

See also 18-Apr-2018 and 16-May-2018 at  
<http://phyloseminar.org/recorded.html>

# Likelihood in Phylogenetics

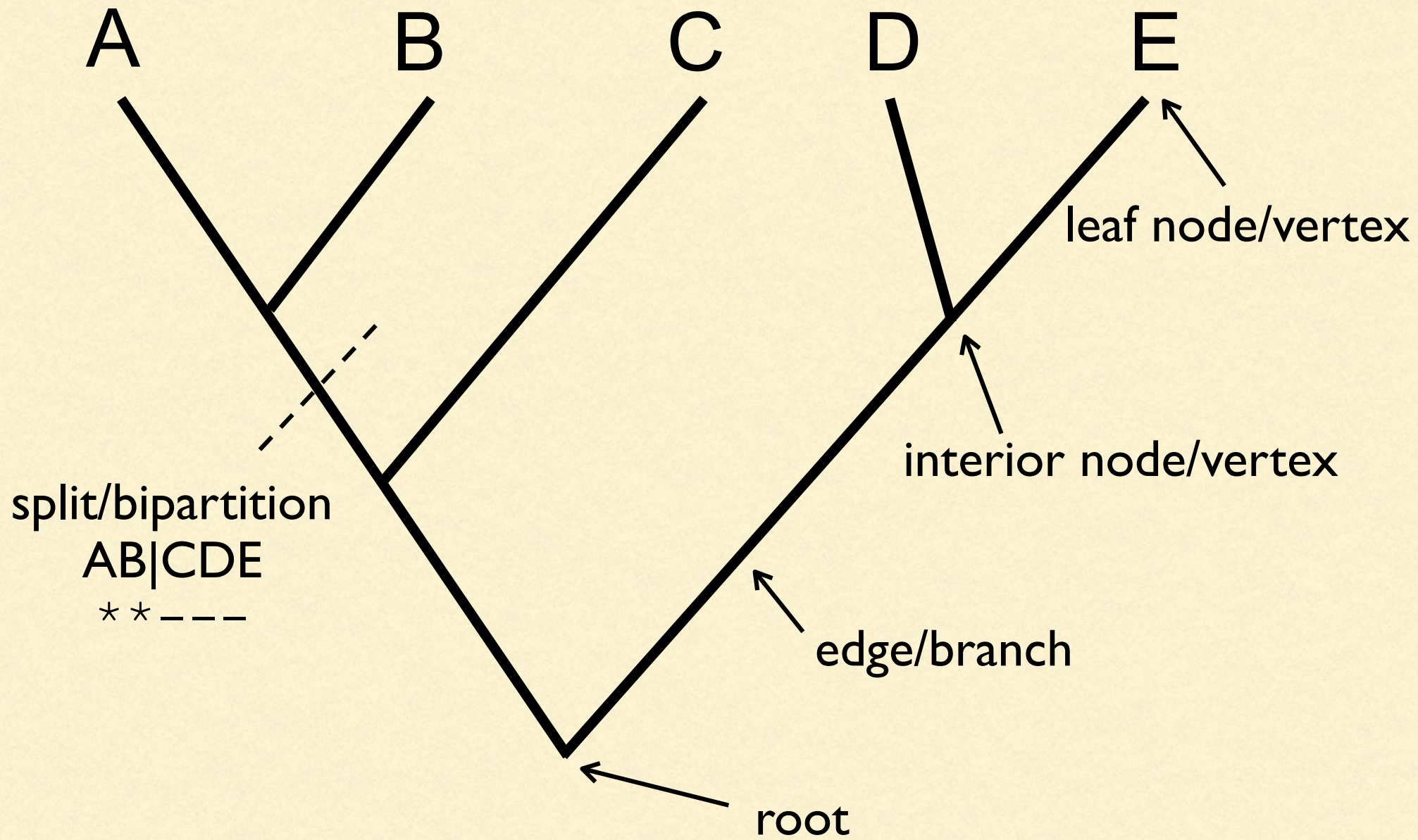
Workshop on Molecular Evolution  
Woods Hole, Massachusetts

2-3 Aug 2019

Paul O. Lewis

Department of Ecology & Evolutionary Biology







---

# Probability

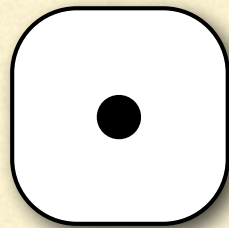
---

---

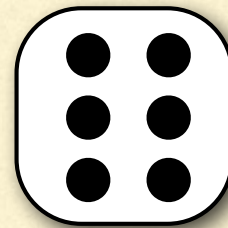
# Probabilities: the AND rule

---

Rolling 2 dice, what is the probability of seeing (simultaneously)  
a 1 on the first die and a 6 on the second die?



AND



$(1/6)$

$\times$

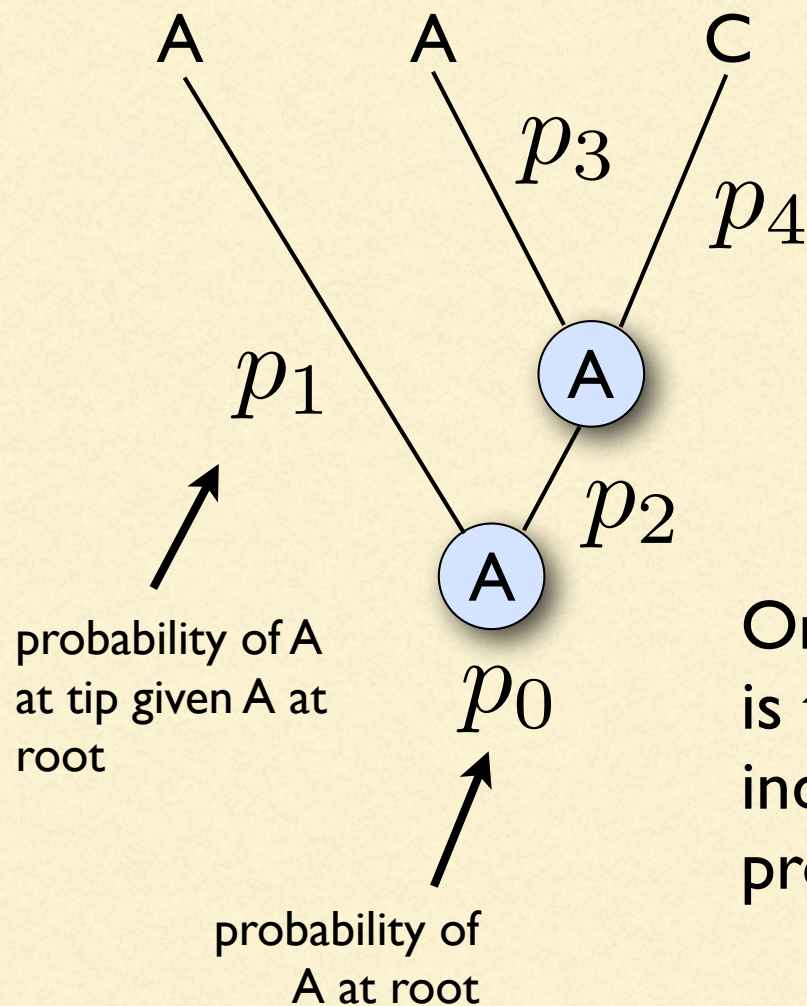
$(1/6)$

$=$

$1/36$



# AND rule in phylogenetics



we have observed these states

pretend (for now) we also know these ancestral states

$$\Pr(A, A, C, A, A) = p_0 p_1 p_2 p_3 p_4$$

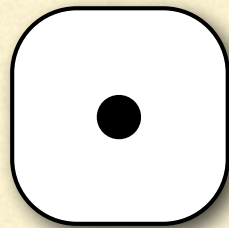
One use of the AND rule in phylogenetics is to combine probabilities associated with individual branches to produce the overall probability of the data for one site.

---

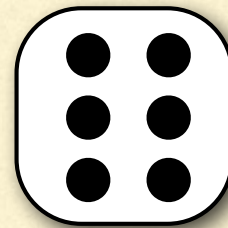
# Probabilities: the OR rule

---

Rolling 1 die, what is the probability of seeing either a 1 or a 6?



OR



$(1/6)$

+

$(1/6)$

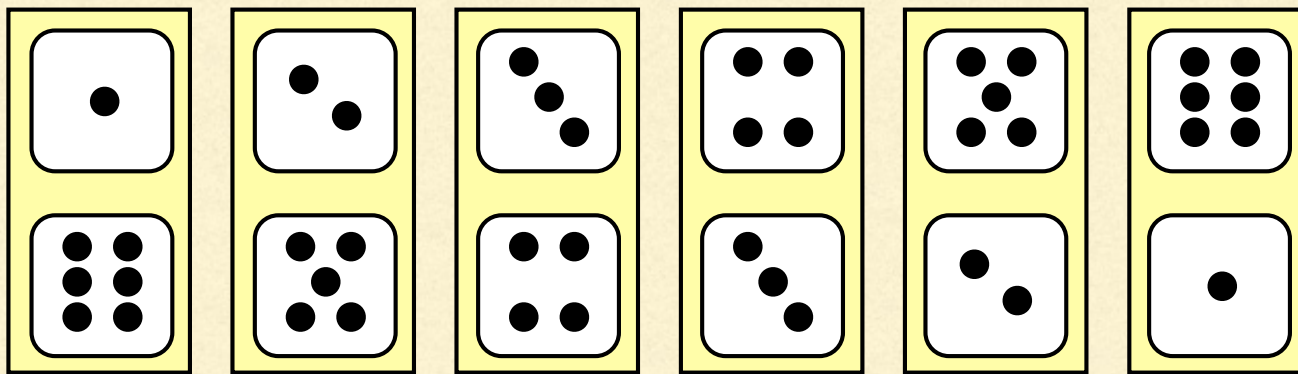
=

$1/3$



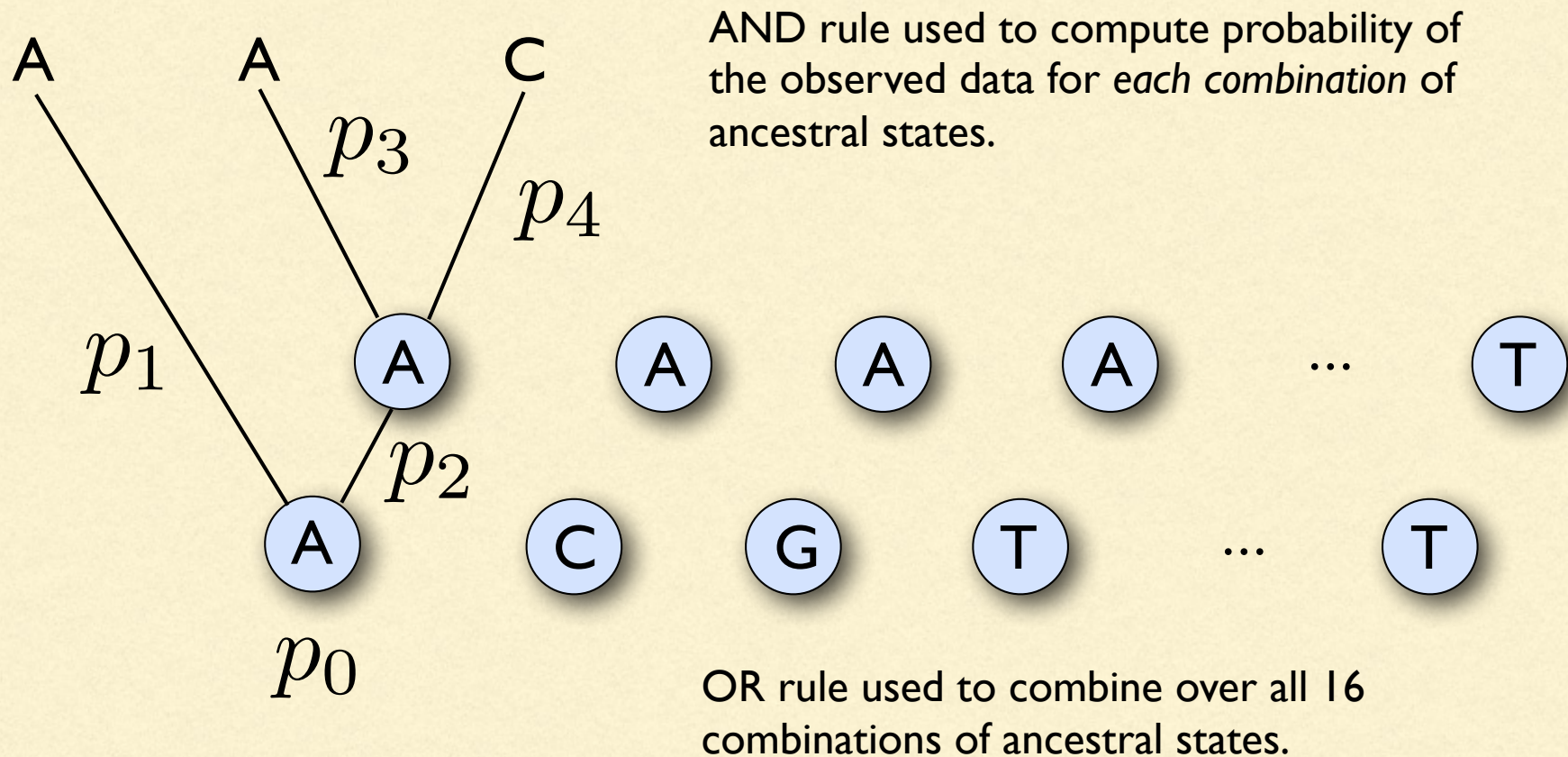
# Combining AND and OR

What is the probability that the sum of two dice is 7?



$$(1/36) + (1/36) + (1/36) + (1/36) + (1/36) + (1/36) = 1/6$$

# Using both AND and OR in phylogenetics



$$\Pr(\mathbf{A}, \mathbf{A}, \mathbf{C}) = \Pr(\mathbf{A}, \mathbf{A}, \mathbf{C}, \mathbf{A}, \mathbf{A}) + \Pr(\mathbf{A}, \mathbf{A}, \mathbf{C}, \mathbf{A}, \mathbf{C}) + \dots + \Pr(\mathbf{A}, \mathbf{A}, \mathbf{C}, \mathbf{T}, \mathbf{T})$$



---

# Independence

---

$$\Pr(A, B) = \Pr(A) \Pr(B)$$

Probability of flipping a coin twice and  
getting heads both times:

$$\Pr(H, H) = \Pr(H) \Pr(H)$$

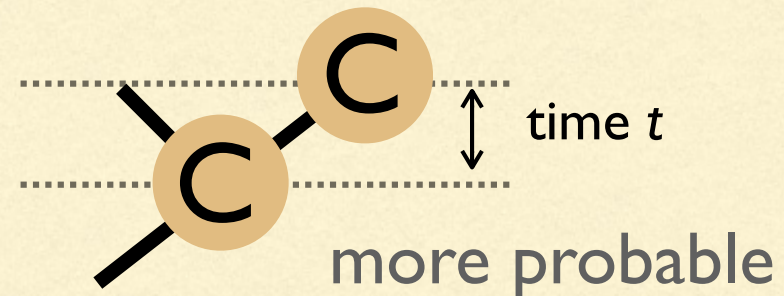
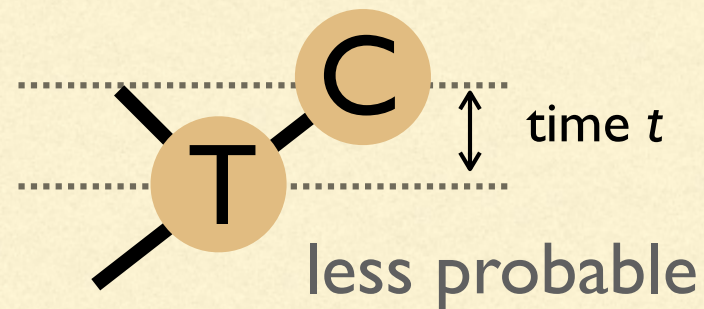




# Non-independence in phylogenies

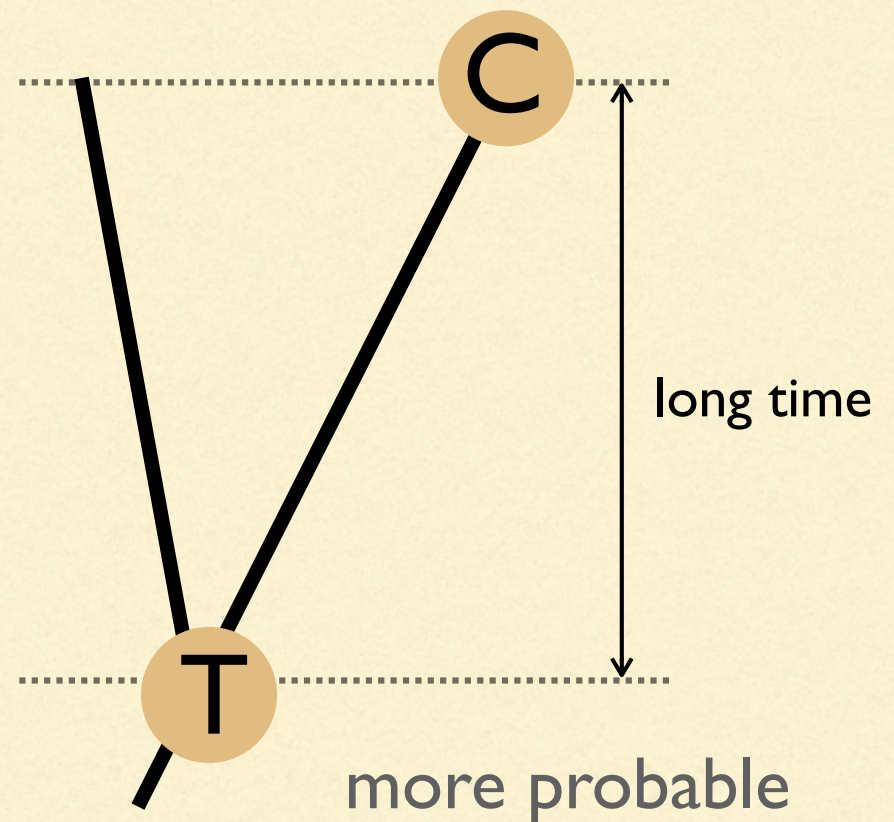
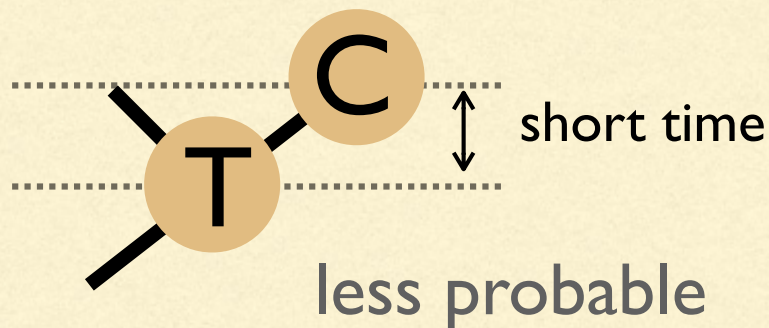
Normally, for a given rate of substitution and time, the probability of the end state is *dependent* on the starting state

$$\Pr(C|C) > \Pr(C|T)$$



# Non-independence in phylogenies

For a given rate of substitution and starting state, the probability of the end state is *dependent* on time



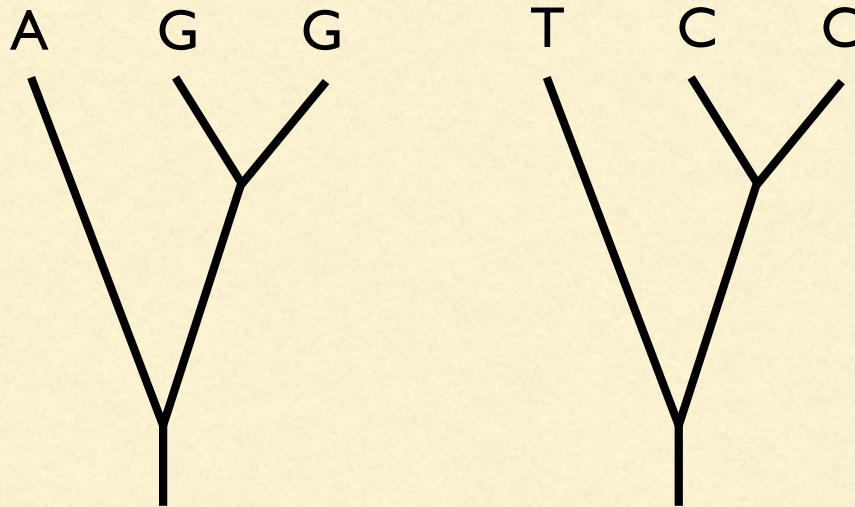


---

# Conditional Independence

---

$$\Pr(A,B|C) = \Pr(A|C) \Pr(B|C)$$



$$\Pr(\text{AGG}, \text{TCC} | \text{tree}) = \Pr(\text{AGG} | \text{tree}) \Pr(\text{TCC} | \text{tree})$$

---

# Likelihood

---

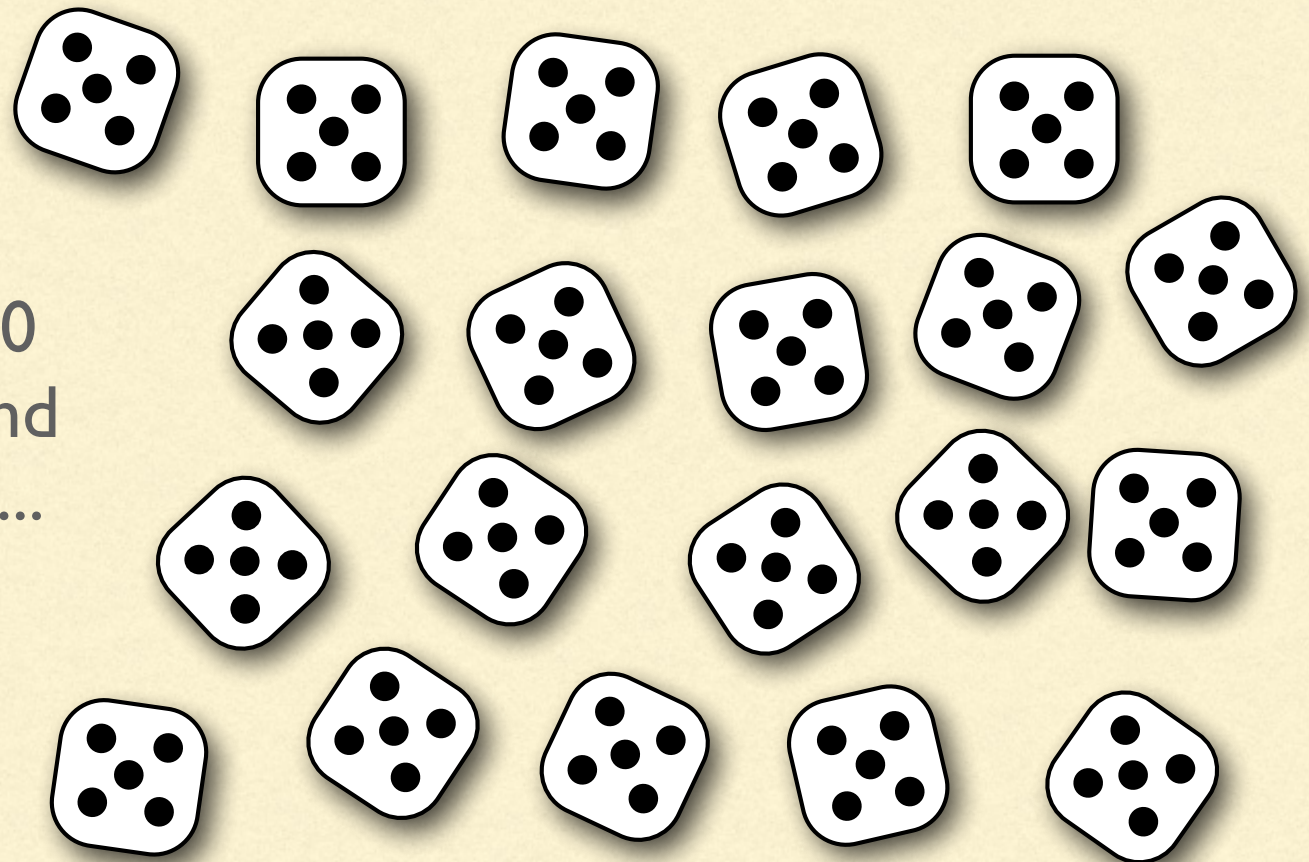


---

# The Likelihood Criterion

---

Suppose I threw 20  
dice on the table and  
this was the result...

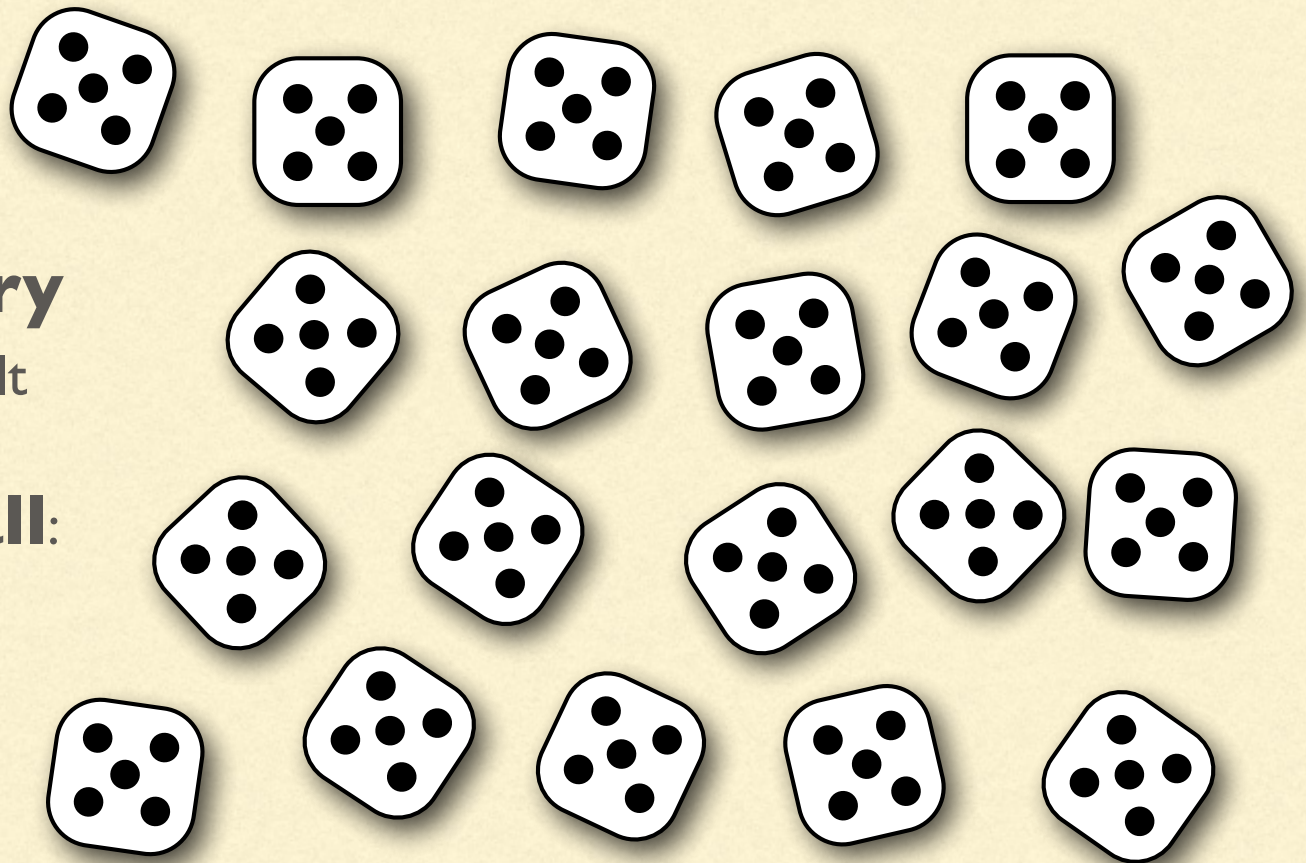


---

# The Fair Dice model

---

You should have been **very surprised** at this result because the probability of this event is **very small**: only 1 in 3.6 quadrillion!





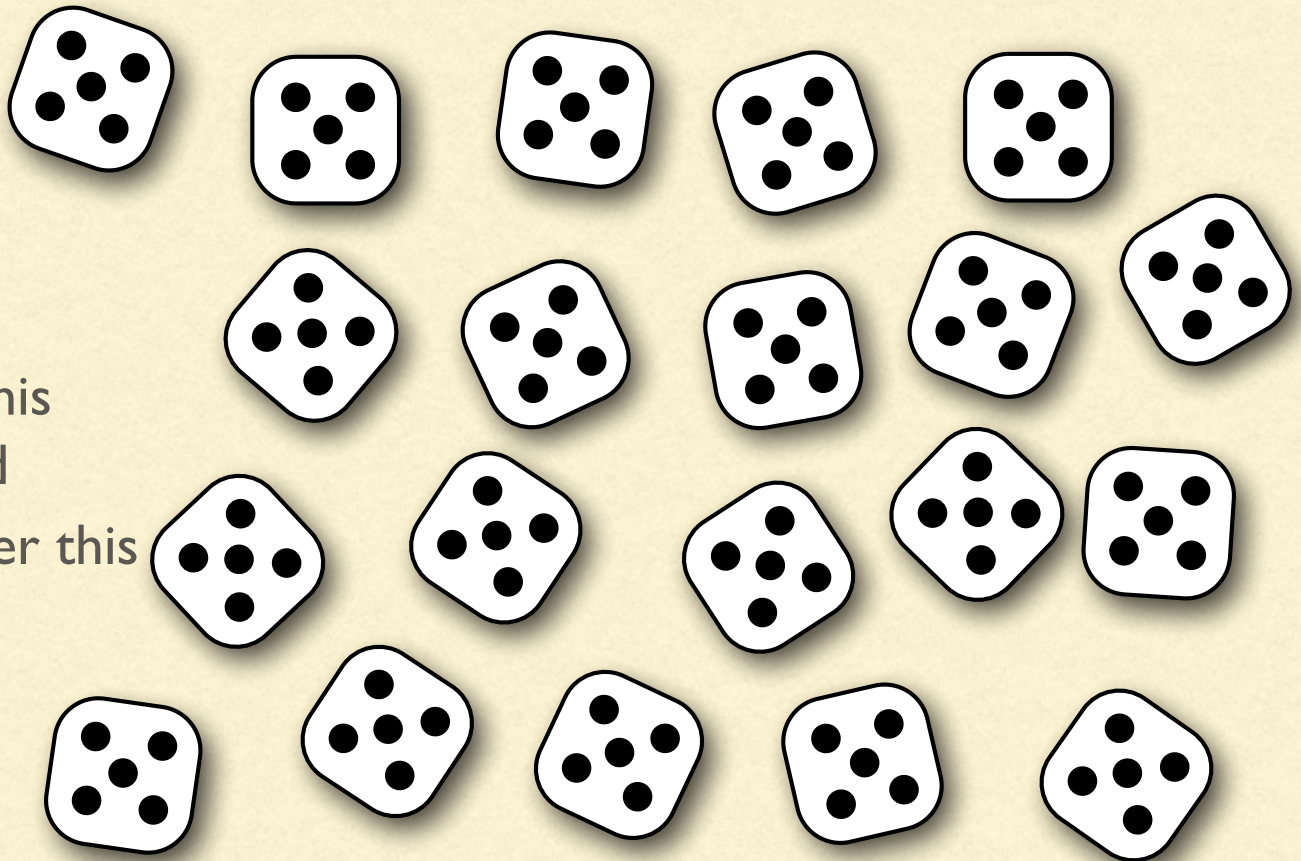
---

# The Trick Dice model

(assumes dice each have 5 on every side)


---

You should **not be surprised at all** at this result because the observed outcome is **certain** under this model.



# What's changed? (the model)

The **winning model** makes us **least surprised** at the data we've observed



Model	Likelihood	Surprise level
Fair Dice	$\frac{1}{3,656,158,440,062,976}$	Very, very, <b>very</b> surprised
→ Trick Dice	1	Not surprised at all



# Why do we need the term **likelihood**?

Outcome	Fair coin model	Two-heads model
H	0.5	1
T	0.5	0

Likelihoods of models  
given one particular data  
set are not expected to  
sum to 1.0

Probabilities of data  
outcomes given one  
particular model sum to  
1.0

Probability of the data  
given the model  
  
Likelihood of the model  
given the data

# Likelihood of a single vertex

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

● **GAAGTCCTTGAGAAATAAACTGCACACACTGG**

$$L = \Pr(G) \Pr(A) \Pr(A) \Pr(G) \Pr(T) \cdots \Pr(G)$$

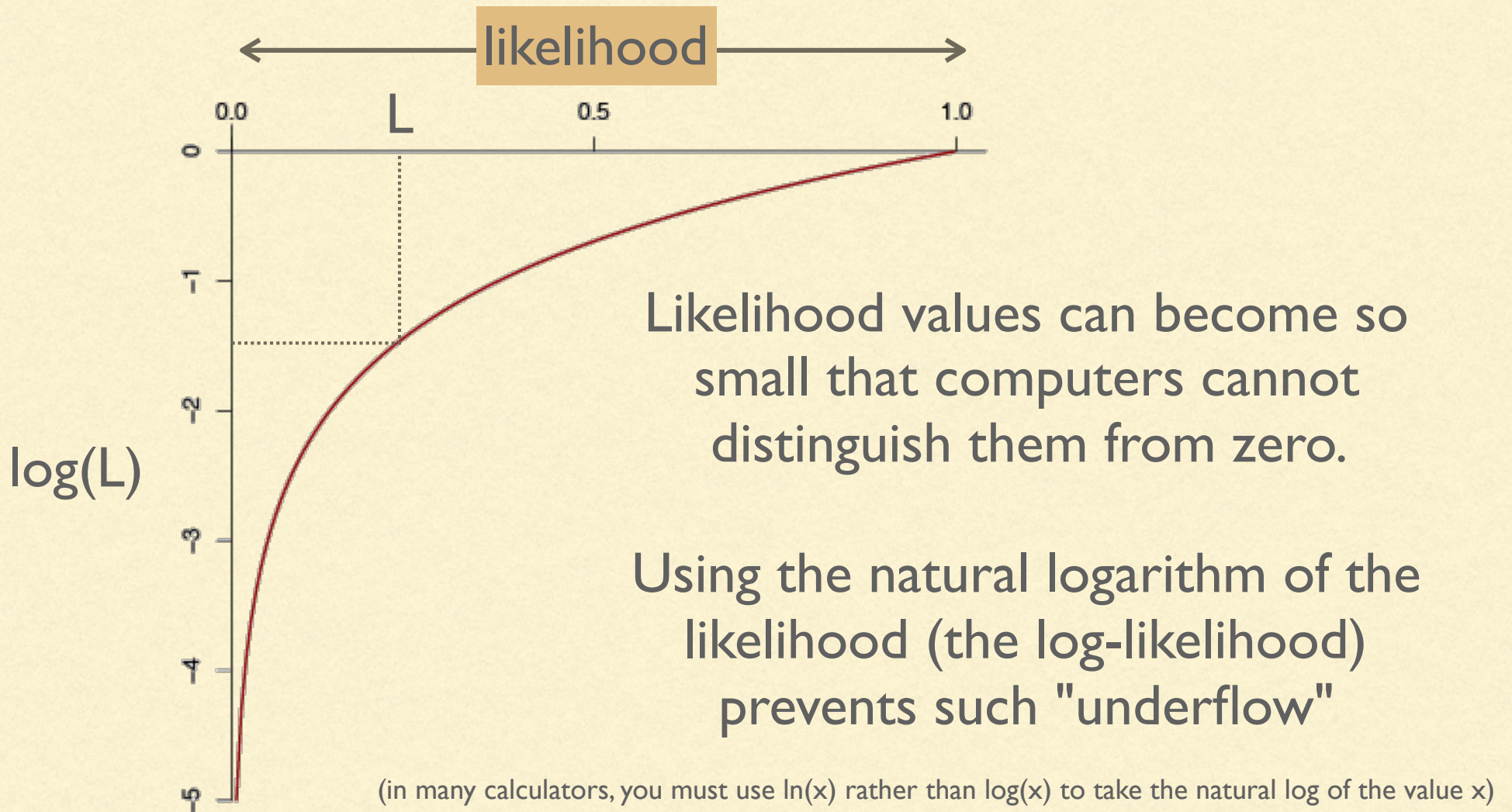
$$L = \pi_G \pi_A \pi_A \pi_G \pi_T \cdots \pi_G$$

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

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

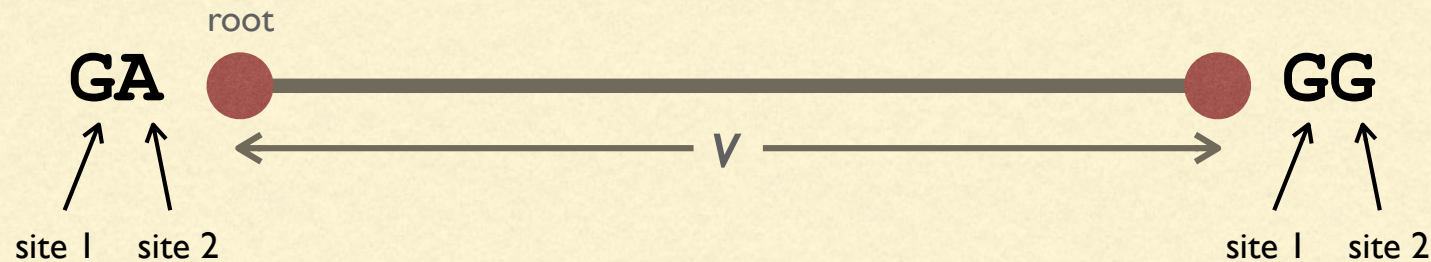


# Natural logarithm



# Likelihood of a single-edge tree

Two nodes have sequence data (but only for two sites)



$$L = \left[ \left( \frac{1}{4} \right) \left( \frac{1}{4} + \frac{3}{4} e^{-4v/3} \right) \right] \left[ \left( \frac{1}{4} \right) \left( \frac{1}{4} - \frac{1}{4} e^{-4v/3} \right) \right]$$

$\longleftrightarrow$  site 1  $\longleftrightarrow$  site 2  $\longleftrightarrow$

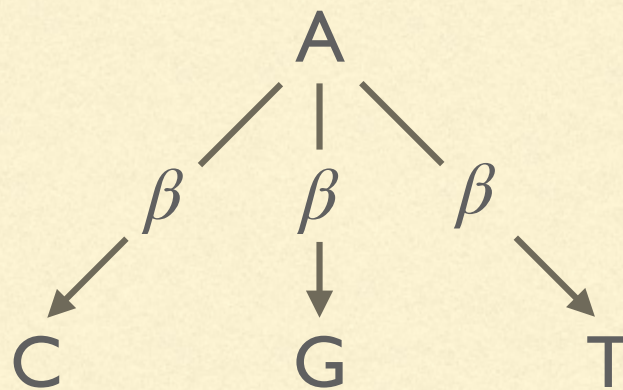
Each **site likelihood** is the probability of the **starting state** at the root (1/4) times the **transition probability** (probability of the end state given the starting state)



# What is the edge length $v$ ?

**number** of  
substitutions  
per site = substitution  
**rate** per site  $\times$  **time**

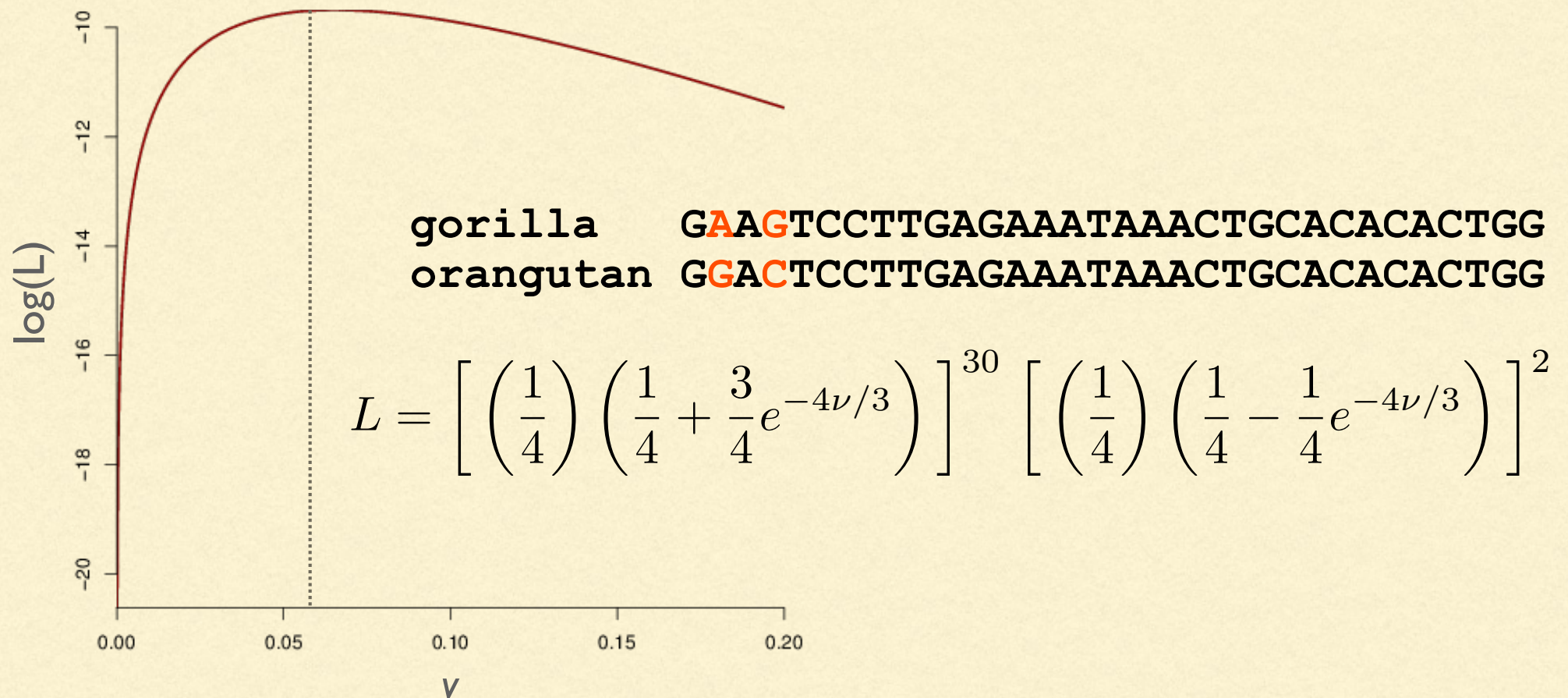
$$v = 3\beta \times t$$



3 possible substitutions, each of which happens with rate  $\beta$

# Maximum likelihood estimation

0.065 is the maximum likelihood estimate (MLE) of  $\nu$





---

# Transition probabilities

---

$$\frac{1}{4} + \frac{3}{4}e^{-4\nu/3}$$

same state

$$\frac{1}{4} - \frac{1}{4}e^{-4\nu/3}$$

different states

Conditional probability of end state given  
starting state and edge length

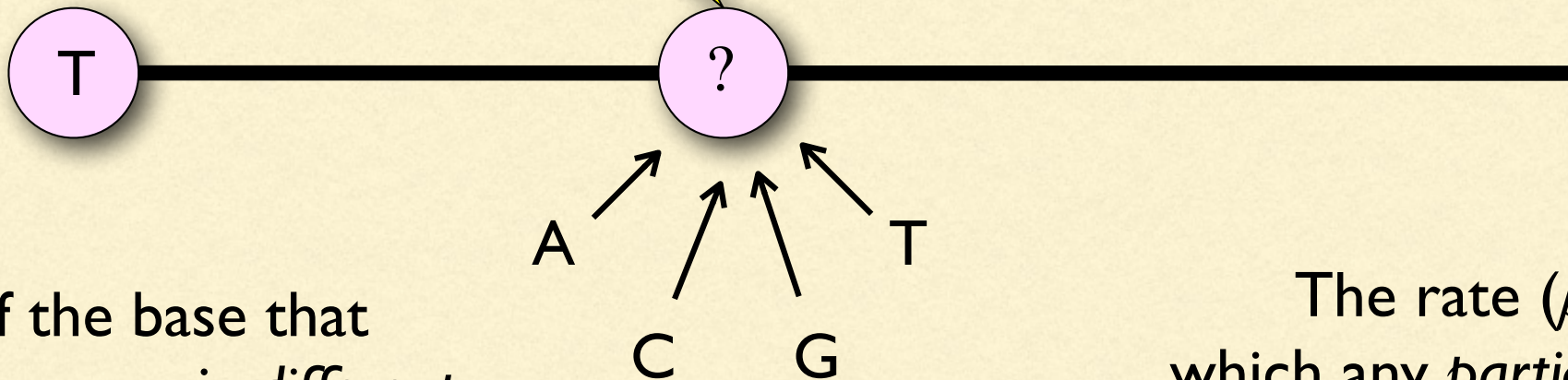
How do these formulas arise?

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pages 21-132 in H. N. Munro (ed.), Mammalian Protein Metabolism. Academic Press, New York.

# "ACHNyons" vs. substitutions

I made up this term  
(Anything Can  
Happen Now)

When an **achnyon** occurs, *any* base can appear in a sequence.



If the base that appears is *different* from the base that was already there, then a **substitution** event has occurred.

The rate ( $\beta$ ) at which any *particular* substitution occurs will be 1/4 the achnyon rate ( $\mu$ ). That is,  $\beta = \mu/4$  (or  $\mu = 4\beta$ )



---

# Deriving a transition probability

---

Calculate the probability that a site currently T will change to G over time  $t$  when the rate of this particular substitution is  $\beta$ :

$$\text{Pr}(0 \text{ achnyons}) = e^{-\mu t} \quad (\text{Poisson probability of zero events})$$

$$\text{Pr}(\text{at least 1 achnyon}) = 1 - e^{-\mu t}$$

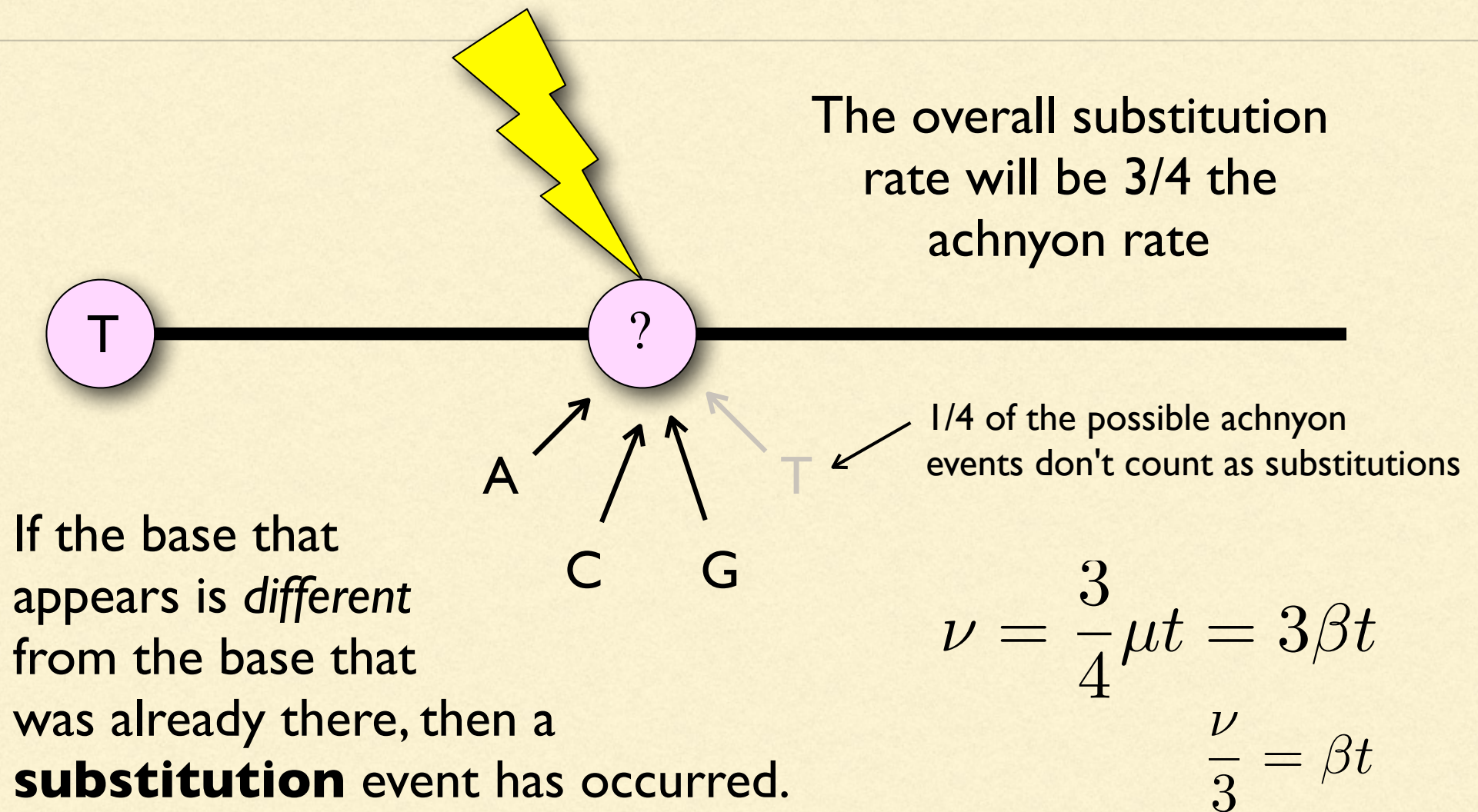
$$\text{Pr}(\text{last achnyon results in base G}) = \frac{1}{4}$$

$$\text{Pr}(\text{end in G} \mid \text{start in T}) = \frac{1}{4} (1 - e^{-\mu t})$$

Remember that the achnyon rate ( $\mu$ ) is 4 times the rate ( $\beta$ ) of any *particular* substitution:

$$P_{TG}(t) = \frac{1}{4} (1 - e^{-4\beta t})$$

# Expected number of substitutions





---

# Deriving a transition probability

---

Calculate the probability that a site currently T will change to G over time  $t$  when the rate of this particular substitution is  $\beta$ :

⋮

Remember that the achnyon rate ( $\mu$ ) is 4 times the rate ( $\beta$ ) of any *particular* substitution:

$$P_{TG}(t) = \frac{1}{4} (1 - e^{-4\beta t})$$

Substitute  $v/3$  for  $\beta t$ :

$$P_{TG}(t) = \frac{1}{4} \left( 1 - e^{-4v/3} \right)$$

# Transition Probabilities: Remarks

$$\left. \begin{aligned} P_{TA}(t) &= \frac{1}{4} \left( 1 - e^{-4\nu/3} \right) \\ P_{TC}(t) &= \frac{1}{4} \left( 1 - e^{-4\nu/3} \right) \\ P_{TG}(t) &= \frac{1}{4} \left( 1 - e^{-4\nu/3} \right) \\ P_{TT}(t) &= \frac{1}{4} \left( 1 - e^{-4\nu/3} \right) \end{aligned} \right\} \begin{array}{l} \text{These should add to} \\ \text{1.0 because } T \text{ *must*} \\ \text{change to something!} \end{array}$$

---

$$1 - e^{-4\nu/3}$$

Doh! Something must be wrong here...



---

# Transition Probabilities: Remarks

---

$$P_{TA}(t) = \frac{1}{4} \left( 1 - e^{-4\nu/3} \right)$$

$$P_{TC}(t) = \frac{1}{4} \left( 1 - e^{-4\nu/3} \right)$$

$$P_{TG}(t) = \frac{1}{4} \left( 1 - e^{-4\nu/3} \right)$$

$$P_{TT}(t) = \frac{1}{4} \left( 1 - e^{-4\nu/3} \right) + e^{-4\nu/3}$$

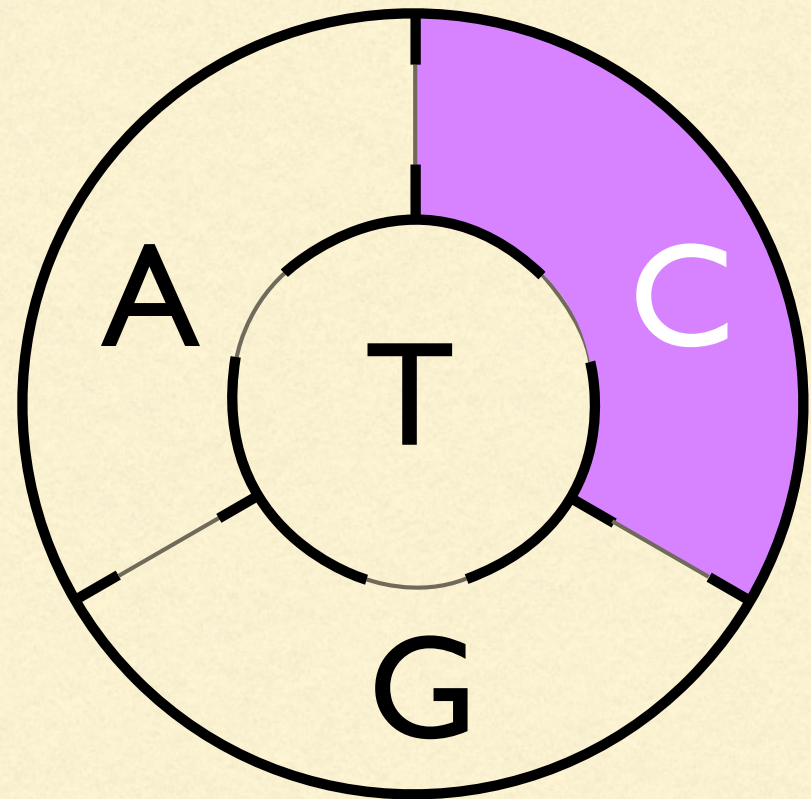
I forgot to account for the possibility of *no* achnyons over time  $t$

# Equilibrium Frequencies

Imagine a bottle of perfume has been spilled in room C.

The doors to the other rooms are closed, so the perfume has, thus far, not been able to spread.

What would happen if we opened all the doors?



Architect: Joe Bielawski



# Equilibrium Frequencies

At the **instant the doors open**, perfume molecules...

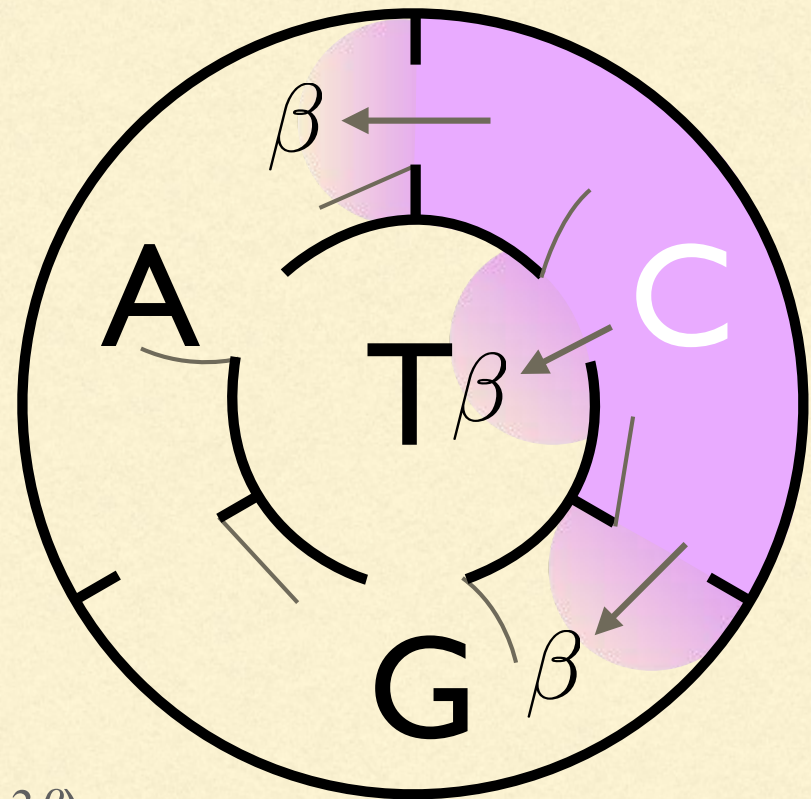
enter room A at rate  $\beta$

enter room T at rate  $\beta$

enter room G at rate  $\beta$

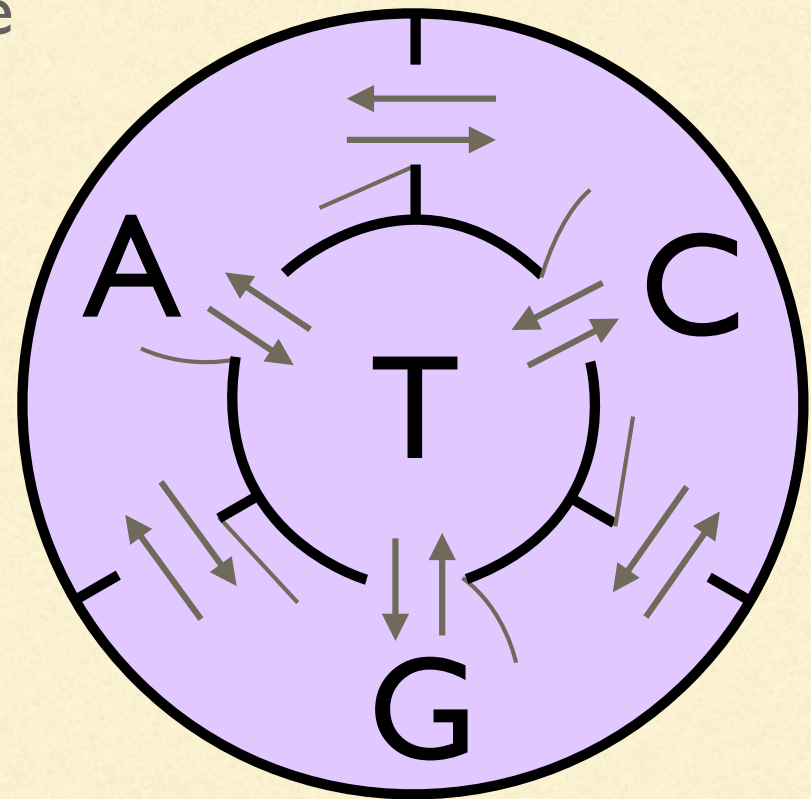
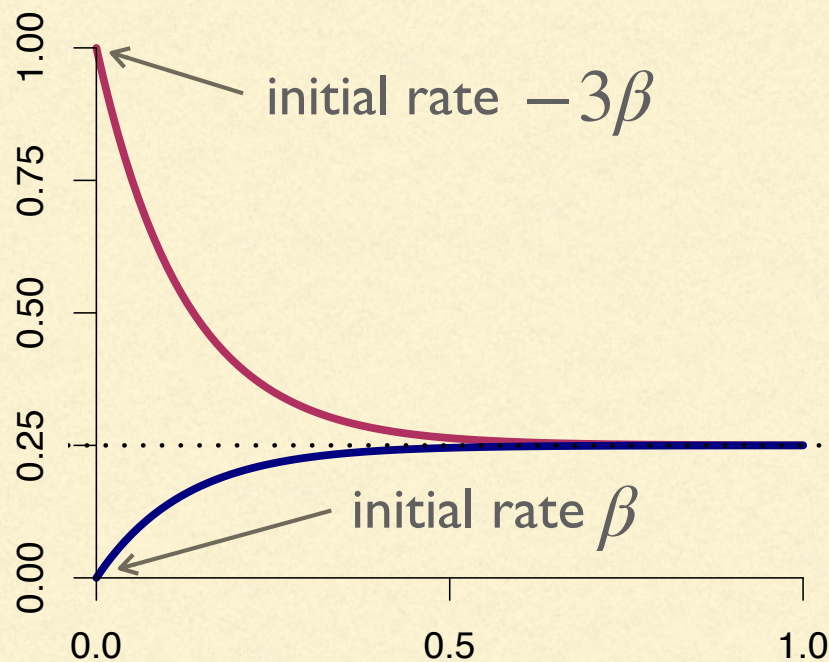
enter room C at rate  $-3\beta$

(you could also say they *leave* C at rate  $3\beta$ )



# Equilibrium Frequencies

At **equilibrium**, the relative concentration of perfume is **equal** in all rooms



$$\pi_A = \pi_C = \pi_G = \pi_T = \frac{1}{4}$$



---

# Transition probability demo

---

<https://phylogeny.uconn.edu/transition-probability/>

---

# Sequence data for four taxa

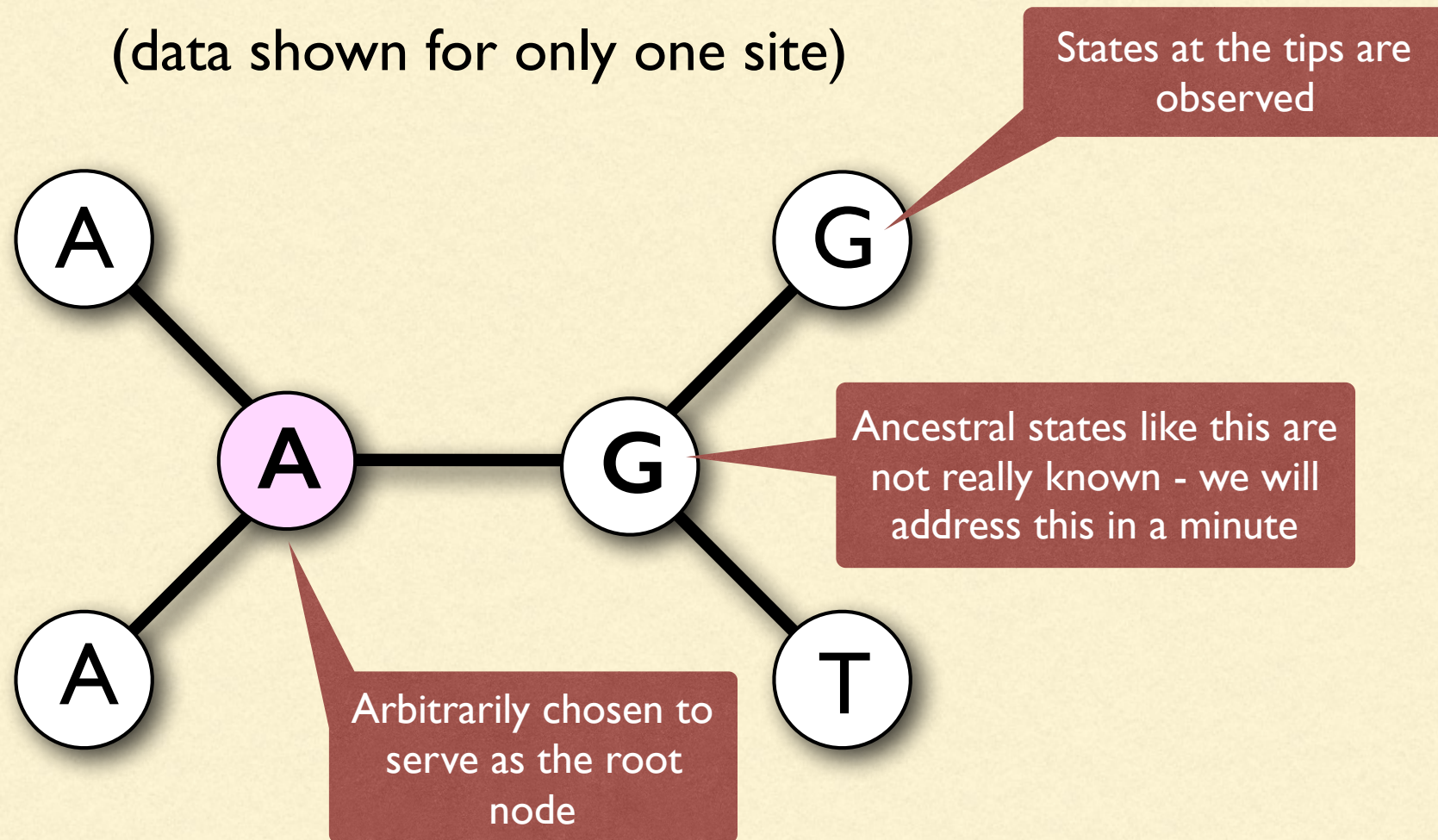
---

one site

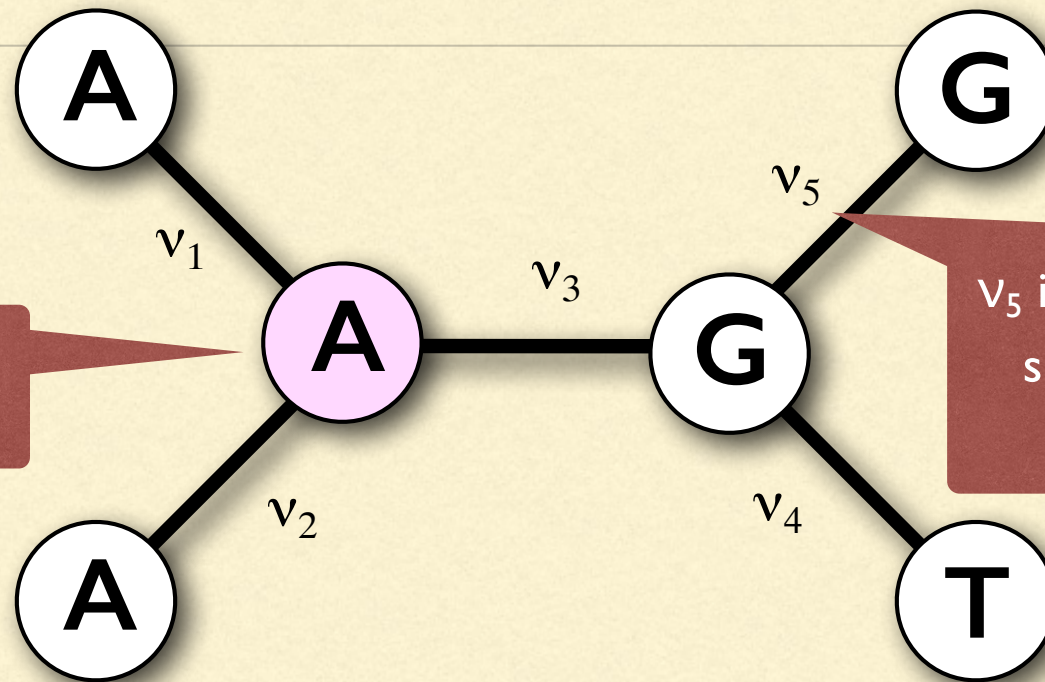
Sphagnum	GGCAGCATTTCGAATGACTCCTCAACCTGGAGT	A	TCACCCG...
Asplenium	GGCAGCTTTCCGGATGACCCCAACAACCCGGAGT	A	TCAGCTG...
Picea	GGCAGCATTCCGAGTAACTCCTCAACCAGGGGT	G	TCGCCCG...
Avena	GGCAGCATTCCGAGTAACTCCTCAACCTGGGGT	T	TCGCCGG...



# Likelihood for tree (one site)



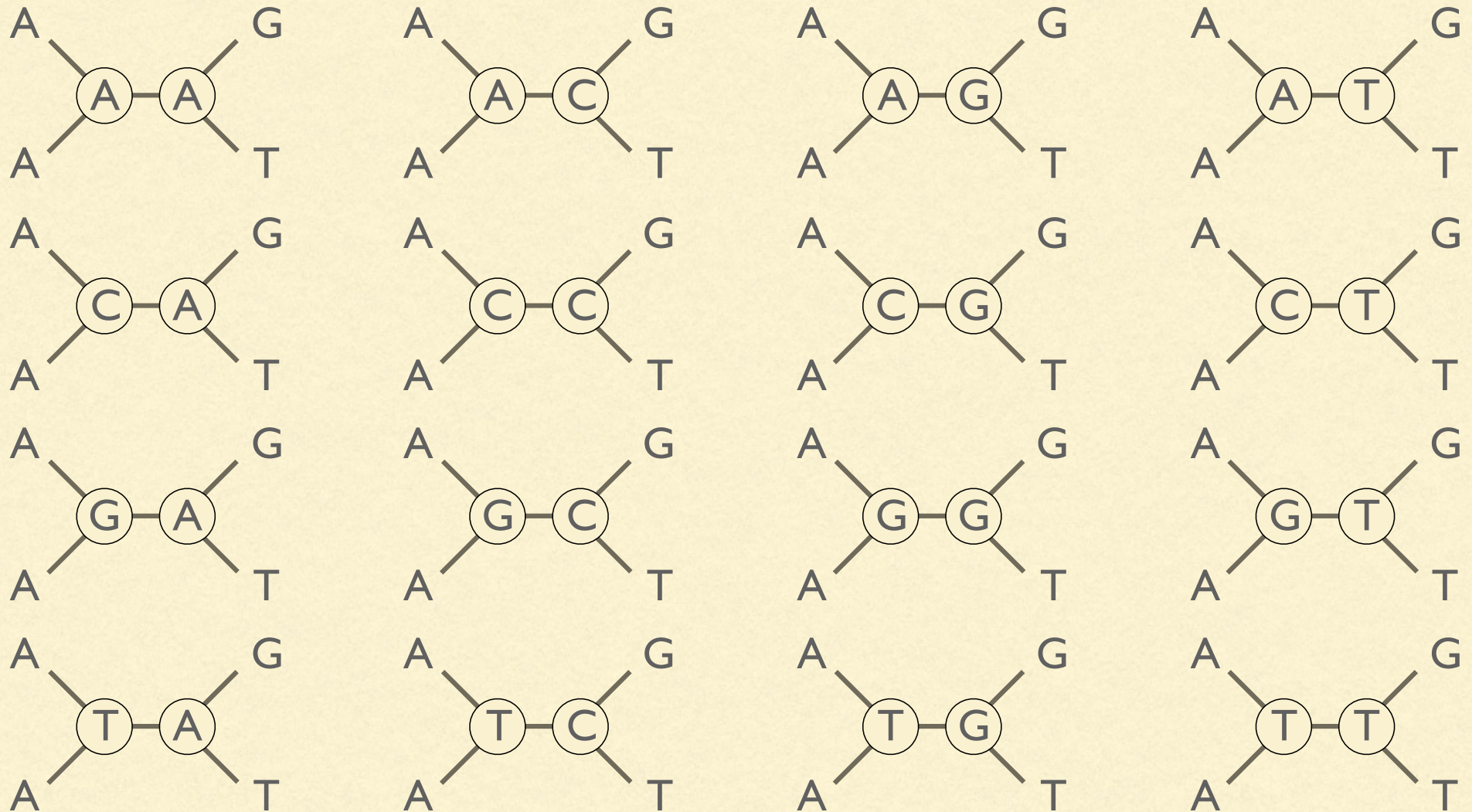
# Likelihood for tree (one site)



$$L = \frac{1}{4} \left[ \frac{1}{4} + \frac{3}{4} e^{-4\nu_1/3} \right] \left[ \frac{1}{4} + \frac{3}{4} e^{-4\nu_2/3} \right] \left[ \frac{1}{4} - \frac{1}{4} e^{-4\nu_3/3} \right] \left[ \frac{1}{4} - \frac{1}{4} e^{-4\nu_4/3} \right] \left[ \frac{1}{4} + \frac{3}{4} e^{-4\nu_5/3} \right]$$

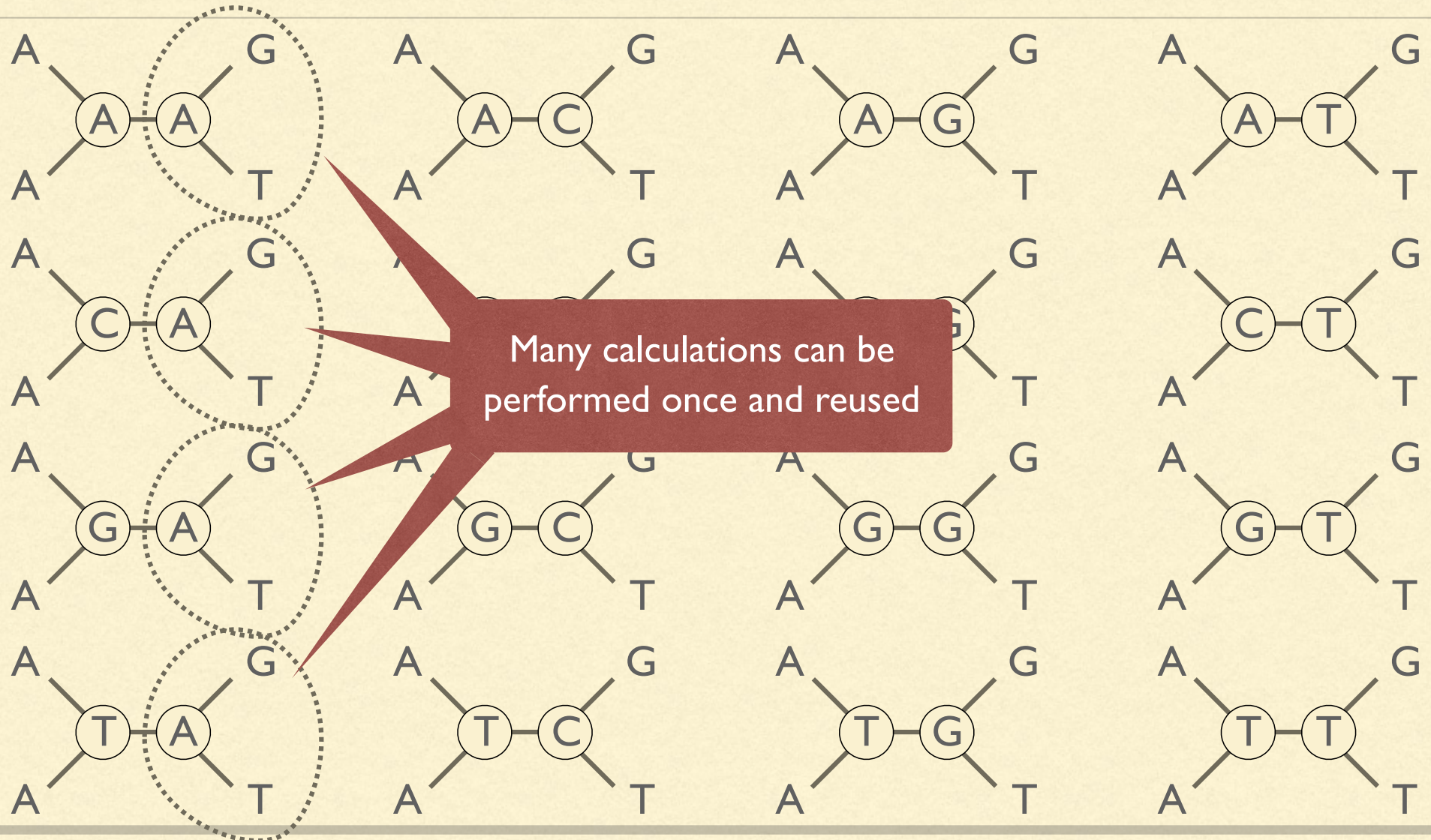


# Brute force approach



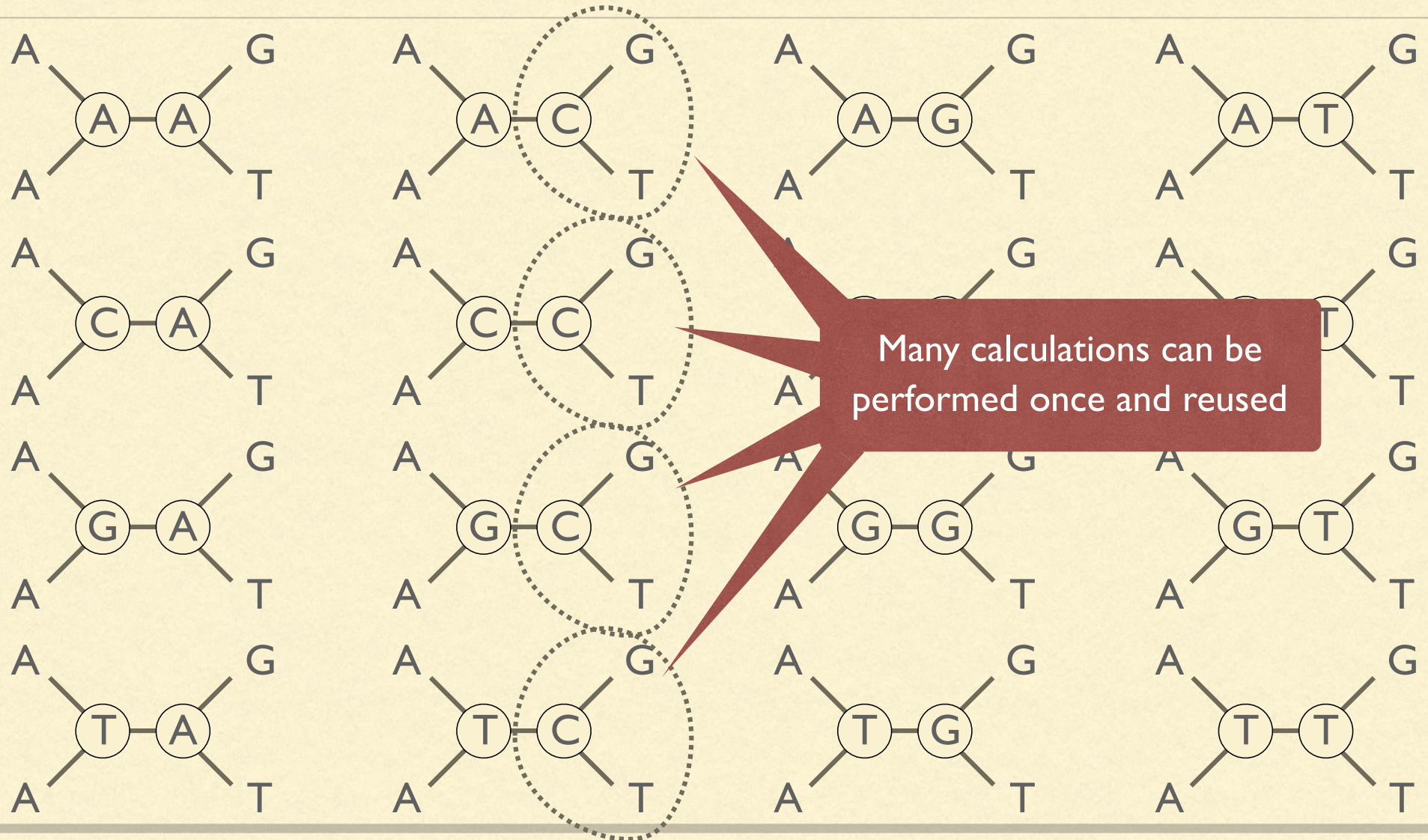
# Pruning algorithm

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

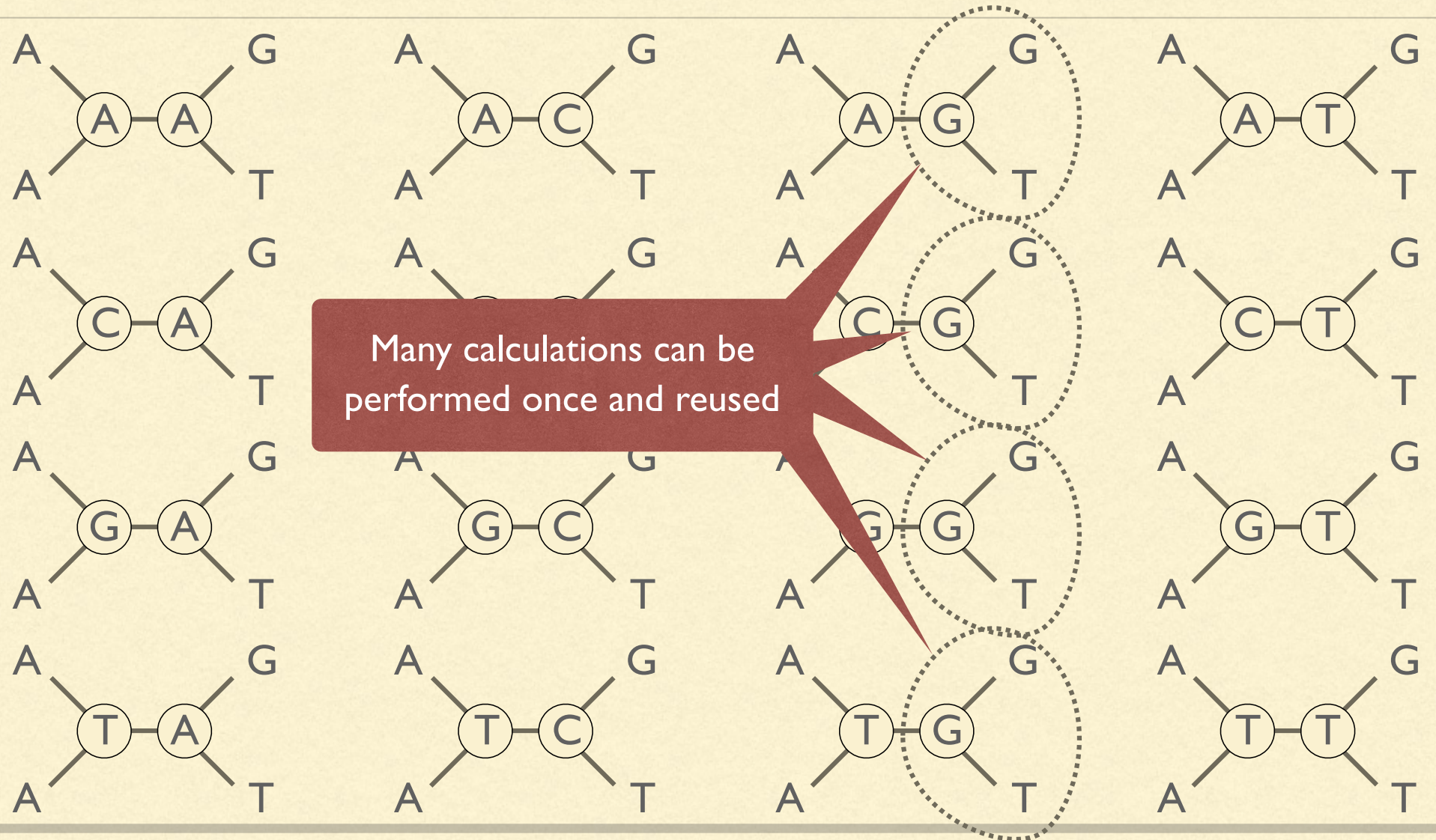




# Pruning algorithm

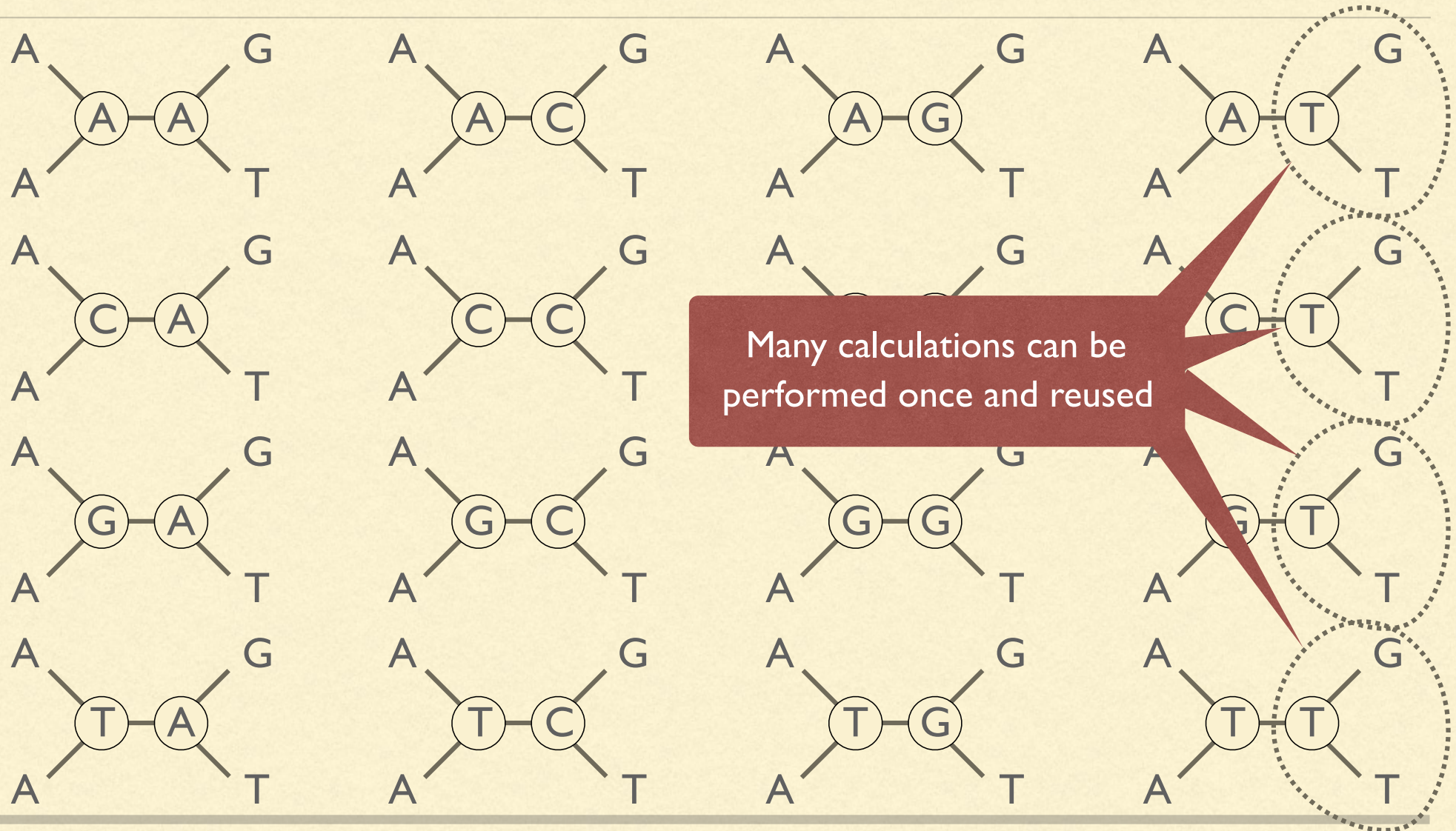


# Pruning algorithm



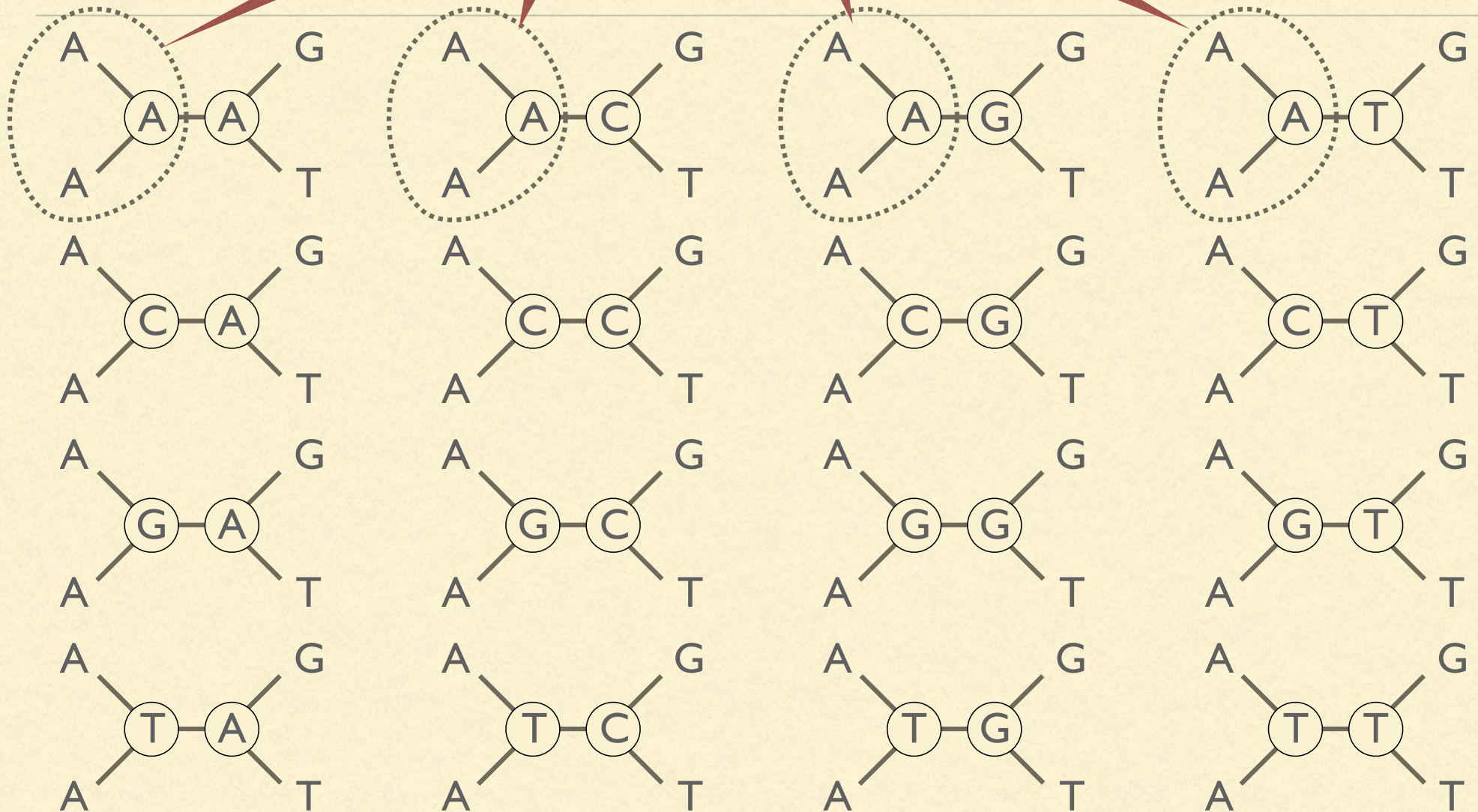


# Pruning algorithm



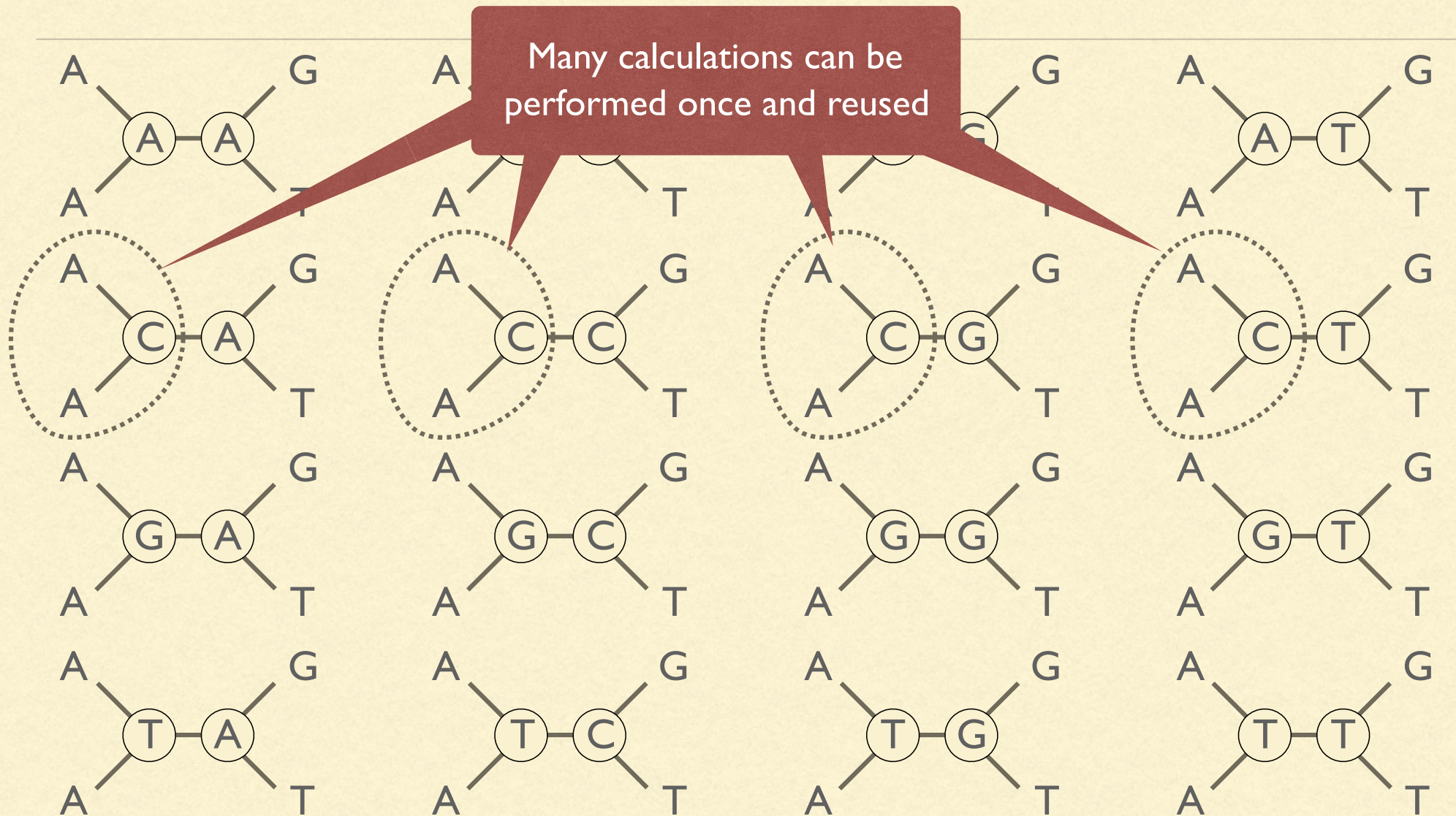
# Pruning alg

Many calculations can be performed once and reused





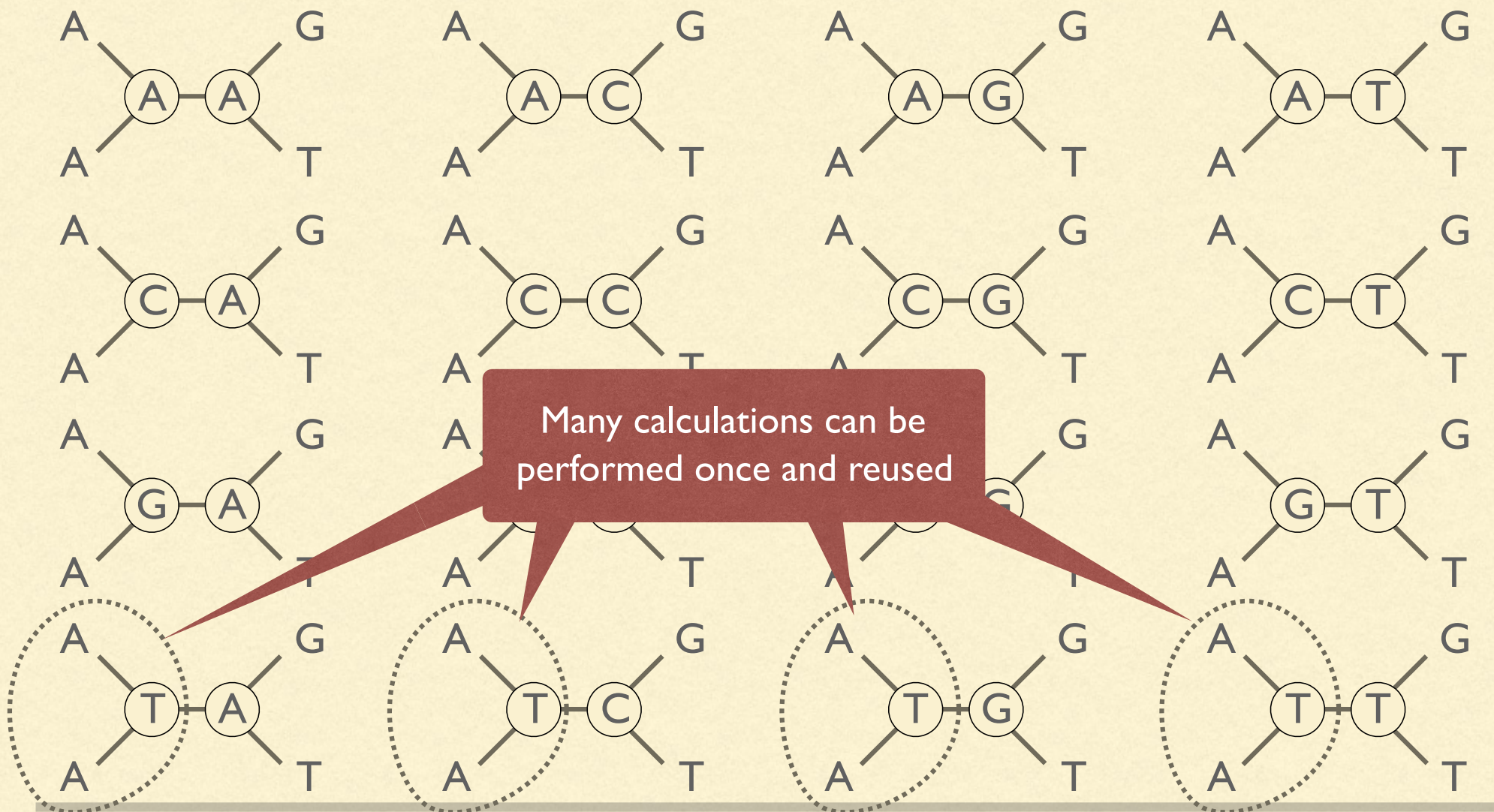
# Pruning algorithm







# Pruning algorithm



---

# Total likelihood

---

$$L = L_1 L_2 \cdots L_n$$

↑      ↑                      ↑  
site 1 site 2                      site n


$$\log L = \log L_1 + \log L_2 + \cdots + \log L_n$$



# Jukes and Cantor (1969)

JC69 model

to:

Parameters:  $\beta$

from:

$$\begin{array}{c} \text{A} \\ \text{C} \\ \text{G} \\ \text{T} \end{array} \left[ \begin{array}{cccc} \text{A} & \text{C} & \text{G} & \text{T} \\ -3\beta & \beta & \beta & \beta \\ \beta & -3\beta & \beta & \beta \\ \beta & \beta & -3\beta & \beta \\ \beta & \beta & \beta & -3\beta \end{array} \right]$$

# Kimura (1980)

K80 (or K2P) model

Parameters:  $\alpha, \beta$

$$\begin{array}{c} \text{A} \\ \text{C} \\ \text{G} \\ \text{T} \end{array} \begin{bmatrix} -\alpha - 2\beta & \beta & \alpha & \beta \\ \beta & -\alpha - 2\beta & \beta & \alpha \\ \alpha & \beta & -\alpha - 2\beta & \beta \\ \beta & \alpha & \beta & -\alpha - 2\beta \end{bmatrix}$$



# Kimura (1980)

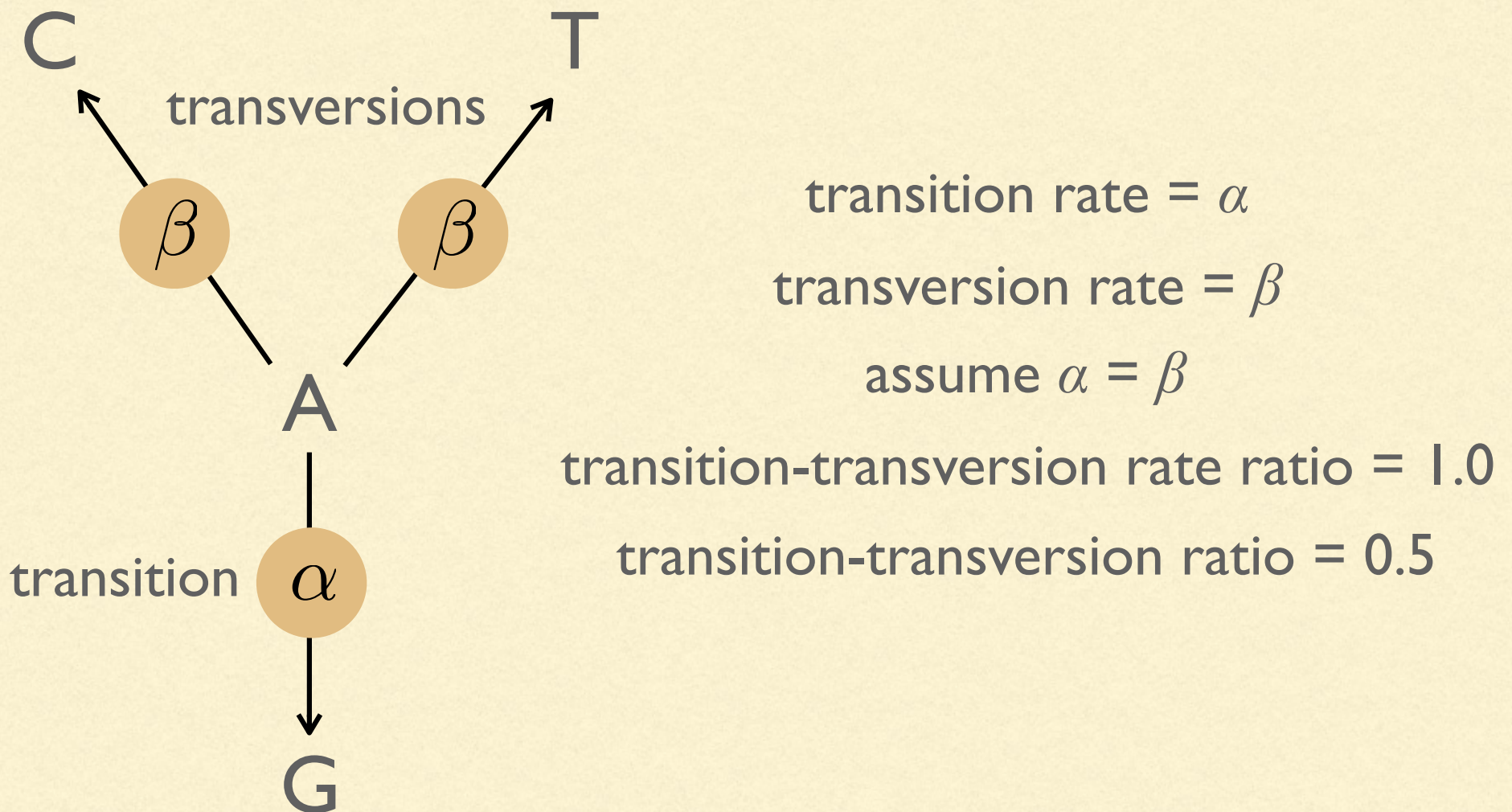
K80 (or K2P) model

$$\kappa = \alpha/\beta$$

Parameters:  $\kappa, \beta$

	A	C	G	T
A	$-\beta(\kappa + 2)$	$\beta$	$\kappa\beta$	$\beta$
C	$\beta$	$-\beta(\kappa + 2)$	$\beta$	$\kappa\beta$
G	$\kappa\beta$	$\beta$	$-\beta(\kappa + 2)$	$\beta$
T	$\beta$	$\kappa\beta$	$\beta$	$-\beta(\kappa + 2)$

# Transition-transversion (rate) ratio





# Felsenstein (1981)

F81 model

Parameters:  $\mu, \pi_A, \pi_C, \pi_G$

	A	C	G	T
A	$-\mu(1 - \pi_A)$	$\pi_C \mu$	$\pi_G \mu$	$\pi_T \mu$
C	$\pi_A \mu$	$-\mu(1 - \pi_C)$	$\pi_G \mu$	$\pi_T \mu$
G	$\pi_A \mu$	$\pi_C \mu$	$-\mu(1 - \pi_G)$	$\pi_T \mu$
T	$\pi_A \mu$	$\pi_C \mu$	$\pi_G \mu$	$-\mu(1 - \pi_T)$

---

JC69 is a special case of F81

---

	A	C	G	T		A	C	G	T
A	$-\frac{3}{4}\mu$	$\frac{1}{4}\mu$	$\frac{1}{4}\mu$	$\frac{1}{4}\mu$	$\left[ \begin{array}{cccc} -3\beta & \beta & \beta & \beta \\ \beta & -3\beta & \beta & \beta \\ \beta & \beta & -3\beta & \beta \\ \beta & \beta & \beta & -3\beta \end{array} \right]$	$-3\beta$	$\beta$	$\beta$	$\beta$
C	$\frac{1}{4}\mu$	$-\frac{3}{4}\mu$	$\frac{1}{4}\mu$	$\frac{1}{4}\mu$		$\beta$	$-3\beta$	$\beta$	$\beta$
G	$\frac{1}{4}\mu$	$\frac{1}{4}\mu$	$-\frac{3}{4}\mu$	$\frac{1}{4}\mu$		$\beta$	$\beta$	$-3\beta$	$\beta$
T	$\frac{1}{4}\mu$	$\frac{1}{4}\mu$	$\frac{1}{4}\mu$	$-\frac{3}{4}\mu$		$\beta$	$\beta$	$\beta$	$-3\beta$

$$\beta = \frac{1}{4}\mu$$



# Hasegawa, Kishino, and Yano (1985)

## HKY85 model

Parameters:  $\mu, \kappa, \pi_A, \pi_C, \pi_G$

these are global  
parameters  
(apply to all  
edge lengths)

one parameter in each model is  
associated with the length of an edge

	A	C	G	T
A	$-\mu(\pi_C + \pi_G\kappa + \pi_T)$	$\pi_C\mu$	$\pi_G\mu\kappa$	$\pi_T\mu$
C	$\pi_A\mu$	$-\mu(\pi_A + \pi_G + \pi_T\kappa)$	$\pi_G\mu$	$\pi_T\mu\kappa$
G	$\pi_A\mu\kappa$	$\pi_C\mu$	$-\mu(\pi_A\kappa + \pi_C + \pi_T)$	$\pi_T\mu$
T	$\pi_A\mu$	$\pi_C\mu\kappa$	$\pi_G\mu$	$-\mu(\pi_A + \pi_C\kappa + \pi_G)$

# Hasegawa, Kishino, and Yano (1985)

HKY85 model

Parameters:  $\mu, \kappa, \pi_A, \pi_C, \pi_G, \pi_T$

sum of the circled rates equals the total rate given that we start with an A

	A	C	G	T
A	$-\mu(\pi_C + \pi_G\kappa + \pi_T)$	$\pi_C\mu$	$\pi_G\mu\kappa$	$\pi_T\mu$
C	$\pi_A\mu$	$-\mu(\pi_A + \pi_G + \pi_T\kappa)$	$\pi_G\mu$	$\pi_T\mu\kappa$
G	$\pi_A\mu\kappa$	$\pi_C\mu$	$-\mu(\pi_A\kappa + \pi_C + \pi_T)$	$\pi_T\mu$
T	$\pi_A\mu$	$\pi_C\mu\kappa$	$\pi_G\mu$	$-\mu(\pi_A + \pi_C\kappa + \pi_G)$



# Hasegawa, Kishino, and Yano (1985)

HKY85 model

Parameters:  $\mu, \kappa, \pi_A, \pi_C, \pi_G$

The diagonal element conveniently equals the negative of the total rate away from A

$$\begin{array}{c} \text{A} \\ \text{C} \\ \text{G} \\ \text{T} \end{array} \begin{bmatrix} \text{A} & \text{C} & \text{G} & \text{T} \\ \begin{array}{c} -\mu(\pi_C + \pi_G\kappa + \pi_T) \\ \pi_A\mu \\ \pi_A\mu\kappa \\ \pi_A\mu \end{array} & \begin{array}{c} \pi_C\mu \\ -\mu(\pi_A + \pi_G + \pi_T\kappa) \\ \pi_C\mu \\ \pi_C\mu\kappa \end{array} & \begin{array}{c} \pi_G\mu\kappa \\ \pi_G\mu \\ -\mu(\pi_A\kappa + \pi_C + \pi_T) \\ \pi_G\mu \end{array} & \begin{array}{c} \pi_T\mu \\ \pi_T\mu\kappa \\ \pi_T\mu \\ -\mu(\pi_A + \pi_C\kappa + \pi_G) \end{array} \end{array} \right]$$

---

~ Coffee Break ~

---