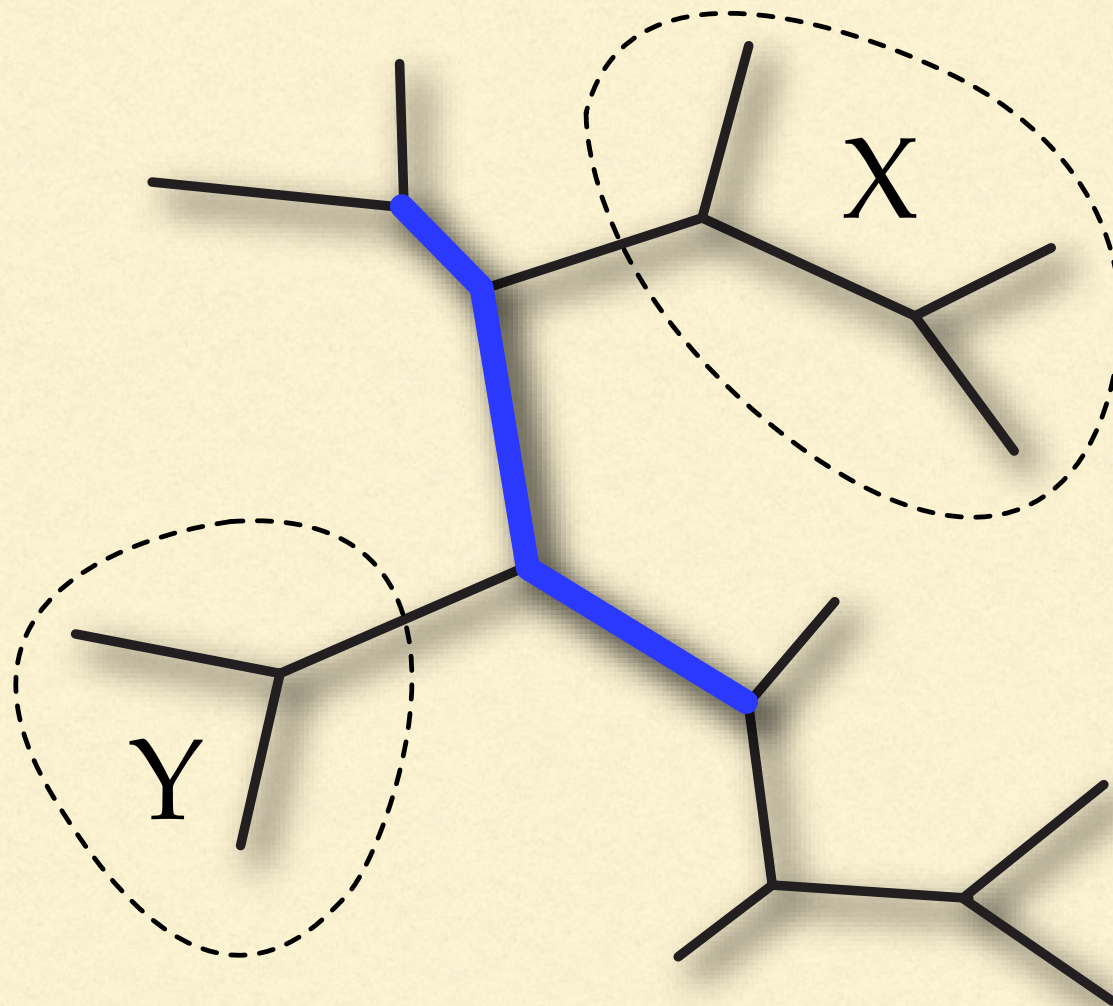


---

# MCMC proposals

---

# Moving through treespace



## The Larget-Simon move

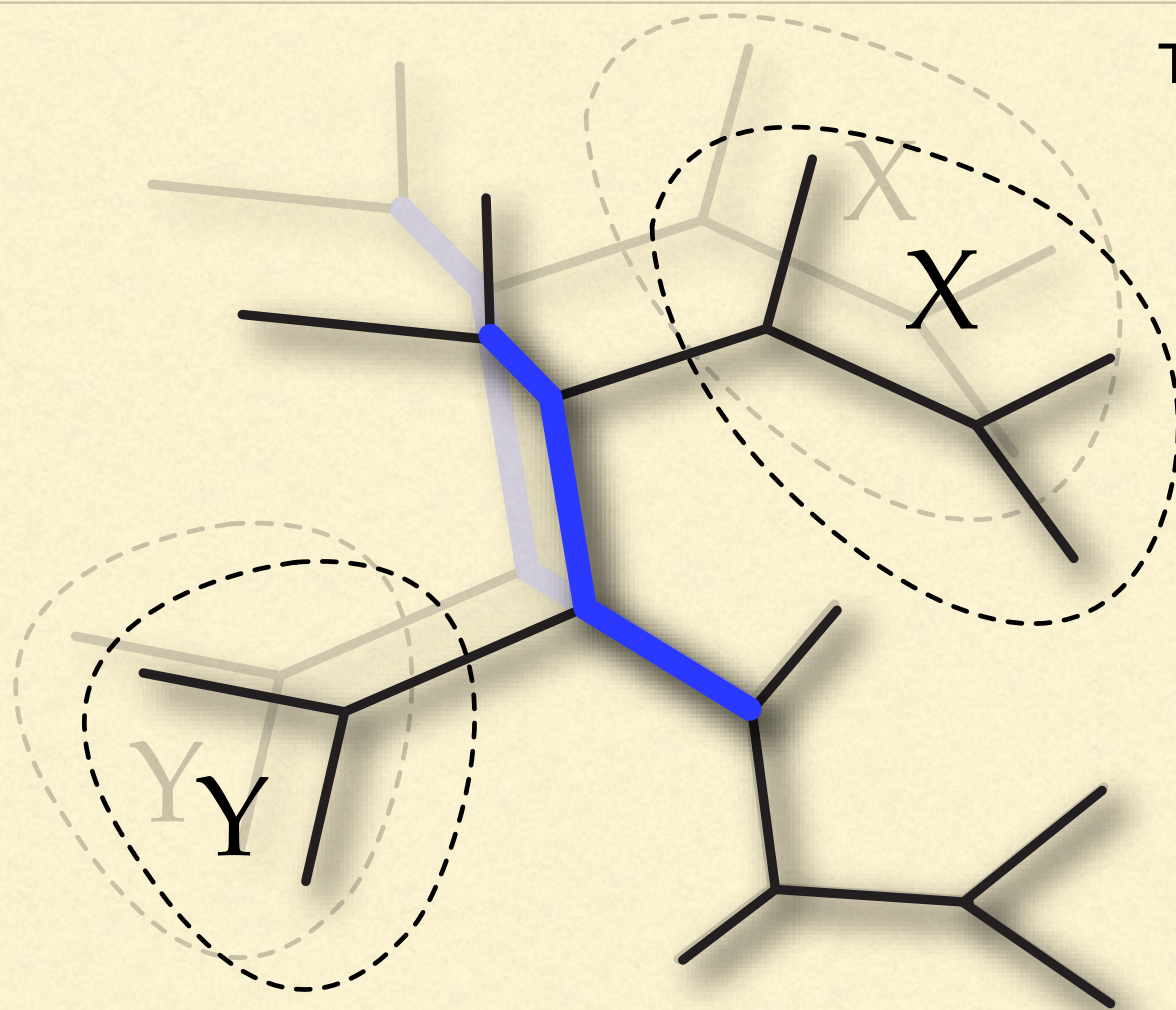
### **Step 1:**

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:

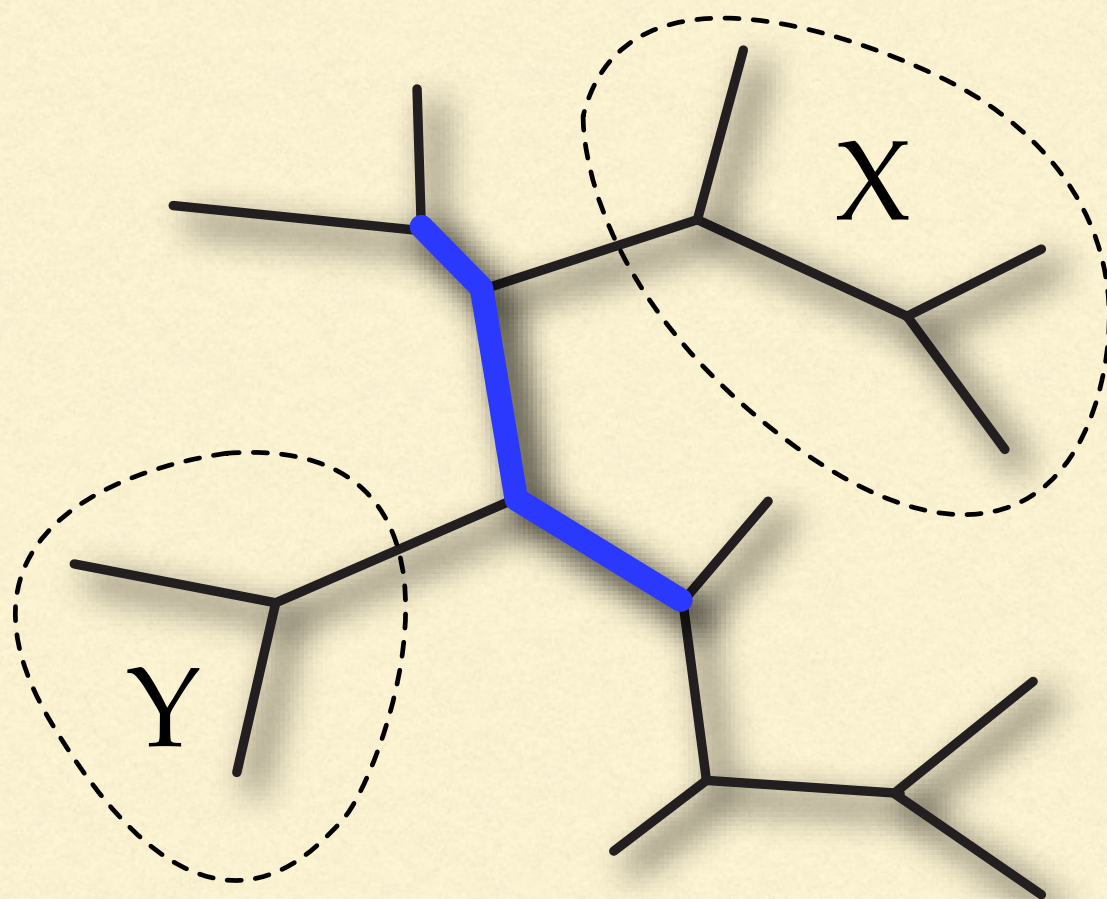
Shrink or grow selected 3-edge segment by a random amount

---

# Moving through treespace

---

## The 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

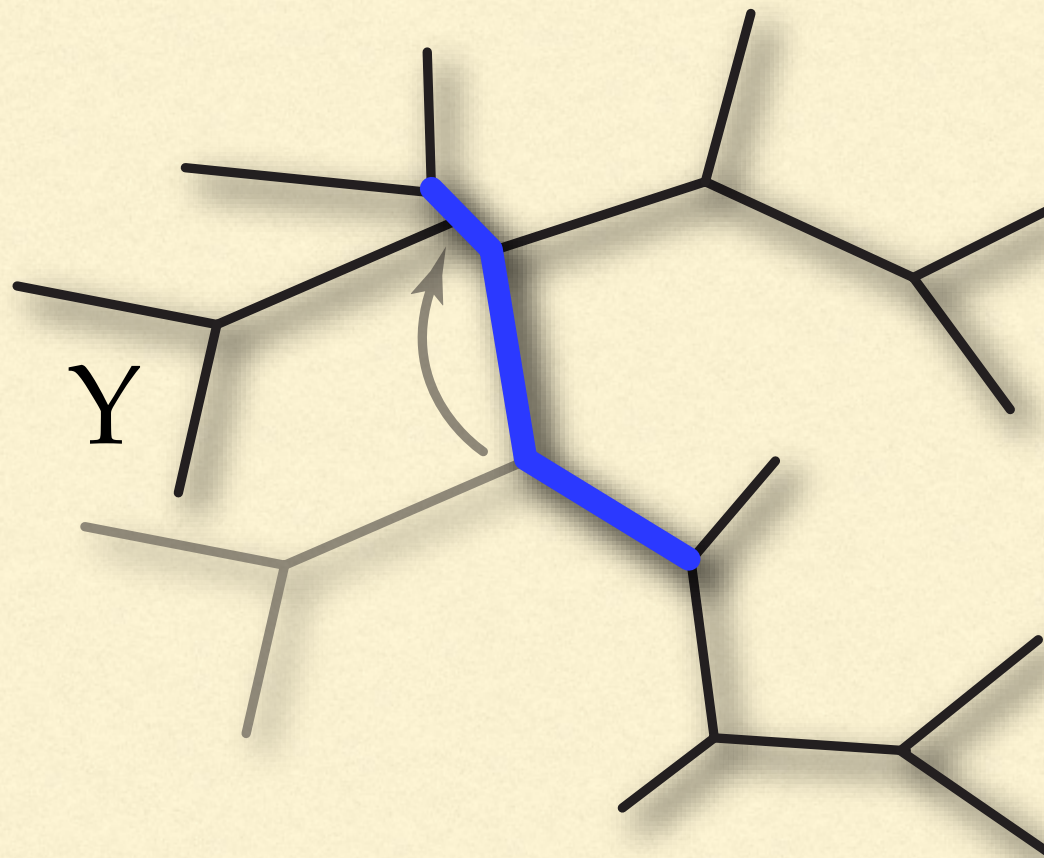


---

# Moving through treespace

---

## The 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:

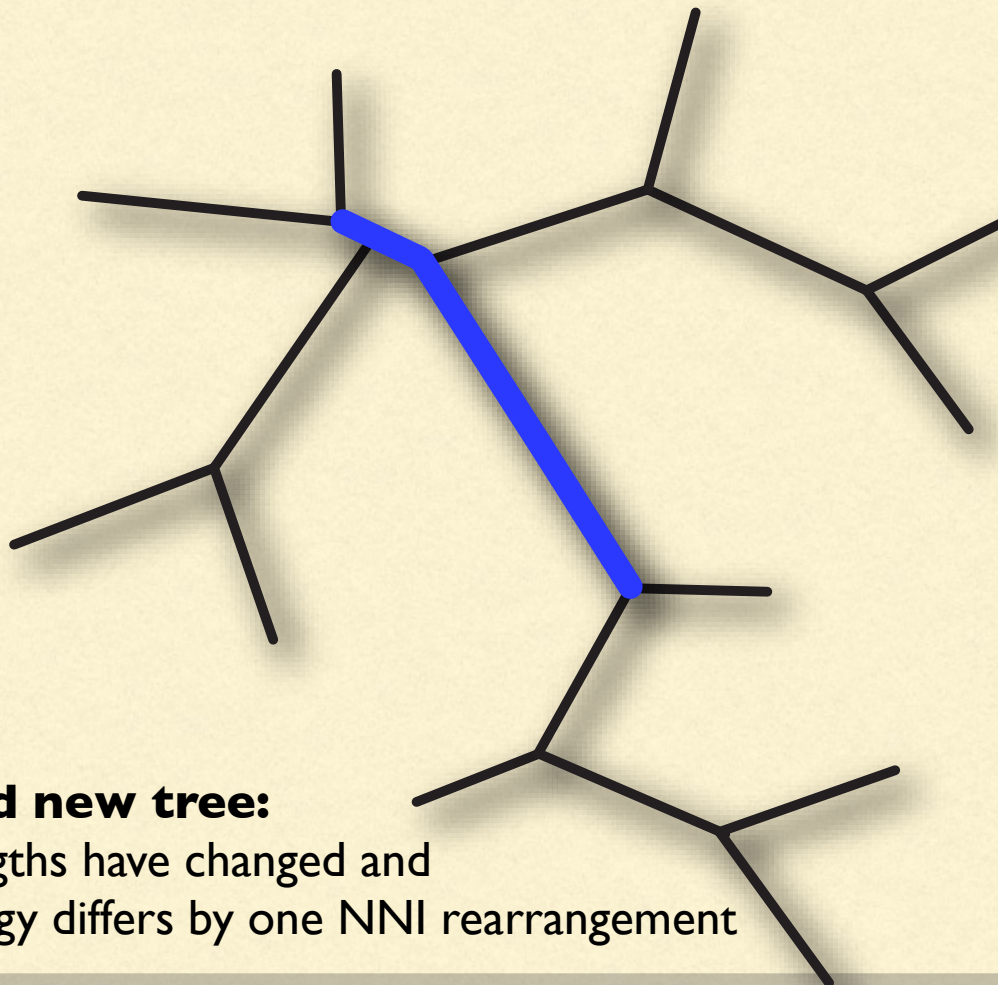
Choose X or Y randomly, then reposition randomly

---

# Moving through treespace

---

## The Target-Simon move



### **Proposed new tree:**

3 edge lengths have changed and the topology differs by one NNI rearrangement

### **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:**

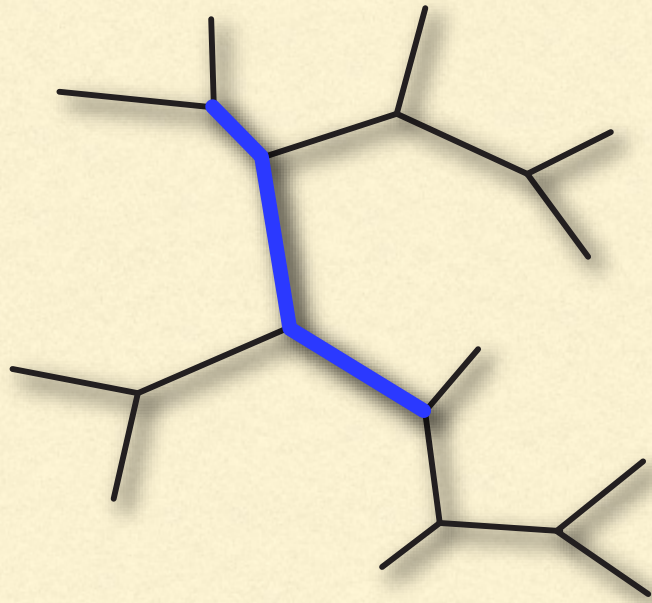
Choose X or Y randomly, then reposition randomly



---

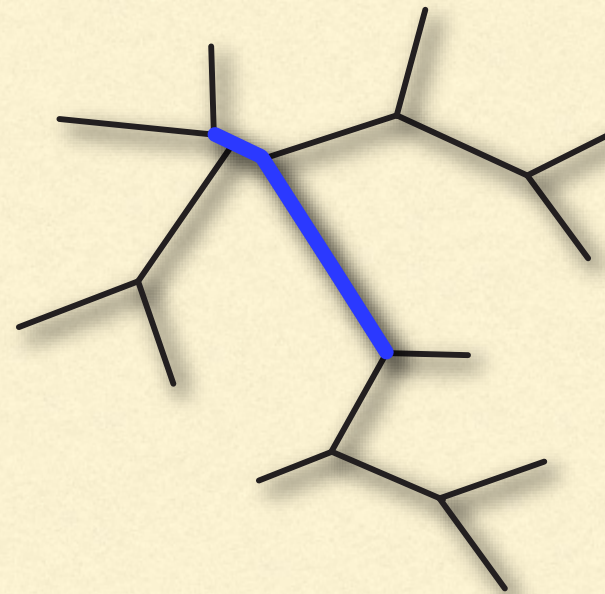
# Moving through treespace

---



Current tree

log-posterior = -34256

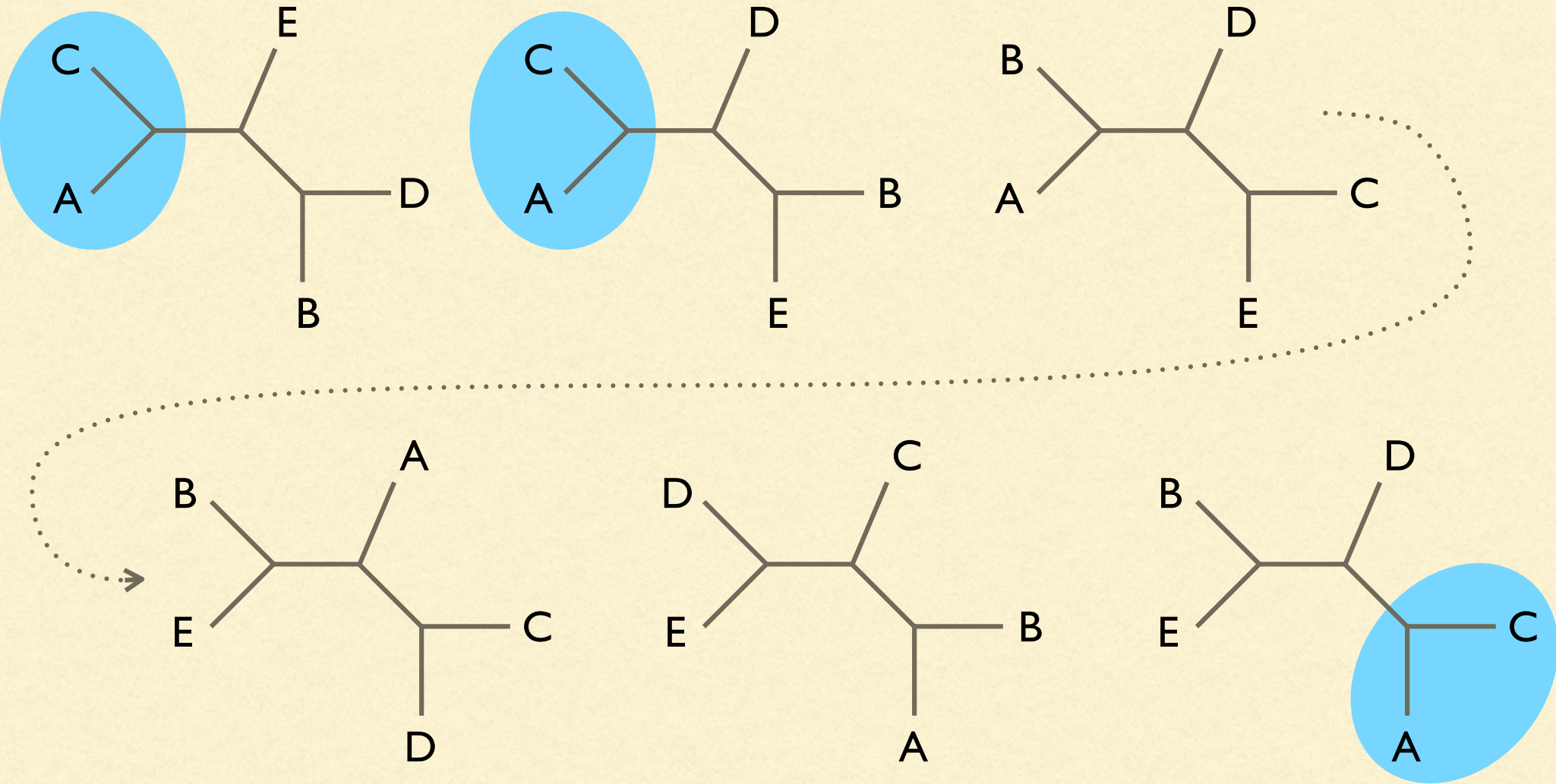


Proposed tree

log-posterior = -32519

(better, so accept)

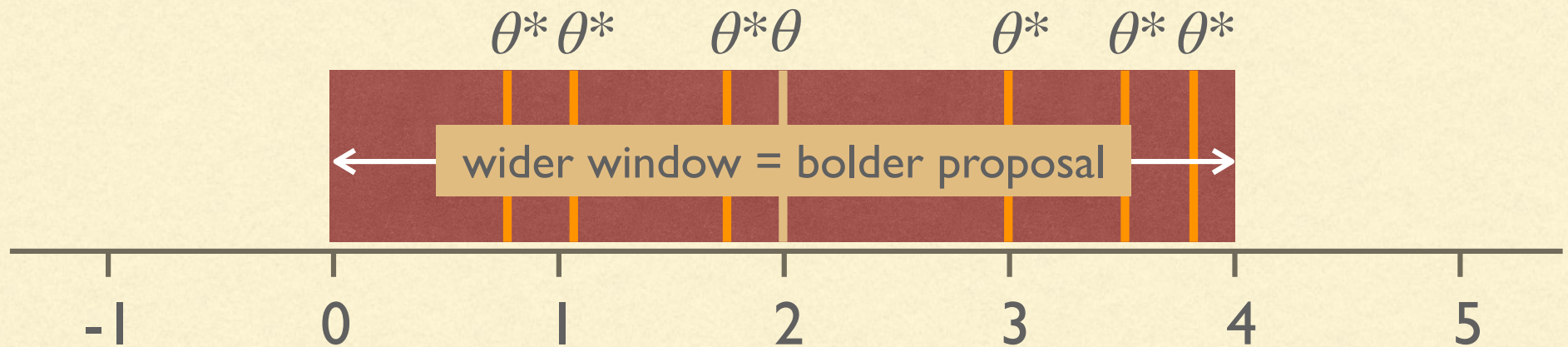
# Marginal split posterior probabilities





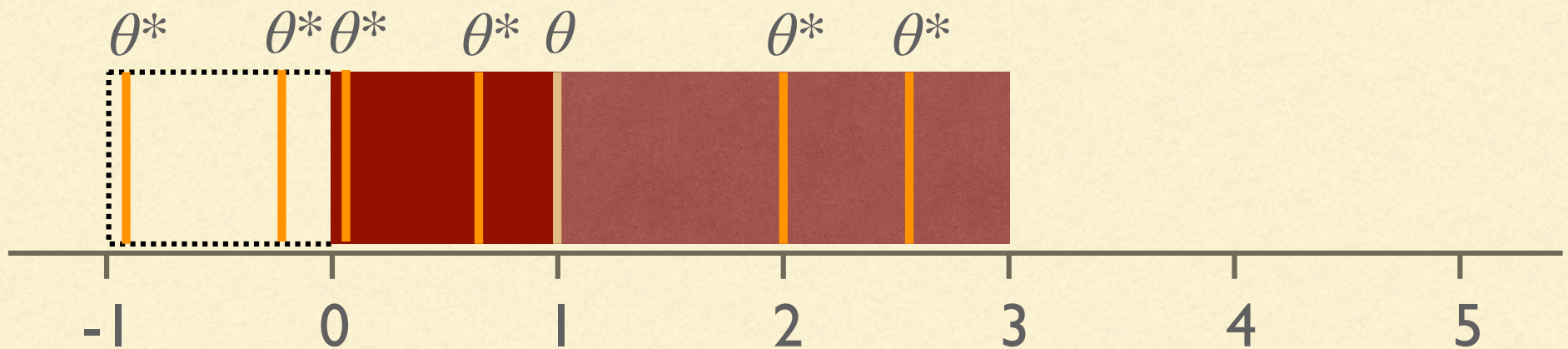
# Sliding window proposal

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



# Sliding window proposal

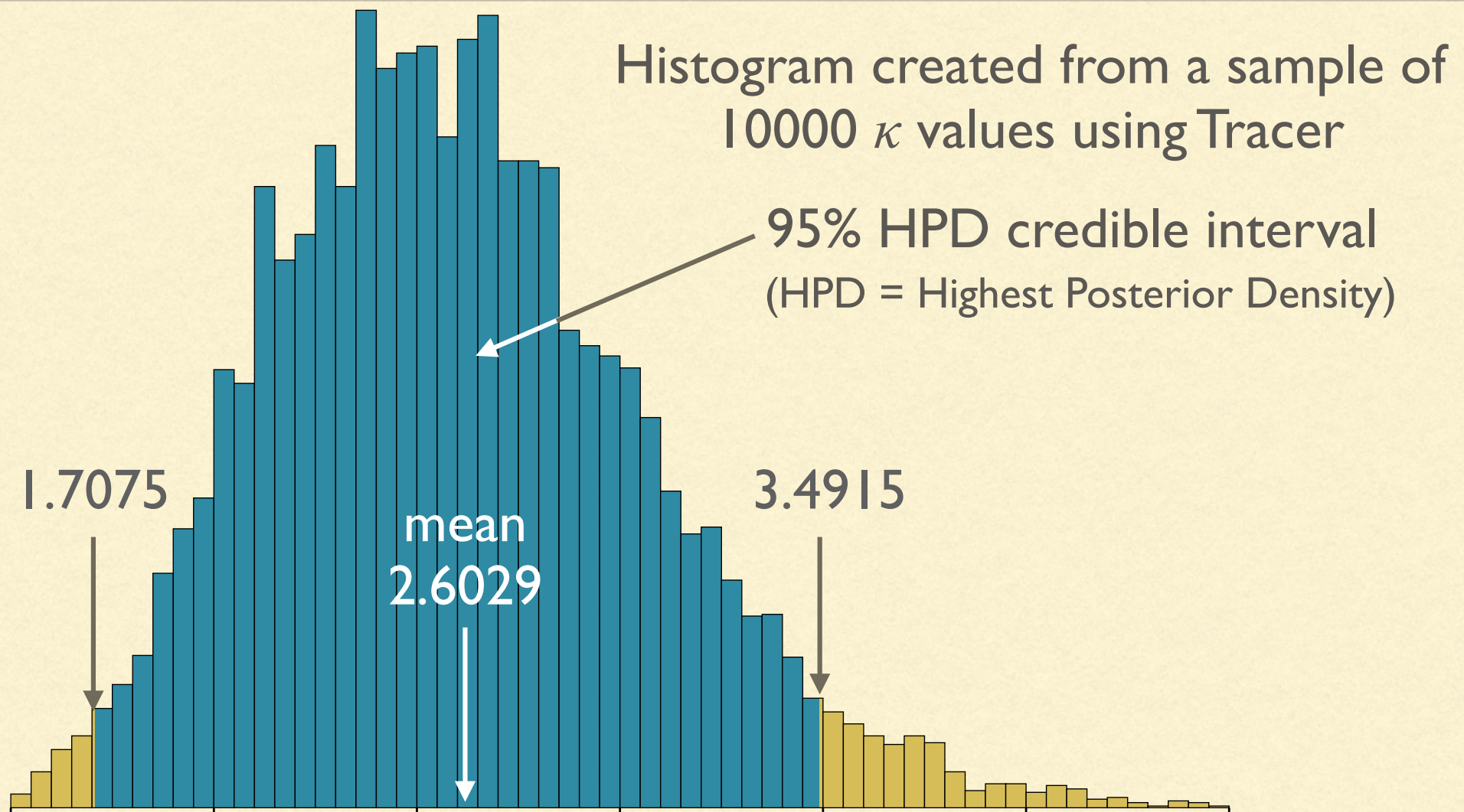
If proposed value is out of bounds,  
reflect it back in bounds



Surprisingly, this is a symmetric  
proposal, so the Hastings ratio = 1.0



# Marginal distributions and credible intervals



---

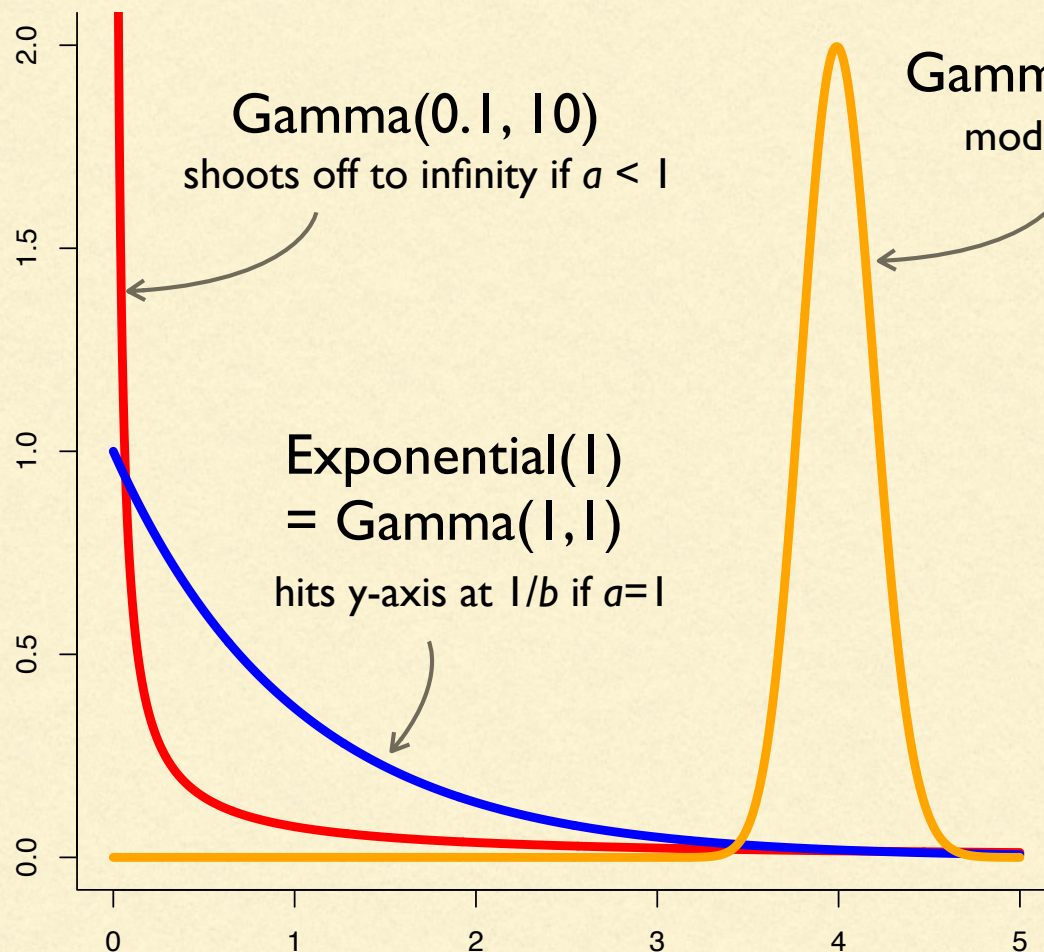
# Prior distributions

---

$$p(\theta | D) = \frac{p(D | \theta) p(\theta)}{p(D)}$$



# Gamma( $a,b$ ) distribution



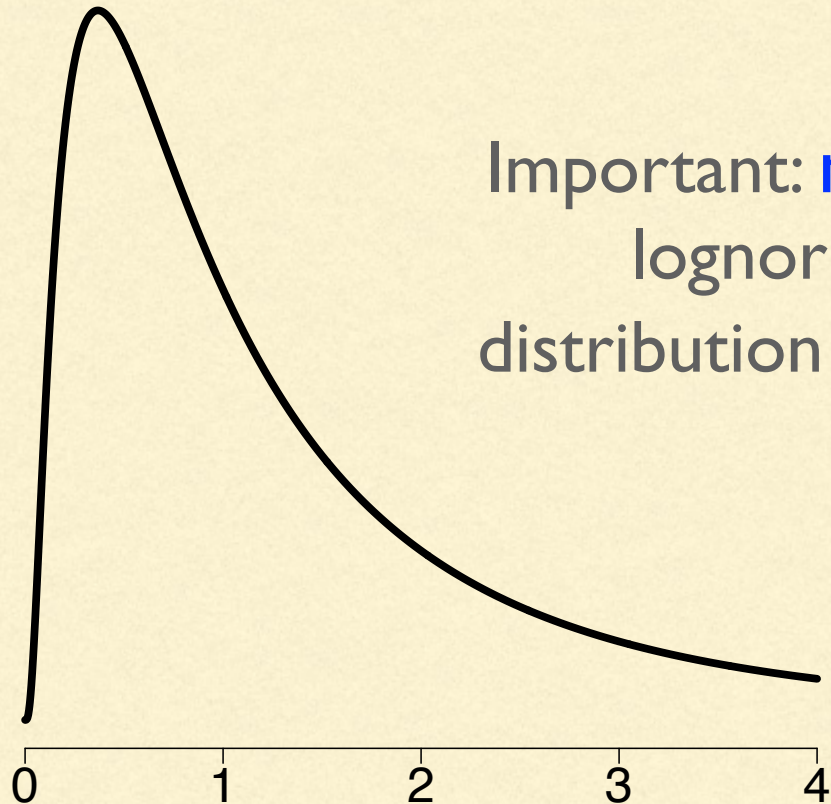
**Gamma( $a,b$ )**  
distributions are  
appropriate for  
parameters that range  
from 0 to infinity (e.g.  
branch lengths)

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

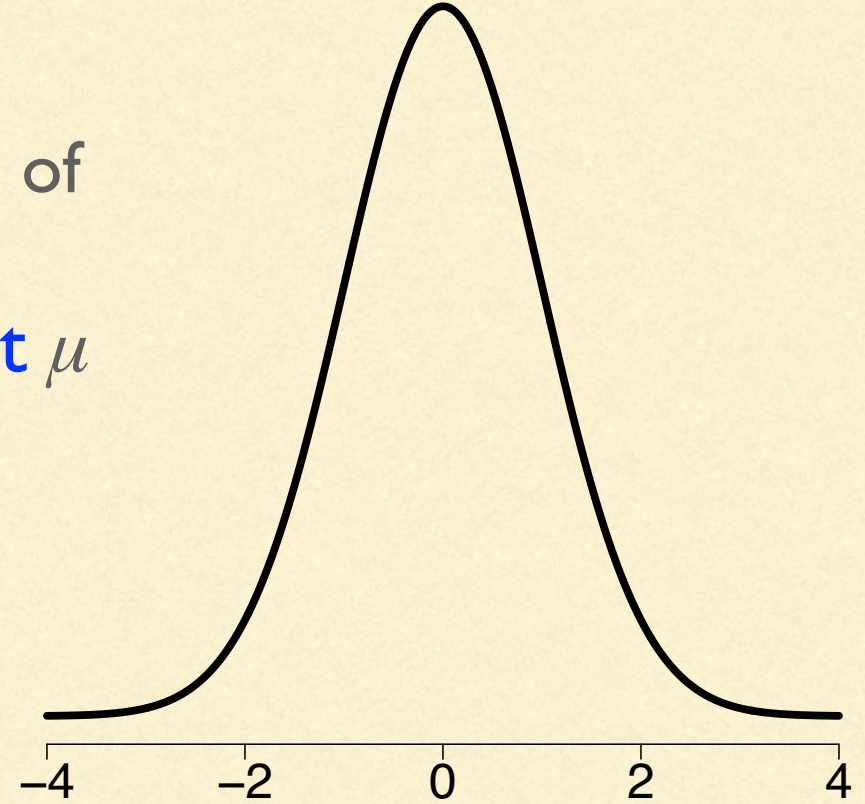
\*Note: be aware that in many papers the Gamma distribution is defined such that the second parameter is the rate (*inverse* of the scale  $b$  used in this slide) rather than the scale! In this case, the mean and variance would be  $a/b$  and  $a/b^2$ , respectively.

# Lognormal( $\mu, \sigma$ ) distribution

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



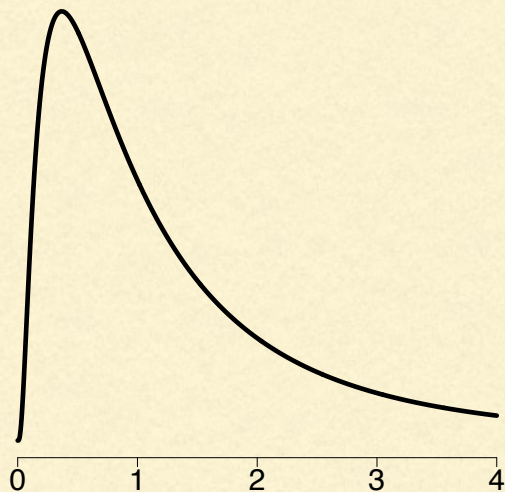
Normal( $\mu = 0, \sigma = 1$ )



Important: **mean** of  
lognormal  
distribution is **not**  $\mu$



# Lognormal( $\mu, \sigma$ ) distribution



$$\mu = 0, \sigma = 1$$

$$1.65$$

$$4.67$$

$$0.37$$

$$1.0$$

$$\text{mean} = e^{\mu + \sigma^2/2}$$

$$\text{variance} = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

$$\text{mode} = e^{\mu - \sigma^2}$$

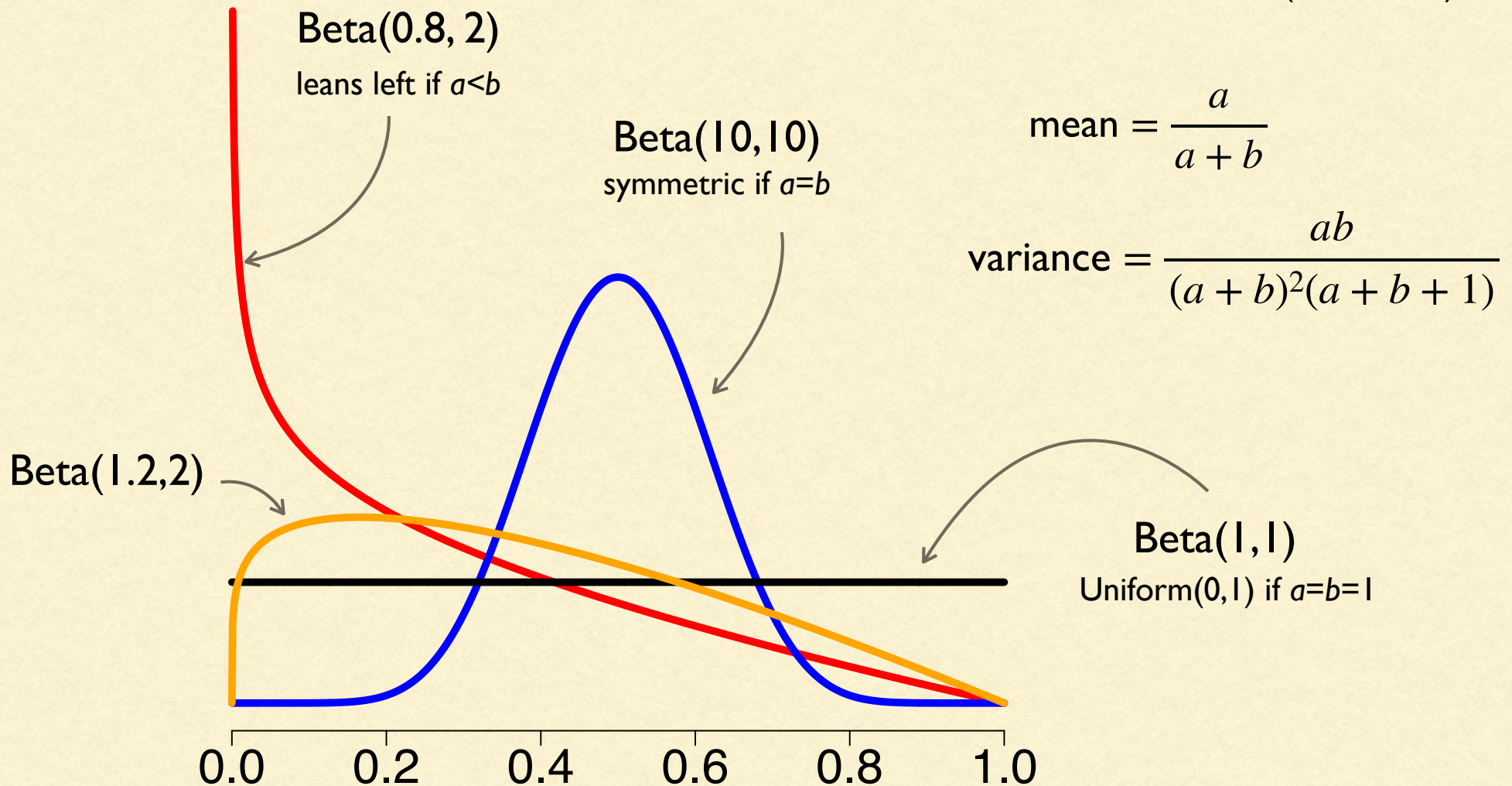
$$\text{median} = e^{\mu}$$

To choose  $\mu$  and  $\sigma$  to yield a particular mean ( $m$ ) and variance ( $v$ ) for a lognormal prior, use these formulas (log is natural logarithm):

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

# Beta( $a,b$ ) distribution

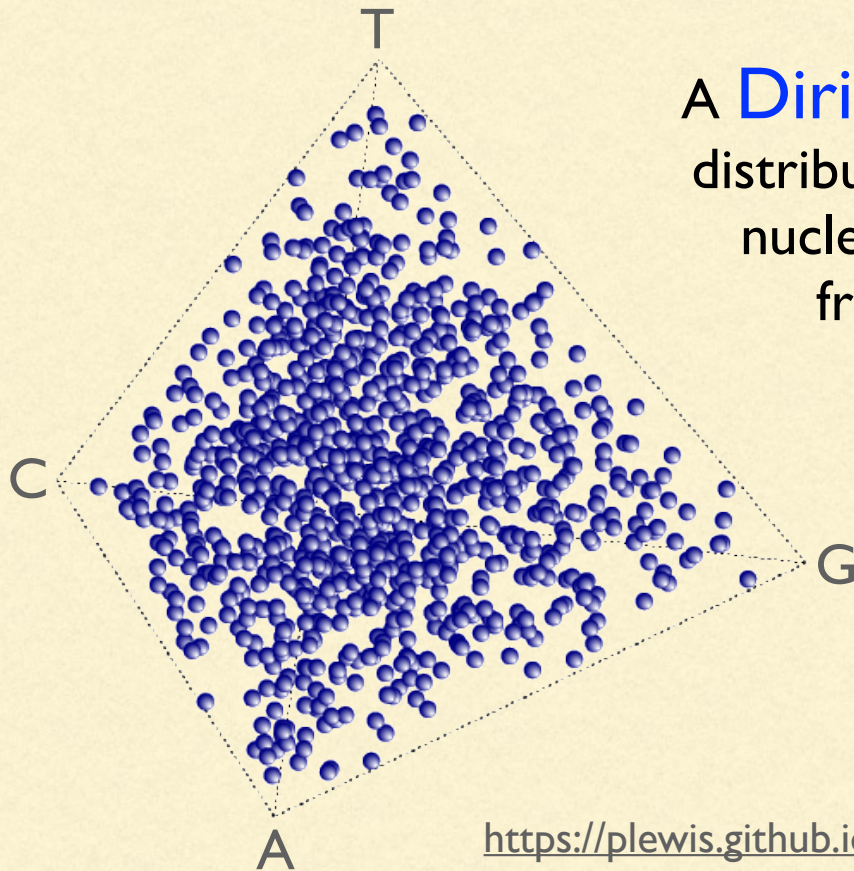
Beta( $a,b$ ) distributions are appropriate for proportions, which must lie between 0 and 1 (inclusive).



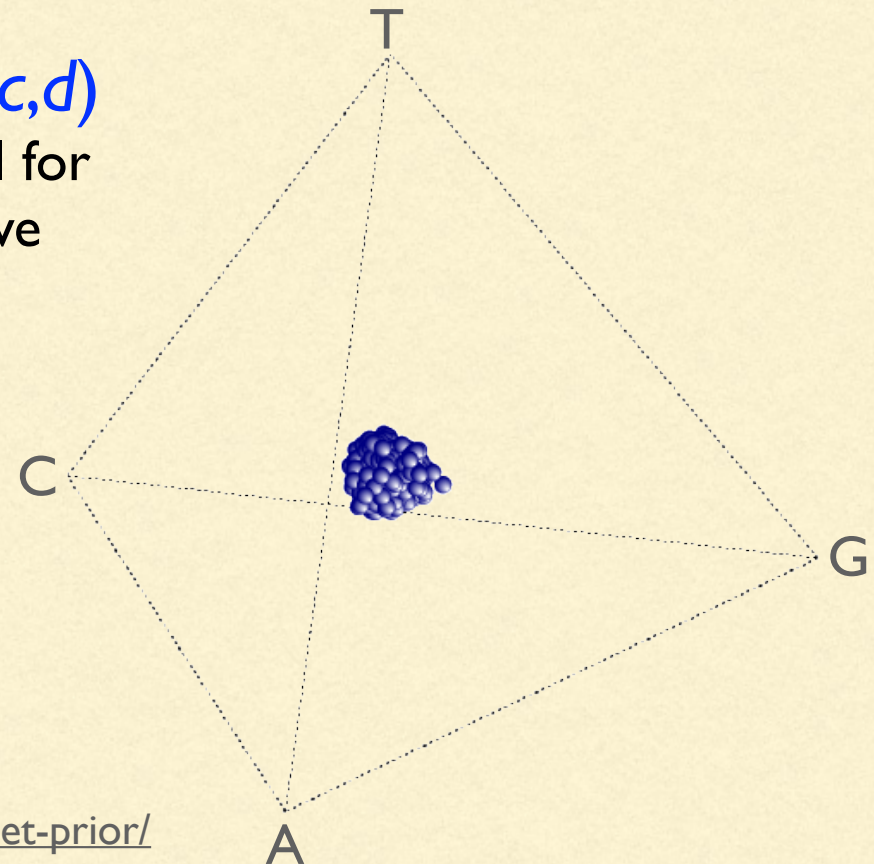


# Dirichlet( $a, b, c, d$ ) distribution

**Flat:**  $a = b = c = d = 1$   
(every combination equally probable)



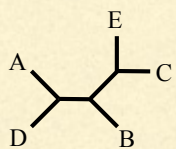
**Informative:**  $a = b = c = d = 100$   
(frequencies tend to be nearly equal)



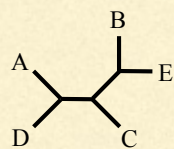
A **Dirichlet( $a, b, c, d$ )** distribution is ideal for nucleotide relative frequencies.

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

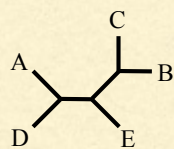
# Topology: discrete uniform



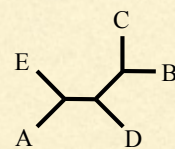
$$\frac{1}{15}$$



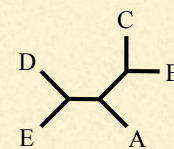
$$\frac{1}{15}$$



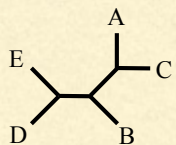
$$\frac{1}{15}$$



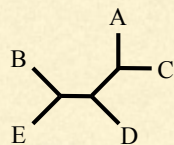
$$\frac{1}{15}$$



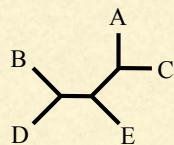
$$\frac{1}{15}$$



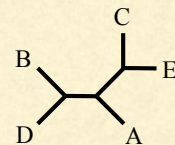
$$\frac{1}{15}$$



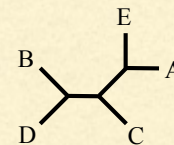
$$\frac{1}{15}$$



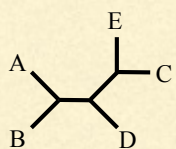
$$\frac{1}{15}$$



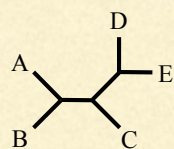
$$\frac{1}{15}$$



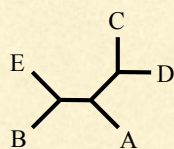
$$\frac{1}{15}$$



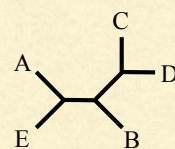
$$\frac{1}{15}$$



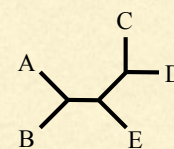
$$\frac{1}{15}$$



$$\frac{1}{15}$$



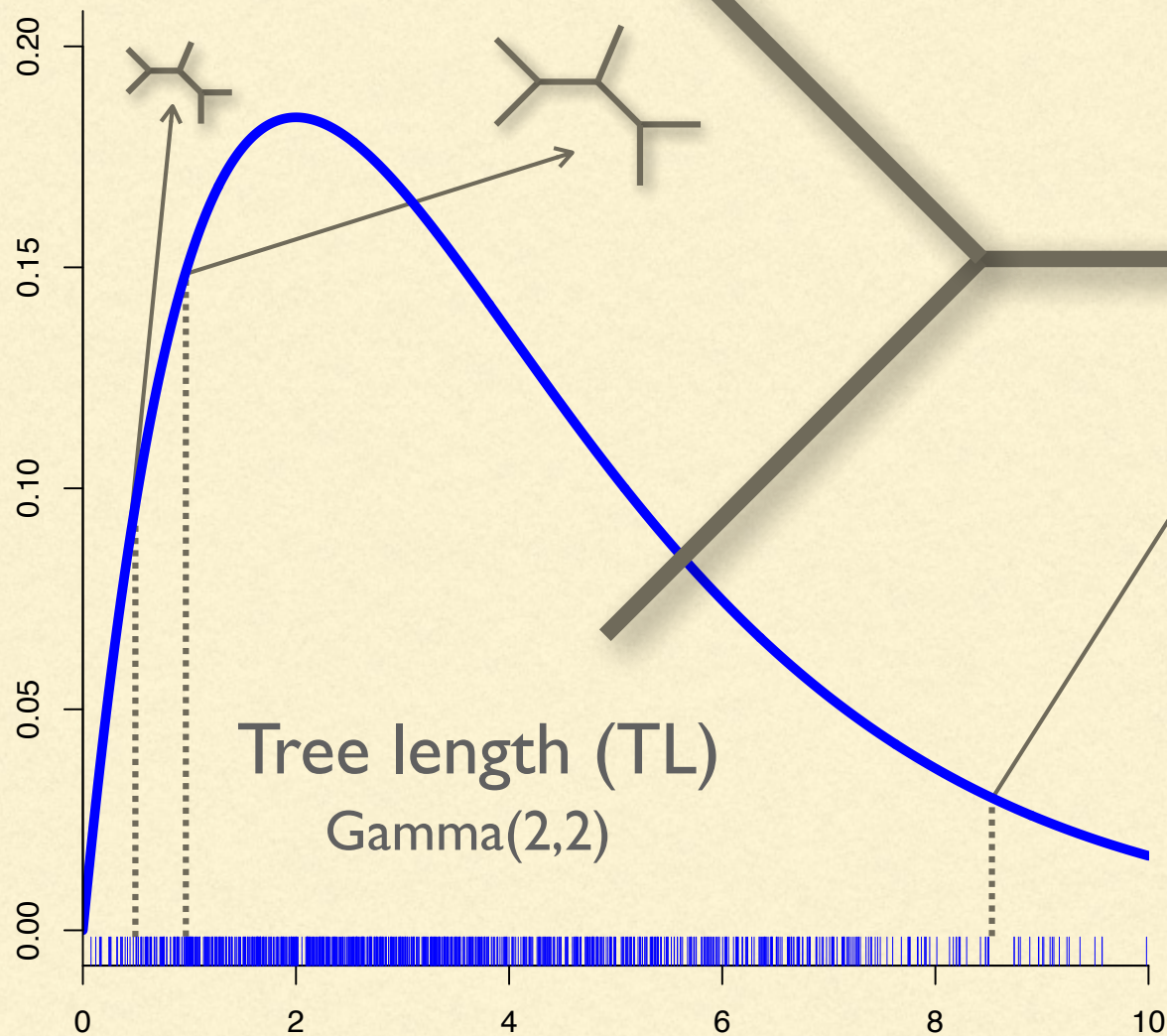
$$\frac{1}{15}$$



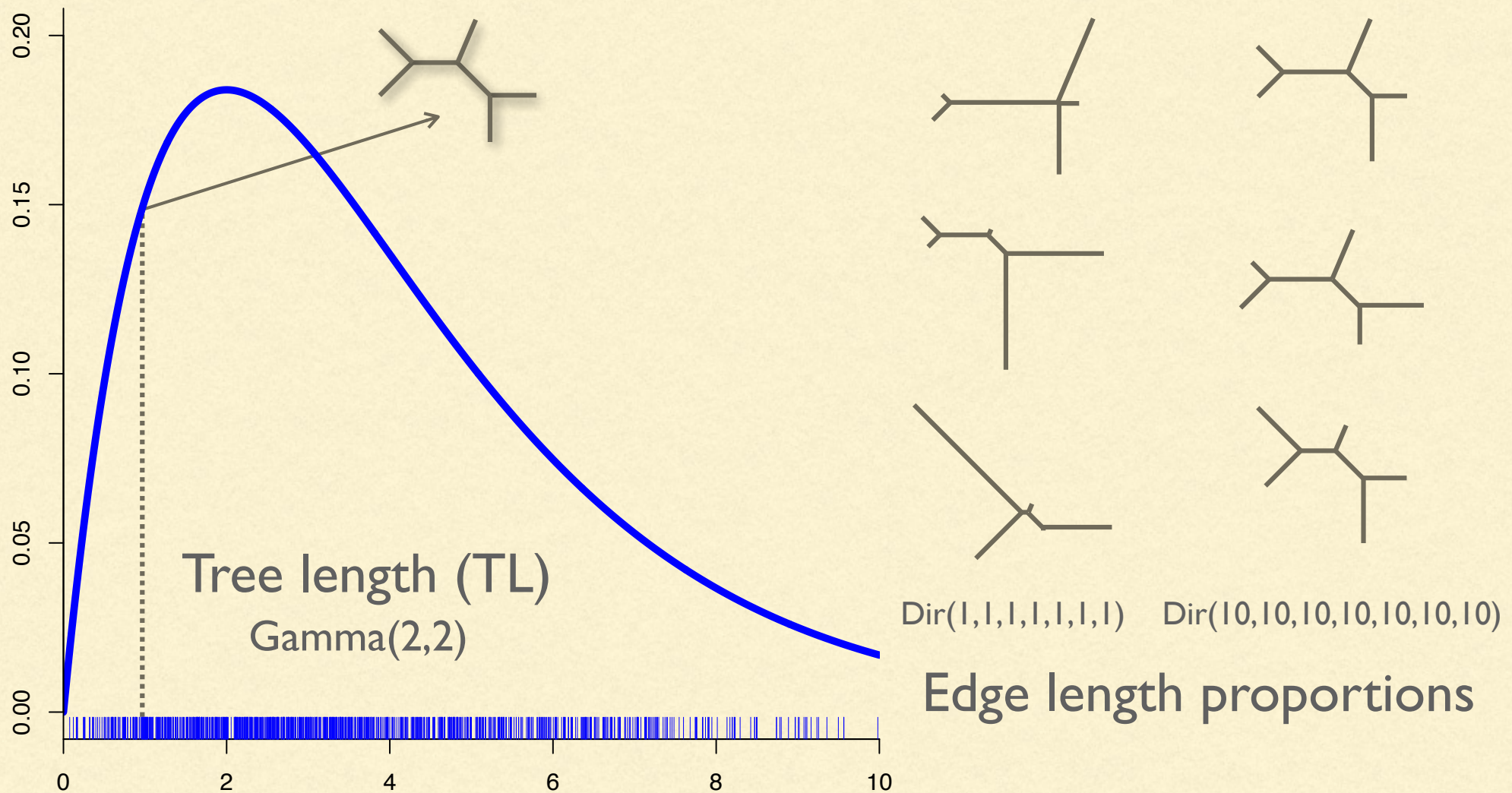
$$\frac{1}{15}$$



# Edge lengths: Gamma-Dirichlet

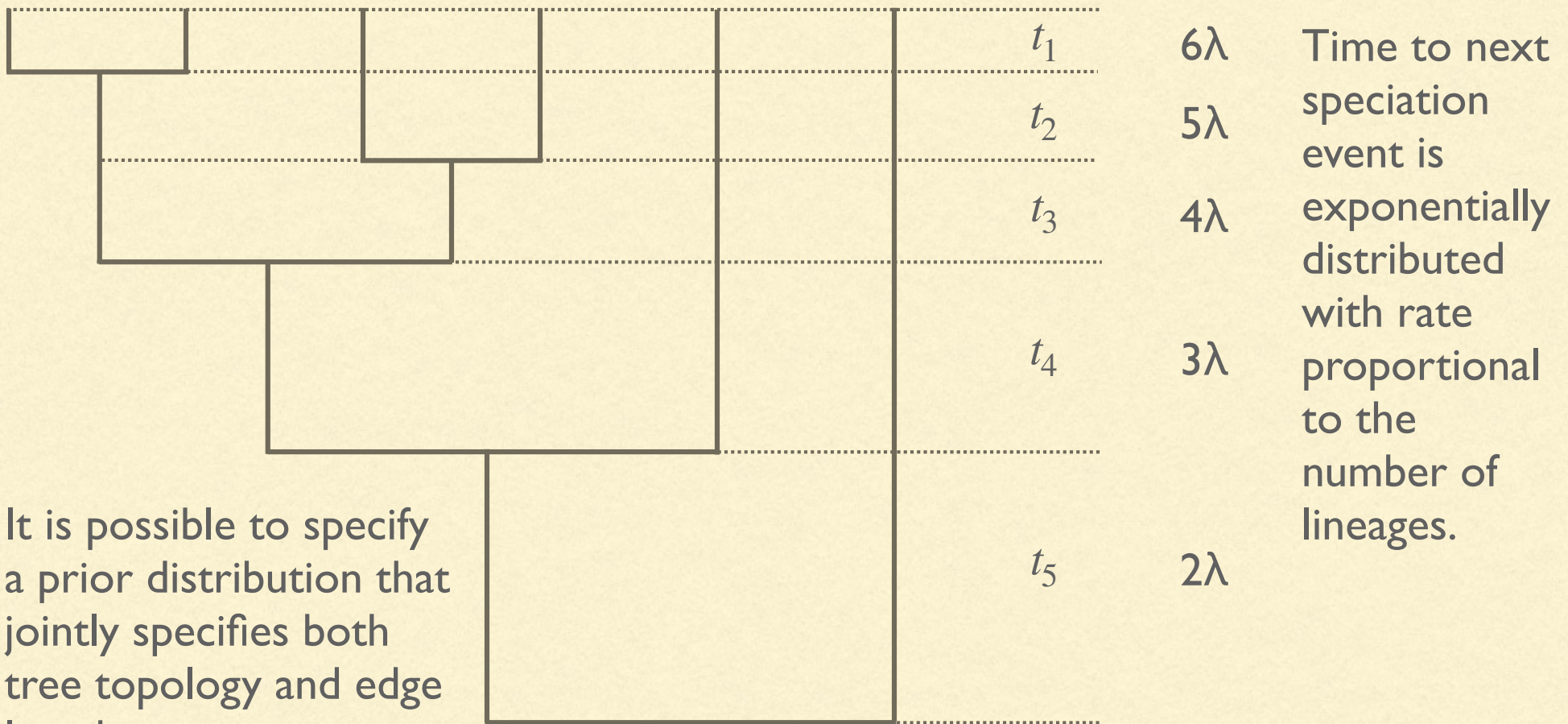


# Edge lengths: Gamma-Dirichlet





# Yule (pure birth) prior



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

---

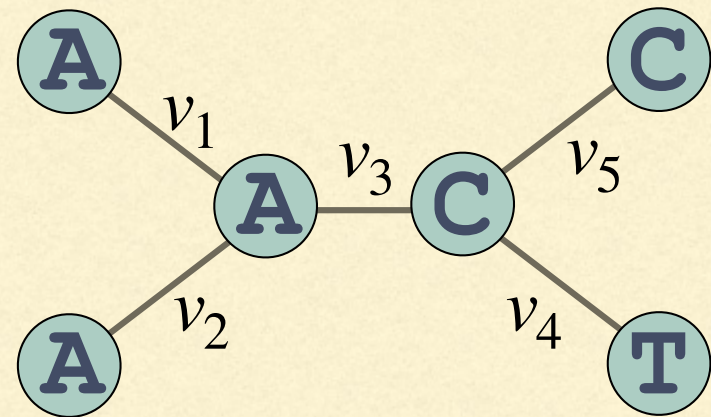
# Hierarchical models

---



# Non-hierarchical model

All model parameters can be found in the likelihood function



Prior mean 0.1

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

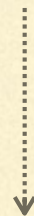
JC69 likelihood for site  $k$

# Hierarchical model

A prior for a hyperparameter is a hyperprior

$\mu$  is a hyperparameter

Hyperprior



Prior mean  $\mu$

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

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

JC69 likelihood for site  $k$



---

# Empirical Bayes

---

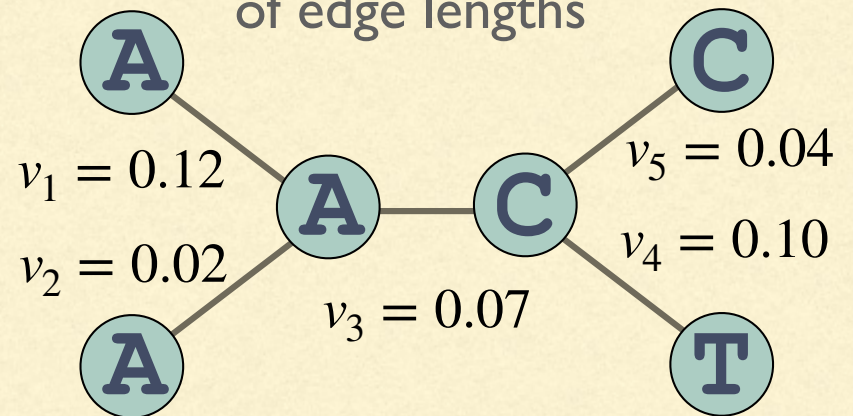
# Empirical Bayes approach

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  $k$

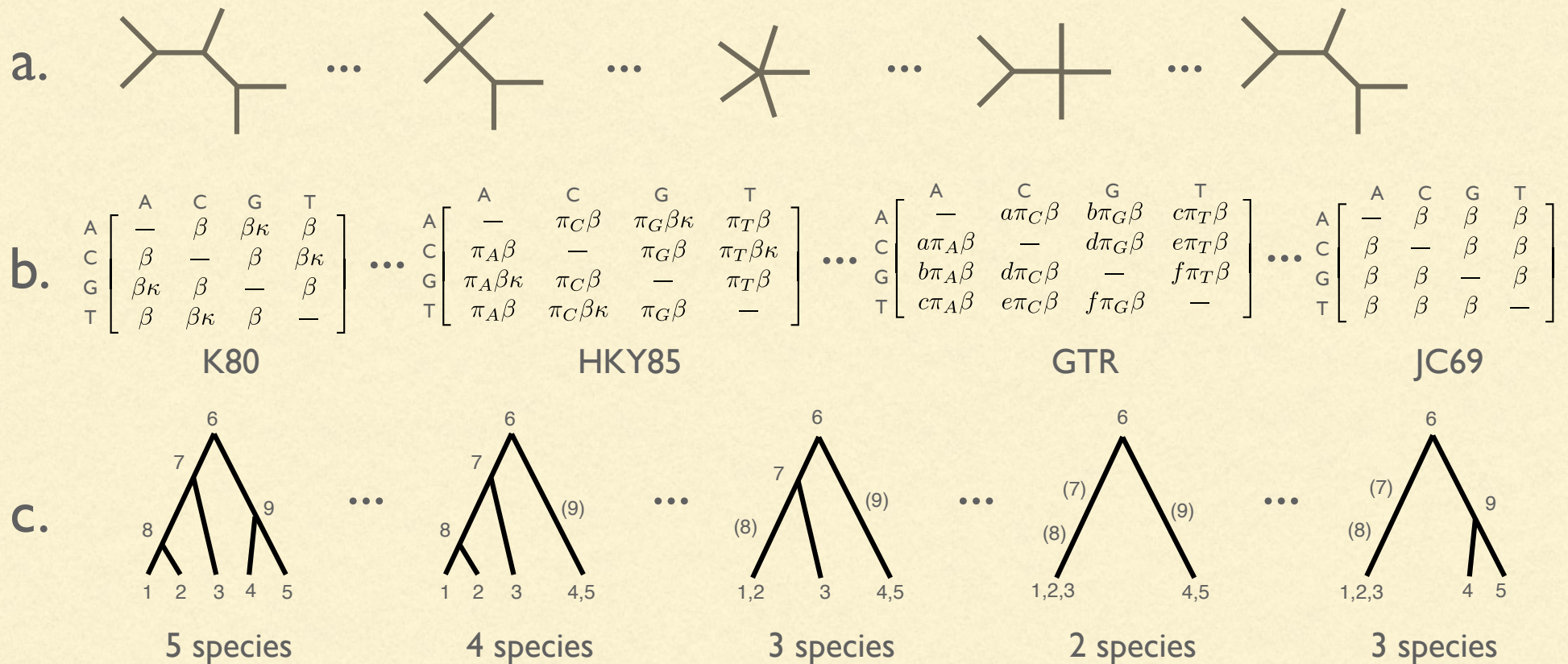


---

rjMCMC  
(reversible-jump MCMC)

---

# Examples of rjMCMC analyses



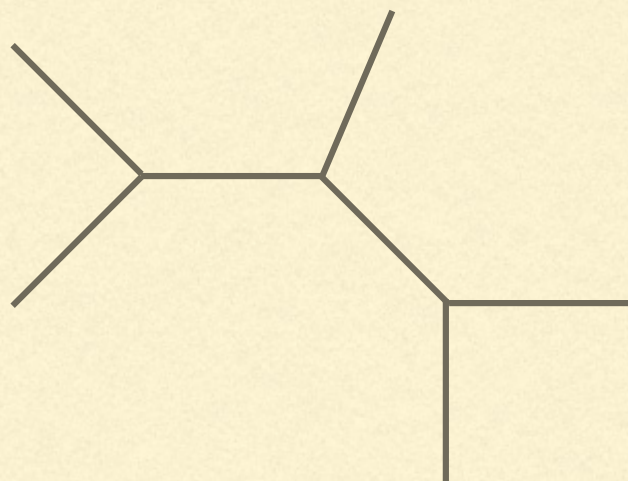
a. Lewis, Holder, & Holsinger. 2005. Systematic Biology 54:241–253.

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

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



# rjMCMC polytomy model



7 edge lengths

a. choose 1 of 2 edges  
to delete

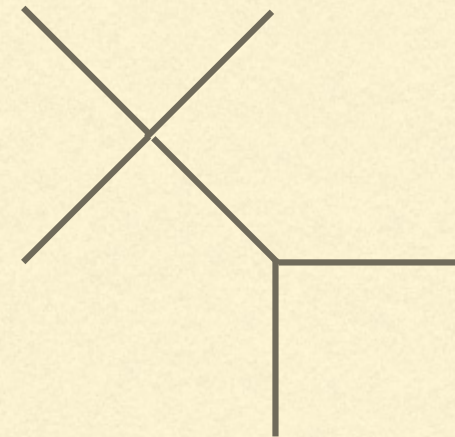


forward

reverse



a. choose 1 of 1 polytomies  
in which to insert an edge  
b. invent a new edge length  
to go with the new edge



6 edge lengths

The probability of accepting the forward proposal must take into account the proposal that exactly reverses it to ensure that at equilibrium the chain visits each model with the appropriate frequency

---

# Bayes' factors

---

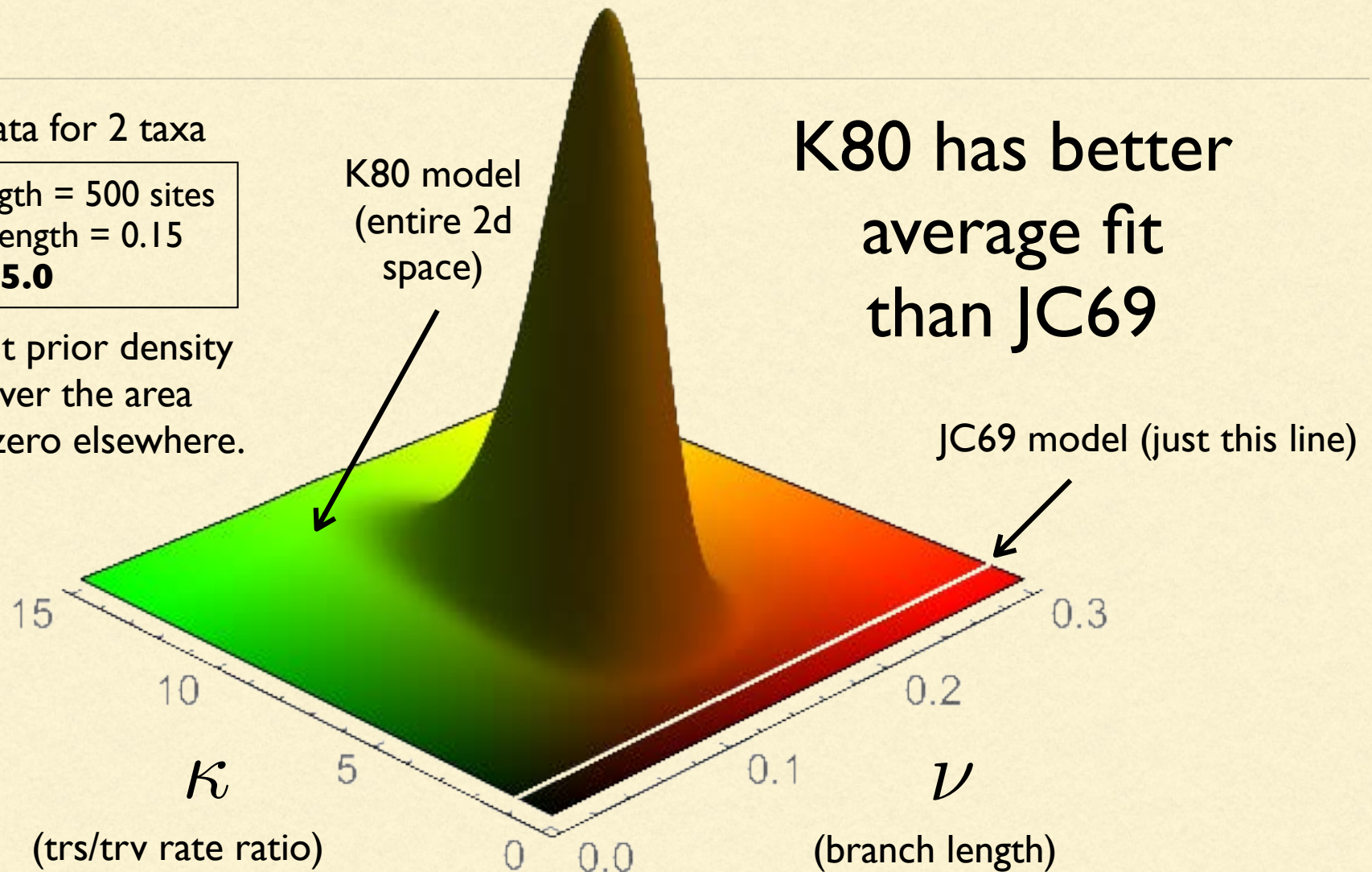


# Likelihood surface when K80 true

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.

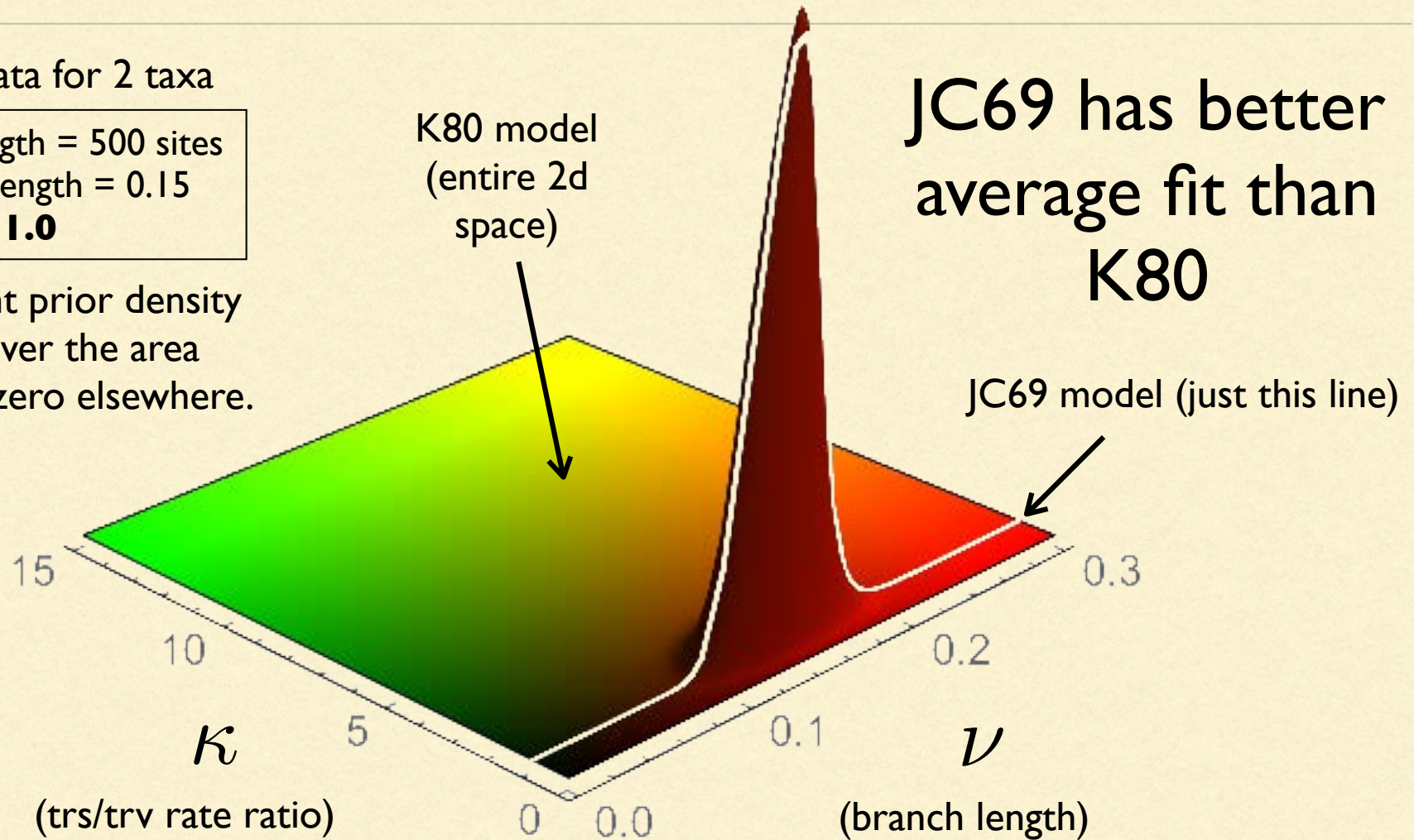


# Likelihood surface when JC69 true

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.





---

# Dirichlet process (DP) prior

---

all genes 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

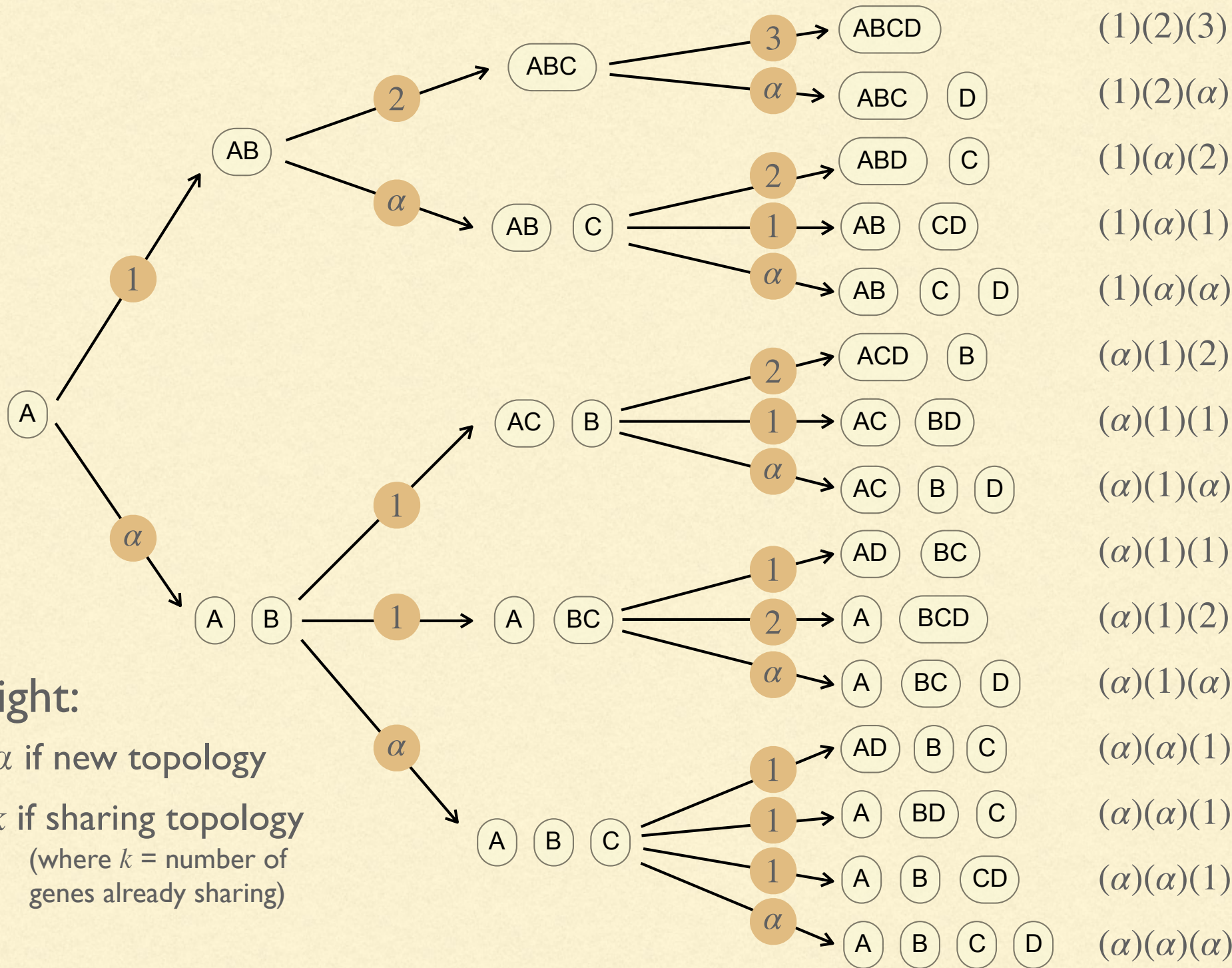
each gene has a different tree topology

A B C D

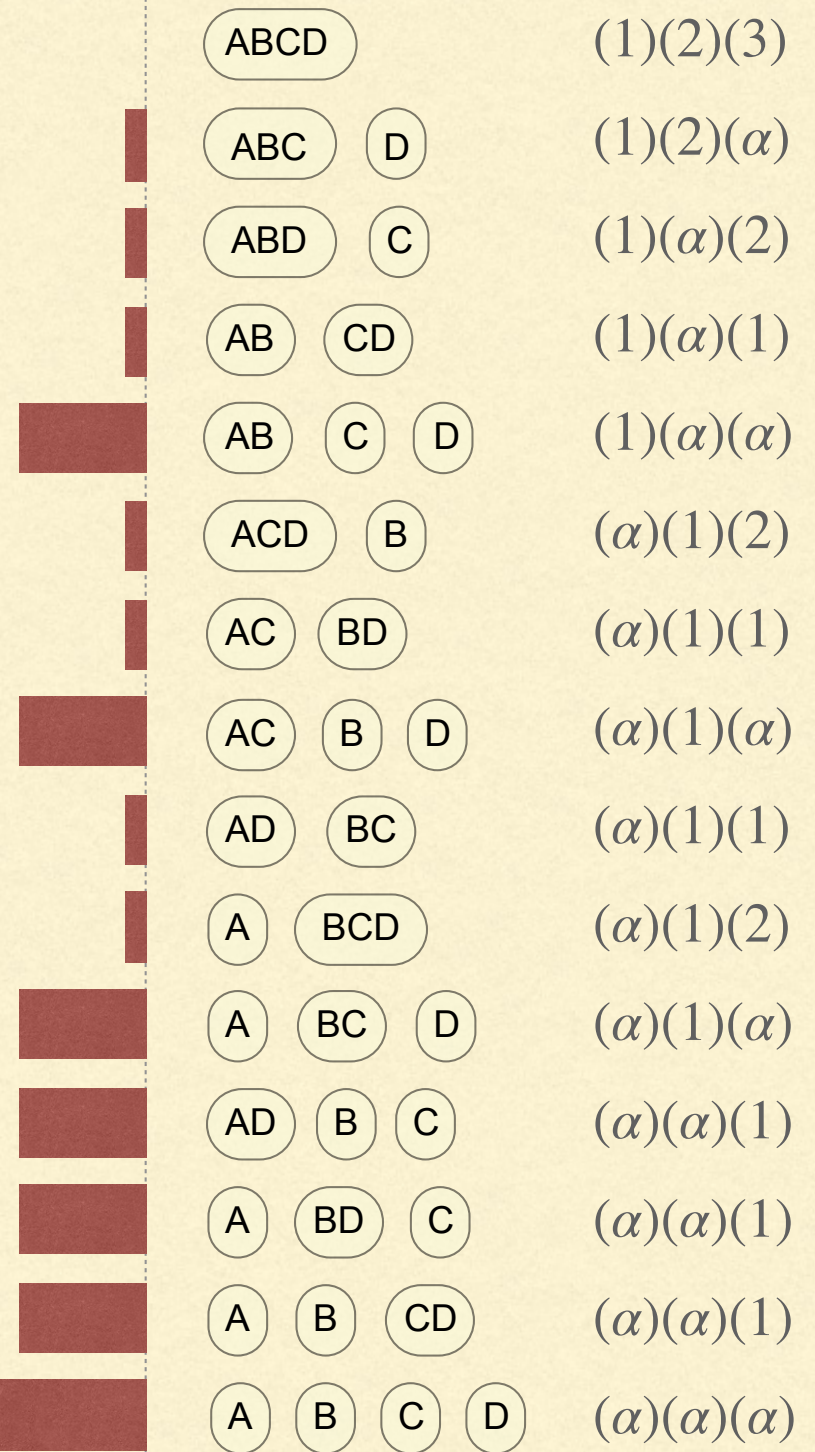
Suppose we have data for 4 genes:  
A, B, C, D

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





$\alpha = 10$   
 Large  $\alpha$  values favor  
 more, smaller groups

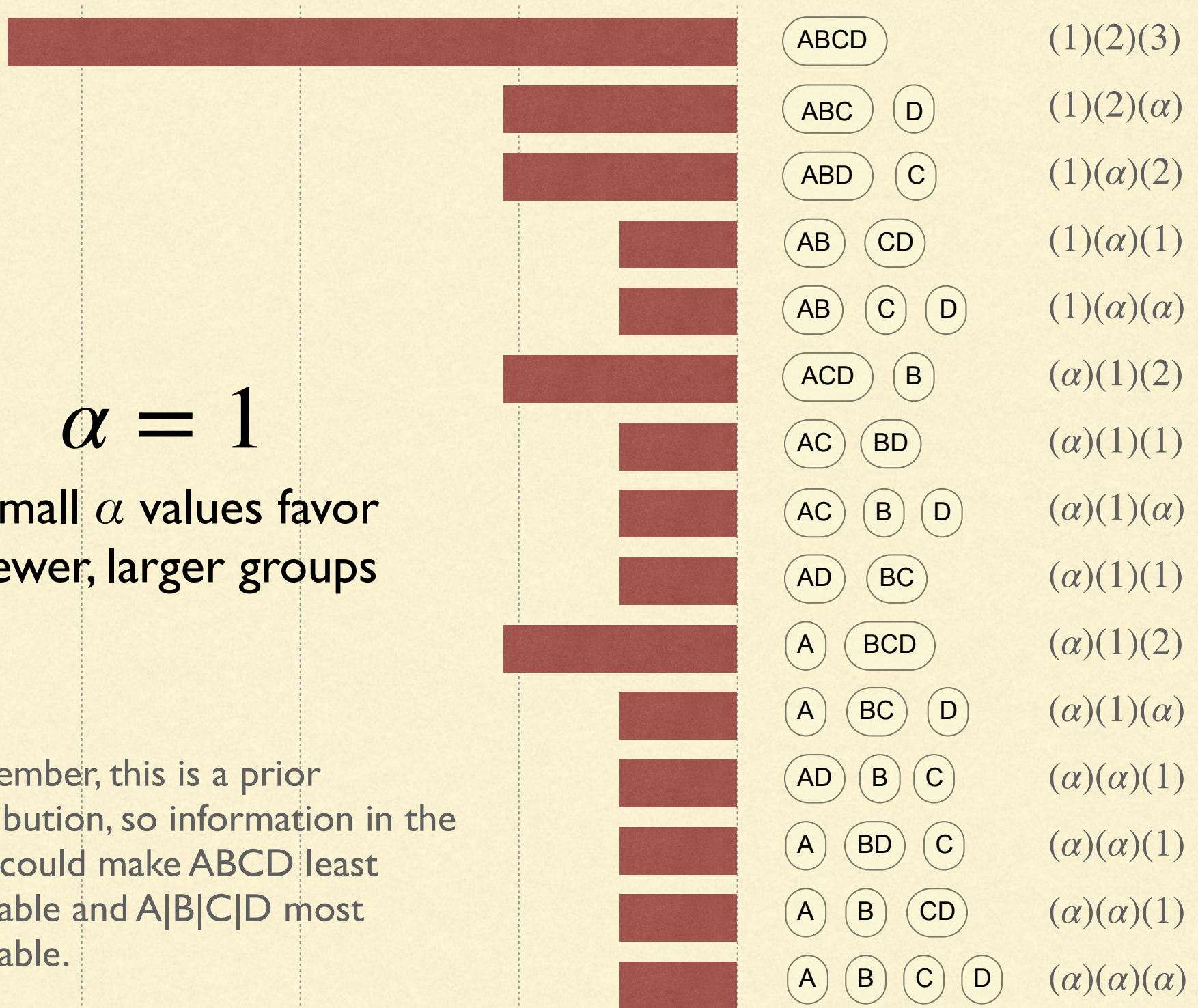




$$\alpha = 1$$

Small  $\alpha$  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.



## Dirichlet Process Prior applet

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