

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

Likelihood in Phylogenetics

Workshop on Molecular Evolution
Woods Hole, Massachusetts

23 May 2025

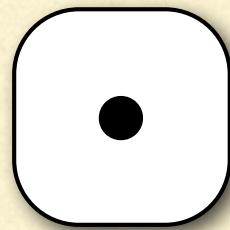
Paul O. Lewis
Department of Ecology & Evolutionary Biology



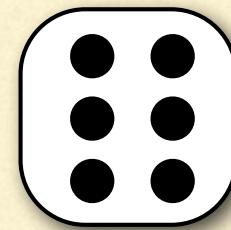
Probability

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



$$(1/6)$$

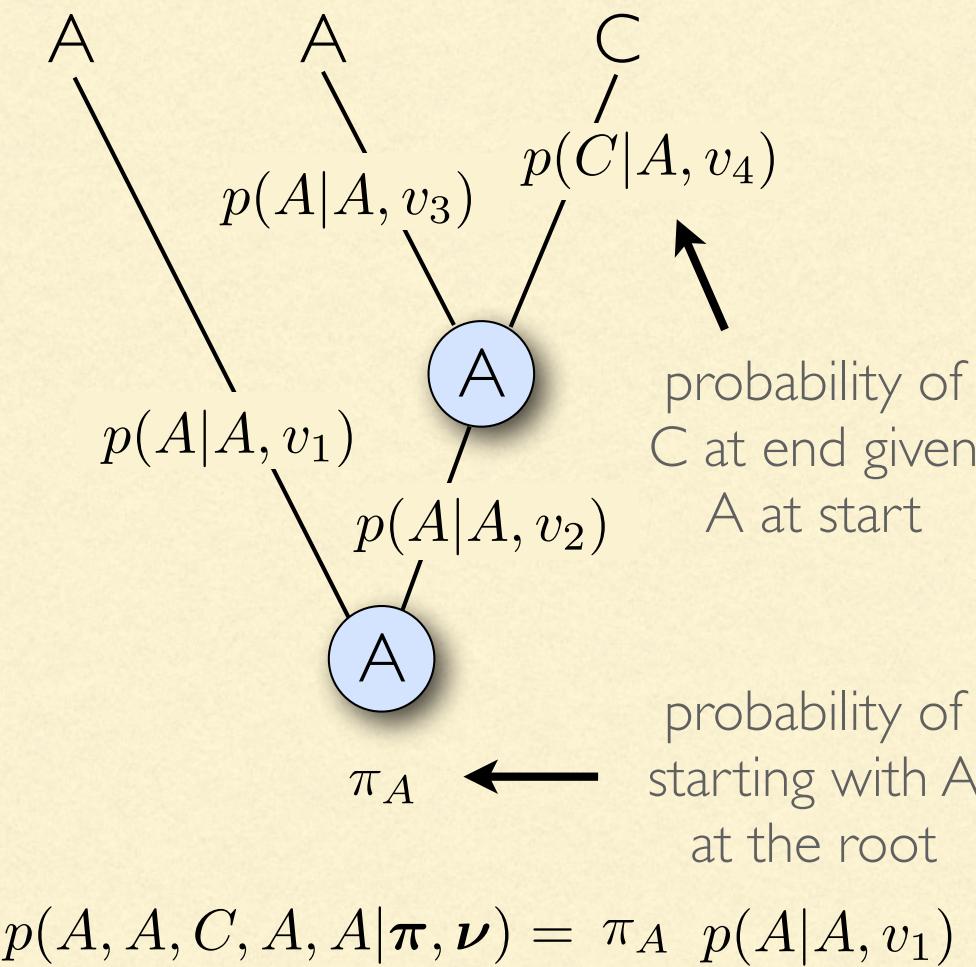
×

$$(1/6)$$

=

$$1/36$$

AND rule in phylogenetics

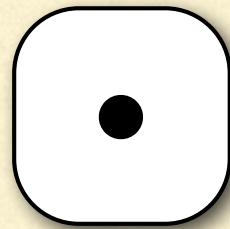


One use of the AND rule in phylogenetics is to combine probabilities associated with individual branches to produce the overall probability of the data for one site.

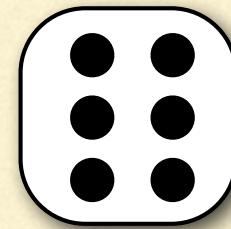
Note: π is just a symbol here; no relation to the famous constant

Probabilities: the OR rule

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



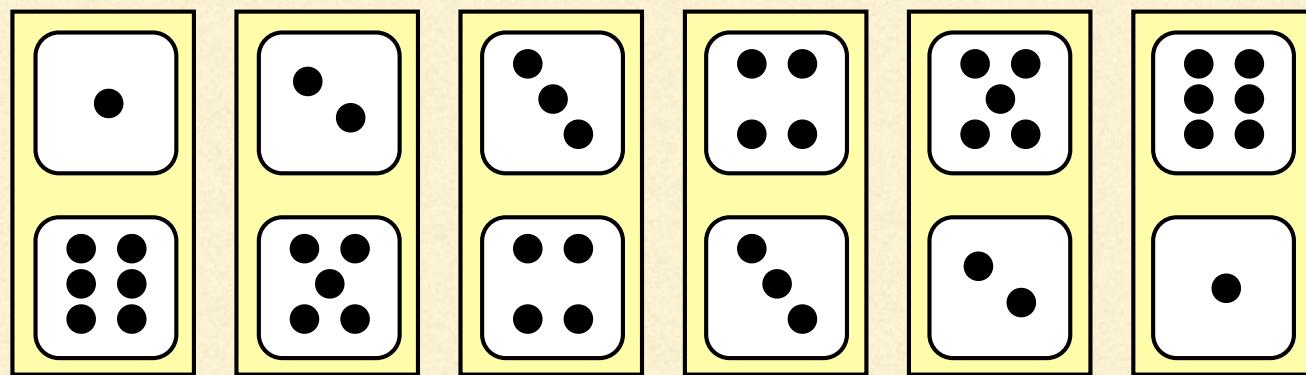
OR



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

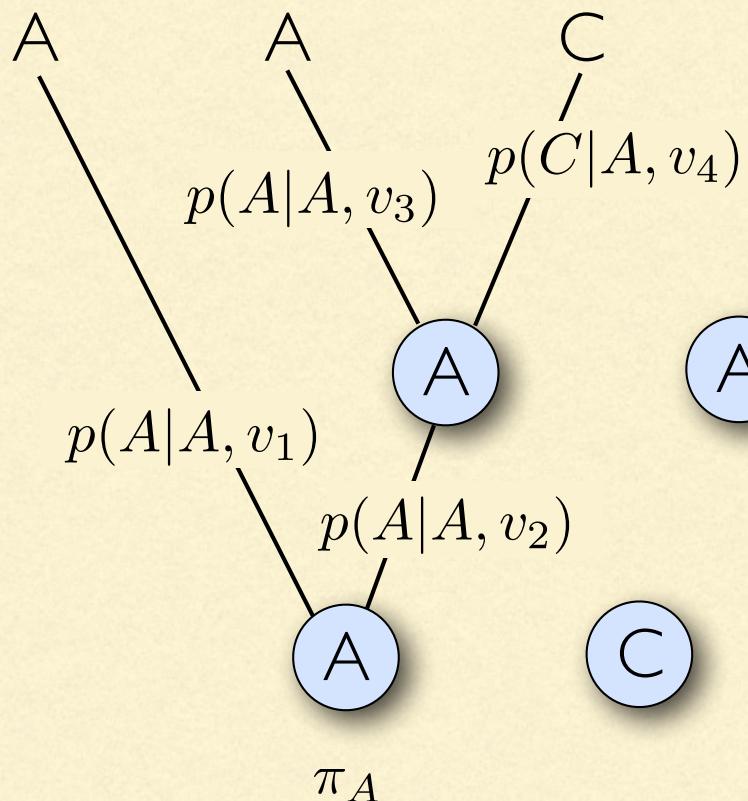
Combining AND and OR

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



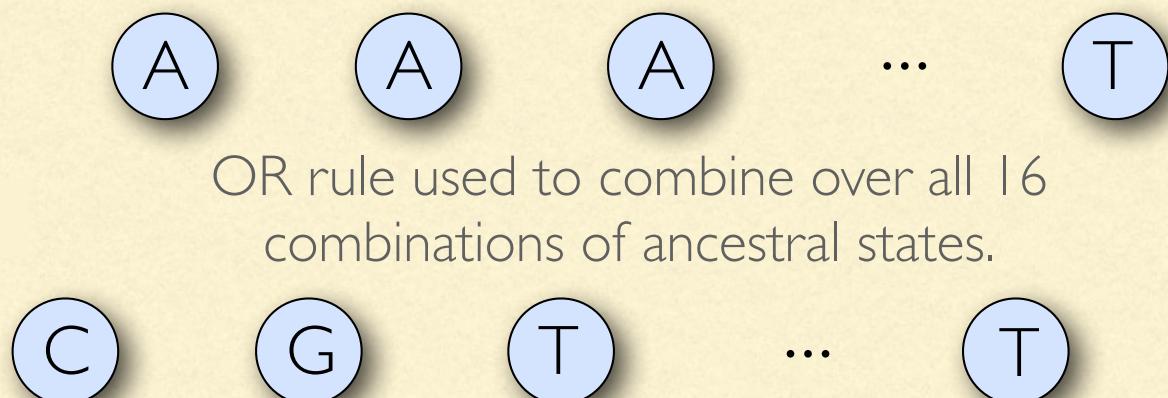
$$(1/36) + (1/36) + (1/36) + (1/36) + (1/36) + (1/36) = 1/6$$

Using both AND and OR in phylogenetics



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.



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

Independence

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

Probability of flipping 2 fair coins and
getting heads on both:

$$p(H,H) = p(H) \ p(H)$$

Non-independence

$$p(A, B) = p(A) p(B|A)$$

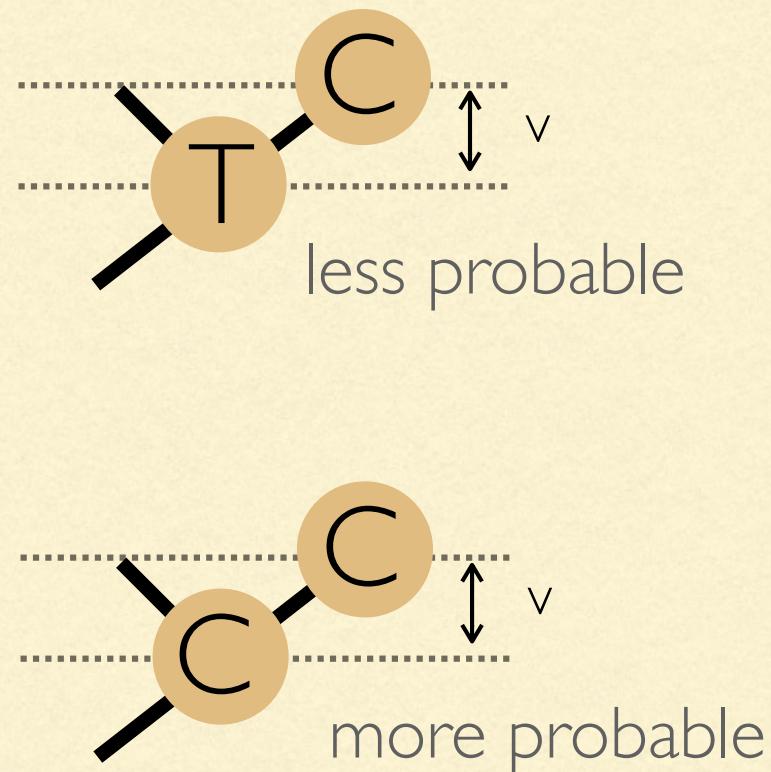
↑ ↑
joint probability conditional probability
of A and B of B given A

$$p(\text{walk to work}|\text{sunny}) = 0.99$$
$$p(\text{walk to work}|\text{rainy}) = 0.50$$

Non-independence in phylogenies

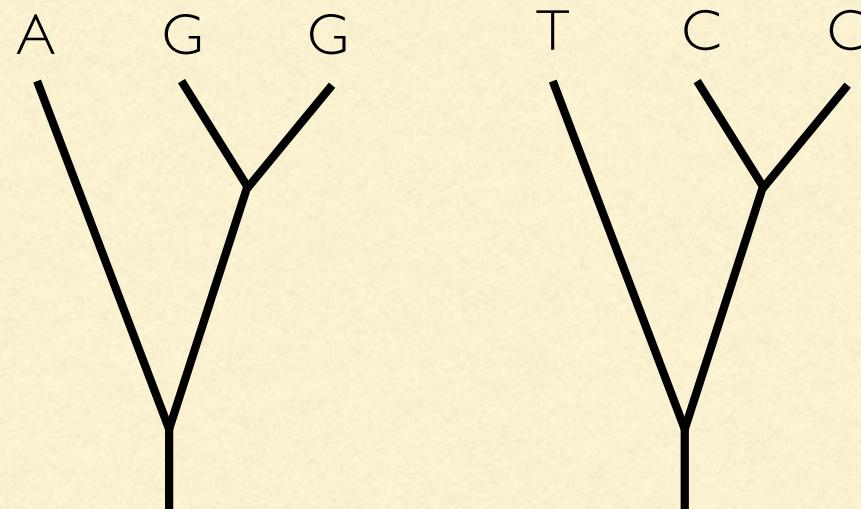
Normally, for a given rate of substitution and time, the probability of the end state is dependent on the starting state

$$p(C|C, v) > p(C|T, v)$$



Conditional Independence

$$p(A,B|C) = p(A|C) p(B|C)$$



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

Likelihood

Likelihood tells you **how surprised**
you should be at the observed data

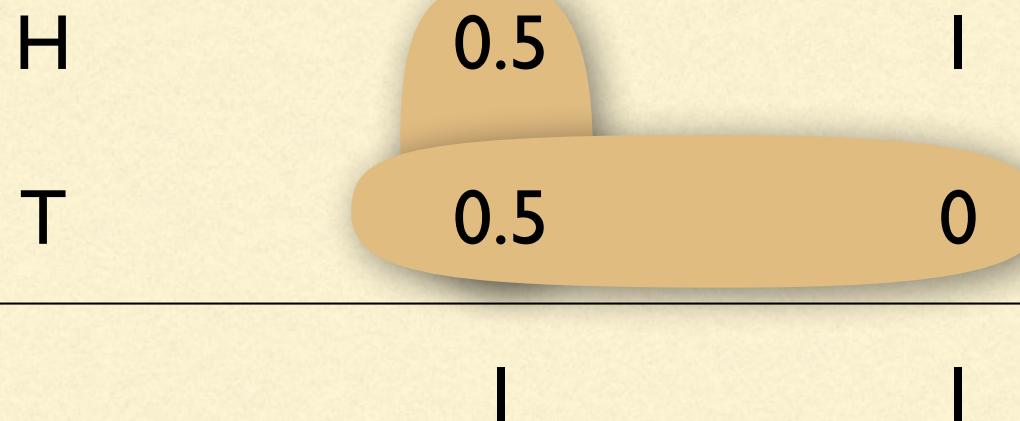
High likelihood \Rightarrow **less** surprised

Low likelihood \Rightarrow **more** surprised

Let's take a second to
test your knowledge of
likelihood so far...

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

Outcome	Fair coin model	Two-heads model
---------	-----------------	-----------------



Probabilities of data outcomes given one particular model sum to 1.0

Likelihoods of models given one particular data outcome 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 $\Psi\eta$ -globin gene of gorilla:

● **GAAGTCCTTGAGAAATAAACTGCACACACTGG**

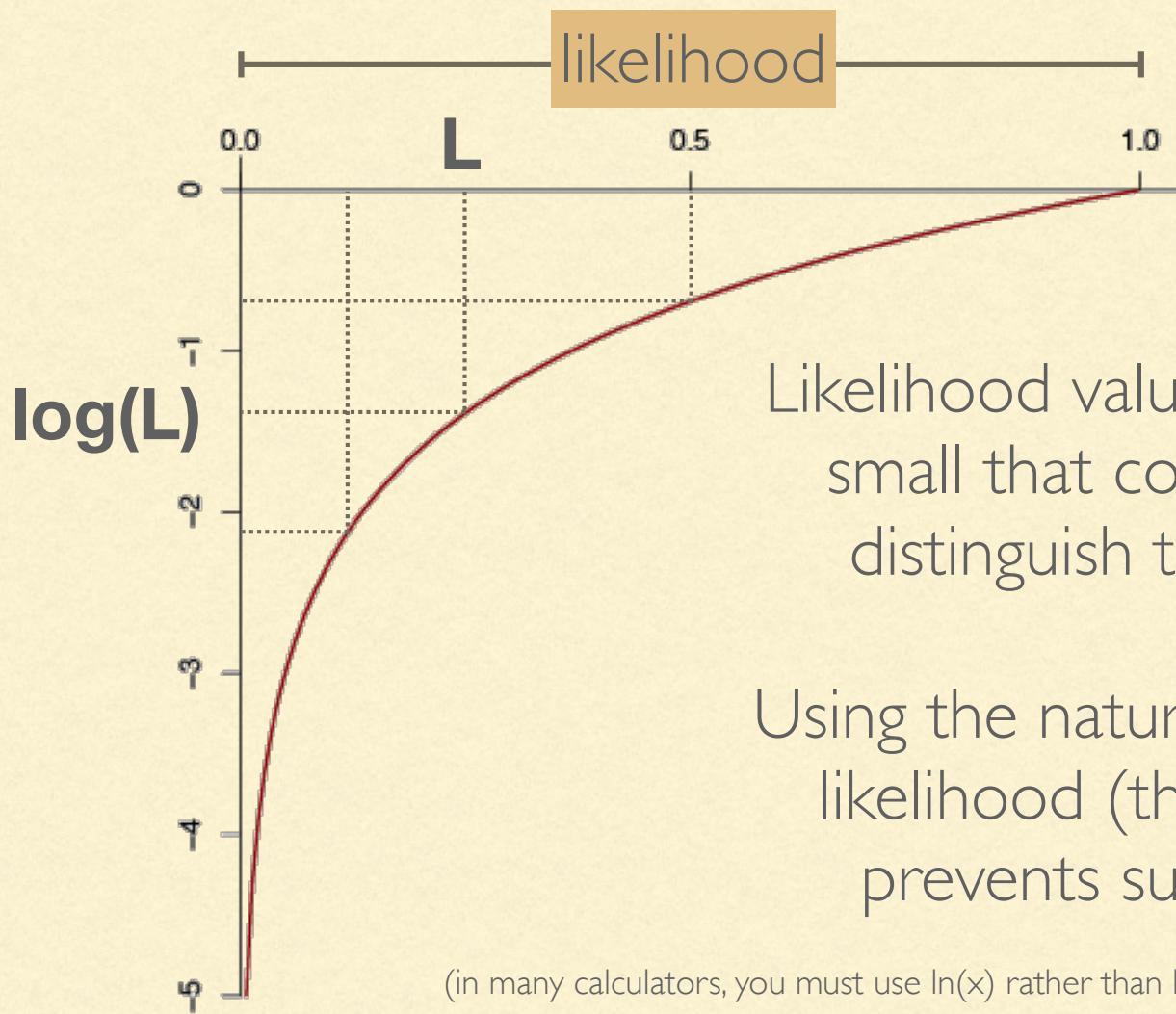
$$L = p(G) \quad p(A) \quad p(A) \quad p(G) \quad p(T) \quad \dots \quad p(G)$$

$$L = \pi_G \pi_A \pi_A \pi_G \pi_T \dots \pi_G$$

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

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

Natural logarithm



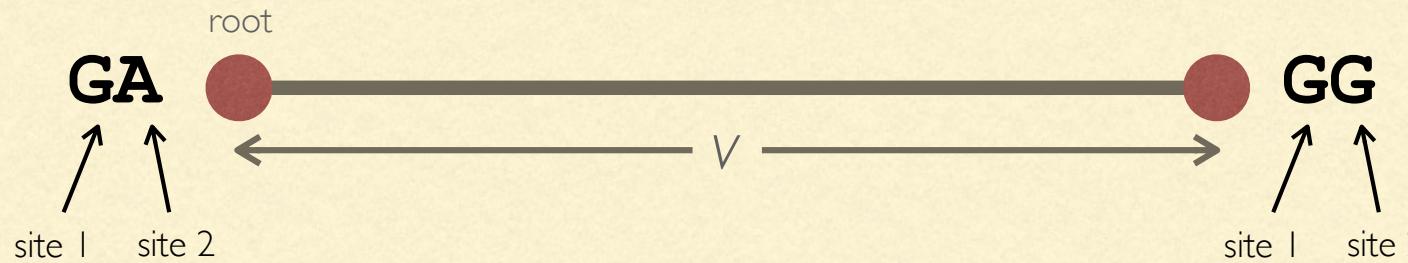
Likelihood values can become so small that computers cannot distinguish them from zero.

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)



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

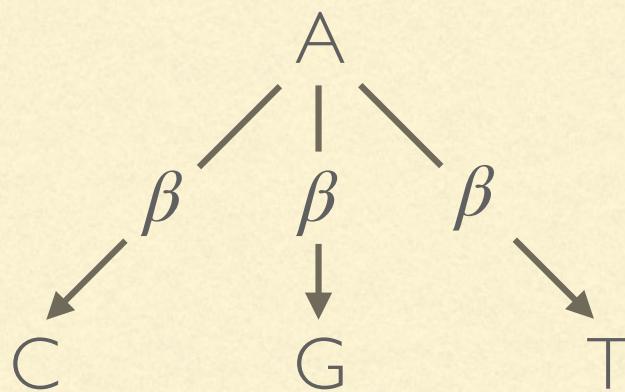
\longleftrightarrow site 1 \longleftrightarrow site 2 \longleftrightarrow

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

What is the edge length v ?

expected
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 β

Jukes and Cantor (1969)

JC69 model

to:

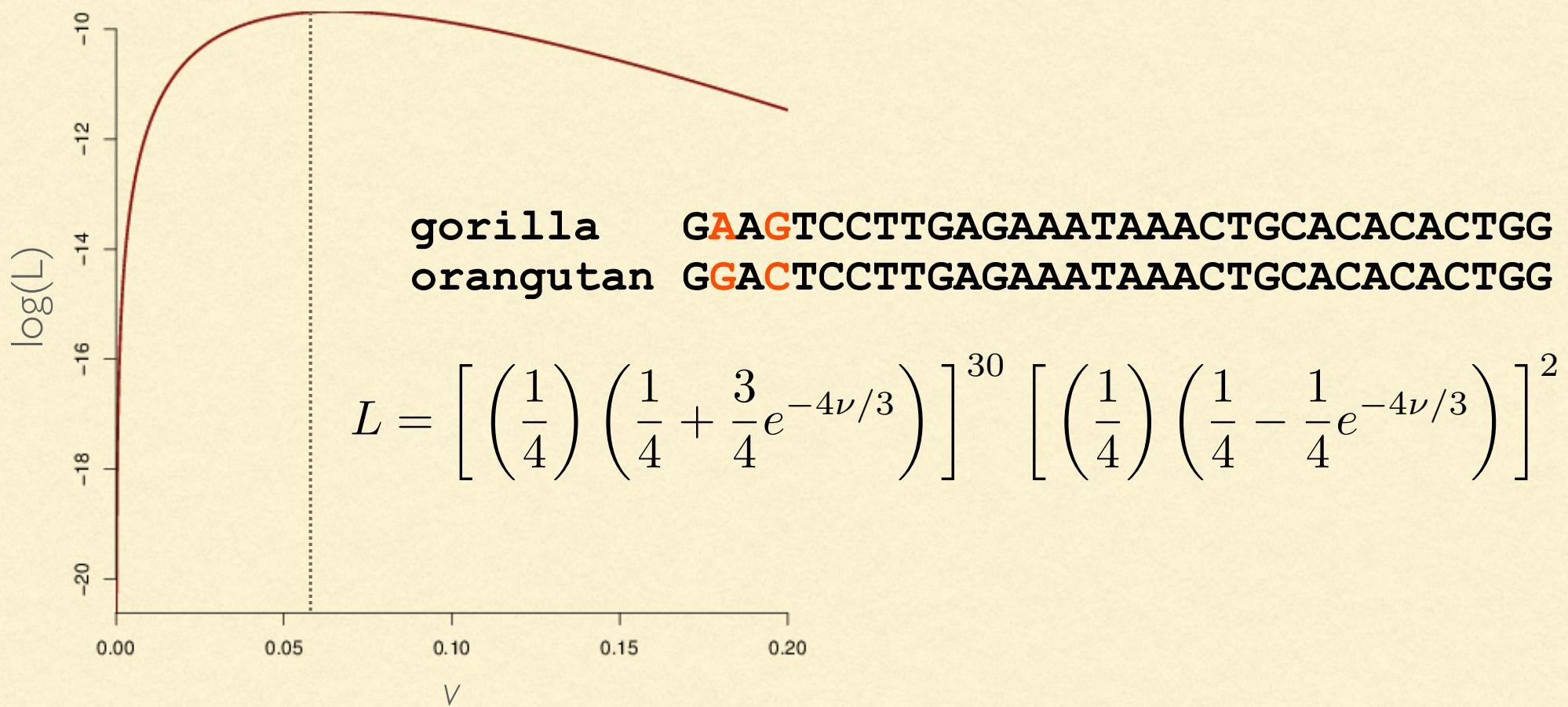
Parameters: β

from:

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

Maximum likelihood estimation

0.065 is the maximum likelihood estimate (MLE) of ν

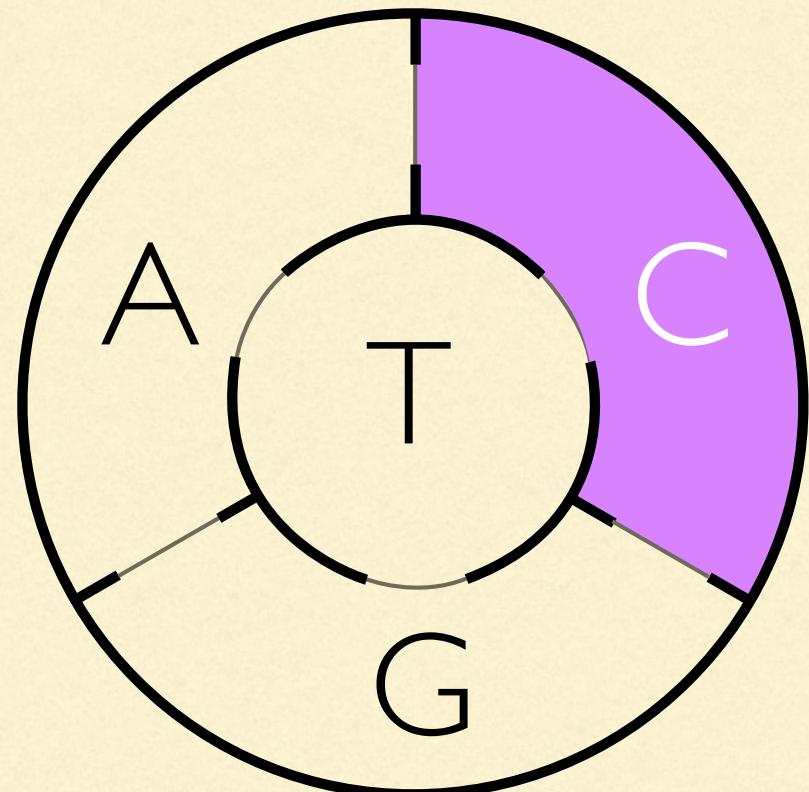


Equilibrium Frequencies

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...

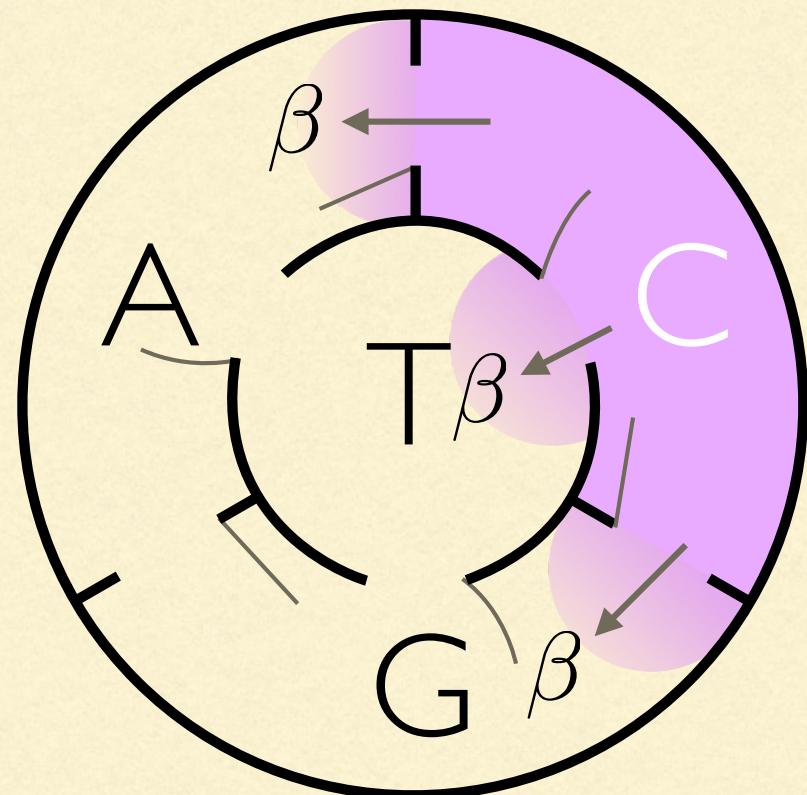
enter room A at rate β

enter room T at rate β

enter room G at rate β

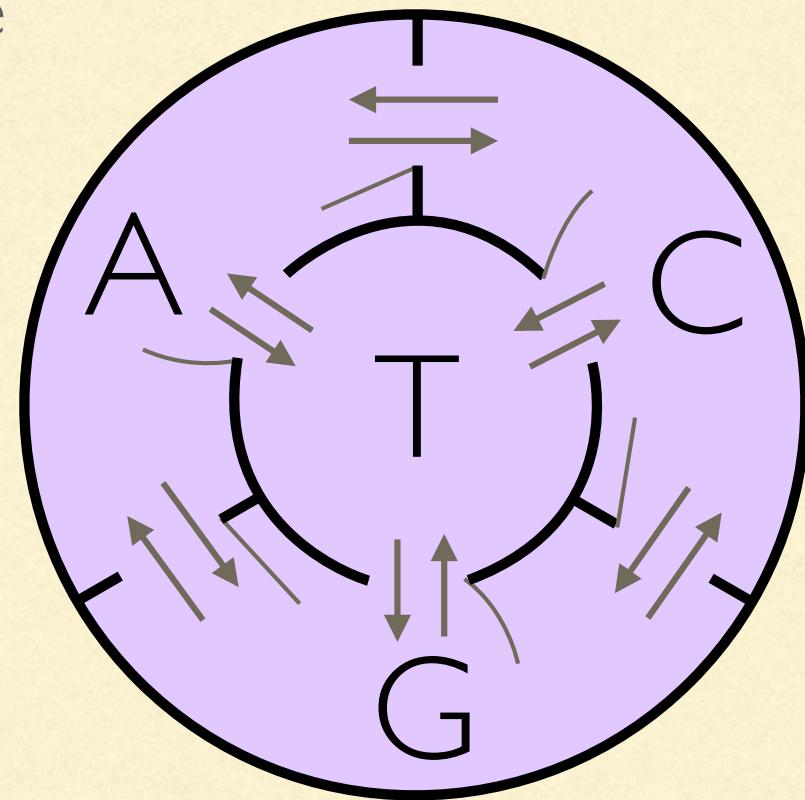
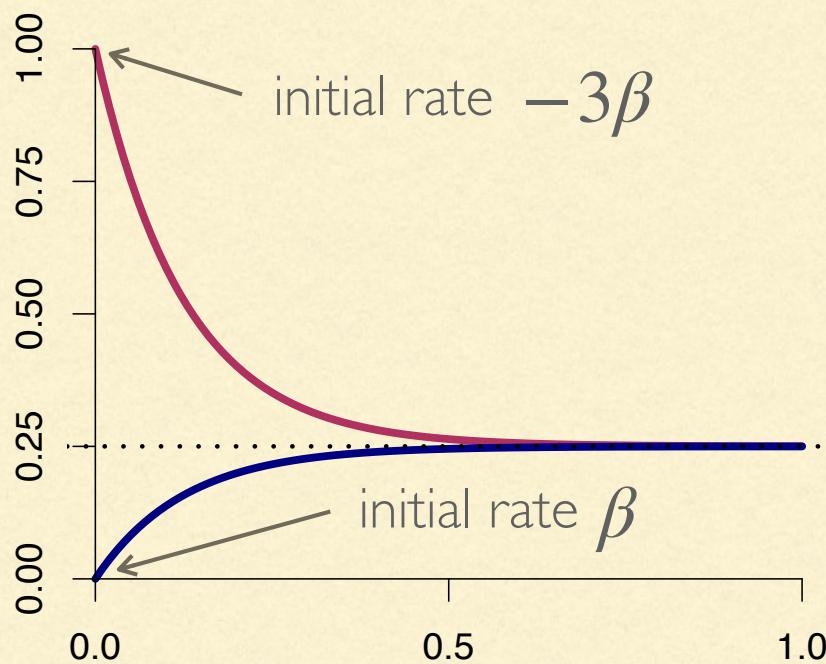
enter room C at rate -3β

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



Equilibrium Frequencies

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



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

Transition probability demo

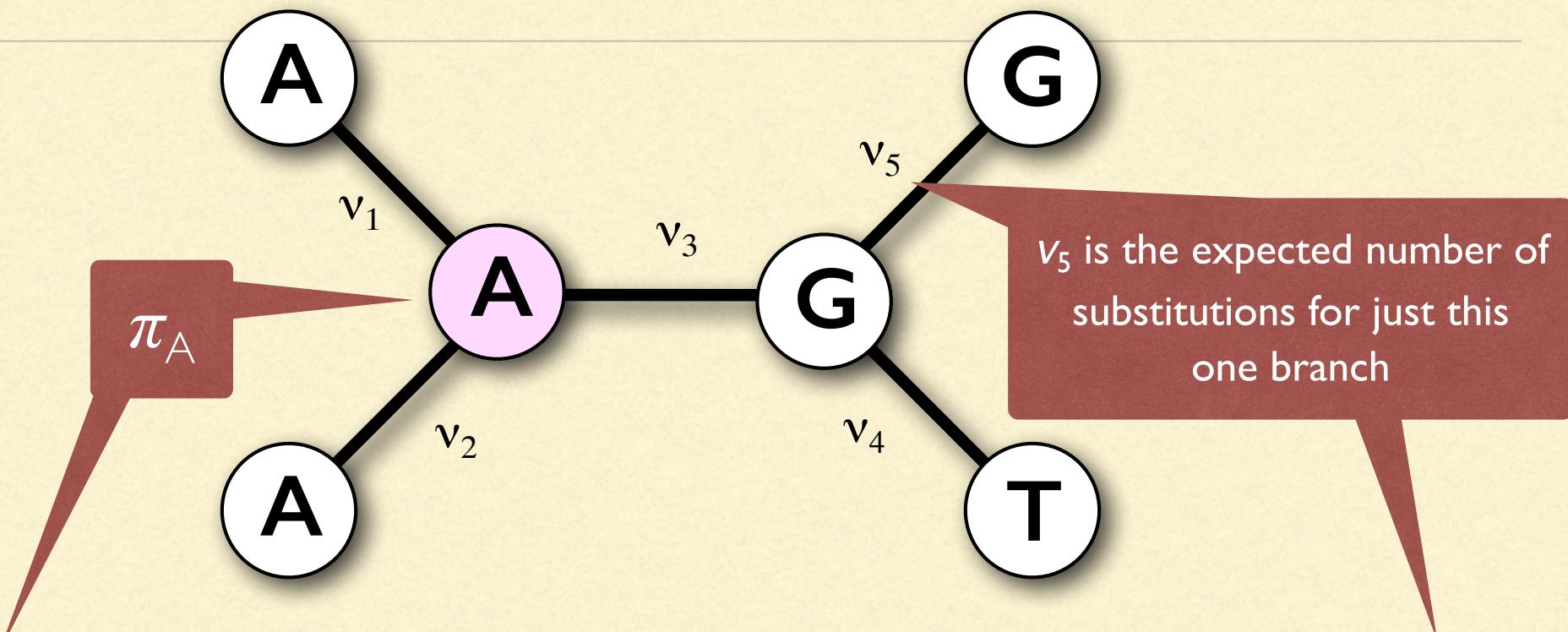
<https://plewis.github.io/applets/jc-transition-probabilities/>

Sequence data for four taxa

Sphagnum	GGCAGCATTTCGAATGACTCCTCAACCTGGAGT	A	CCACCCG...
Asplenium	GGCAGCTTCCGGATGACCCCACAACCCGGAGT	A	CAGCTG...
Picea	GGCAGCATTCCGAGTAACTCCTCAACCAGGGGT	G	CGCCCCG...
Avena	GGCAGCATTCCGAGTAACTCCTCAACCTGGGGT	T	CGCCGG...

one site

Likelihood for tree (one site)



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

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

from:

$$\begin{matrix} & \text{A} & \text{C} & \text{G} & \text{T} \\ \text{A} & -3\beta & \beta & \beta & \beta \\ \text{C} & \beta & -3\beta & \beta & \beta \\ \text{G} & \beta & \beta & -3\beta & \beta \\ \text{T} & \beta & \beta & \beta & -3\beta \end{matrix}$$

Kimura (1980)

K80 (or K2P) model

Parameters: α, β

	A	C	G	T
A	$-\alpha - 2\beta$	β	α	β
C	β	$-\alpha - 2\beta$	β	α
G	α	β	$-\alpha - 2\beta$	β
T	β	α	β	$-\alpha - 2\beta$

Kimura (1980)

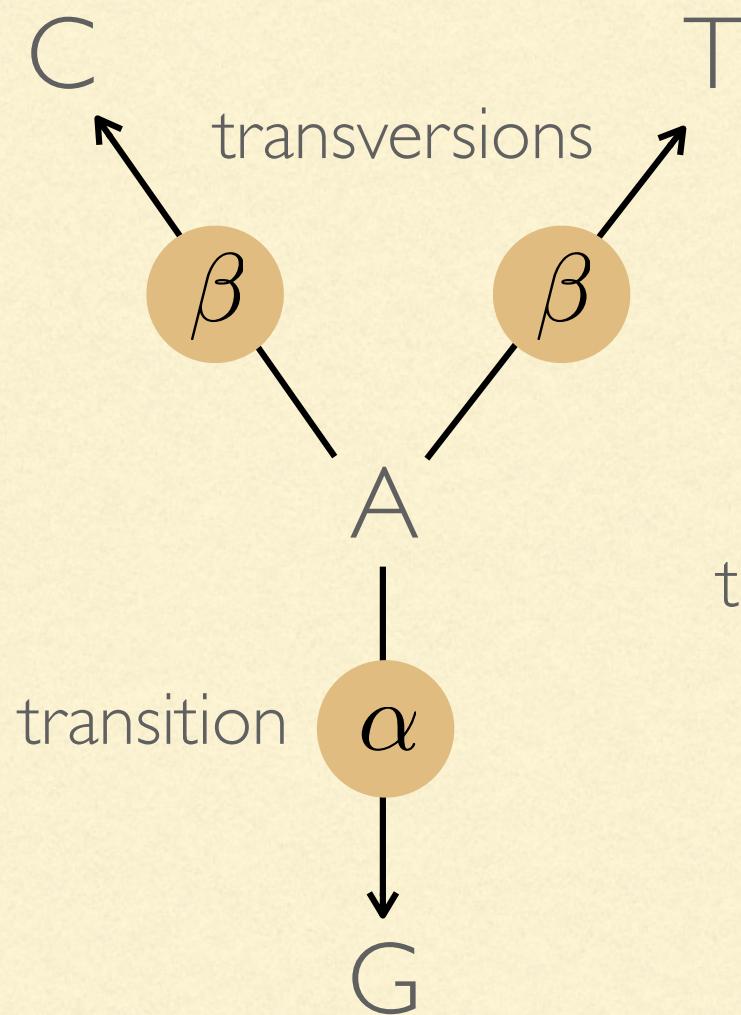
K80 (or K2P) model

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

Parameters: κ, β

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

Transition-transversion (rate) ratio



transition rate = α

transversion rate = β

assume $\alpha = \beta$

transition-transversion rate ratio = 1.0

transition-transversion ratio = 0.5

Felsenstein (1981)

F81 model

Parameters: μ, π_A, π_C, π_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 F8 |

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

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

Hasegawa, Kishino, and Yano (1985)

HKY85 model

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



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

these are global parameters
(apply to all edge lengths)

	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	$\pi_A\mu$	$\pi_C\mu\kappa$	$\pi_G\mu$	$-\mu(\pi_A + \pi_C\kappa + \pi_G)$

Tavaré (1986)

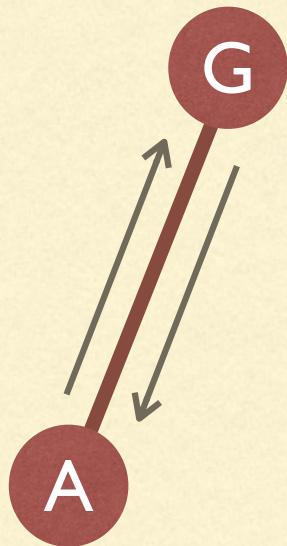
GTR model

Parameters: ?

	A	C	G	T
A	—	$\pi_C \mu a$	$\pi_G \mu b$	$\pi_T \mu c$
C	$\pi_A \mu a$	—	$\pi_G \mu d$	$\pi_T \mu e$
G	$\pi_A \mu b$	$\pi_C \mu d$	—	$\pi_T \mu f$
T	$\pi_A \mu c$	$\pi_C \mu e$	$\pi_G \mu f$	—

exchangeability
parameters are
circled

Time reversibility



Time reversibility means...

$$\Pr(A) \Pr(G|A, v) = \Pr(G) \Pr(A|G, v)$$

Time reversibility allows any point on the tree to serve as the root, and thus has some practical advantages, but time reversibility is not a requirement for substitution models used in phylogenetics

Rate heterogeneity

Green plant *rbcL* gene

First 88 amino acids (translation is for *Zea mays*)

M--S--P--Q--T--E--T--K--A--S--V--G--F--K--A--G--V--K--D--Y--K--L--T--Y--T--P--E--Y--E--T--K--D--T--D--I--L--A--A--F--R--V--T--P--	
Chara (green alga; land plant lineage)	AAAGATTACAGATTAACCTACTATACTCCTGAGTATAAAACTAAAGATACTGACATTTAGCTGCATTCGTGTAACCCA
Chlorella (green alga)C...C.T.....T..CC..C.A.....C.....T...C.T..A..G..C..A.G.....T
Volvox (green alga)TC.T.....A.....C..A.....C..GT.GTA.....C.....C.....A.....A.G.....
Conocephalum (liverwort)TC.....T.....G..T.....G.....G..T.....A.....A..AA.G.....T
Bazzania (moss)T.....C..T.....G.....A..G.G..C.....G..A..T.....G..A.....A.G.....C
Anthoceros (hornwort)T.....CC.T.....C.....T..CG.G..C..G.....T.....G..A..G.C.T..AA.G.....T
Osmunda (fern)TC.....G..C.....C..T.....G.G..C..G.....T.....G..A.....C..AA.G.....C
Lycopodium (club "moss")	.GG.....C.T..C.....T.....G..C.....A..C..T..C.G..A.....AA.G.....T
Ginkgo (gymnosperm; <i>Ginkgo biloba</i>)G.....T.....A..C..C.....T..C..G..A.....C..A.....T
Picea (gymnosperm; spruce)T.....A..C.G..C.....G..T.....G..A.....C..A.....T
Iris (flowering plant)G.....T.....T..CG..C.....T..C..G..A.....C..A.....T
Asplenium (fern; spleenwort)TC..C.G.....T..C..C..C..A..C..G..C.....C..T..C..G..A..T..C..GA.G..C..
Nicotiana (flowering plant; tobacco)G..A.....T.....CC..C..G.....T..A..G..A.....C..A.....T
Q--L--G--V--P--E--E--A--G--A--V--A--E--S--S--T--G--T--W--T--T--V--W--T--D--G--L--T--S--L--D--R--Y--K--G--R--C--Y--H--I--E--	
CAACCTGGCGTTCCACCTGAAGAACGGGGCTGCAGTAGCTGCAGAACATCTTCTACTGGTACATGGACTACTGTTGGACTGACGGATTAACTAGTTGGACCGATACAAAGGAAGATGCTACGATATTGAA	
....A..T.....A.....G..T..G.....A.....A..A.....T.....G..A.....T..T.....A.....T.....TC..T..T..T..C..C..G	
....A..T.....TGT..T.....T..T..T.....A..A..A.....T.....A..A.....T..T.....A..C..T.....T.....TC..T..T..T..C..C..G	
..G...G..A.....G..A.....A..A.....T.....T.....A.....A.....T..TC..T.....ACC..T..T..T..T.....TC.....T..G.....C	
....G..A..A.....A..G.....T.....A..C.....G.....C..G.....C..T..GC..T..A..C..C..C..T..T.....TC.....T..C..C..	
T....A..G..G.....A..C.....T.....A.....A.....C..T..C..CC..T.....T.....TC.....C..C..	
....C..A..A..GG..G..T..A.....G.....A..G.....C.....A.....G..T..C..T..C..C..T..T..T..G..TC.....C..C..	
....T..A..A.....C..G..G..A..C.....T.....C.....A.....C..T..C..T..C..C..C..T..C.....TC..G.....T..A..	
....A..G.....G..G..A.....C.....C.....C.....C.....C..T..C..T..C..C..T..T..T.....G.....T..C..C..G	
....A..G..G..G..C..G..G..A..A.....T.....C..C.....C.....C..T..C..T..C..T..T..T.....G..GC.....T..C..C..G	
....C..A.....TG.....G..C..G.....C.....A.....A..A..G.....T..C..T..C..C..T..T..T.....C.....C..C..C..G	
....C..A..A..G.....C..A.....G..C.....A.....C..G.....A.....A..G..G..G..C..CC..T.....T..G..CC.....C..C..G	
....A.....C..G.....C.....A.....A.....C..T..C..T..C..CC..T..T..T.....GC.....CGC..C..G	

All 4 bases are observed at some sites...

...while at other sites, only 1 base is observed

Site-specific rates

Each defined subset (e.g. 1st+2nd pos. versus 3rd pos.) has its own relative rate

CACCGGGTCCCCAGAGCGGGCGTGCAGATCTCACGGACTGACACGTTGACGAGGTTACAGTTGACGTAAAGGAGTGTAGAATGA	ATCTATAAAGTAATAATTAGTTGTACATTGCACAAACCTTA
.....TG.....	.AT...A..GTG..A..AA..T.G.A..TT..A.T..TTTTCCG
.....G.....	.AT.....TT..TT..T.AAA..T.A.A..TT..A..T..TTTTCCG
T.....	G.GA.A...AA..T..T.....A..TTT..CTTTT..T..T..C
.....G.....	.GAA.....AG..T..AC..G..CG..CGTTA..CTT..T..TCC.
....T.....	.AGG.....AC..T..A.....C..TTCT..T..T..C..
.....G.....	CAAG..G..TA.....G..A..G..C..A..G..TTC..TTTTGT..
.....G.....	..AA..CG..GAC..T..C.....C..TTC..CTC..TG..TA..
.....G.....	..AG..G..GA.....C..C.....C..C..TTC..TTT..G..TCCG
.....G.....	AGGGCG..GAA.....T..CC..C..C..TT..TTT..GG..TCCG
.....G.....	CA..T..G..CC..C.....AAG..TTC..TTT..CCG
.....G.....	CAA.....CA.....GC..A..C..G..AG..GCCT..T..GC..CG
.....	A.....CG..C.....A..A..C..TTCCTT..G..CCG

r_1 applies to subset 1
1st+2nd codon positions
(sites 1 - 88)

r_2 applies to subset 2
3rd codon positions
(sites 89-132)

Relative rates have mean 1.0: $r_1 p(r_1) + r_2 p(r_2) = 1$

2/3

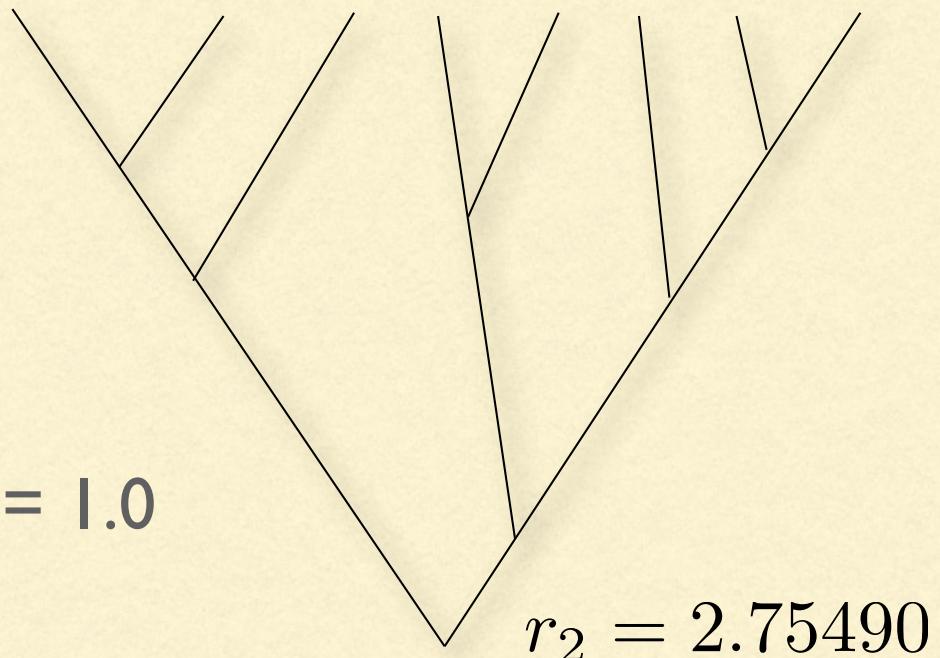
1/3

Site-specific rates

$$L = \underbrace{p(\mathbf{y}_1|r_1) \cdots p(\mathbf{y}_{88}|r_1)}_{\text{1st+2nd codon positions}} \underbrace{p(\mathbf{y}_{89}|r_2) \cdots p(\mathbf{y}_{132}|r_2)}_{\text{3rd codon positions}}$$

ψψ

$$r_1 = 0.12255$$



mean relative rate:

$$(0.12255)(2/3) + (2.75490)(1/3) = 1.0$$

$$r_2 = 2.75490$$

Site-specific rates

JC69 transition probabilities that would be used for every site if rate *homogeneity* were assumed:

$$P_{ii}(t) = \frac{1}{4} + \frac{3}{4}e^{-4\beta t} \quad \text{C} \xrightarrow{\text{identity}} \text{C}$$

$$P_{ij}(t) = \frac{1}{4} - \frac{1}{4}e^{-4\beta t} \quad \text{C} \xrightarrow{\text{difference}} \text{T}$$

Site specific rates

JC69 transition probabilities that would be used for sites in **subset 1**:

$$P_{ii}(t) = \frac{1}{4} + \frac{3}{4}e^{-4r_1\beta t}$$

$$P_{ij}(t) = \frac{1}{4} - \frac{1}{4}e^{-4r_1\beta t}$$

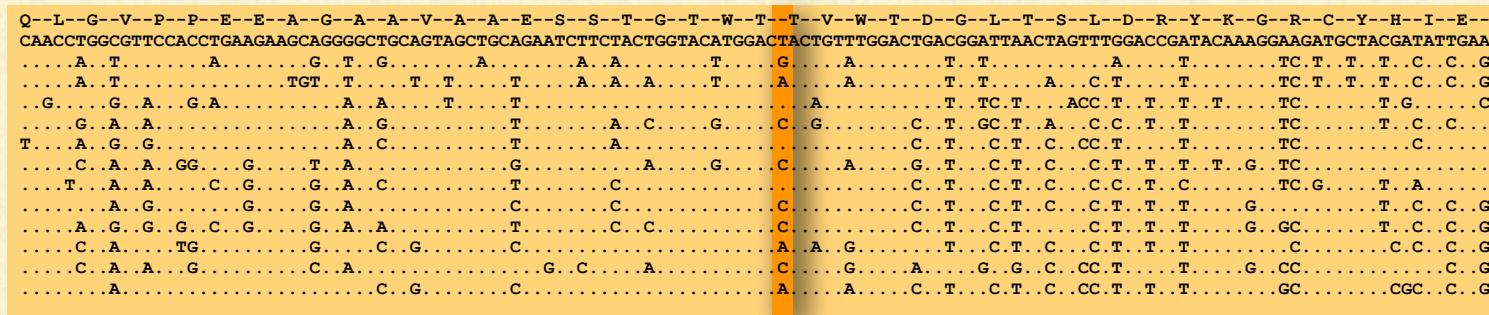
JC69 transition probabilities that would be used for sites in **subset 2**:

$$P_{ii}(t) = \frac{1}{4} + \frac{3}{4}e^{-4r_2\beta t}$$

$$P_{ij}(t) = \frac{1}{4} - \frac{1}{4}e^{-4r_2\beta t}$$

Mixture models

All k relative rates applied to every site



$$L_i = p(\mathbf{y}_i | r_1)p(r_1) + p(\mathbf{y}_i | r_2)p(r_2) + \dots + p(\mathbf{y}_i | r_k)p(r_k)$$

Common examples

{ Invariable sites (I) model
Discrete Gamma (G) model

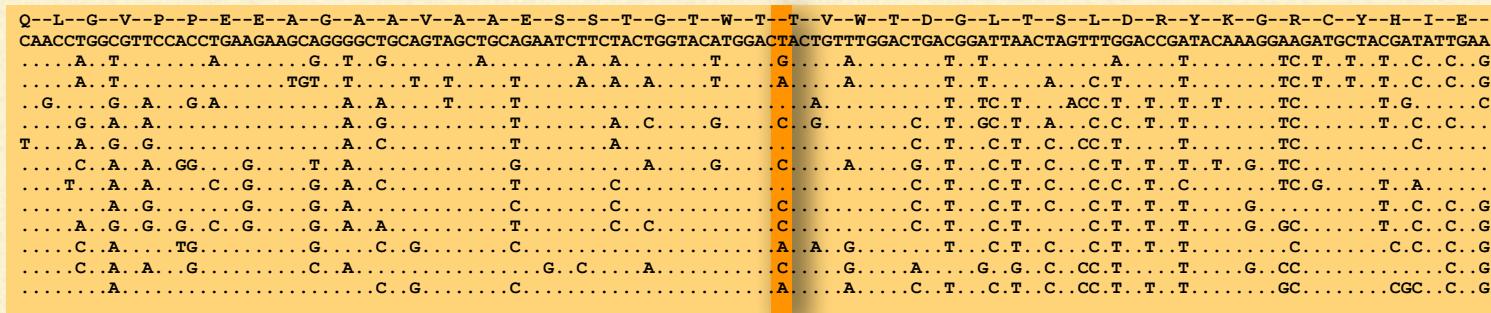
Invariable sites model (Reeves 1992)

$$L_i = p(\mathbf{y}_i|r_1)p(r_1) + p(\mathbf{y}_i|r_2)p(r_2)$$

$$L_i = p(\mathbf{y}_i|0.0)p_{\text{invar}} + p(\mathbf{y}_i|r_2)(1 - p_{\text{invar}})$$

Discrete Gamma model (Yang 1994)

No relative rate is exactly 0.0, and all are equally probable



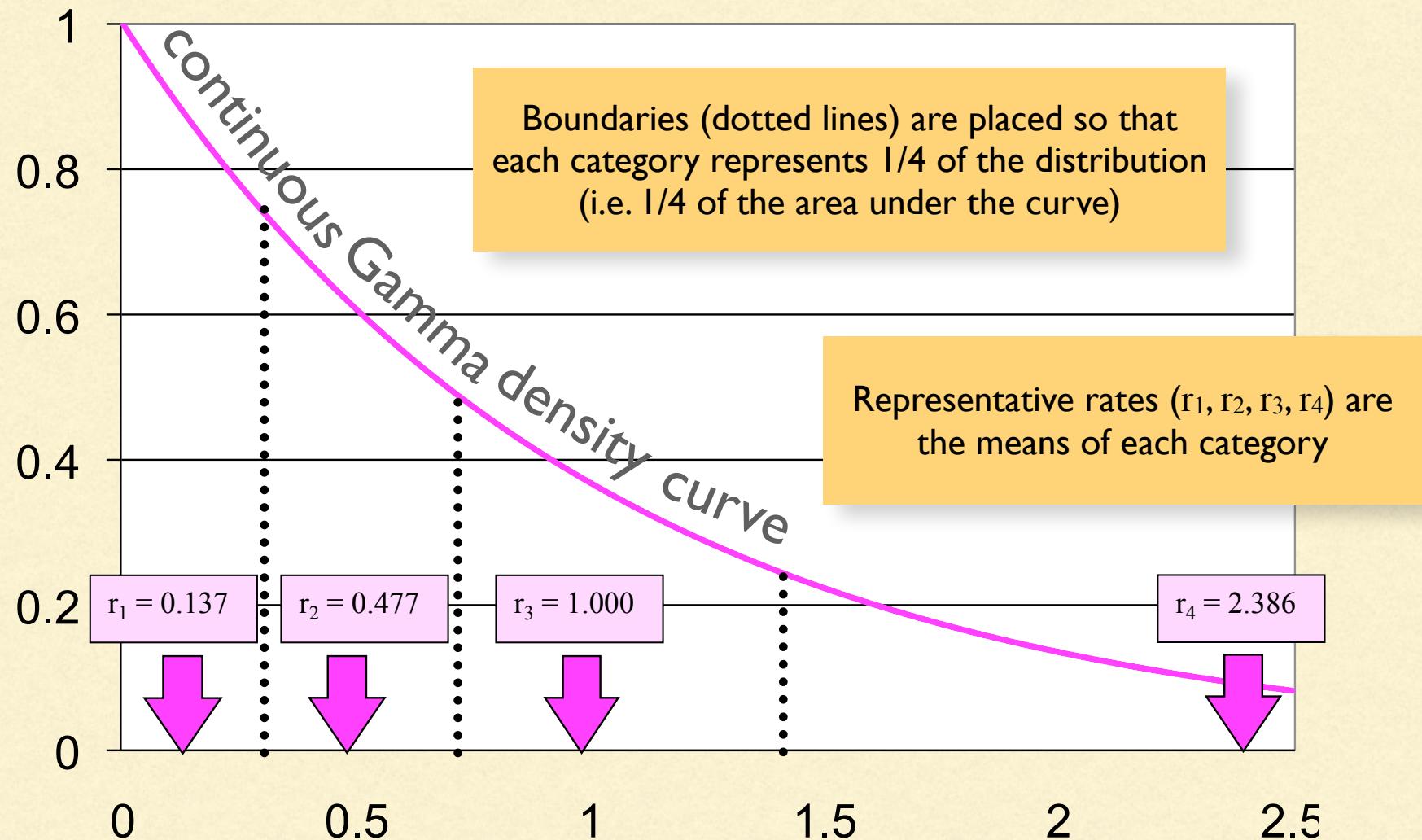
site *i*

$$L_i = p(\mathbf{y}_i | r_1) \left(\frac{1}{4} \right) + p(\mathbf{y}_i | r_2) \left(\frac{1}{4} \right) + p(\mathbf{y}_i | r_3) \left(\frac{1}{4} \right) + p(\mathbf{y}_i | r_4) \left(\frac{1}{4} \right)$$

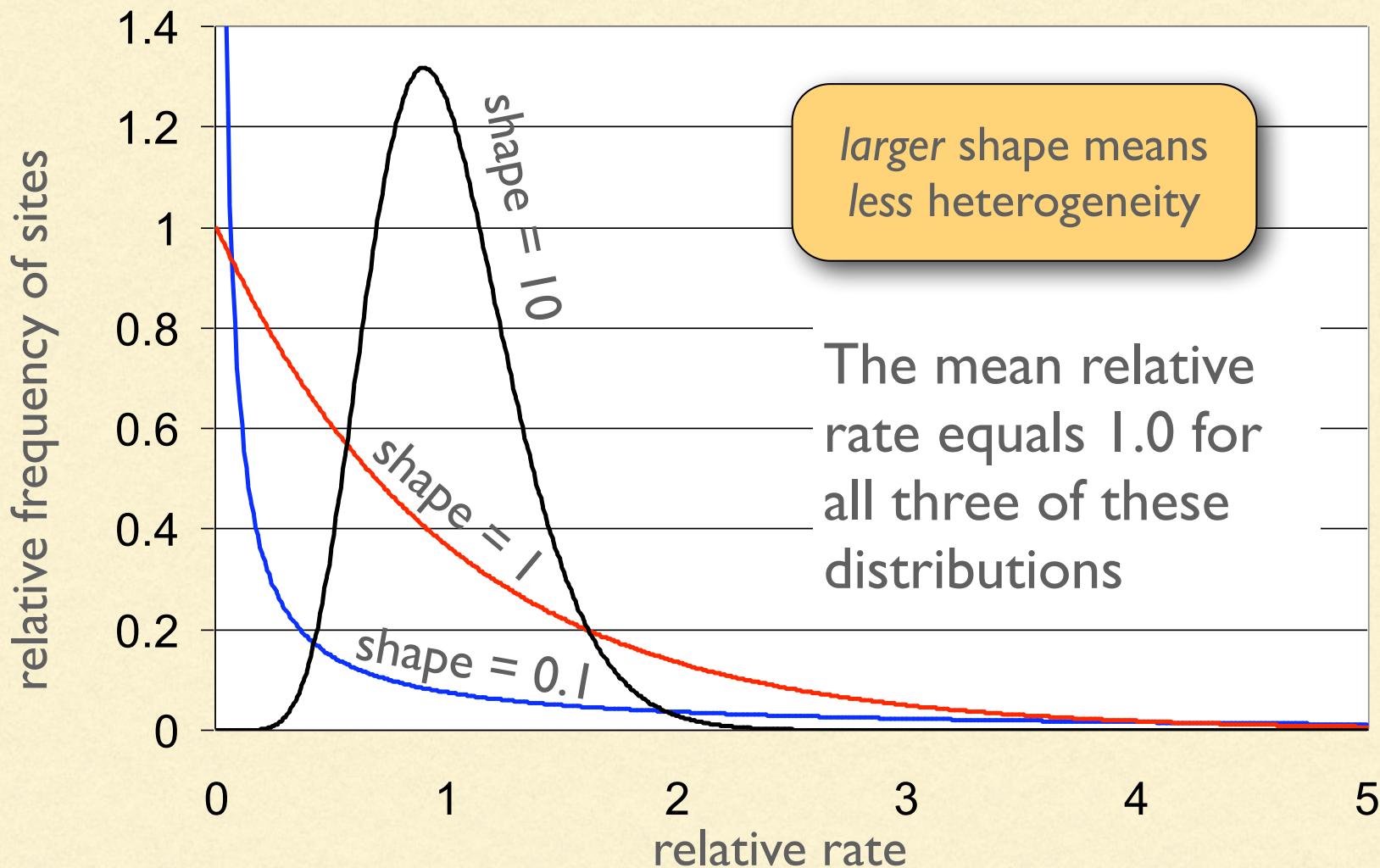
Relative rates are determined by a discrete gamma distribution

Number of rate categories can vary (4 used here)

Relative rates in 4-category case



Gamma distributions



~ Coffee Break ~
