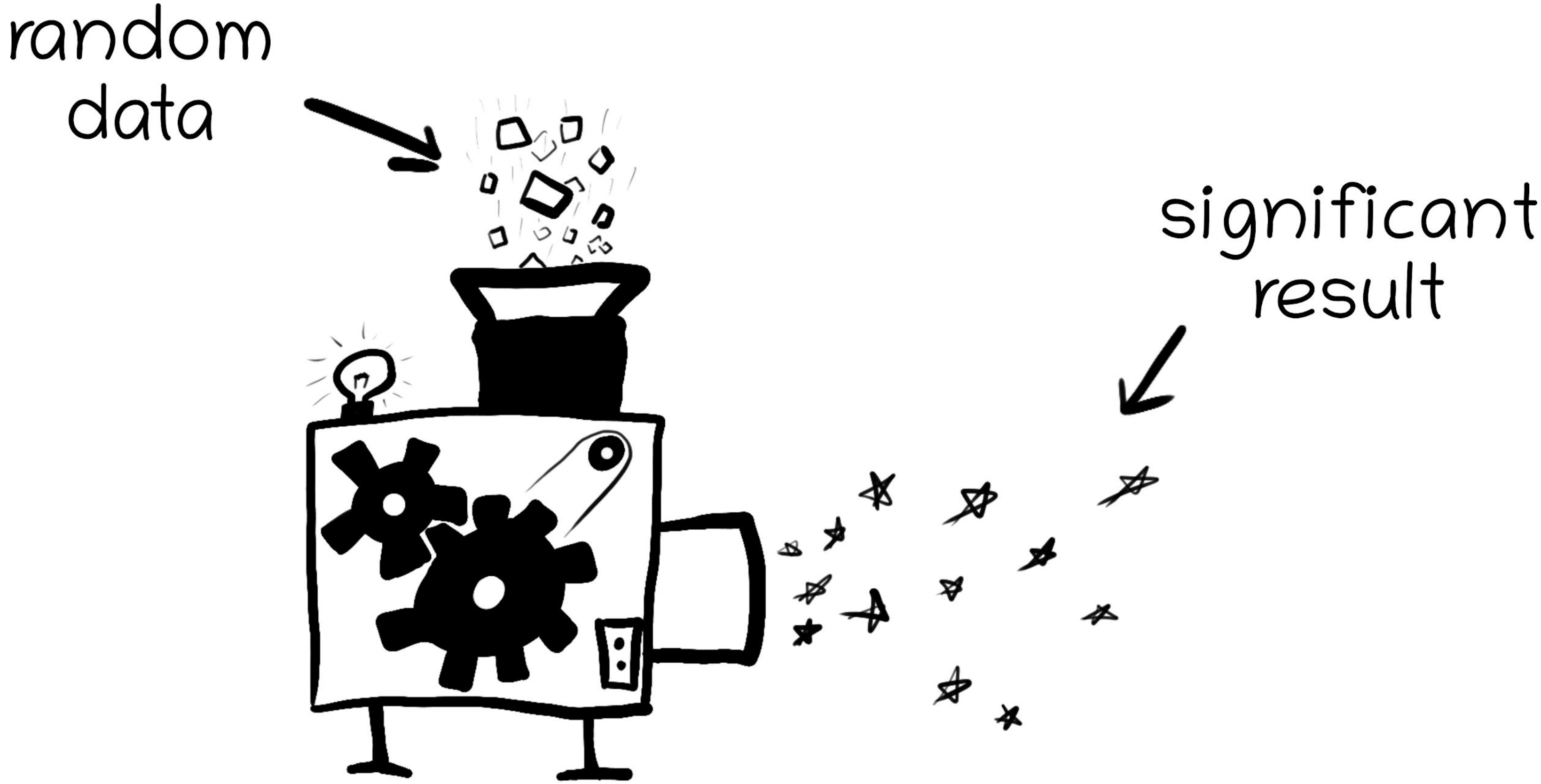
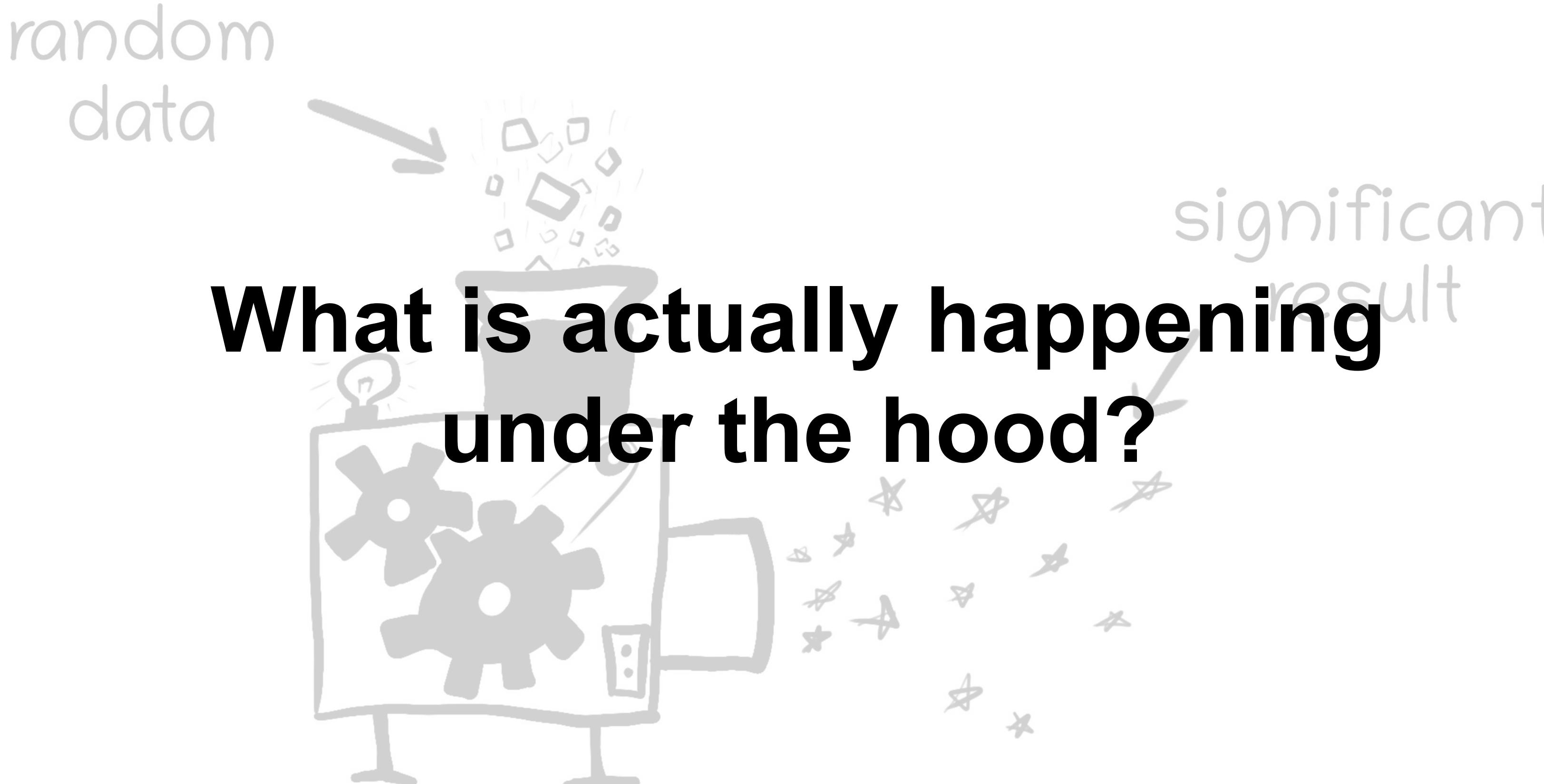
Hidden States SSE models

CIDs - Character Independent models

The crisis with SSE models:



The crisis with SSE models:



What is actually happening under the hood?

whales & dolphins

(Steeman et al. 2009)

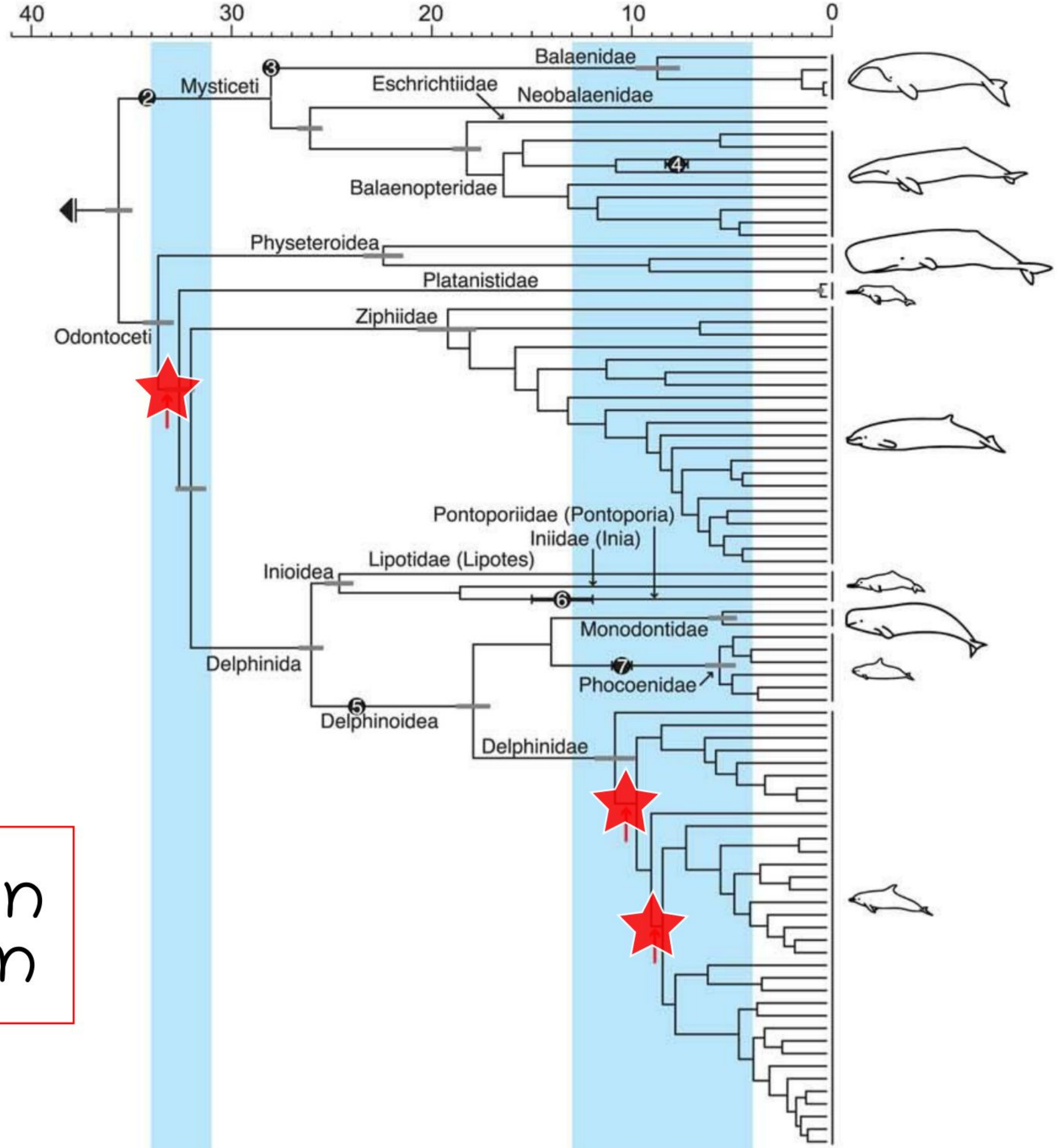
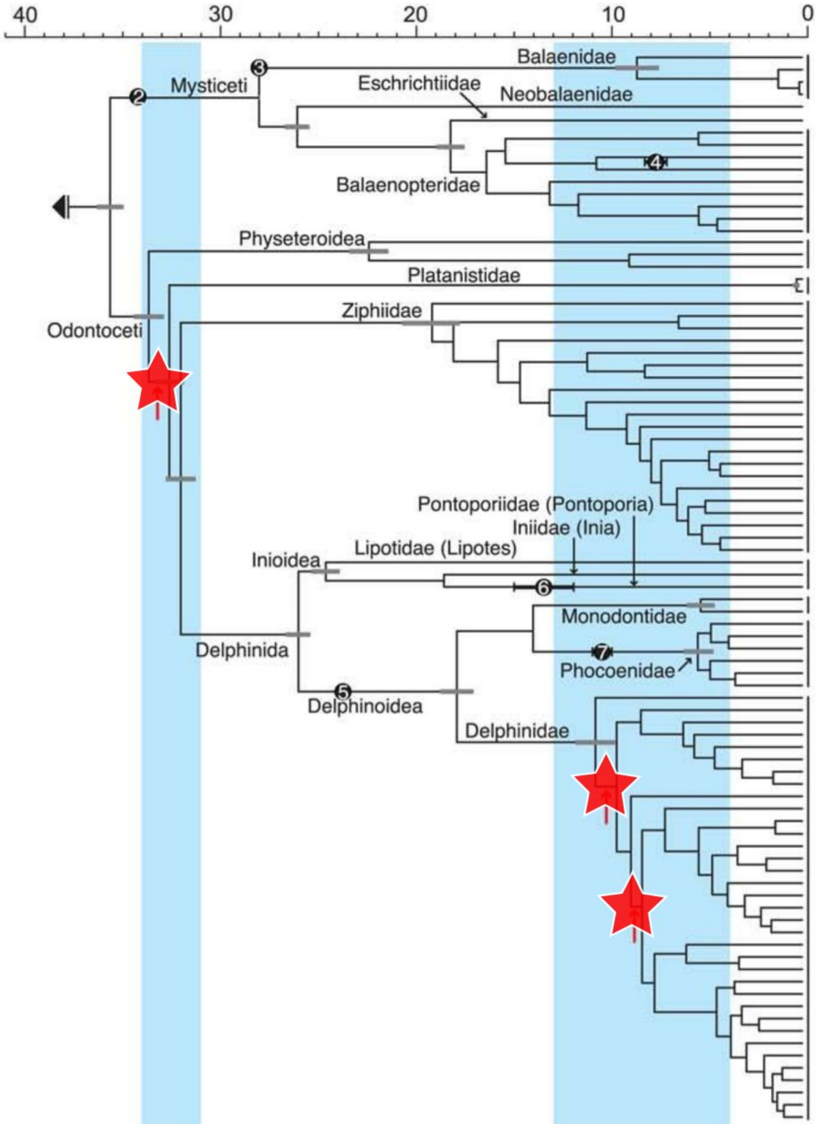
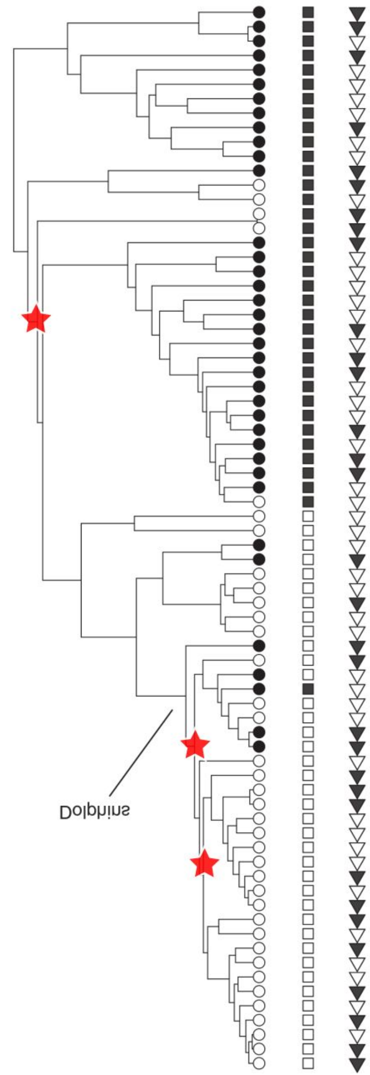


figure extracted from Steeman et al. 2009

★ major shifts in diversification

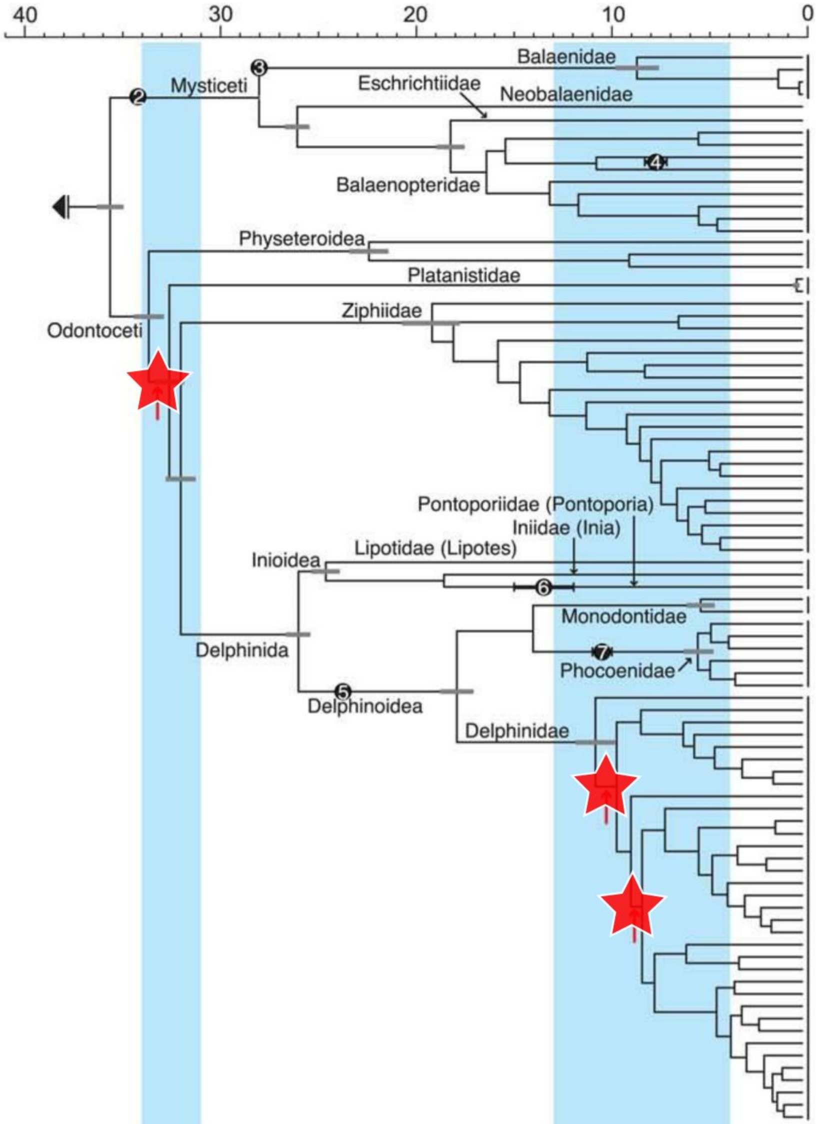


(Steeman et al. 2009)

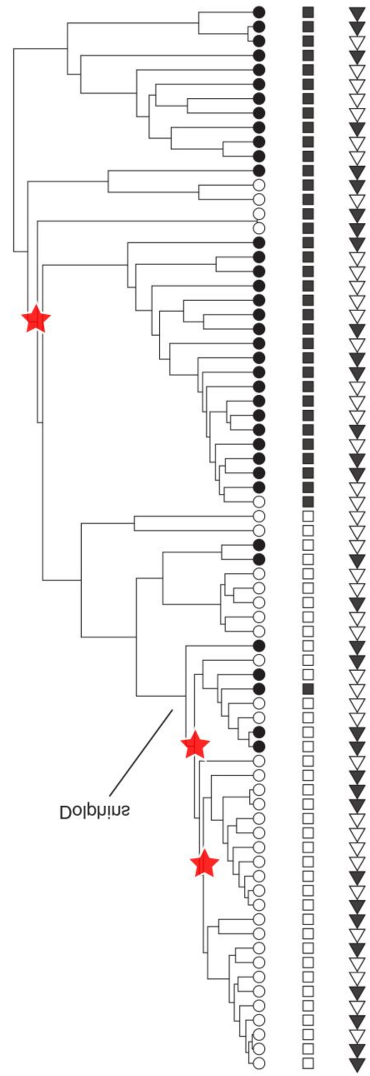


(Rabosky and Goldberg, 2015)

How useful is to reject a model that assumes a single rate of diversification?



(Steeman et al. 2009)



(Rabosky and Goldberg, 2015)

How useful is to reject a model that assumes a single rate of diversification?

State-dependent diversification?

or

Multiple rates of diversification?

Using empirical phylogenies and random traits we get a very high rate of false positives with standard BiSSE models

TABLE 1. Type I error rates for neutral character simulations on 186 empirical phylogenies of birds, squamates, amphibians, and fishes

State frequency	N	$P < 0.05$	$P < 0.001$
$0.10 \leq x < 0.15$	1562	0.404	0.164
$0.15 \leq x < 0.20$	713	0.456	0.212
$0.20 \leq x < 0.25$	444	0.617	0.387
$0.25 \leq x < 0.30$	241	0.631	0.402
$0.30 \leq x < 0.35$	291	0.632	0.423
$0.35 \leq x < 0.40$	171	0.608	0.298
$0.40 \leq x < 0.45$	163	0.669	0.423
$0.45 \leq x < 0.50$	128	0.656	0.414

Notes: Results are binned by frequency of the rarer character state and are pooled across transition rates and clades. Data are identical to those presented in Figure 5. The second column gives the number of simulated realizations. The third and fourth columns report the proportion of realizations for which state-independent diversification is rejected.

Rabosky and Goldberg, 2015

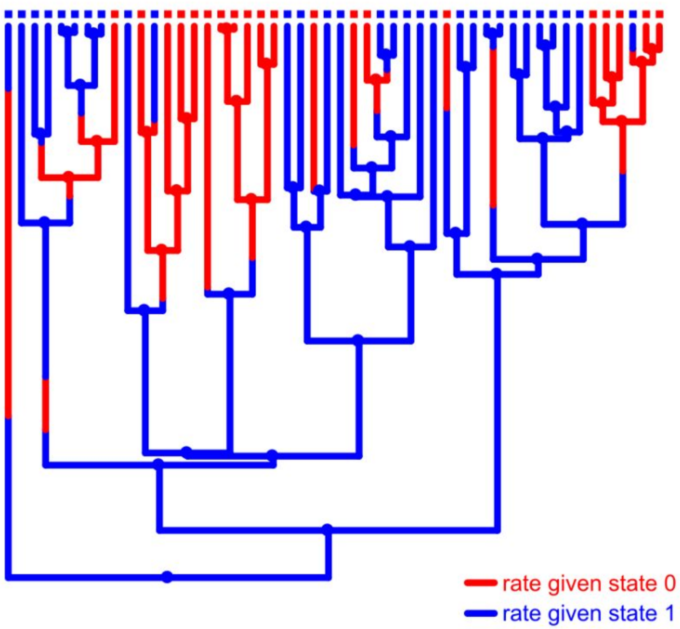
Hidden rates SSE models

state-dependent diversification

$\lambda_{0A} = \lambda_{0B}$ $\quad$ $\mu_{0A} = \mu_{0B}$

$\lambda_{1A} = \lambda_{1B}$ $\quad$ $\mu_{1A} = \mu_{1B}$

4 free diversification parameters



state-independent diversification

$\lambda_{0A} = \lambda_{1A}$ $\quad$ $\mu_{0A} = \mu_{1A}$

$\lambda_{0B} = \lambda_{1B}$ $\quad$ $\mu_{0B} = \mu_{1B}$

4 free diversification parameters

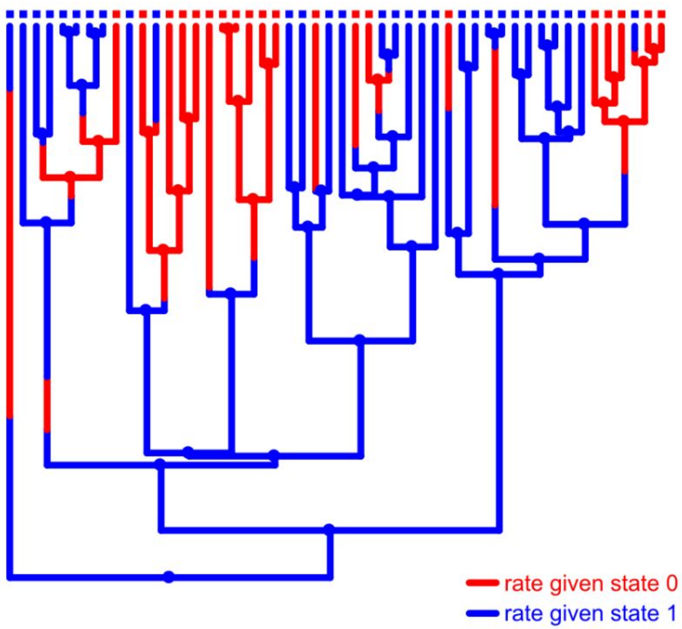
Hidden rates SSE models

Effect model

$$\lambda_{0A} = \lambda_{0B} \qquad \mu_{0A} = \mu_{0B}$$

$$\lambda_{1A} = \lambda_{1B} \qquad \mu_{1A} = \mu_{1B}$$

4 free diversification parameters

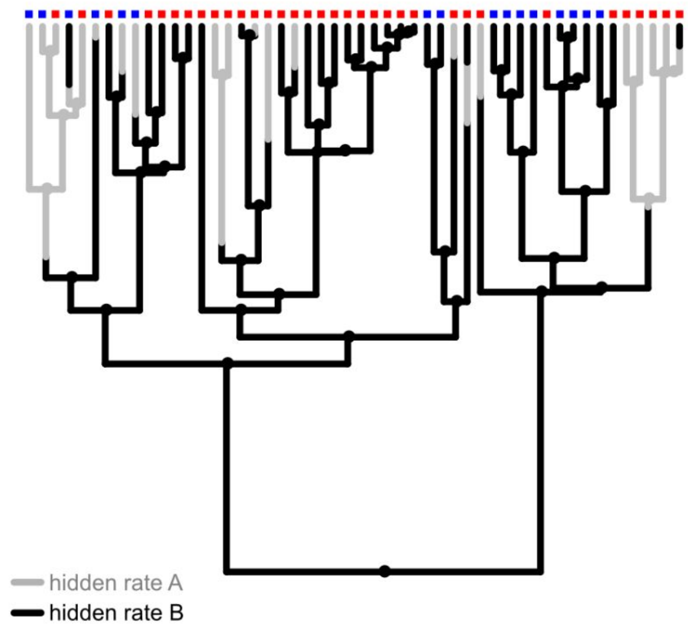


CID model

$$\lambda_{0A} = \lambda_{1A} \qquad \mu_{0A} = \mu_{1A}$$

$$\lambda_{0B} = \lambda_{1B} \qquad \mu_{0B} = \mu_{1B}$$

4 free diversification parameters



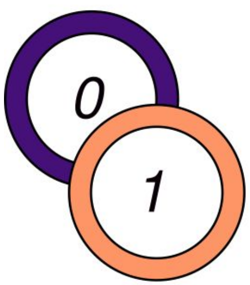
GeoSSE

Geographic SSE models

d_0 - dispersion

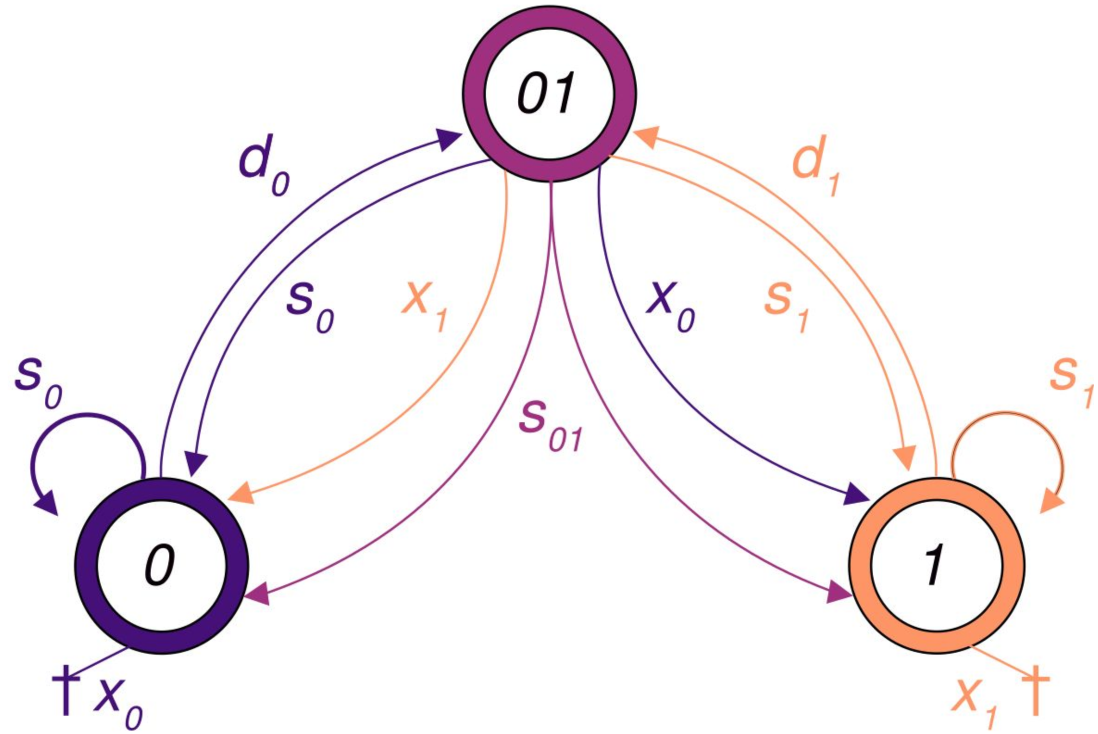
s_0 - speciation

x_0 - extinction



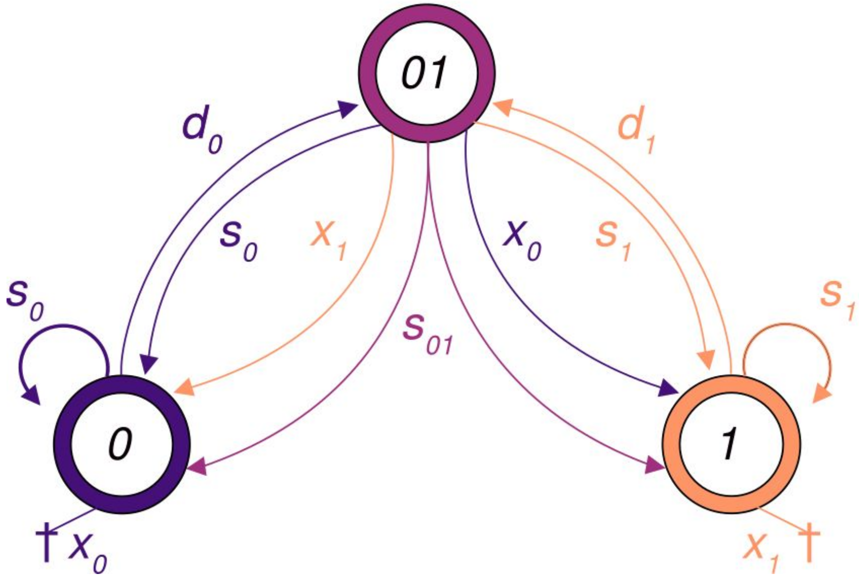
endemic areas

widespread areas



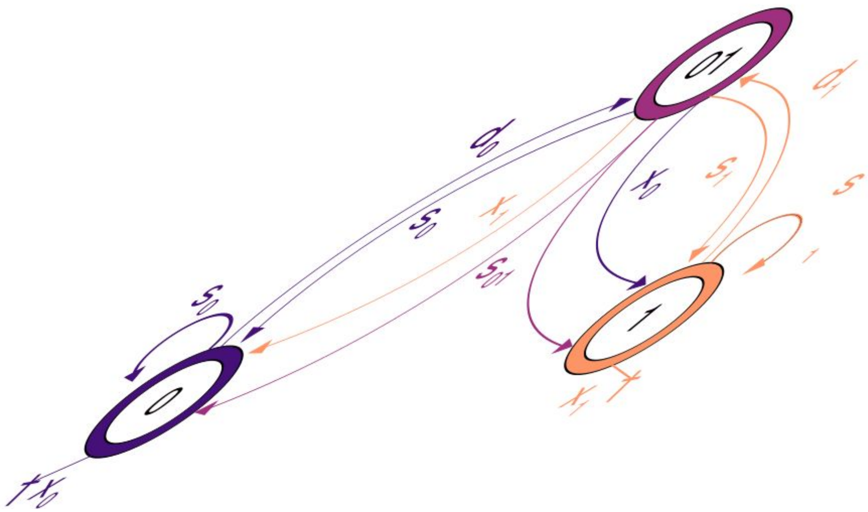
GeoHiSSE

Geographic hidden state SSE models

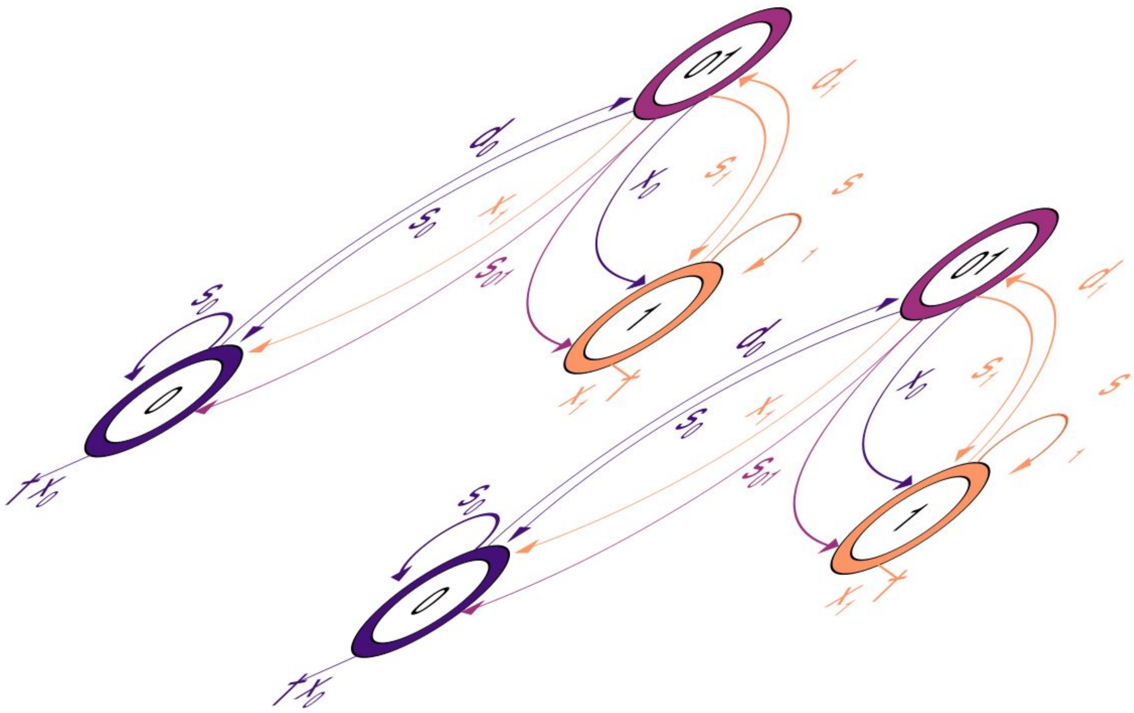


GeoHiSSE

Geographic hidden state SSE models

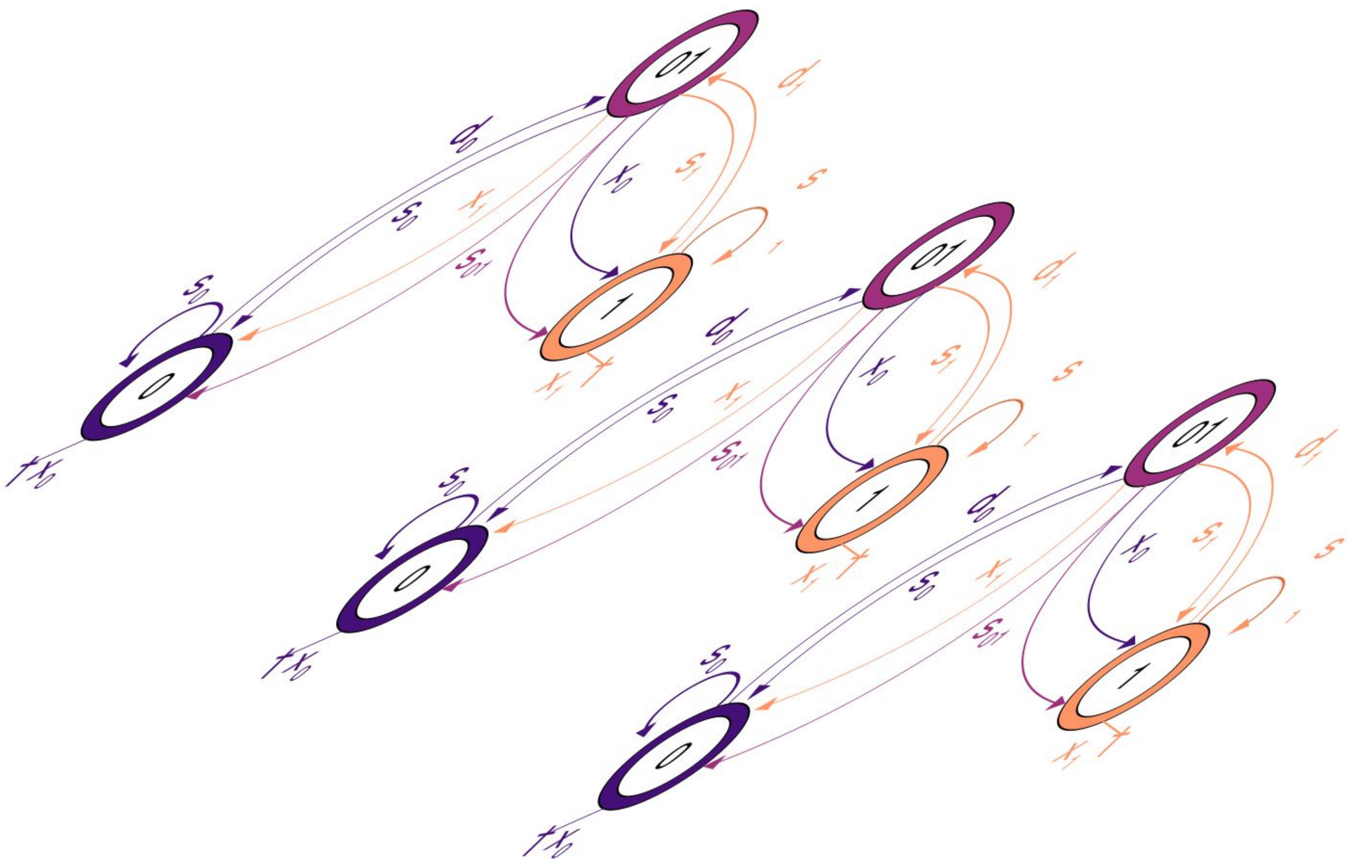


GeoHiSSE
Geographic hidden state SSE models

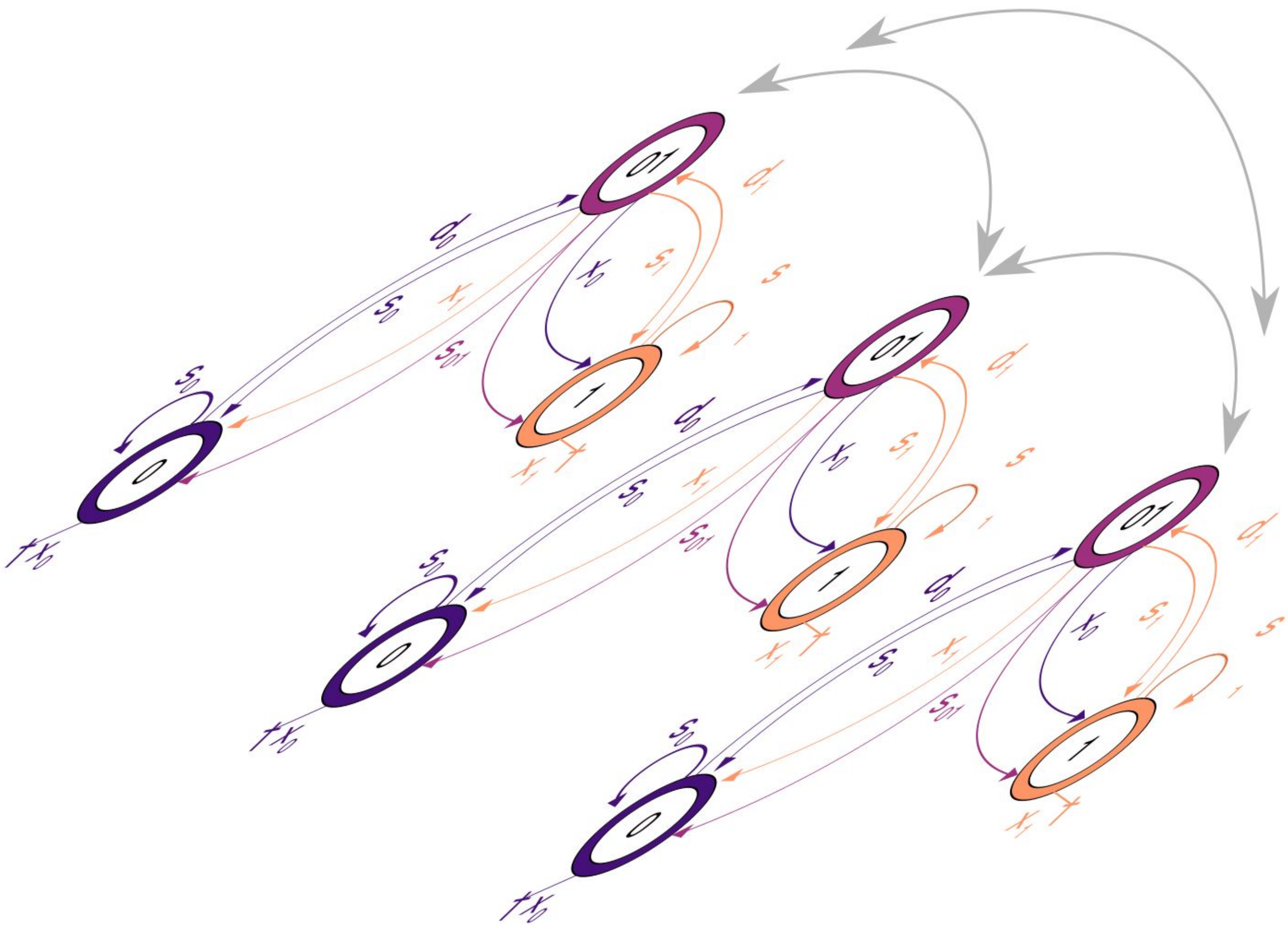


GeoHiSSE

Geographic hidden state SSE models

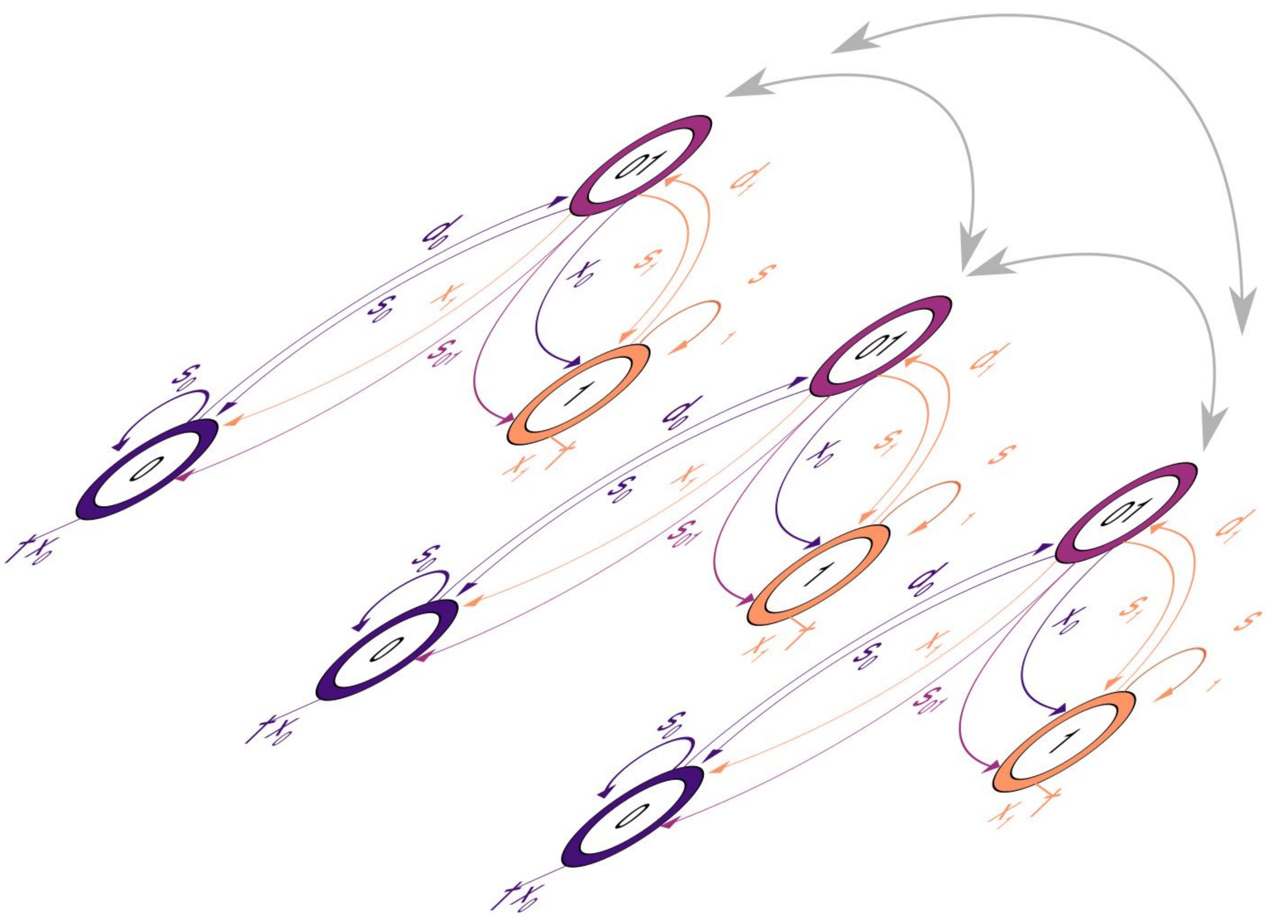


GeoHiSSE
Geographic hidden state SSE models



GeoHiSSE

Geographic hidden state SSE models



Area-dependent diversification

partition diversification shifts *between* areas

Area-independent diversification

partition diversification shifts *within* areas

Review of how hidden states work and what do they mean:

SSE

SOCIETY for the STUDY of EVOLUTION

Hidden state models improve state-dependent diversification approaches, including biogeographical models

Daniel S. Caetano,[1,2] Brian C. O'Meara,[3] and Jeremy M. Beaulieu[1]

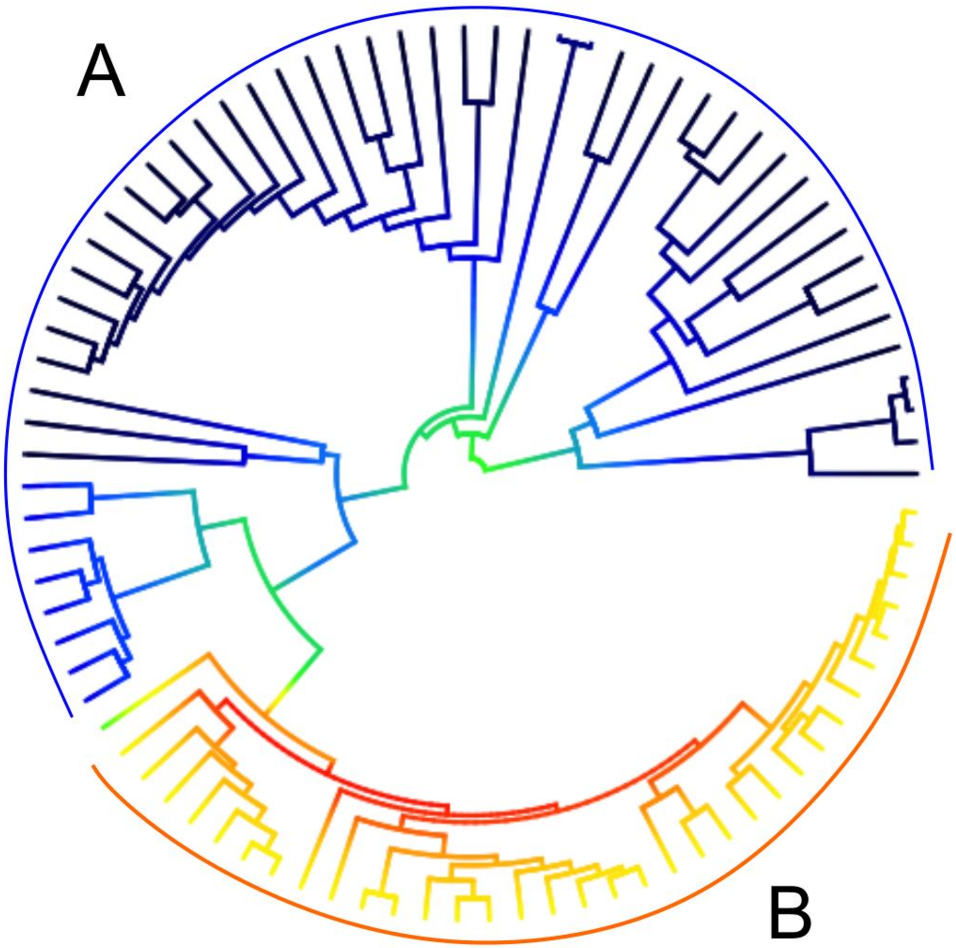
[1] *Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701*

[2] *E-mail: dcaetano@uark.edu*

[3] *Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996-1610*

Common questions when using Hidden states SSE models

"... hidden state model was favored, thus there are unobserved traits influencing diversification."



there are *always* multiple traits influencing diversification

hidden states are a tool to accomodate variation in rates across the tree

hidden states ***simplify*** the real variation on rates

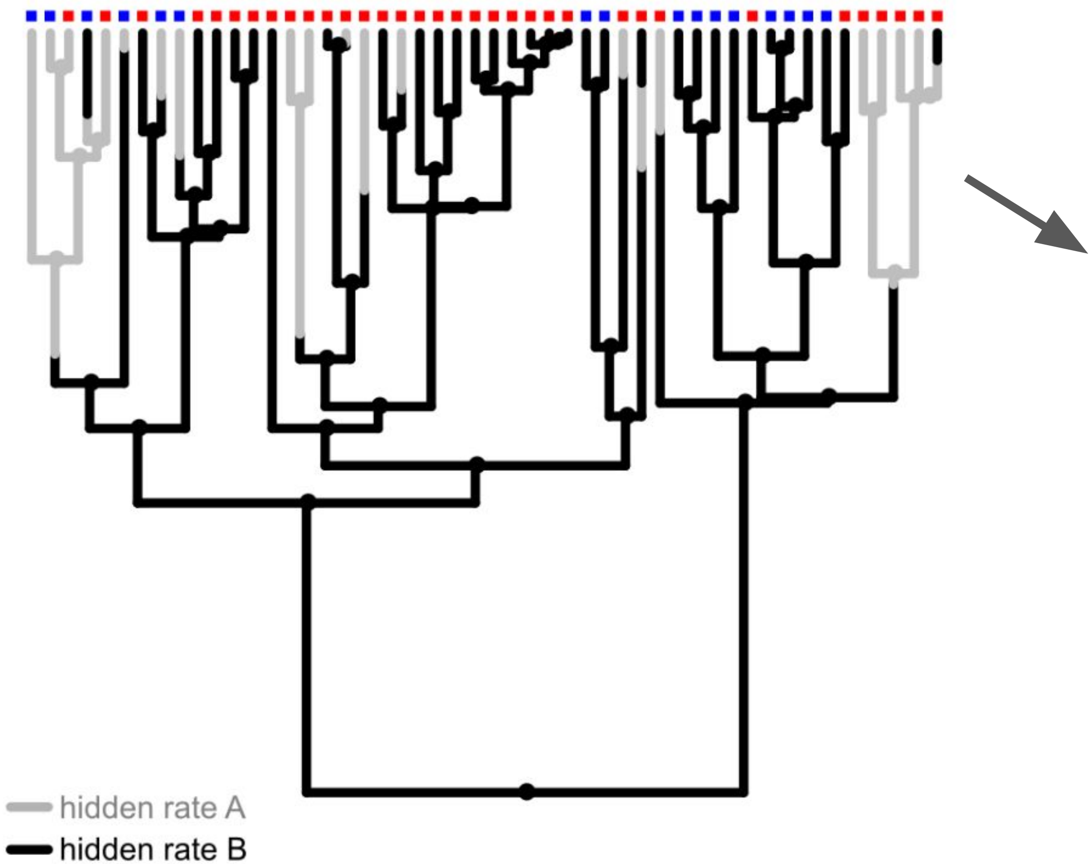
"Too many models to work with!"

or

"Model A and model B have the very similar AIC values."

Model averaging allows to take into account
multiple models that explain well the observed data.

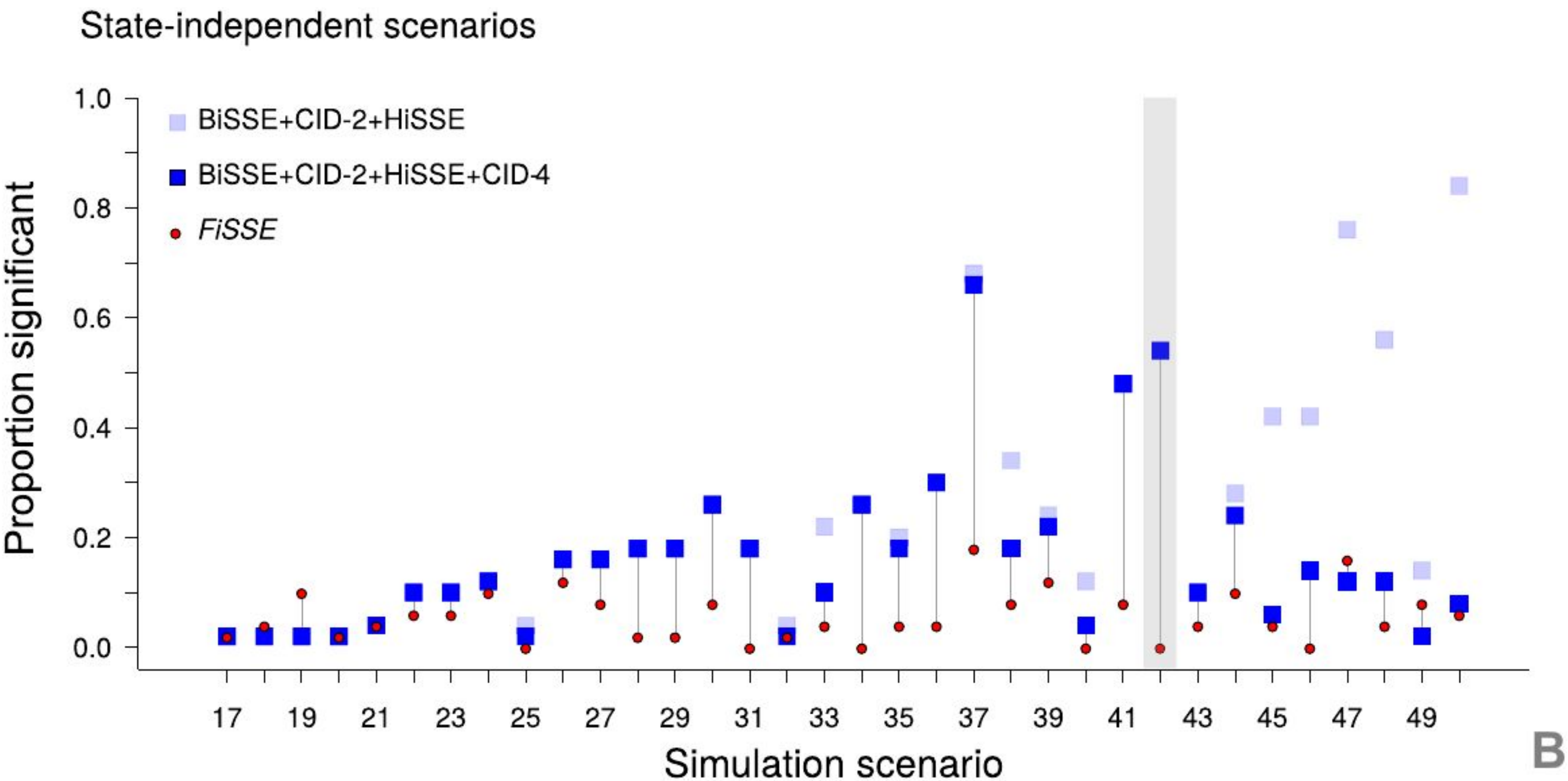
“How many hidden rate categories do we need to use?”



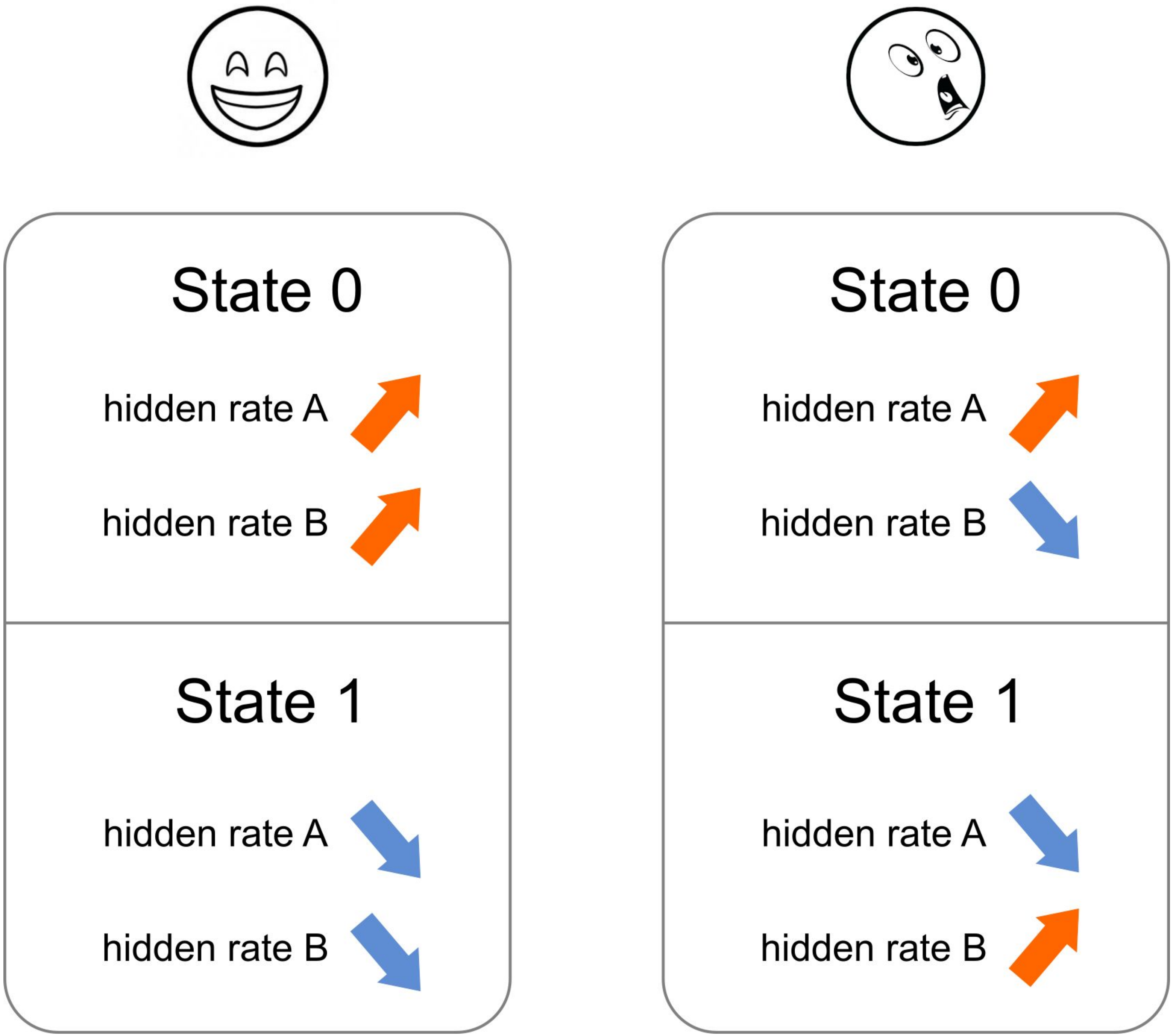
CID model with 2 hidden rate categories

… but a model with more hidden rate categories could be fitted.

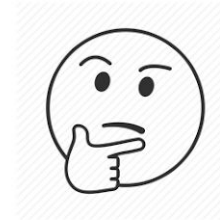
The number of **hidden rate categories** needs to match the **rate categories of the state-dependent model**



Caetano et al., 2018



State 0
hidden rate A
hidden rate B
State 1
hidden rate A
hidden rate B
State 0
hidden rate A
hidden rate B
State 1
hidden rate A
hidden rate B



Hidden states **do not** occur with the same frequency across the tree.

The weight of a hidden rate is proportional to the frequency of that hidden state in the data.

We can use equilibrium frequencies to compute the expected frequency of each hidden rate among tips and nodes.