

Exploring trait-dependent models of multi-lineage processes of diversification with machine learning

Jamie Oaks
Auburn University

phyletica.org

 [joaks1](https://github.com/joaks1) & [phyletica](https://github.com/phyletica)

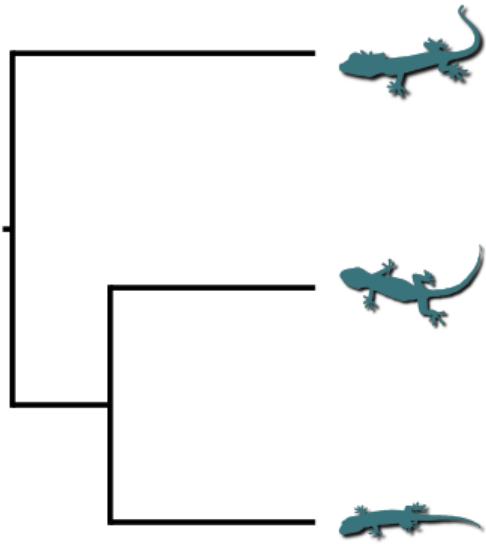
 @jamieoaks.bsky.social

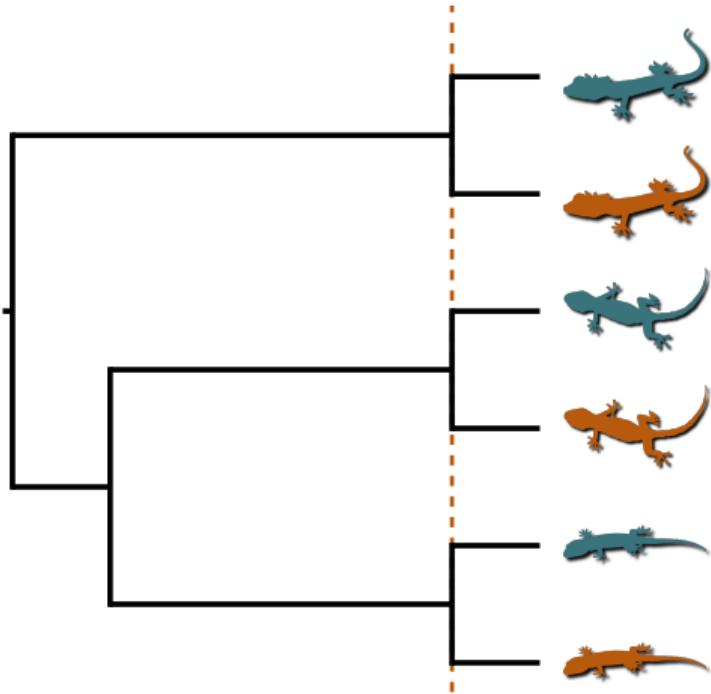


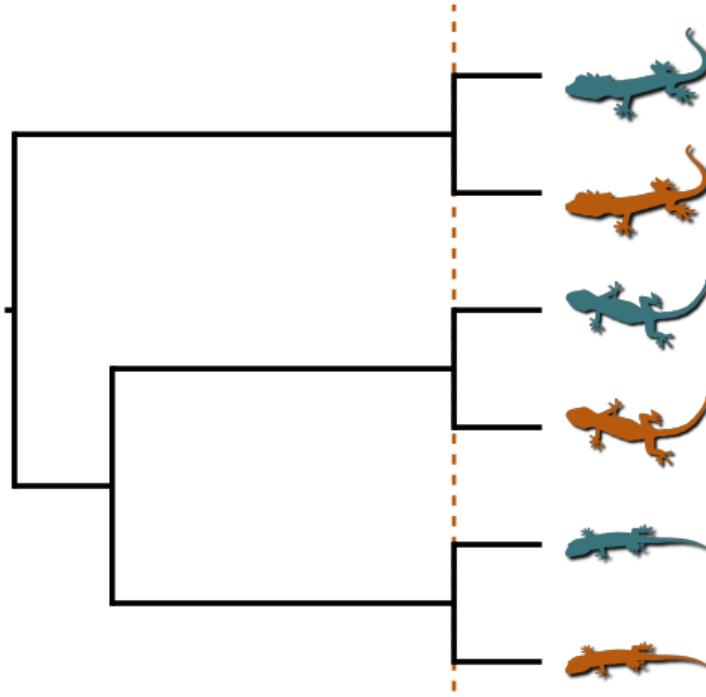
phyletica.org/slides/evol2025.pdf



© 2007 Boris Kulikov boris-kulikov.blogspot.com







These “**shared**” or “**burst**” divergences violate assumption of independent divergences

Biogeography

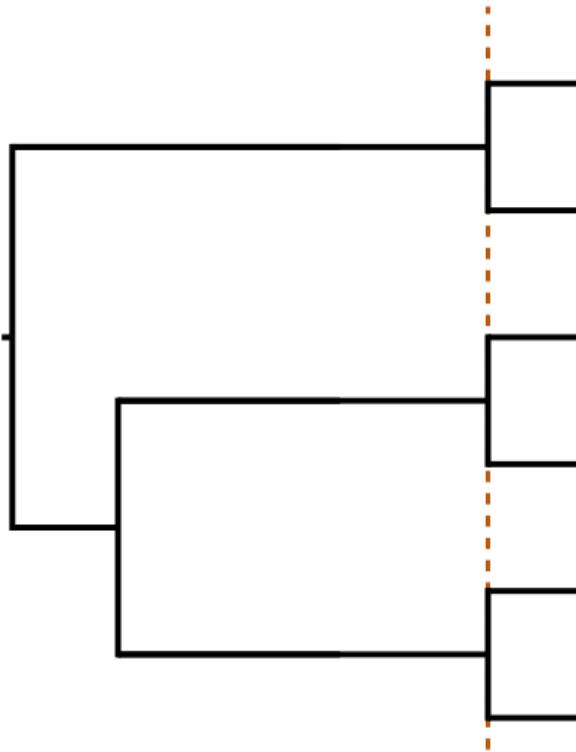
- ▶ Environmental changes that affect whole communities of species

Epidemiology

- ▶ Transmission at social gatherings

Genome evolution

- ▶ Duplication of a chromosome segment harboring a gene family



Biogeography

- ▶ Environmental changes that affect whole communities of species

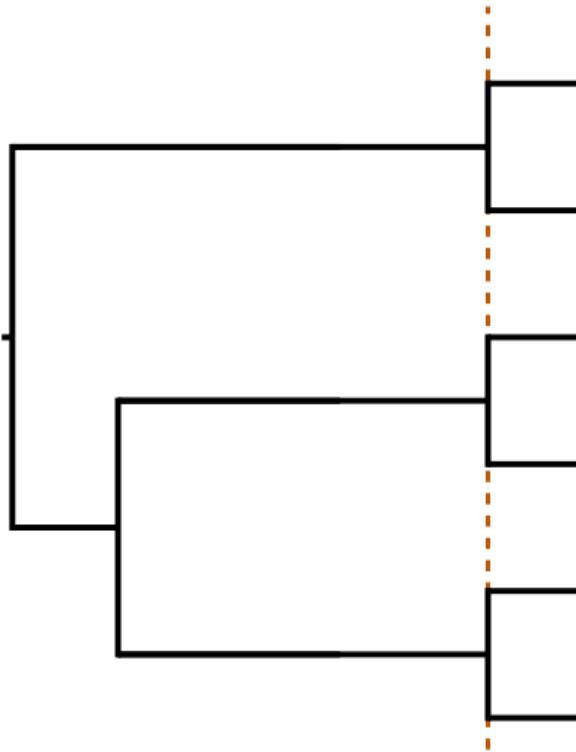
Epidemiology

- ▶ Transmission at social gatherings

Genome evolution

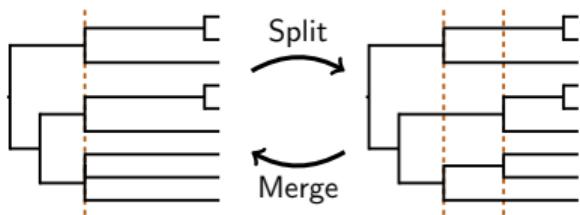
- ▶ Duplication of a chromosome segment harboring a gene family

These processes are interesting!

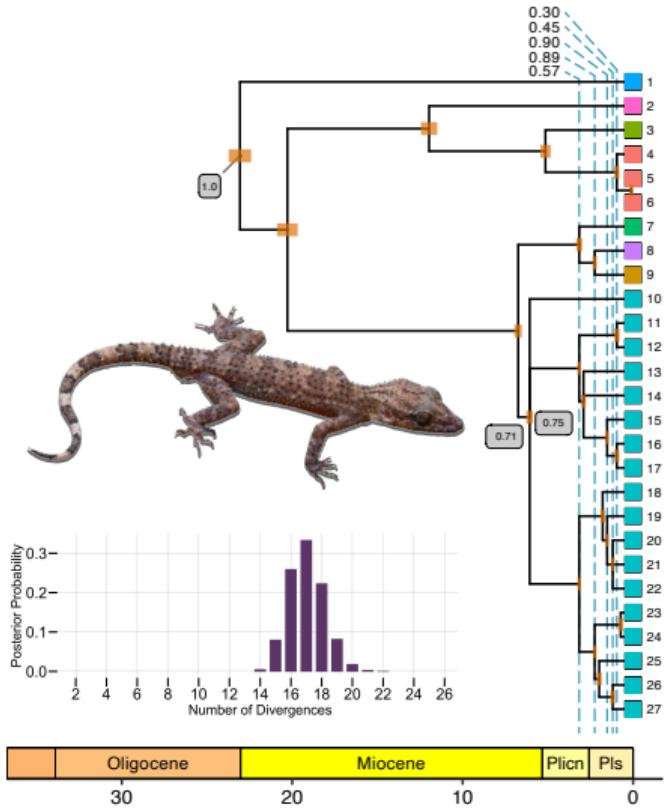


Progress on shared divergences: Phycoeval

PhycoEval



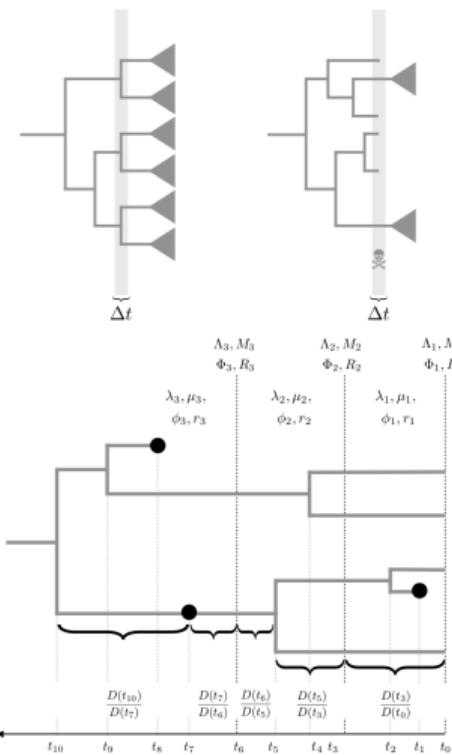
- ▶ Bayesian inference of trees with shared divergences



Progress on shared divergences: Birth-Death + Bursts

- ▶ Andrew Magee & Sebastian Höhna
- ▶ Likelihood of tree under BD model with “burst” events

(e) Burst (f) Mass Extinction



Motivating questions

- ▶ Did geckos adapted to karst diversify via fragmentation of karst formations?



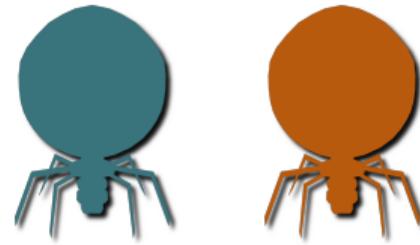
Credit: Perry Wood, Jr.; phylopic.org CC0

Motivating questions

- ▶ Did geckos adapted to karst diversify via fragmentation of karst formations?



- ▶ Do viral strains differ in ability to spread at social gatherings?



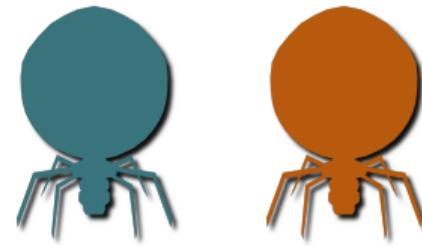
Credit: Perry Wood, Jr.; phylopic.org CC0

Motivating questions

- ▶ Did geckos adapted to karst diversify via fragmentation of karst formations?



- ▶ Do viral strains differ in ability to spread at social gatherings?

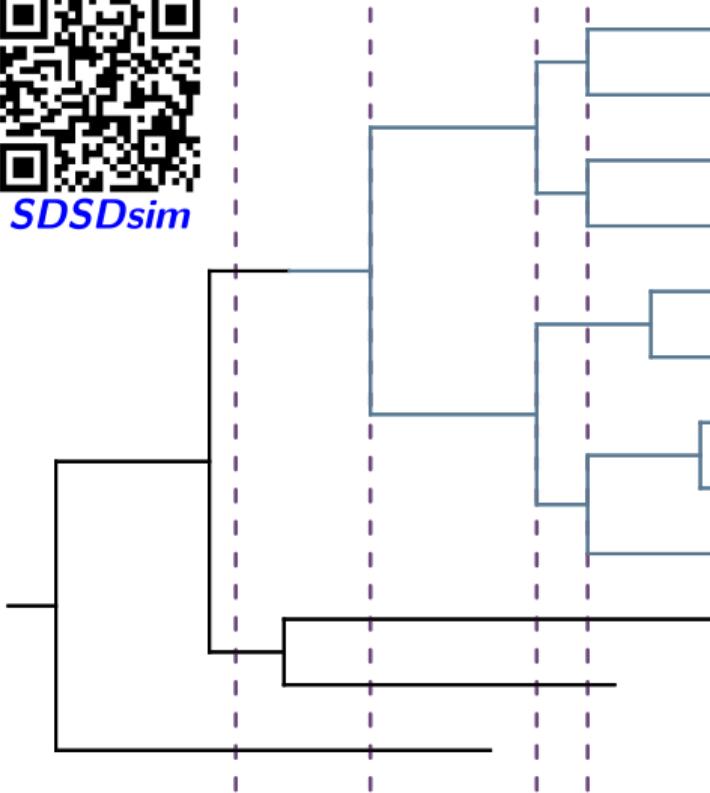


These questions require state-dependent models

State-dependent model of shared divergences



SDSsim



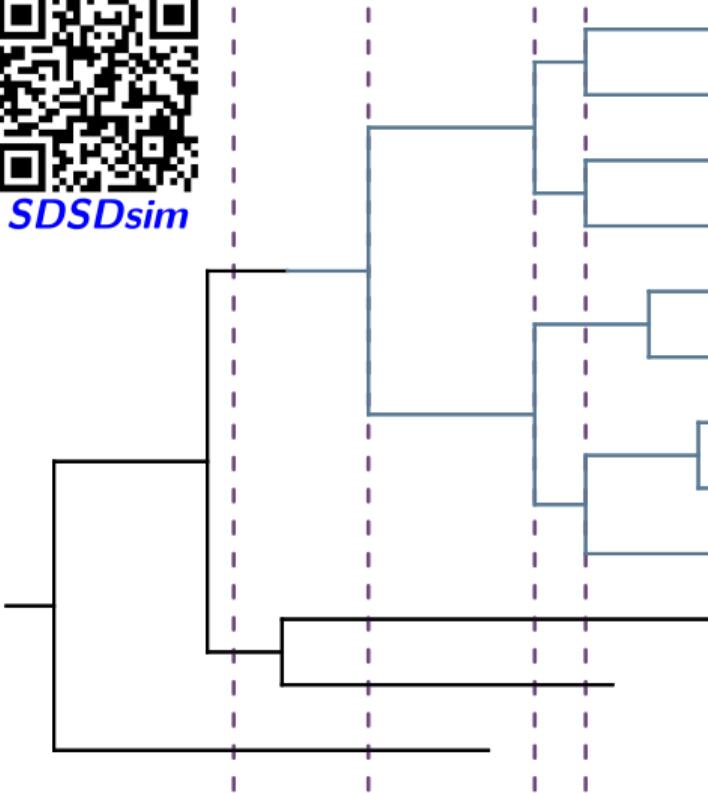
Parameters:

- ▶ Transition rates between character
State 0 \rightleftharpoons **State 1**
- ▶ Speciation rate, λ (per-lineage)
- ▶ Extinction rate, μ (per-lineage)
- ▶ Rate of “burst events”, β (tree-wide)
- ▶ **State 0** probability of speciation at events, p_0
- ▶ **State 1** probability of speciation at events, p_1

State-dependent model of shared divergences



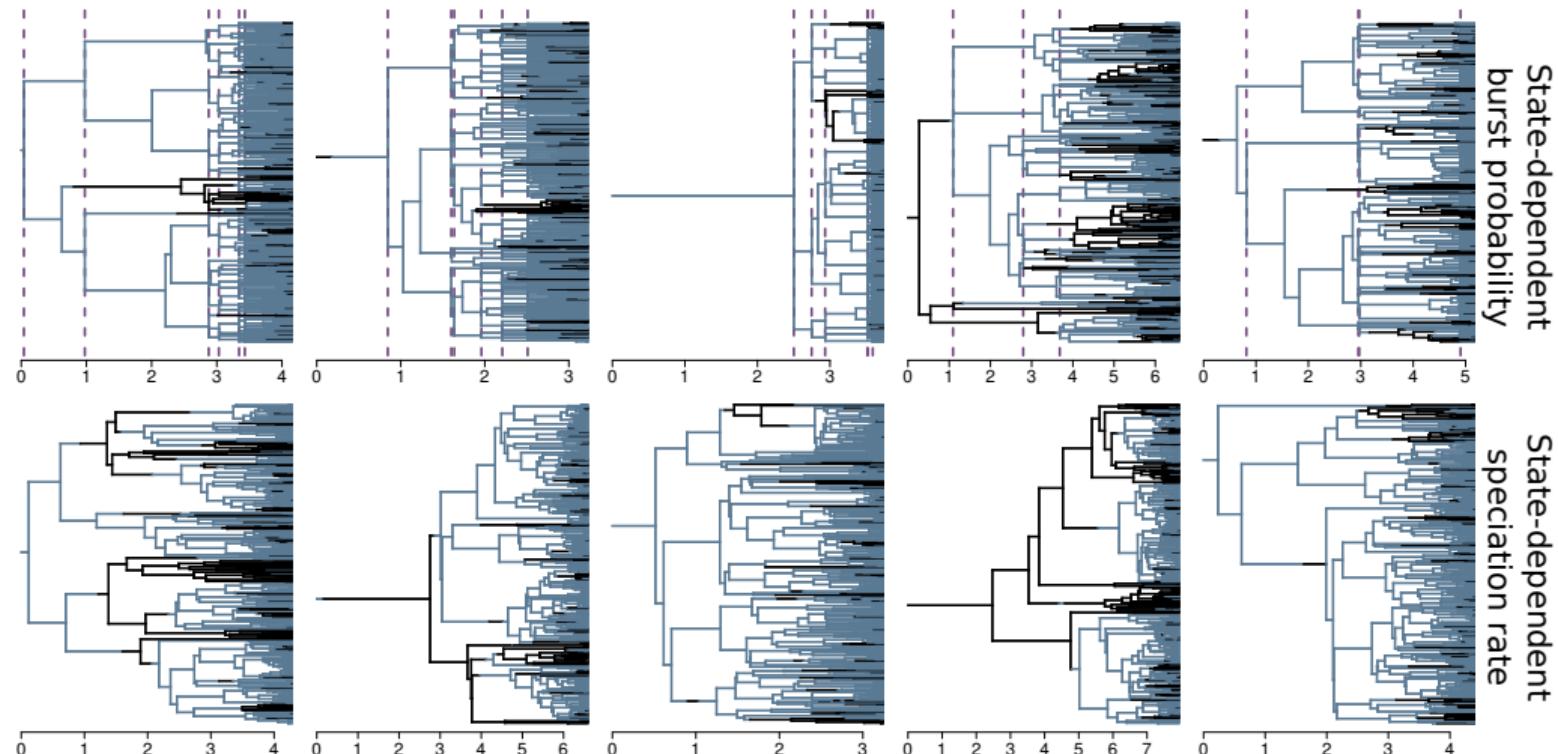
SDSsim



Parameters:

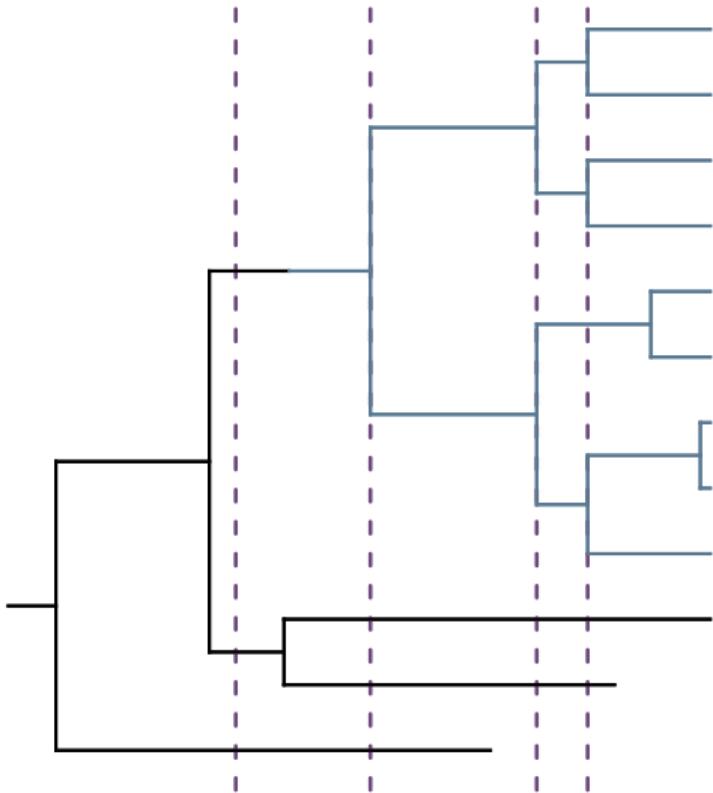
- ▶ Transition rates between character
State 0 \rightleftharpoons **State 1**
- ▶ Speciation rate, λ (per-lineage)
- ▶ Extinction rate, μ (per-lineage)
- ▶ Rate of “burst events”, β (tree-wide)
- ▶ **State 0** probability of speciation at events, p_0
- ▶ **State 1** probability of speciation at events, p_1
 - ▶ $p_1\beta =$ Per-lineage expected rate of burst divergences

Some random example trees



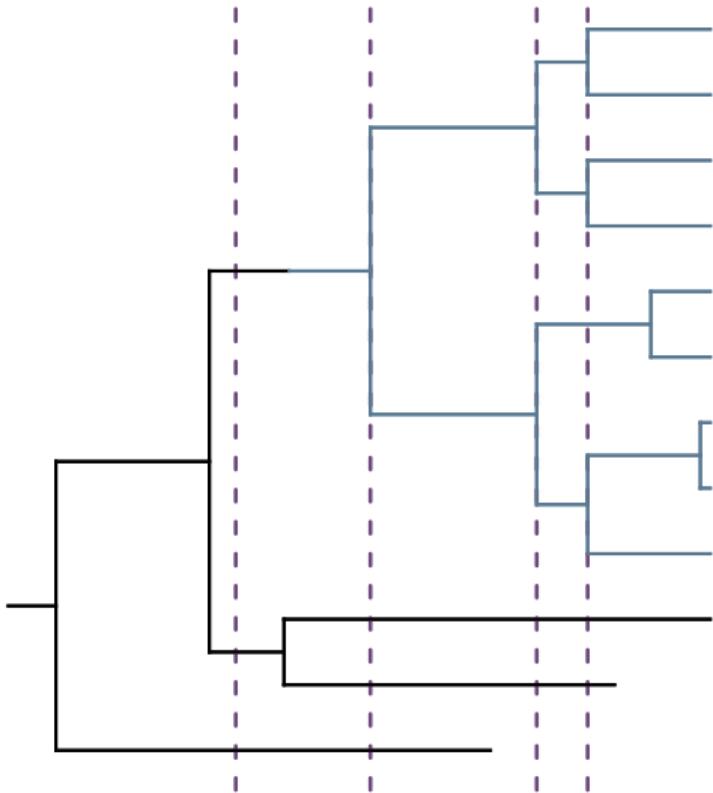
Both models have same expected rates of state-dependent diversification

What next?



Ideally, full Bayesian inference of state-dependent
burst model and tree

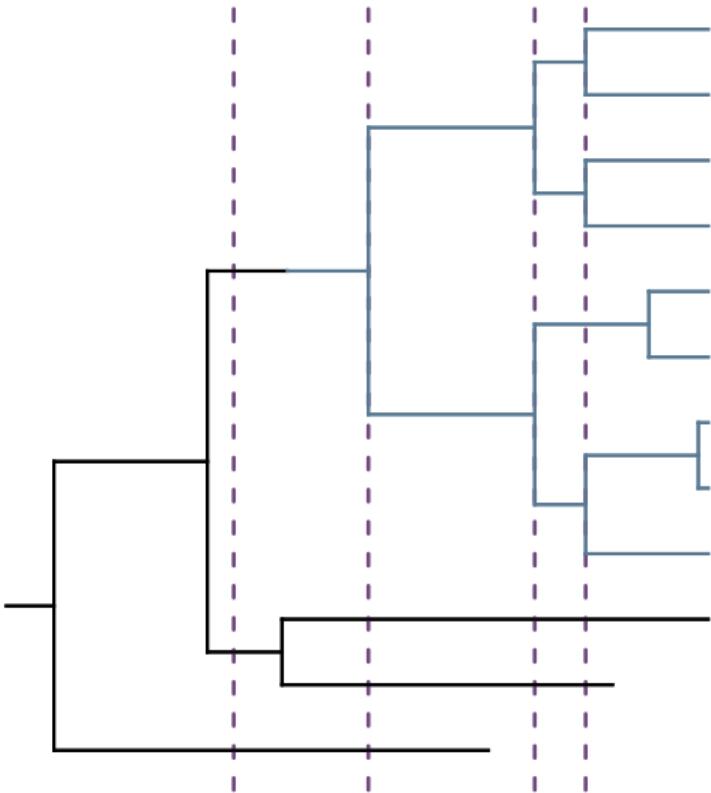
What next?



Ideally, full Bayesian inference of state-dependent burst model and tree

That's a lot of work...

What next?



Ideally, full Bayesian inference of state-dependent burst model and tree

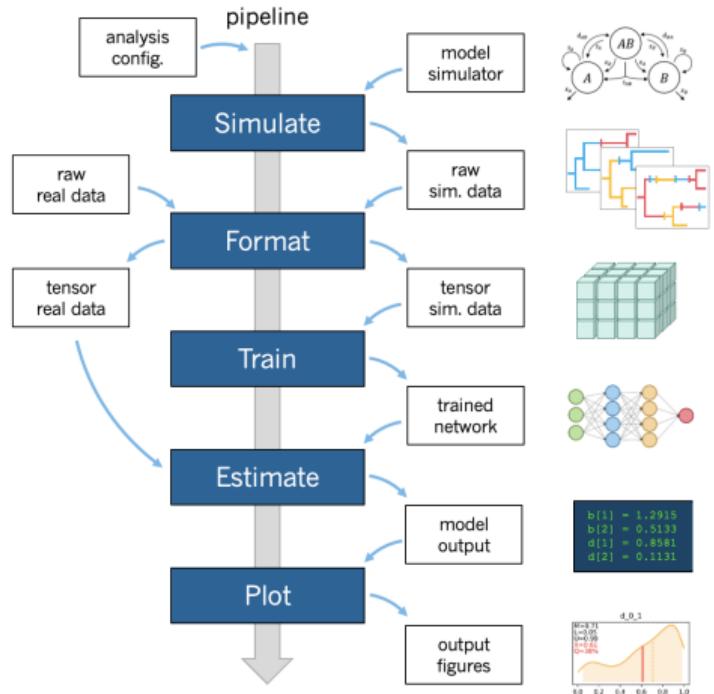
That's a lot of work...

Is it worth it?

- ▶ Is there information in the trees to detect shared divergence rates and state-dependent differences?

phyddle

- ▶ User-friendly deep learning with trees
- ▶ Ammon Thompson & Michael Landis



phyddle analyses

- ▶ Simulated 100k trees (95k training, 5k holdout) with 500 extant tips
 - ▶ Extinct lineages removed
- ▶ Speciation rate, $\lambda \sim \text{Log-uniform}(0.01, 1)$
- ▶ Extinction rate $\sim \text{Log-uniform}(0.01\lambda, \lambda)$
- ▶ Two character states
 - ▶ Transition rate $\sim \text{Uniform}(0.1\lambda, 0.7\lambda)$

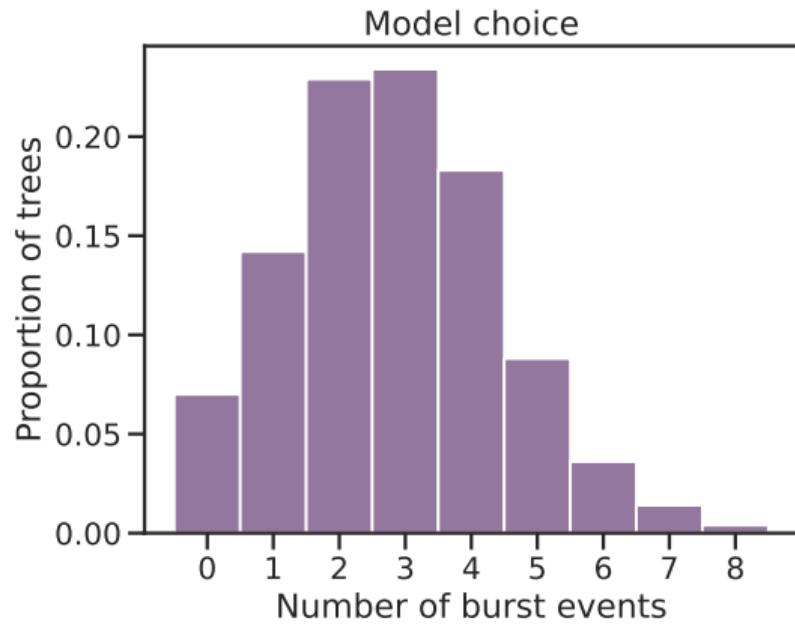
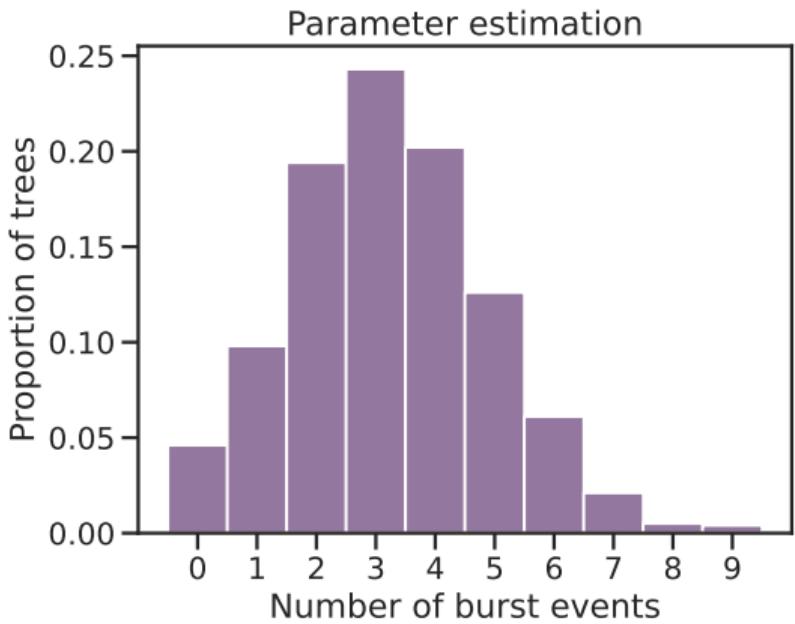
Parameter estimation

- ▶ Burst rate $\sim \text{Uniform}(0.3\lambda, 1.2\lambda)$
- ▶ State 0 burst prob $\sim \text{Uniform}(0.65, 1)$
- ▶ State 1 burst prob $\sim \text{Uniform}(0.1, 0.65)$

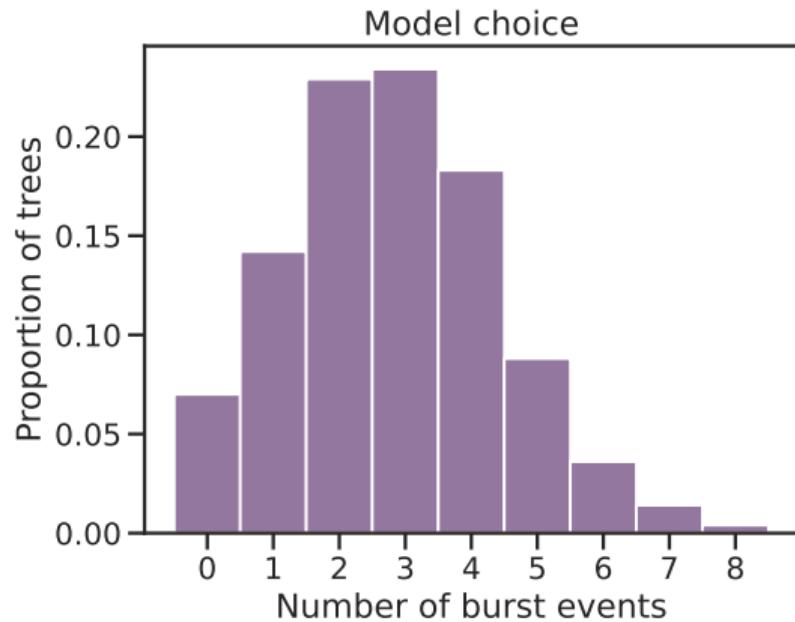
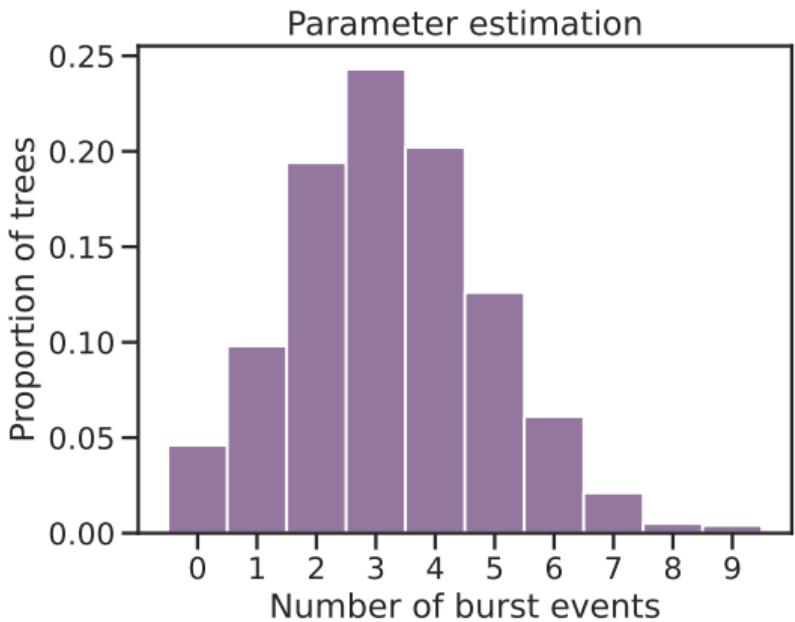
Model choice

- ▶ Burst rate $\sim \text{Uniform}(0.2\lambda, \lambda)$
- ▶ State 0 burst prob $\sim \text{Uniform}(0.6, 1)$
- ▶ State 1 burst prob = State 0
OR $\sim \text{Uniform}(0.0, 0.3)$

Results: Number of burst events

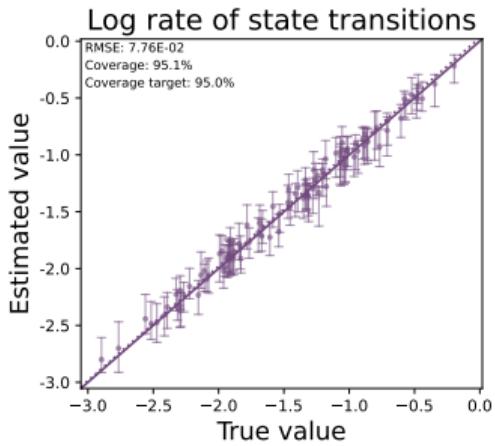
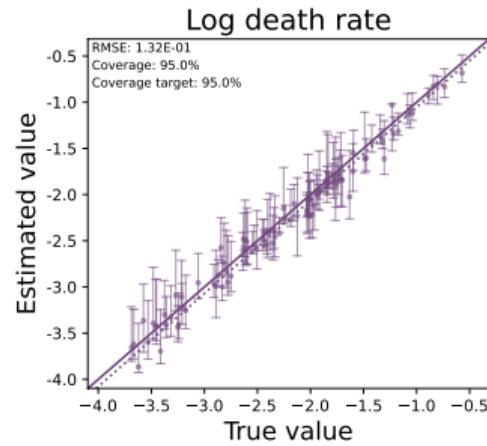
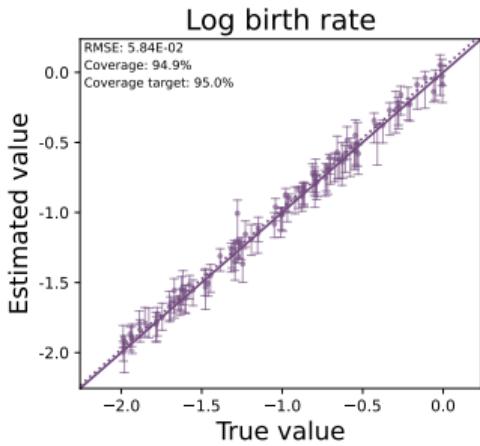


Results: Number of burst events



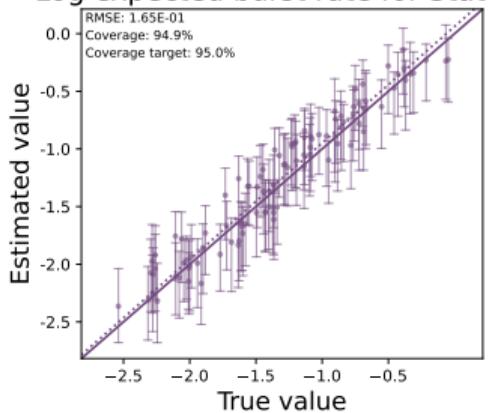
The simulated trees do not have large numbers of burst events to learn from

Results: State-dependent burst *parameter estimation*

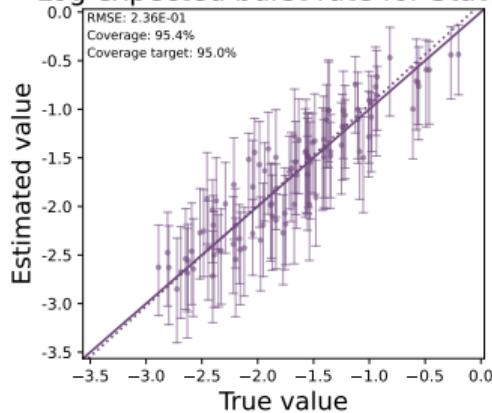


Results: State-dependent burst *parameter estimation*

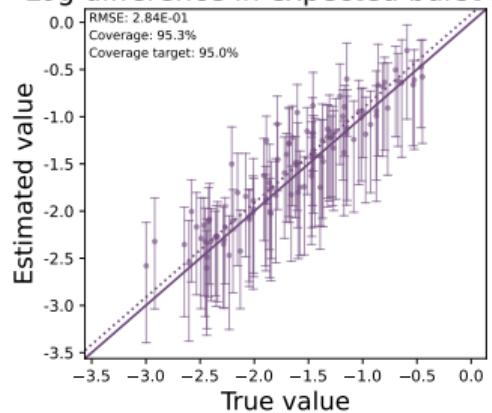
Log expected burst rate for State 0



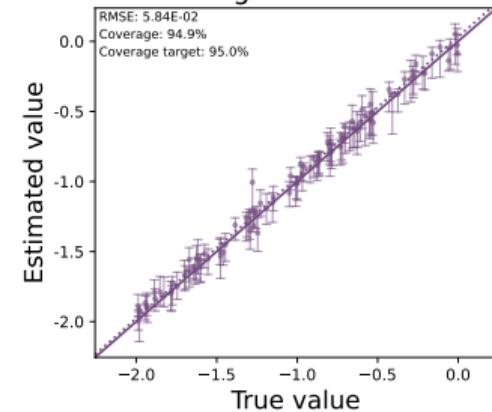
Log expected burst rate for State 1



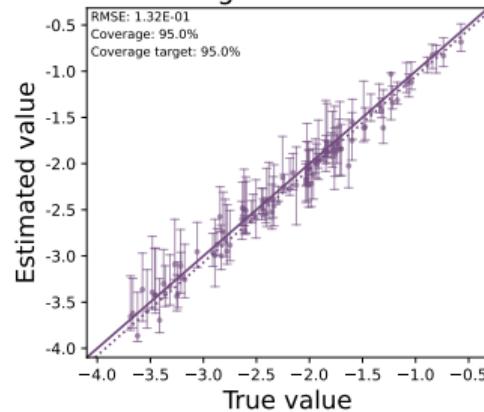
Log difference in expected burst rate



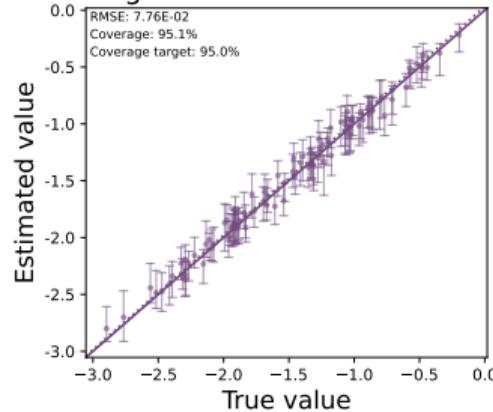
Log birth rate



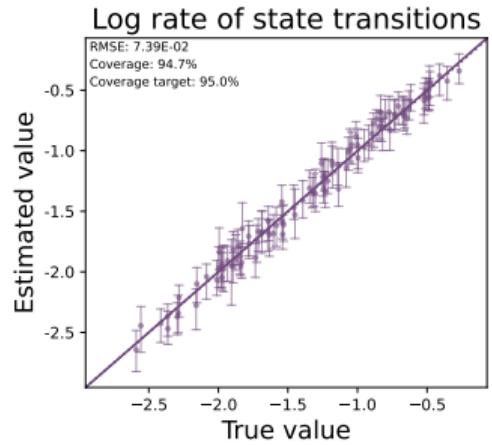
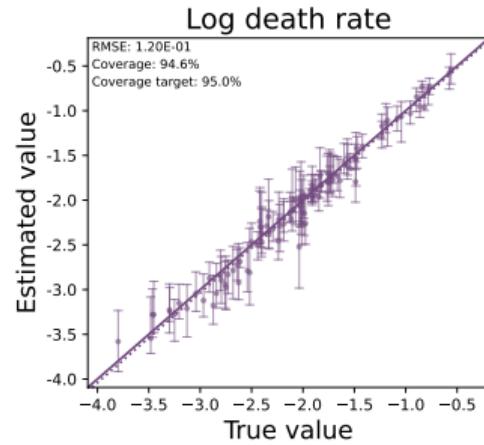
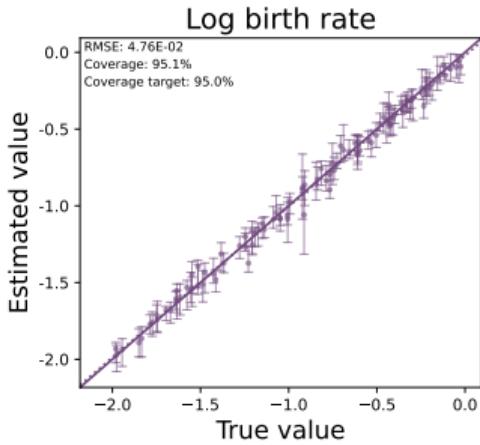
Log death rate



Log rate of state transitions

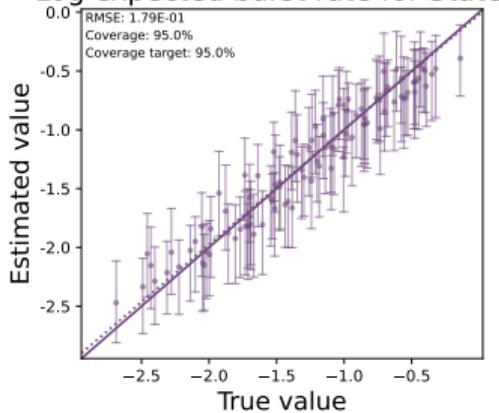


Results: State-dependent burst *model choice*

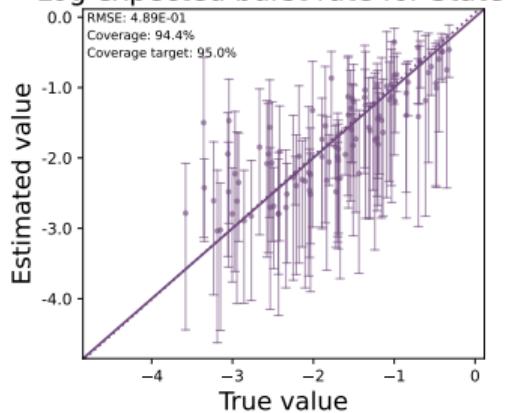


Results: State-dependent burst *model choice*

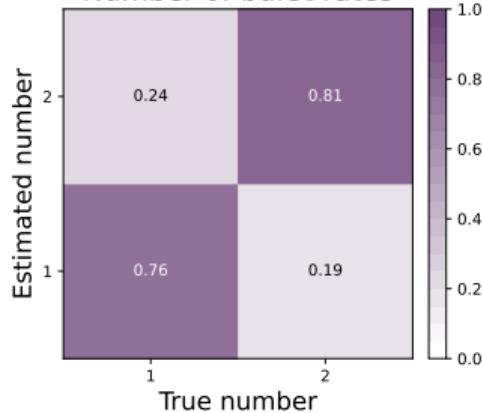
Log expected burst rate for State 0



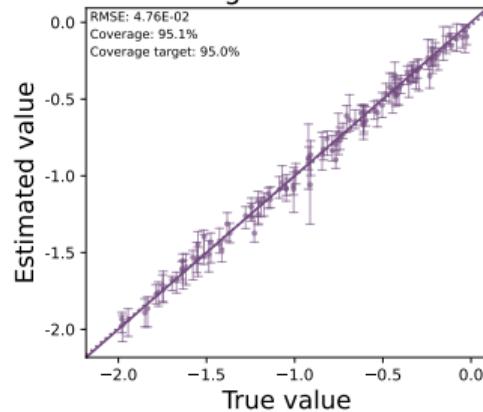
Log expected burst rate for State 1



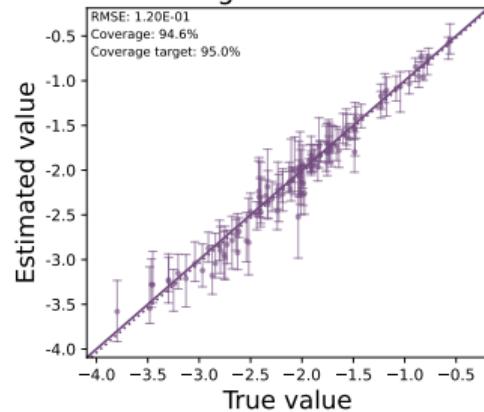
Number of burst rates



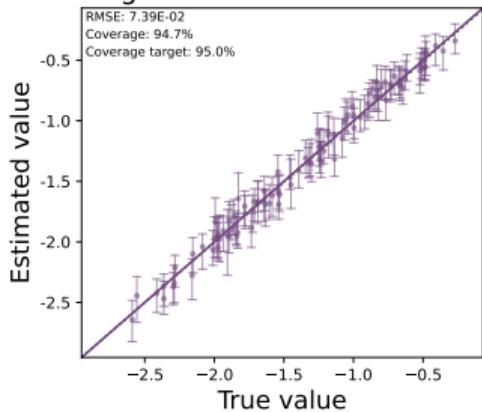
Log birth rate



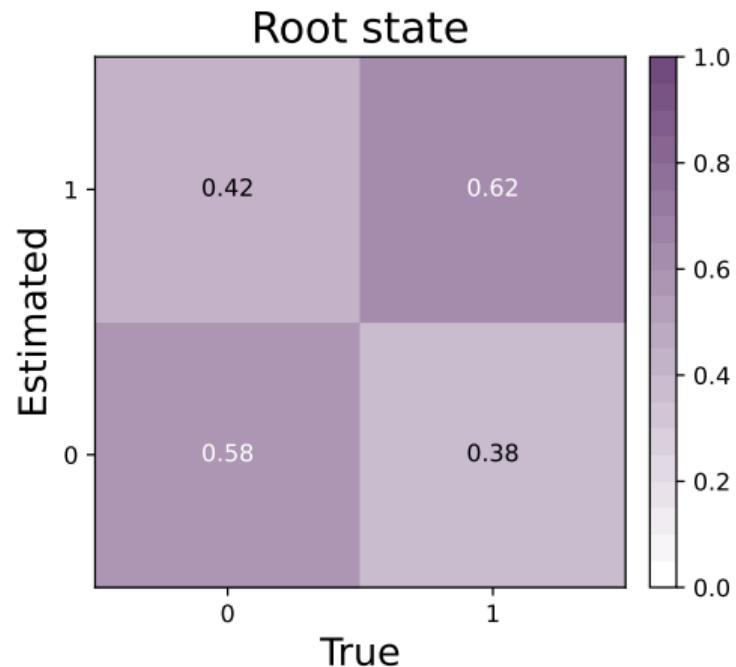
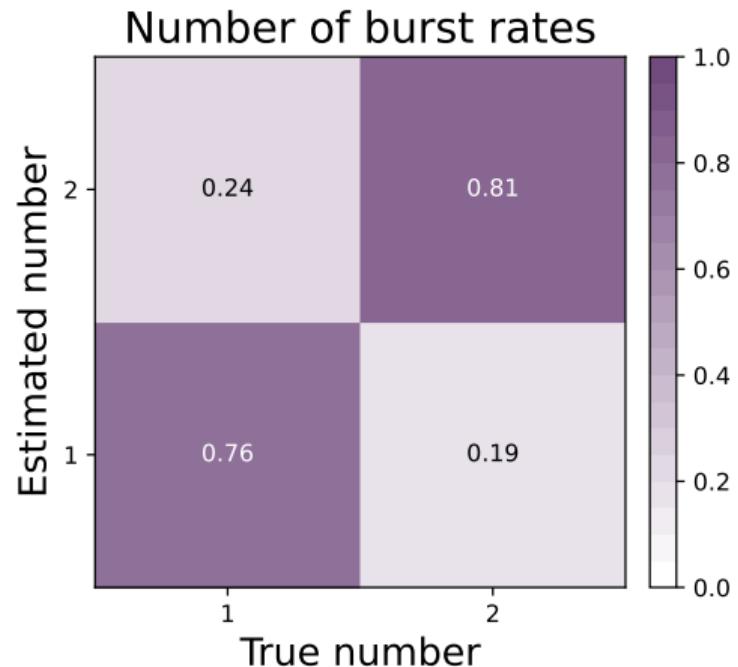
Log death rate



Log rate of state transitions



Results: State-dependent burst *model choice*



*Estimating burst model better than ancestral state***

Take-homes

Trees generated by a BD model with state-dependent burst rates have information about those burst rates

phyddle is a user-friendly tool for quickly experimenting with new (hair-brained) phylogenetic models

Caveats:

- ▶ Analyses assumed no model violations
- ▶ Trees aren't observable!

Next steps

Introduce polytomies into burst events

- ▶ Burst-y processes predict them
 - ▶ *E.g.*, Do rising sea levels always split one island into two?
 - ▶ *E.g.*, Does a carrier always infect one other individual at a social gathering?

Next steps

Introduce polytomies into burst events

- ▶ Burst-y processes predict them
 - ▶ *E.g.*, Do rising sea levels always split one island into two?
 - ▶ *E.g.*, Does a carrier always infect one other individual at a social gathering?

Develop full Bayesian implementation of state-dependent shared divergence models

- ▶ Learn from actual data (sequences) while integrating phylo uncertainty
- ▶ Can make probability statements about events within the tree
 - ▶ The neural network is trained on random trees

Open science: everything is available...

Simulator:



github.com/phyletica/SDSDsim

Open-science notebook:



github.com/phyletica/SDSDsim-phyddle-experiments

Shout-outs

- ▶ [Phyletica Lab](#) (the Phyleticians)
- ▶ Perry Wood, Jr.
- ▶ Ammon Thompson
- ▶ Michael Landis
- ▶ Sebastian Höhna

Computation:

- ▶ Alabama Supercomputer Authority
- ▶ Auburn University Easley Cluster

Photo credits:

- ▶ Perry Wood, Jr.
- ▶ [PhyloPic](#)



MUSEUM OF
NATURAL
HISTORY

Questions?

joaks@auburn.edu

Slides:



phyletica.org/slides/evol2025.pdf



© 2007 Boris Kulikov boris-kulikov.blogspot.com