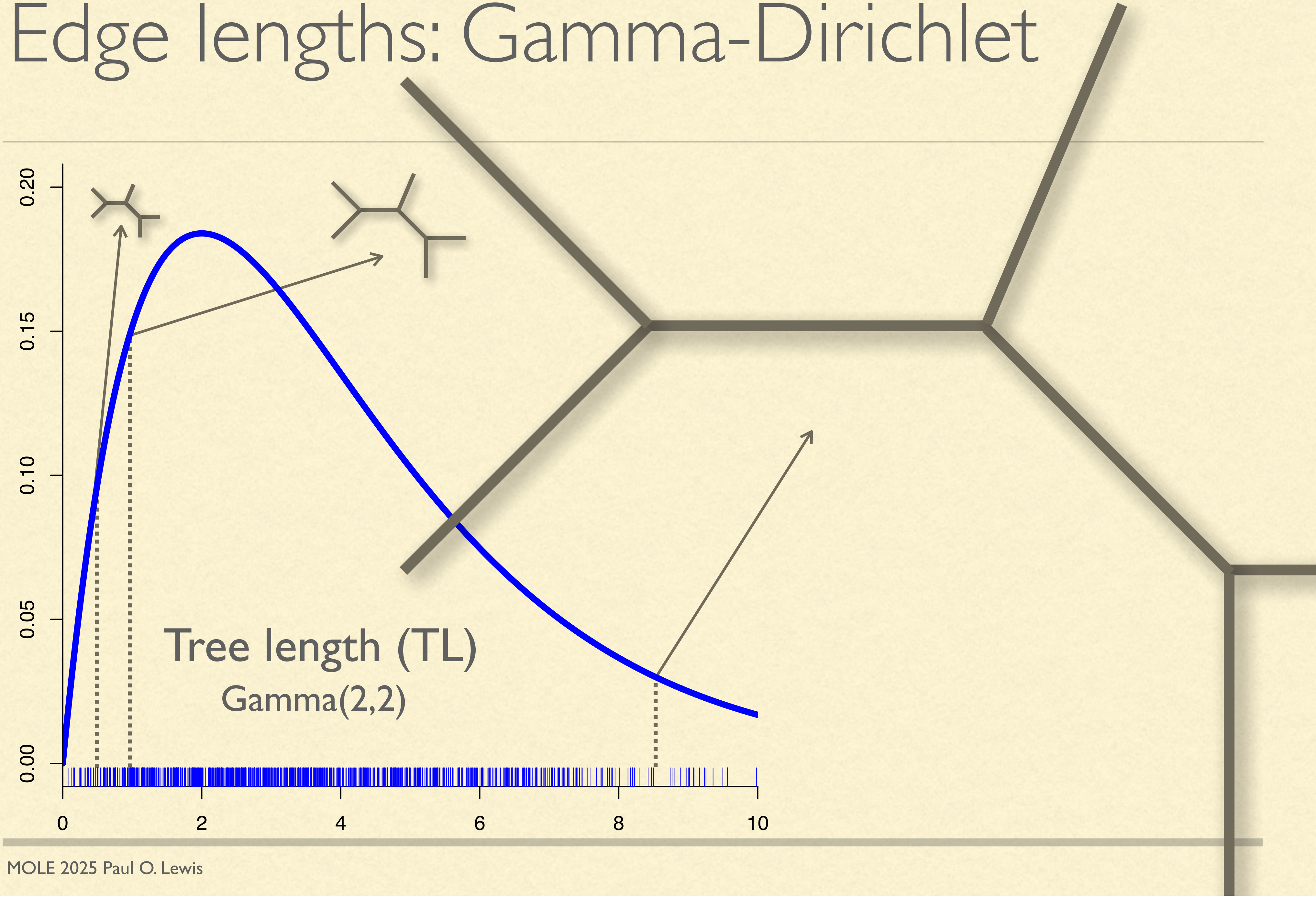


$\frac{1}{15}$



$\frac{1}{15}$

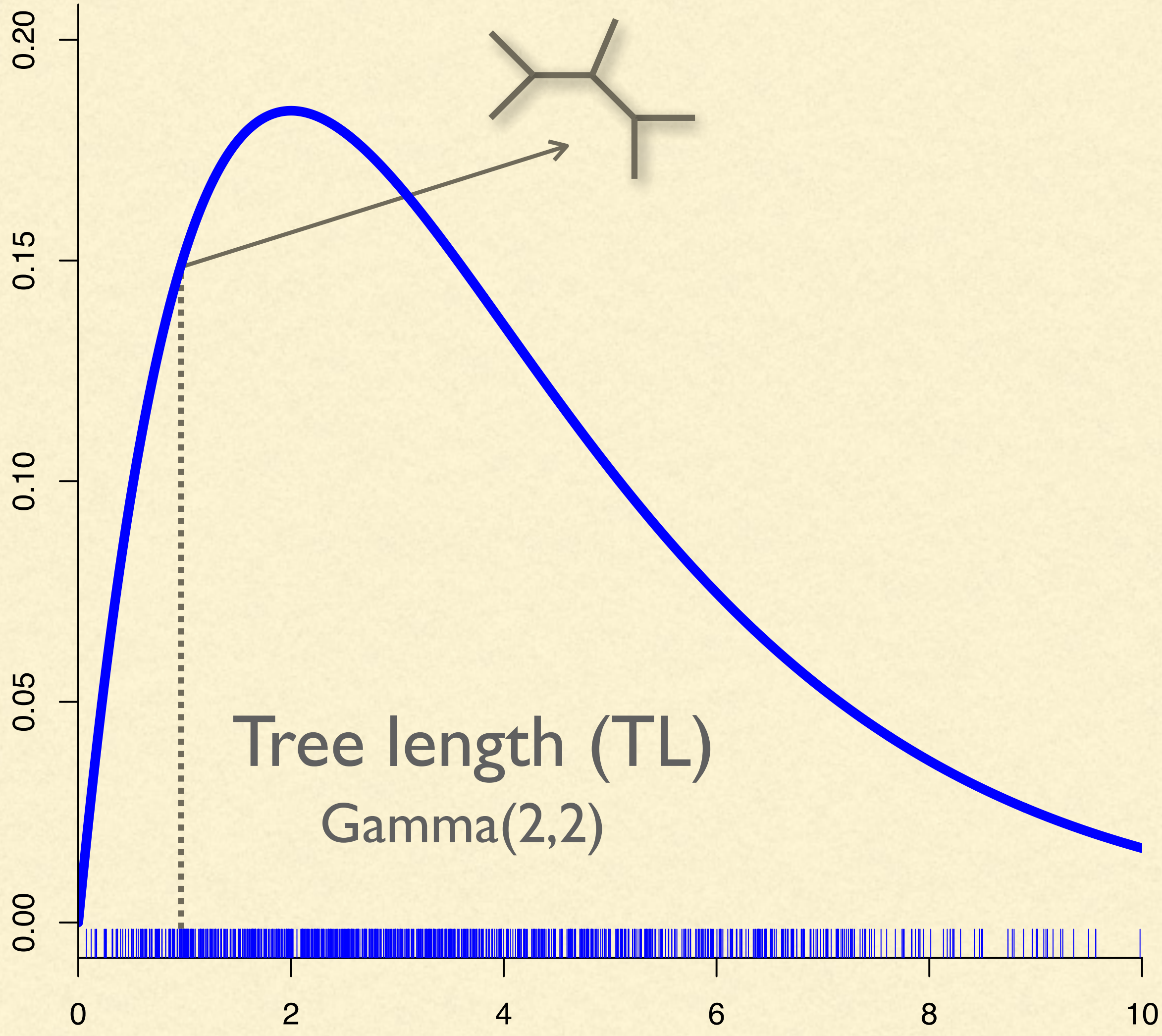
Edge lengths: Gamma-Dirichlet

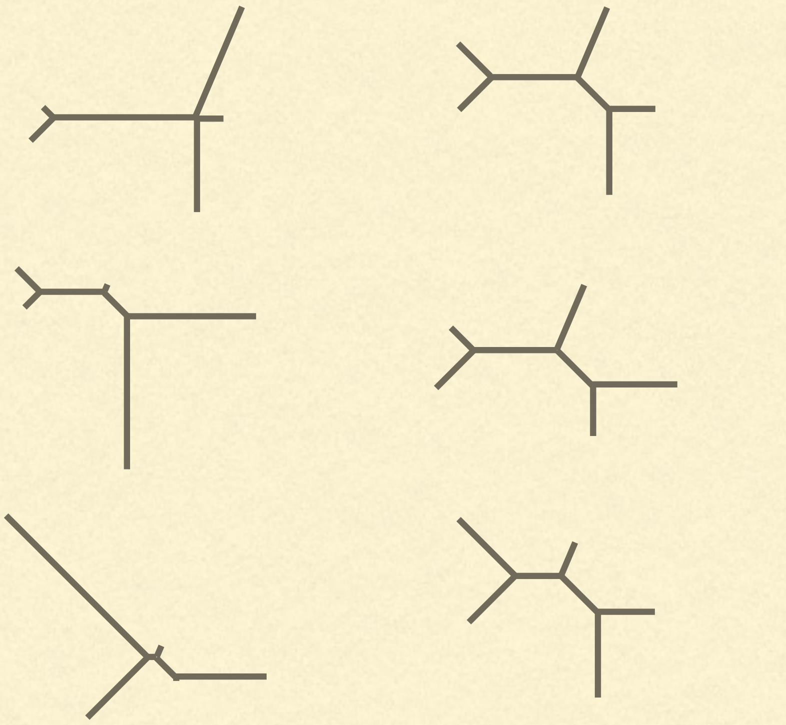


Tree length (TL)

Gamma(2,2)

Edge lengths: Gamma-Dirichlet

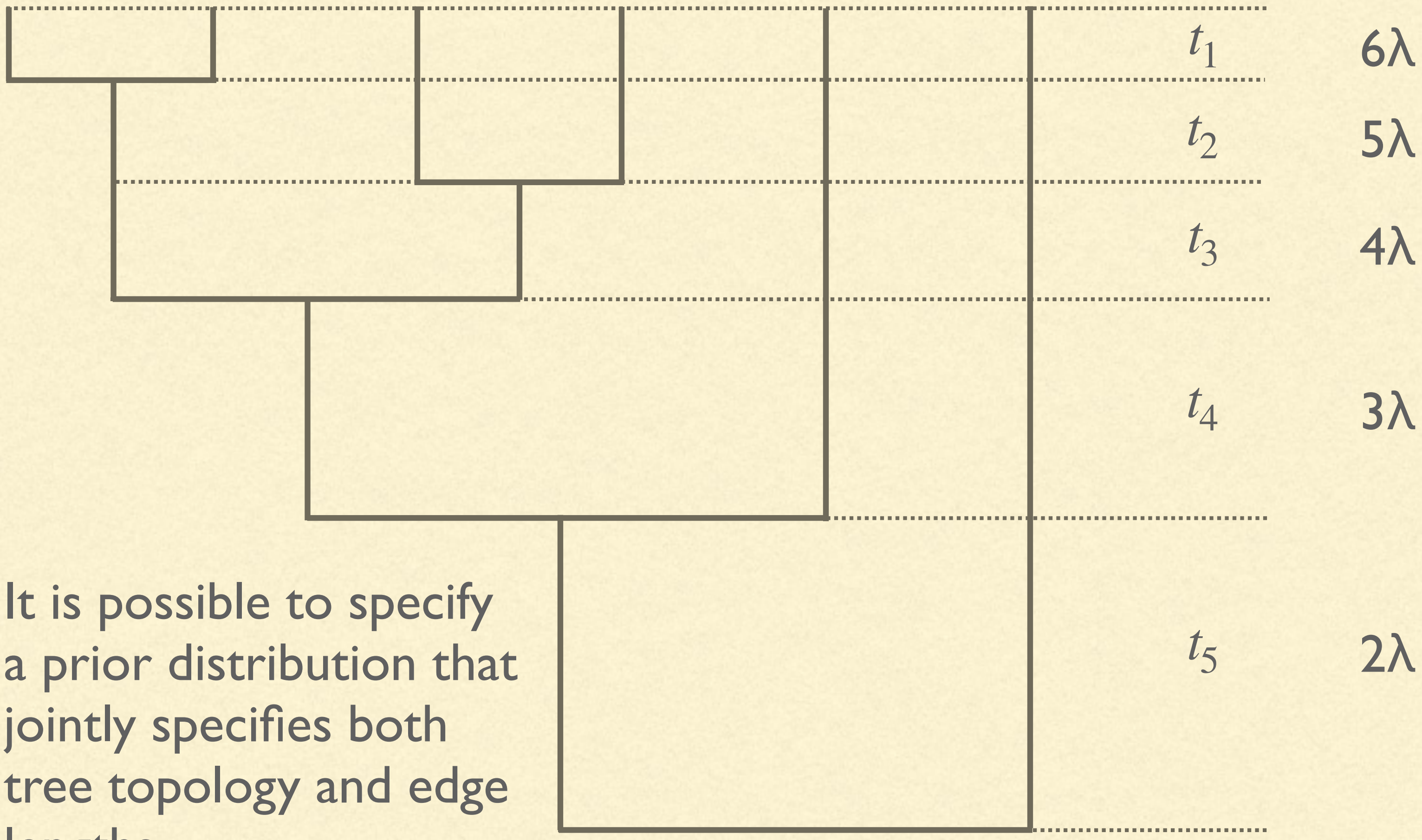




Dir(1,1,1,1,1,1,1)   Dir(10,10,10,10,10,10,10)

Edge length proportions

Yule (pure birth) prior



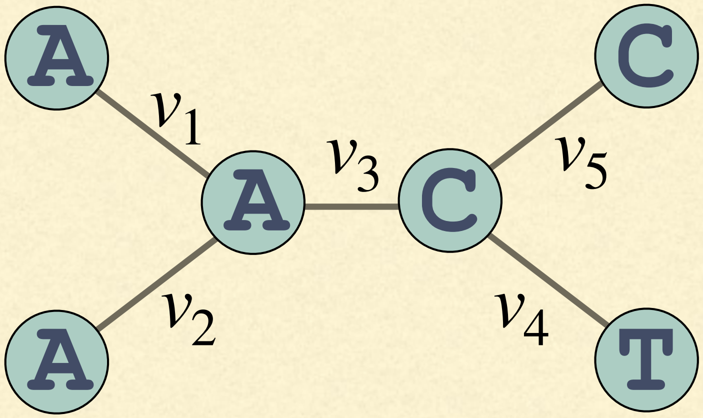
Time to next speciation event is exponentially distributed with rate proportional to the number of lineages.

It is possible to specify a prior distribution that jointly specifies both tree topology and edge lengths.

Hierarchical models

Non-hierarchical model

All model parameters can be found in the likelihood function



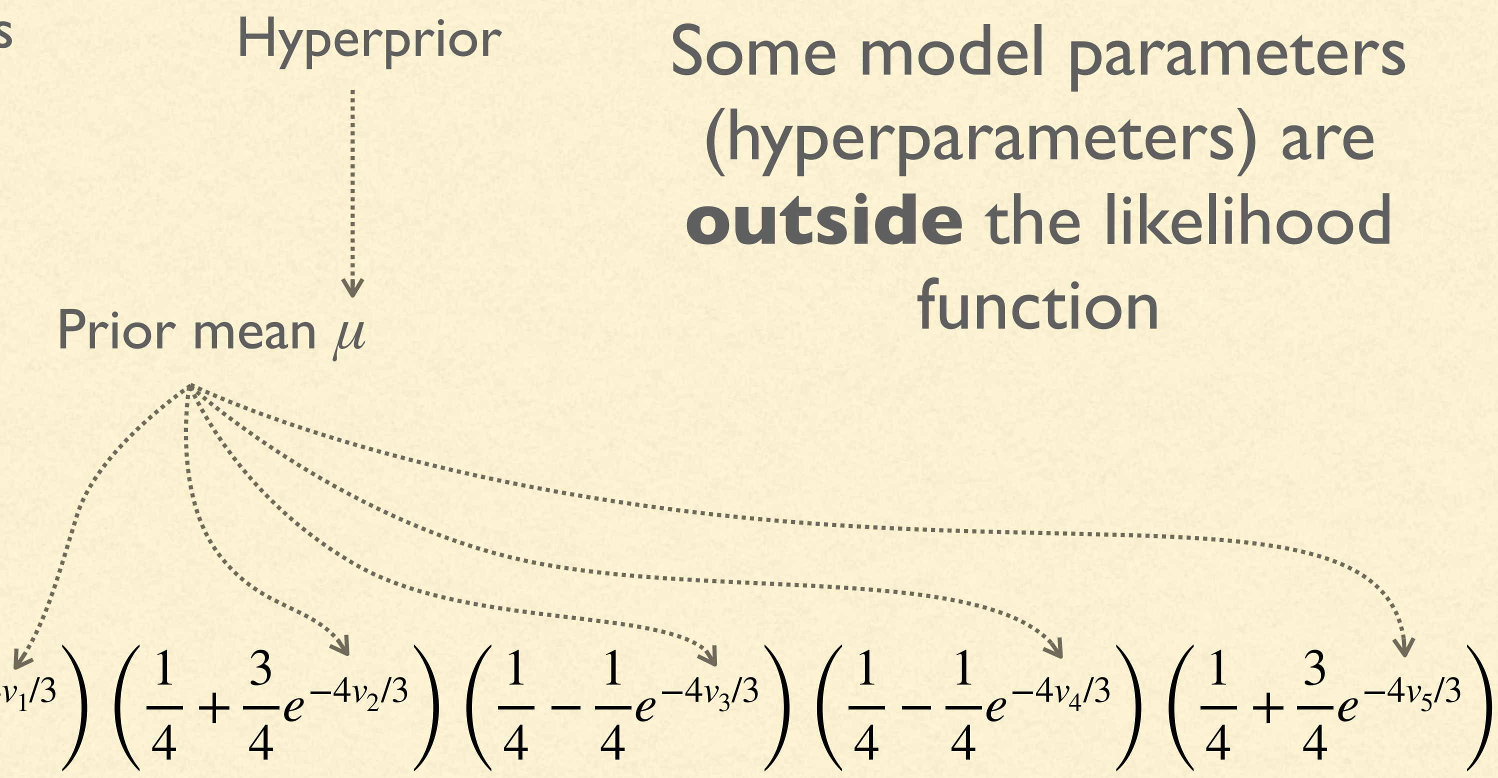
Prior mean 0.1

$$L_k = \frac{1}{4}\left(\frac{1}{4}+\frac{3}{4}e^{-4v_1/3}\right)\left(\frac{1}{4}+\frac{3}{4}e^{-4v_2/3}\right)\left(\frac{1}{4}-\frac{1}{4}e^{-4v_3/3}\right)\left(\frac{1}{4}-\frac{1}{4}e^{-4v_4/3}\right)\left(\frac{1}{4}+\frac{3}{4}e^{-4v_5/3}\right)$$

JC69 likelihood for site k

Hierarchical model

A prior for a hyperparameter is a hyperprior



Hyperprior

Some model parameters (hyperparameters) are **outside** the likelihood function

μ is a hyperparameter

Prior mean μ

$$L_k = \frac{1}{4}\left(\frac{1}{4}+\frac{3}{4}e^{-4v_1/3}\right)\left(\frac{1}{4}+\frac{3}{4}e^{-4v_2/3}\right)\left(\frac{1}{4}-\frac{1}{4}e^{-4v_3/3}\right)\left(\frac{1}{4}-\frac{1}{4}e^{-4v_4/3}\right)\left(\frac{1}{4}+\frac{3}{4}e^{-4v_5/3}\right)$$

JC69 likelihood for site k

Empirical Bayes

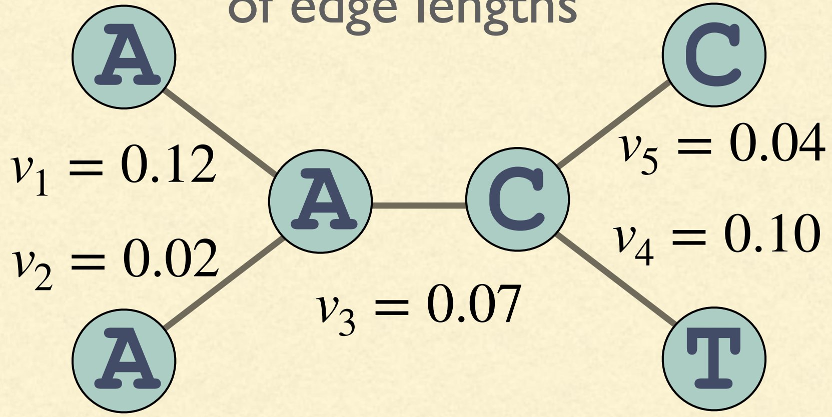
Empirical Bayes approach

Average edge length MLE used as the mean of the prior.

0.07 = (0.12+0.02+0.07+0.04+0.10)/5

Prior mean = 0.07

Maximum likelihood estimates (MLEs) of edge lengths



$$L_k = \frac{1}{4}\left(\frac{1}{4} + \frac{3}{4}e^{-4v_1/3}\right)\left(\frac{1}{4} + \frac{3}{4}e^{-4v_2/3}\right)\left(\frac{1}{4} - \frac{1}{4}e^{-4v_3/3}\right)\left(\frac{1}{4} - \frac{1}{4}e^{-4v_4/3}\right)\left(\frac{1}{4} + \frac{3}{4}e^{-4v_5/3}\right)$$

JC69 likelihood for site k

rjMCMC
(reversible-jump MCMC)

Examples of rjMCMC analyses

a.

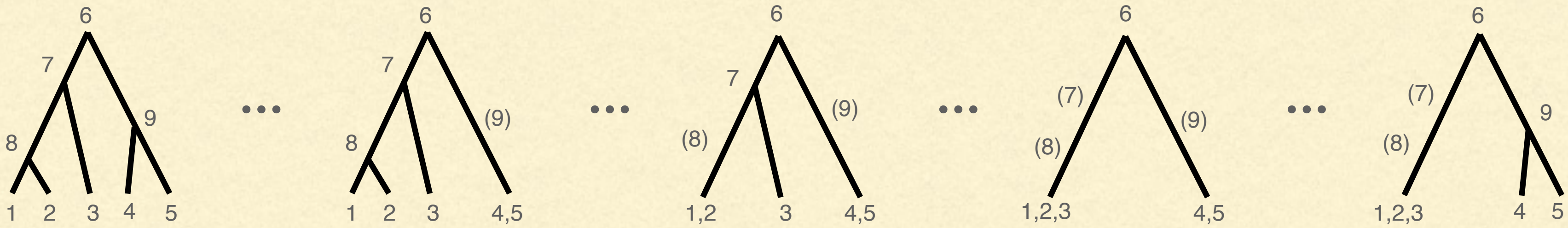
b.

$$\begin{array}{cc} & \begin{array}{cccc} A & C & G & T \end{array} \\ \begin{array}{c} A \\ C \\ G \\ T \end{array} & \left[\begin{array}{cccc} - & \beta & \beta\kappa & \beta \\ \beta & - & \beta & \beta\kappa \\ \beta\kappa & \beta & - & \beta \\ \beta & \beta\kappa & \beta & - \end{array}\right] \end{array} \cdots \begin{array}{cc} & \begin{array}{cccc} A & C & G & T \end{array} \\ \begin{array}{c} A \\ C \\ G \\ T \end{array} & \left[\begin{array}{cccc} - & \pi_C\beta & \pi_G\beta\kappa & \pi_T\beta \\ \pi_A\beta & - & \pi_G\beta & \pi_T\beta\kappa \\ \pi_A\beta\kappa & \pi_C\beta & - & \pi_T\beta \\ \pi_A\beta & \pi_C\beta\kappa & \pi_G\beta & - \end{array}\right] \end{array} \cdots$$

$$\begin{array}{cc} & \begin{array}{cccc} A & C & G & T \end{array} \\ \begin{array}{c} A \\ C \\ G \\ T \end{array} & \left[\begin{array}{cccc} - & a\pi_C\beta & b\pi_G\beta & c\pi_T\beta \\ a\pi_A\beta & - & d\pi_G\beta & e\pi_T\beta \\ b\pi_A\beta & d\pi_C\beta & - & f\pi_T\beta \\ c\pi_A\beta & e\pi_C\beta & f\pi_G\beta & - \end{array}\right] \end{array} \cdots \begin{array}{cc} & \begin{array}{cccc} A & C & G & T \end{array} \\ \begin{array}{c} A \\ C \\ G \\ T \end{array} & \left[\begin{array}{cccc} - & \beta & \beta & \beta \\ \beta & - & \beta & \beta \\ \beta & \beta & - & \beta \\ \beta & \beta & \beta & - \end{array}\right] \end{array}$$

K80 HKY85 GTR JC69

c.



5 species 4 species 3 species 2 species 3 species

a. Lewis, Holder, & Holsinger. 2005.
Systematic Biology 54:241–253.

b. Huelsenbeck, Larget, & Alfaro. 2004.
Molecular Biology and Evolution 21:1123–1133.

c. Rannala & Yang. 2013.
Genetics 194:245–253.

Marginal likelihood and Bayes factors

$$p(D) = \int p(D|\theta)p(\theta)d\theta$$

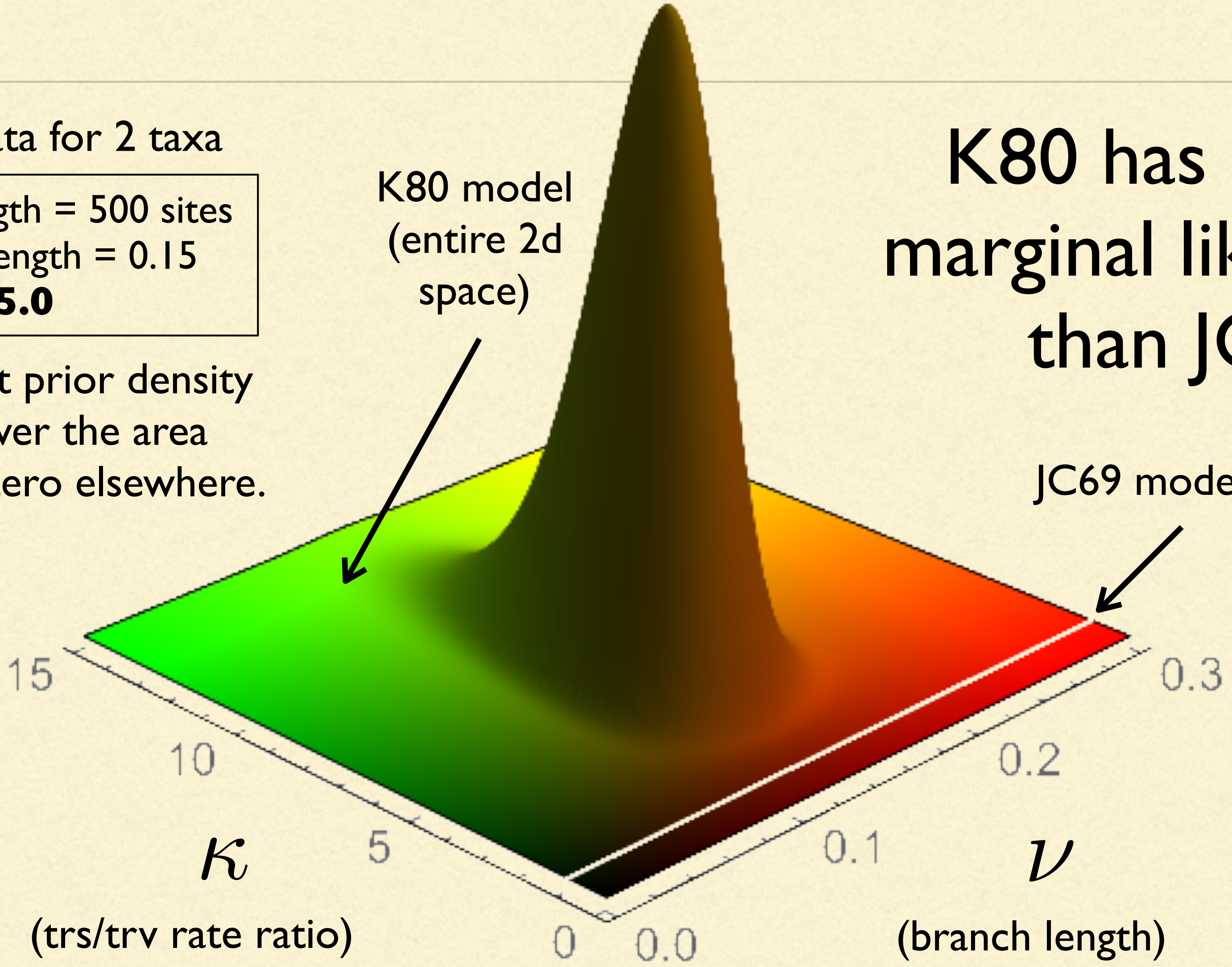
The **marginal likelihood** is a weighted **average** of the **likelihood** where the weights are provided by the prior

Likelihood surface when K80 true

Simulated data for 2 taxa

sequence length = 500 sites
true branch length = 0.15
true kappa = **5.0**

Assume joint prior density is uniform over the area shown and zero elsewhere.



K80 model (entire 2d space)

JC69 model (just this line)

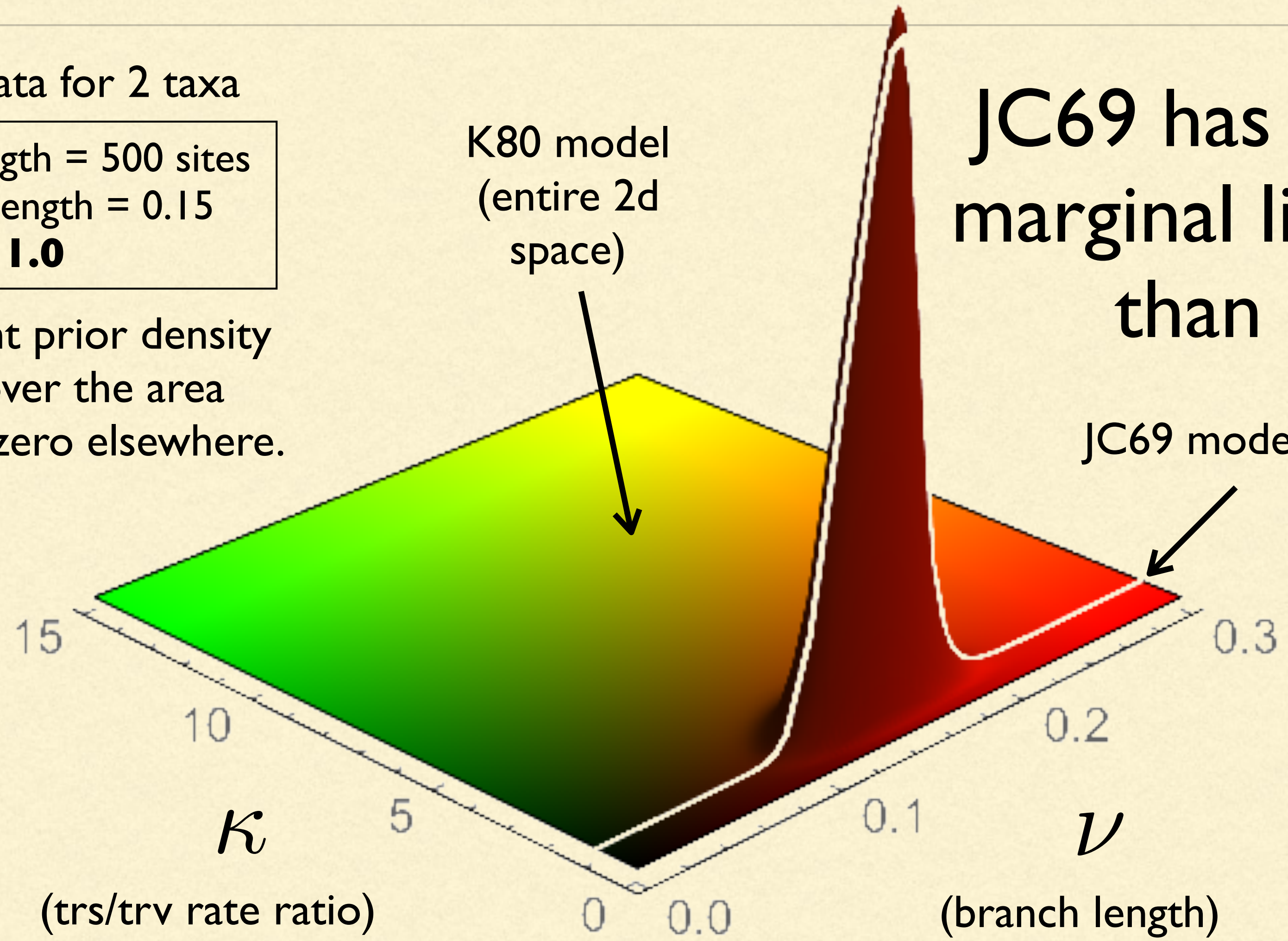
K80 has higher marginal likelihood than JC69

Likelihood surface when JC69 true

Simulated data for 2 taxa

sequence length = 500 sites
true branch length = 0.15
true kappa = **1.0**

Assume joint prior density is uniform over the area shown and zero elsewhere.



K80 model (entire 2d space)

JC69 model (just this line)

(trs/trv rate ratio)

(branch length)

JC69 has a higher marginal likelihood than K80

Dirichlet process (DP) prior

Suppose we have data for 4 loci:
A, B, C, D

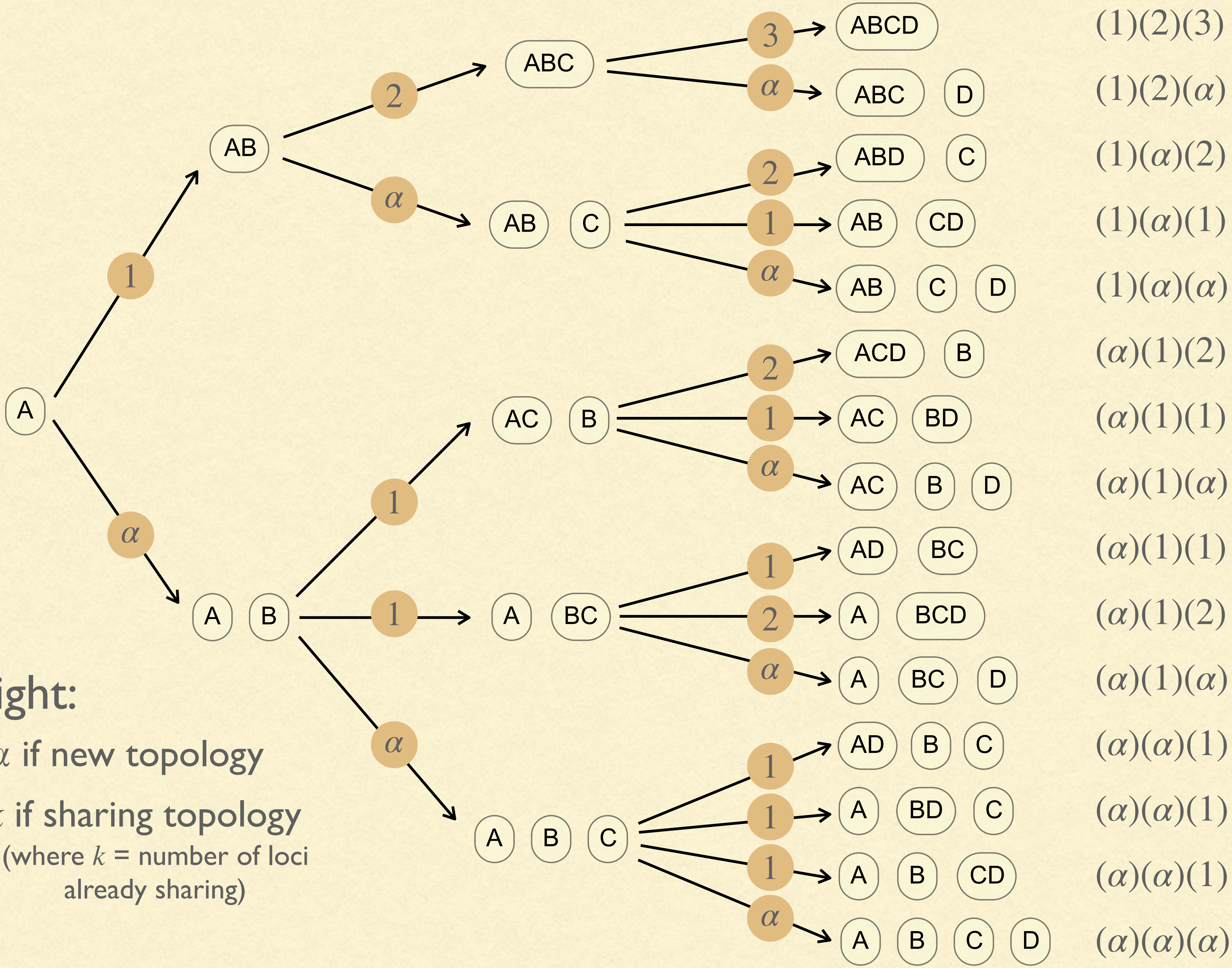
Would like a prior that encourages concordance (i.e. loci have same tree topology) but allows discordance

all loci share same tree topology

ABCD
ABC D
ABD C
AB CD
AB C D
ACD B
AC BD
AC B D
AD BC
A BCD
A BC D
AD B C
A BD C
A B CD
A B C D

each locus has a different tree topology

BUCKy model: Ané et al. 2007. Molecular Biology and Evolution 24:412–426.



ABCD $(1)(2)(3)$

ABC D $(1)(2)(\alpha)$

ABD C $(1)(\alpha)(2)$

AB CD $(1)(\alpha)(1)$

AB C D $(1)(\alpha)(\alpha)$

ACD B $(\alpha)(1)(2)$

AC BD $(\alpha)(1)(1)$

AC B D $(\alpha)(1)(\alpha)$

AD BC $(\alpha)(1)(1)$

A BCD $(\alpha)(1)(2)$

A BC D $(\alpha)(1)(\alpha)$

AD B C $(\alpha)(\alpha)(1)$

A BD C $(\alpha)(\alpha)(1)$

A B CD $(\alpha)(\alpha)(1)$

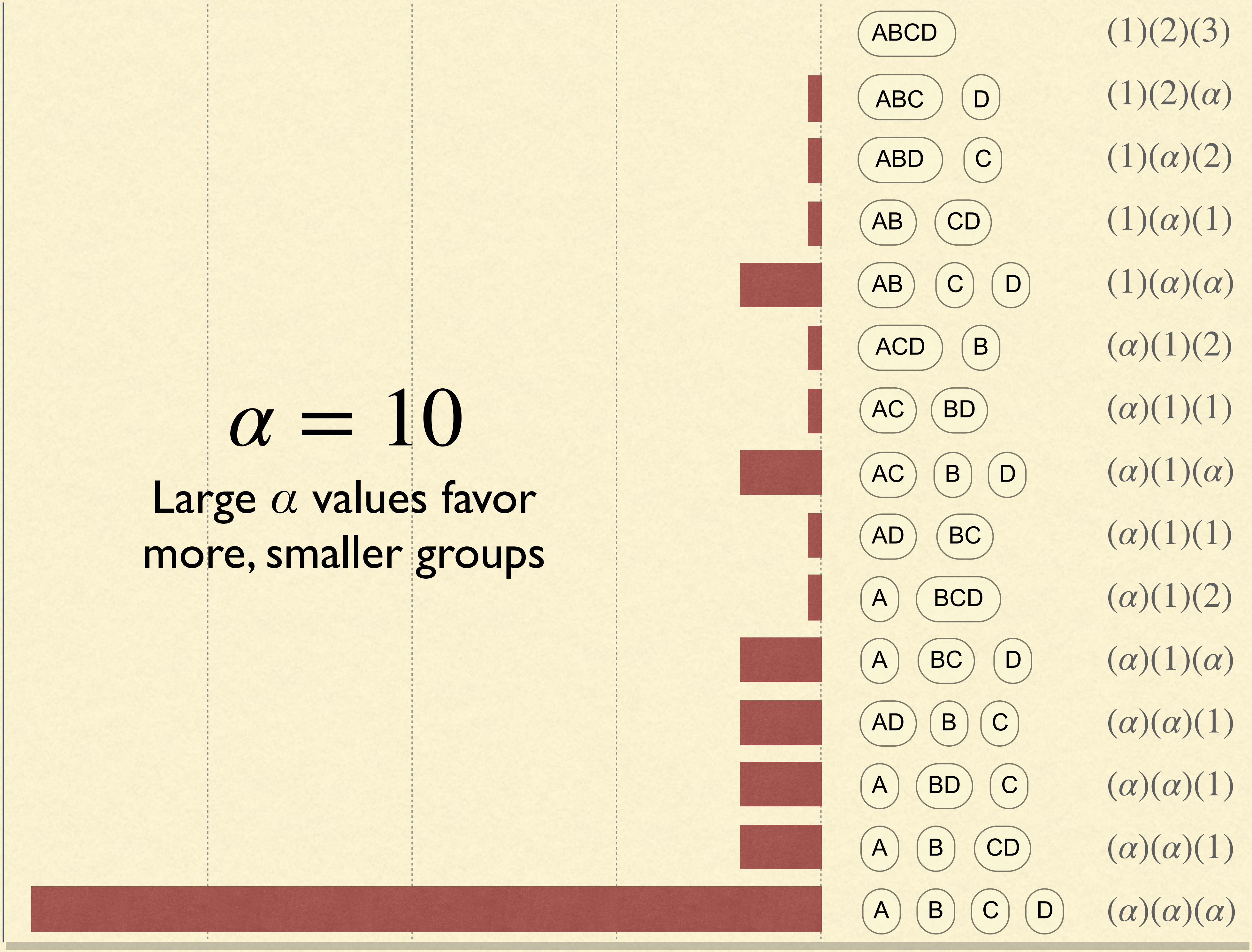
A B C D $(\alpha)(\alpha)(\alpha)$

weight:

α if new topology

k if sharing topology
(where k = number of loci already sharing)

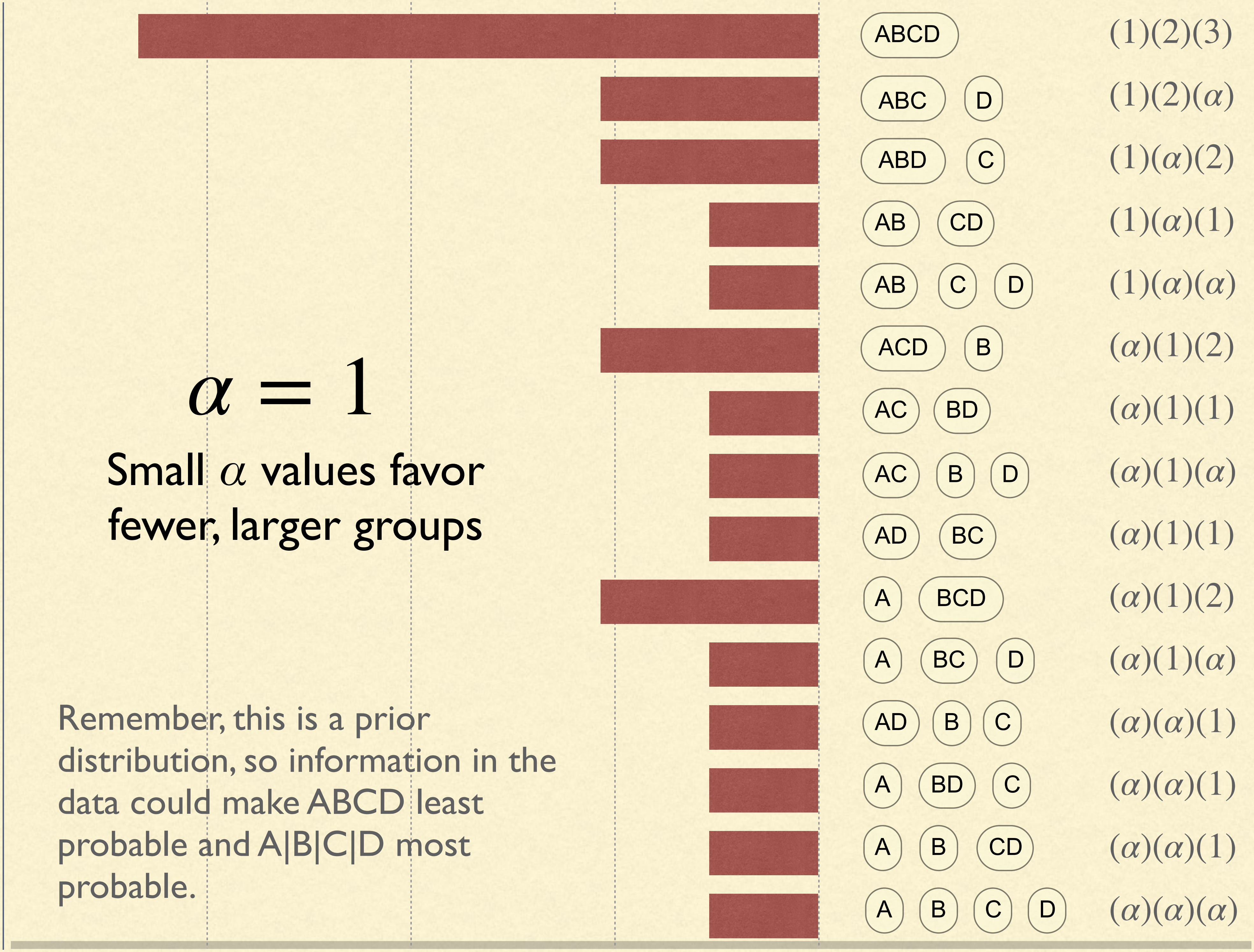
BUCKy model: Ané et al. 2007. Molecular Biology and Evolution 24:412–426.



$$\alpha = 10$$

Large α values favor more, smaller groups

BUCKy model: Ané et al. 2007. Molecular Biology and Evolution 24:412–426.



$$\alpha = 1$$

Small α values favor fewer, larger groups

Remember, this is a prior distribution, so information in the data could make ABCD least probable and A|B|C|D most probable.

BUCKy model: Ané et al. 2007. Molecular Biology and Evolution 24:412–426.

Dirichlet Process Prior applet

https://plewis.github.io/applets/dpp/