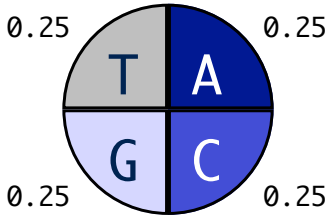
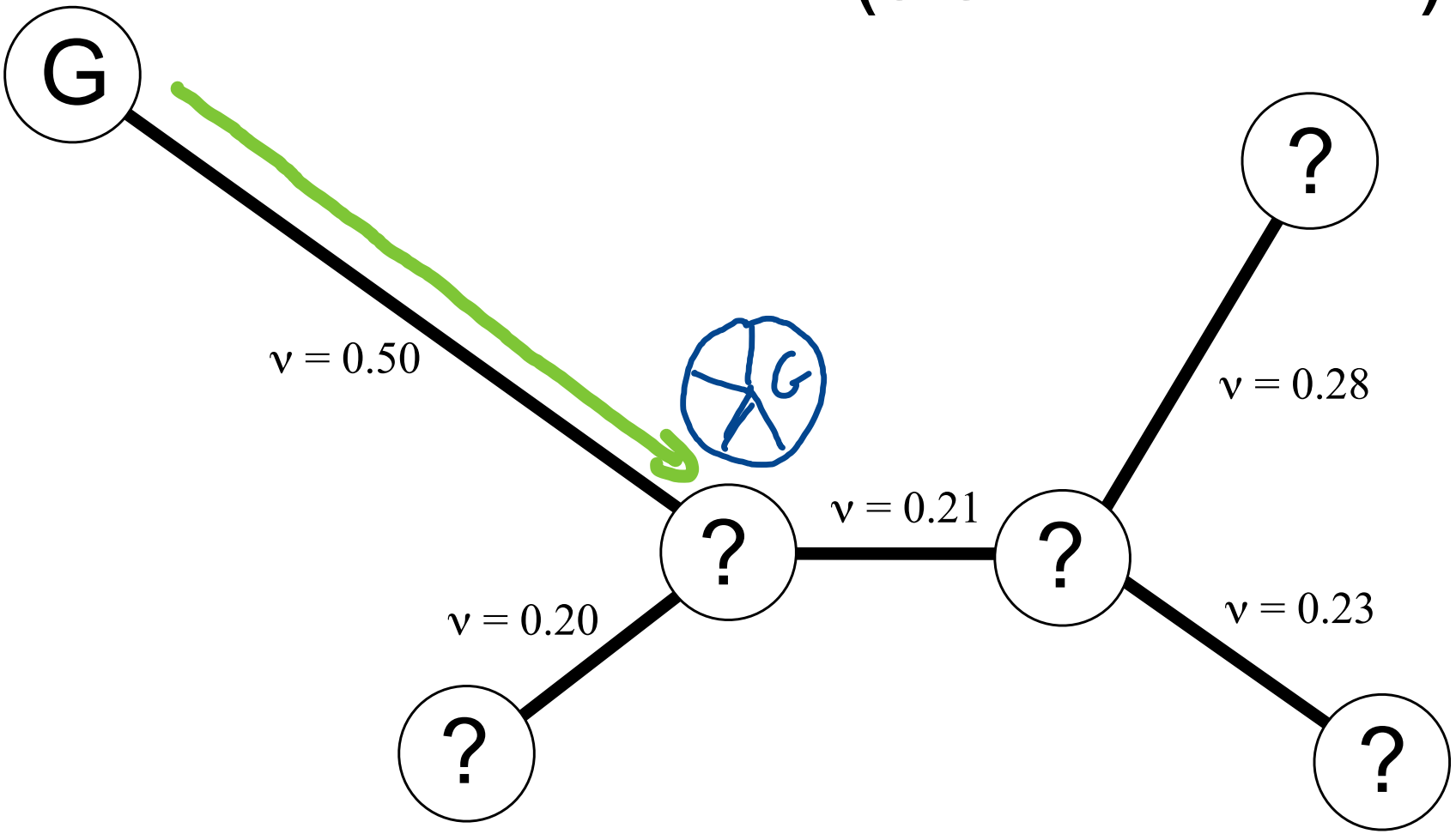
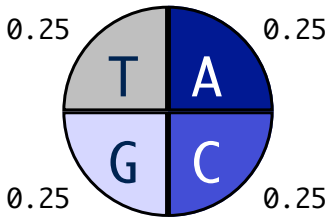
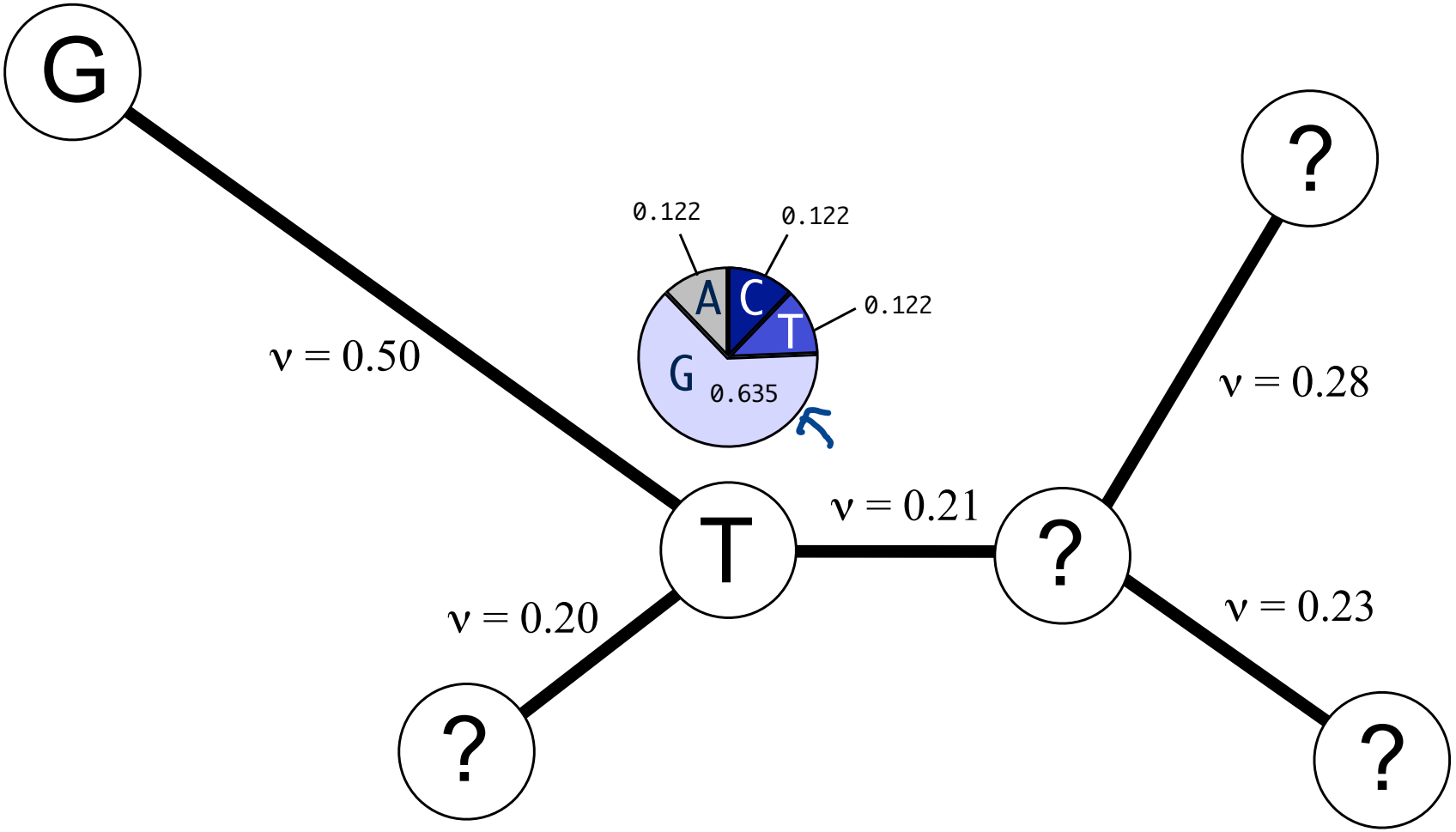


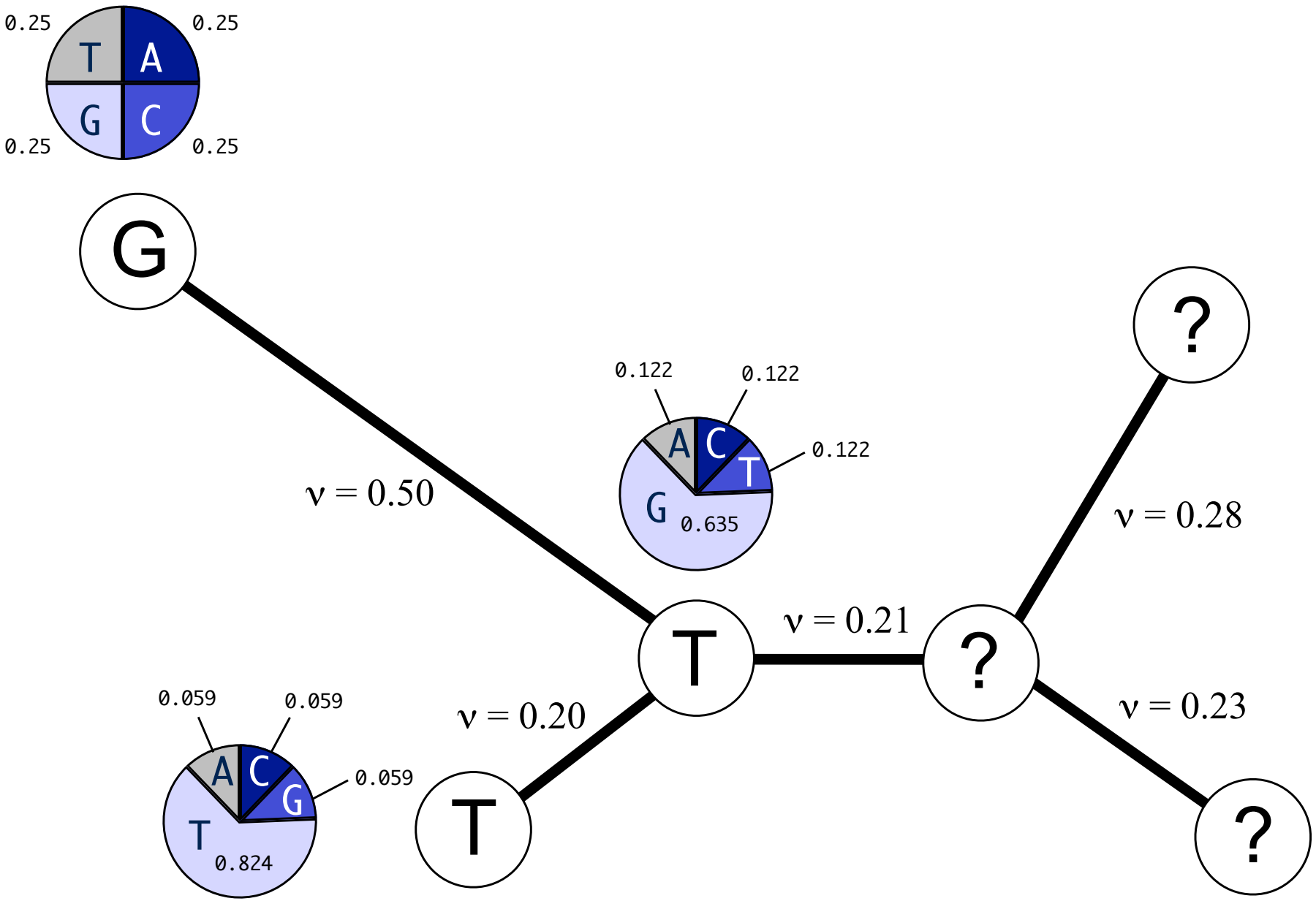
How to simulate data for one site (JC69 model)

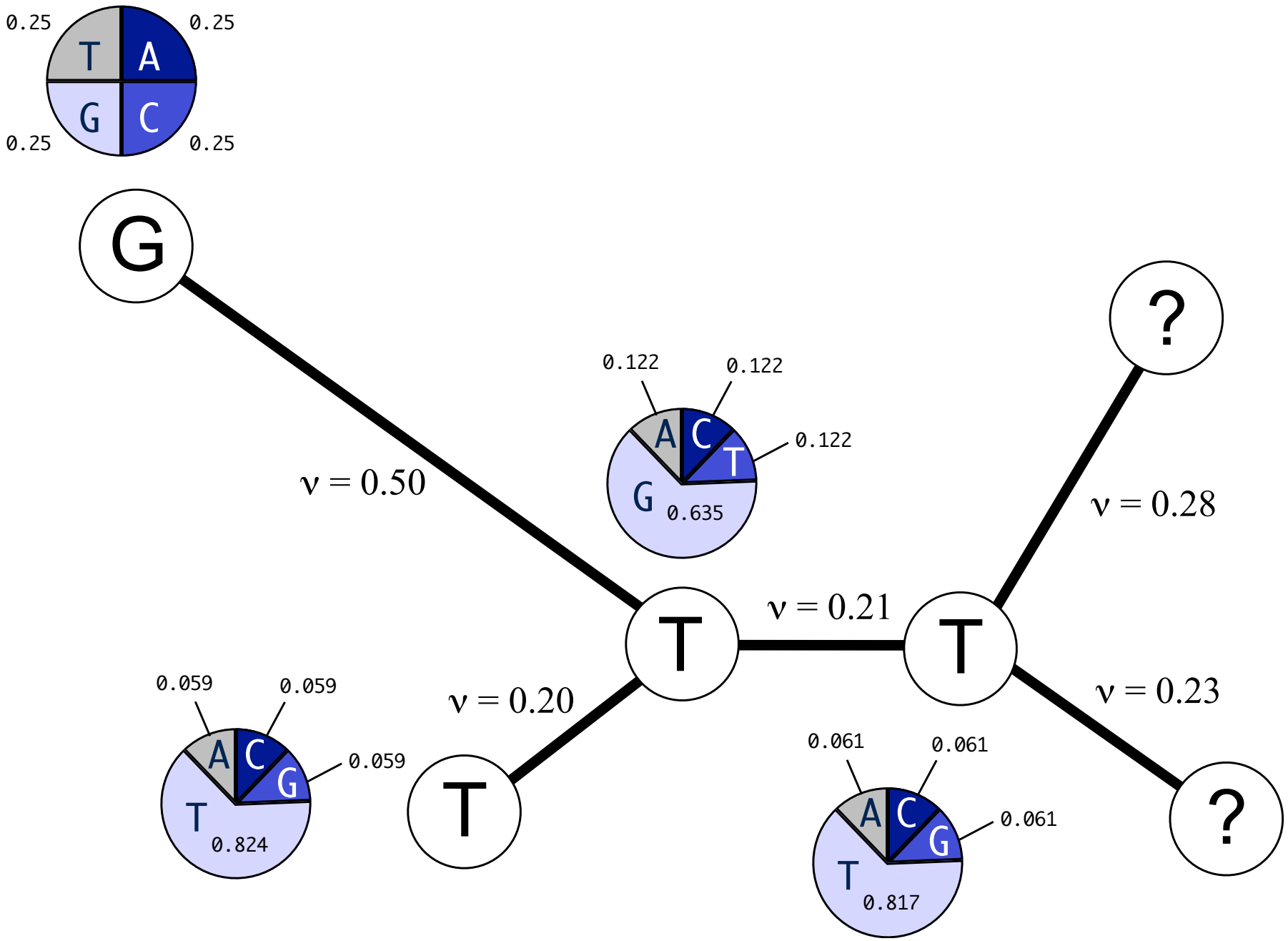


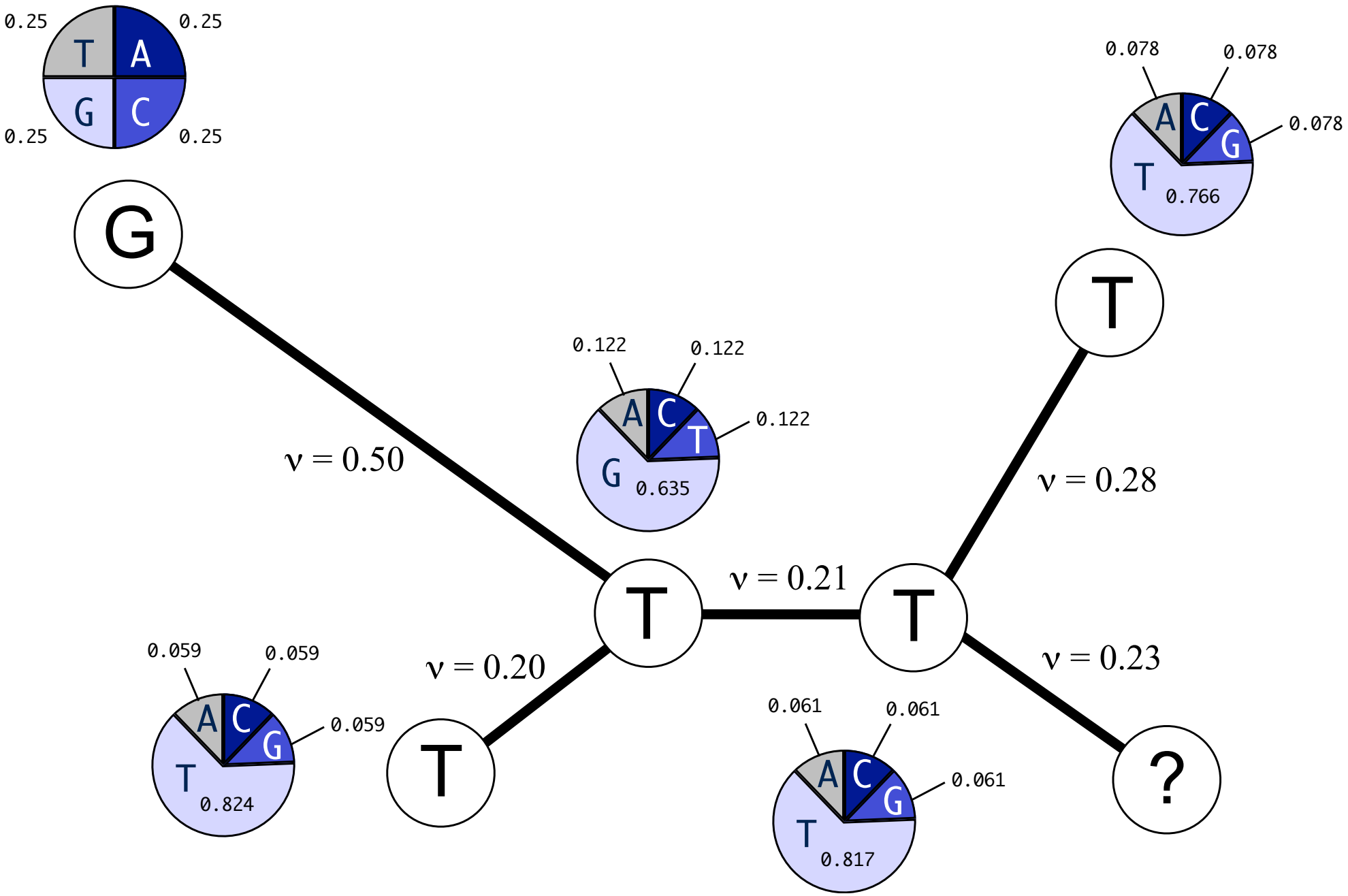


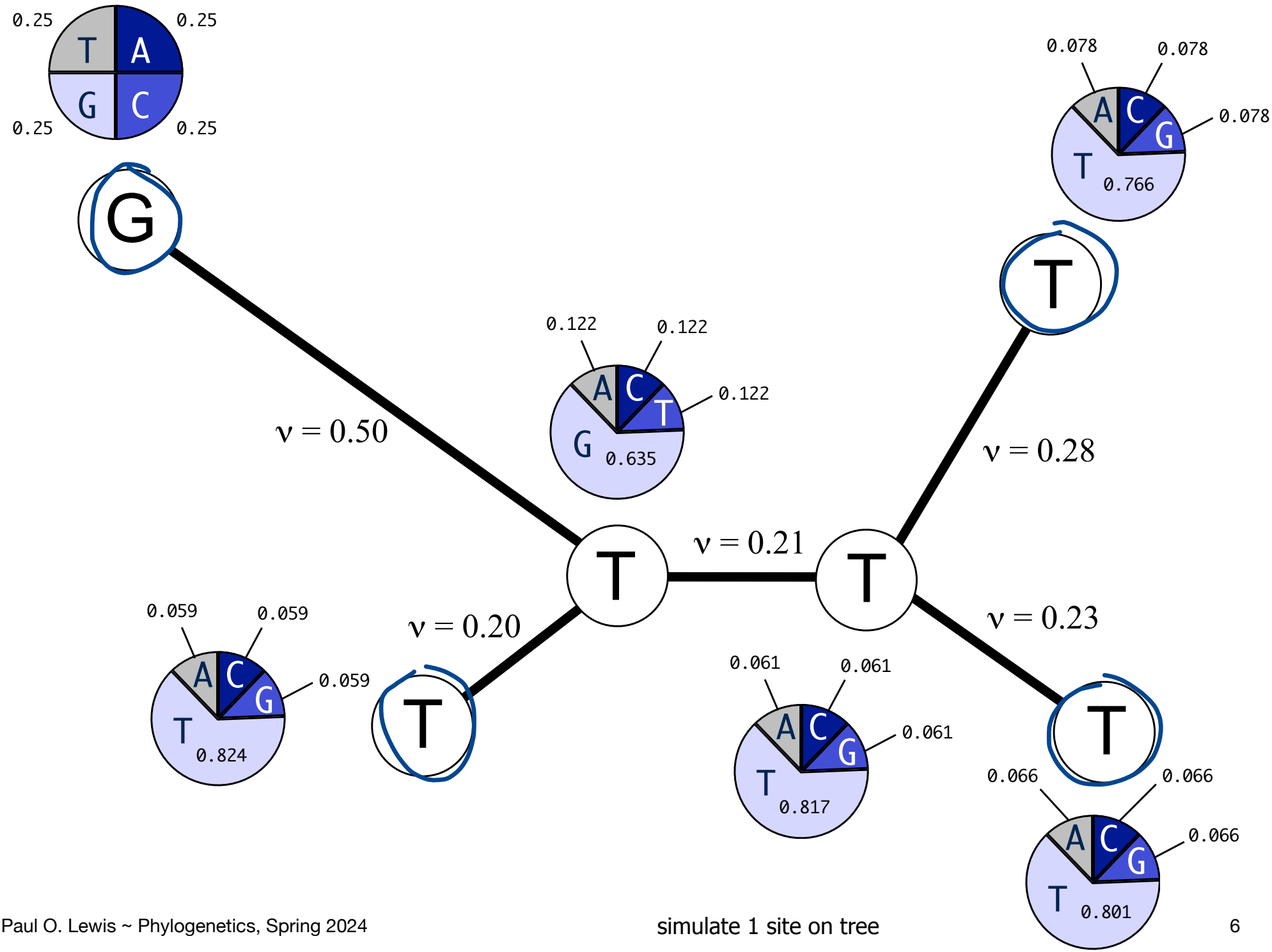
$$.635 = \frac{1}{4} + \frac{3}{4} e^{-4v/3}$$




0.25
0.25
T
A
G
C
0.25
0.25
G
ν = 0.50
0.122
0.122
0.122
A
C
T
G
0.635
T
ν = 0.21
?
?
ν = 0.28
ν = 0.23
?
ν = 0.20
T
0.059
0.059
0.059
A
C
G
T
0.824

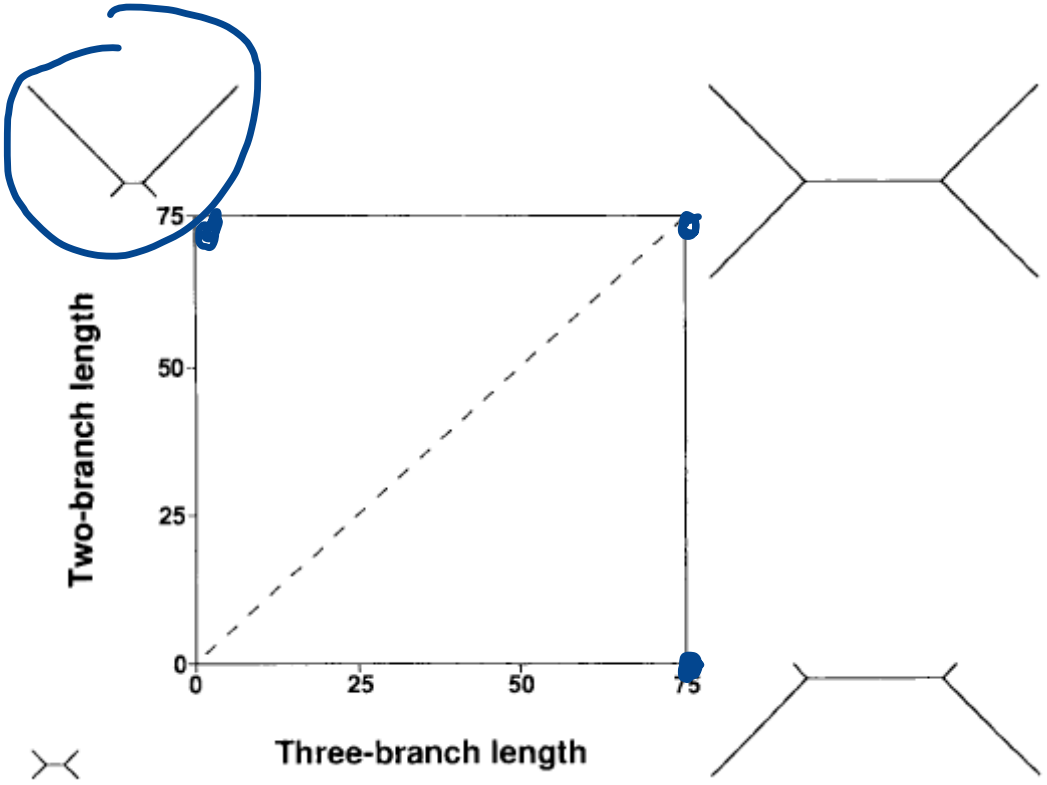

0.25
0.25
T
A
G
C
0.25
0.25
G
ν = 0.50
0.122
0.122
0.122
A
C
T
G
0.635
?
ν = 0.28
T
ν = 0.21
T
ν = 0.20
ν = 0.23
0.059
0.059
0.059
A
C
G
T
0.824
T
0.061
0.061
0.061
A
C
G
T
0.817
?

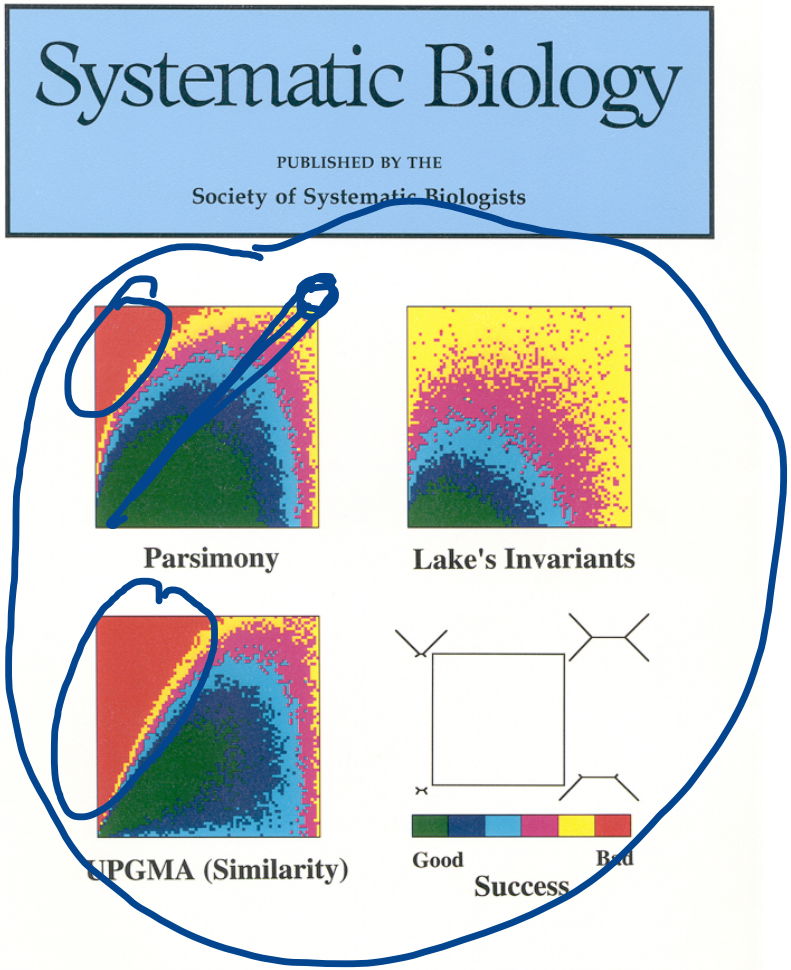

0.25
0.25
T
A
G
C
0.25
0.25
G
ν = 0.50
0.122
0.122
A
C
T
0.122
G
0.635
T
ν = 0.21
ν = 0.28
0.078
0.078
A
C
G
0.078
T
0.766
T
ν = 0.20
ν = 0.23
0.059
0.059
A
C
G
0.059
T
0.824
T
0.061
0.061
A
C
G
0.061
T
0.817
?


0.25
0.25
T
A
G
C
0.25
0.25
G
ν = 0.50
0.122
0.122
A
C
T
0.122
G
0.635
T
ν = 0.21
T
ν = 0.28
0.078
0.078
A
C
G
0.078
T
0.766
T
ν = 0.20
0.059
0.059
A
C
G
0.059
T
0.824
T
ν = 0.23
0.061
0.061
A
C
G
0.061
T
0.817
T
0.066
0.066
A
C
G
0.066
T
0.801

Uses of simulation in phylogenetics

- Investigate effects of violating the assumptions of models





Felsenstein (1978)
Huelsenbeck and Hillis (1993)

Uses of simulation in phylogenetics

- Test hypotheses

https://imgur.com/gallery/eH0aCzk

megabat

https://www.milkwood.net/2015/09/21/how-to-build-a-microbat-box/

microbat

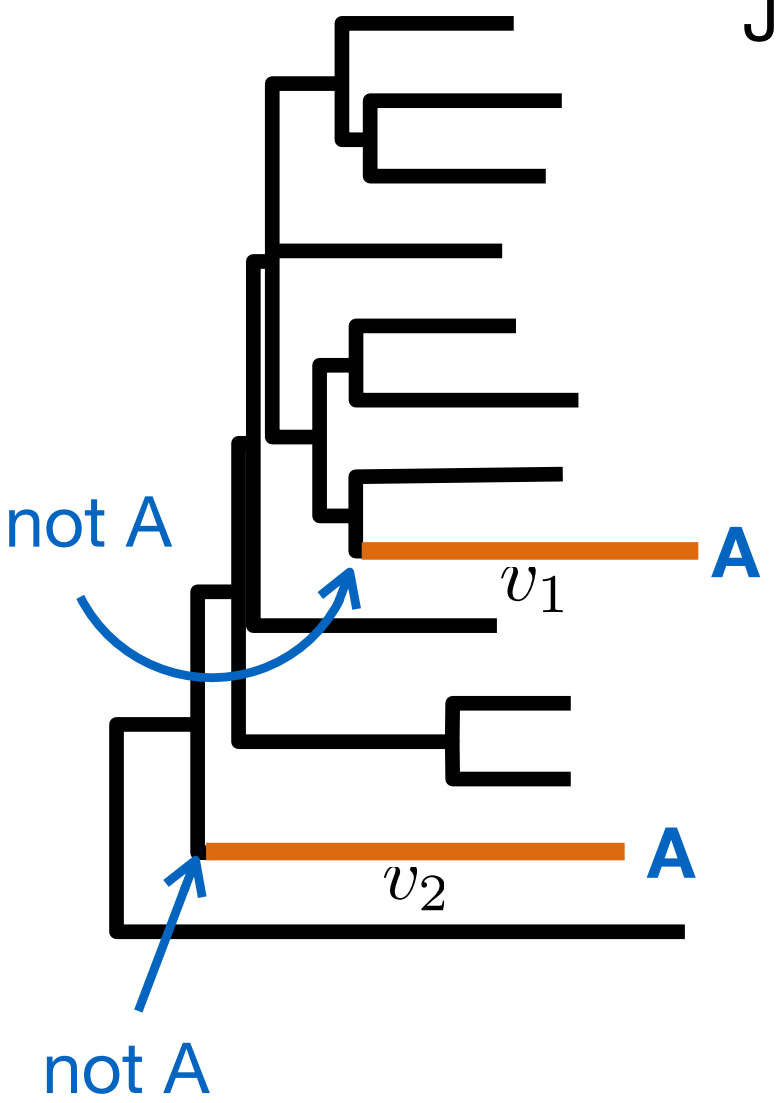
one or two origins of bats?

Bat monophyly test: background

- Most phylogenetic studies concluded bats monophyletic (microbats and megabats together)
- Pettigrew argued megabats related instead to primates (megabats were “flying primates”)
- Pettigrew: apparent monophyly is an artifact: high AT composition of all bats leads to “base compositional attraction”
- Test by simulating data under null (flying primate) hypothesis (assuming strong AT bias in bat lineages) and see how often bat monophyly occurs in phylogenetic analyses

Pettigrew (1991)

Base-compositional attraction



JC69 model applies to entire tree

Probability both lineages *converge* on A under JC69 model

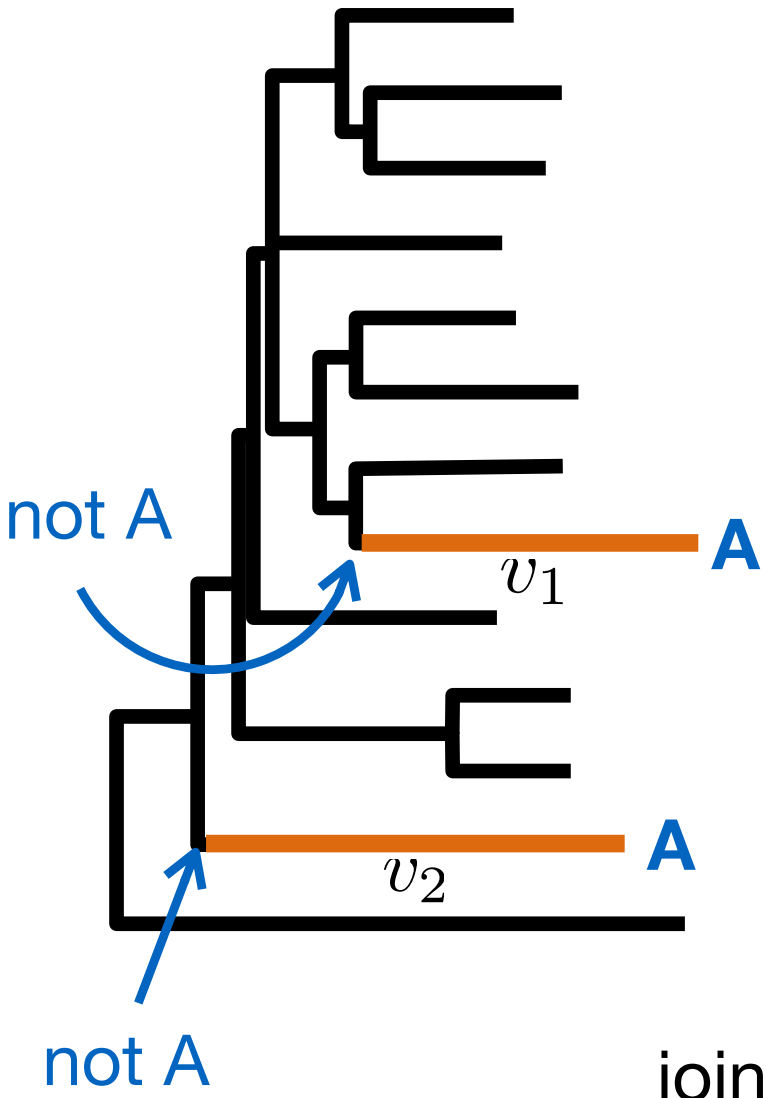
$$\frac{1}{4}\left(1 - e^{-4\beta t}\right)$$

$$v = 3\beta t$$

joint prob. approx. 0.001 if $v_1 = v_2 = 0.1$

Base-compositional attraction

JC69 model applies to black lineages only



Probability both lineages *converge* on A under F81 model with **high A+T bias** and **non-stationarity**

$$\pi_A \left(1 - e^{-\mu t}\right)$$

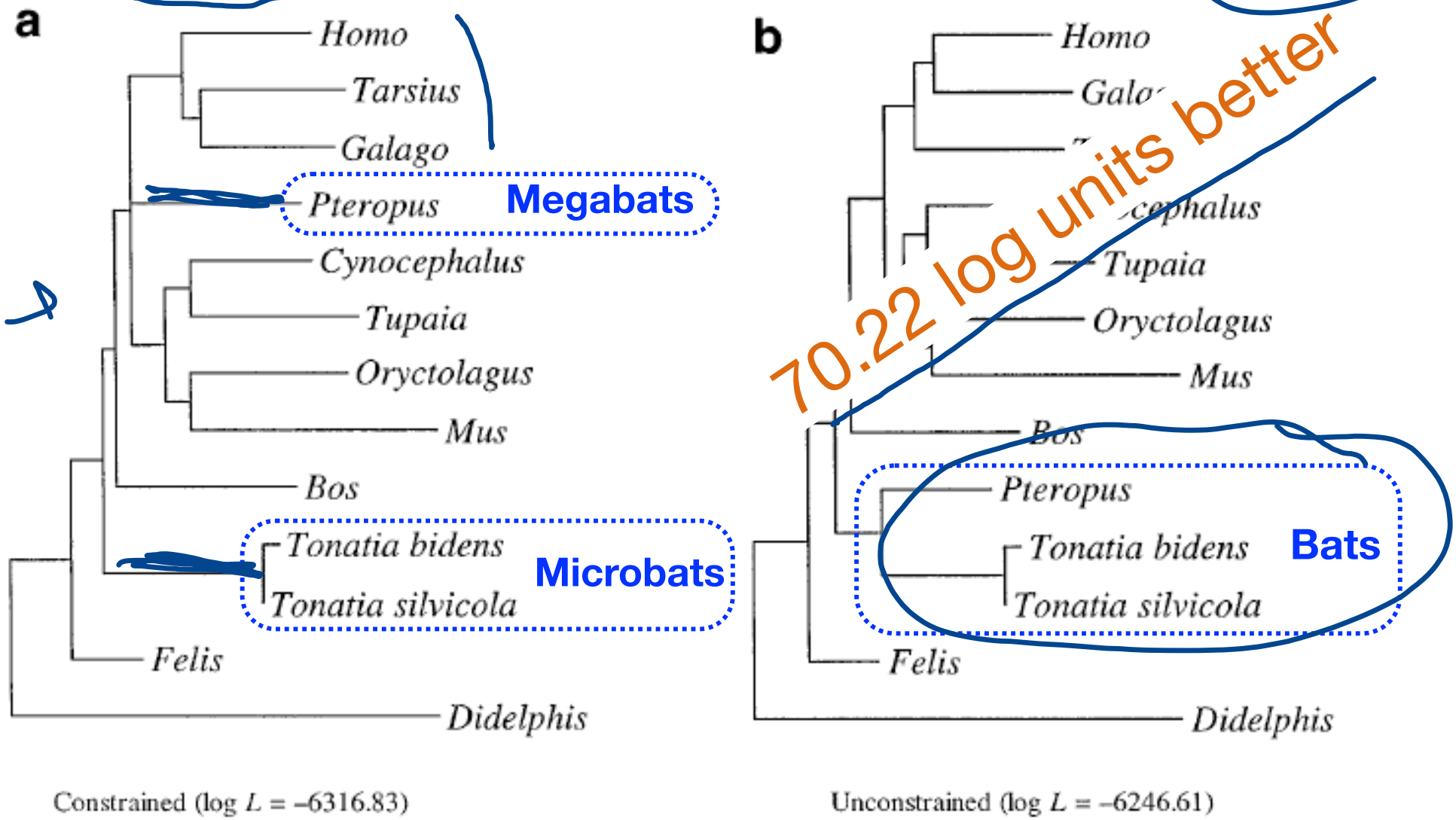
$$v = \mu t(1 - \pi_A^2 - \pi_C^2 - \pi_G^2 - \pi_T^2)$$

joint prob. approx. 0.003 if $v_1 = v_2 =$ 0.1 and $\pi_A = \pi_T =$ 0.4, $\pi_C = \pi_G =$ 0.1

Bat monophyly example

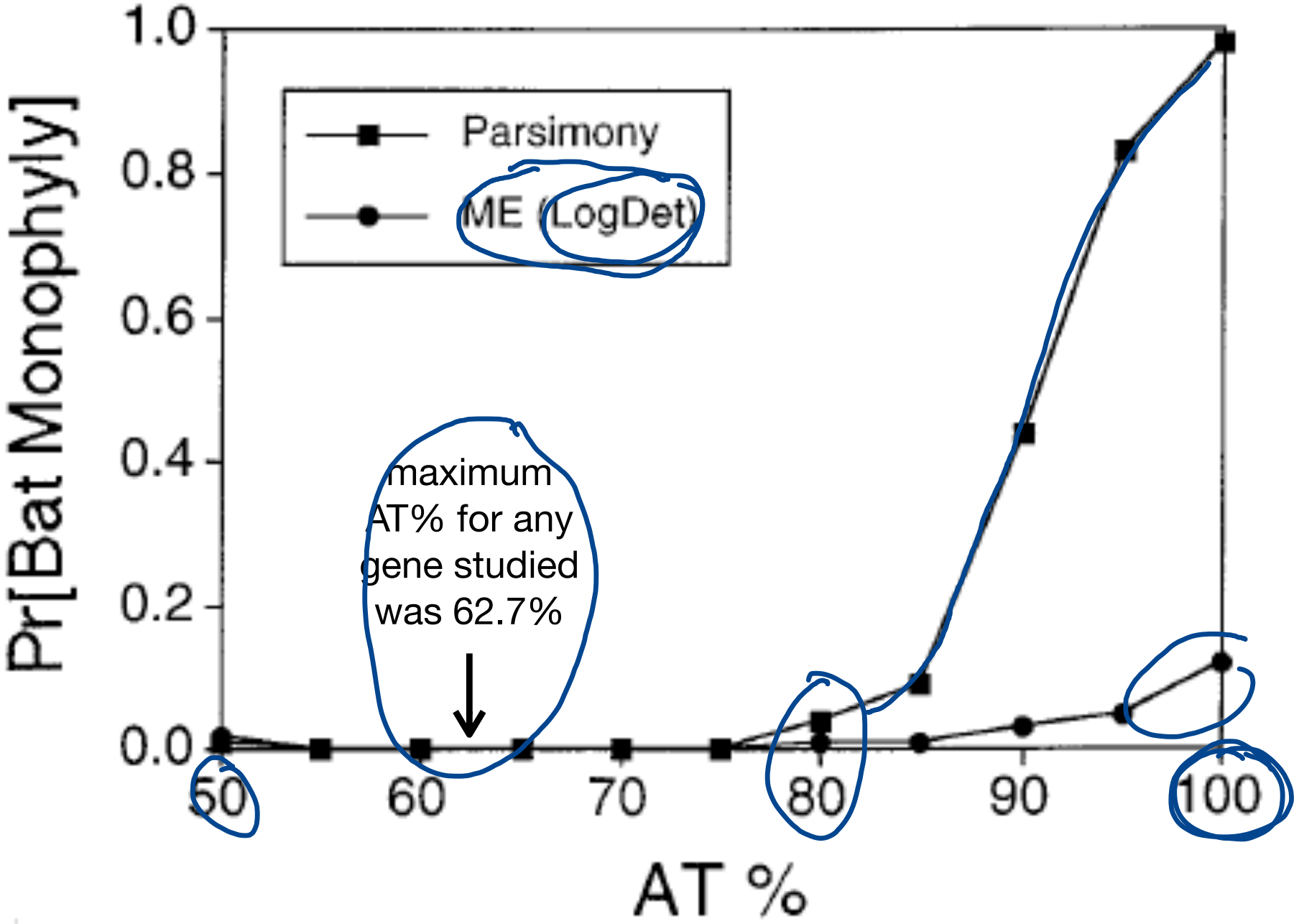
Null hypothesis

Alternative hypothesis (ML tree)



Constrained (log L = –6316.83)

Unconstrained (log L = –6246.61)

Van Den Bussche et al. (1998)

Bat monophyly example