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

28 May 2022

Paul O. Lewis
Department of Ecology & Evolutionary Biology
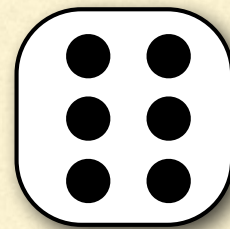
UCONN
UNIVERSITY OF CONNECTICUT

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?
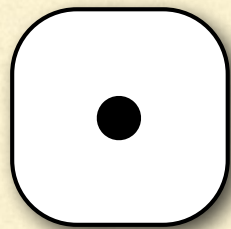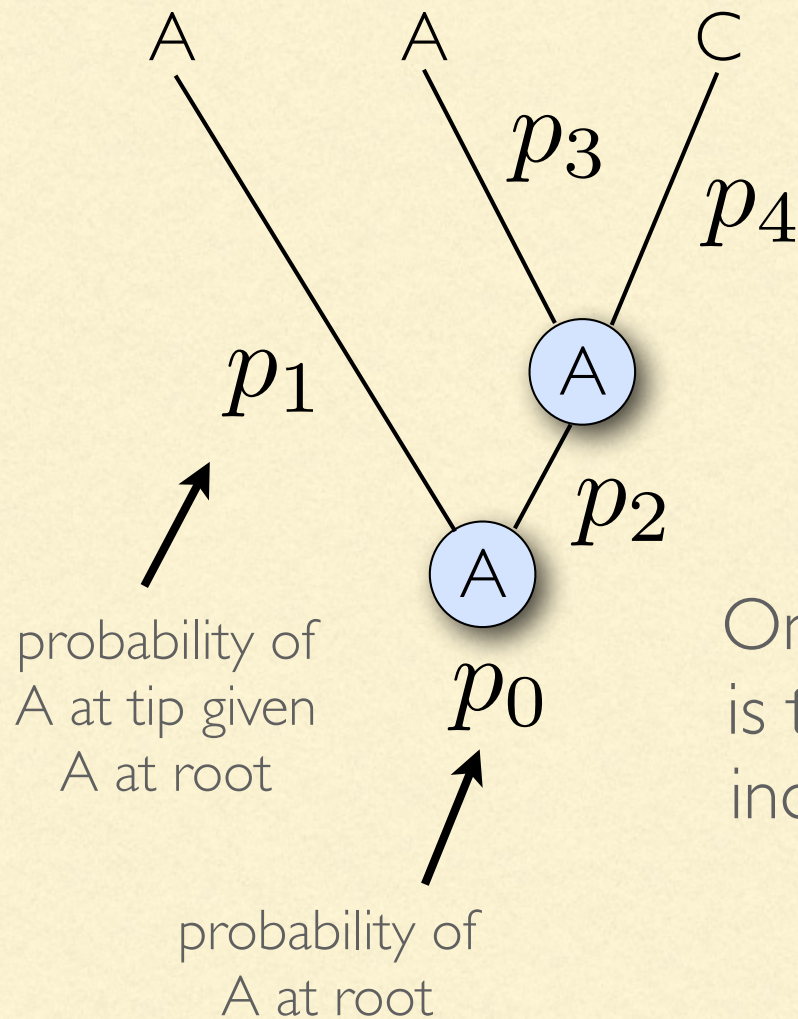
 AND

(1/6) × (1/6) = 1/36

AND rule in phylogenetics

A   A   C

p_3

p_4

p_1

A

p_2

A

p_0

we have observed these states

pretend (for now) we also
know these ancestral states

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

probability of
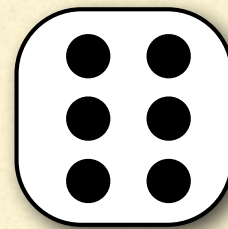A at tip given
A at root

probability of
A at root

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?
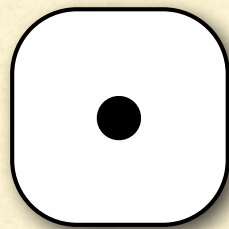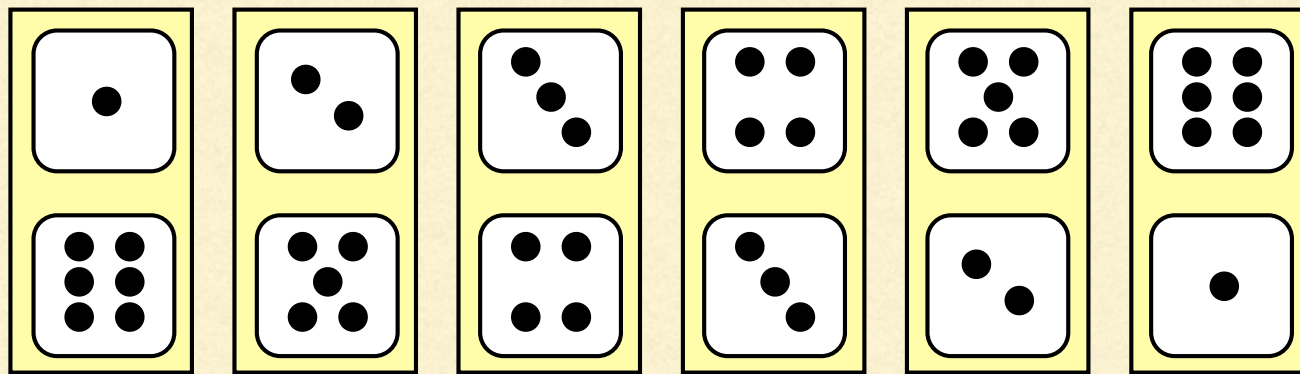

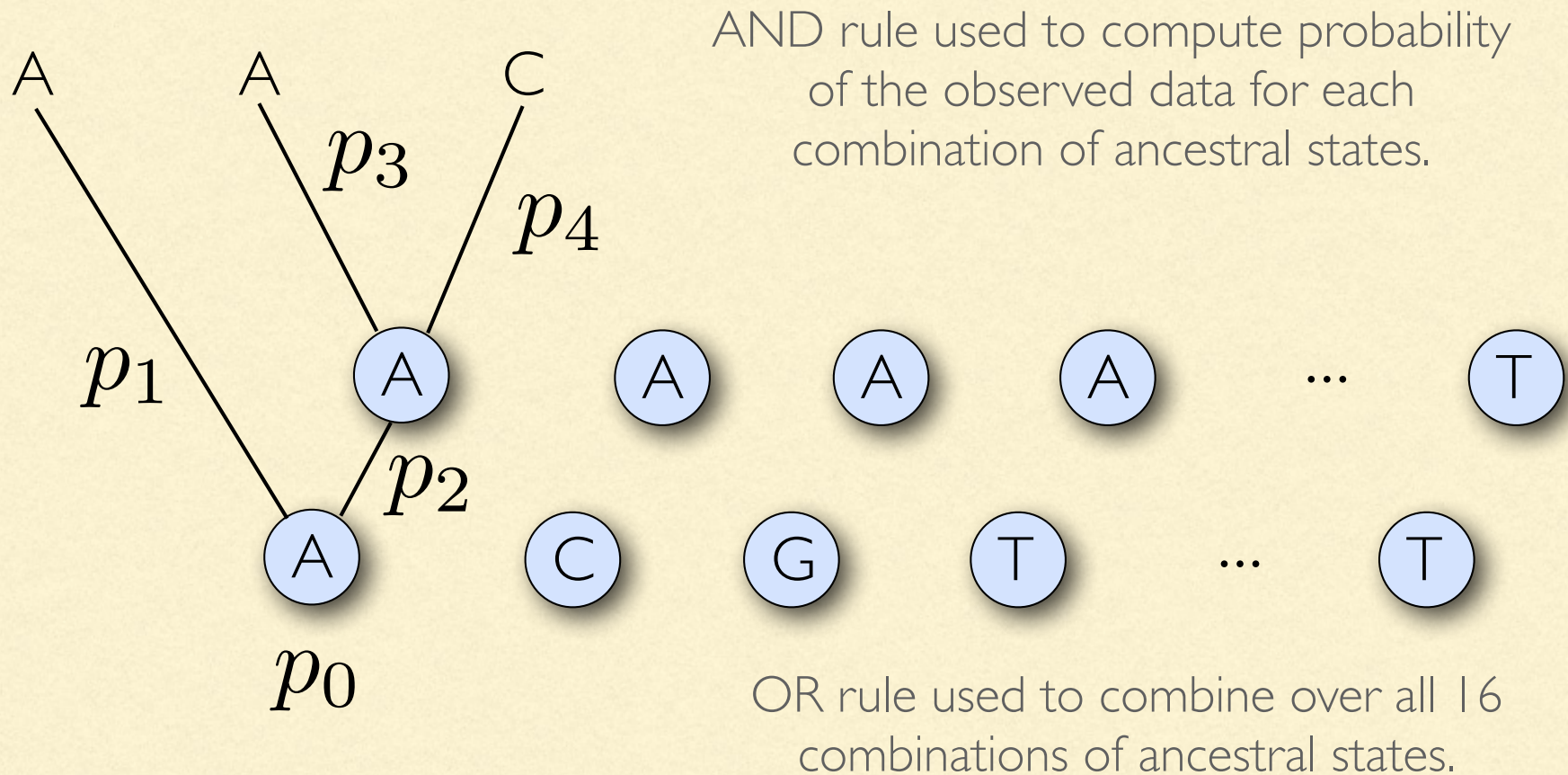(1/6)   +   (1/6)   =   1/3

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

AND rule used to compute probability of the observed data for each combination of ancestral states.

OR rule used to combine over all 16 combinations of ancestral states.

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

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

$$\mathbf{Pr}(A, B) = \mathbf{Pr}(A)\,\mathbf{Pr}(B|A)$$

joint probability
of A and B

conditional probability
of B given A

Pr(walk to work|sunny) = 0.99
Pr(walk to work|raining) = 0.50

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

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

time *t*

less probable

time *t*

more probable
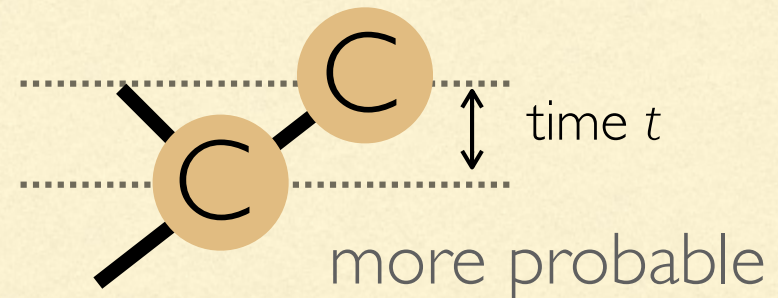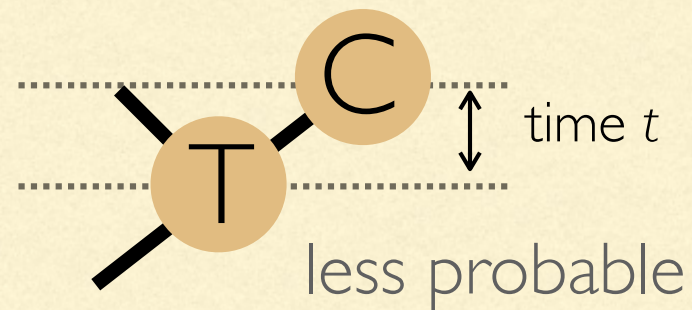
Conditional Independence

$$Pr(A, B | C) = Pr(A | C)\, Pr(B | C)$$
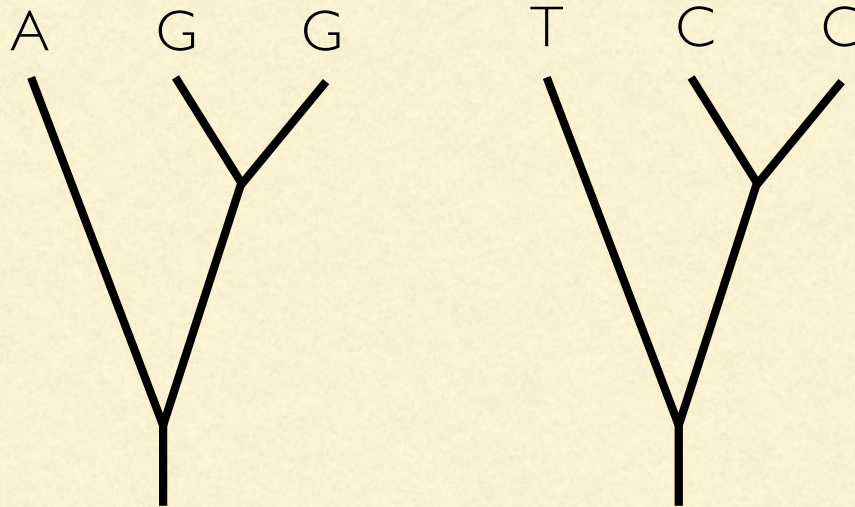
$$Pr(AGG, TCC | tree) = Pr(AGG | tree)\, Pr(TCC | tree)$$

Likelihood

The Likelihood Criterion

Suppose I threw 20 dice on the table and this was the result...
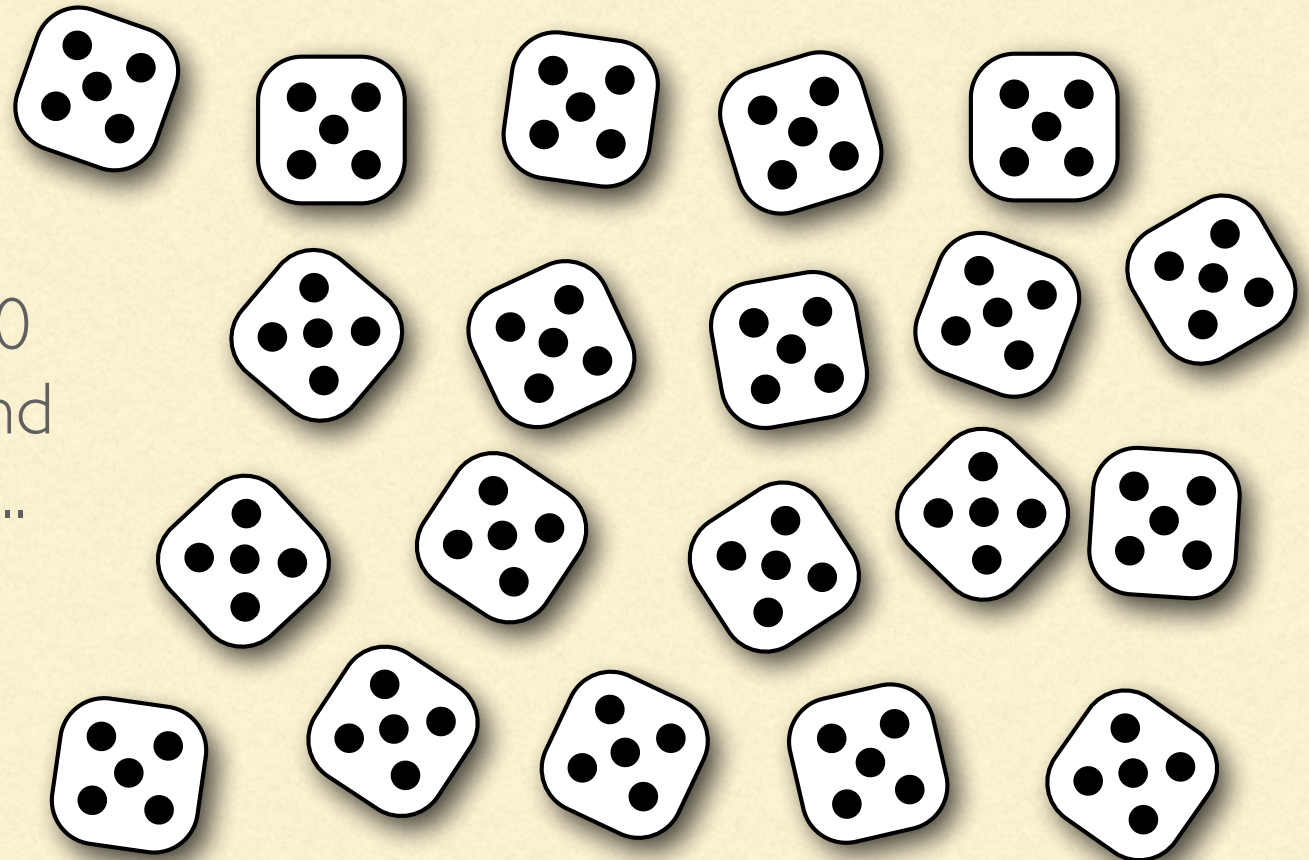
The Fair Dice model

You should have been **very surprised** at this result because the probability of this event is **very small**: only 1 in 3.6 quadrillion!
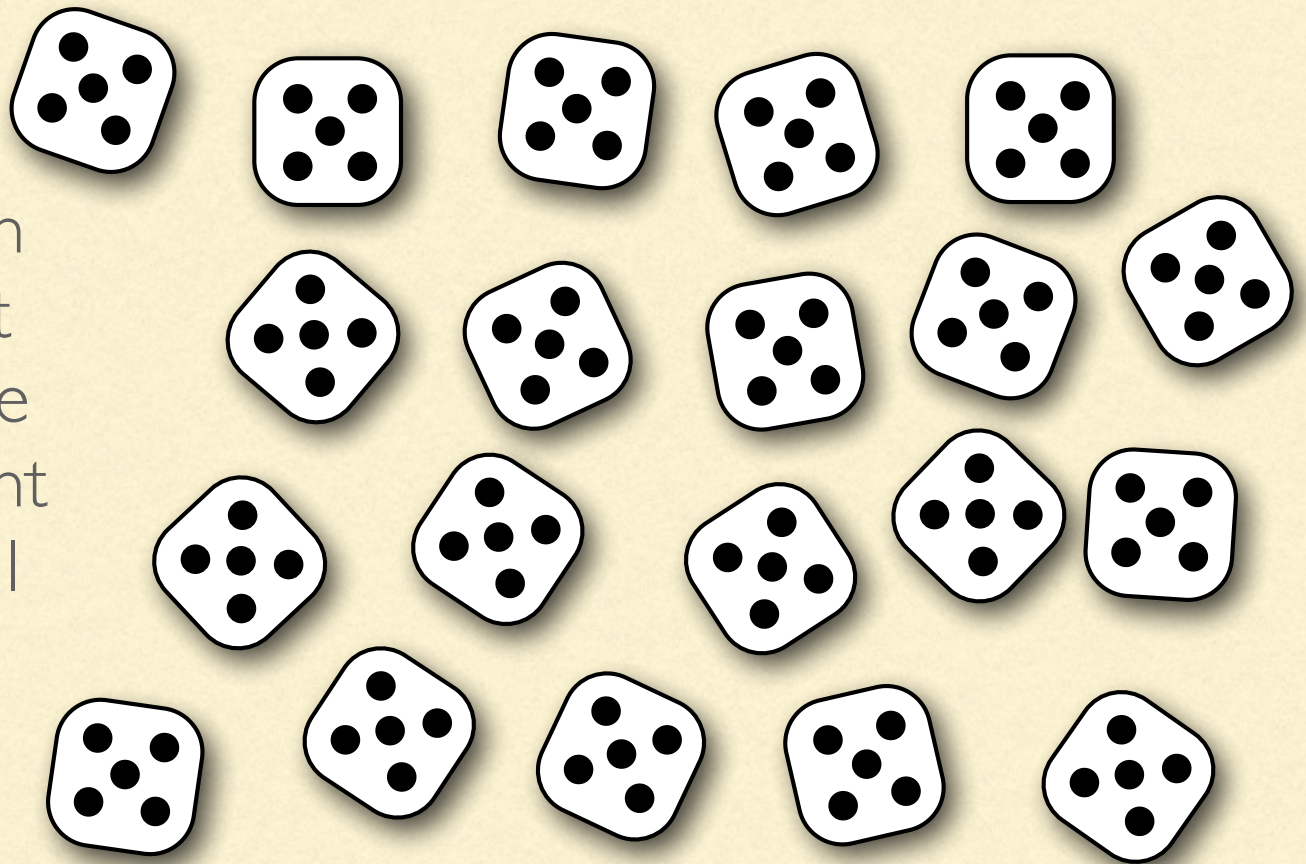
The Trick Dice model
(assumes dice each have 5 on every side)

You should be **not surprised** at all at this result because the observed outcome is **certain** under this model.
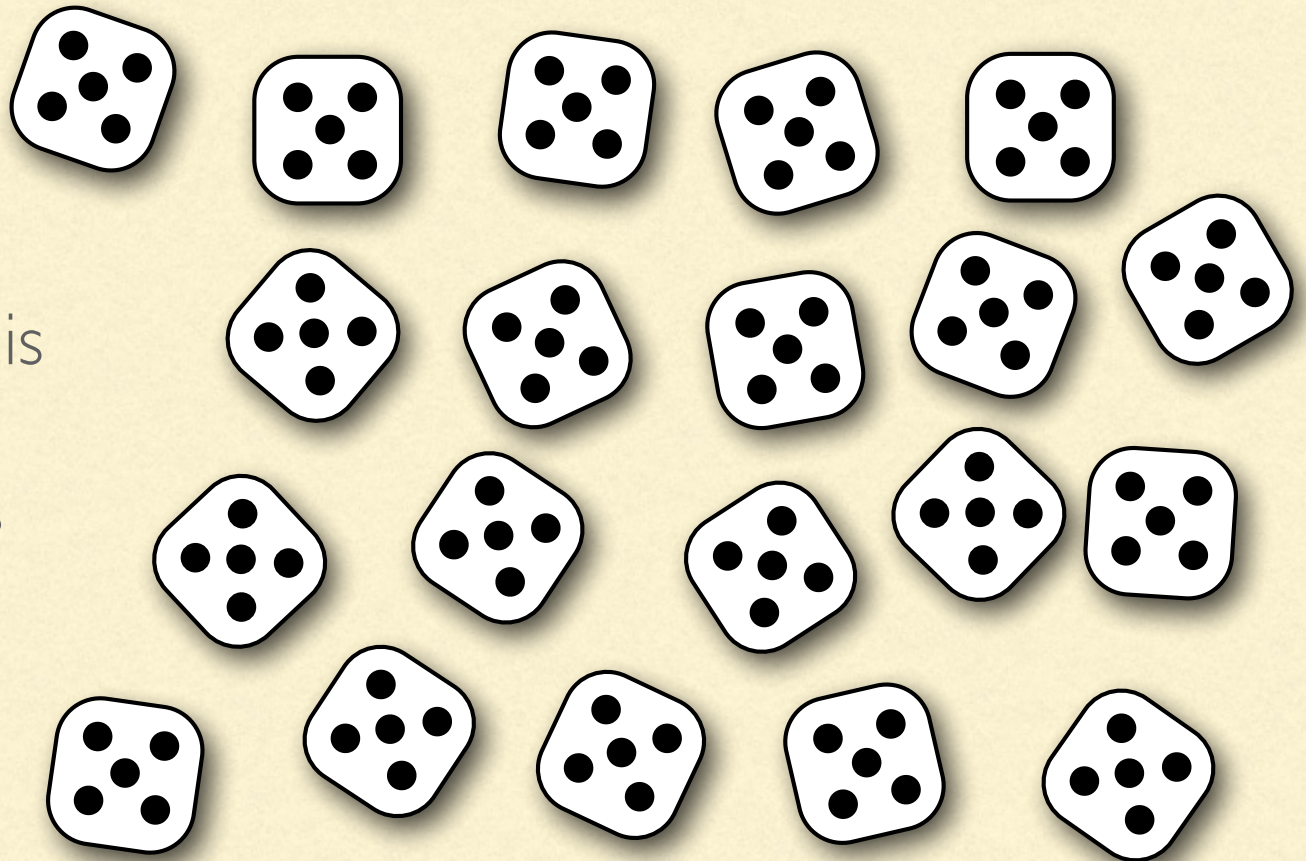
What's changed? (the model)

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

Model	Likelihood	Surprise level
Fair Dice	$\dfrac{1}{3{,}656{,}158{,}440{,}062{,}976}$	Very, *very*, **very** 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
	1	1

Probabilities of data outcomes given one particular model sum to 1.0

Likelihoods of models given one particular data set are not expected to 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 **ψη**-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
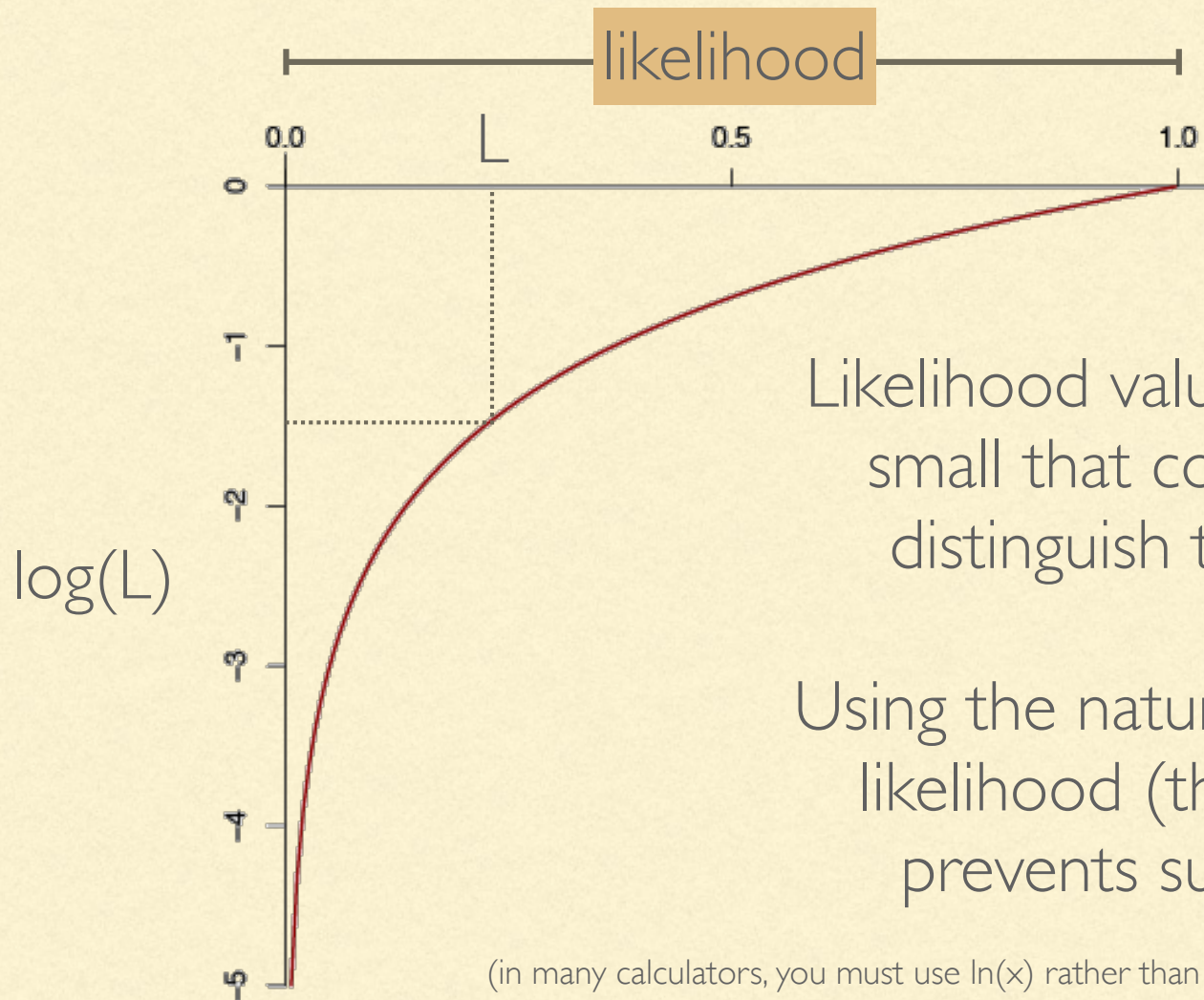
likelihood

Likelihood values can become so small that computers cannot distinguish them from zero.

Using the natural logarithm of the likelihood (the log-likelihood) prevents such "underflow"

(in many calculators, you must use ln(x) rather than log(x) to take the natural log of the value x)

Likelihood of a single-edge tree

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

root

GA • ―――――――――――――――――― • **GG**

$\longleftarrow v \longrightarrow$

↗ ↗   site 1   site 2   ↗ ↗   site 1   site 2
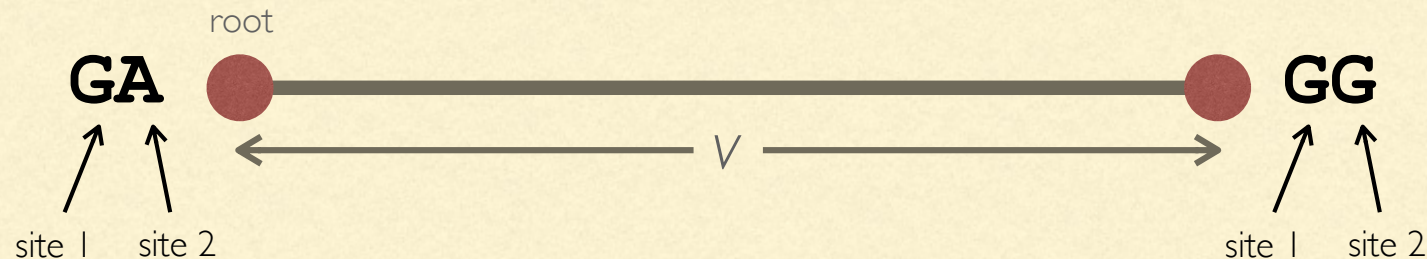
$$L = \left[\left(\tfrac{1}{4} \right) \left(\tfrac{1}{4} + \tfrac{3}{4} e^{-4v/3} \right) \right] \left[\left(\tfrac{1}{4} \right) \left(\tfrac{1}{4} - \tfrac{1}{4} e^{-4v/3} \right) \right]$$

$\longleftarrow$ site 1 $\longrightarrow$ $\longleftarrow$ site 2 $\longrightarrow$

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**

$$v \quad = \quad 3\beta \quad \times \quad t$$
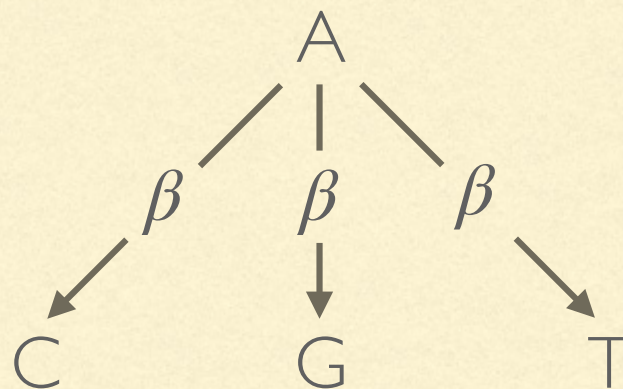
3 possible substitutions, each of which happens with rate β

Jukes and Cantor (1969)

JC69 model   to: Parameters: β

$$
\begin{array}{c}
 \\
\text{from:}
\end{array}
\begin{array}{cccc}
 & \text{A} & \text{C} & \text{G} & \text{T} \\
\text{A} & -3\beta & \beta & \beta & \beta \\
\text{C} & \beta & -3\beta & \beta & \beta \\
\text{G} & \beta & \beta & -3\beta & \beta \\
\text{T} & \beta & \beta & \beta & -3\beta
\end{array}
$$

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.

Maximum likelihood estimation

0.065 is the maximum likelihood estimate (MLE) of v


```
gorilla    GAAGTCCTTGAGAAATAAACTGCACACACTGG
orangutan  GGACTCCTTGAGAAATAAACTGCACACACTGG
```

$$L = \left[\left(\frac{1}{4}\right) \left(\frac{1}{4} + \frac{3}{4} e^{-4\nu/3}\right) \right]^{30} \left[\left(\frac{1}{4}\right) \left(\frac{1}{4} - \frac{1}{4} e^{-4\nu/3}\right) \right]^{2}$$
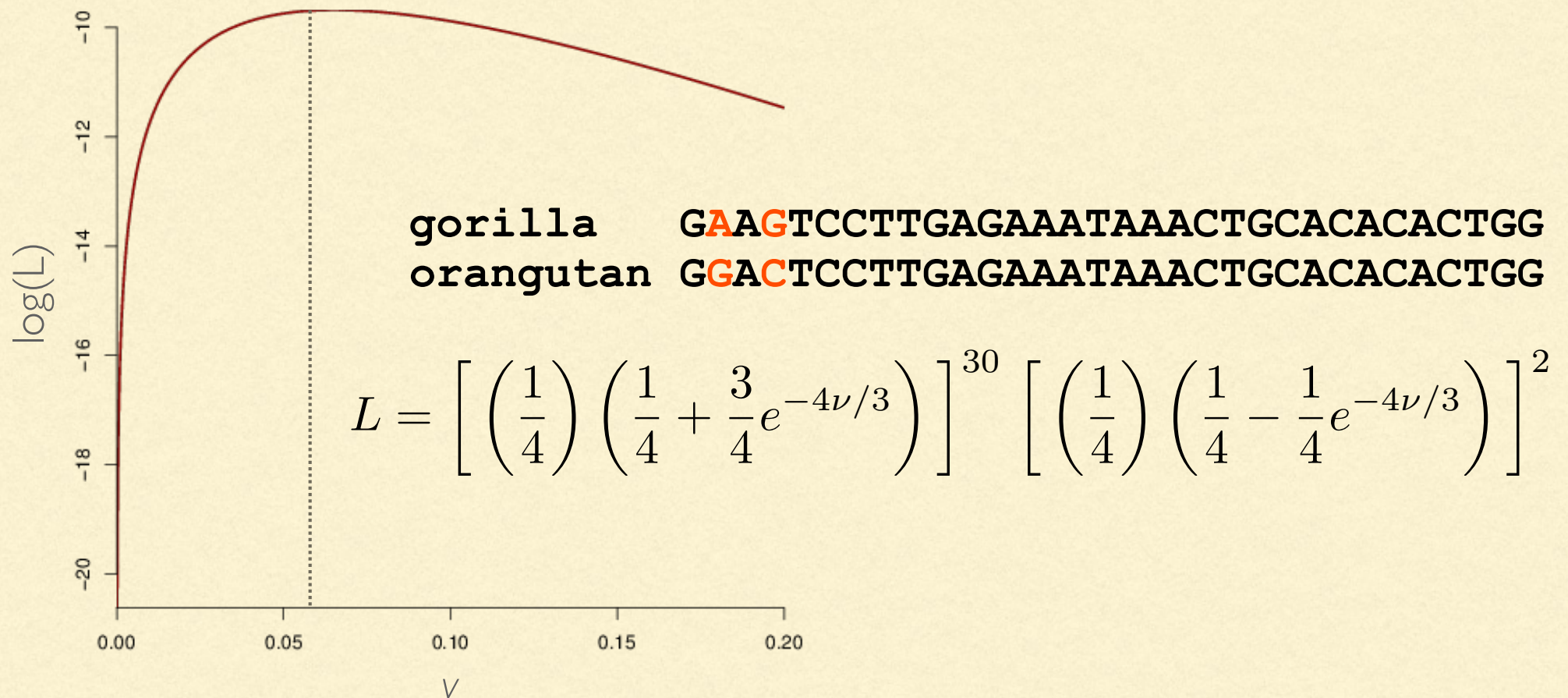
Transition probabilities

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

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?

"ACHNyons" vs. substitutions

I made up this term (Anything Can Happen Now)

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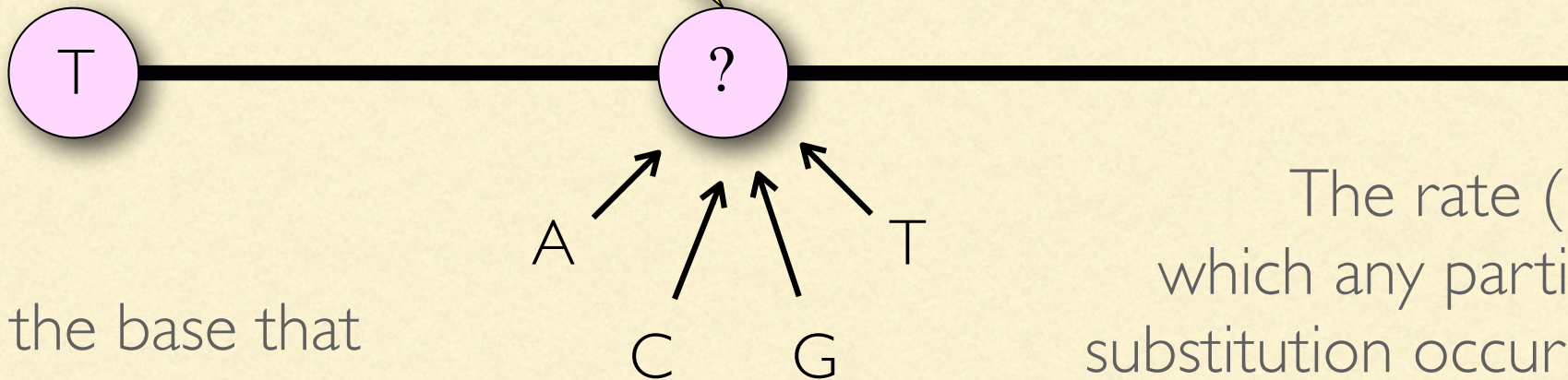
T ———————— ?

A   C   G   T

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

The rate (β) at which any particular substitution occurs will be 1/4 the achnyon rate (μ). 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 β:

$$\Pr(0 \text{ achnyons}) = e^{-\mu t} \quad \text{(Poisson probability of zero events)}$$
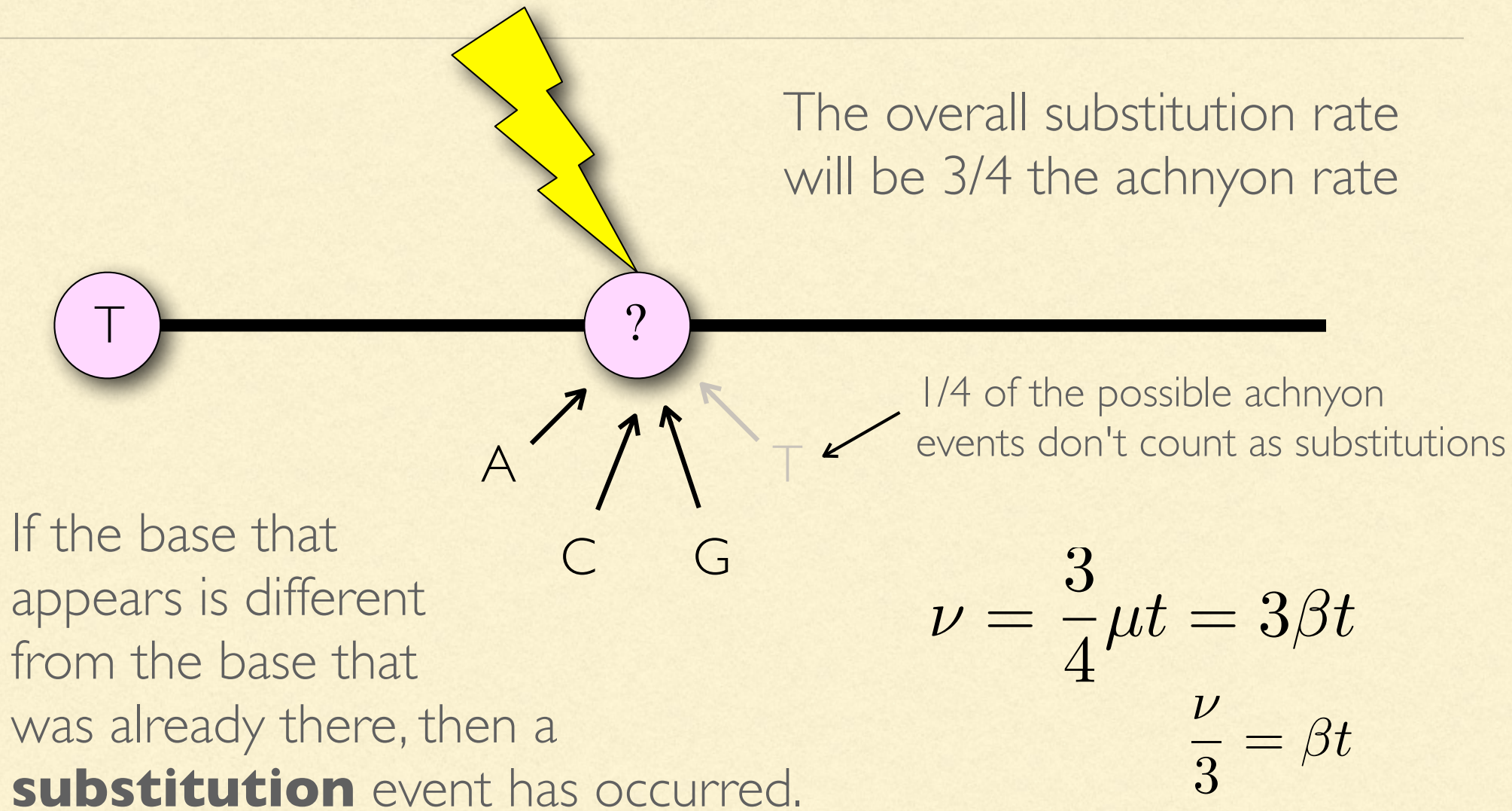
$$\Pr(\text{at least 1 achnyon}) = 1 - e^{-\mu t}$$

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

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

Remember that the achnyon rate (μ) is 4 times the rate (β) of any *particular* substitution:

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

Expected number of substitutions

The overall substitution rate will be 3/4 the achnyon rate

T ————————— ? —————————

A C G T

1/4 of the possible achnyon events don't count as substitutions

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

$$\nu = \frac{3}{4}\mu t = 3\beta t$$

$$\frac{\nu}{3} = \beta t$$

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 β:

$$\vdots$$

Remember that the achnyon rate (μ) is 4 times the rate (β) of any *particular* substitution:

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

Substitute $v/3$ for βt:

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

Transition Probabilities: Remarks

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

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

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

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

These should add to 1.0 because T *must* change to something!

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

Doh! Something must be wrong here...

Transition Probabilities: Remarks

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

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

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

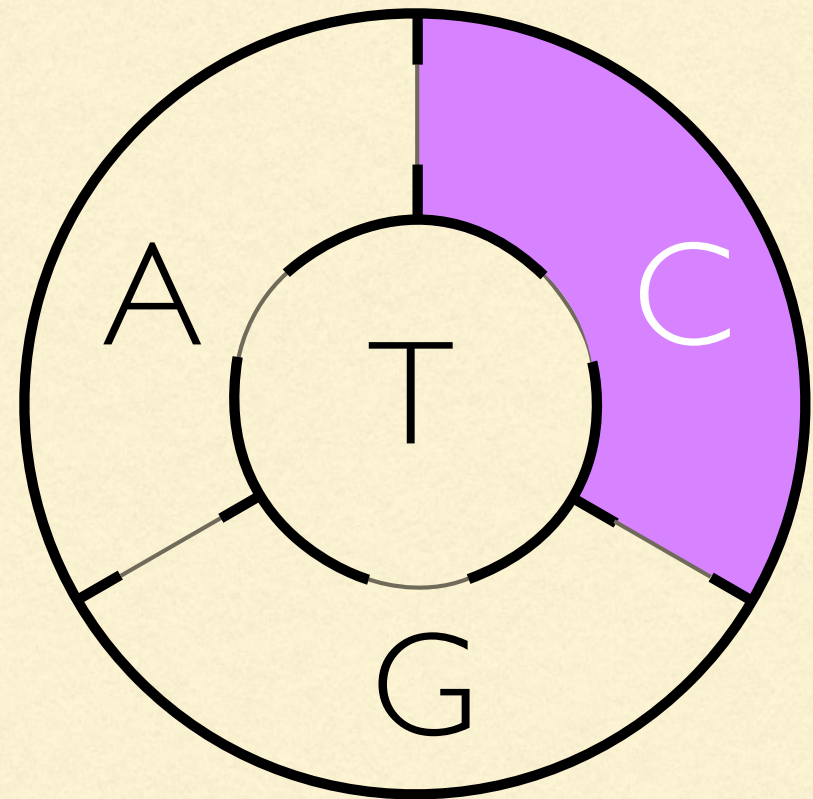$$P_{TT}(t) = \frac{1}{4}\left(1 - e^{-4\nu/3}\right) + e^{-4\nu/3}$$

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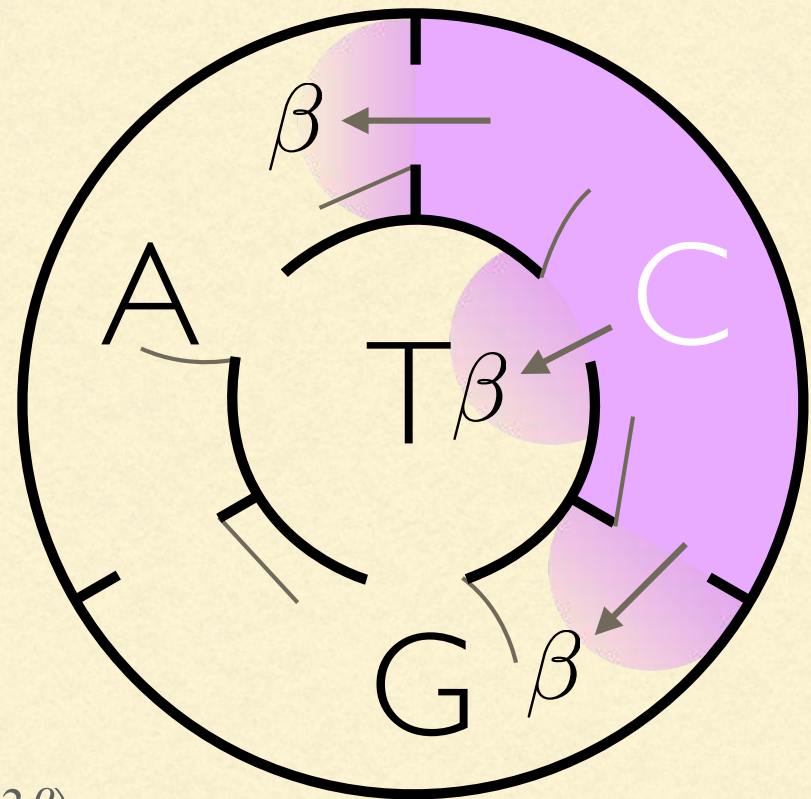
At the **instant the doors open**, perfume molecules...

enter room A at rate β
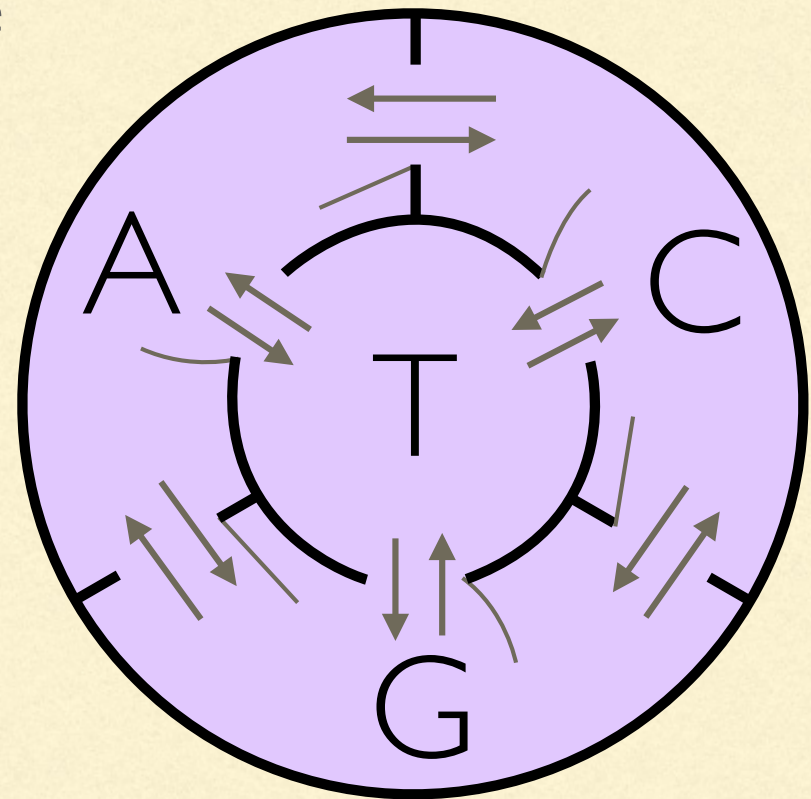
enter room T at rate β
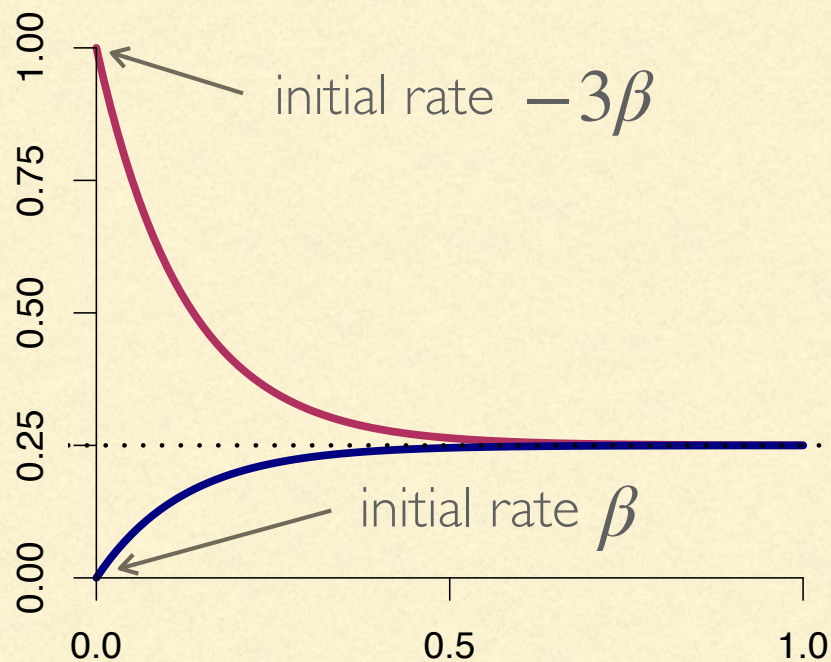
enter room G at rate β

enter room C at rate -3β

(you could also say they *leave* C at rate 3β)

Equilibrium Frequencies

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

initial rate -3β

initial rate β

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

Transition probability demo

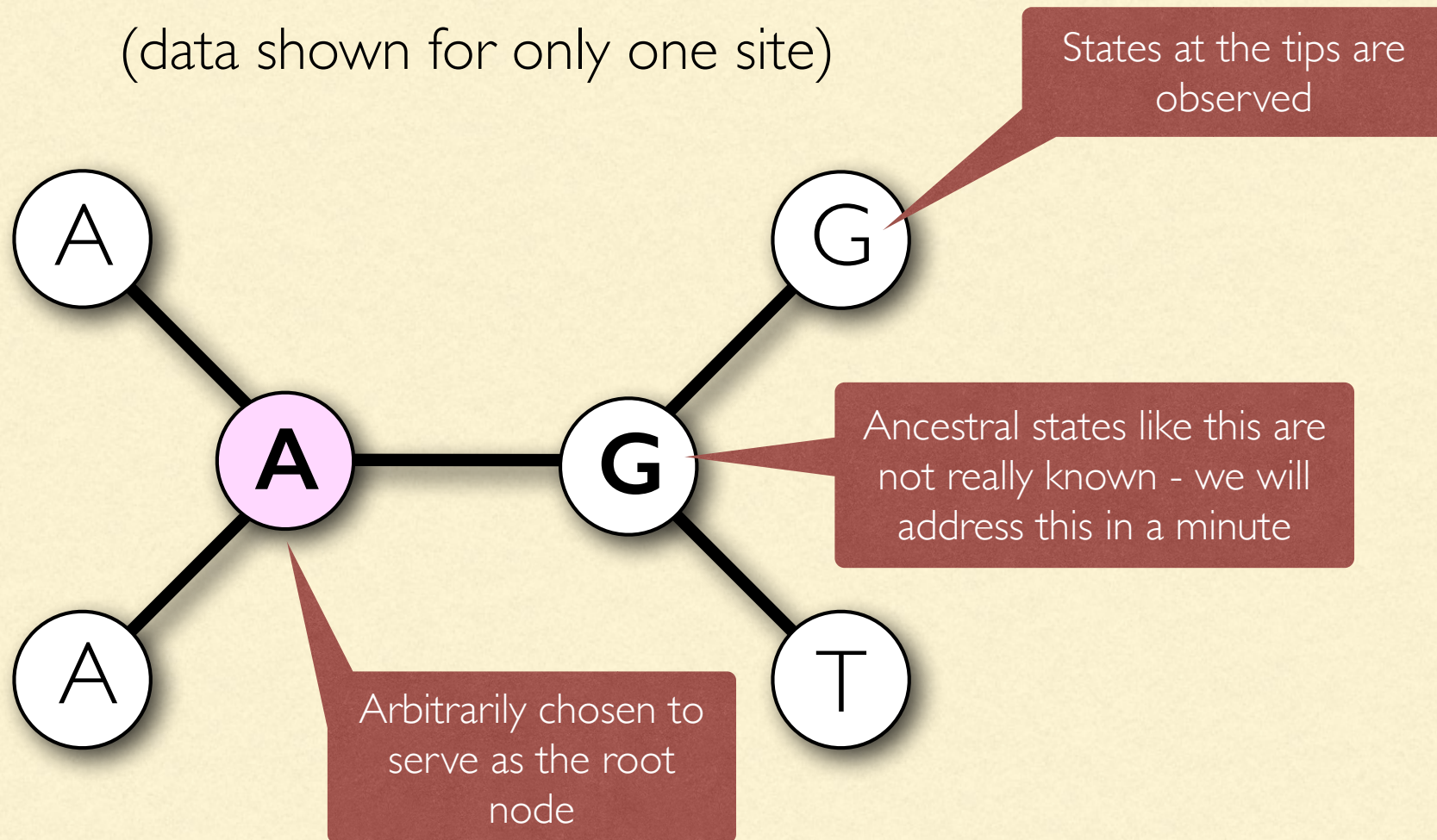https://plewis.github.io/applets/jc-transition-probabilities/
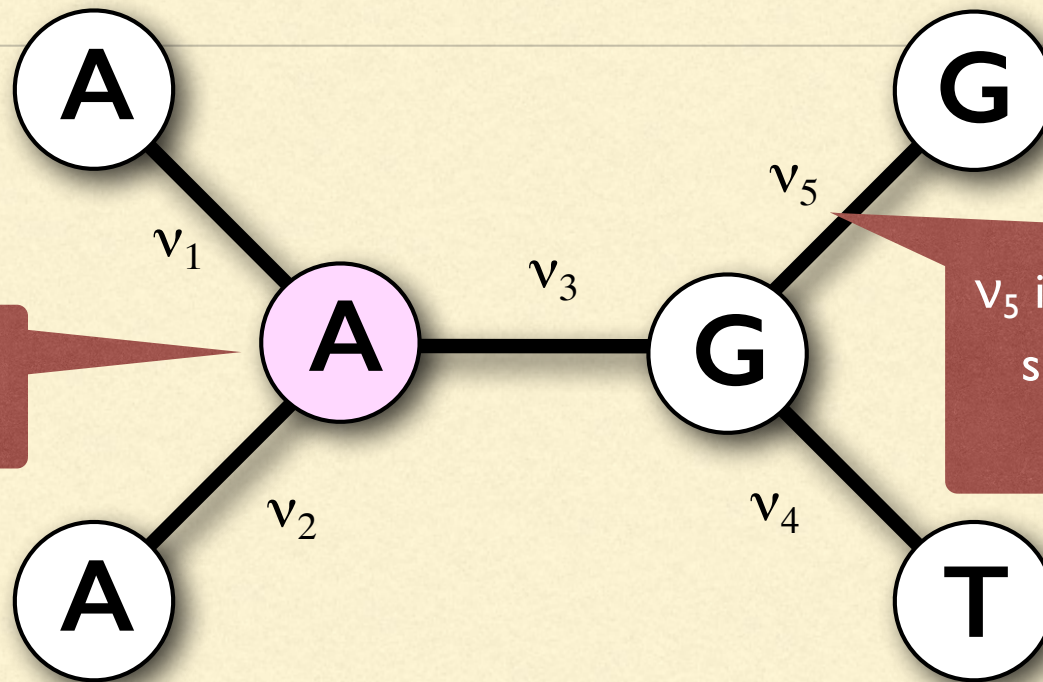
Sequence data for four taxa

one site

```
Sphagnum    GGCAGCATTTCGAATGACTCCTCAACCTGGAGTA CACCCG...
Asplenium   GGCAGCTTTCCGGATGACCCCACAACCCGGAGTA CAGCTG...
Picea       GGCAGCATTCCGAGTAACTCCTCAACCAGGGGTG CGCCCG...
Avena       GGCAGCATTCCGAGTAACTCCTCAACCTGGGGTT CGCCGG...
```

The highlighted site reads (top to bottom): A, A, G, T

Likelihood for tree (one site)

(data shown for only one site)

States at the tips are observed

Ancestral states like this are not really known - we will address this in a minute

Arbitrarily chosen to serve as the root node

Likelihood for tree (one site)

$\pi_\mathbb{A}$

ν_5 is the expected number of substitutions for just this one branch
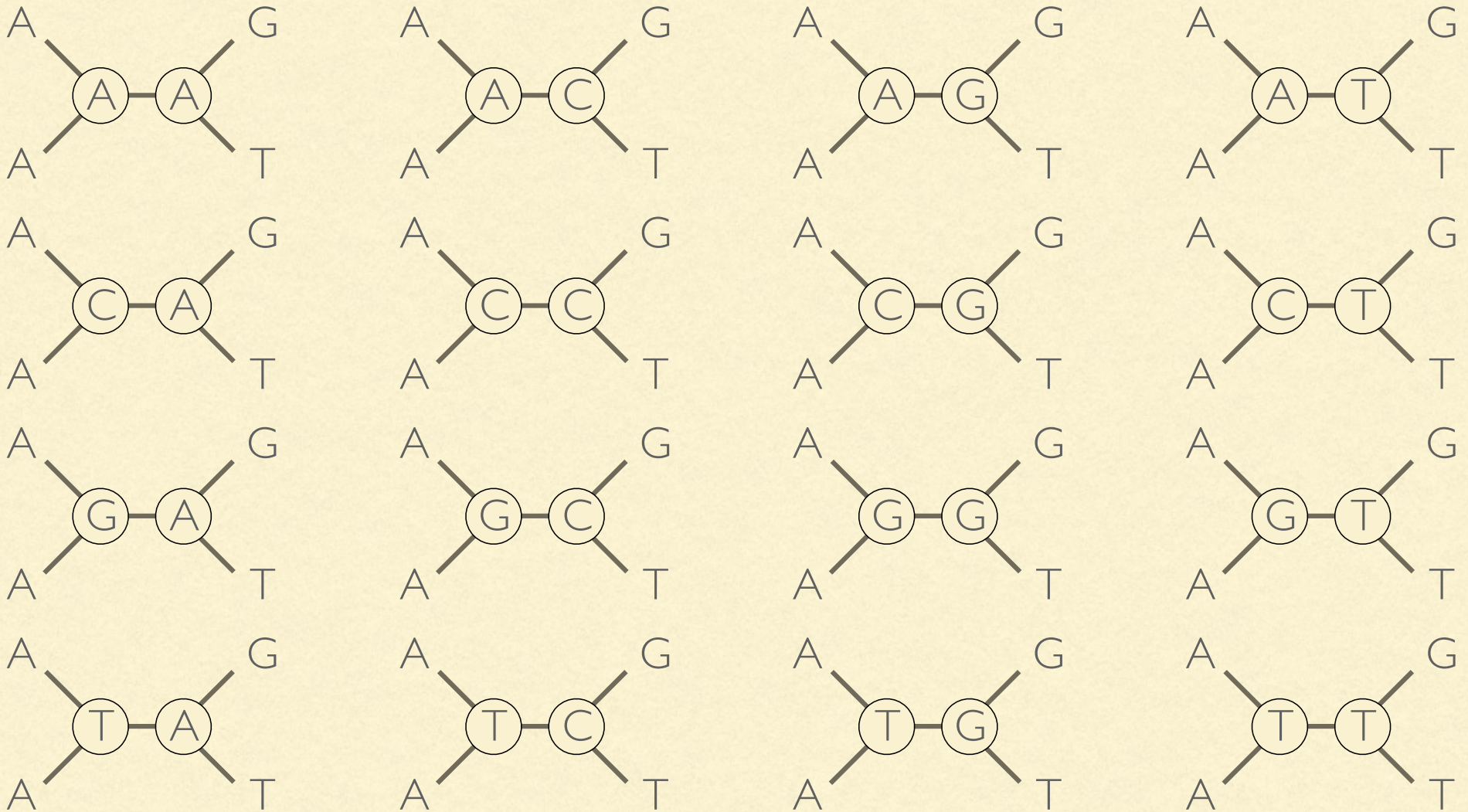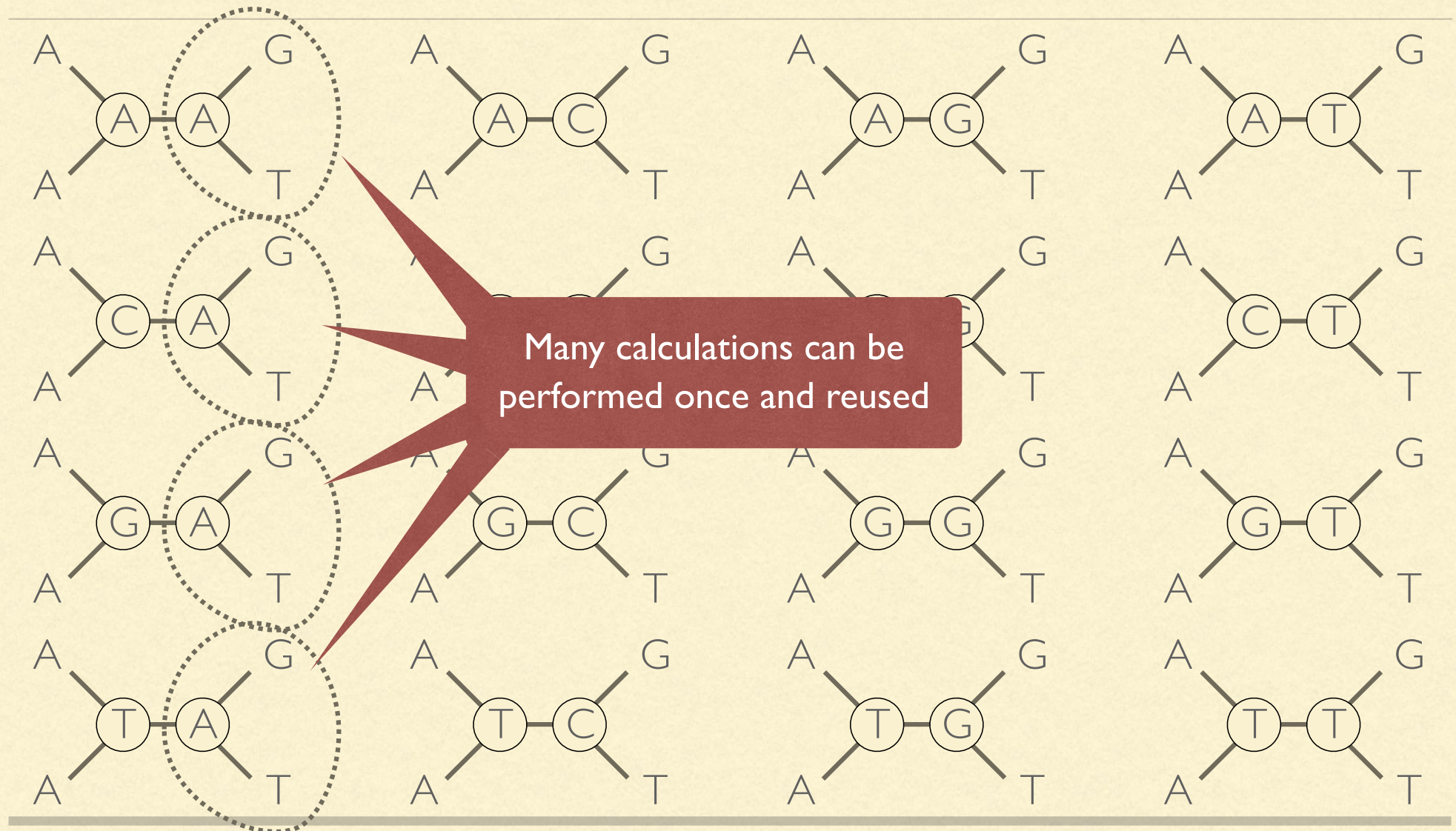
$$L = \frac{1}{4}\left[\frac{1}{4} + \frac{3}{4}e^{-4\nu_1/3}\right]\left[\frac{1}{4} + \frac{3}{4}e^{-4\nu_2/3}\right]\left[\frac{1}{4} - \frac{1}{4}e^{-4\nu_3/3}\right]\left[\frac{1}{4} - \frac{1}{4}e^{-4\nu_4/3}\right]\left[\frac{1}{4} + \frac{3}{4}e^{-4\nu_5/3}\right]$$

Brute force approach

Pruning algorithm

Many calculations can be performed once and reused

Pruning algorithm

Many calculations can be performed once and reused
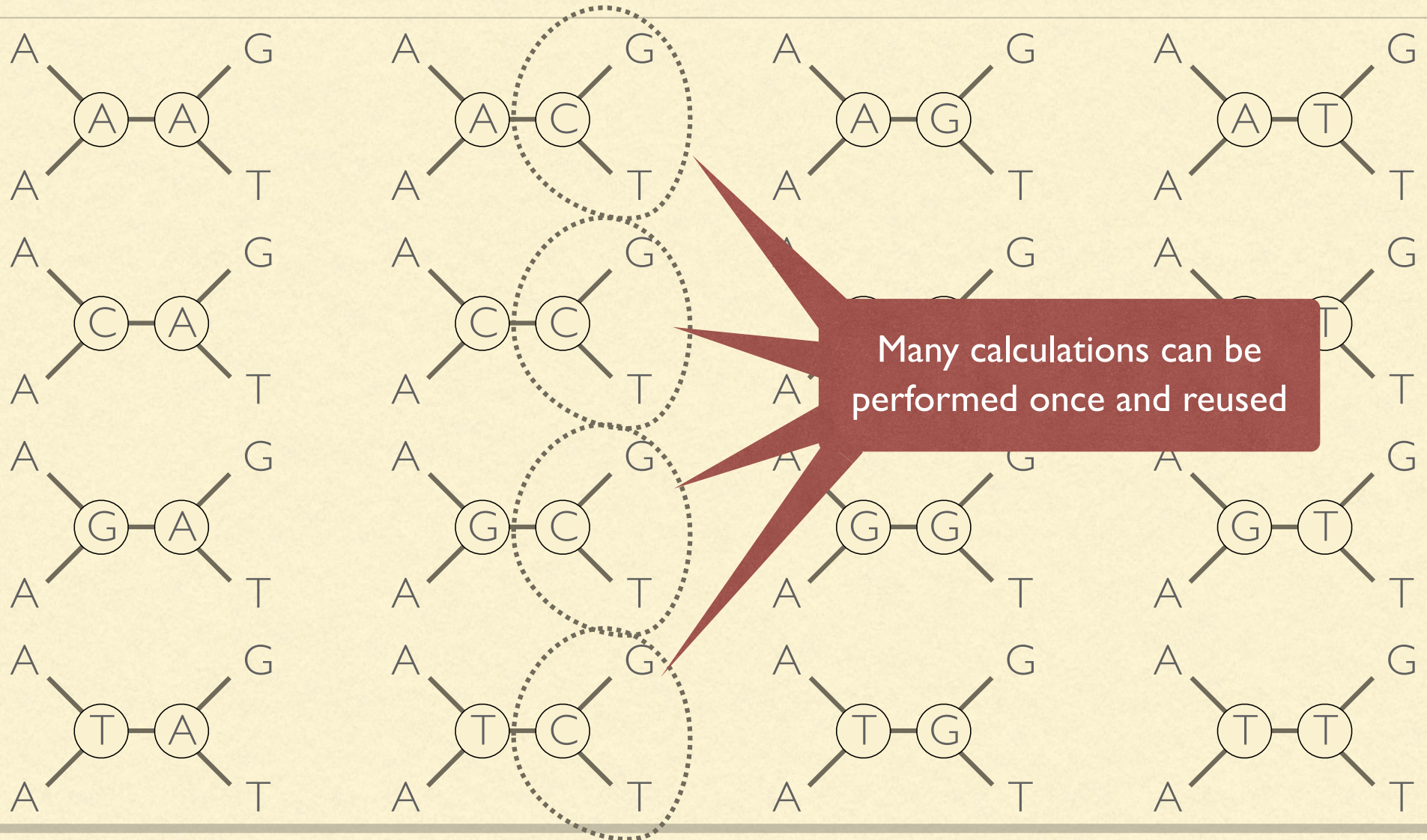
Pruning algorithm

Pruning algorithm

Many calculations can be performed once and reused

Pruning algorithm

Many calculations can be performed once and reused

Pruning algorithm

Many calculations can be performed once and reused

Pruning algorithm

Many calculations can be performed once and reused
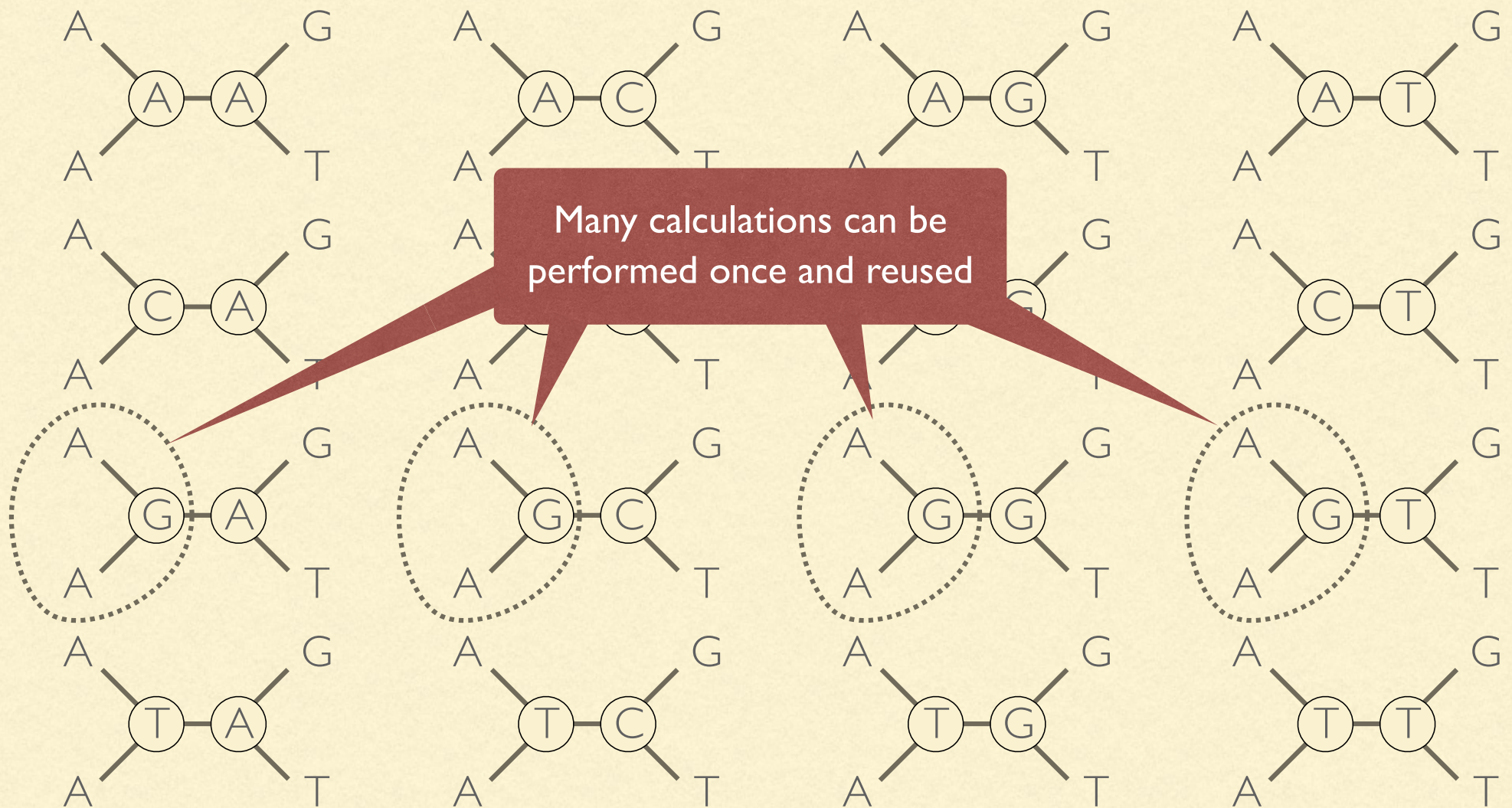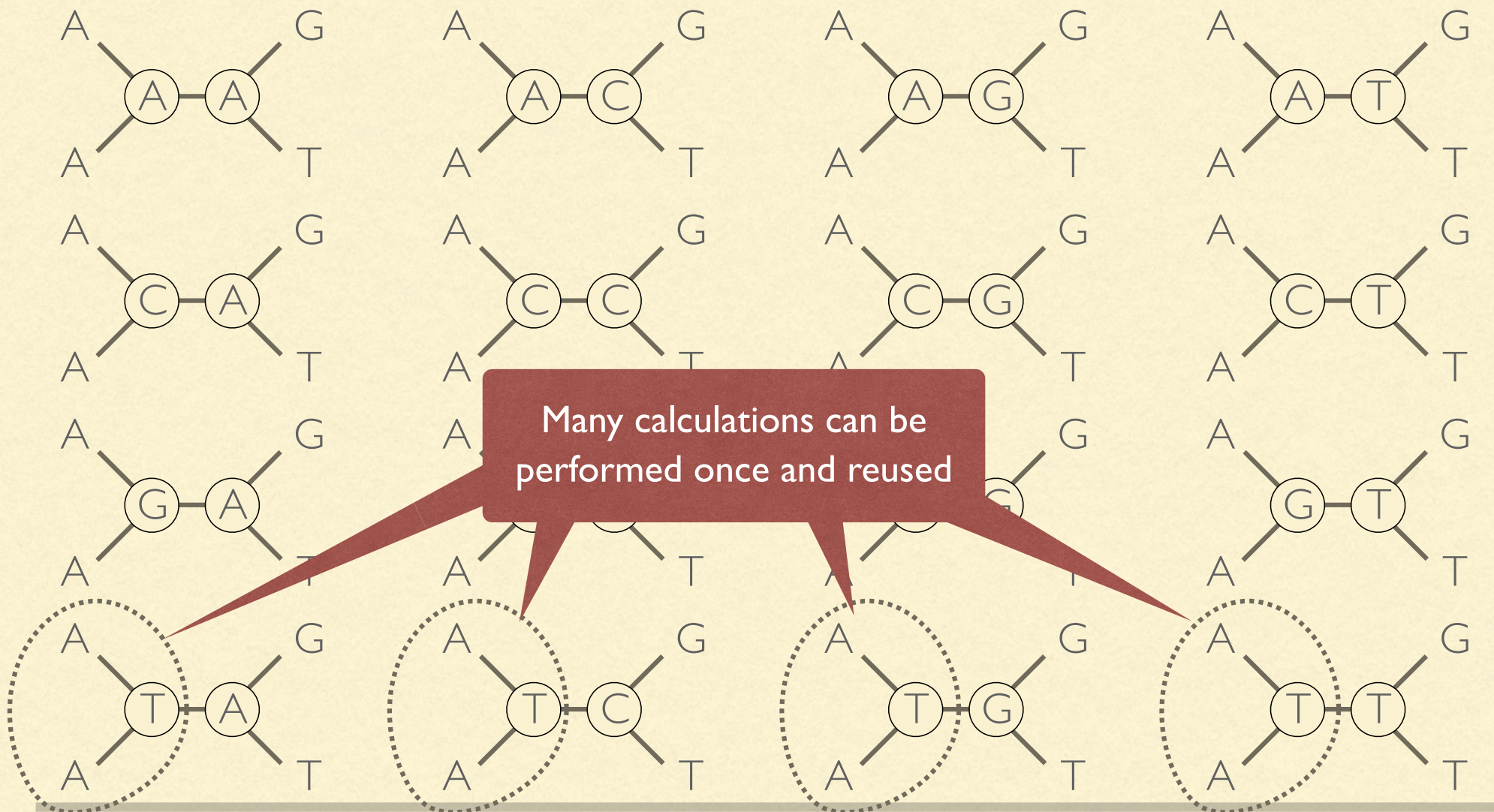
Total likelihood

$$L = L_1 L_2 \cdots L_n$$

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: β

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

Kimura (1980)

K80 (or K2P) model   Parameters: α, β

$$
\begin{array}{c|cccc}
 & A & C & G & T \\
\hline
A & -\alpha - 2\beta & \beta & \alpha & \beta \\
C & \beta & -\alpha - 2\beta & \beta & \alpha \\
G & \alpha & \beta & -\alpha - 2\beta & \beta \\
T & \beta & \alpha & \beta & -\alpha - 2\beta
\end{array}
$$

Kimura (1980)

K80 (or K2P) model $\qquad \kappa = \alpha/\beta \qquad$ Parameters: κ, β

$$
\begin{array}{c c c c c}
 & A & C & G & T \\
A & -\beta(\kappa+2) & \beta & \kappa\beta & \beta \\
C & \beta & -\beta(\kappa+2) & \beta & \kappa\beta \\
G & \kappa\beta & \beta & -\beta(\kappa+2) & \beta \\
T & \beta & \kappa\beta & \beta & -\beta(\kappa+2)
\end{array}
$$

Transition-transversion (rate) ratio

C   T

transversions

β β

A

transition α

G

transition rate = α

transversion rate = β

assume α **=** β

transition-transversion rate ratio = 1.0

transition-transversion ratio = 0.5

Felsenstein (1981)

F81 model   Parameters: μ, π_A, π_C, π_G

$$
\begin{array}{c c c c c}
 & A & C & G & T \\
A & -\mu(1 - \pi_A) & \pi_C\mu & \pi_G\mu & \pi_T\mu \\
C & \pi_A\mu & -\mu(1 - \pi_C) & \pi_G\mu & \pi_T\mu \\
G & \pi_A\mu & \pi_C\mu & -\mu(1 - \pi_G) & \pi_T\mu \\
T & \pi_A\mu & \pi_C\mu & \pi_G\mu & -\mu(1 - \pi_T)
\end{array}
$$

JC69 is a special case of F81

$$
\begin{array}{cccc}
 & \text{A} & \text{C} & \text{G} & \text{T} \\
\text{A} & -\frac{3}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu \\
\text{C} & \frac{1}{4}\mu & -\frac{3}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu \\
\text{G} & \frac{1}{4}\mu & \frac{1}{4}\mu & -\frac{3}{4}\mu & \frac{1}{4}\mu \\
\text{T} & \frac{1}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu & -\frac{3}{4}\mu
\end{array}
\qquad
\begin{array}{cccc}
\text{A} & \text{C} & \text{G} & \text{T} \\
-3\beta & \beta & \beta & \beta \\
\beta & -3\beta & \beta & \beta \\
\beta & \beta & -3\beta & \beta \\
\beta & \beta & \beta & -3\beta
\end{array}
$$

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

Hasegawa, Kishino, and Yano (1985)

HKY85 model

Parameters: μ, $\boxed{\kappa, \pi_A, \pi_C, \pi_G}$

these are global parameters (apply to all edge lengths)

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

$$
\begin{array}{ccccc}
 & A & C & G & T \\
A & -\mu\left(\pi_C + \pi_G\kappa + \pi_T\right) & \pi_C\mu & \pi_G\mu\kappa & \pi_T\mu \\
C & \pi_A\mu & -\mu\left(\pi_A + \pi_G + \pi_T\kappa\right) & \pi_G\mu & \pi_T\mu\kappa \\
G & \pi_A\mu\kappa & \pi_C\mu & -\mu\left(\pi_A\kappa + \pi_C + \pi_T\right) & \pi_T\mu \\
T & \pi_A\mu & \pi_C\mu\kappa & \pi_G\mu & -\mu\left(\pi_A + \pi_C\kappa + \pi_G\right)
\end{array}
$$

Hasegawa, Kishino, and Yano (1985)

HKY85 model

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

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

$$
\begin{array}{c@{\quad}c@{\quad}c@{\quad}c@{\quad}c}
 & A & C & G & T \\
A & -\mu\left(\pi_C + \pi_G\kappa + \pi_T\right) & \boxed{\pi_C\mu} & \boxed{\pi_G\mu\kappa} & \boxed{\pi_T\mu} \\
C & \pi_A\mu & -\mu\left(\pi_A + \pi_G + \pi_T\kappa\right) & \pi_G\mu & \pi_T\mu\kappa \\
G & \pi_A\mu\kappa & \pi_C\mu & -\mu\left(\pi_A\kappa + \pi_C + \pi_T\right) & \pi_T\mu \\
T & \pi_A\mu & \pi_C\mu\kappa & \pi_G\mu & -\mu\left(\pi_A + \pi_C\kappa + \pi_G\right)
\end{array}
$$

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

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