

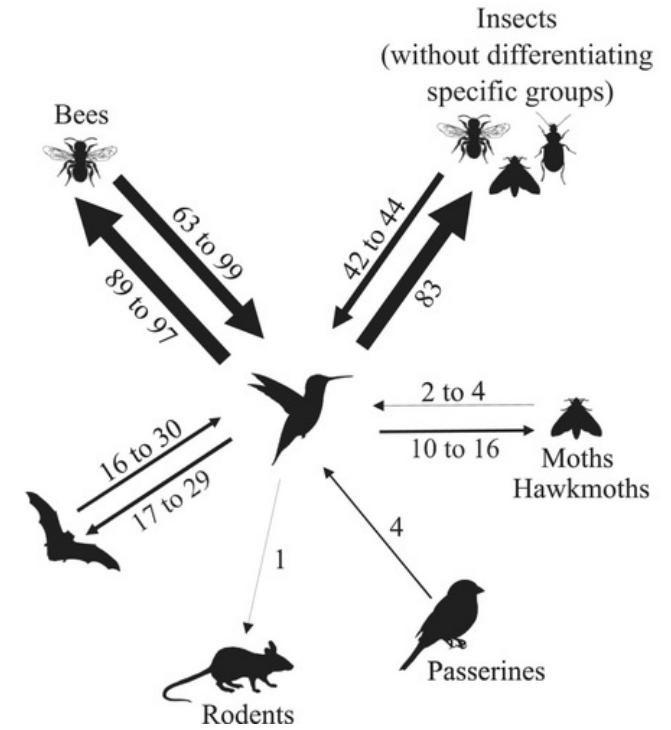
Discrete trait models

Introduction to phylogenetic comparative methods

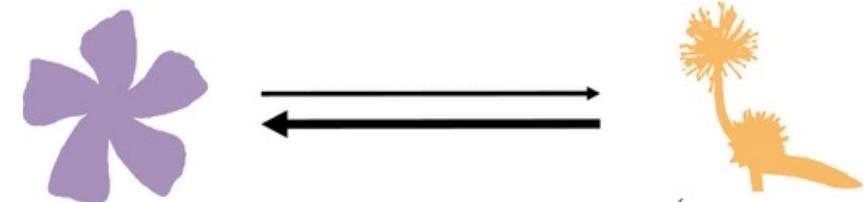
Rosana Zenil-Ferguson
Assistant Professor. University of Kentucky
MOLE 2025 Workshop. Woods Hole, MA.

Pollination

- Plants depend on pollination to start the breeding process
- Pollen can be transferred from plant to plant by insects, wind, vertebrates, or even water
- There are multiple types of pollination, and frequently within a flowering plant/genus or family we can find different strategies for pollination.
- Often flower shape tells us something about the type of pollination we can expect.



Barreto et al. 2024



How do we estimate the number of transitions and the tempo?

Cumberland Falls State Park, KY



Thalictrum
multiple changes between
insect and wind pollination

Verónica di Stilio
University of Washington



Thalictrum pubescens

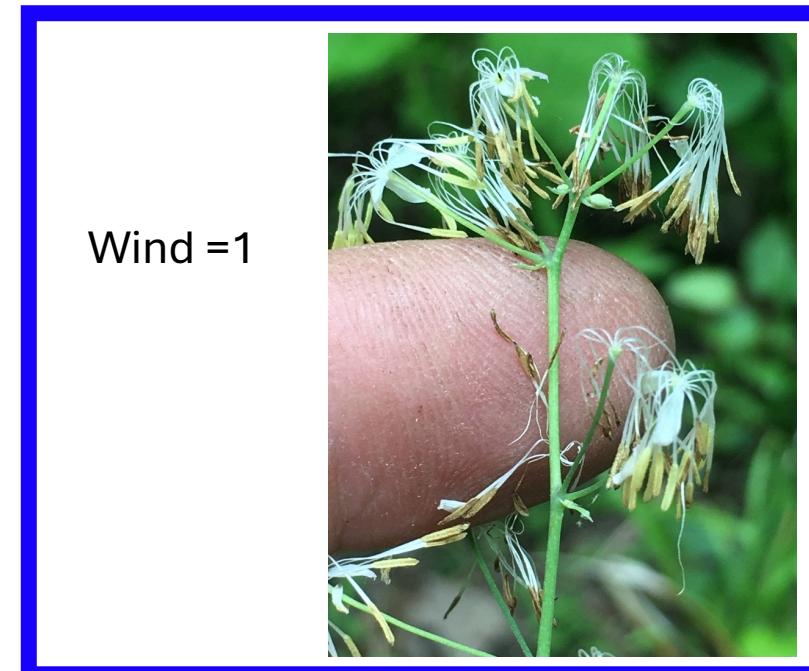


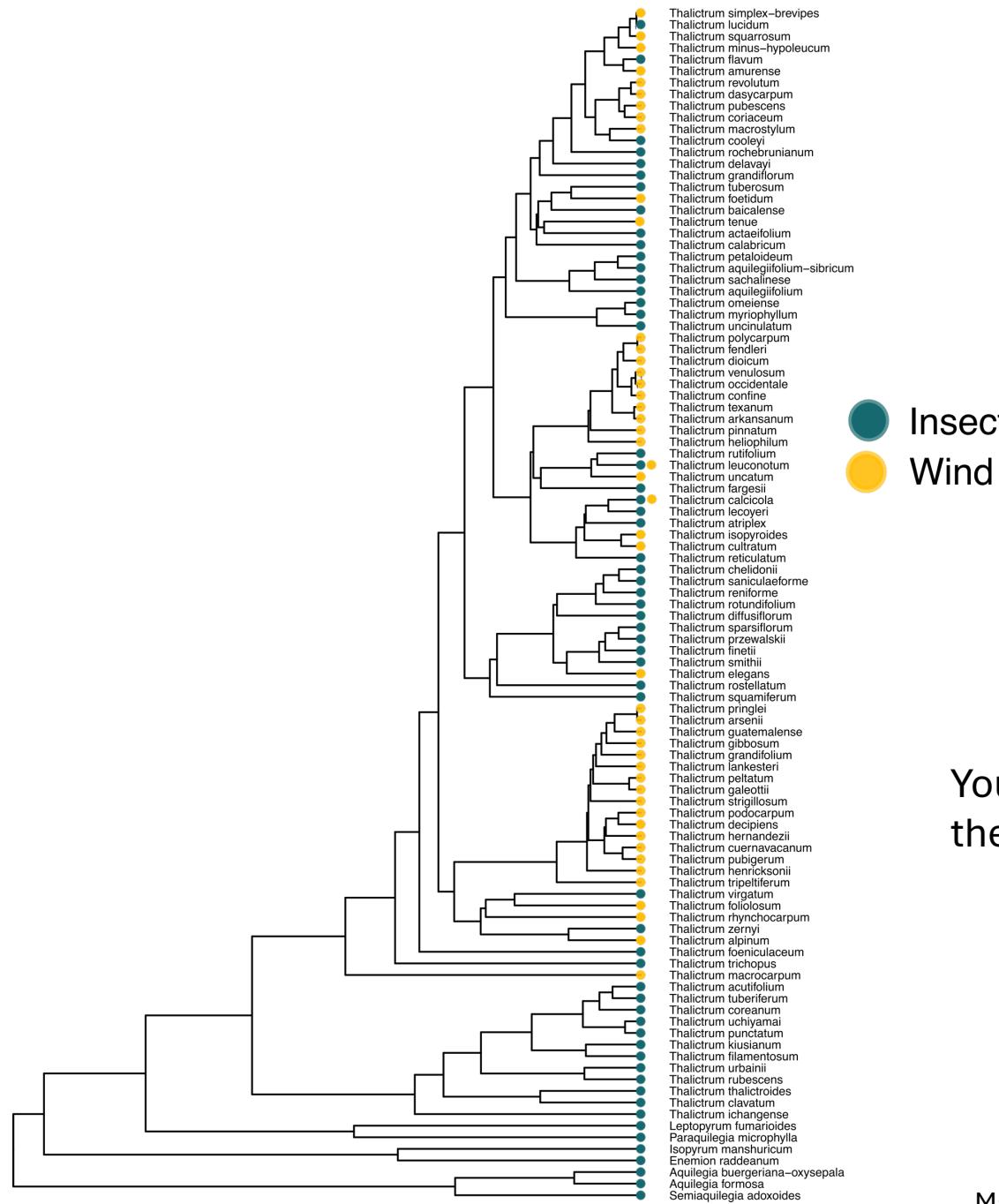
Thalictrum thalictroides

Pollination modes



many pollination
mode changes





- Always plot your data!
- Check for nestedness of trait!
- Transitions are the effective sample size for these analyses

You cannot apply discrete models if you are in the "Felsenstein worst case scenario"



Maddison and Fitzjohn, 2014



Uyeda, Zenil-Ferguson, and Pennell, 2018

Random variable definition

Continuous-Time Markov Chains (CTMC)

$$\{X(t), t \geq 0\}$$

Stochastic models that follow change in time with an associated **probability**

$X(t)$ = phenotype (trait) value at time t

t= millions of years or expected number of substitutions

Discrete models of evolution

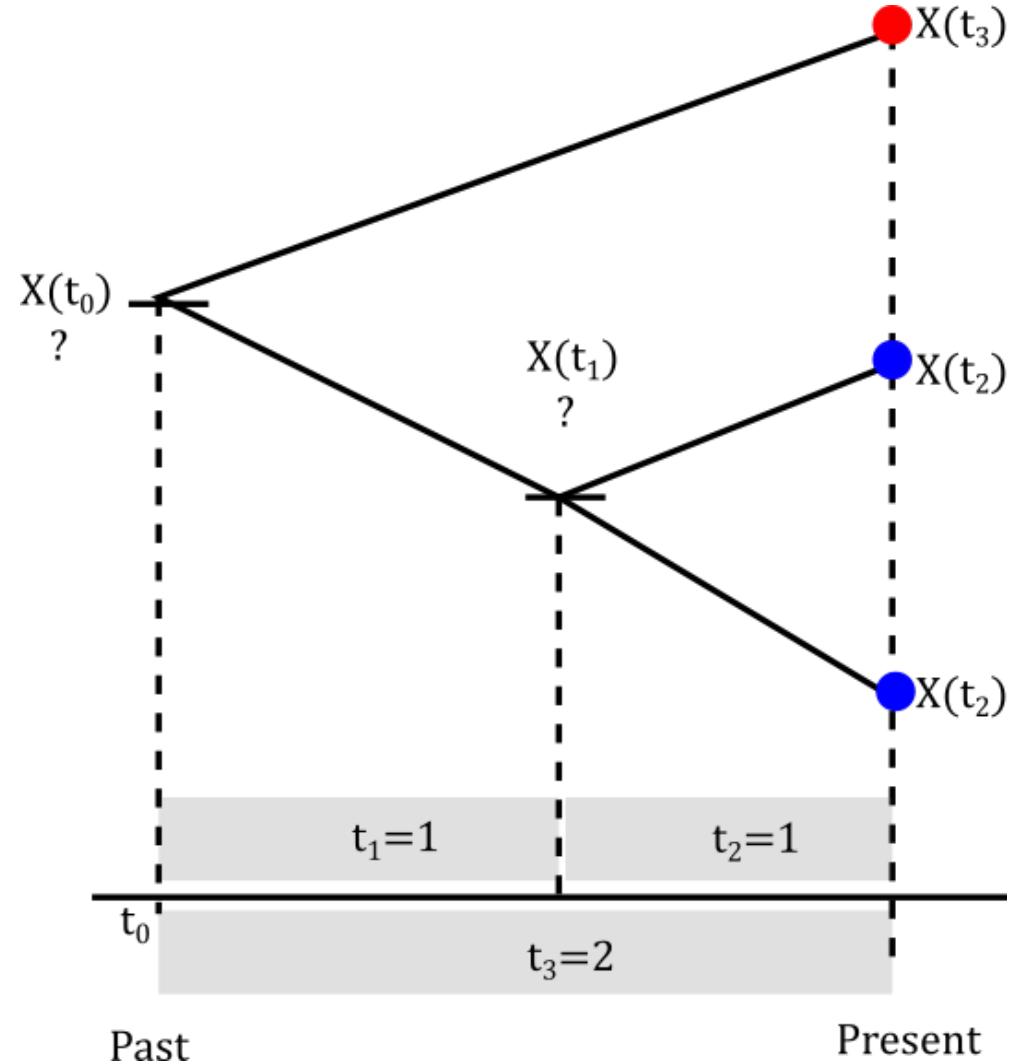
$X(t)$ = phenotype (trait) value at time t

$X(t)$ = Insect (0), Wind (1)

We use conditional probabilities. For example:

$$P(X(t) = 1 | X(0) = 0)$$

Data: Phylogenetic tree and traits

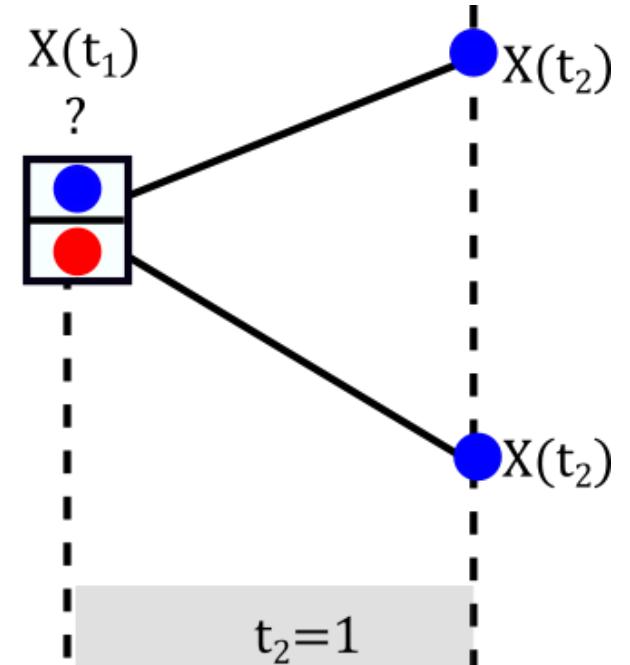


Calculating transition probabilities

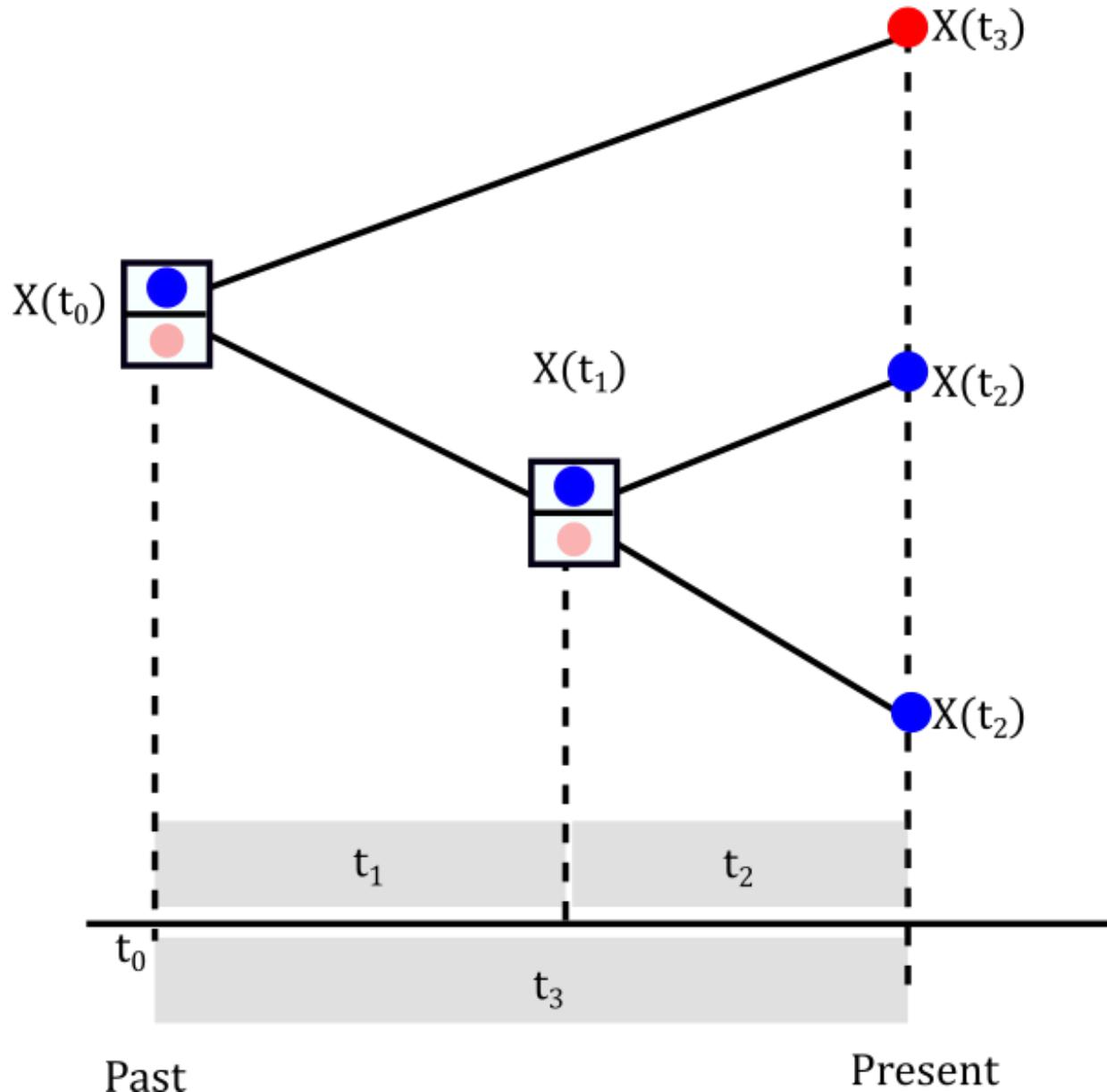
$$P(X(t) = 1 | X(0) = 0) = \frac{1}{q_{01} + q_{10}} (q_{10} - q_{01} e^{-(q_{01} + q_{10})t})$$

Past

Present



What about the rest of the tree?



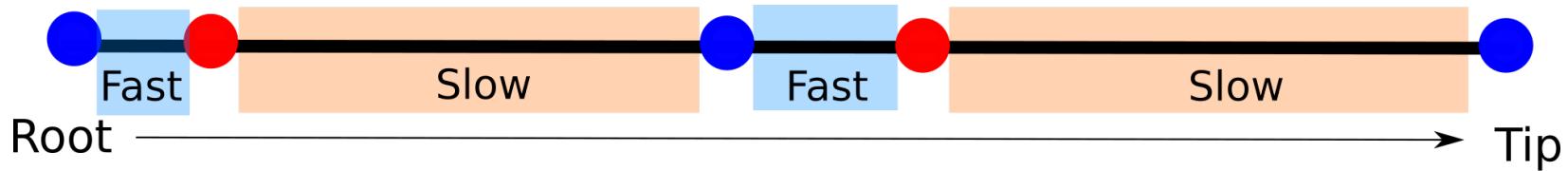
Probabilities and the Q-matrix

$$P(t) = \frac{1}{q_{01} + q_{10}} \begin{pmatrix} q_{10} + q_{01}e^{-(q_{01}+q_{10})t} & q_{01} - q_{01}e^{-(q_{01}+q_{10})t} \\ q_{10} - q_{10}e^{-(q_{01}+q_{10})t} & q_{01} + q_{10}e^{-(q_{01}+q_{10})t} \end{pmatrix}$$

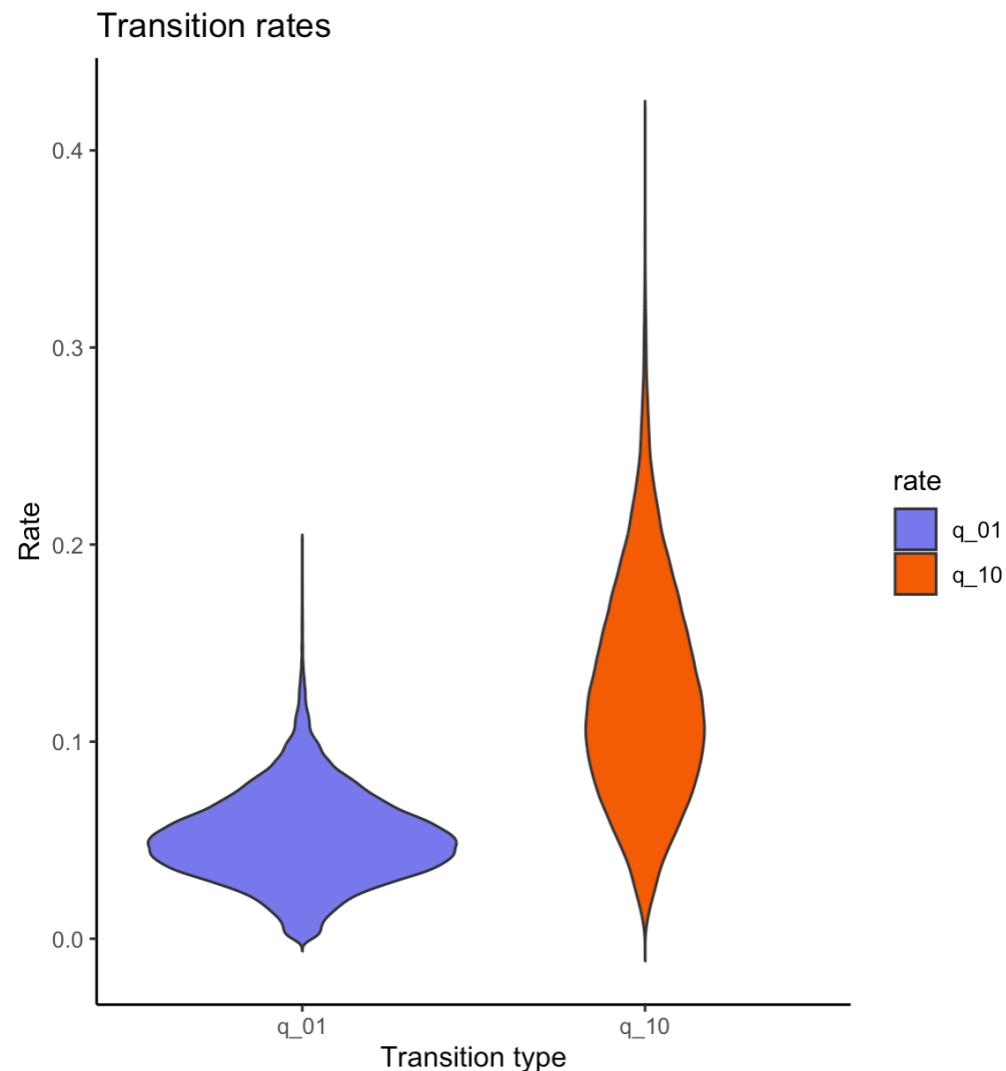
$$P(t) = e^{Qt}$$

Drawing the model

Rates: Change per time



Posterior distribution: Transition rates

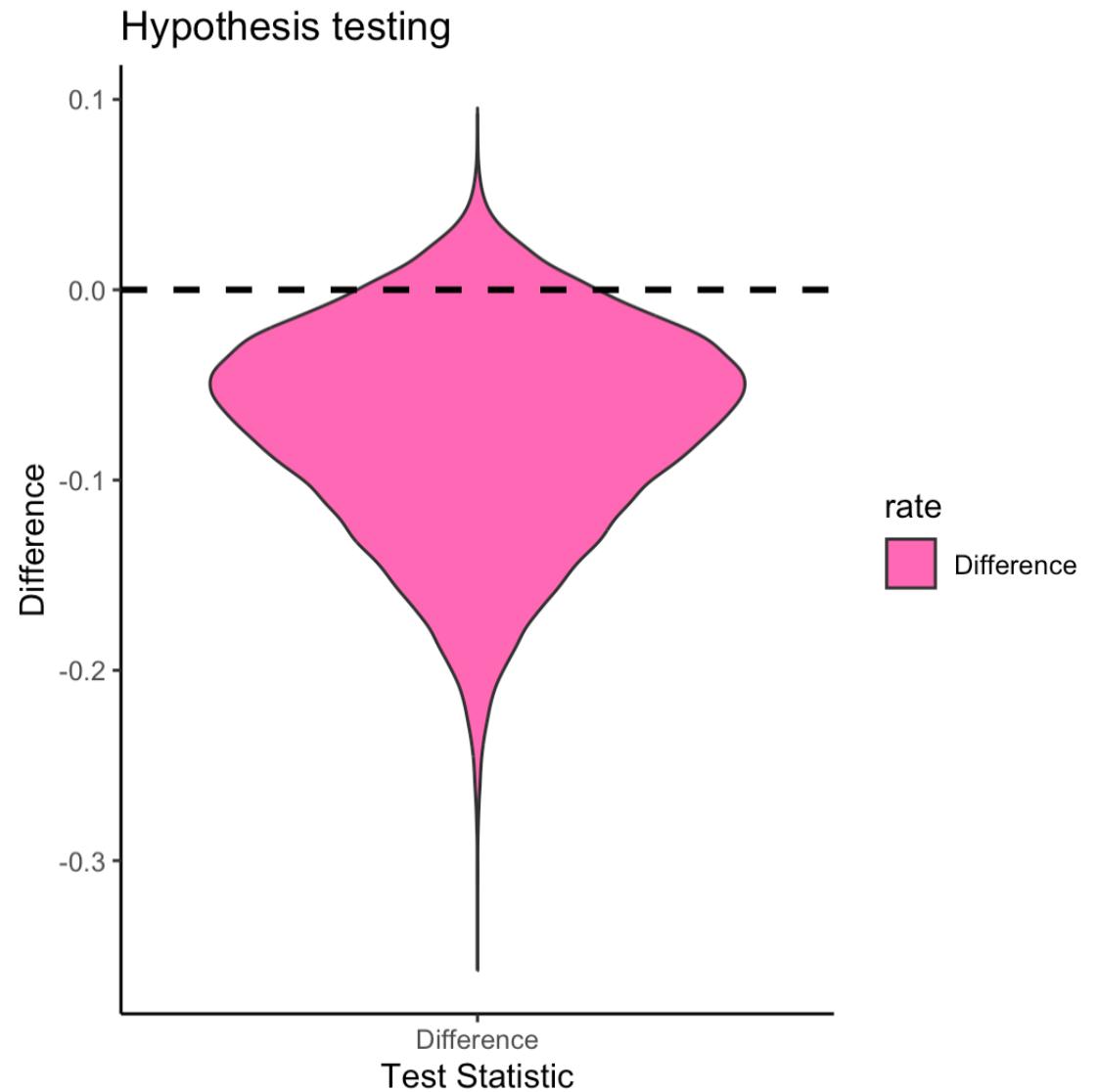


Hypothesis testing- Bayesian framework

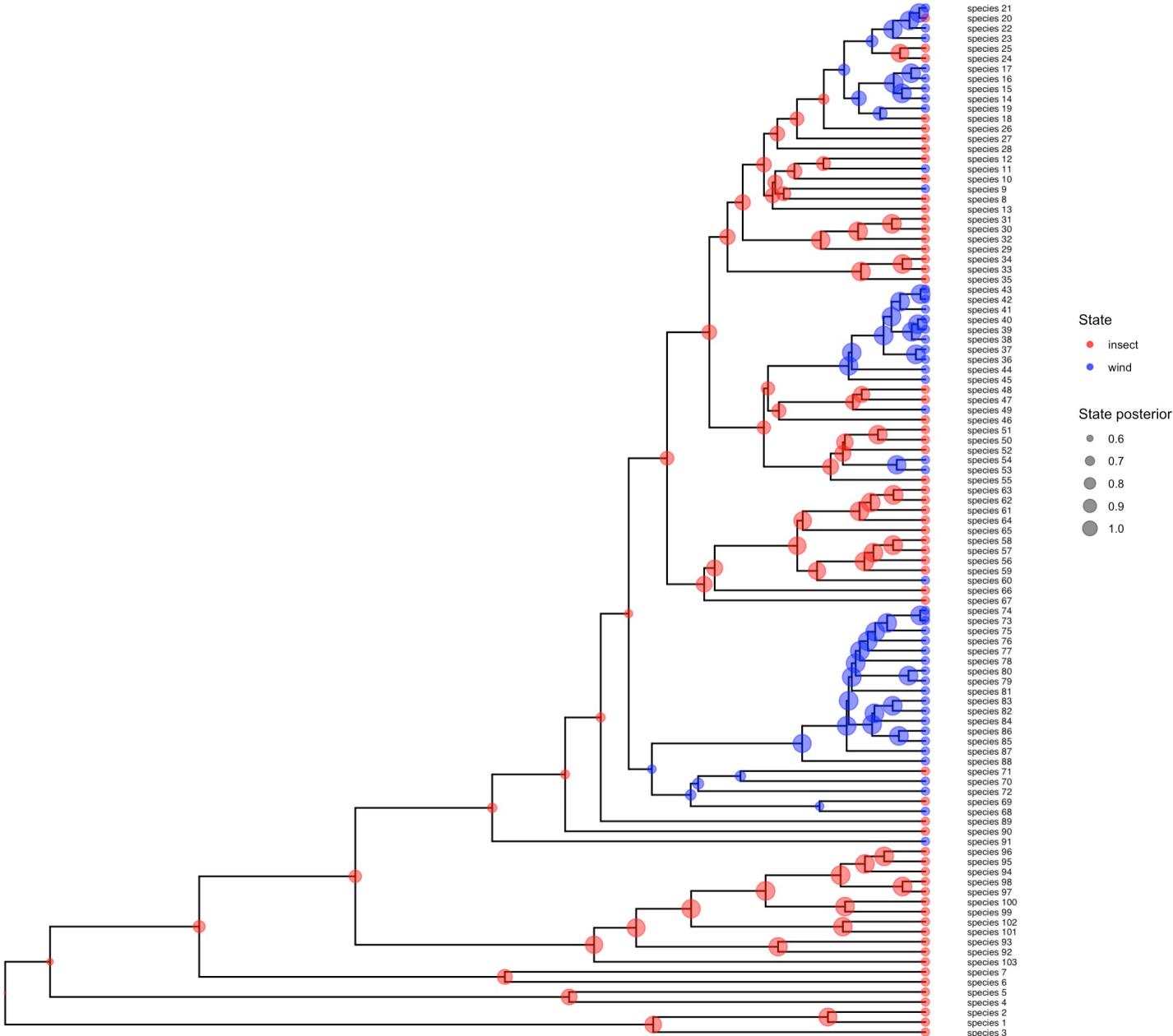
Null hypothesis $H_0: q_{01} = q_{10}$

Test statistic $D = q_{01} - q_{10}$

HPD (D)=



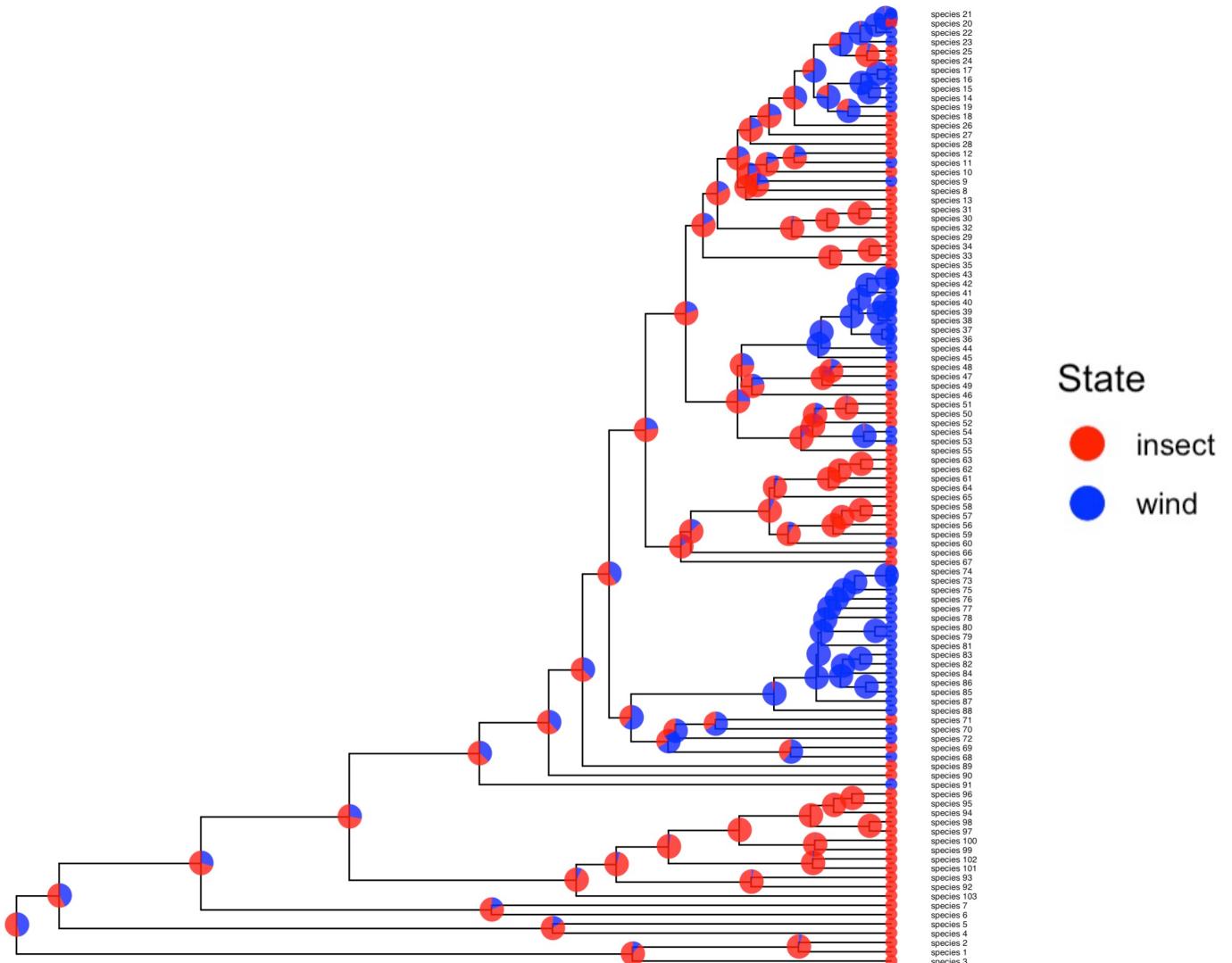
Ancestral state reconstructions



- Bayesian: Posterior distribution at each node.
- Marginal: Focus on one node at a time summing (integrating) over the rest of the nodes and maximizing the probability of the given node
- Global: Find the combination of node values that maximizes the total probability

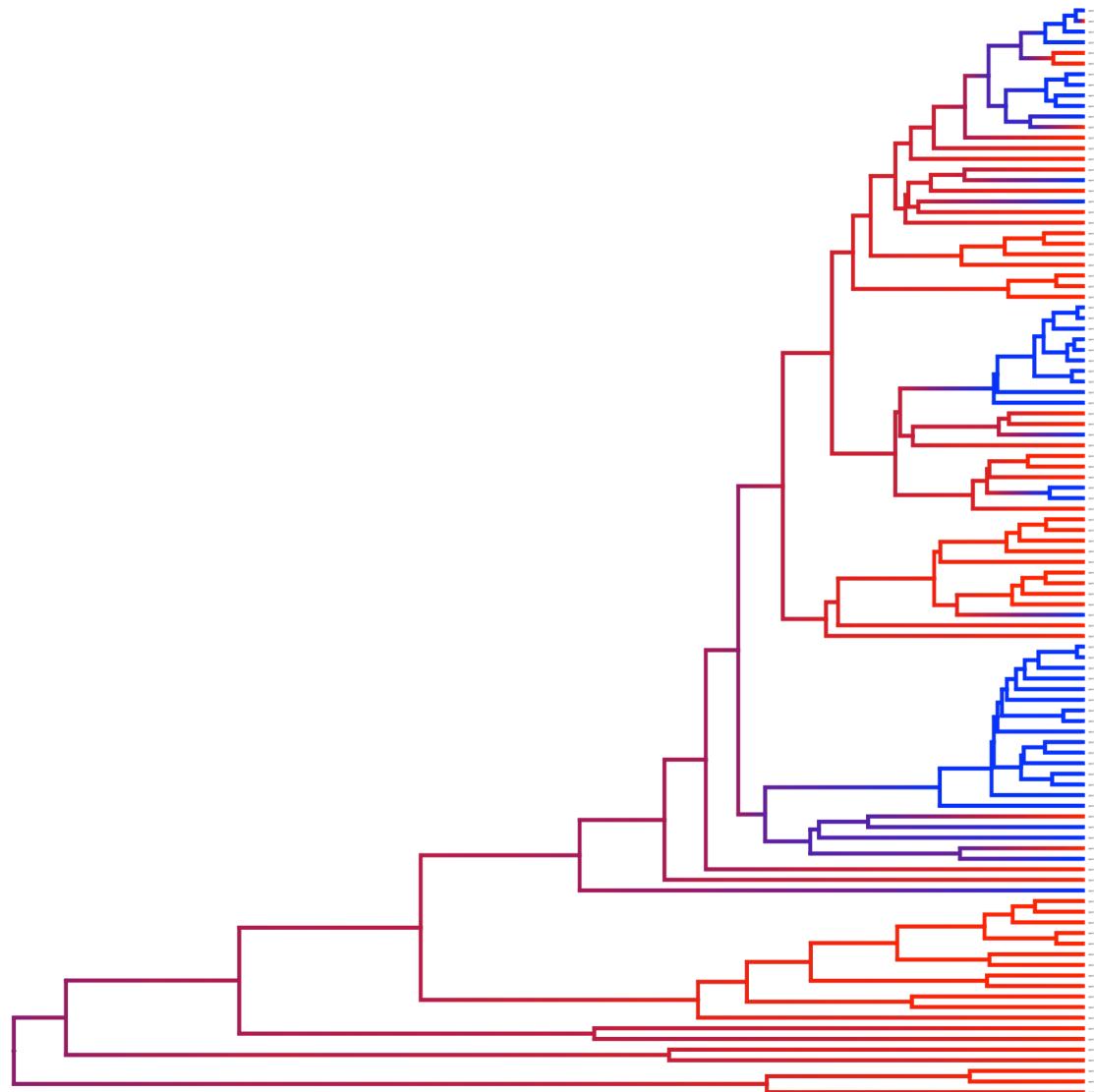
Can you talk about the number of transitions with an ancestral reconstruction?

Ancestral state reconstruction (marginal)



- **Warning:** Pies are often used by other non Bayesian software to do marginal ancestral state reconstructions. Those packages use the **marginal likelihood** function

Stochastic mapping



- Many simulations= stochastic maps
- Posterior distribution in small time slices along branches (summarized as the MAP in each of the chunks)
- *Can be summarized with other statistics

Can you talk about the number of transitions with an stochastic map?