

ForBio course: Introduction to phylogenetic methods

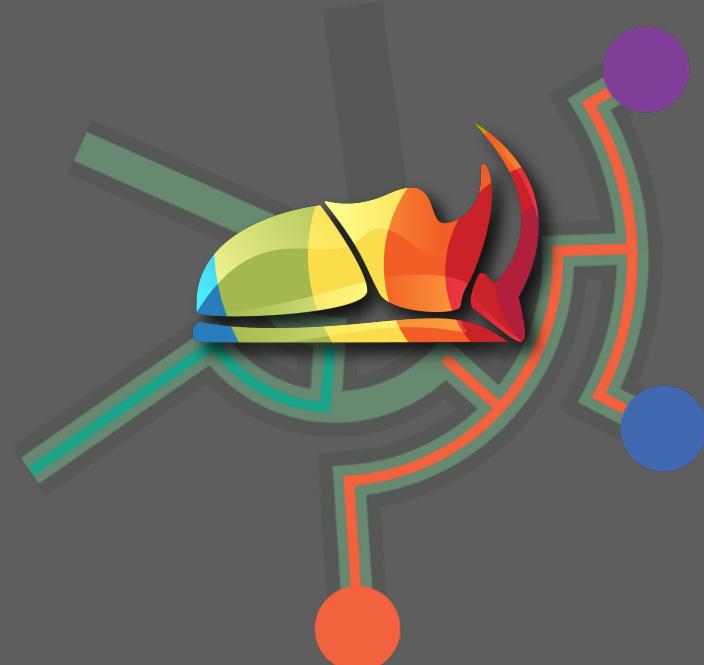
# Modeling Evolution of Discrete Traits on Trees

---

Sergei Tarasov

Beetle curator

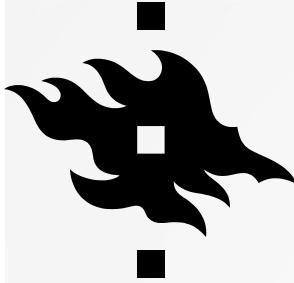
Finnish Museum of Natural History, University of Helsinki



- @tarasov\_sergio

- sergei.tarasov@helsinki.fi

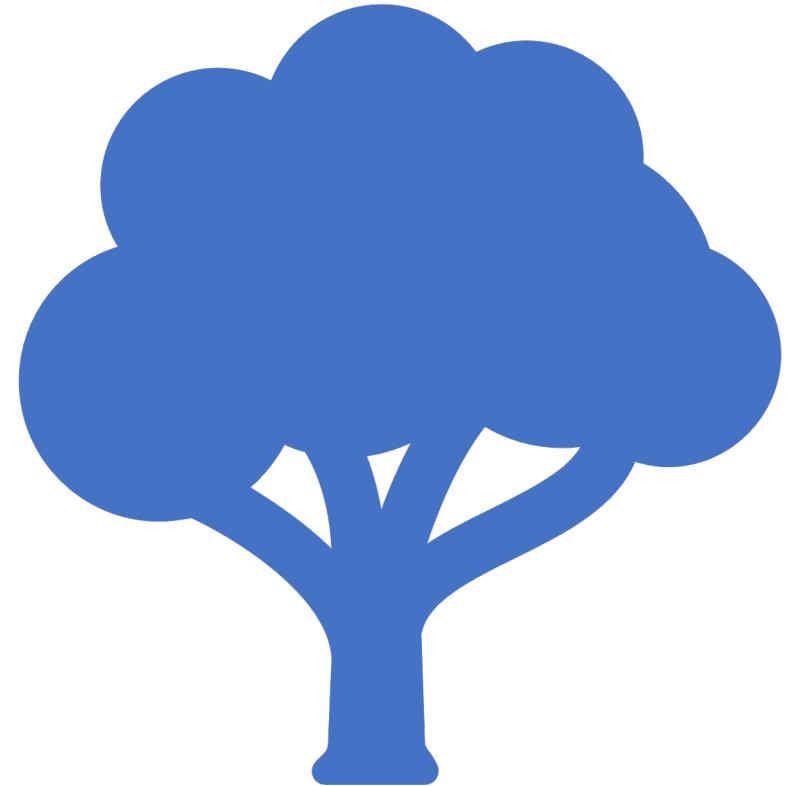
- <https://www.tarasovlab.com>



# PLAN OF THE TODAY'S LECTURE

**Aim:** Learn how to model Discrete Traits in phylogenetic context

1. General Intro
2. Ancestral character state reconstruction
3. Character correlation
4. Hidden Markov models
5. Reconstructing ancestral anatomies



Discrete Traits

# Discrete Traits



Squamates, legged and legless.

[https://lukejharmon.github.io/pcm/chapter7\\_introdiscrete/](https://lukejharmon.github.io/pcm/chapter7_introdiscrete/)

# Categorical traits: biological questions to address

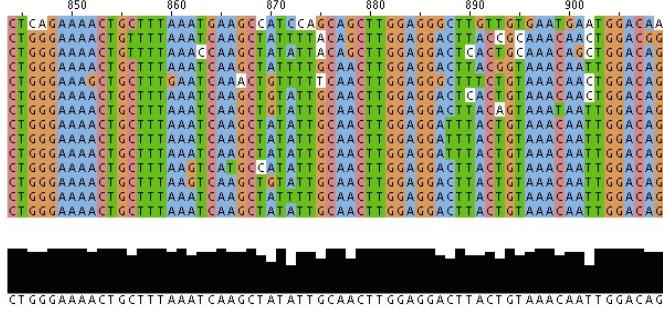
Reconstruct ancestral states

Rate of trait evolution over tree (equal rate vs. heterogeneous rate)

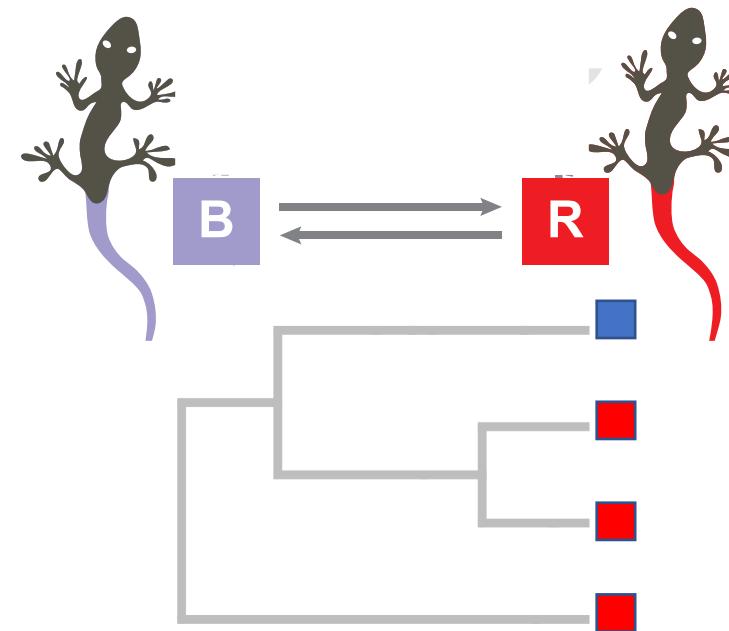
Trait correlation

Correlation between trait and diversification (extinctions, speciation)

# Molecular data vs. Traits



Many sites



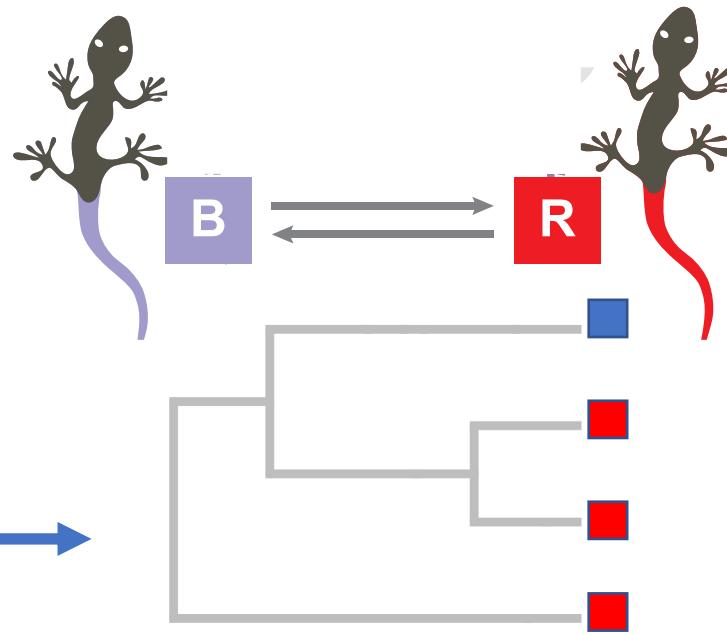
One site

# Phylogeny is given prior to analysis

850      860      870      880      890      900  
CTCAGAAAACTCTTTAATGAAGCATCAGAGCTGGAGGCTTGTGAATGAAAGGACCG  
CTGGGGAAACTGTTTAAATCAAGCTATTACAGCTTGGAGGACTTACCGCAAAACAGCTGGACAG  
CTGGGGAAACTGCTTTAAATCAAGCTATTACAGCTTGGAGGACTTACCGCAAAACAGCTGGACAG  
CTGGGGAAAGCTCTTGAAATCAAACGTGTTTCAGACTGGAGGCTTCTGTAAACCAACTGGACAG  
CTGGGGAAACTGCTTTAAATCAAGCTGTATTGCAACTTGGAGGACTTACAGTAACAAATTGGACAG  
CTGGGGAAACTGCTTTAAATCAAGCTGTATTGCAACTTGGAGGACTTACAGTAACAAATTGGACAG  
CTGGGGAAACTGCTTTAAATCAAGCTGTATTGCAACTTGGAGGACTTACAGTAACAAATTGGACAG  
CTGGGGAAACTGCTTTAAATCAAGCTGTATTGCAACTTGGAGGACTTACAGTAACAAATTGGACAG  
CTGGGGAAACTGCTTTAAATCAAGCTGTATTGCAACTTGGAGGACTTACAGTAACAAATTGGACAG  
CTGGGGAAACTGCTTTAAATCAAGCTGTATTGCAACTTGGAGGACTTACAGTAACAAATTGGACAG  
CTGGGGAAACTGCTTTAAATCAAGCTGTATTGCAACTTGGAGGACTTACAGTAACAAATTGGACAG  
CTGGGGAAACTGCTTTAAATCAAGCTGTATTGCAACTTGGAGGACTTACAGTAACAAATTGGACAG



CTGGGGAAACTGCTTTAAATCAAGCTATATTGCAACTTGGAGGACTTACTGTAAACAAATTGGACAG



Many sites

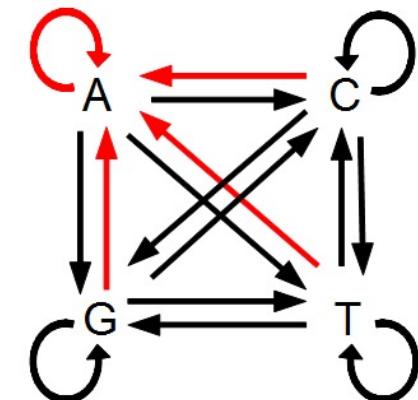
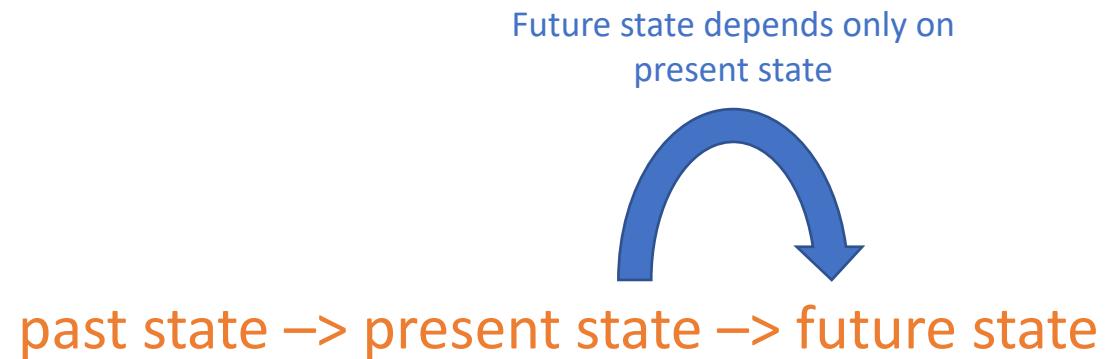
One site

# Modeling phylogenetic process: Markov models (Markov chains)

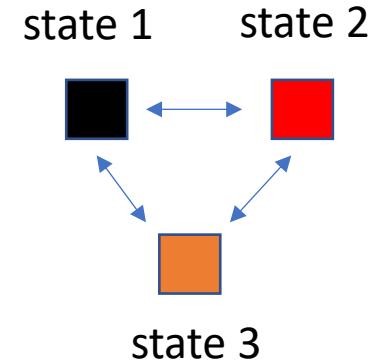
- In probability theory, a Markov model is a stochastic model used to model randomly changing systems
- Markov model assumes that future states depend only on the current state, not on the events that occurred before it (Markov property)



Andrey Markov (1856 – 1922)



# Continuous-time Markov models: transition rate matrix

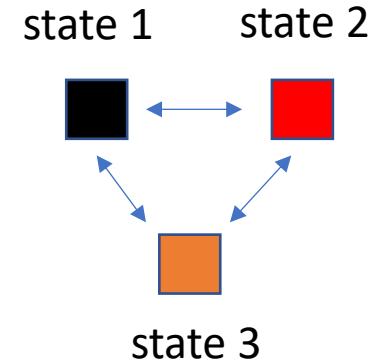


To:

$$\begin{array}{ccc} & \text{[black square]} & \text{[red square]} \\ \text{From:} & \text{[black square]} & \left[ \begin{array}{ccc} -\lambda_1 & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & -\lambda_2 & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & -\lambda_3 \end{array} \right] \\ & \text{[red square]} & \\ & \text{[orange square]} & \end{array}$$

**Transition rate matrix  
consist of infinitesimal  
rates of change**

# Continuous-time Markov models: transition rate matrix



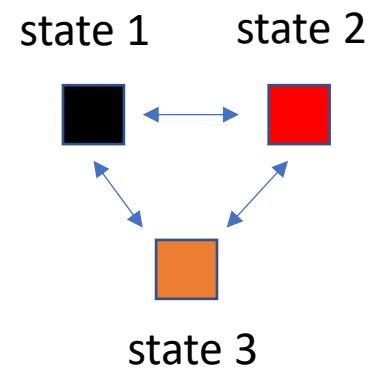
To:

From:

$$\begin{matrix} & \begin{matrix} \blacksquare & \square & \square \end{matrix} \\ \blacksquare & \left[ \begin{matrix} -\lambda_1 & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & -\lambda_2 & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & -\lambda_3 \end{matrix} \right] \\ \square & \\ \square & \end{matrix}$$

Transition rate matrix  
consist of infinitesimal  
rates of change

Meaning of the rates?



# From rates to probabilities

- **Transition rate matrix.** Infinitesimal rates

$$Q = \begin{bmatrix} -0.5 & 0.4 & 0.1 \\ 0.8 & -1 & 0.2 \\ 0.96 & 0.24 & -1.2 \end{bmatrix}$$

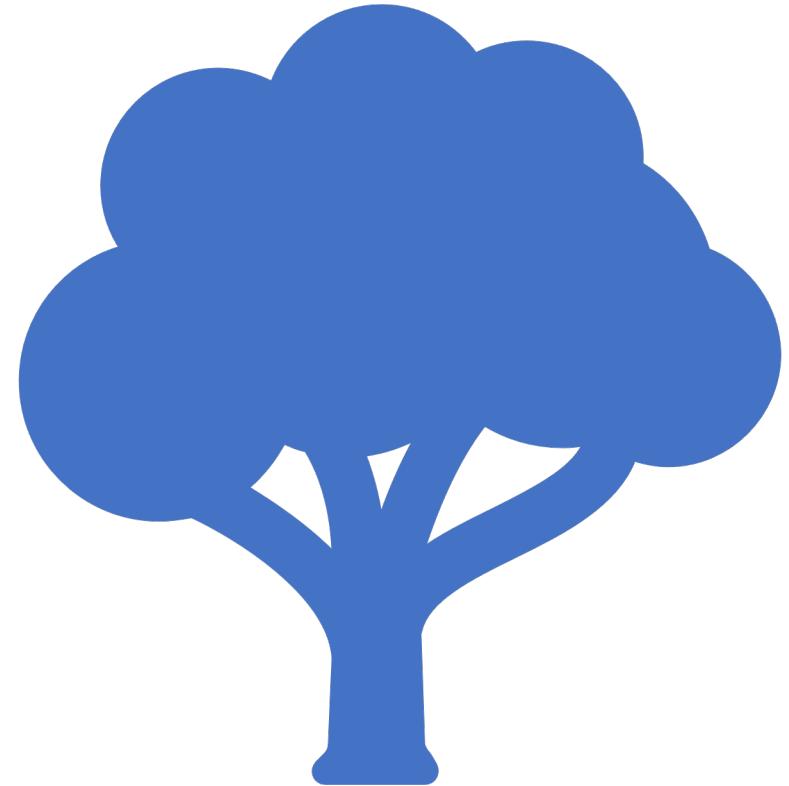
Exponentiate rate matrix

- **Probability transition matrix.**

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

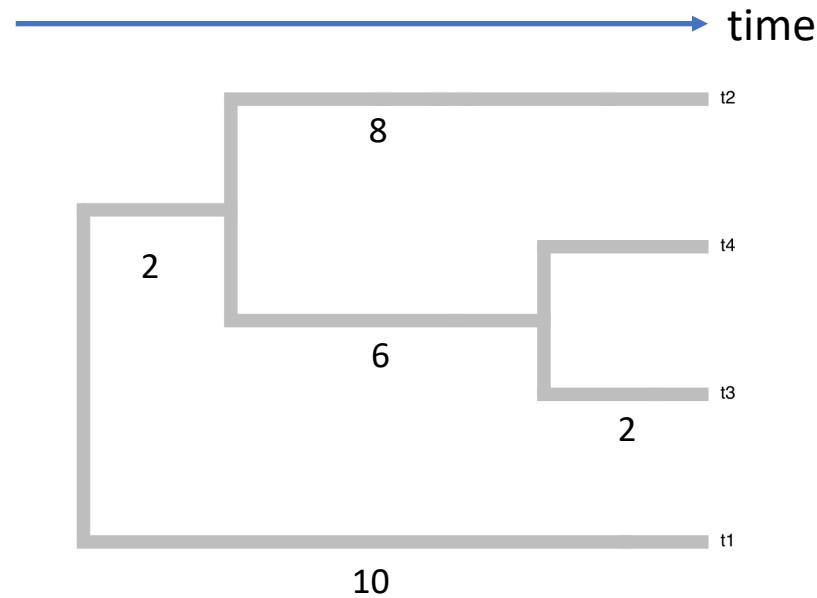
$$e^{Q*1} = \begin{bmatrix} 0.72 & 0.2 & 0.08 \\ 0.46 & 0.46 & 0.08 \\ 0.46 & 0.2 & 0.34 \end{bmatrix}$$

Matrix exponential:  $e^{Qt} = 1 + \frac{Qt^1}{1!} + \frac{Qt^2}{2!} + \frac{Qt^3}{3!} + \dots$



Markov modes:  
Simulation and Inference

# Simulating data under Markov models on a tree



Random number  
generator

$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

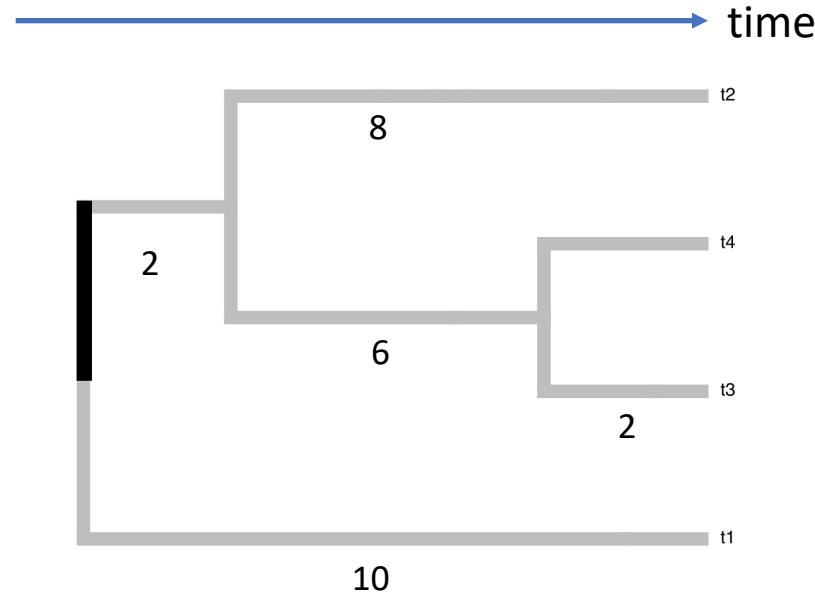
state 1  $\xrightarrow{\quad}$  state 2



Initial vector

$$\pi = (1/2, 1/2)$$

# Simulating data under Markov models on a tree



Random number  
generator

1. Randomly select state at the root from a uniform distribution.  
RND=0.4 (starting state 1)

$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

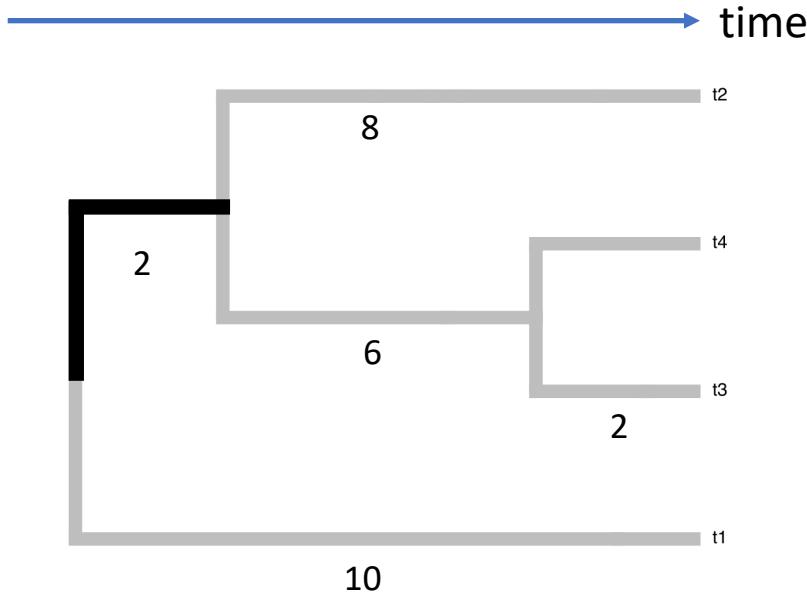
state 1      state 2

A diagram showing a transition between two states. On the left is a black square labeled "state 1". An arrow points from it to a red square labeled "state 2". Another arrow points back from the red square to the black square, indicating a reversible transition.

Initial vector  
 $\pi = (1/2, 1/2)$

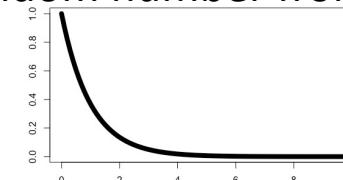
A diagram showing the initial vector  $\pi = (1/2, 1/2)$ . It consists of two squares side-by-side: a black square on the left and a red square on the right.

# Simulating data under Markov models on a tree



Random number  
generator

1. Randomly select state at the root from a uniform distribution.  
RND=0.4 (starting state 1)
2. Draw a random number from Exponential distribution with  $\lambda = 1$ .  
RND=2.4



$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

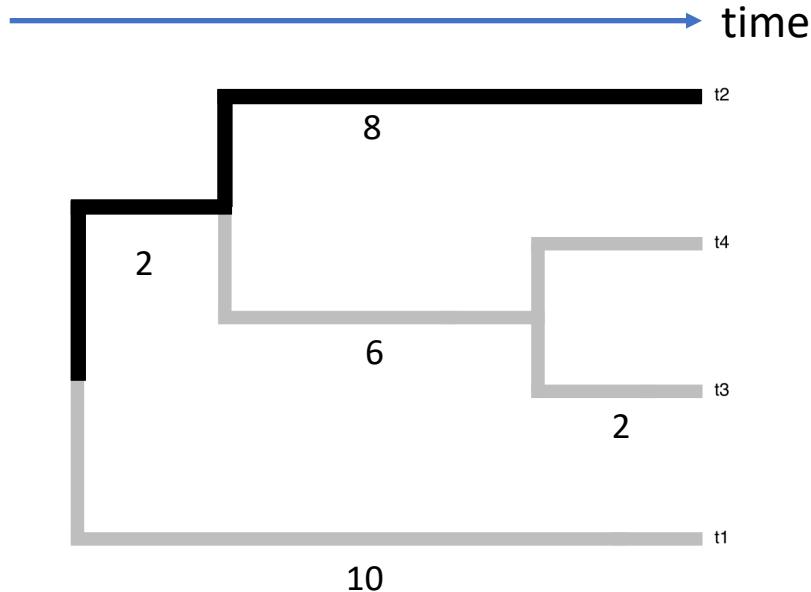
state 1      state 2

A diagram showing state transitions between two states. A black square labeled "state 1" is connected by a blue arrow pointing right to a red square labeled "state 2". Another blue arrow points back from "state 2" to "state 1".

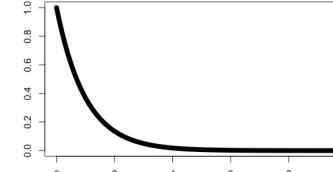
Initial vector  
 $\pi = (1/2, 1/2)$

A diagram showing the initial vector  $\pi = (1/2, 1/2)$  as a mixture of black and red squares. It consists of two black squares and two red squares arranged side-by-side.

# Simulating data under Markov models on a tree



Random number generator

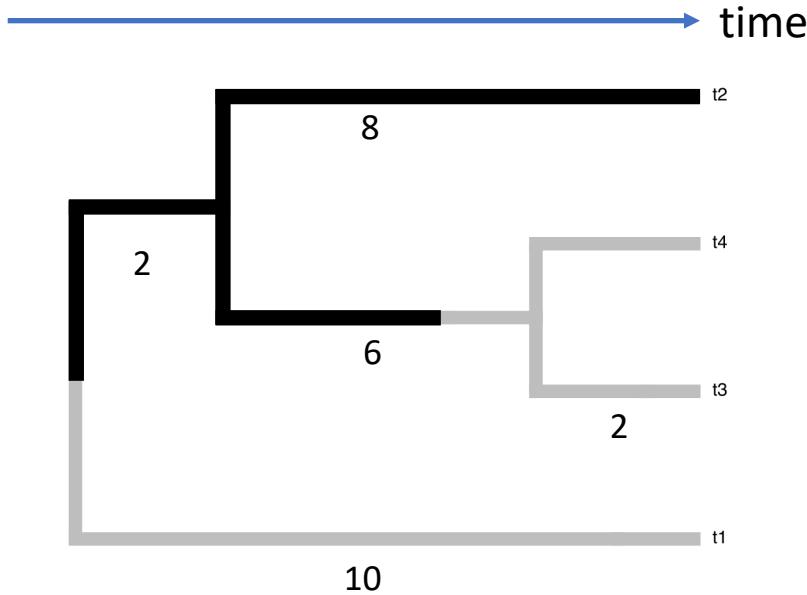
1. Randomly select state at the root from a uniform distribution.  
RND=0.4 (starting state 1)
2. Draw a random number from Exponential distribution with  $\lambda = 1$ .  
RND=2.4
3. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=8.4

$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

state 1  state 2 

Initial vector  
 $\pi = (1/2, 1/2)$   
 

# Simulating data under Markov models on a tree



$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

state 1      state 2

Two small squares, one black and one red, are connected by a double-headed blue arrow, indicating a reversible transition between state 1 and state 2.

Initial vector  
 $\pi = (1/2, 1/2)$

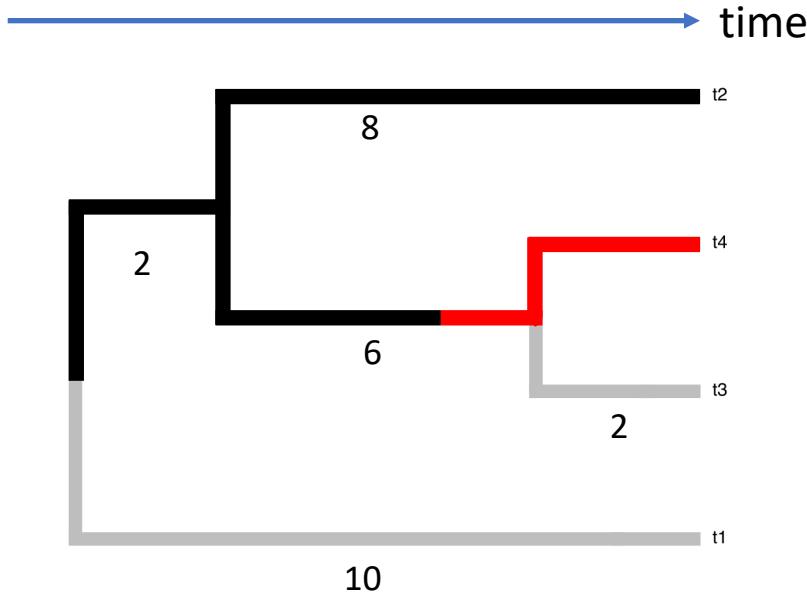


Random number generator

1. Randomly select state at the root from a uniform distribution.  
RND=0.4 (starting state 1)
2. Draw a random number from Exponential distribution with  $\lambda = 1$ .  
RND=2.4

A graph showing the probability density function of an exponential distribution with  $\lambda = 1$ . The x-axis ranges from 0 to 10, and the y-axis ranges from 0.0 to 1.0. The curve starts at (0, 1.0) and decays rapidly towards zero as x increases.
3. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=8.4
4. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=4.2 (to state 2)

# Simulating data under Markov models on a tree



$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

state 1      state 2

Two states are shown: state 1 (black square) and state 2 (red square). A blue arrow points from state 1 to state 2, and another blue arrow points from state 2 back to state 1.

Initial vector  
 $\pi = (1/2, 1/2)$

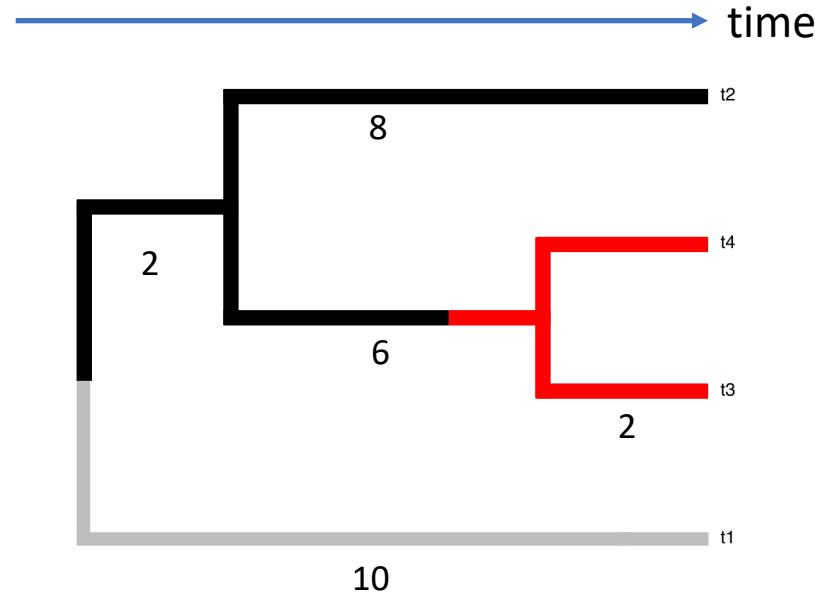


Random number generator

1. Randomly select state at the root from a uniform distribution.  
RND=0.4 (starting state 1)
2. Draw a random number from Exponential distribution with  $\lambda = 1$ .  
RND=2.4

A graph showing the probability density function of an exponential distribution with  $\lambda = 1$ . The x-axis ranges from 0 to 10, and the y-axis ranges from 0.0 to 1.0. The curve starts at (0, 1.0) and decays rapidly towards zero as x increases.
3. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=8.4
4. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=4.2 (to state 2)
5. Draw a random number from  $\text{Exp}(\lambda = 2)$ . RND=4.6

# Simulating data under Markov models on a tree



$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

state 1      state 2

A diagram showing two states: state 1 (black square) and state 2 (red square). A blue arrow points from state 1 to state 2, and another blue arrow points from state 2 back to state 1.

Initial vector  
 $\pi = (1/2, 1/2)$

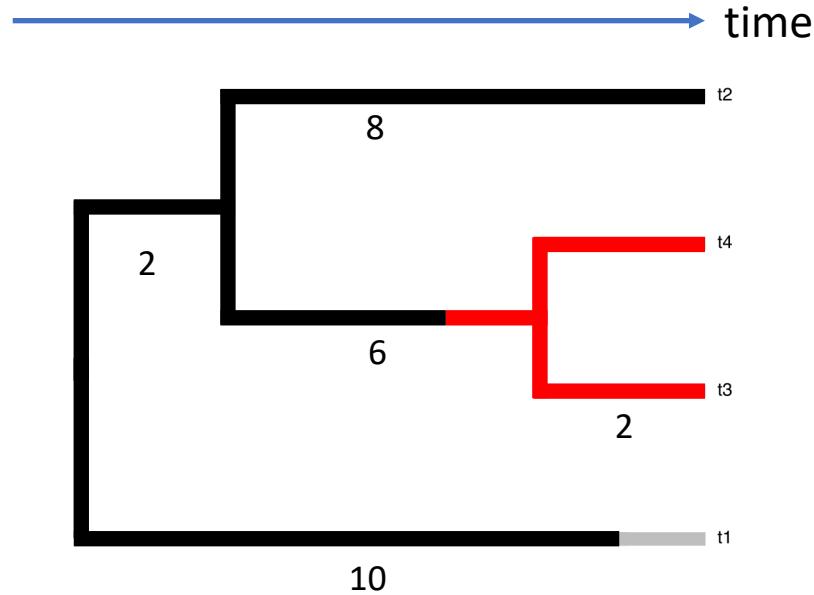


Random number generator

1. Randomly select state at the root from a uniform distribution.  
RND=0.4 (starting state 1)
2. Draw a random number from Exponential distribution with  $\lambda = 1$ .  
RND=2.4

A graph of the exponential distribution probability density function  $f(x) = e^{-x}$  for  $x \geq 0$ . The x-axis ranges from 0 to 10, and the y-axis ranges from 0 to 1.0. The curve starts at (0, 1) and decays rapidly towards zero.
3. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=8
4. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=4.2 (to state 2)
5. Draw a random number from  $\text{Exp}(\lambda = 2)$ . RND=4.6
6. Draw a random number from  $\text{Exp}(\lambda = 2)$ . RND=4.9

# Simulating data under Markov models on a tree



$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

state 1      state 2

A diagram showing two states: state 1 (represented by a black square) and state 2 (represented by a red square). A blue arrow points from state 1 to state 2, indicating a transition probability of 1. A blue arrow also points from state 2 back to state 1, indicating a self-loop transition probability of 2.

Initial vector  
 $\pi = (1/2, 1/2)$

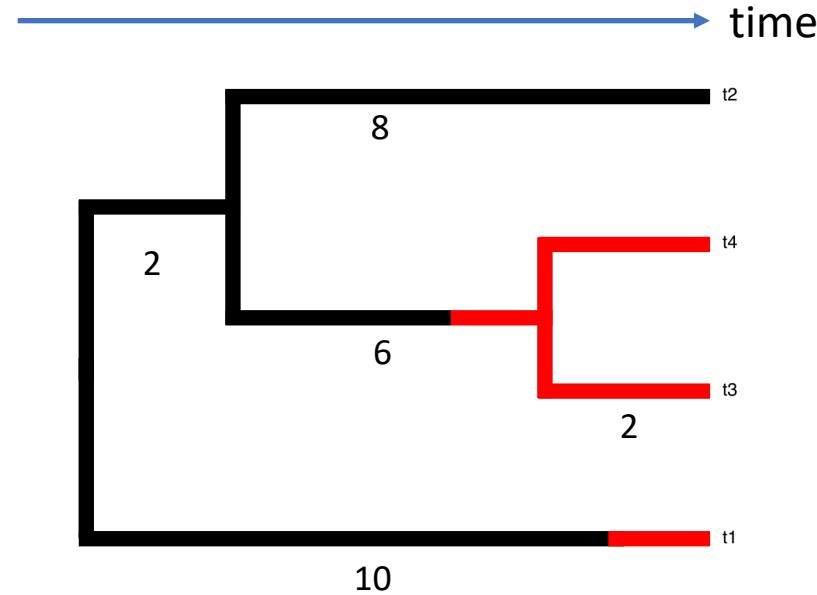


Random number generator

1. Randomly select state at the root from a uniform distribution.  
RND=0.4 (starting state 1)
2. Draw a random number from Exponential distribution with  $\lambda = 1$ .  
RND=2.4

A graph showing the probability density function of an exponential distribution with  $\lambda = 1$ . The x-axis ranges from 0 to 10, and the y-axis ranges from 0 to 1.0. The curve starts at (0, 1) and decays rapidly towards zero.
3. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=8
4. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=4.2 (to state 2)
5. Draw a random number from  $\text{Exp}(\lambda = 2)$ . RND=4.6
6. Draw a random number from  $\text{Exp}(\lambda = 2)$ . RND=4.9
7. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=9.1 (to state 2)

# Simulating data under Markov models on a tree



$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

state 1      state 2

Initial vector

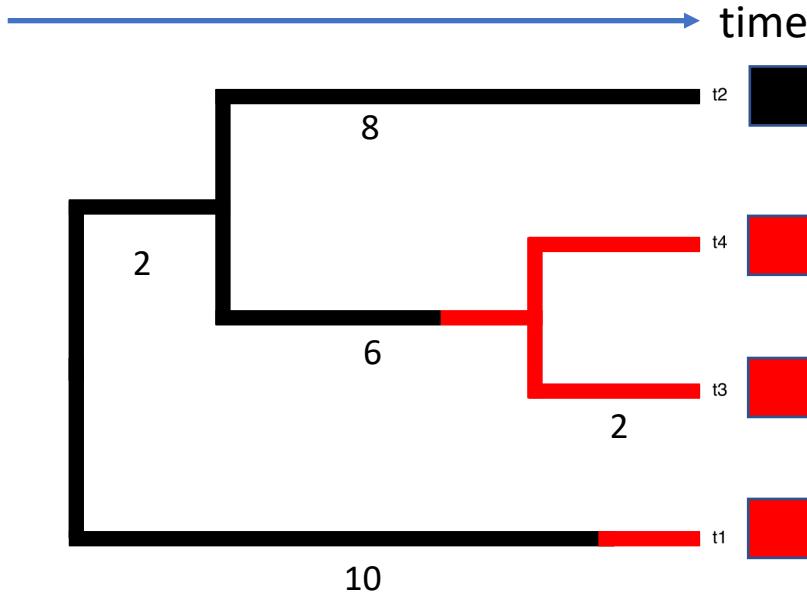
$$\pi = (1/2, 1/2)$$



Random number  
generator

1. Randomly select state at the root from a uniform distribution.  
RND=0.4 (starting state 1)
2. Draw a random number from Exponential distribution with  $\lambda = 1$ .  
RND=2.4
3. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=8
4. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=4.2 (to state 2)
5. Draw a random number from  $\text{Exp}(\lambda = 2)$ . RND=4.6
6. Draw a random number from  $\text{Exp}(\lambda = 2)$ . RND=4.9
7. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=9.1 (to state 2)
8. Draw a random number from  $\text{Exp}(\lambda = 2)$ . RND=3.3

# Simulating data under Markov models on a tree



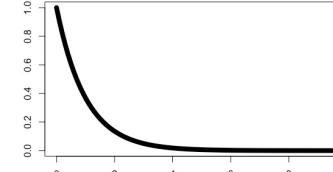
$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

state 1      state 2

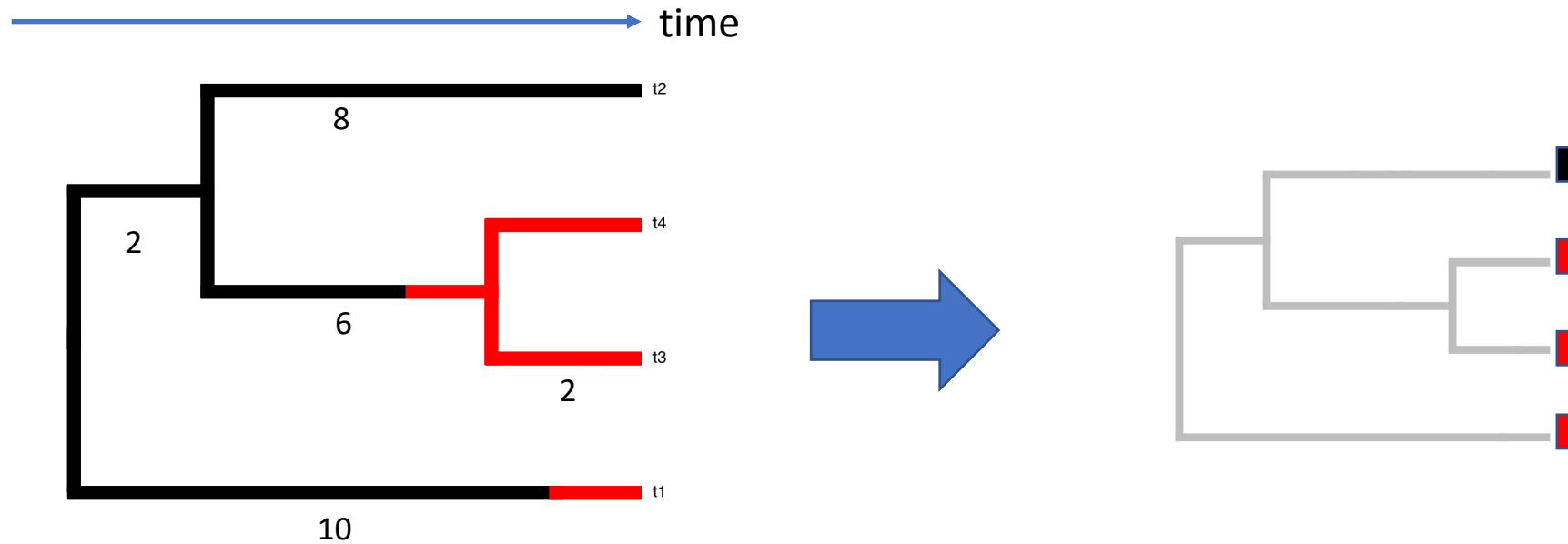
Initial vector  
 $\pi = (1/2, 1/2)$



Random number generator

1. Randomly select state at the root from a uniform distribution.  
RND=0.4 (starting state 1)
2. Draw a random number from Exponential distribution with  $\lambda = 1$ .  
RND=2.4
3. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=8
4. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=4.2 (to state 2)
5. Draw a random number from  $\text{Exp}(\lambda = 2)$ . RND=4.6
6. Draw a random number from  $\text{Exp}(\lambda = 2)$ . RND=4.9
7. Draw a random number from  $\text{Exp}(\lambda = 1)$ . RND=9.1 (to state 2)
8. Draw a random number from  $\text{Exp}(\lambda = 2)$ . RND=3.3

# Now let's run likelihood inference

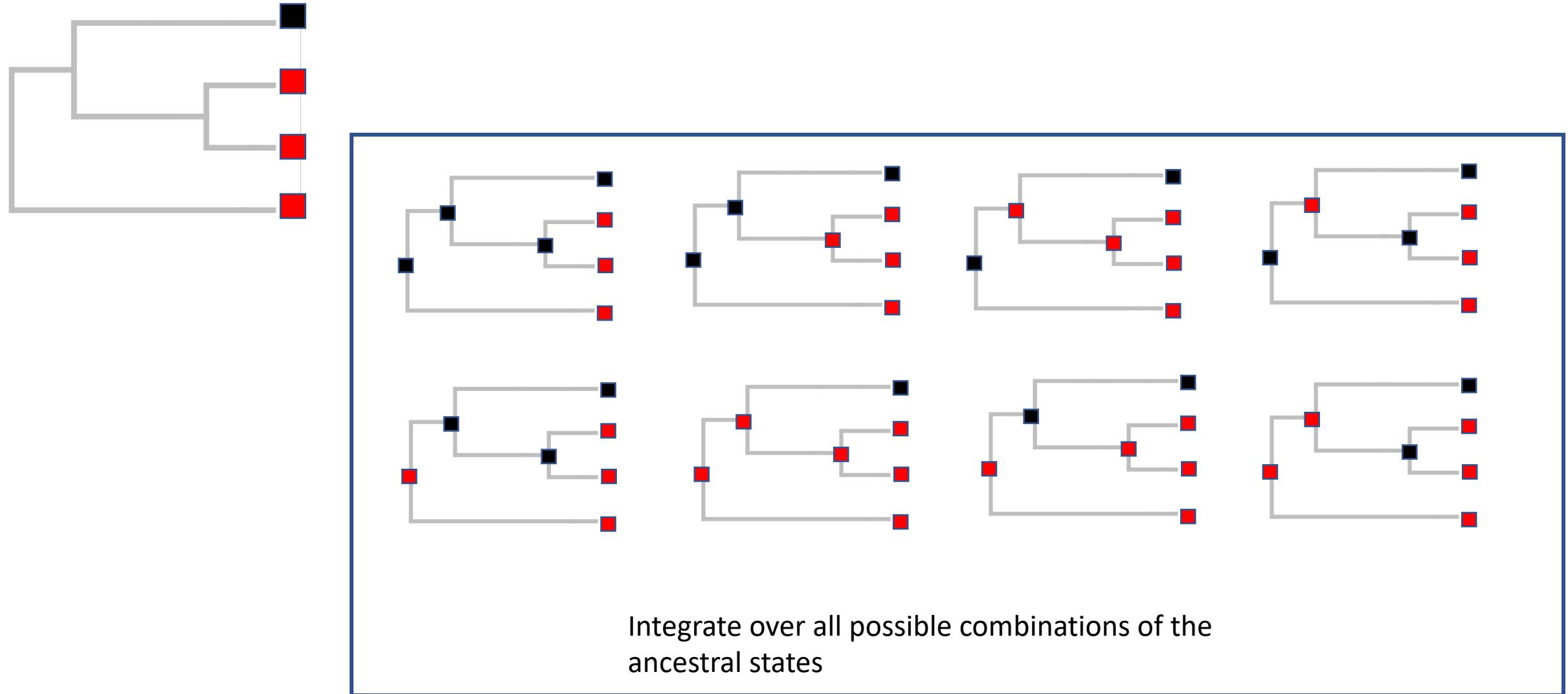


$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

state 1  $\xrightarrow{\hspace{1cm}}$  state 2

Initial vector  
 $\pi = (1/2, 1/2)$

# Inference: estimating tree likelihood



# Felsenstein's pruning algorithm

Given values:

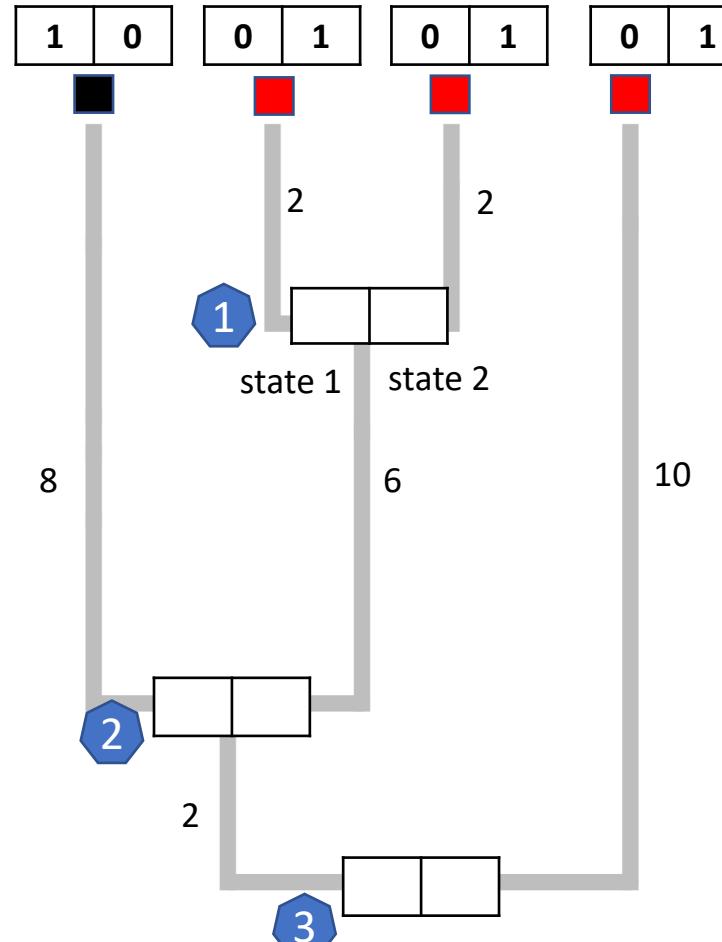
state 1      state 2

1	0
0	1

$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

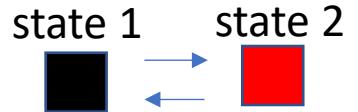


$$\pi = (1/2, 1/2)$$



# Felsenstein's pruning algorithm

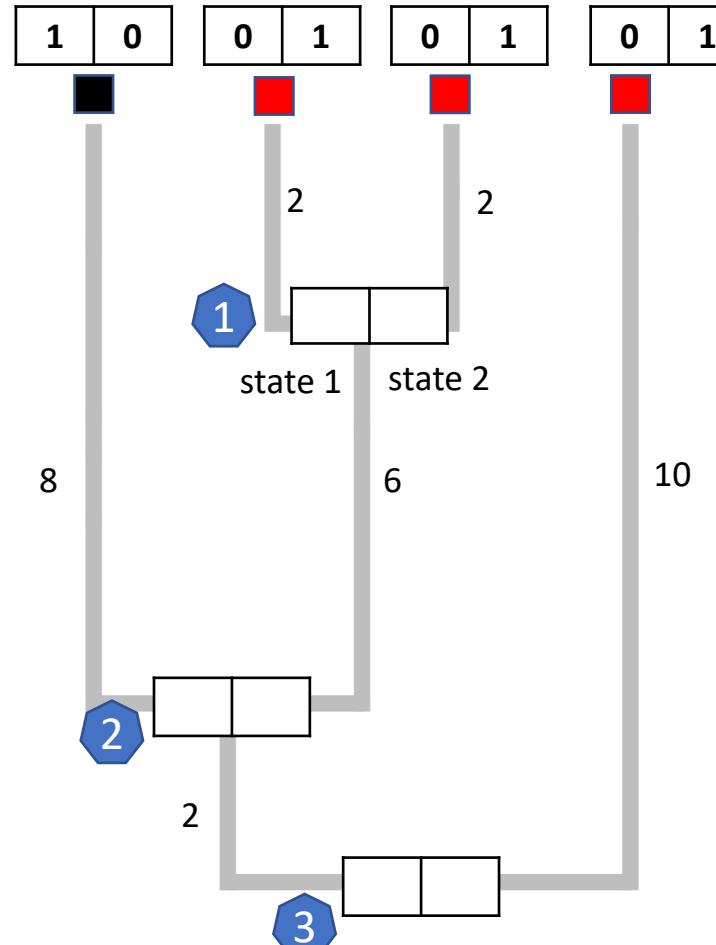
Given values:



1	0
0	1

$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

$$\pi = (1/2, 1/2)$$



Node 1

At state 1 ■

$$\text{Left br. } e^{Q*2} = \begin{bmatrix} 0.66 & 0.34 \\ 0.66 & 0.34 \end{bmatrix}$$

$$\text{Right br. } e^{Q*2} = \begin{bmatrix} 0.66 & 0.34 \\ 0.66 & 0.34 \end{bmatrix}$$

$$\Pr(\text{black}) = 0.34 * 0.34 = 0.12$$

At state 2 ■

$$\text{Left br. } e^{Q*2} = \begin{bmatrix} 0.66 & 0.34 \\ 0.66 & 0.34 \end{bmatrix}$$

$$\text{Right br. } e^{Q*2} = \begin{bmatrix} 0.66 & 0.34 \\ 0.66 & 0.34 \end{bmatrix}$$

$$\Pr(\text{red}) = 0.34 * 0.34 = 0.12$$

# Felsenstein's pruning algorithm

Given values:

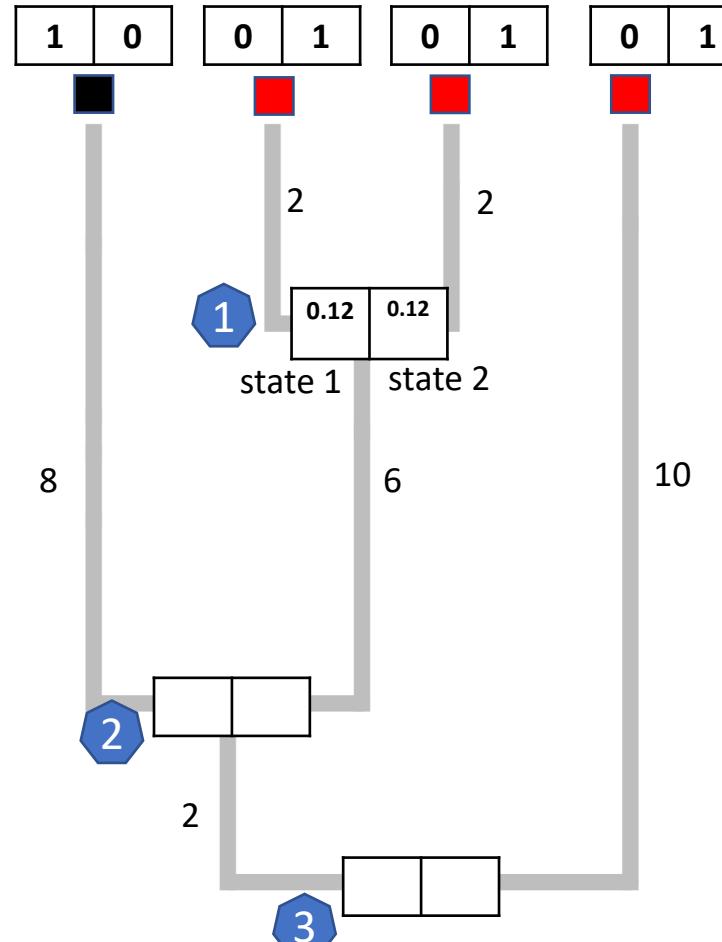
state 1      state 2

1	0
0	1

$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$



$$\pi = (1/2, 1/2)$$



# Felsenstein's pruning algorithm

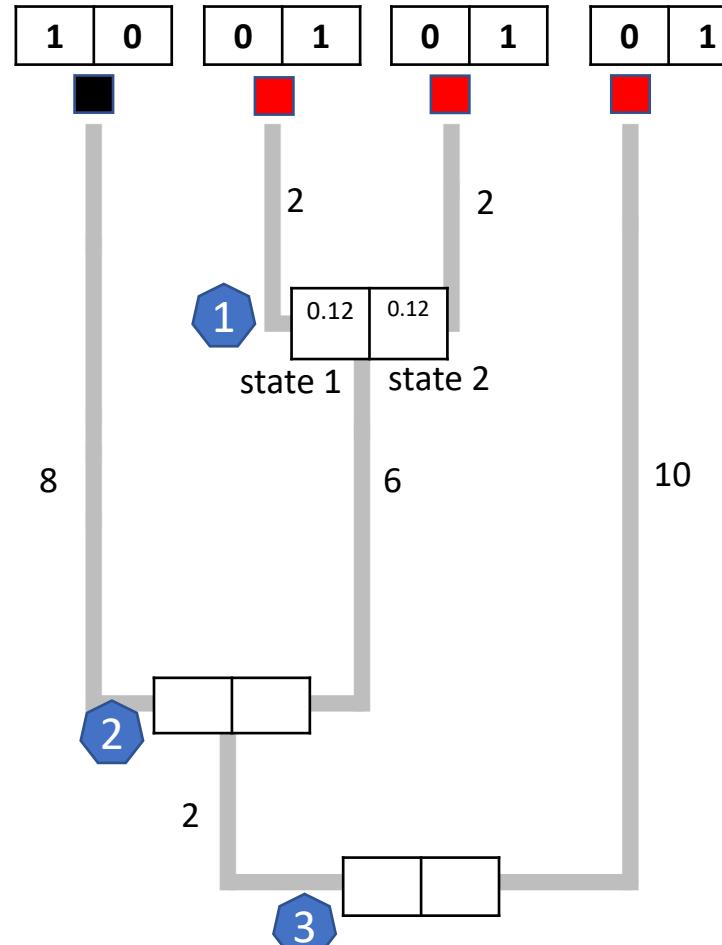
Given values:

state 1      state 2

[1 0]	[0 1]

$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

$$\pi = (1/2, 1/2)$$



Node 2

At state 1

$$\text{Left br. } e^{Q*8} = \begin{bmatrix} 0.66 & 0.33 \\ 0.66 & 0.33 \end{bmatrix}$$

$$\text{Right br. } e^{Q*6} = \begin{bmatrix} 0.66 & 0.33 \\ 0.66 & 0.33 \end{bmatrix}$$

$$\Pr(\text{black}) = 0.66 * (0.66 * 0.12 + 0.33 * 0.12) = 0.08$$

At state 2

$$\text{Left br. } e^{Q*8} = \begin{bmatrix} 0.66 & 0.33 \\ 0.66 & 0.33 \end{bmatrix}$$

$$\text{Right br. } e^{Q*6} = \begin{bmatrix} 0.66 & 0.33 \\ 0.66 & 0.33 \end{bmatrix}$$

$$\Pr(\text{red}) = 0.33 * (0.66 * 0.12 + 0.33 * 0.12) = 0.04$$

# Felsenstein's pruning algorithm

Given values:

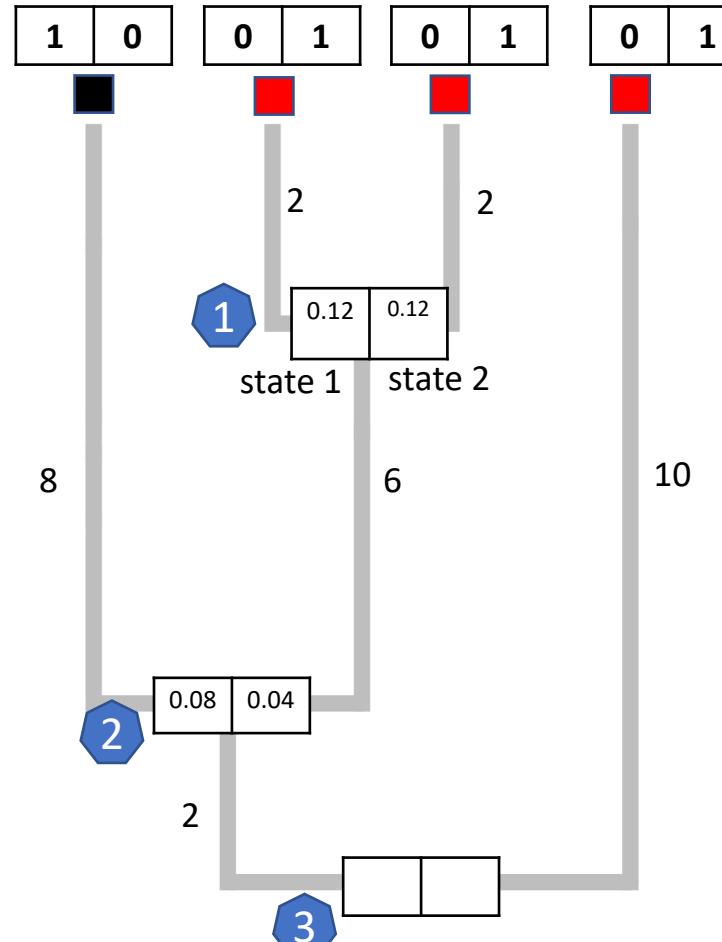
state 1      state 2

1	0
0	1

$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$



$$\pi = (1/2, 1/2)$$



# Felsenstein's pruning algorithm

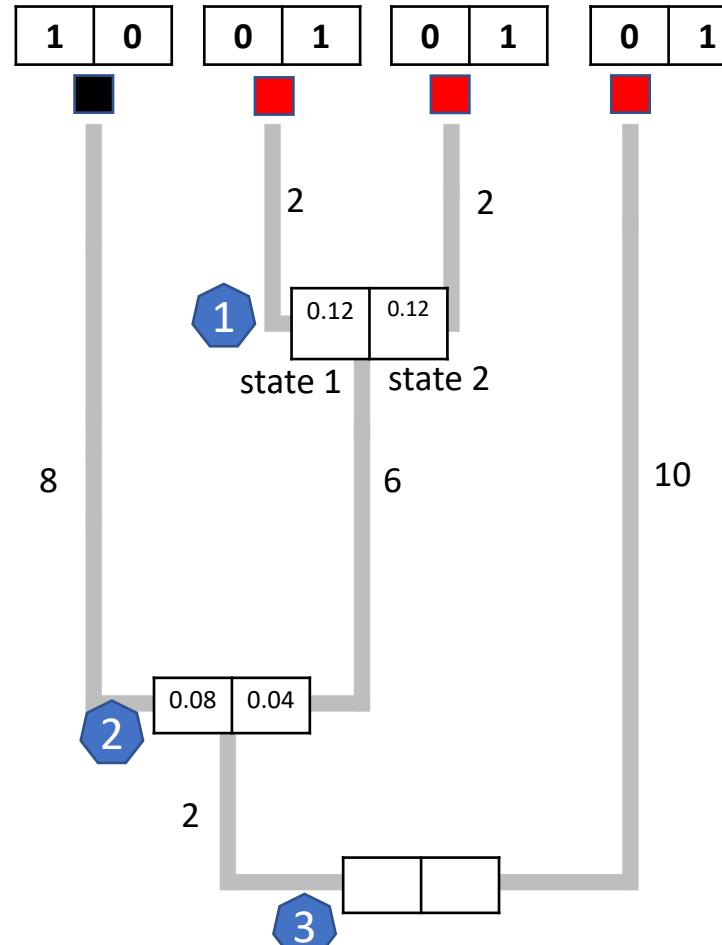
Given values:

state 1  $\rightarrow$  state 2

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

$$\pi = (1/2, 1/2)$$



Node 3

$$\text{Left br. } e^{Q*2} = \begin{bmatrix} 0.66 & 0.34 \\ 0.66 & 0.34 \end{bmatrix}$$

$$\text{Right br. } e^{Q*10} = \begin{bmatrix} 0.66 & 0.33 \\ 0.66 & 0.33 \end{bmatrix}$$

$$\Pr(\text{black}) = (0.08*0.66 + 0.04*0.34)*0.33 = 0.02$$

$$\Pr(\text{red}) = (0.08*0.66 + 0.04*0.34)*0.33 = 0.02$$

# Felsenstein's pruning algorithm

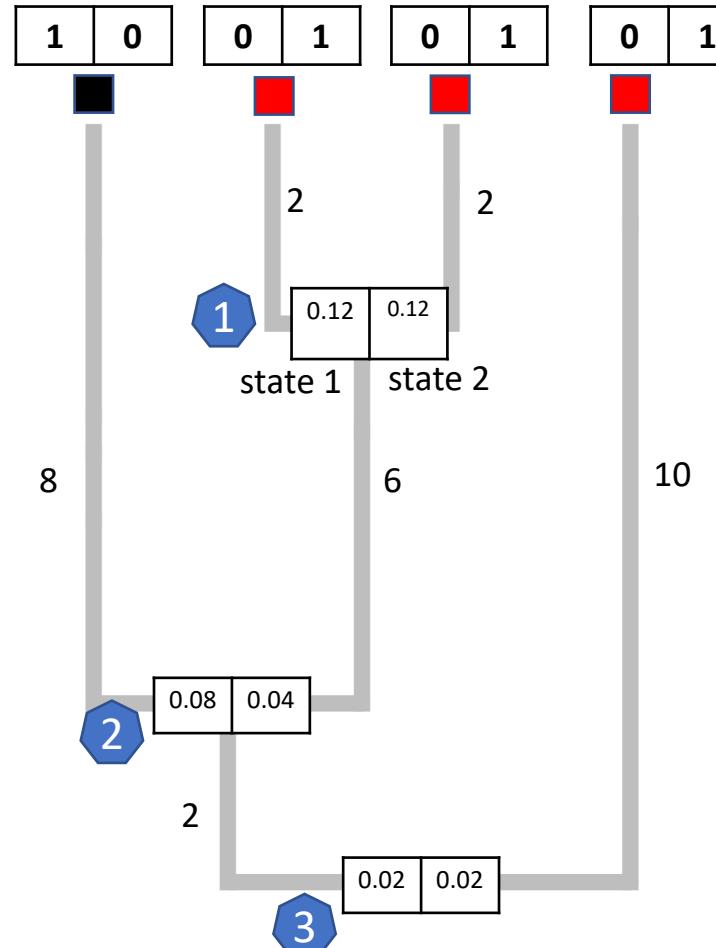
Given values:

state 1      state 2

[1   0]	[0   1]

$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

$$\pi = (1/2, 1/2)$$



Node 3

$$\text{Left br. } e^{Q*2} = \begin{bmatrix} 0.66 & 0.34 \\ 0.66 & 0.34 \end{bmatrix}$$

$$\text{Right br. } e^{Q*10} = \begin{bmatrix} 0.66 & 0.33 \\ 0.66 & 0.33 \end{bmatrix}$$

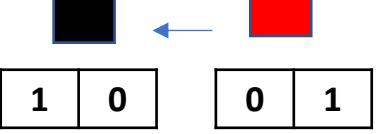
$$\Pr(\text{black}) = (0.08 * 0.66 + 0.04 * 0.34) * 0.33 = 0.02$$

$$\Pr(\text{red}) = (0.08 * 0.66 + 0.04 * 0.34) * 0.33 = 0.02$$

# Felsenstein's pruning algorithm

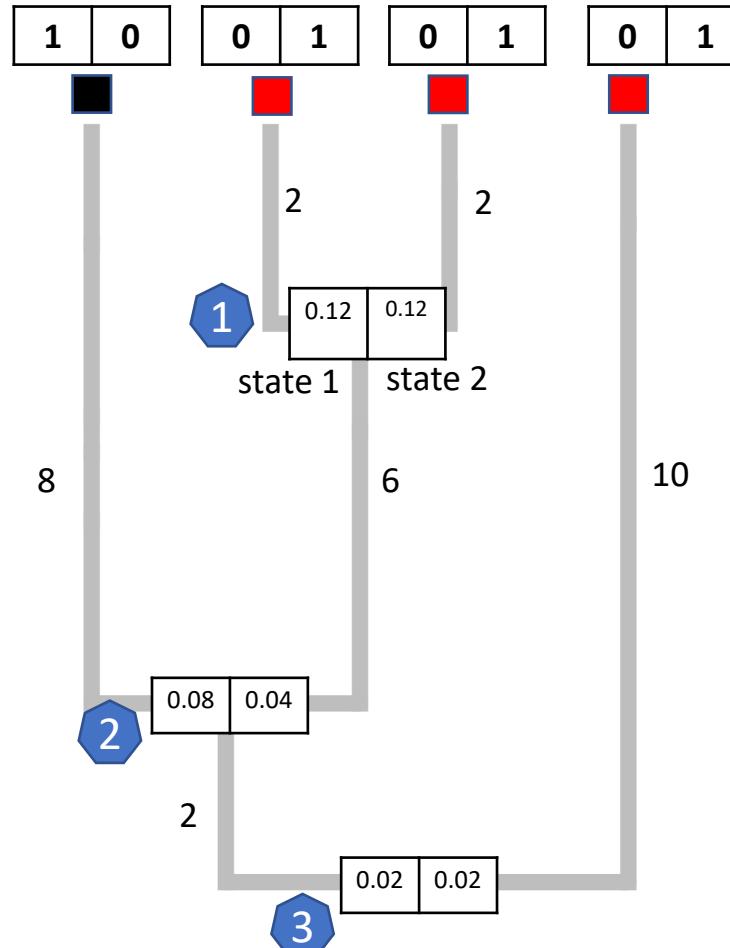
Given values:

state 1      state 2



$$Q = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

$$\pi = (1/2, 1/2)$$



Likelihood (at the root):

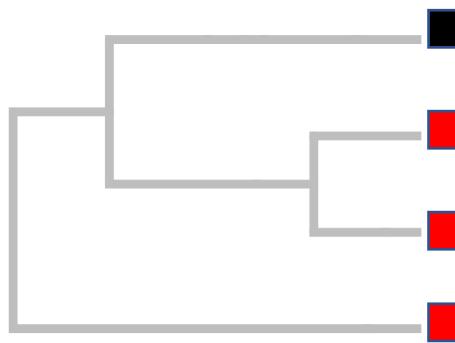
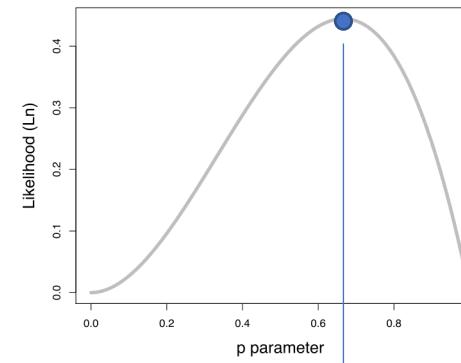
$$L(\text{tree}) = \Pr(\text{black}) * \pi_1 + \Pr(\text{red}) * \pi_2 = 0.02 * 1/2 + 0.02 * 1/2 = 0.02$$

Log Likelihood:

$$\ln(0.02) = -3.91$$

# Maximum Likelihood

Find those values of the following parameters that maximize the likelihood function:



Topology and branch lengths

$$Q = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}$$

Rates of the rate matrix

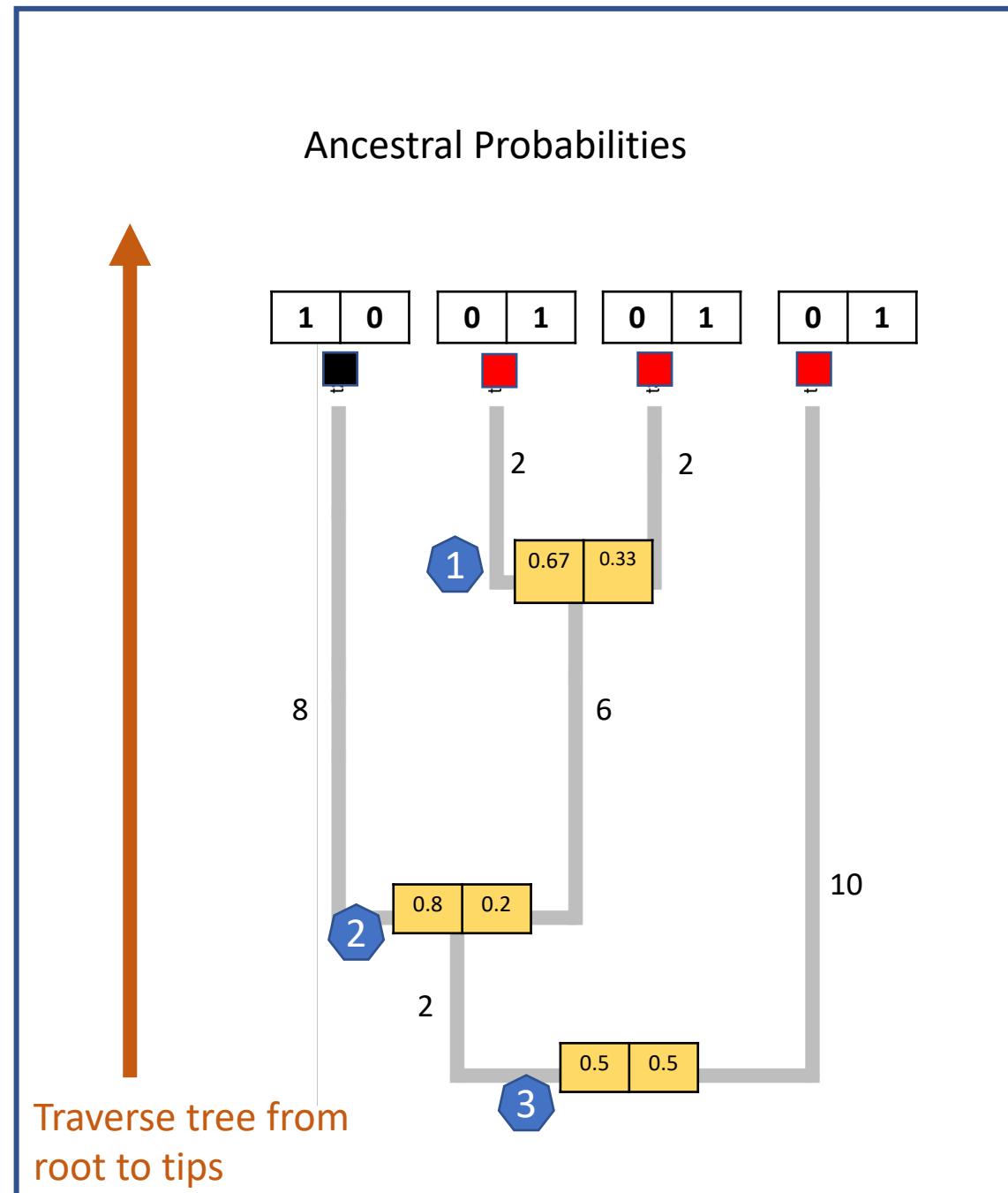


$$\pi = (\pi_1, \pi_2)$$

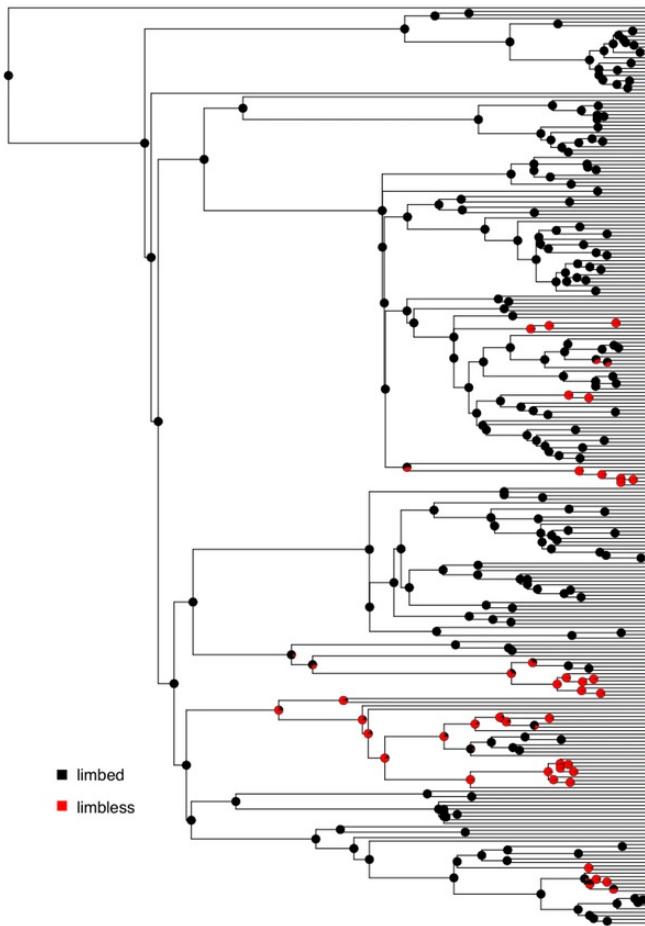
Initial vector at the root of tree

# Ancestral character state reconstruction (marginal reconstruction)

- Felsenstein's pruning algorithm calculates conditional and tree likelihoods
- To calculate state probabilities at each node traverse the tree from the root to the tips and condition on data at tips by successively rerooting the tree



# Test various rate matrices



Suggestions:

- Always run analyses with different rate matrices.
- Select best model using AIC or BIC

$$Q = \begin{bmatrix} -a & a \\ a & -a \end{bmatrix} \quad Q = \begin{bmatrix} -a & a \\ b & -b \end{bmatrix}$$

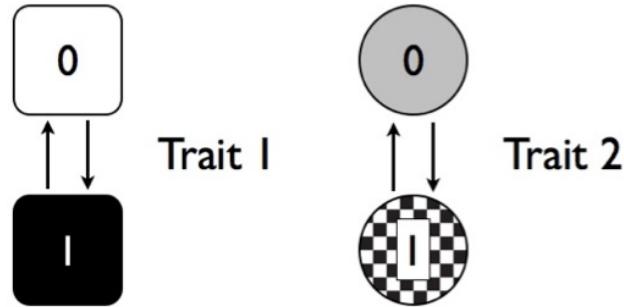
Squamates, legged and legless.

[https://lukejharmon.github.io/pcm/chapter7\\_introdiscrete/](https://lukejharmon.github.io/pcm/chapter7_introdiscrete/)

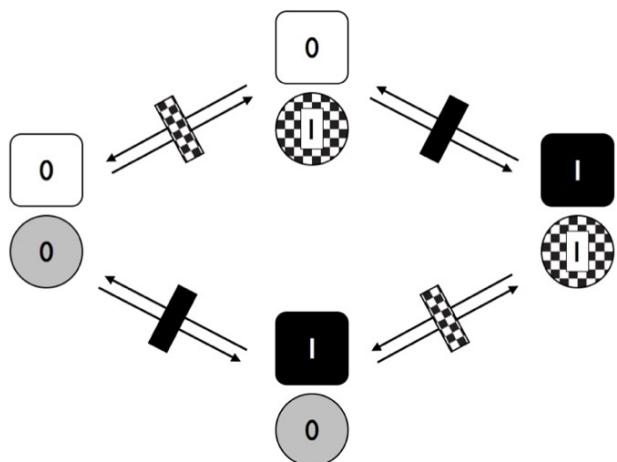
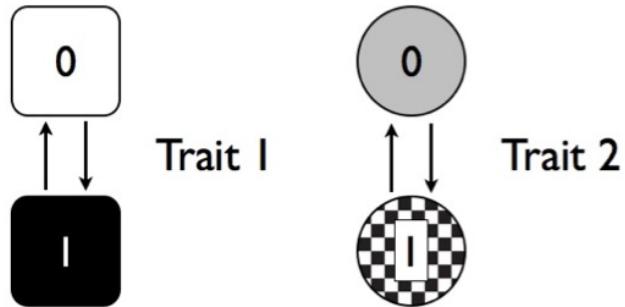


Character correlation

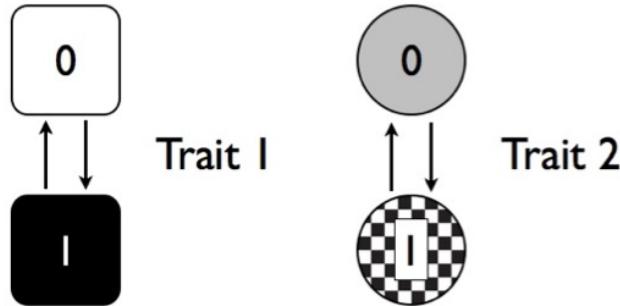
# Character correlation



# Character correlation

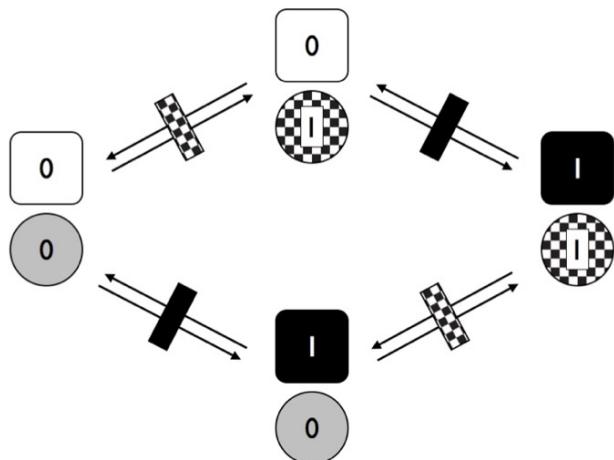


# Character correlation



Uncorrelated Model

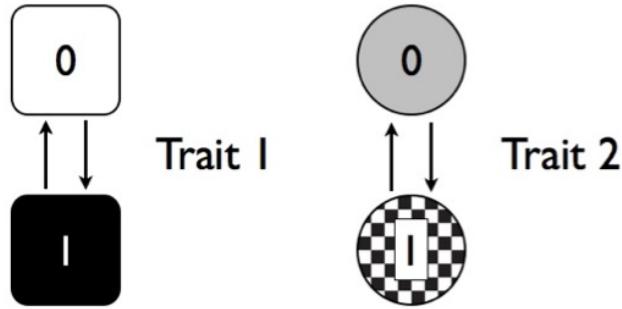
$$Q = \begin{bmatrix} -q_1 - q_2 & q_1 & q_2 & 0 \\ q_1 & -q_1 - q_2 & 0 & q_2 \\ q_2 & 0 & -q_1 - q_2 & q_1 \\ 0 & q_2 & q_1 & -q_1 - q_2 \end{bmatrix}$$



Correlated Model  
(certain rate symmetries violated)

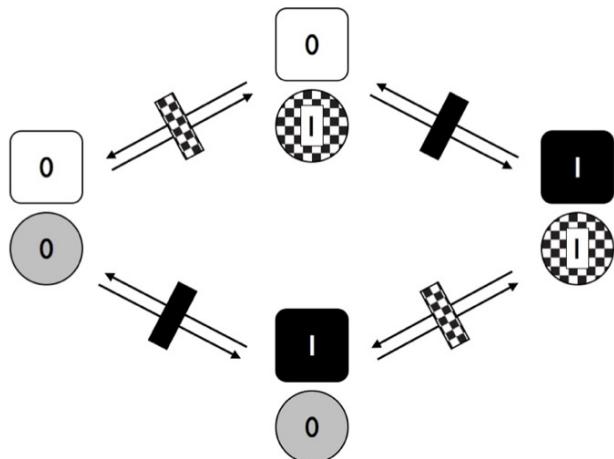
$$Q = \begin{bmatrix} -q_1 - q_2 & q_1 & q_2 & 0 \\ q_1 & -q_1 - q_3 & 0 & q_3 \\ q_2 & 0 & -q_2 - q_4 & q_4 \\ 0 & q_3 & q_4 & -q_3 - q_4 \end{bmatrix}$$

# Character correlation



Suggestions:

- Always run analyses with different rate matrices (correlated vs. uncorrelated).
- Select best model using AIC or BIC



# Unsolved problem: birth of unique traits

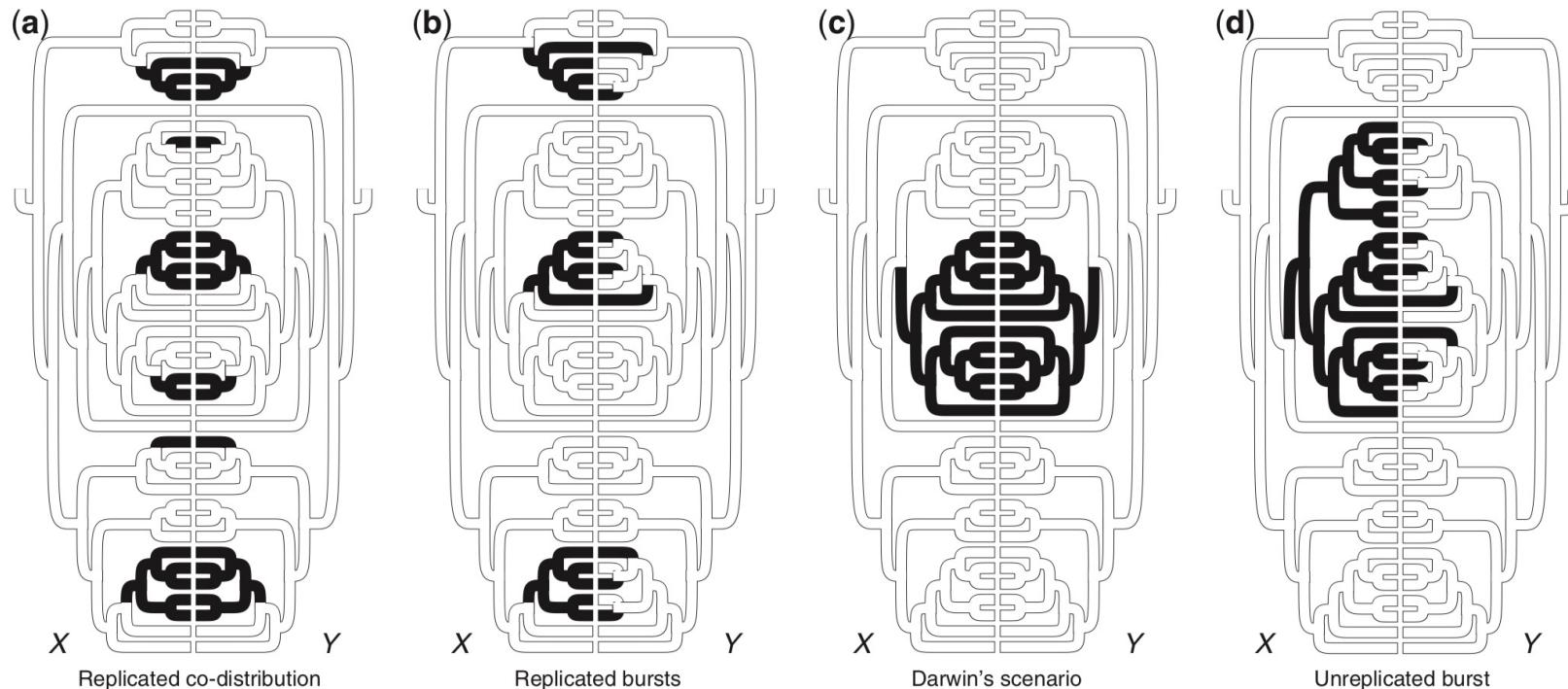
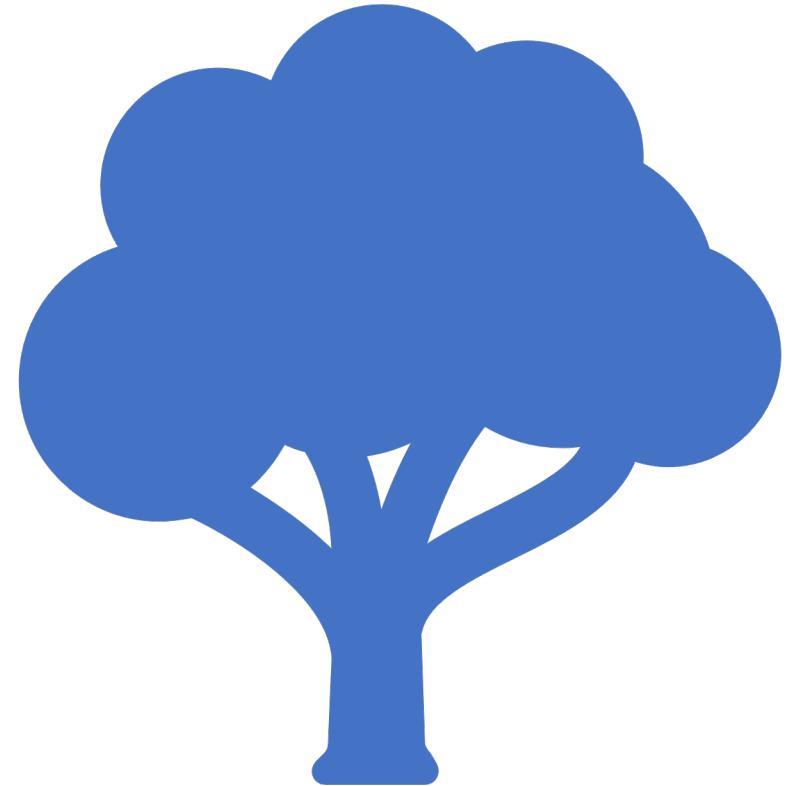
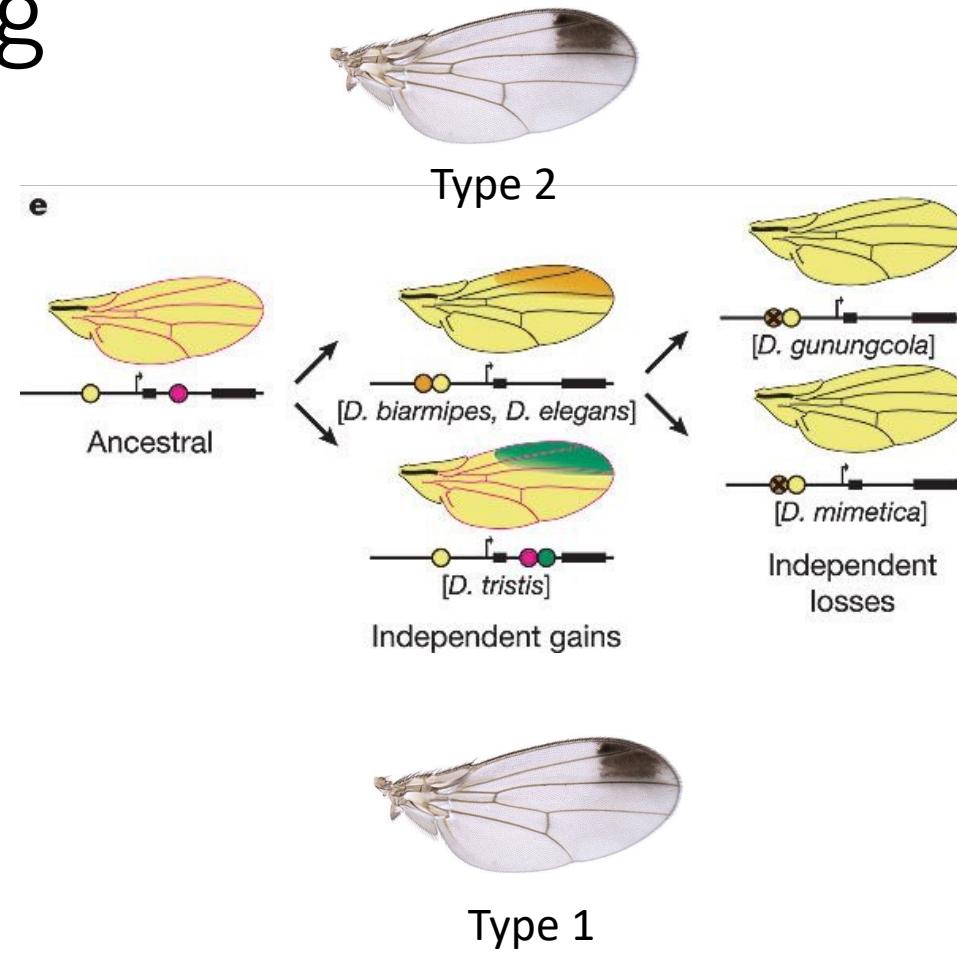
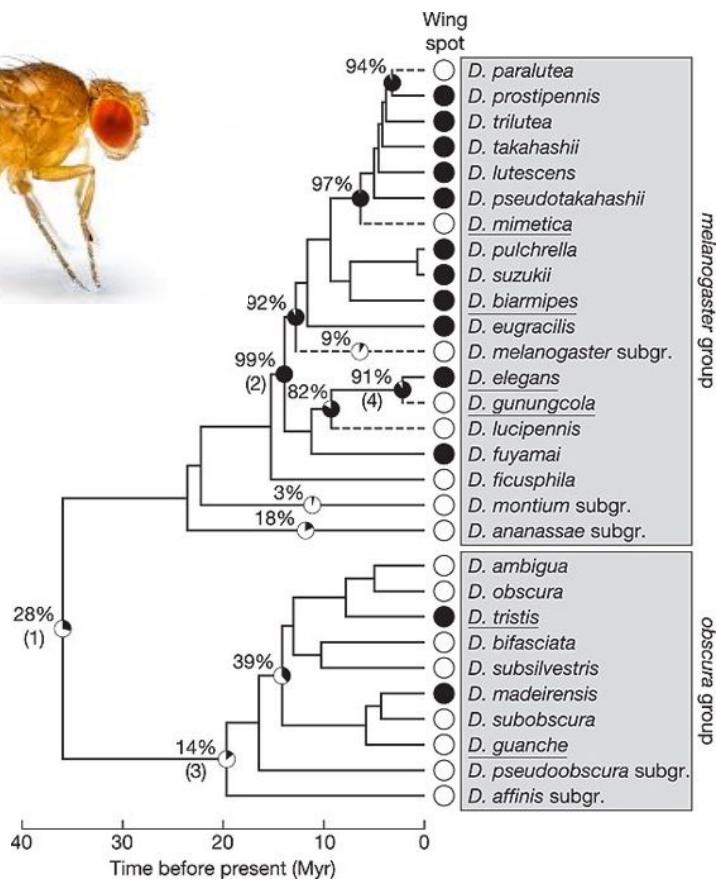


FIGURE 1. Four scenarios for the evolution of states of characters X and Y. In each, the same phylogeny is mirrored to show X at left and Y at right. State 0 = white; State 1 = black. (a) Replicated co-distribution. (b) Replicated bursts. (c) Darwin's scenario. (d) Unreplicated burst. Panels (a) and (b) provide good evidence for an interesting adaptive/functional relationship between X and Y; panels (c) and (d) do not.

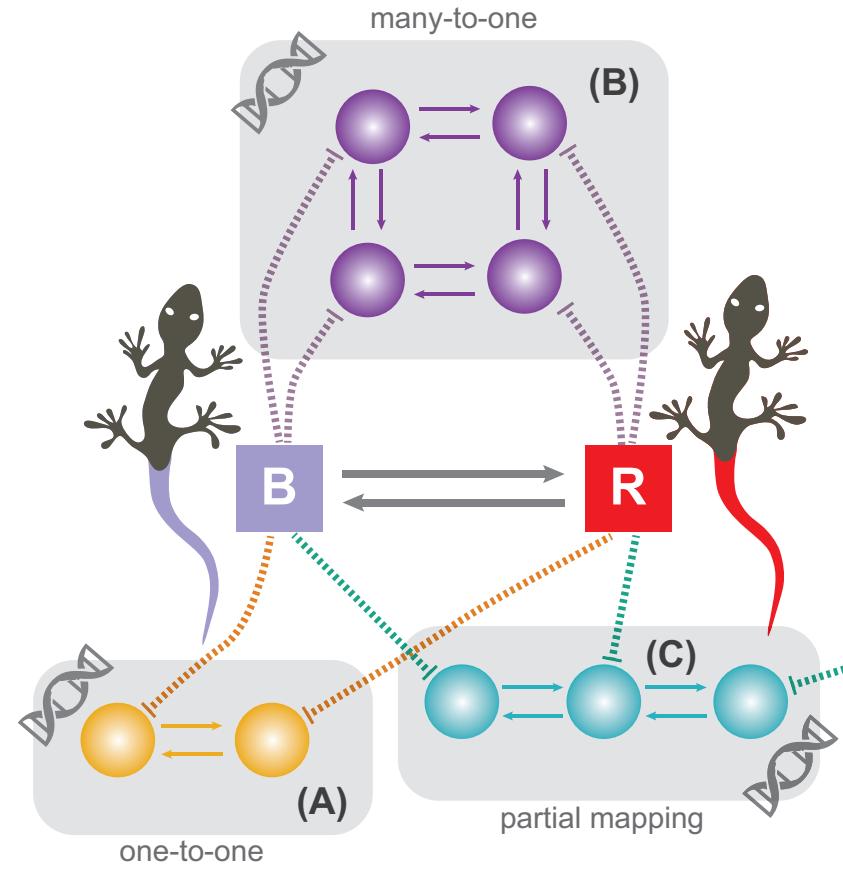


Hidden Markov models

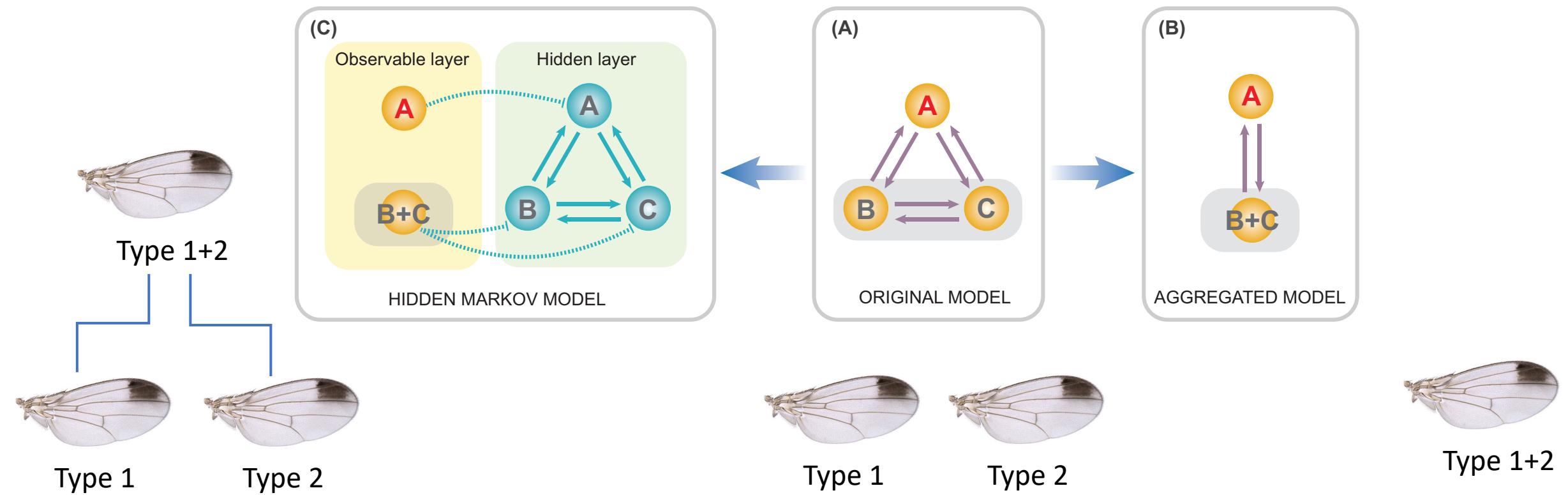
# Spot on Drosophila wing



The real state space upon which evolution acts can be larger than the observable state space

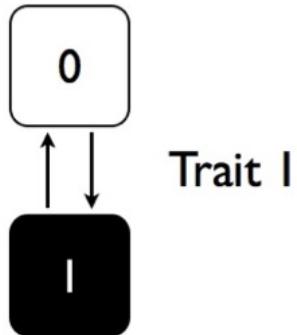


# Hidden Markov models



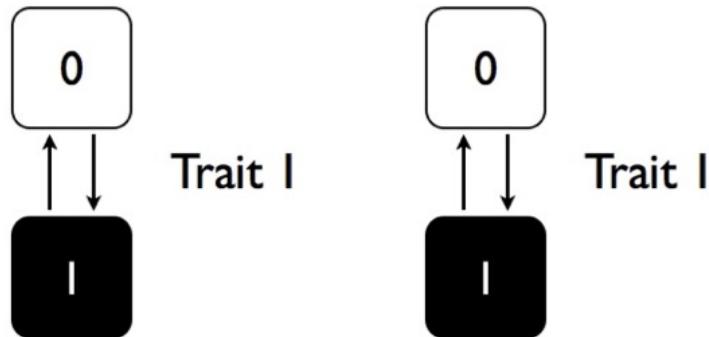
# Hidden Markov models: similar trick as for character correlation

- Amalgamate matrix with itself



# Hidden Markov models: similar trick as for character correlation

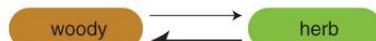
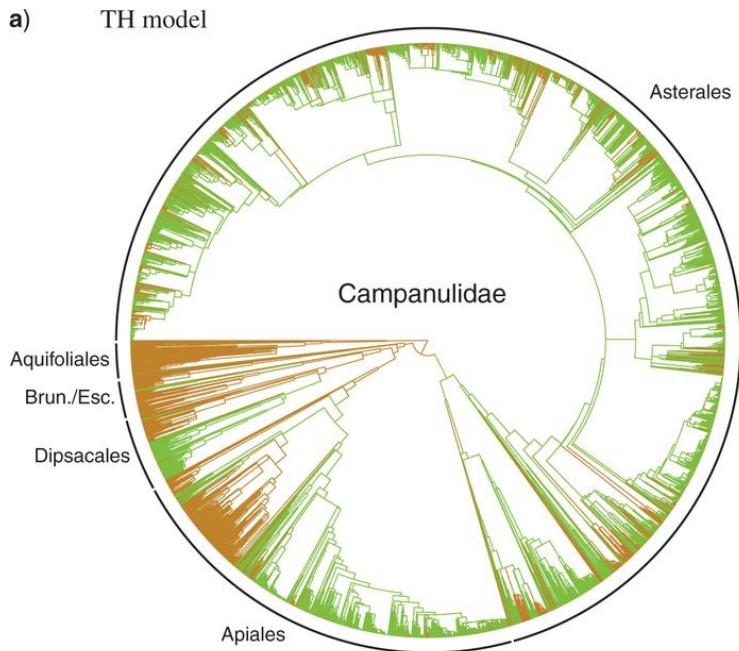
- Amalgamate matrix with itself



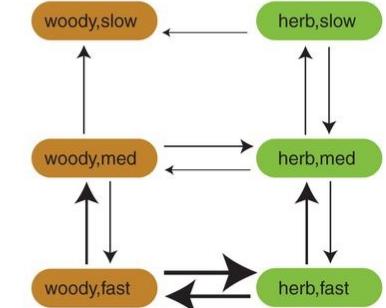
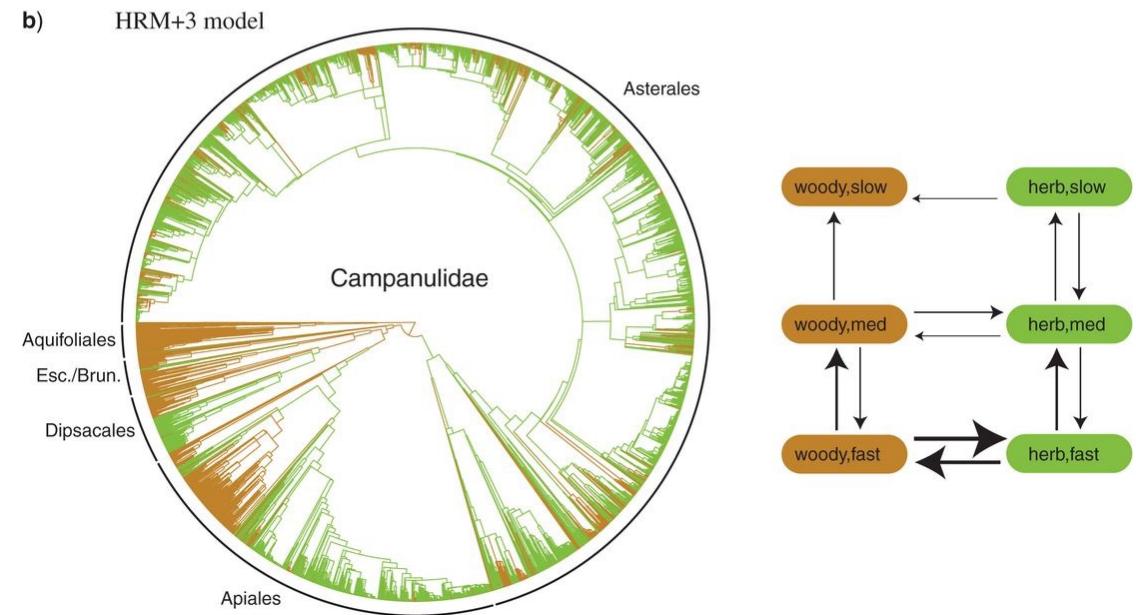
$$Q = \begin{bmatrix} -q_1 - q_2 & q_1 & q_2 & 0 \\ q_1 & -q_1 - q_3 & 0 & q_3 \\ q_2 & 0 & -q_2 - q_4 & q_4 \\ 0 & q_3 & q_4 & -q_3 - q_4 \end{bmatrix}$$

# Evolution of herbaceous vs. wooden traits in plants

a)



b)



# Character Arithmetics in Markov models

- Character Addition

Shape:  
(2 states)



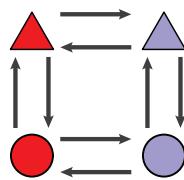
Color:  
(2 states)



+

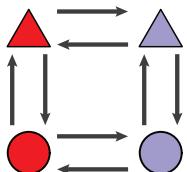
=

Shape + Color:  
(4 states)



- Character Quotient (Direct Quotient)

Shape + Color:  
(4 states)



1)

=

Shape:  
(2 states)

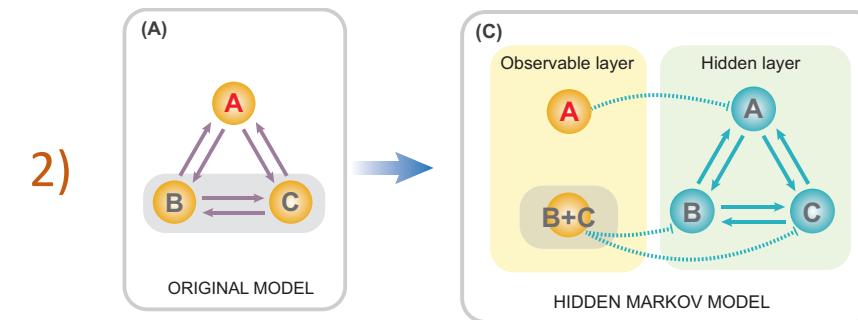


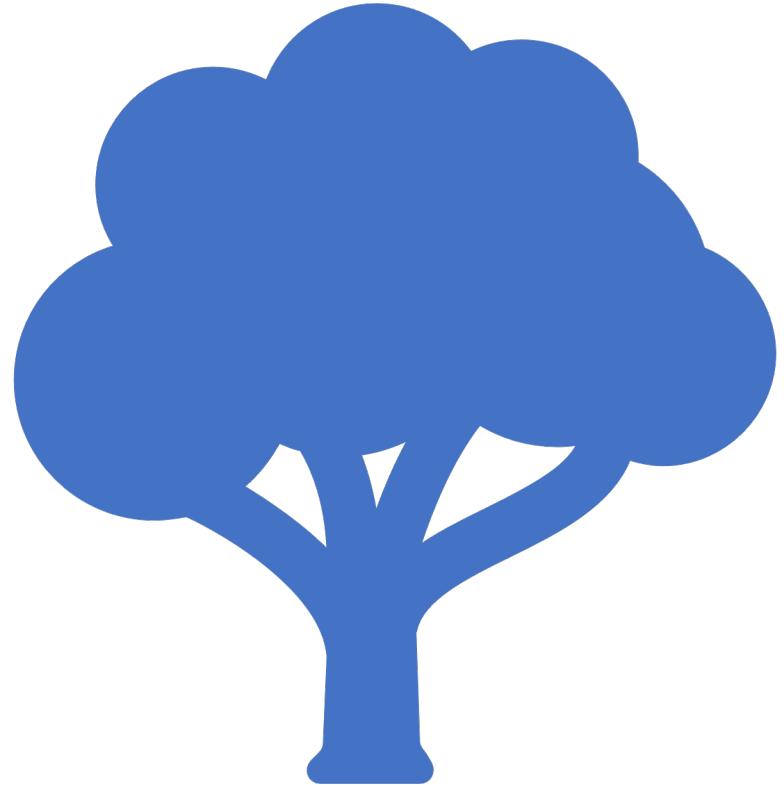
+

Color:  
(2 states)



- Character Quotient (Using HMM)





Going Beyond classical models:  
Inferring Ancestral Anatomies

# Reconstructing ancestral anatomies:

Anatomy is hierarchical

Individual characters

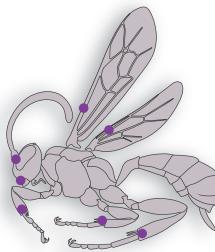


Most studies reconstruct ancestral state for only individual focal traits

# Reconstructing ancestral anatomies:

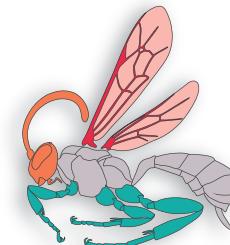
Anatomy is hierarchical

Individual characters



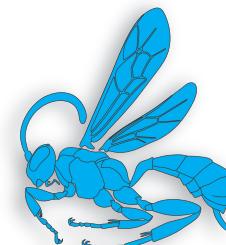
Most studies reconstruct ancestral state for only individual focal traits

Body region characters

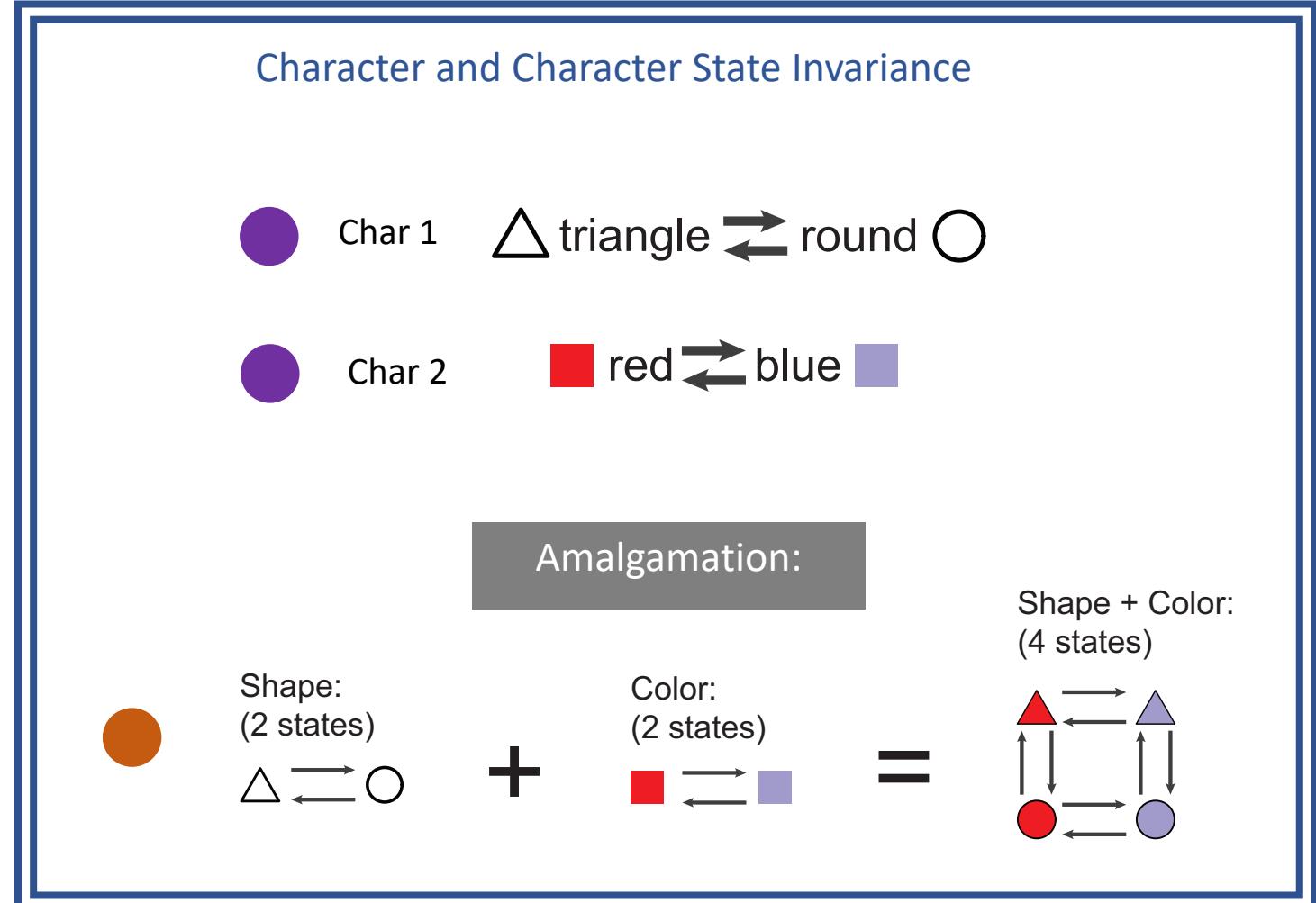
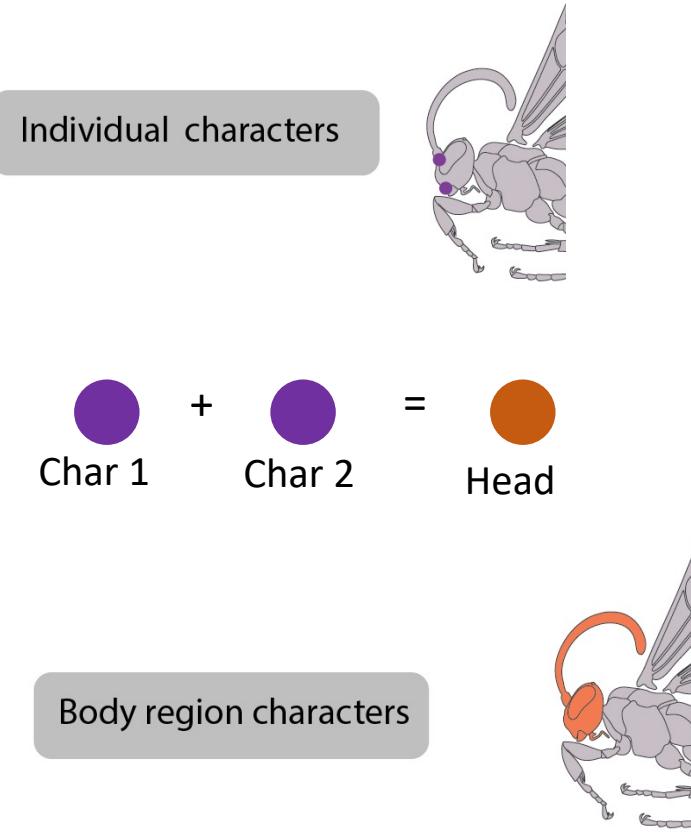


Organismal anatomies (= entire phenotypes) are not merely ensembles of individual traits, they are complex systems

Entire phenotype character



# Character Arithmetics in Markov models: constructing characters for Body Regions



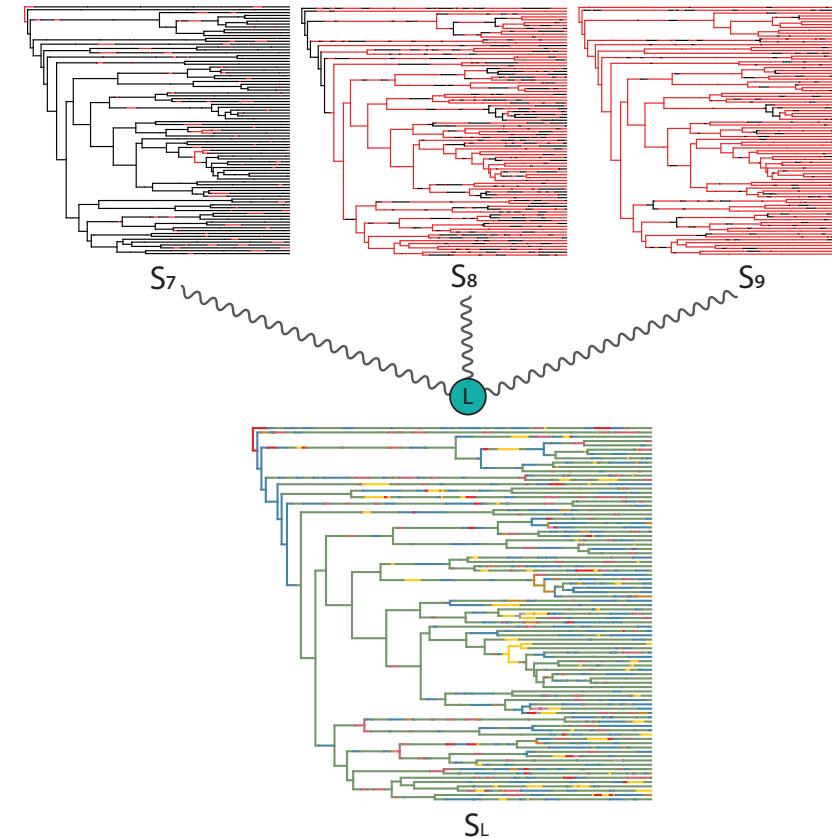
# Character Invariance using Stochastic Mapping

**Solution:** Amalgamate Stochastic Maps

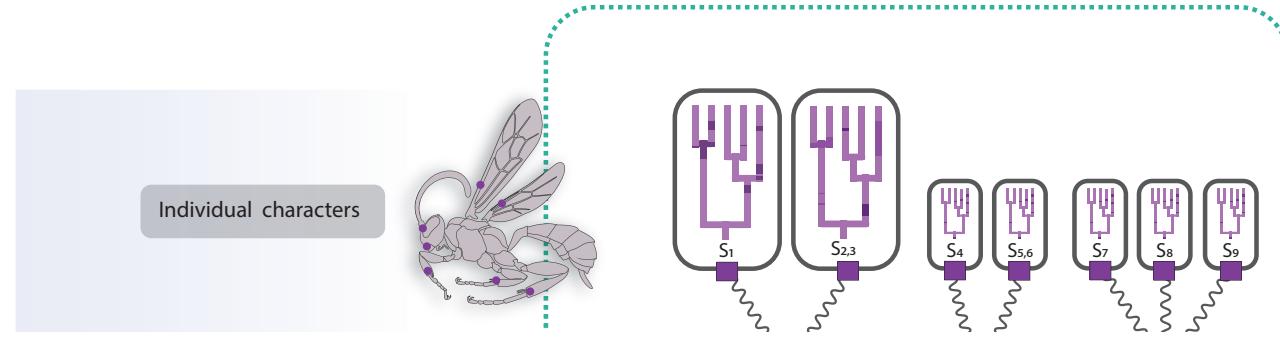
Tree branches

$$\begin{array}{c} S_1 \quad S_2 \\ \oplus \quad = \\ \hline S_{1,2} \end{array}$$

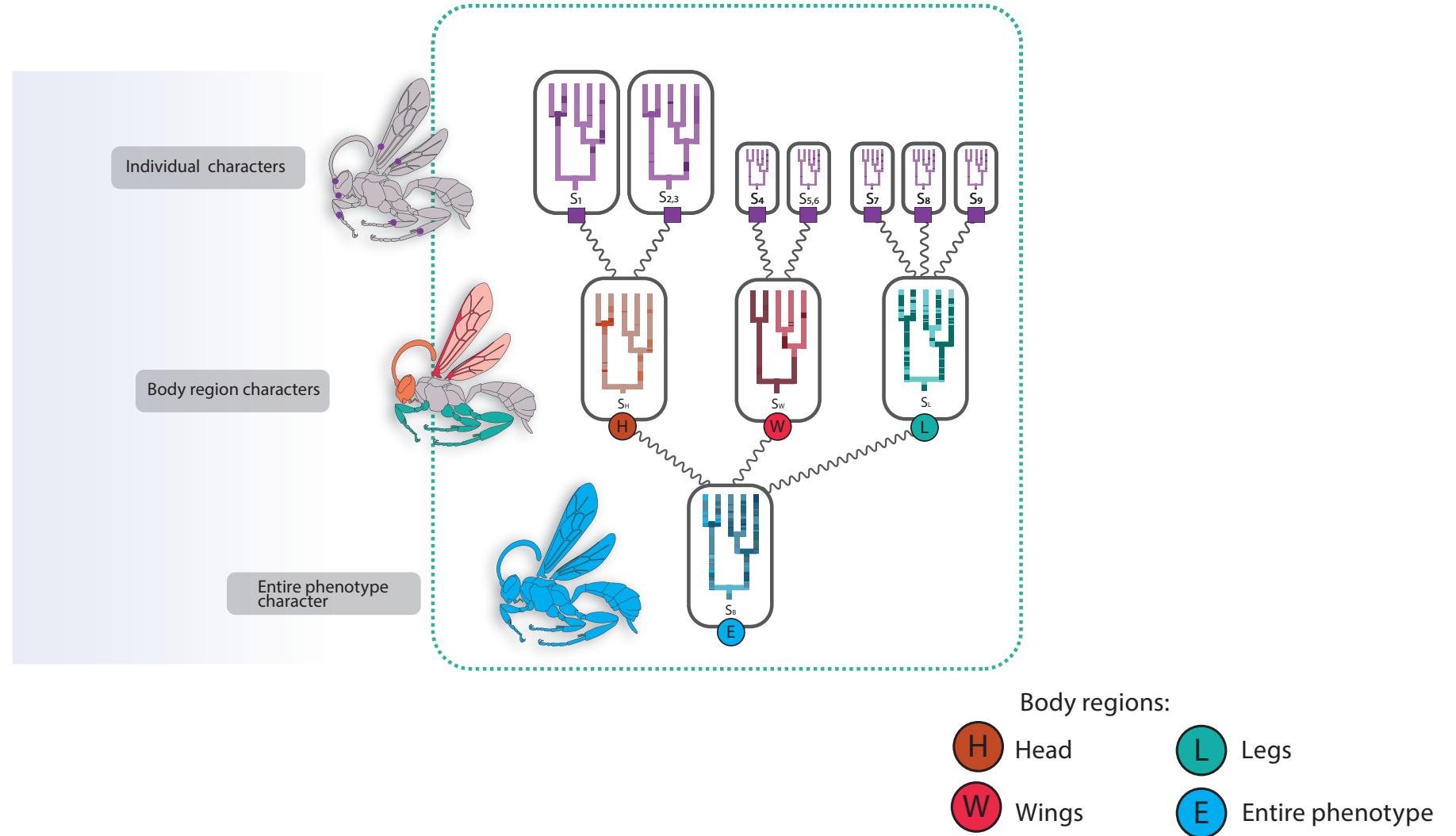
Diagram illustrating the amalgamation of two stochastic maps,  $S_1$  and  $S_2$ , into a single map  $S_{1,2}$ . The vertical axis represents tree branches, with values 0 and 1 indicated at the top and bottom. The horizontal axis represents the amalgamation operation. The resulting map  $S_{1,2}$  is shown as a stack of two columns: the left column has values 0 and 1, and the right column has values 10, 11, 01, and 00. Dashed horizontal lines indicate the boundaries between the stacked columns.



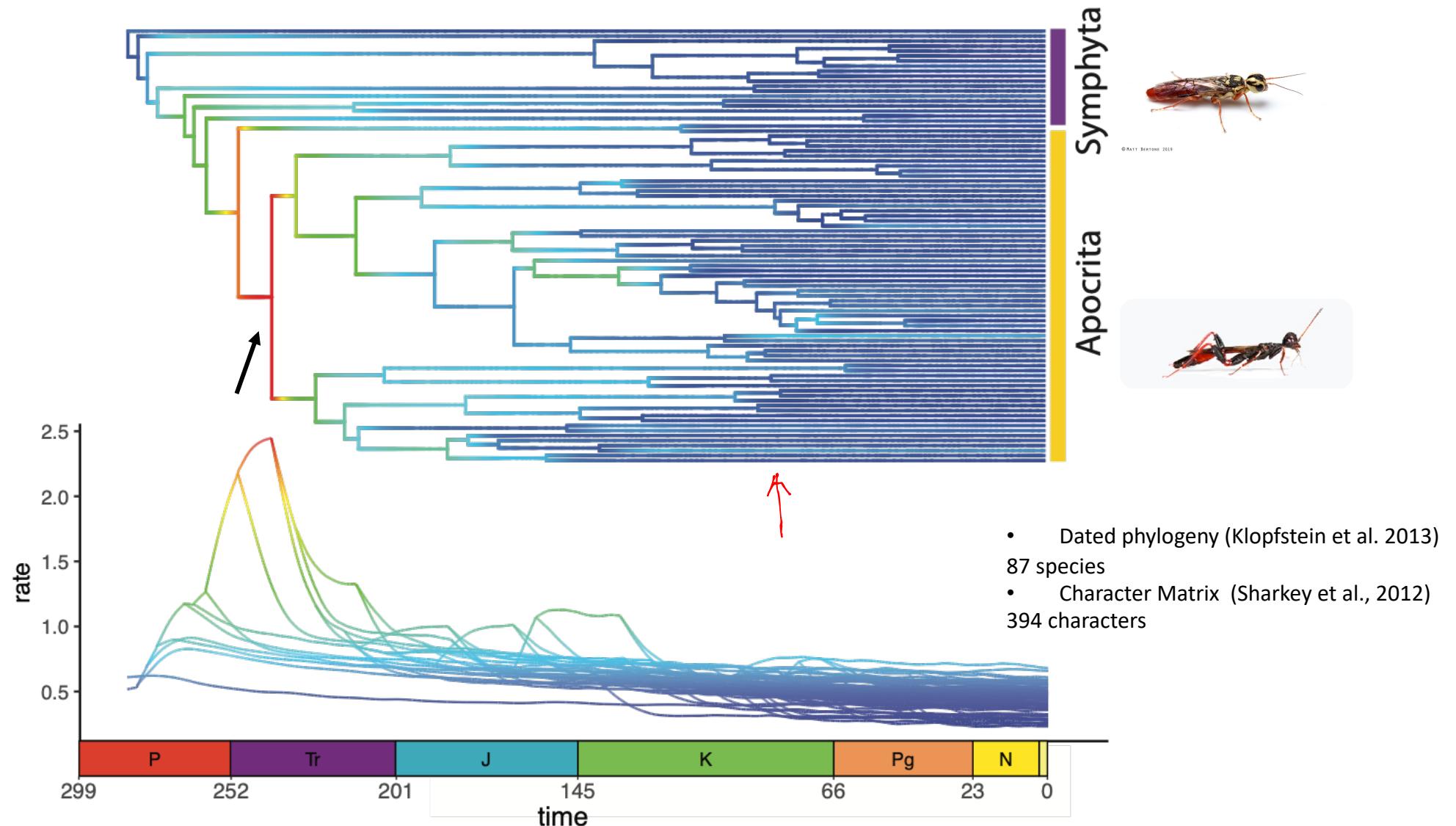
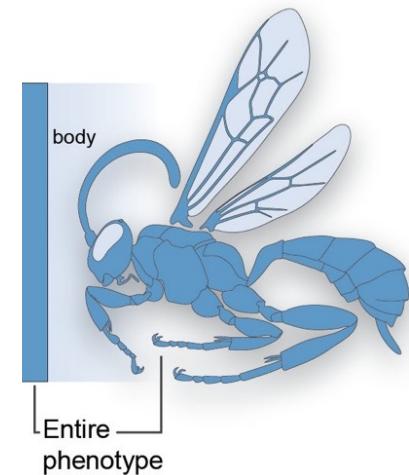
# Ontology-informed amalgamation



# Ontology-informed amalgamation

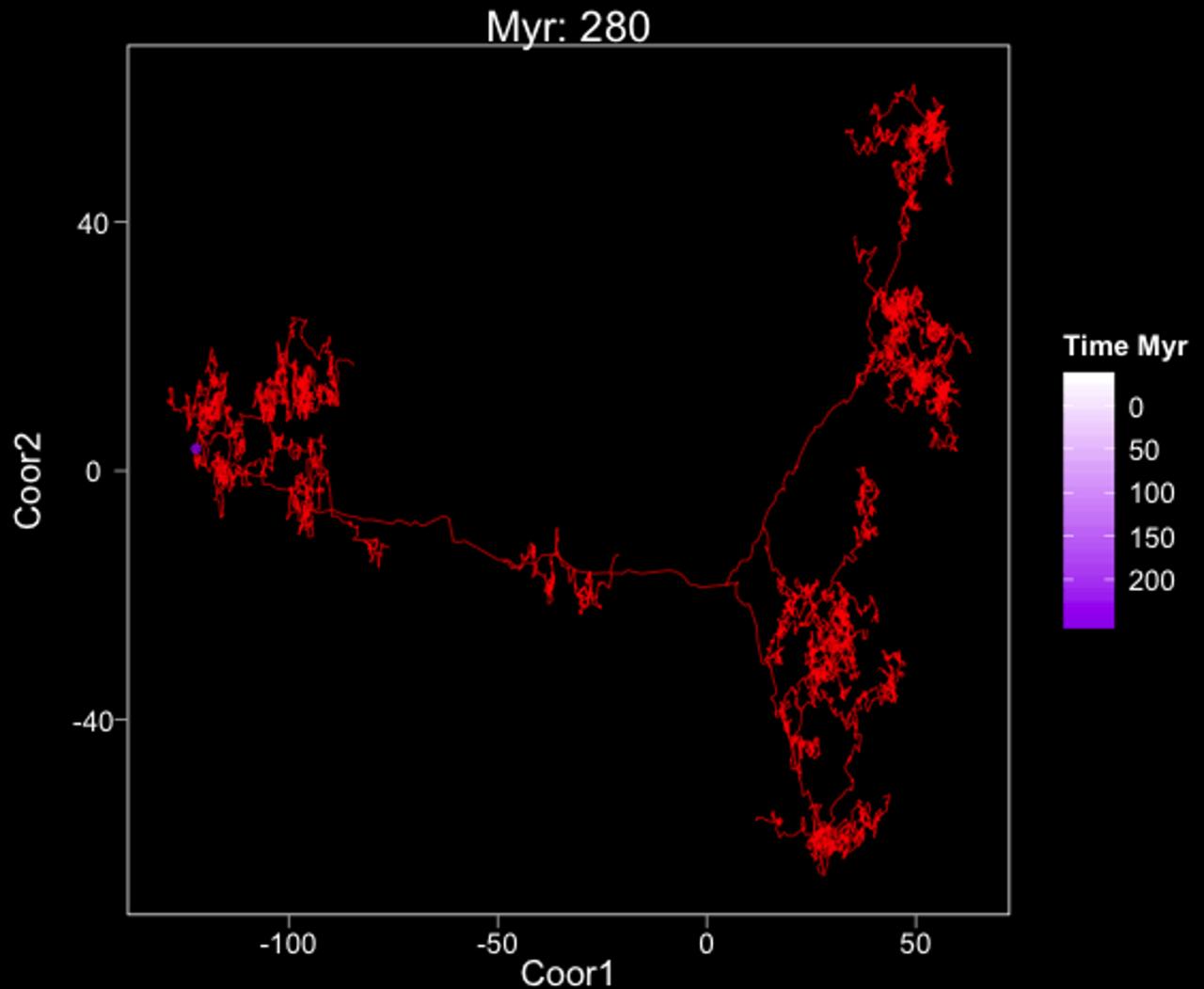
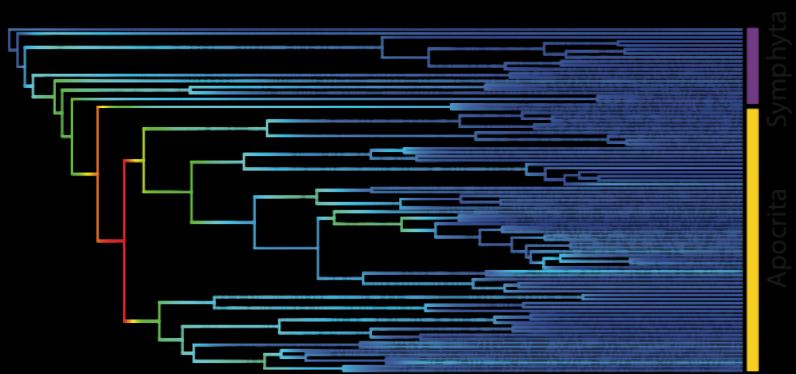


# PARAMO method: Evolutionary Rate of entire Anatomy

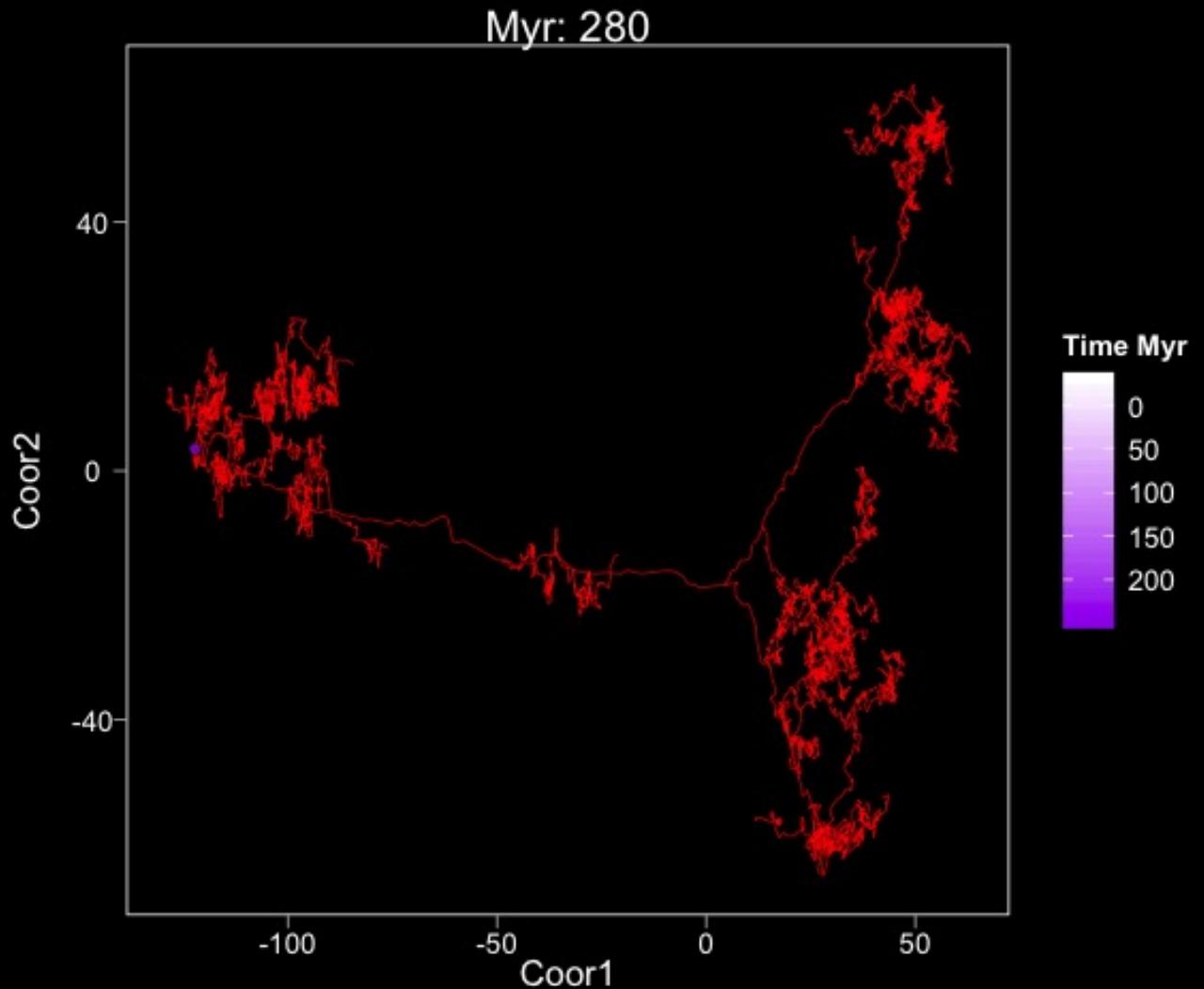
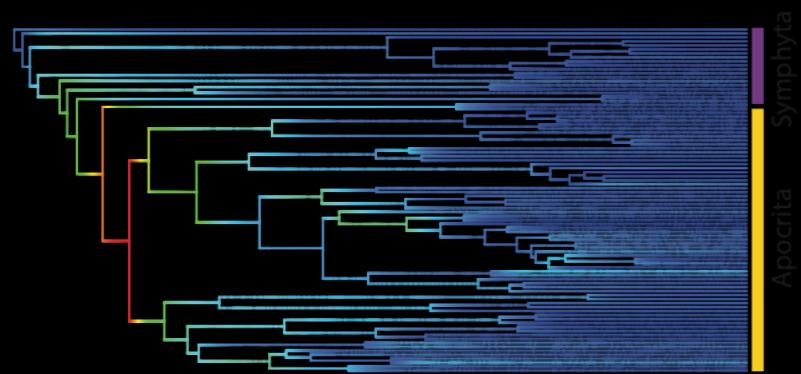


(Tarasov et al., 2019)

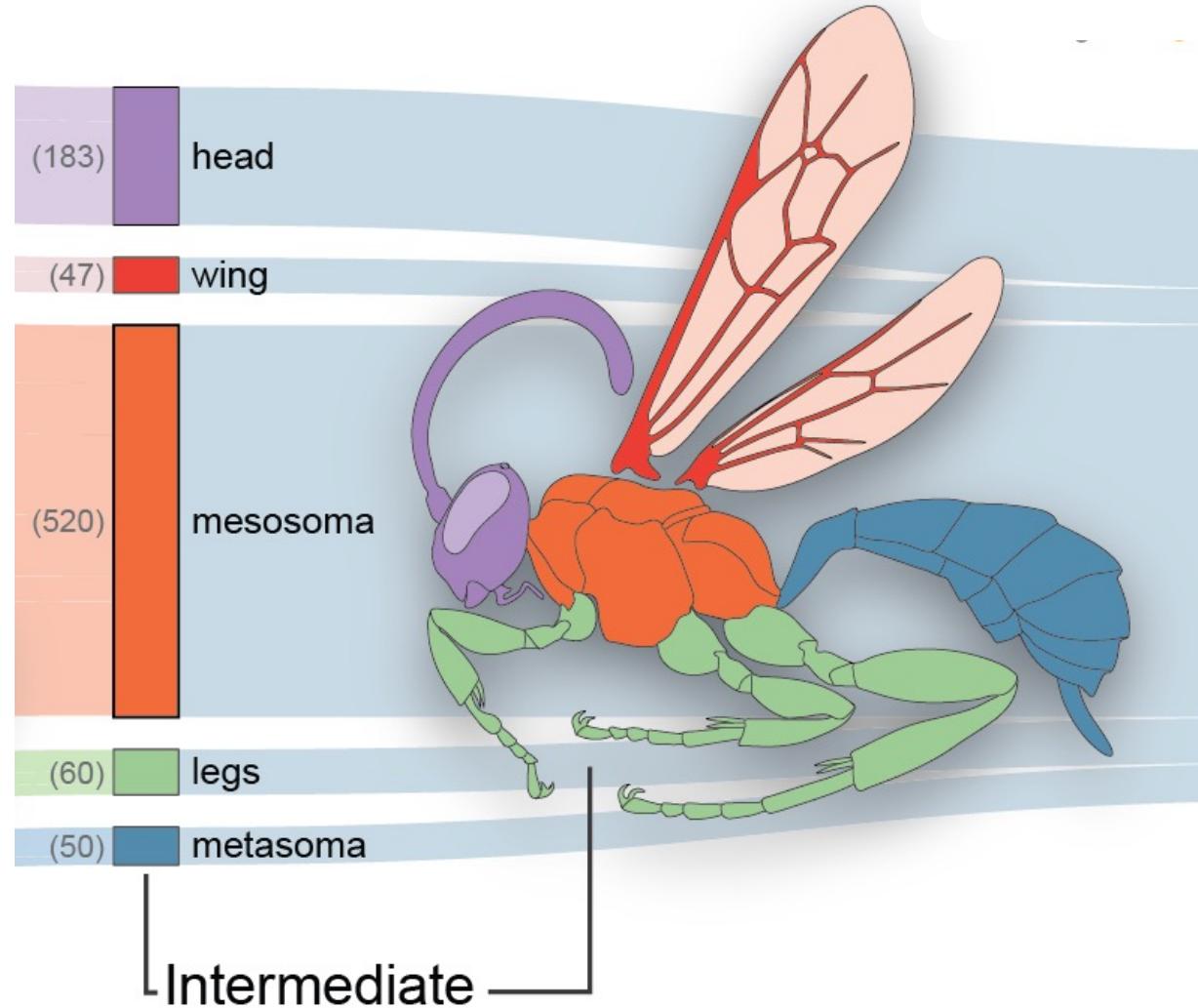
# Explicit Dynamics of Anatomy through time



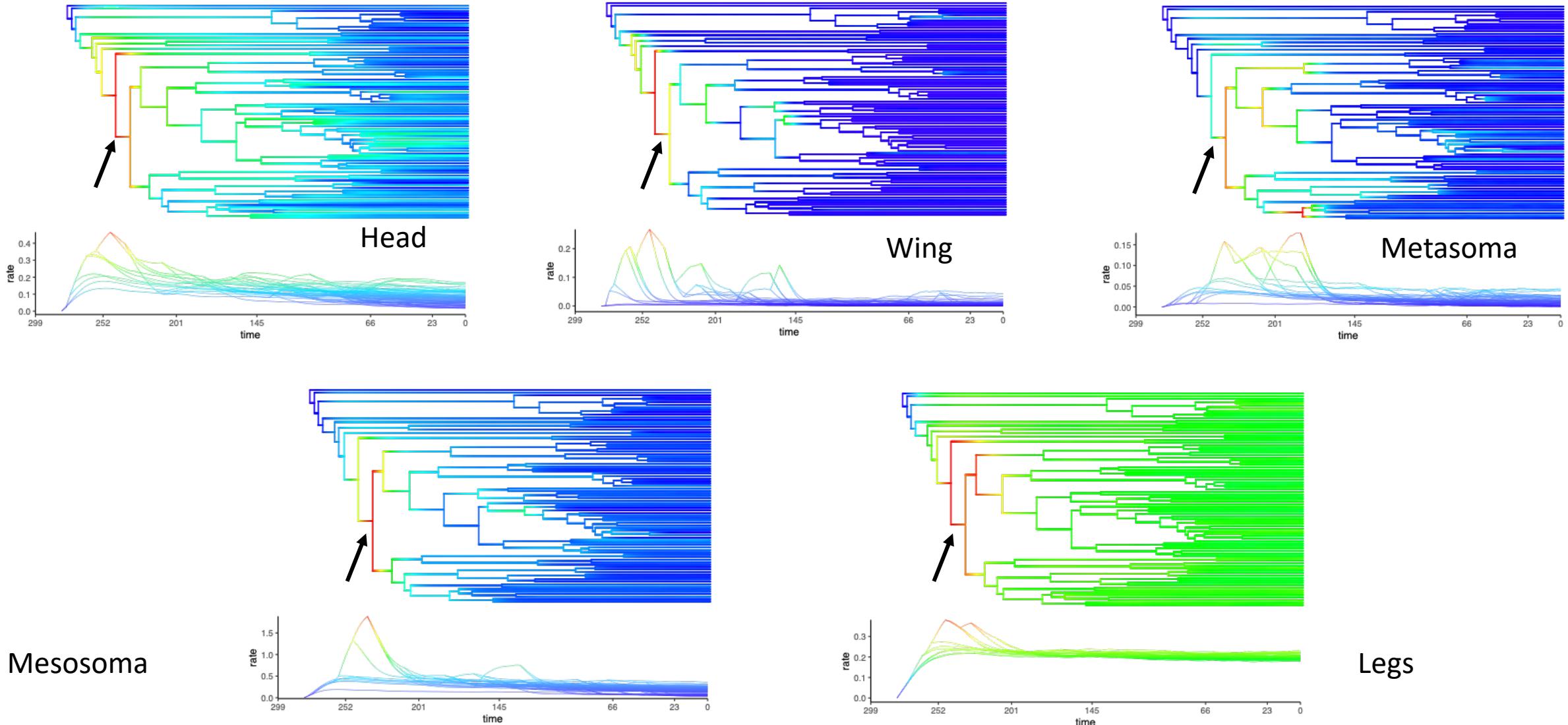
# Explicit Dynamics of Anatomy through time



# Main Body parts

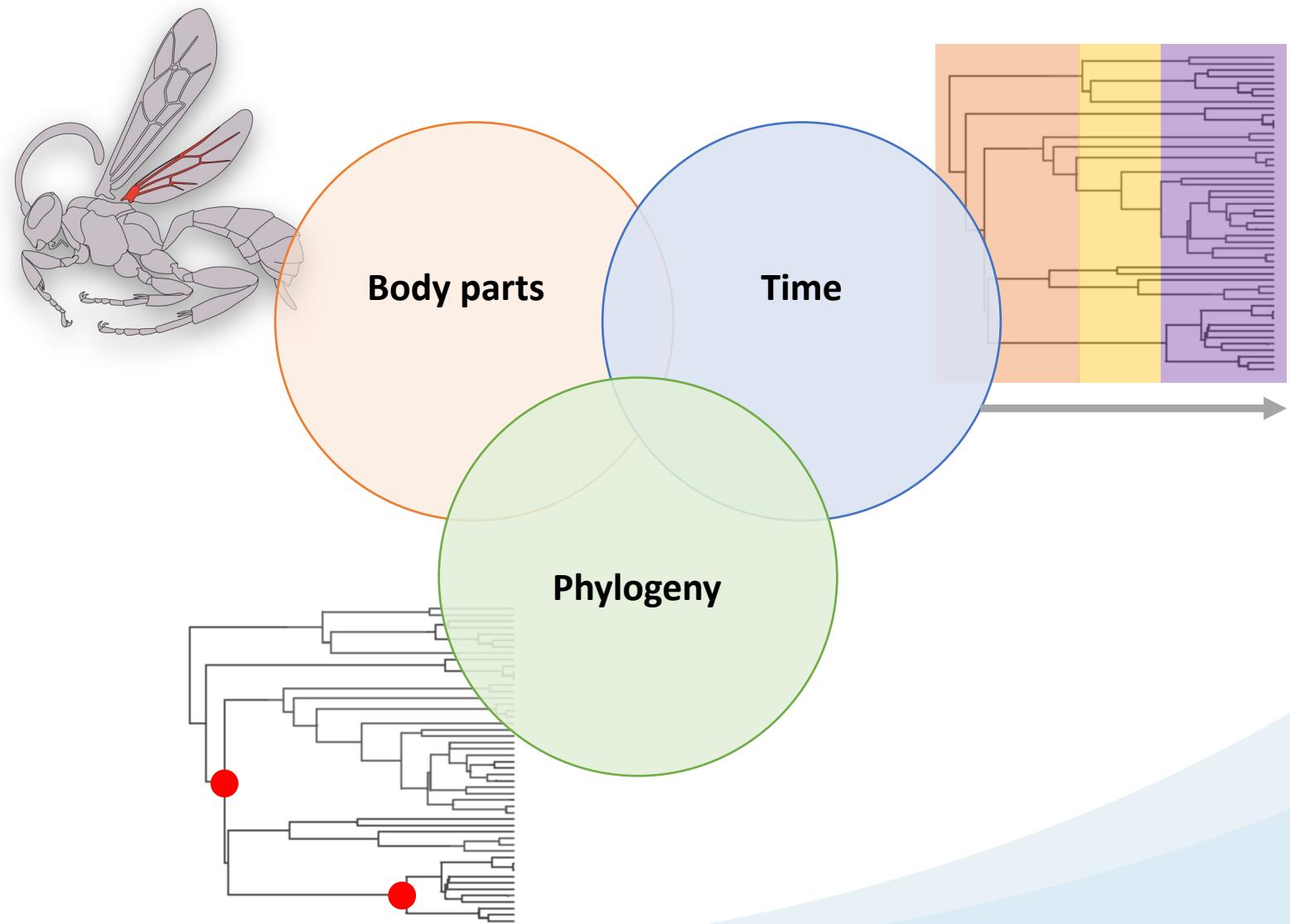


# Evolutionary Rate: main body parts



Detecting rate shifts across phenotypes will open new dimensions to analyze evolutionary radiations

## PARAMO method



# Software

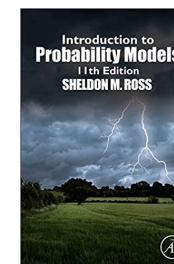
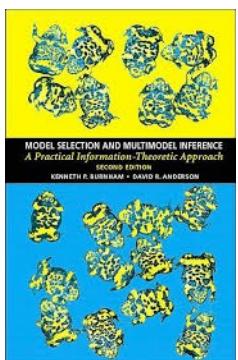
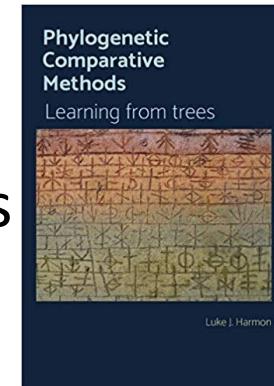
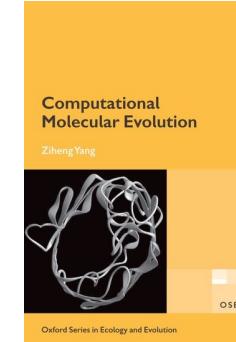
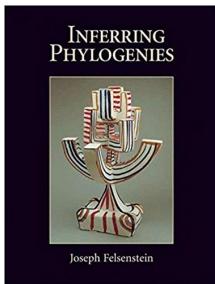
- RStudio <https://www.rstudio.com/products/rstudio/download/>
- R <https://www.r-project.org>
- R package CorHMM
- R package Ape
- R packahe Phytools

# Summary

- Modeling discrete traits helps answering various biological questions about evolutionary rates and character correlations.
- Many ways to construct different models for testing biological hypotheses. So, you need to compare multiple models.

# Suggested literature

- Computational Molecular evolution
- Felsenstein's Inferring phylogenies
- Luke Harmon. Phylogenetic Comparative Methods  
Free at <https://lukejharmon.github.io/pcm/>



- Intro to Probability models
- Model Selection by Burham and Anderson

# Today's Exercises

- Follow this GitHub link

[https://github.com/sergeitarasov/ForBio-Course\\_2021](https://github.com/sergeitarasov/ForBio-Course_2021)