Evolution of discrete characters

Varanus salvator

Bitis gabonica

Gehyra mutilata

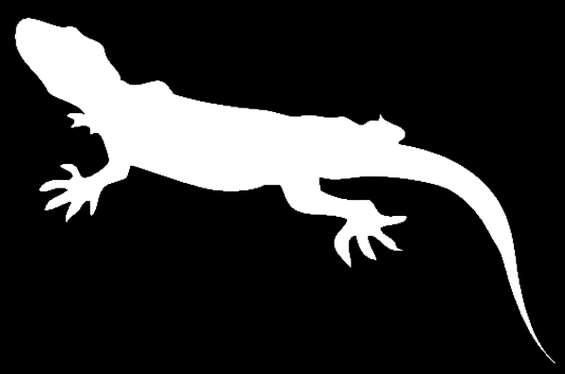
Lialis burtonis

Plestiodon fasciatus

Armand Kok
2009

Acontias meleagris

How do we model the process
by which squamates
(snakes and lizards)
lose their limbs?

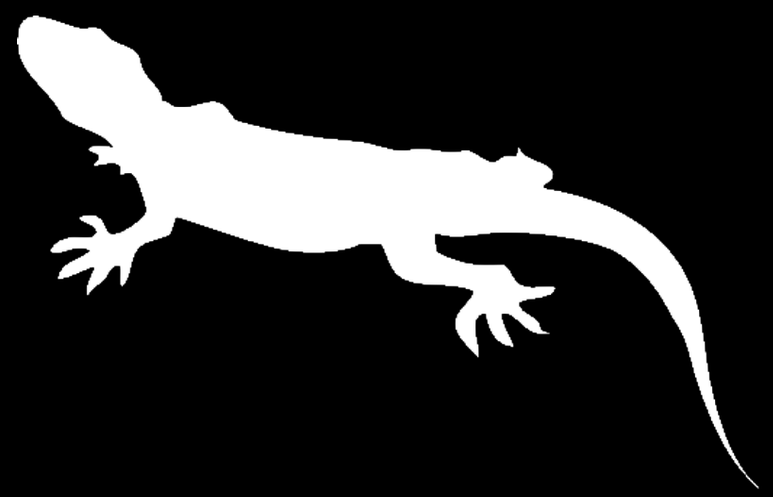


No limbs

State 0

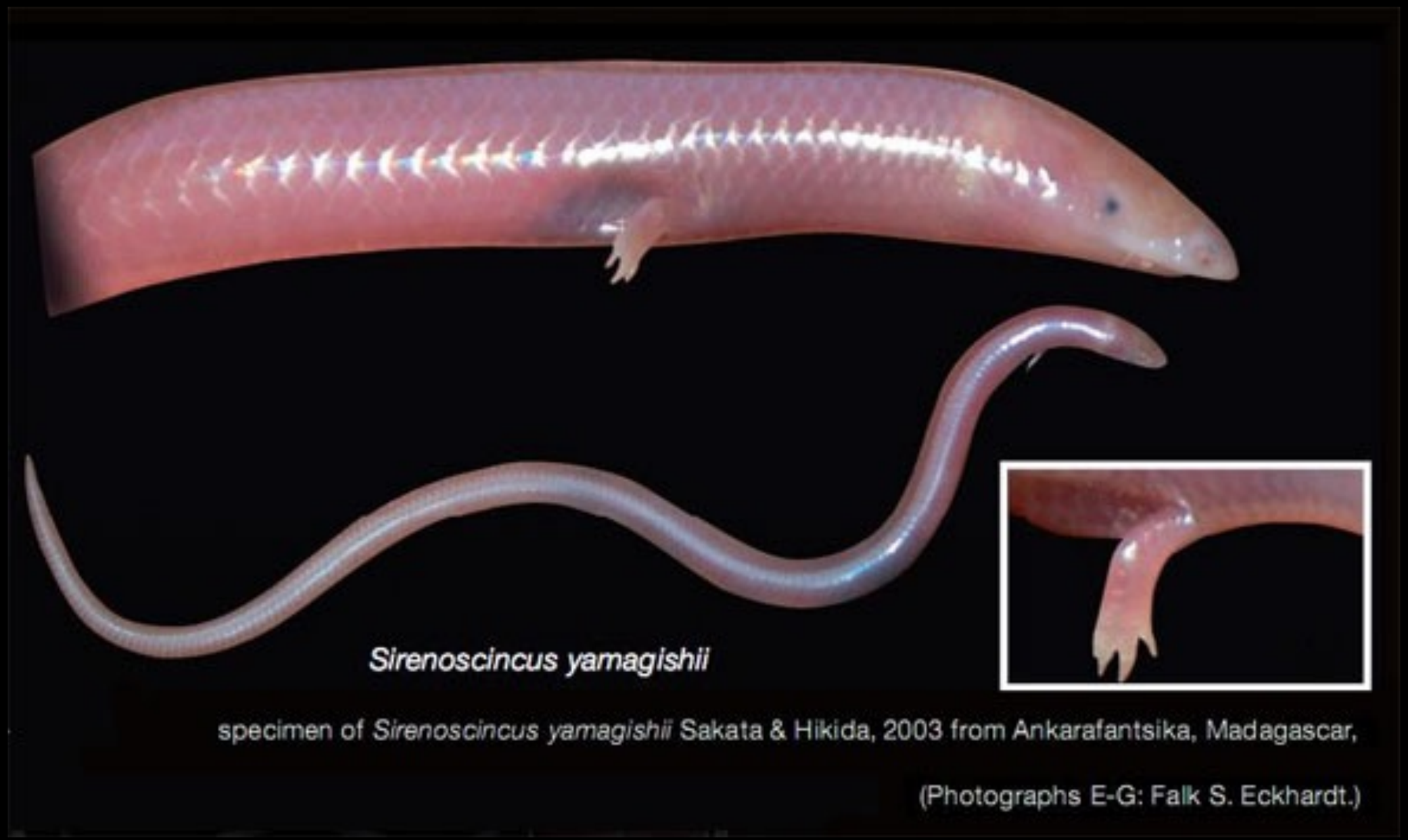
No limbs

State 0



Limbs

State 1

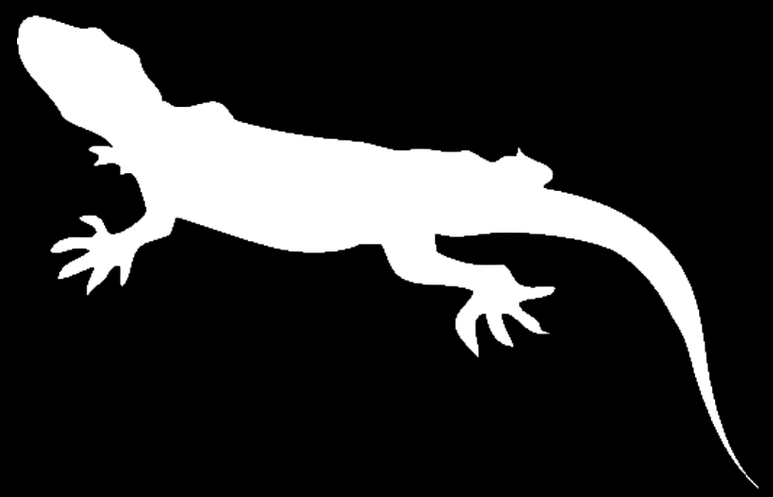


specimen of *Sirenoscincus yamagishii* Sakata & Hikida, 2003 from Ankarafantsika, Madagascar,

(Photographs E-G: Falk S. Eckhardt.)

No limbs

State 0



Limbs

State 1

The Mk model

The Mk model

“Markov k-state”

Lewis 2001 Sys Bio

0

1

0

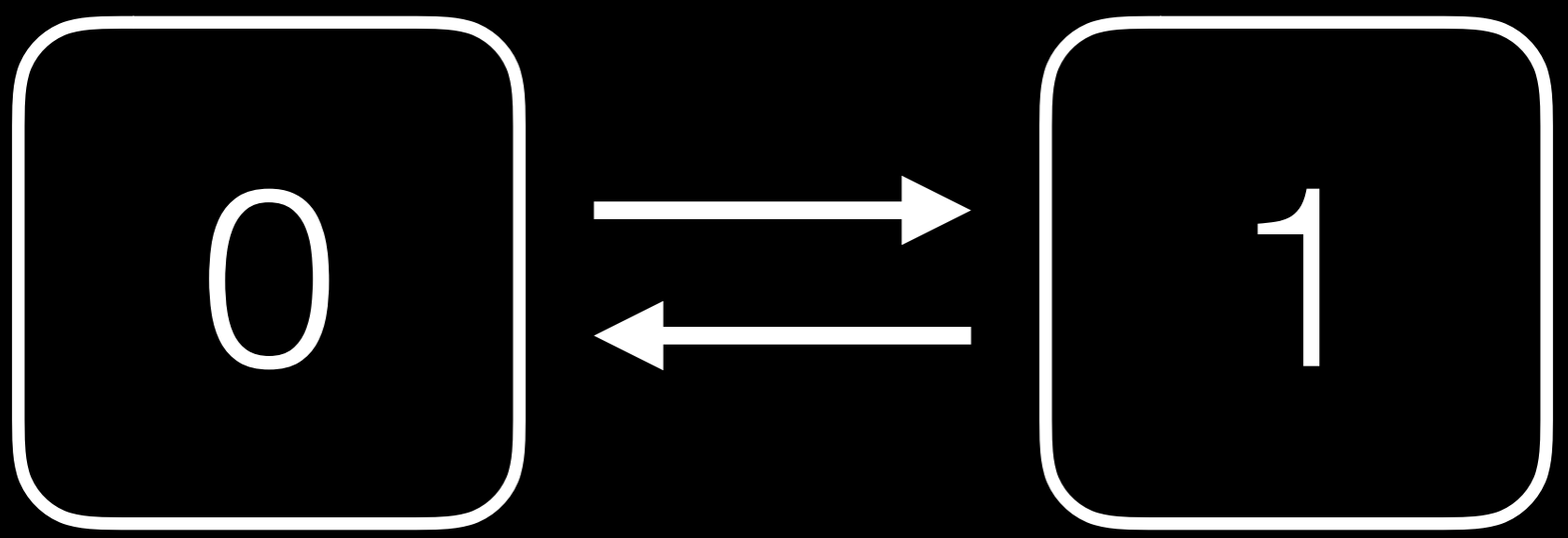
no limbs

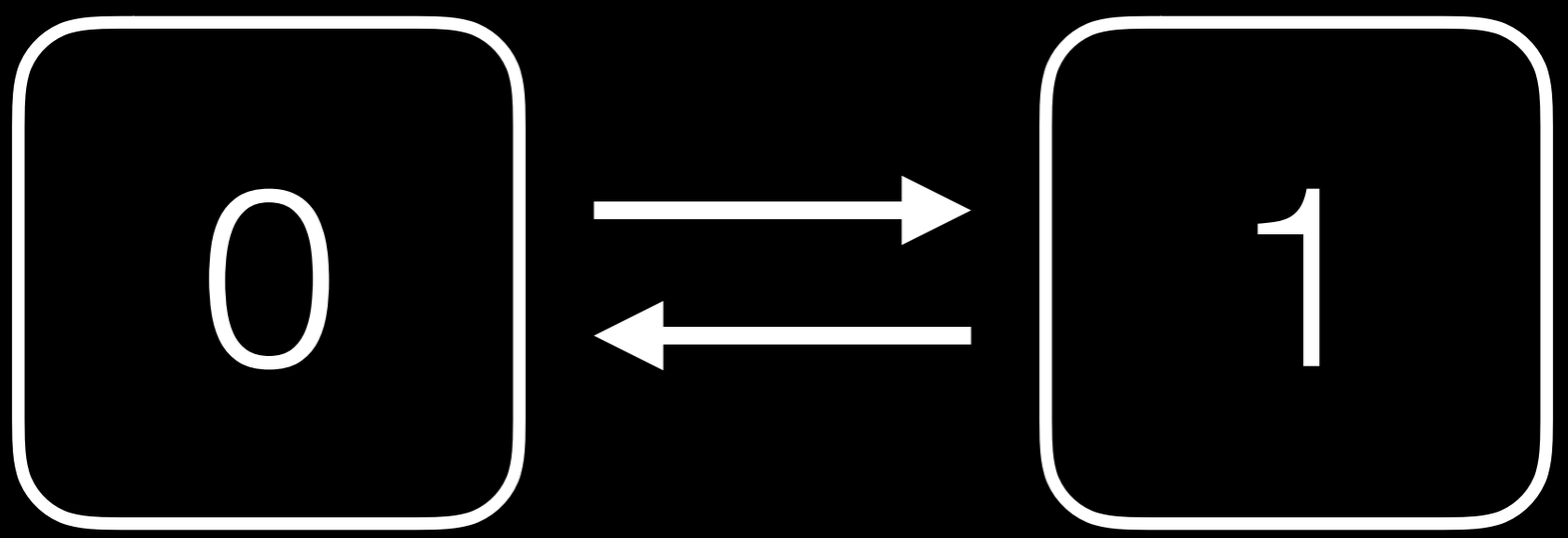
1

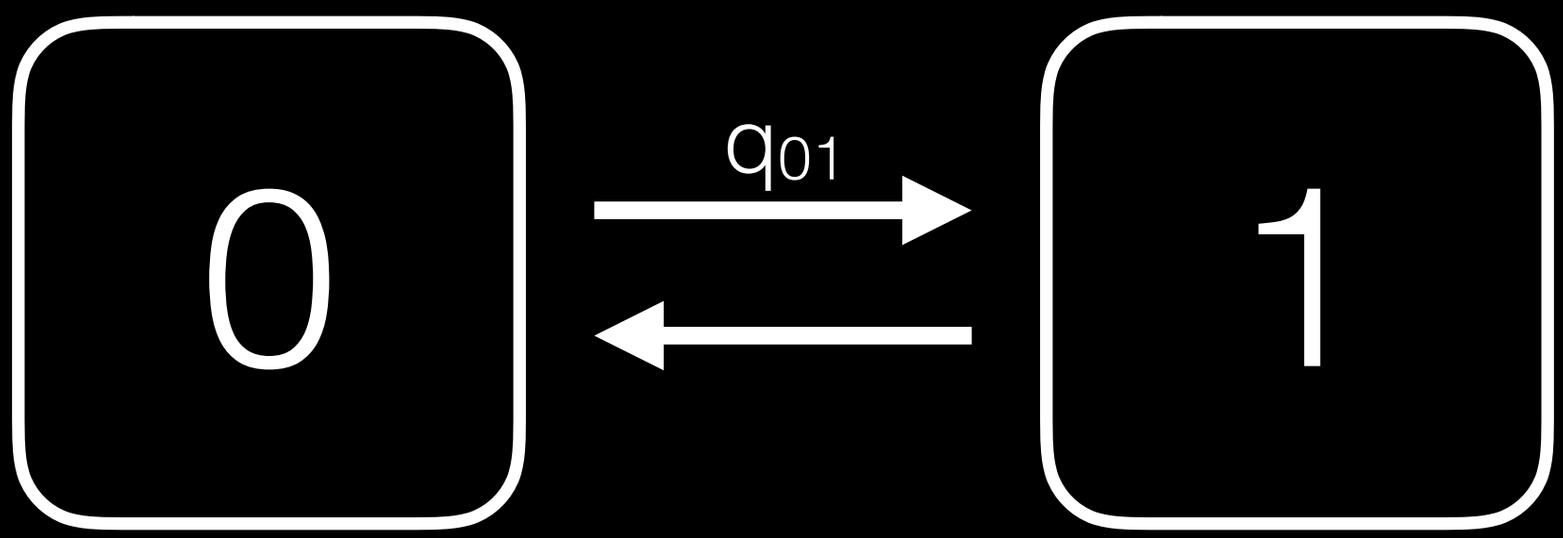
limbs

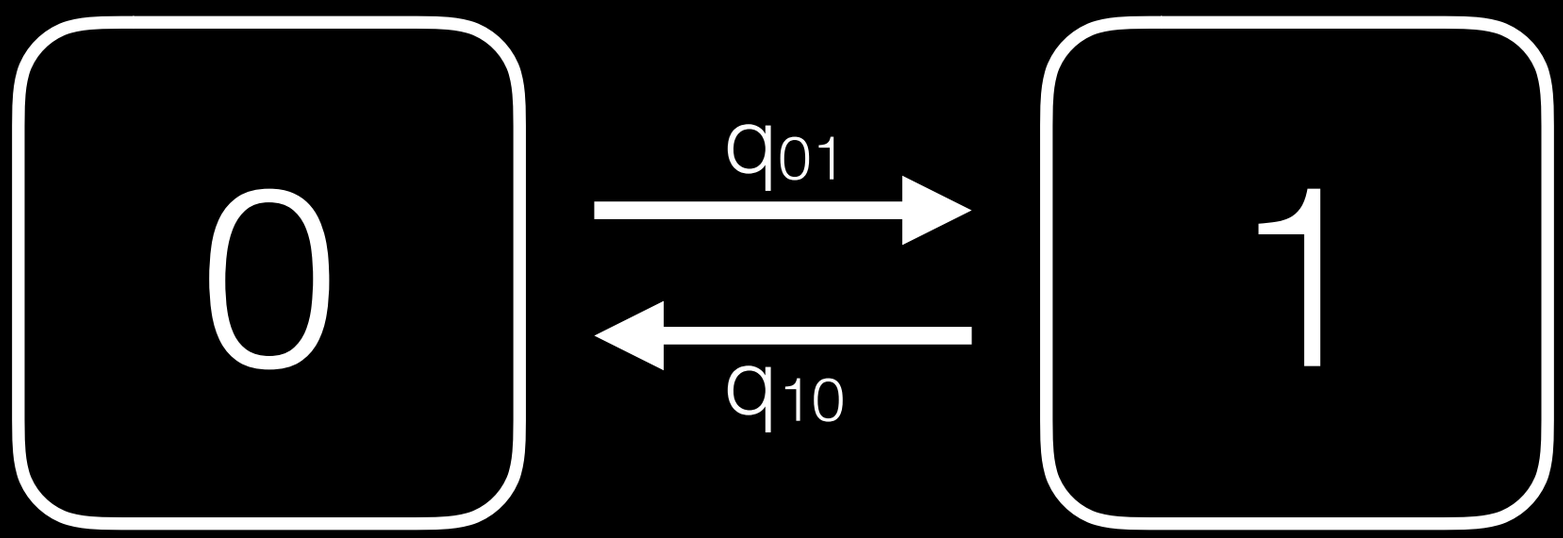
0

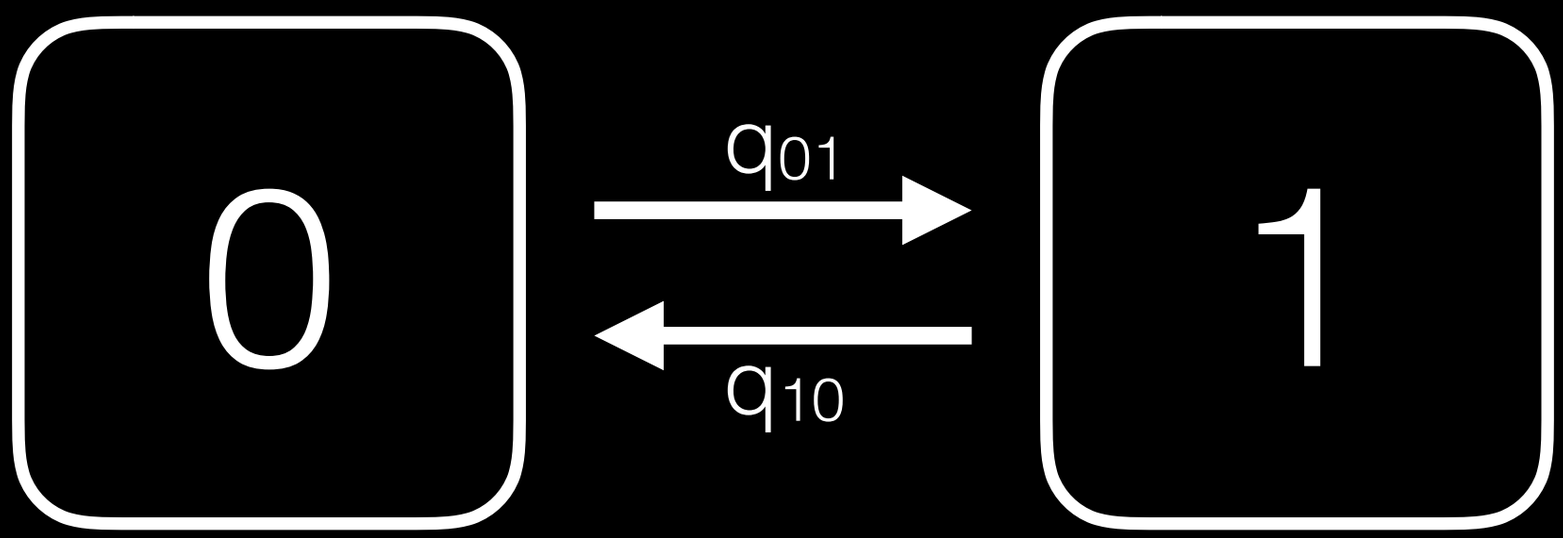
1


0
1


0
1

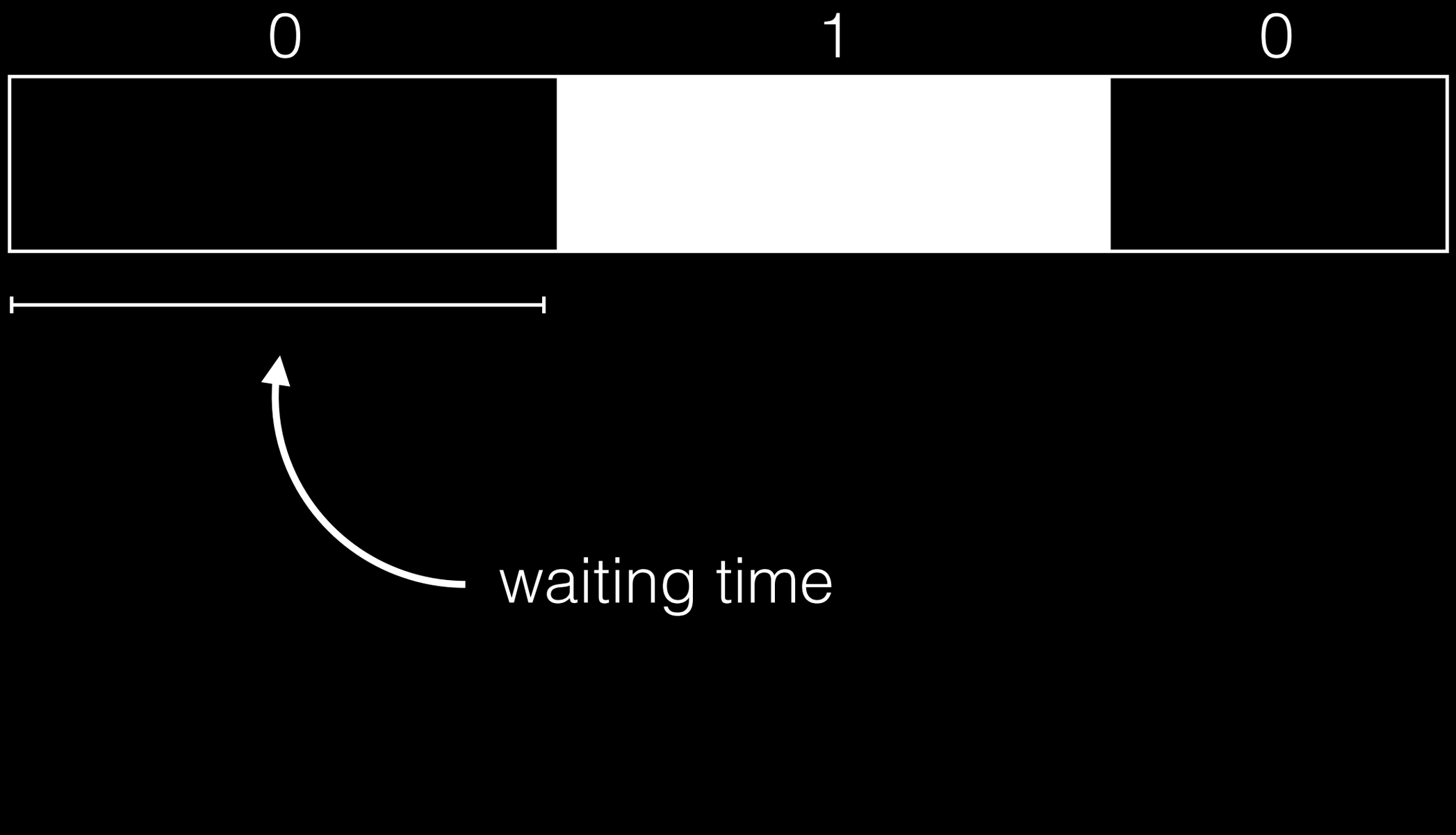

0
1
q_{01}


0
1
q_{01}
q_{10}



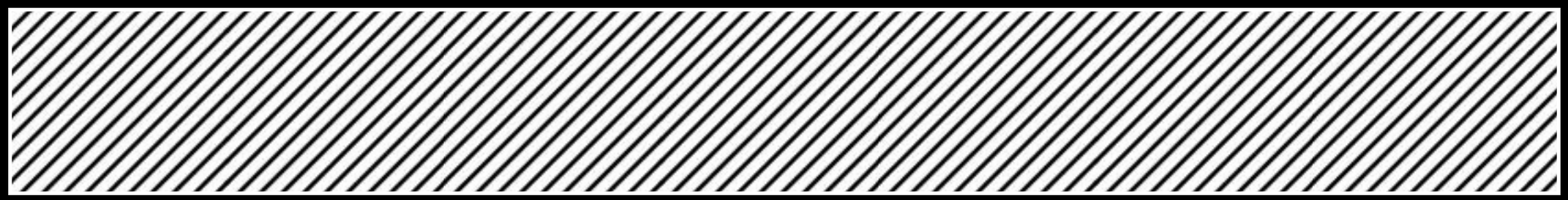
q_{01} = instantaneous rate of change from 0 to 1

0
1
0


0
1
0
waiting time

1

?



t

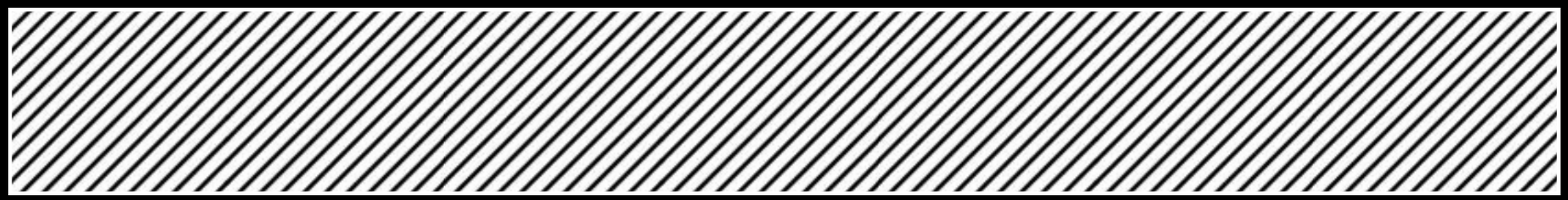
$$Q = \begin{bmatrix} -q_{01} & q_{01} \\ q_{10} & -q_{10} \end{bmatrix}$$

$$Q = \begin{bmatrix} -q_{01} & q_{01} \\ q_{10} & -q_{10} \end{bmatrix}$$

The instantaneous rate matrix

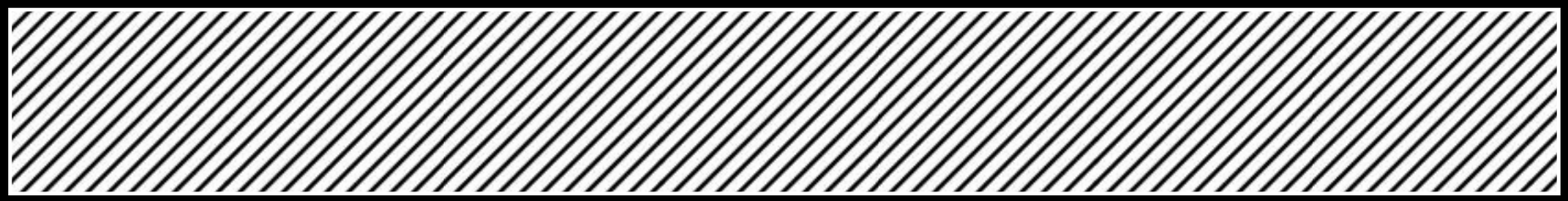
1

?



t

1 ?



t

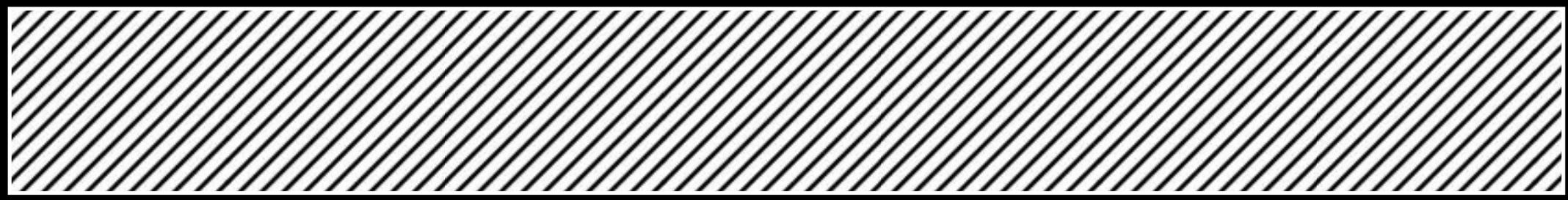
What is $p_{10}(t)$?

$$P(t)=e^{Qt}$$

$$P(t) = \underbrace{e^{Qt}}$$

matrix exponentiation

1 ?



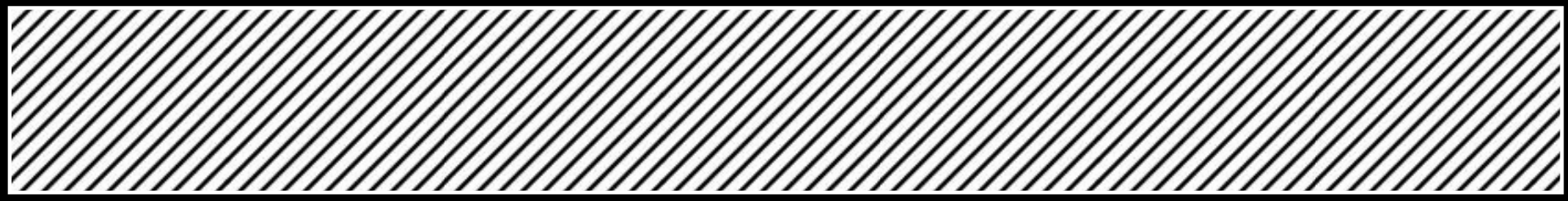
t

$q_{01} = 0.1$

$q_{10} = 0.1$

$t = 5$ my

1 ?



t

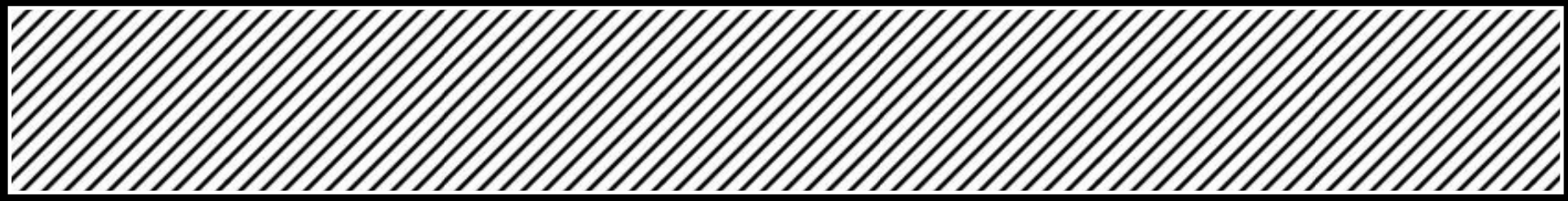
$$Q = \begin{bmatrix} -0.1 & 0.1 \\ 0.1 & -0.1 \end{bmatrix}$$

$q_{01} = 0.1$

$q_{10} = 0.1$

t = 5 my

1 ?



t

$$Q = \begin{bmatrix} -0.1 & 0.1 \\ 0.1 & -0.1 \end{bmatrix}$$

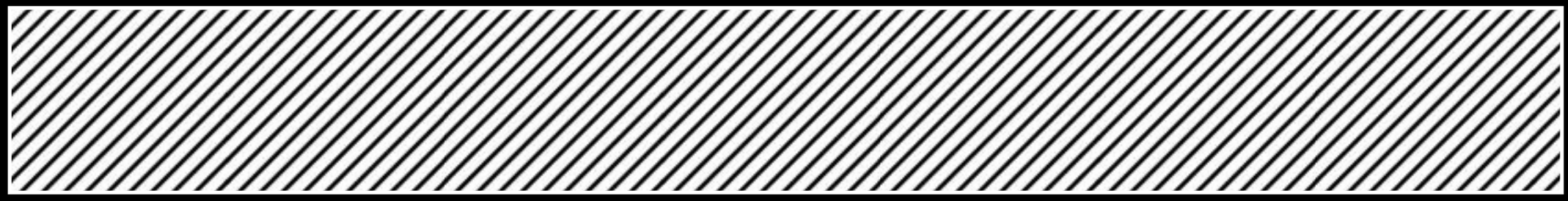
$q_{01} = 0.1$

$q_{10} = 0.1$

t = 5 my

$$Qt = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}$$

1 ?



t

$$Q = \begin{bmatrix} -0.1 & 0.1 \\ 0.1 & -0.1 \end{bmatrix}$$

$q_{01} = 0.1$

$q_{10} = 0.1$

$t = 5$ my

$$Qt = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}$$

$$e^{Qt} = \begin{bmatrix} 0.68 & 0.32 \\ 0.32 & 0.68 \end{bmatrix}$$

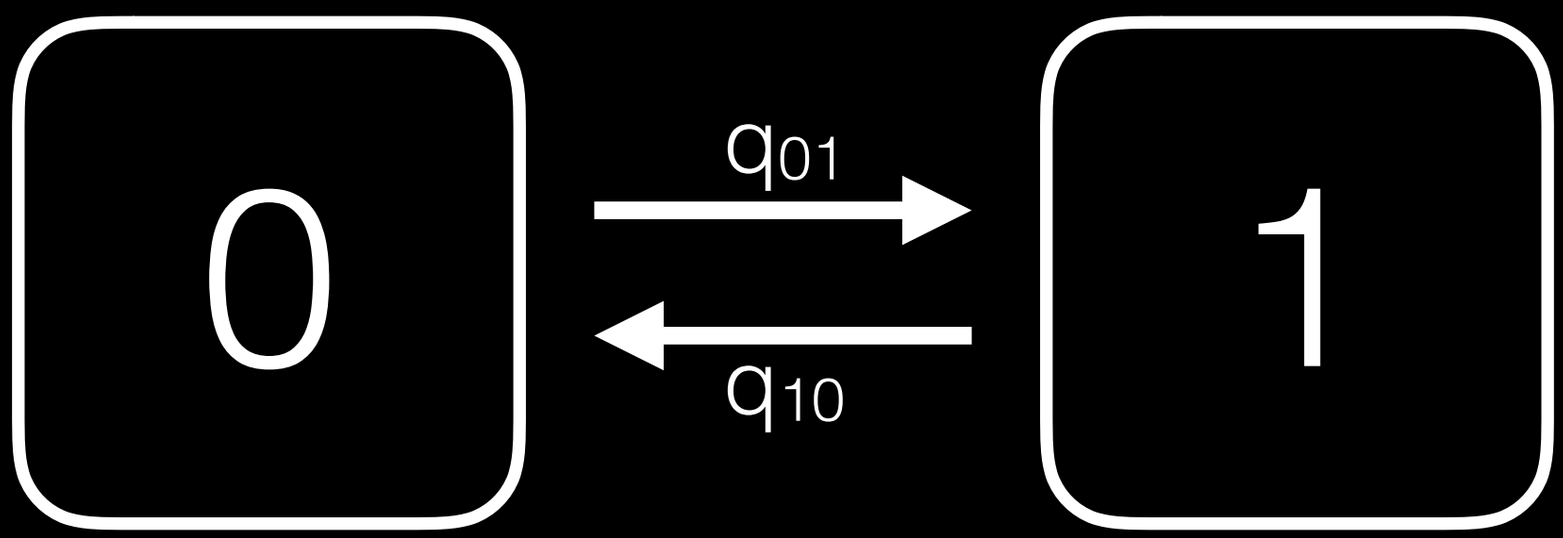
Equal rates (slow)

Equal rates (fast)

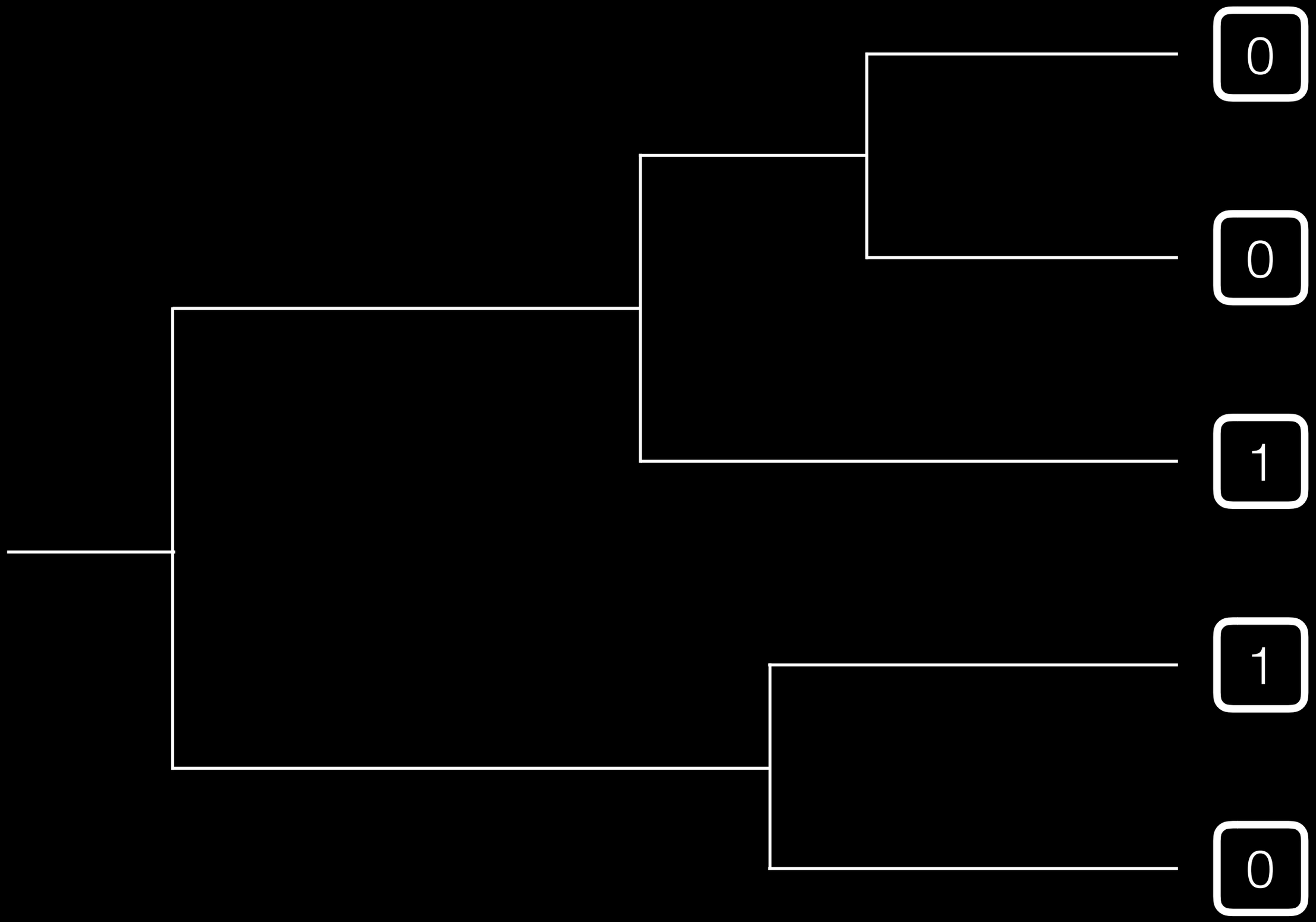
Unequal rates

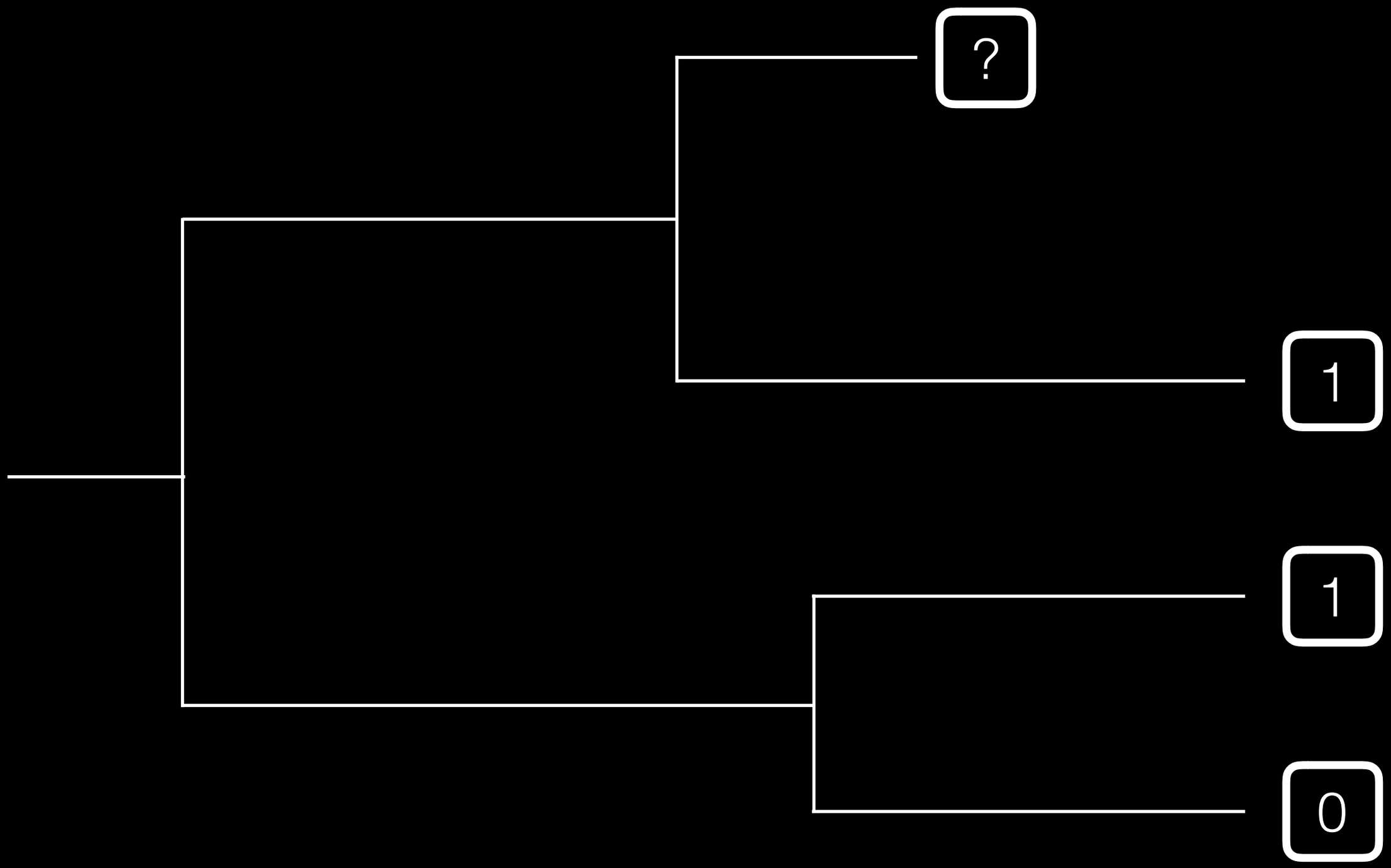
limbed

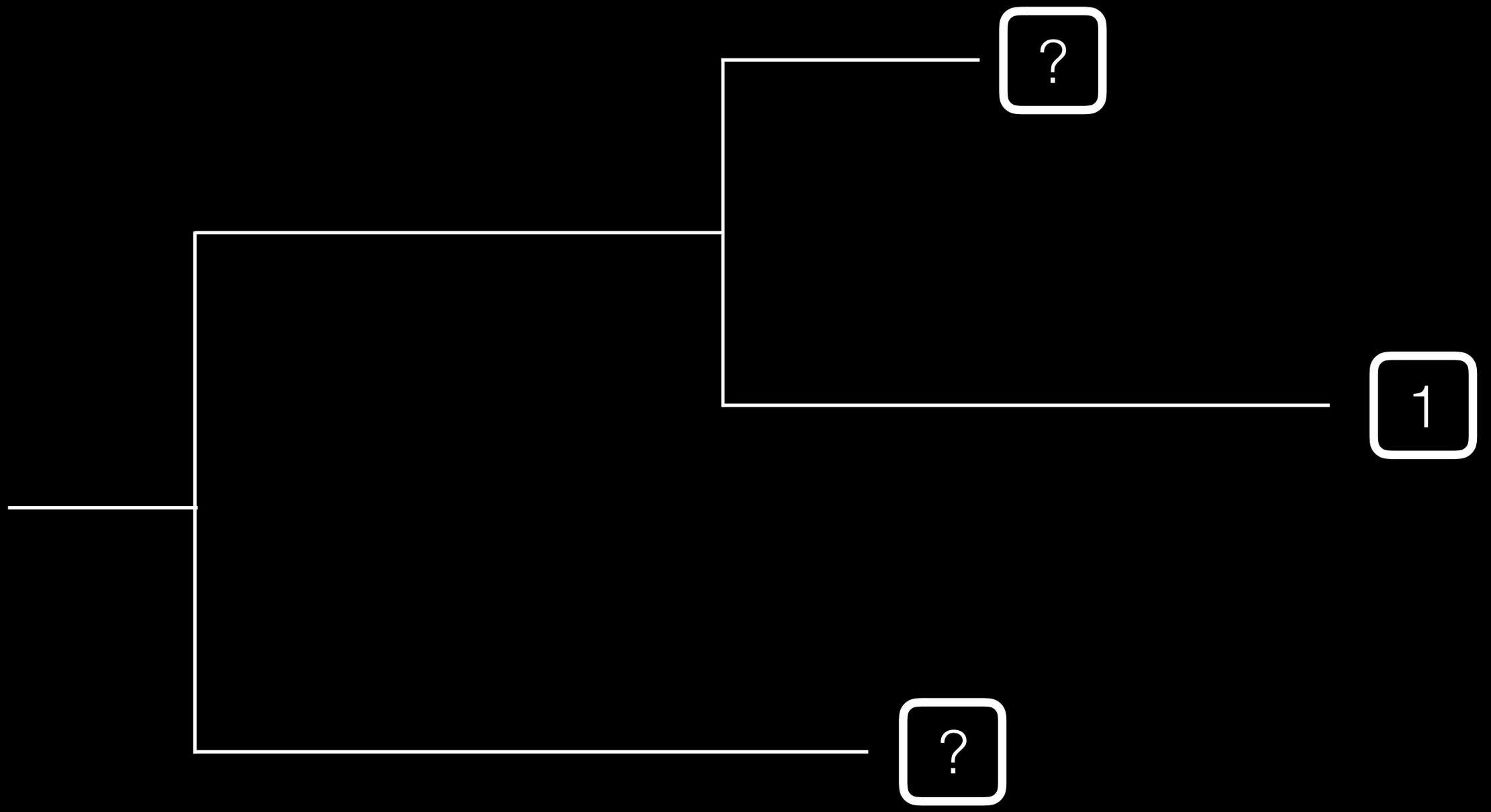
limbless

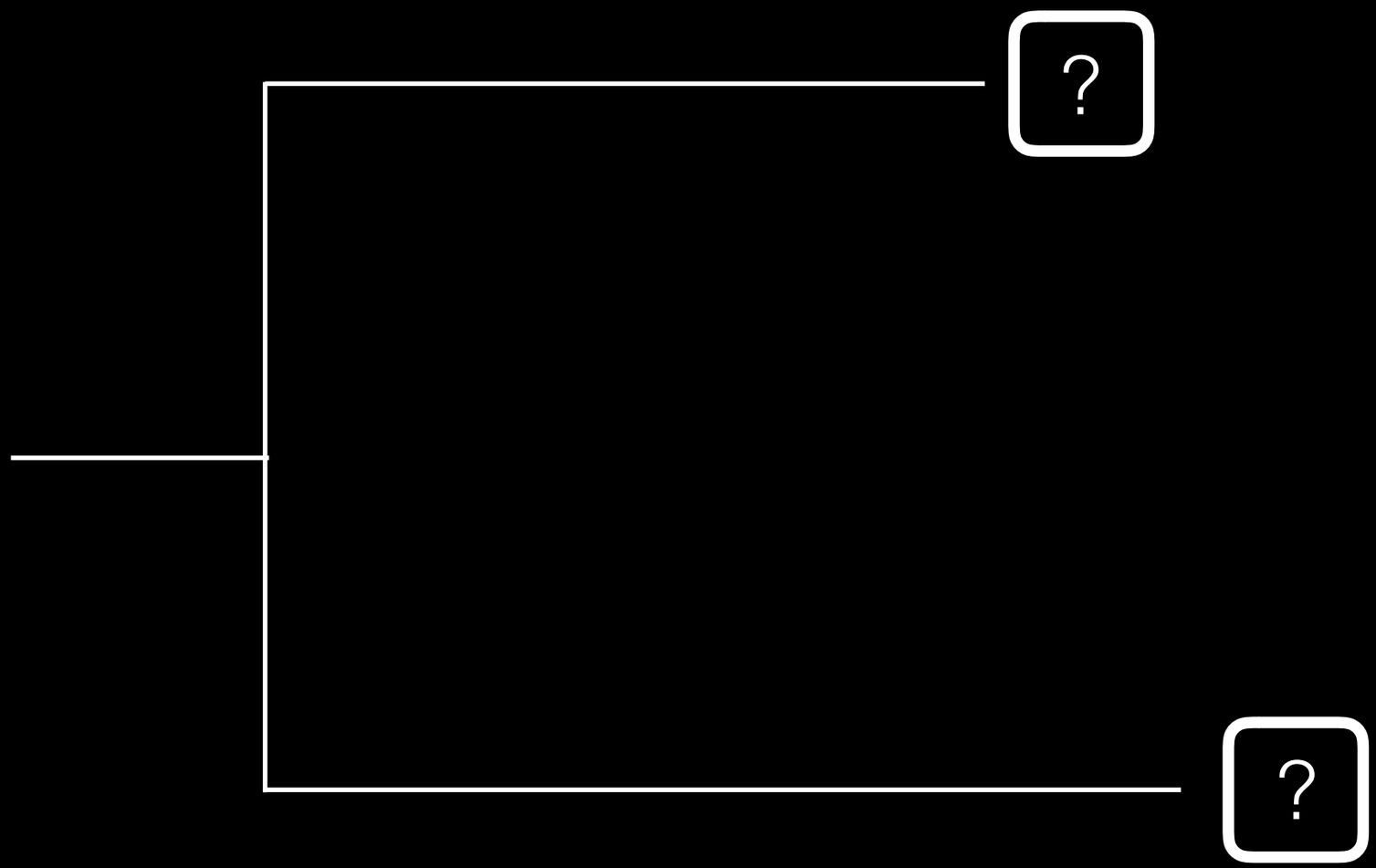

0
1
q_{01}
q_{10}

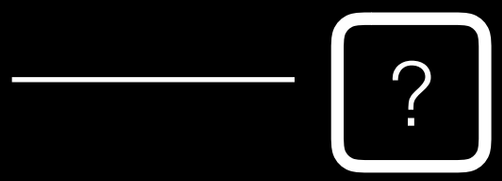
We can calculate the likelihood of the data
given the model and tree
using a **pruning algorithm**


0
0
1
1
0


?
1
1
0

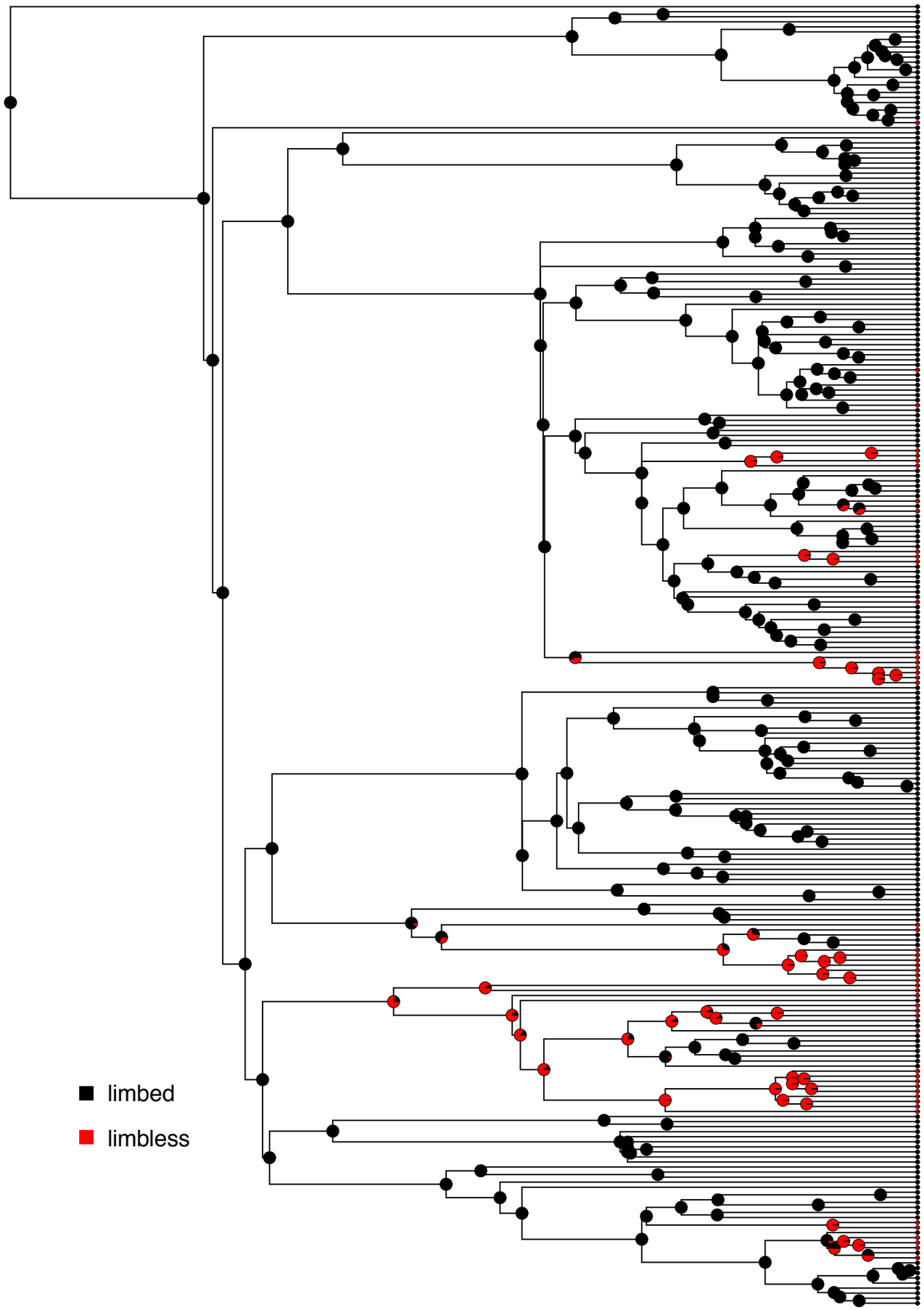

?
1
?





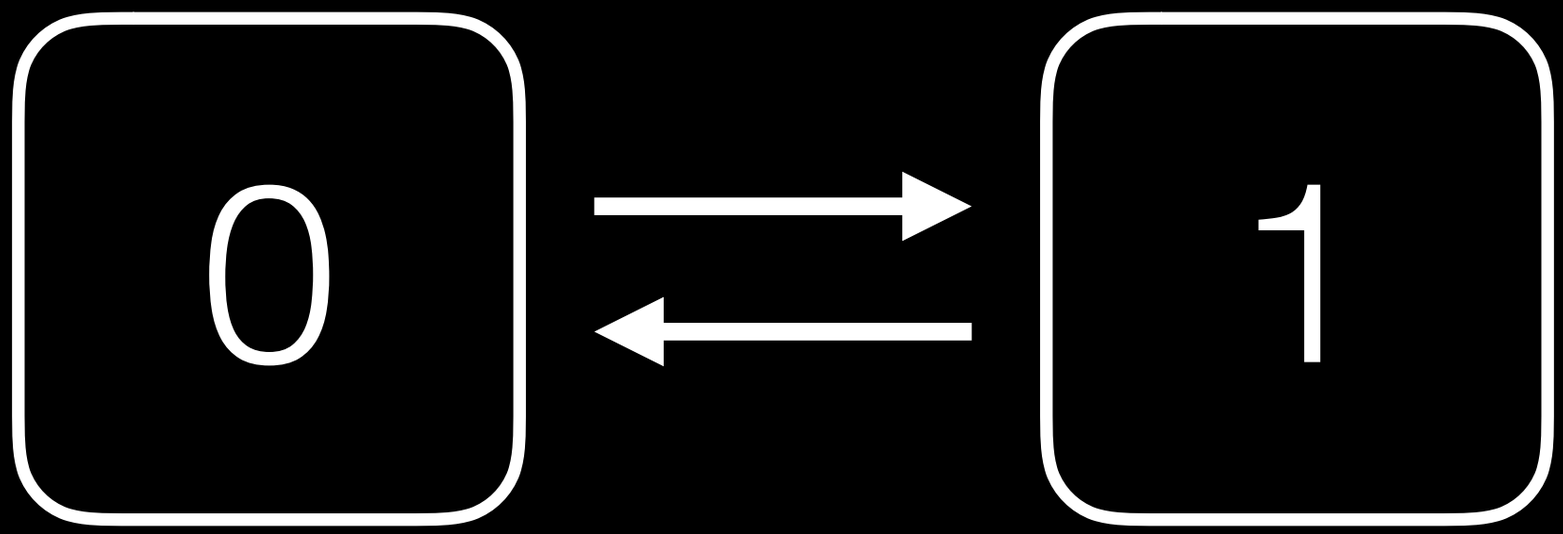
You must be **explicit** about the probability of each state at the root of the tree!

Very common to assume either **equal probability** for all states or that the current state frequencies are in **equilibrium**

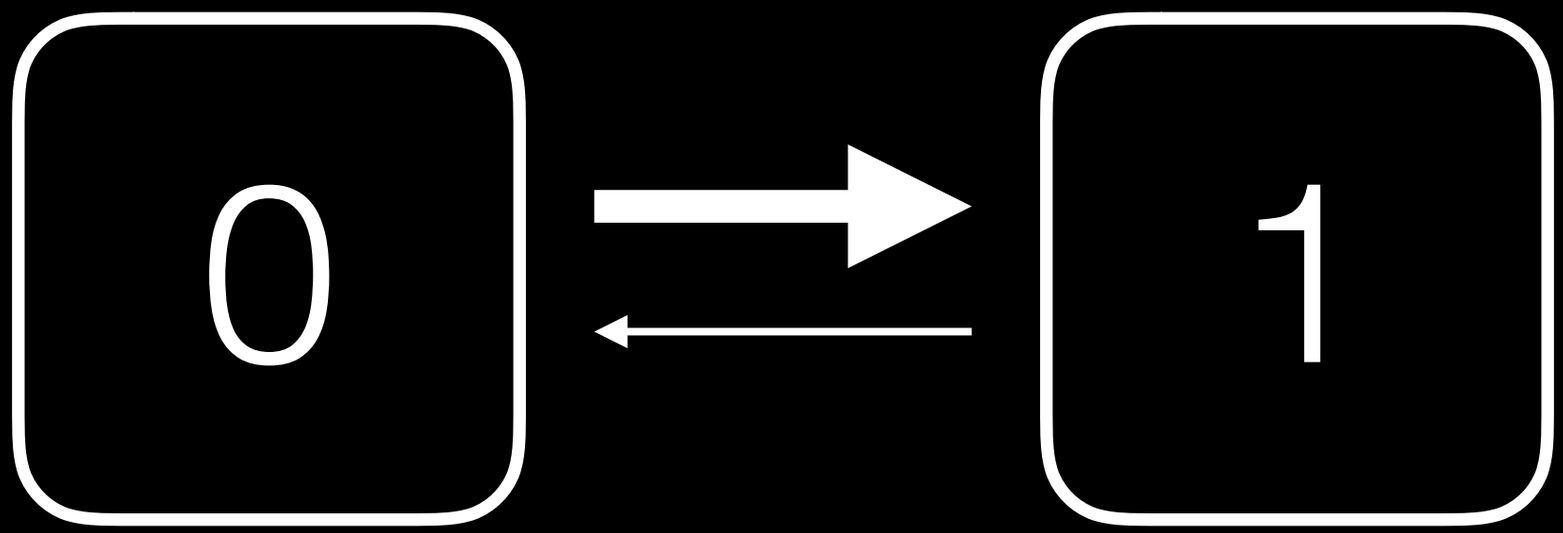

limbed
limbless

Let’s try to fit the model to some example data!

Equal Rates Model



Unequal Rates Model



Equal Rates Model

Unequal Rates Model

Equal Rates Model

lnL = -80.5

$$Q_{ER} = \begin{bmatrix} -0.0019 & 0.0019 \\ 0.0019 & -0.0019 \end{bmatrix}$$

Unequal Rates Model

Equal Rates Model

lnL = -80.5

$$Q_{ER} = \begin{bmatrix} -0.0019 & 0.0019 \\ 0.0019 & -0.0019 \end{bmatrix}$$

Unequal Rates Model

lnL = -79.4

$$Q_{ARD} = \begin{bmatrix} -0.0016 & 0.0016 \\ 0.0038 & -0.0038 \end{bmatrix}$$

Equal Rates Model

lnL = -80.5

$$Q_{ER} = \begin{bmatrix} -0.0019 & 0.0019 \\ 0.0019 & -0.0019 \end{bmatrix}$$

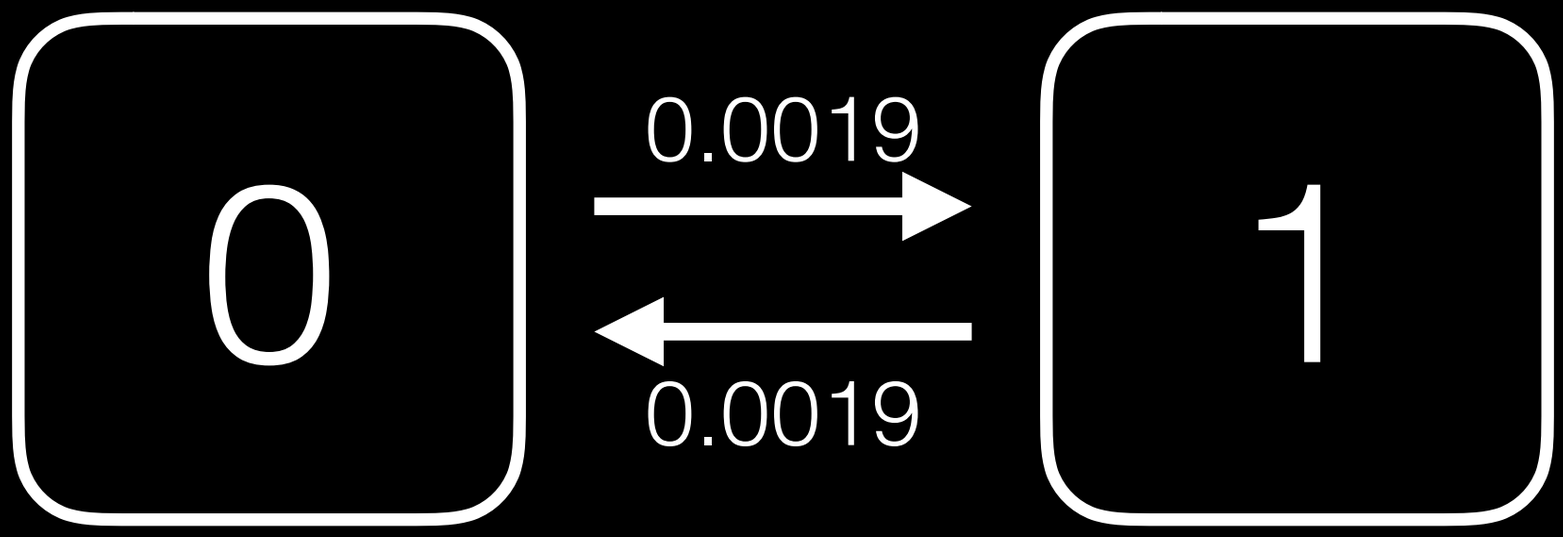
Unequal Rates Model

lnL = -79.4

$$Q_{ARD} = \begin{bmatrix} -0.0016 & 0.0016 \\ 0.0038 & -0.0038 \end{bmatrix}$$

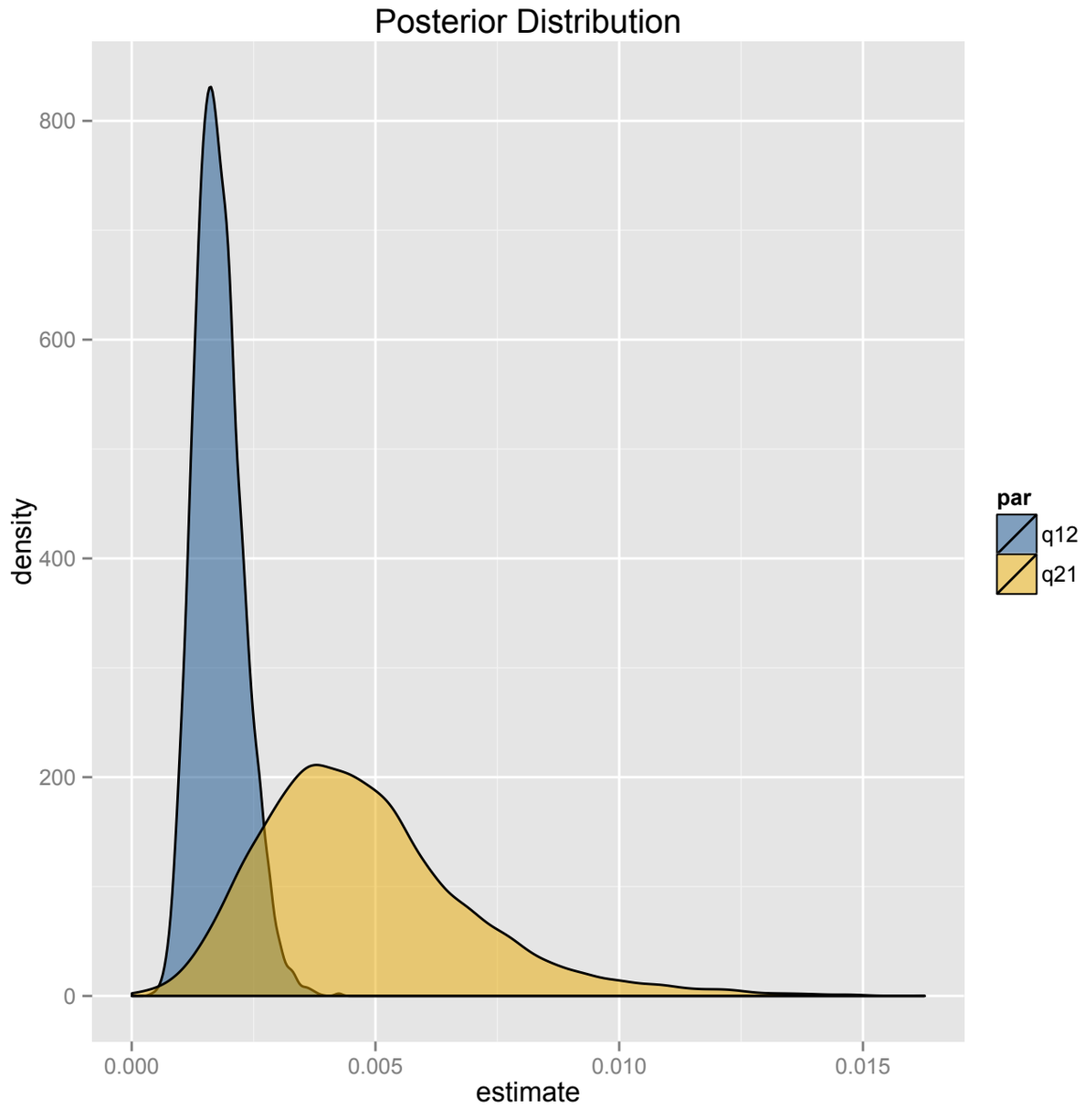
Conclusion: cannot reject ER model

Equal Rates Model



We have to wait about 500 my on average to see a transition

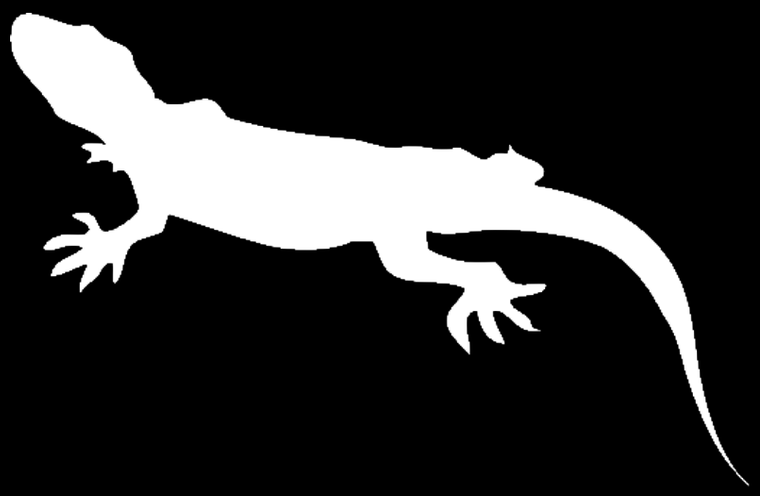
We can also do this with Bayesian analyses

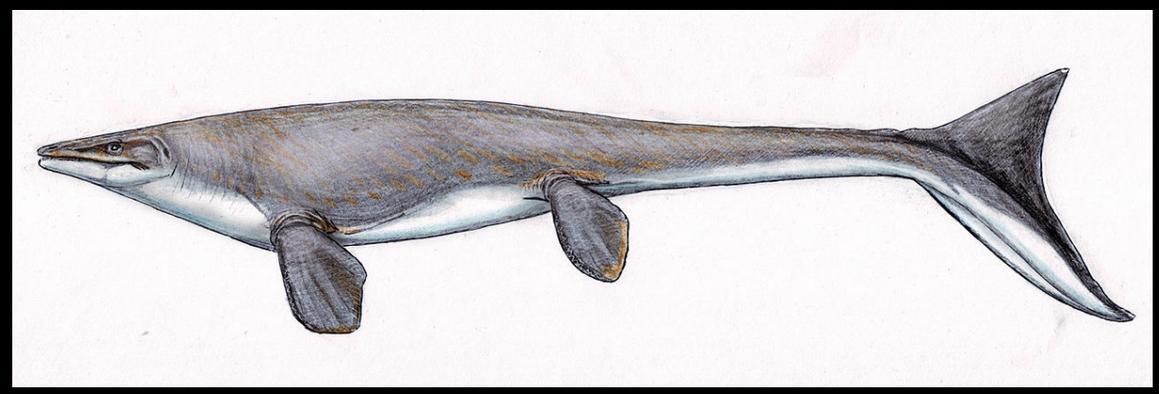


We can also fit a model with more character states

Front limbs only

State 1



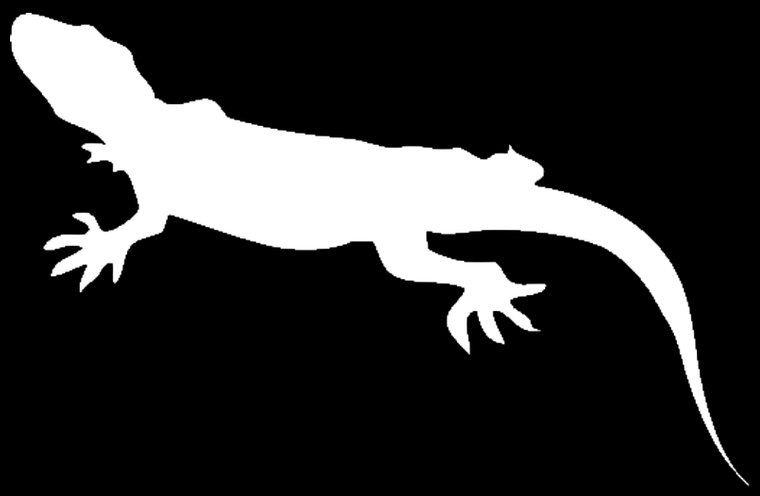


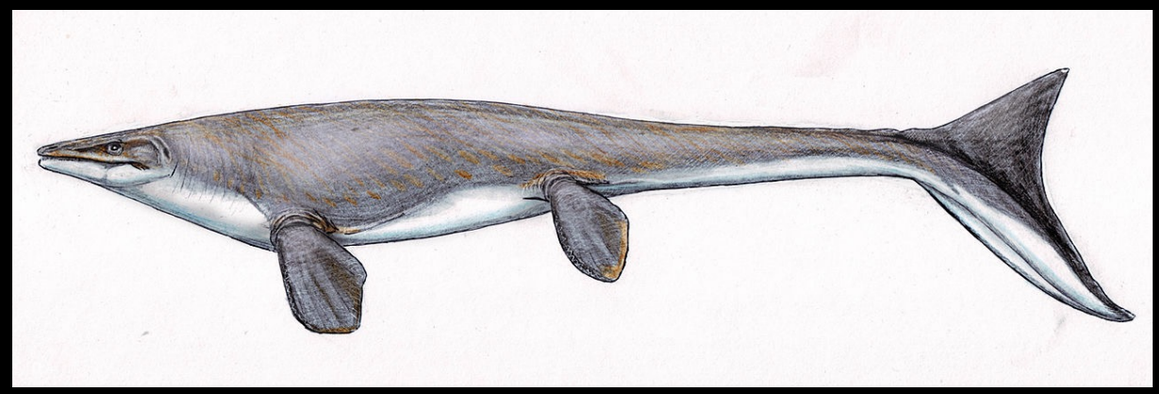
No limbs

State 0

Front limbs only

State 1



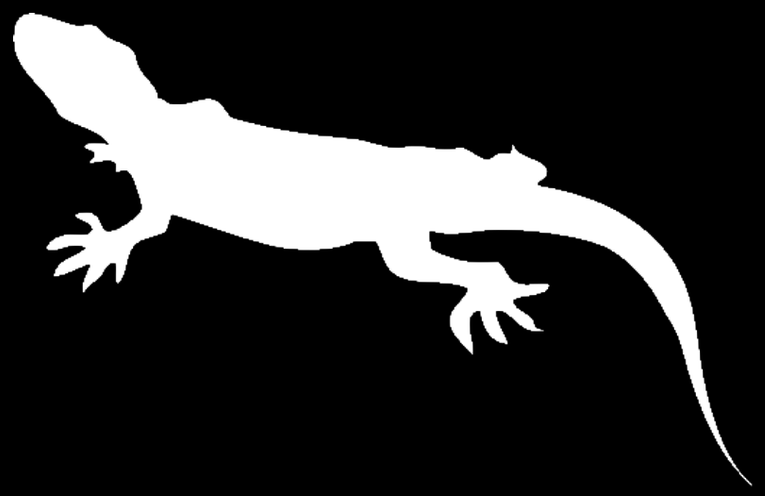


No limbs

State 0

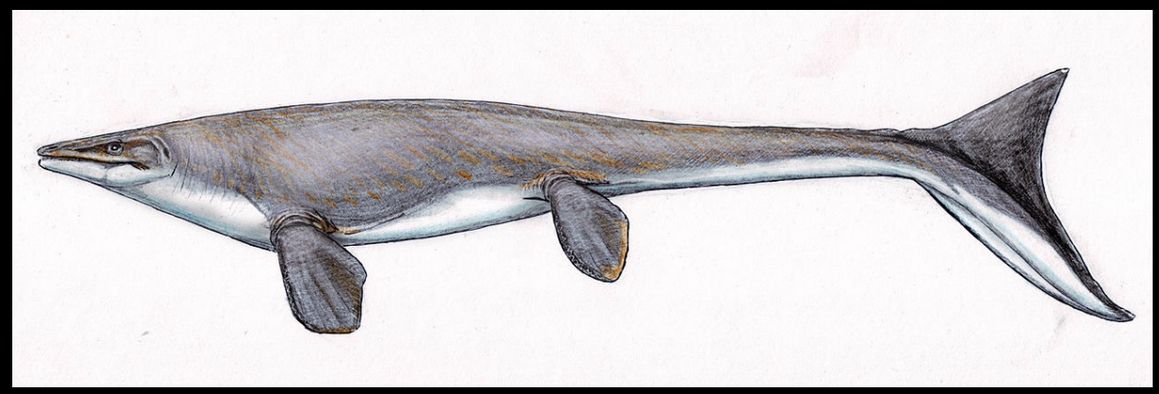
Front limbs only

State 1



Limbs

State 2

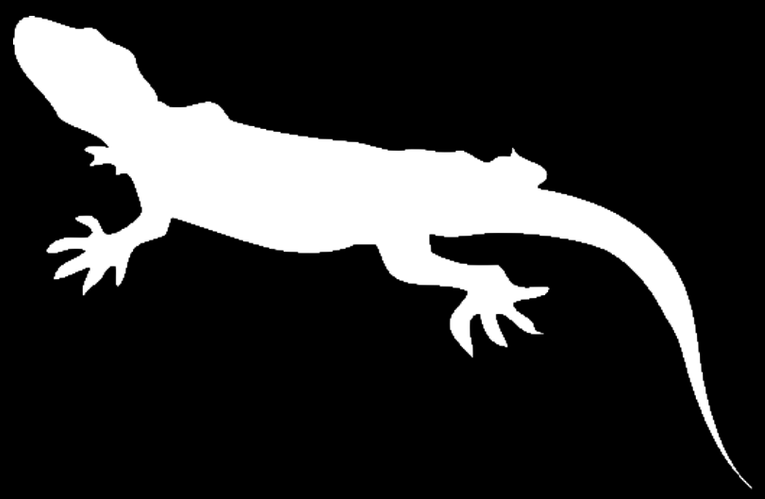


No limbs

State 0

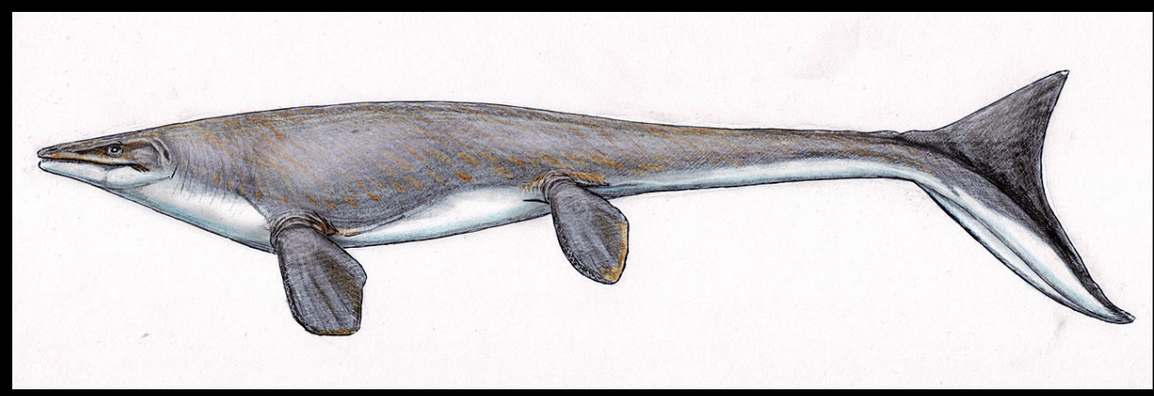
Front limbs only

State 1



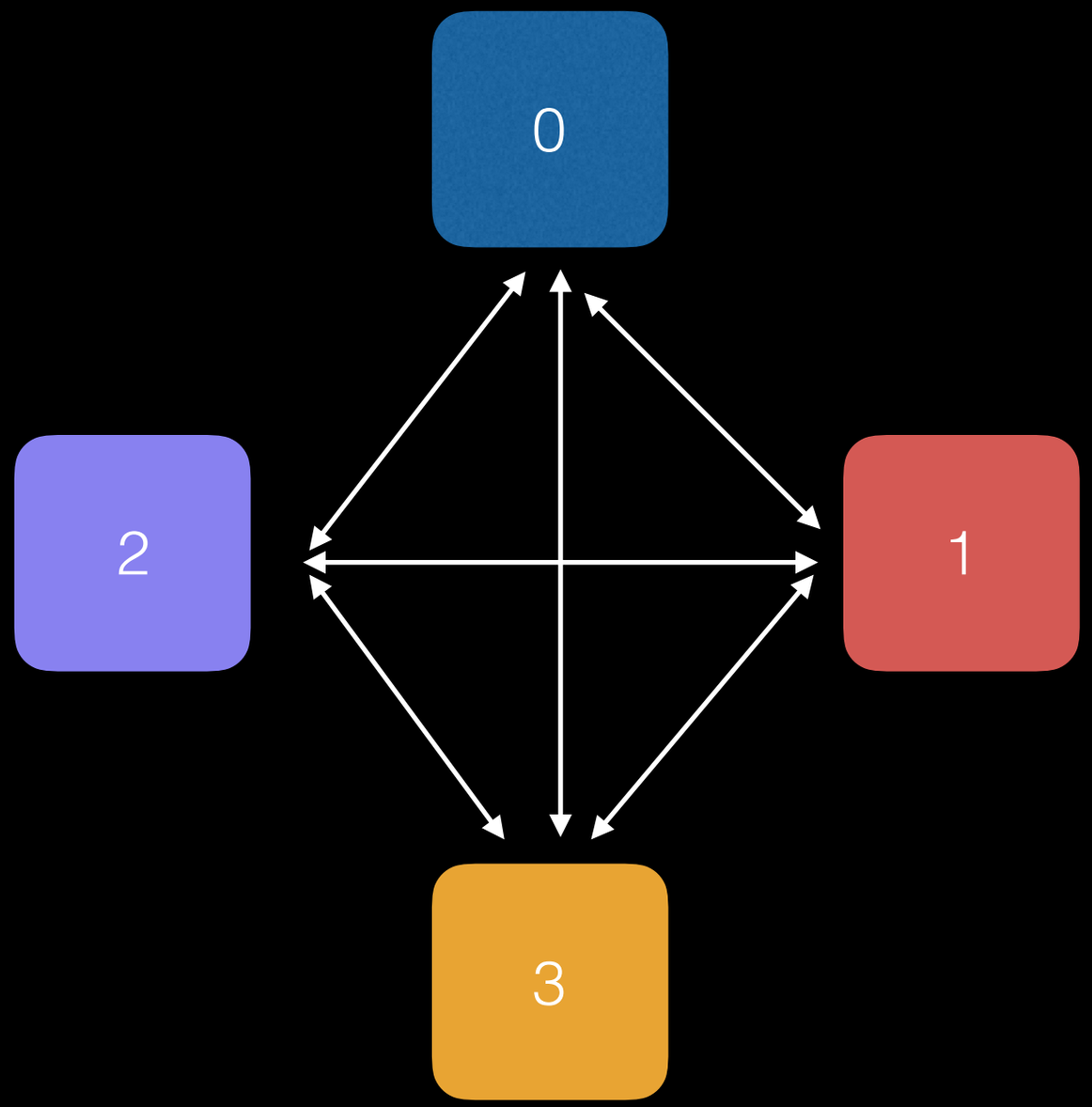
Limbs

State 2



“Fins”

State 3


0
2
1
3

$$Q = \begin{bmatrix} — & q_{01} & q_{02} & q_{03} \\ q_{10} & — & q_{12} & q_{13} \\ q_{20} & q_{21} & — & q_{23} \\ q_{30} & q_{31} & q_{32} & — \end{bmatrix}$$

$$Q = \begin{bmatrix} — & q_{01} & q_{02} & q_{03} \\ q_{10} & — & q_{12} & q_{13} \\ q_{20} & q_{21} & — & q_{23} \\ q_{30} & q_{31} & q_{32} & — \end{bmatrix}$$

— is the number that makes the row sums 0
e.g. $-(q_{01} + q_{02} + q_{03})$

$$Q = \begin{bmatrix} - & q_{01} & q_{02} & q_{03} \\ q_{10} & - & q_{12} & q_{13} \\ q_{20} & q_{21} & - & q_{23} \\ q_{30} & q_{31} & q_{32} & - \end{bmatrix}$$

ER = equal rates

1 rate parameter

$$Q = \begin{bmatrix} - & q_{01} & q_{02} & q_{03} \\ q_{10} & - & q_{12} & q_{13} \\ q_{20} & q_{21} & - & q_{23} \\ q_{30} & q_{31} & q_{32} & - \end{bmatrix}$$

SYM = symmetric

6 rate parameters

$$Q = \begin{bmatrix} - & q_{01} & q_{02} & q_{03} \\ q_{10} & - & q_{12} & q_{13} \\ q_{20} & q_{21} & - & q_{23} \\ q_{30} & q_{31} & q_{32} & - \end{bmatrix}$$

ARD = all rates different

12 rate parameters

We can use the Mk model to test a wide variety of evolutionary hypotheses about discrete characters