



Samantha Price
June 21st 2019 SSB Workshop

Common problems in Phylogenetic Comparative Methods (PCM) & how simulations can help



Samantha Price

June 21st 2019 SSB Workshop



Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models

Lam Si Tung Ho¹ and Cécile Ané^{1, 2*}






¹Department of Statistics, University of Wisconsin, 1300 University Ave., Madison, WI, 53706, USA; and ²Department of Botany, University of Wisconsin, Madison, WI, USA

The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters

Wayne P. Maddison , Richard G. FitzJohn

Systematic Biology, Volume 64, Issue 1, January 2015, Pages 127–136, <https://doi.org/10.1093/sysbio/syu070>

Published: 09 September 2014 Article history ▼






 PDF  Split View  Cite  Permissions  Share ▼




Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation

Daniel L. Rabosky , Emma E. Goldberg


Systematic Biology, Volume 64, Issue 2, March 2015, Pages 340–355, <https://doi.org/10.1093/sysbio/syu131>

Published: 19 January 2015 Article history ▼

 PDF  Split View  Cite  Permissions  Share ▼

Special Feature: 5th Anniversary of *Methods in Ecology and Evolution* |  Open Access |  

Shedding light on the ‘dark side’ of phylogenetic comparative methods

Natalie Cooper , Gavin H. Thomas, Richard G. FitzJohn

First published: 13 June 2016 | <https://doi.org/10.1111/2041-210X.12533> | Cited by: 17

 SECTIONS

 PDF  TOOLS  SHARE



 Free Access

IS YOUR PHYLOGENY INFORMATIVE? MEASURING THE POWER OF COMPARATIVE METHODS

Carl Boettiger, Graham Coop, Peter Ralph

First published: 19 February 2012 | <https://doi.org/10.1111/j.1558-5646.2011.01574.x> | Cited by: 95

- Boettiger, C., Coop, G. and Ralph, P., 2012. Is your phylogeny informative? Measuring the power of comparative methods. *Evolution: International Journal of Organic Evolution*, 66(7), pp.2240-2251.
- Ho, L.S.T. and Ané, C., 2013. Asymptotic theory with hierarchical autocorrelation: Ornstein–Uhlenbeck tree models. *The Annals of Statistics*, 41(2), pp.957-981.
- Ho, L.S.T. and Ané, C., 2014. Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*, 5(11), pp.1133-1146.
- Maddison, W.P. and FitzJohn, R.G., 2014. The unsolved challenge to phylogenetic correlation tests for categorical characters. *Systematic biology*, 64(1), pp.127-136.
- Rabosky, D.L. and Goldberg, E.E., 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic biology*, 64(2), pp.340-355.
- Cooper, N., Thomas, G.H. and FitzJohn, R.G., 2016. Shedding light on the ‘dark side’ of phylogenetic comparative methods. *Methods in ecology and evolution*, 7(6), pp.693-699.
- Cooper, N., Thomas, G.H., Venditti, C., Meade, A. and Freckleton, R.P., 2016. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society*, 118(1), pp.64-77.

Five questions to ask

1. What is the uncertainty associated with the estimated parameter values?
2. Do we have the power to distinguish between different evolutionary models?
3. Can we correctly reject the null hypothesis?
 - What is the correct null hypothesis?
4. Do our models adequately reflect biological reality?
5. Do our methods adequately control for phylogenetic pseudo-replication?

What can cause issues? Interactions between:

- Number of taxa in dataset
- Number of parameters estimated
- Shape of the phylogeny
- Distribution of data upon phylogeny
- **Stochasticity in the process:** Brownian motion, birth-death, Continuous time-Markov models
- Phylogenetic error

And others

1. What is the uncertainty associated with the estimated parameter values?

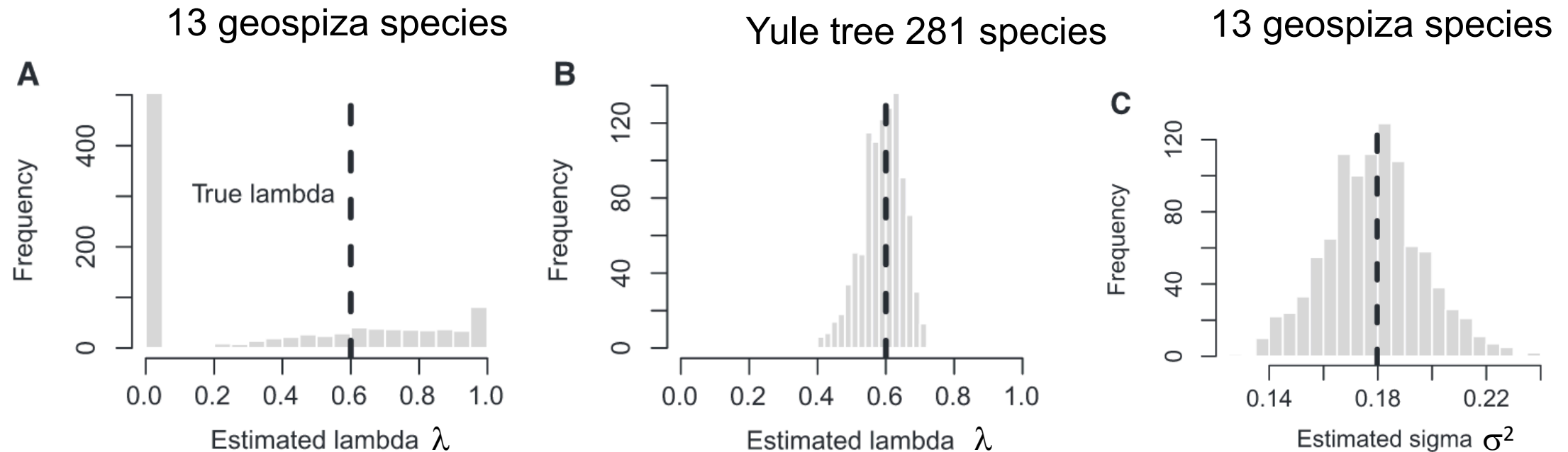
e.g. Simulate data with $\lambda = 0.6$ and $\sigma^2 = 0.18$

‘Phylogenetic signal’ all internal branches shortened by multiplicative factor 0 = star phylogeny and 1 = regular phylogeny = Brownian motion

Brownian motion rate parameter

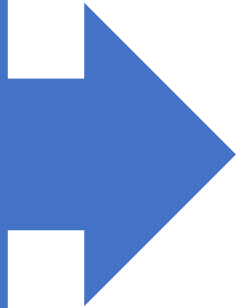
1. What is the uncertainty associated with the estimated parameter values?

e.g. Simulate data with $\lambda = 0.6$ and $\sigma^2 = 0.18 \rightarrow$ estimate λ & σ^2

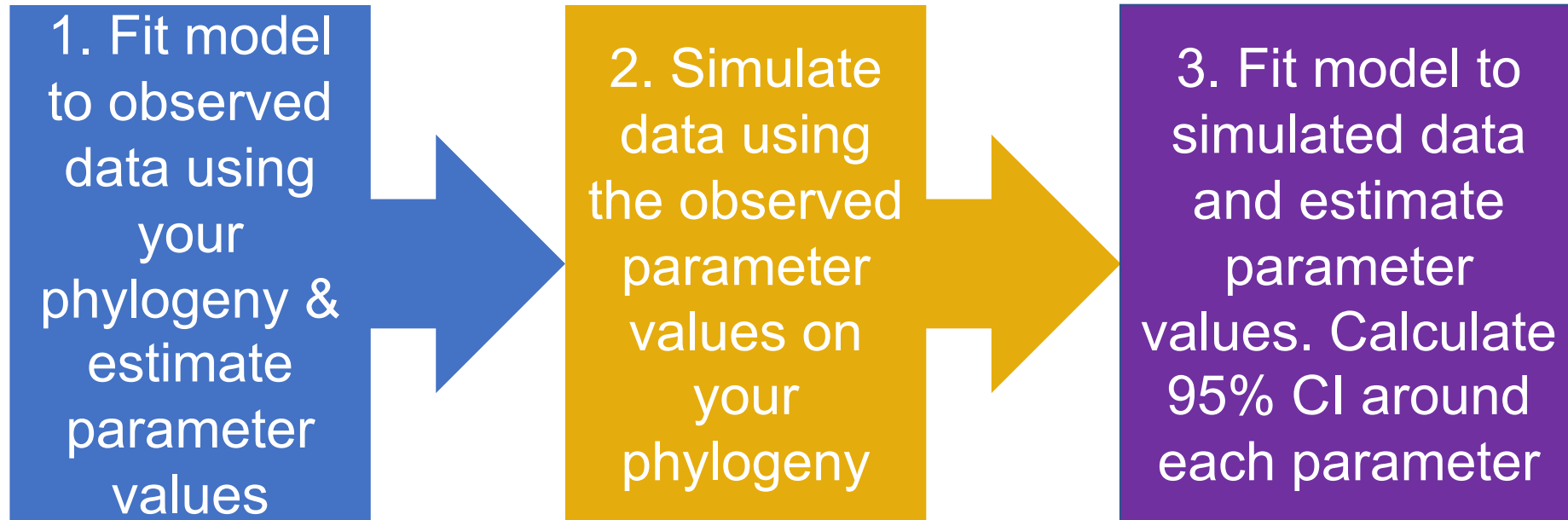


1. What is the uncertainty associated with the estimated parameter values?

1. Fit model
to observed
data using
your
phylogeny &
estimate
parameter
values

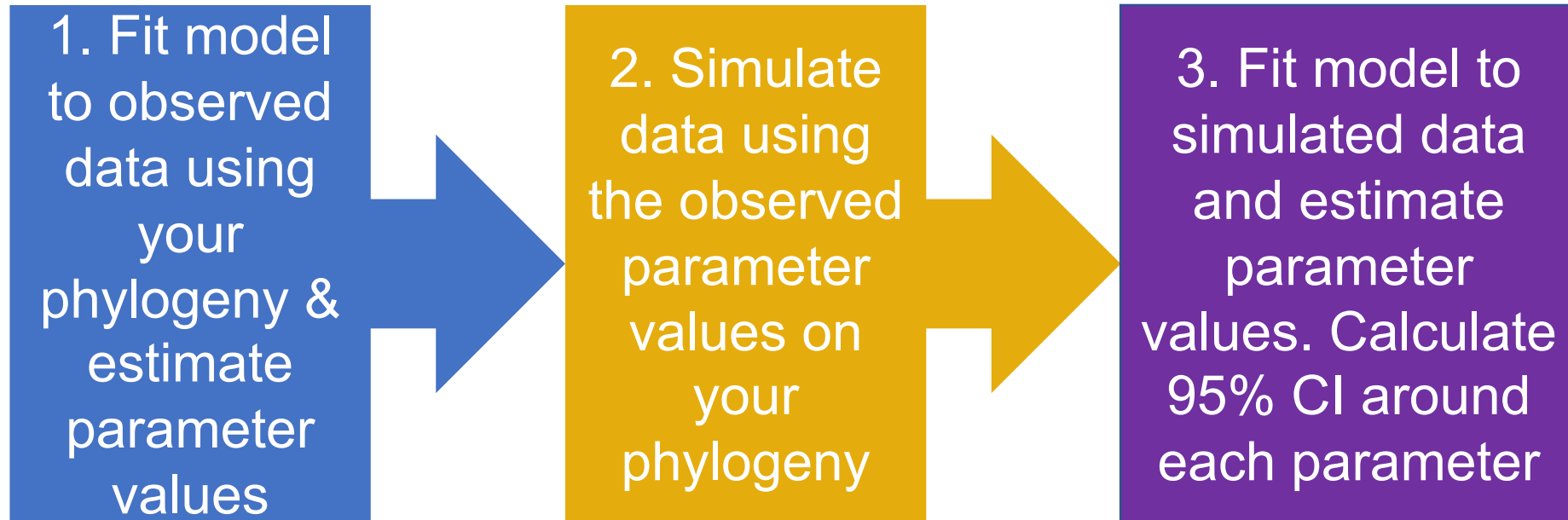


1. What is the uncertainty associated with the estimated parameter values?



Parametric bootstrapping

1. What is the uncertainty associated with the estimated parameter values?



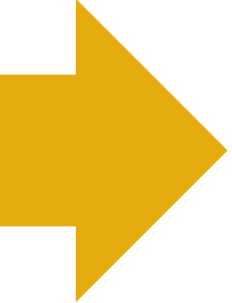
Parametric bootstrapping

OUwie package (BM/OU models with 1 or more regime) has prebuilt parametric bootstrapping function (ouwie.boot)



1. What is the uncertainty associated with the estimated parameter values?

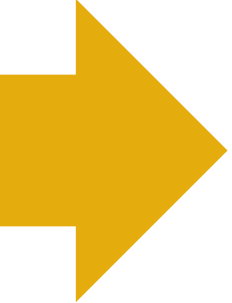
2. Simulate data using the observed parameter values on your phylogeny



For many scenarios and PCMs you don't have to write the simulations from scratch, you can easily do it using pre-existing simulation functions in R.

1. What is the uncertainty associated with the estimated parameter values?

2. Simulate data using the observed parameter values on your phylogeny



Complete list of simulations available:
R phylogenetics task view:

<https://cran.r-project.org/web/views/Phylogenetics.html>

- **Simulating under simple BM**

- Geiger: sim.char
- Phytools: fastBM
- ape: rTraitCont

- **Simulating under simple OU**

- OUwie: OUwie.sim
- ape: rTraitCont

- **Simulating under multivariate BM**

- Geiger: sim.char
- ape: rTraitMult

- **Simulating under multiple rate or peaks**

- OUwie: OUwie.sim
- Phytools: sim.rates (BM) multiOU (OU)

- **Simulating discrete traits**

- Geiger: sim.char
- ape: rTraitDisc

- **Simulating under accelerating or declining rates**

One way to do this is to simulate under BM on a tree that has been transformed to have longer branches towards the tips or root, which can be done in Geiger using the rescale function and the early burst or delta model.

- **Simulating phylogenies**

- Basic birth-death models

- ape: rlineage, rbdtree & rphylo
- geiger: sim.bdtree (if death rate >0 then will contain extinct tips – can be removed using drop.extinct)

- Mass extinction

- TESS: tess.sim.age, tess.sim.taxa, tess.sim.taxa.age

- State-dependent models

- Diversitree: tree.bisse, tree.musse, tree.classe, tree.geosse

- Time-dependent models

- TESS: tess.sim.age, tess.sim.taxa, tess.sim.taxa.age
- ape: rlineage, rbdtree & rphylo

1. What is the uncertainty associated with the estimated parameter values?



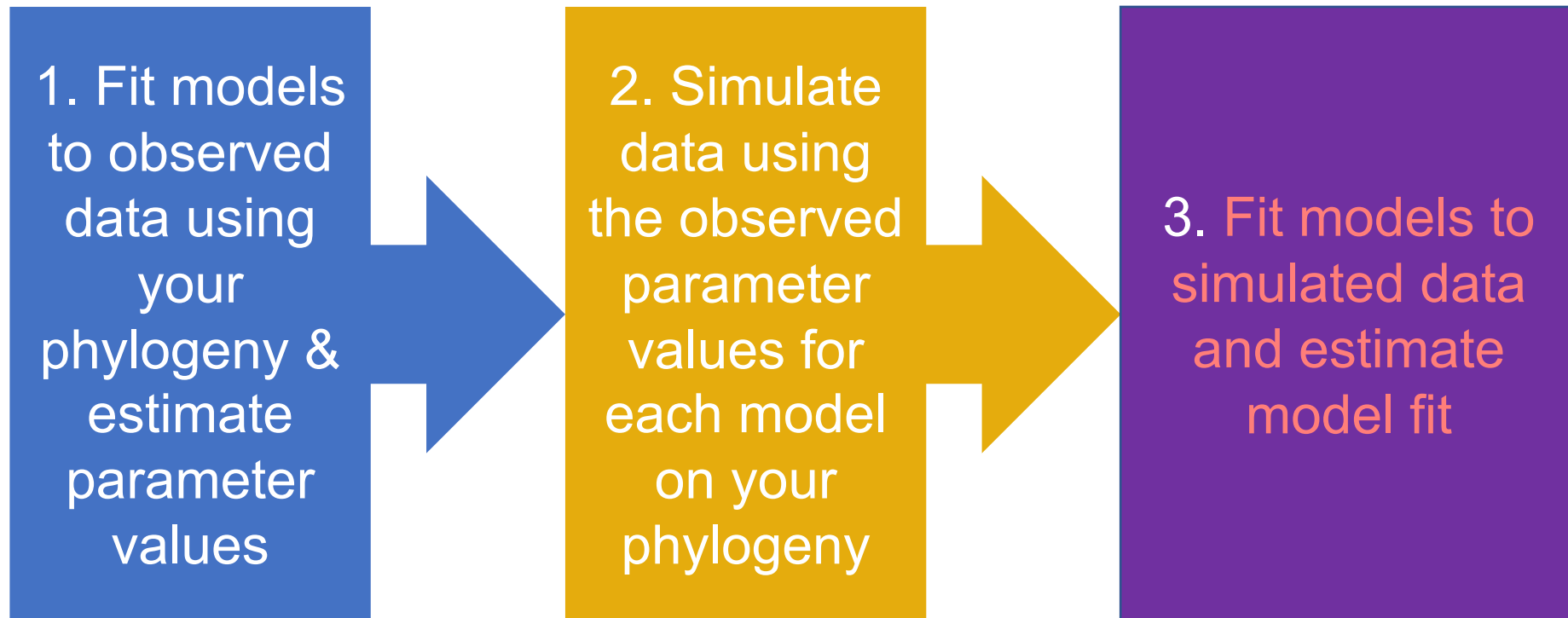
simulationexercise_SSBworkshop.R

- 1) Example simulation to estimate confidence intervals around Pagel's λ**
- 2) Example using ouwie.boot to estimate confidence intervals around OU multi-peak parameters**

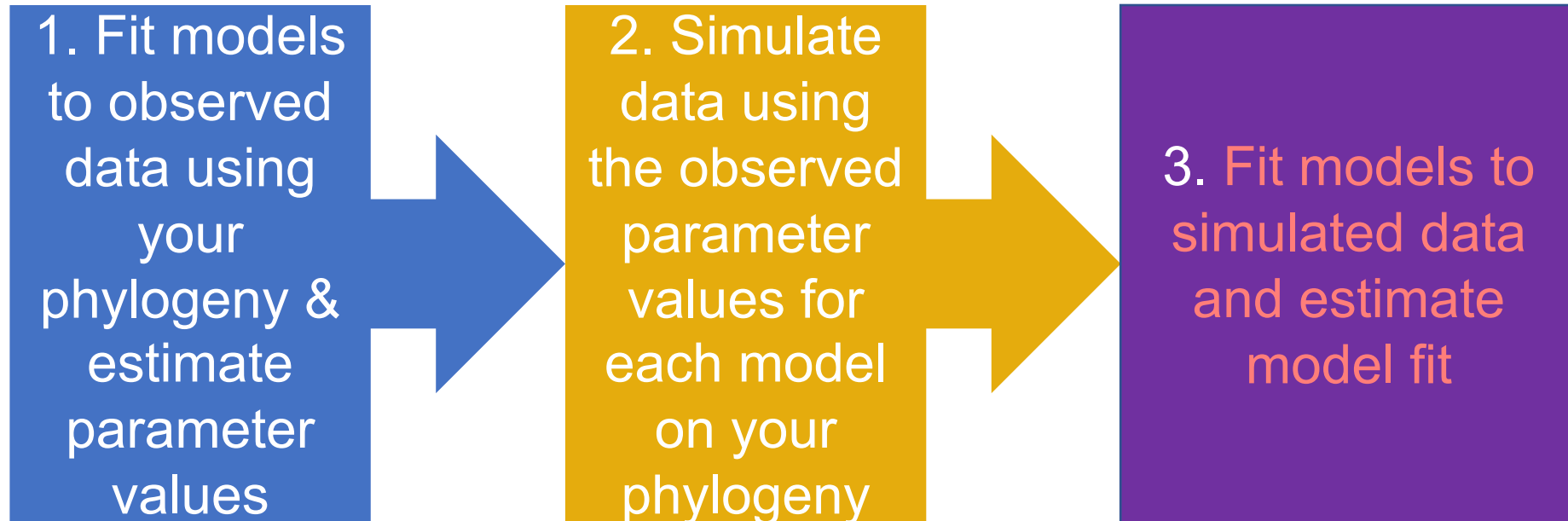
Five questions to ask

1. What is the uncertainty associated with the estimated parameter values?
2. Do we have the power to distinguish between different evolutionary models?
3. Can we correctly reject the null hypothesis?
 - What is the correct null hypothesis?
4. Do our models adequately reflect biological reality?
5. Do our methods adequately control for phylogenetic pseudo-replication?

2. Do we have the power to distinguish between different evolutionary models?

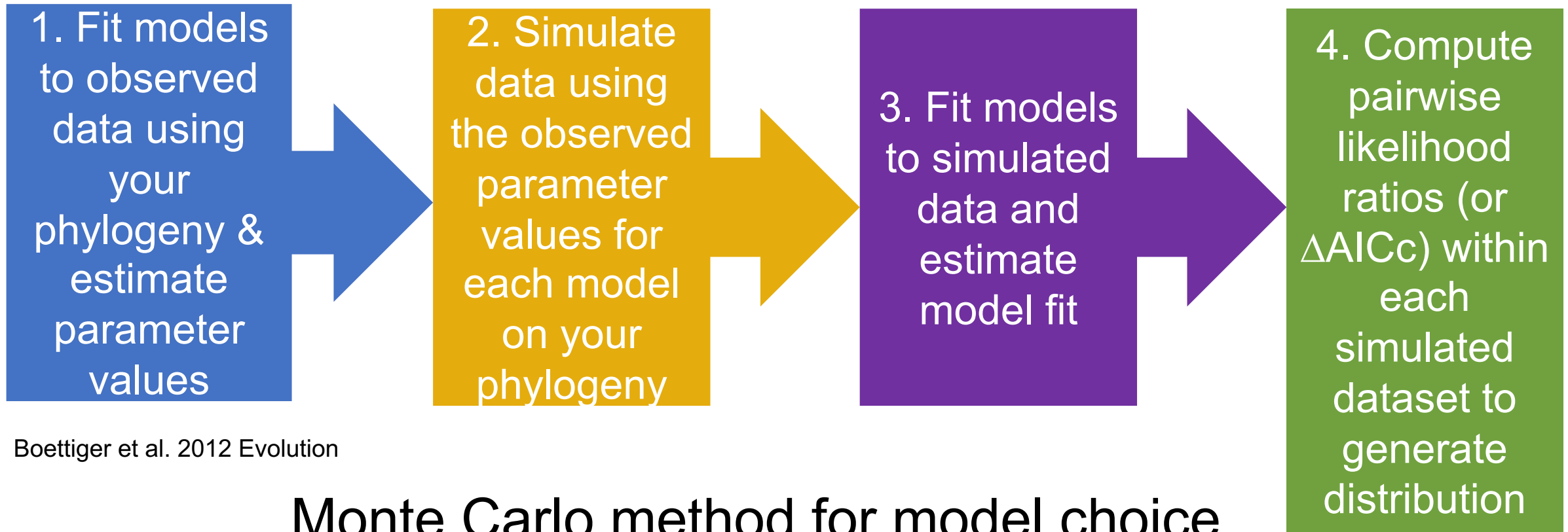


2. Do we have the power to distinguish between different evolutionary models?



Ideally 100% (or at least 95%) of the time the model you simulated under is the best-fitting model

2. Do we have the power to distinguish between different evolutionary models?



Boettiger et al. 2012 Evolution

Monte Carlo method for model choice

2. Do we have the power to distinguish between different evolutionary models?

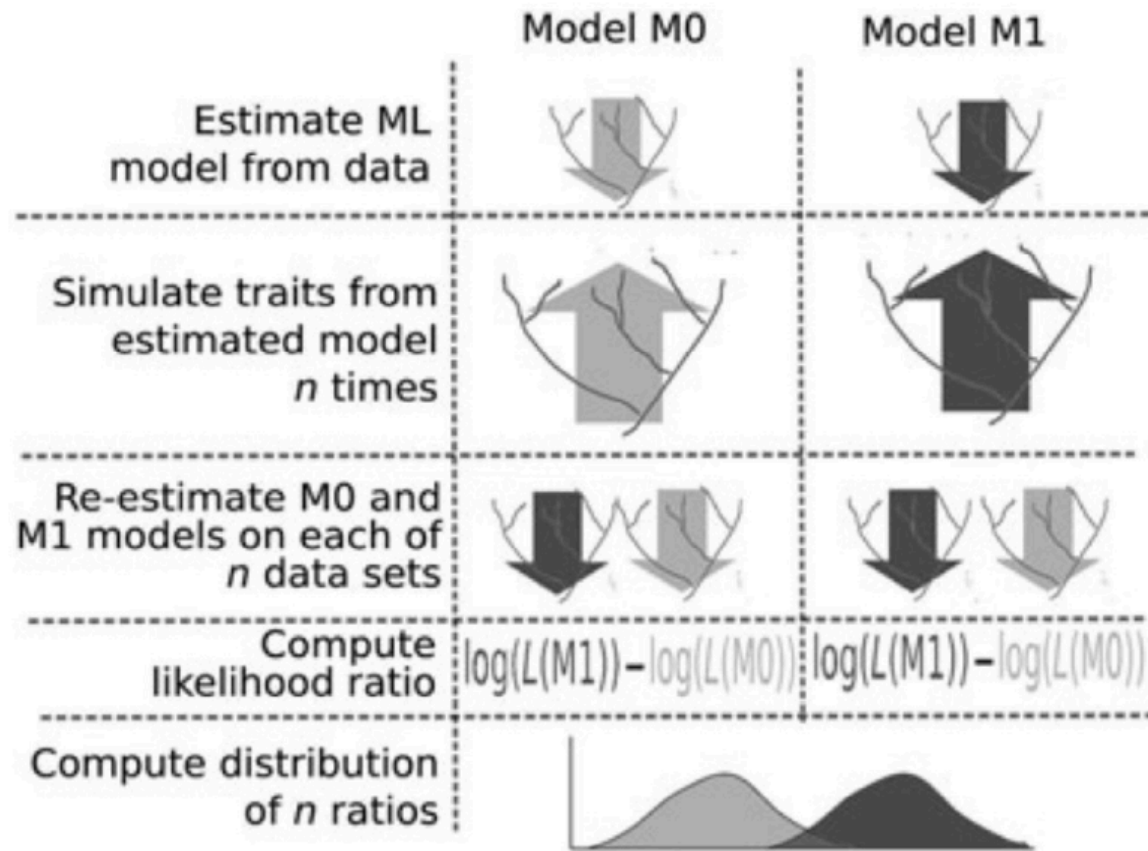


Figure 2. Conceptual diagram of the Monte Carlo method for model choice. First, parameters for both models are estimated from the original data. Then, n simulated datasets are created from each model at these parameters, and on each dataset, the parameters for both models are reestimated and the likelihood ratio statistic is computed. The collection of likelihood ratio statistics generates the corresponding distribution. This involves a process of $4n$ fits by maximum likelihood, instead of only two fits required for information criteria.

2. Do we have the power to distinguish between different evolutionary models?

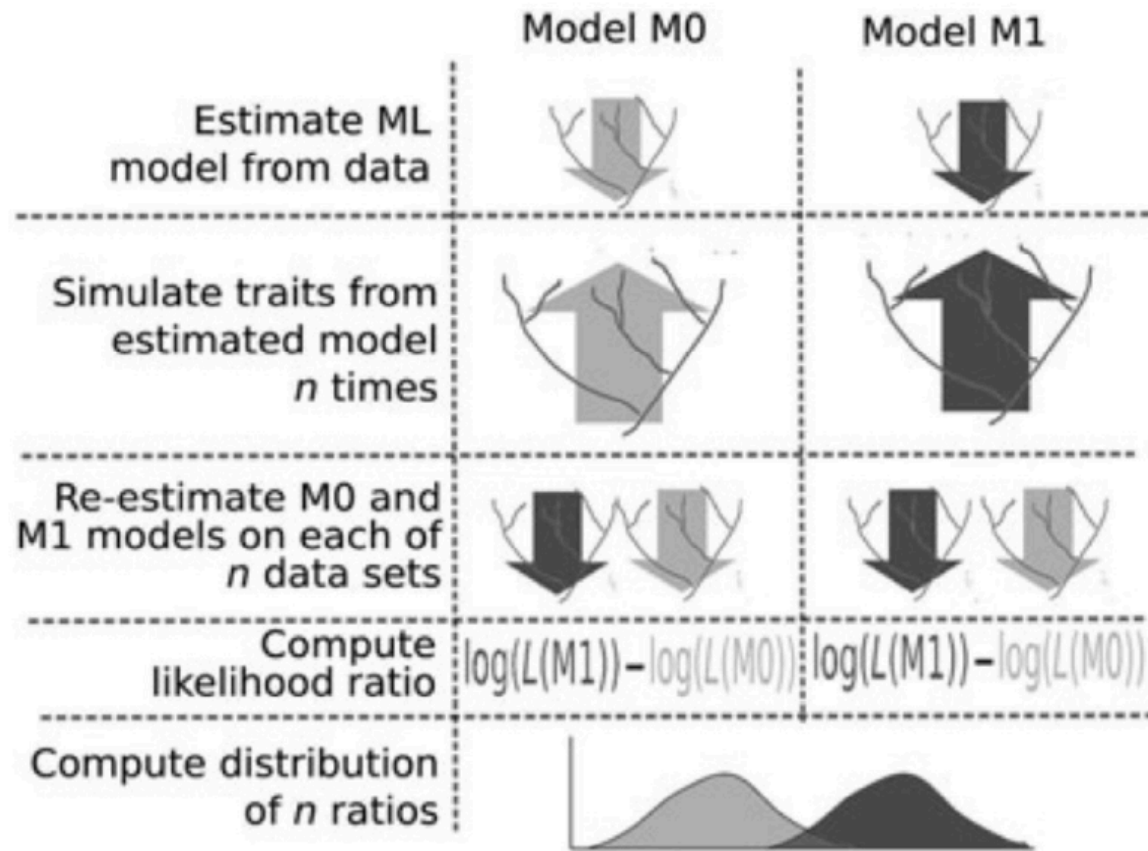
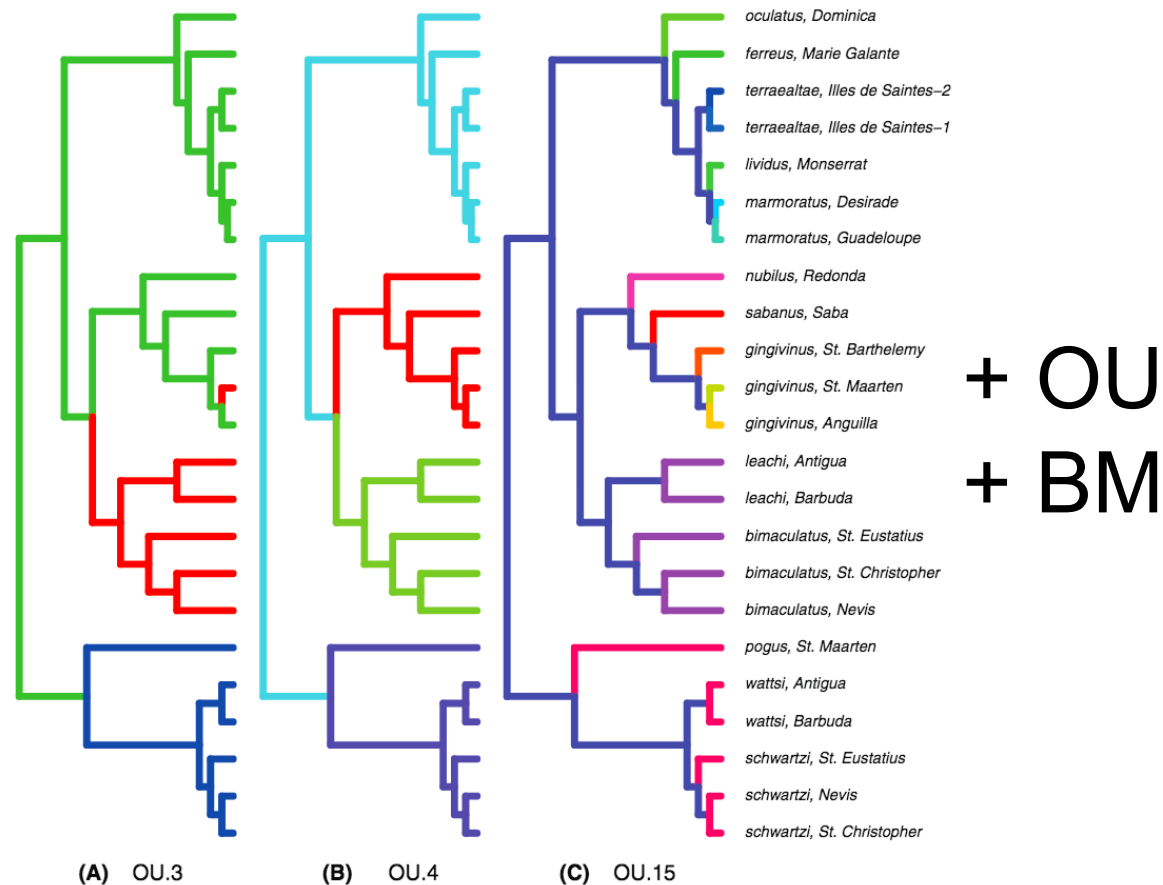


Figure 2. Conceptual diagram of the Monte Carlo method for model choice. First, parameters for both models are estimated from the original data. Then, n simulated datasets are created from each model at these parameters, and on each dataset, the parameters for both models are reestimated and the likelihood ratio statistic is computed. The collection of likelihood ratio statistics generates the corresponding distribution. This involves a process of $4n$ fits by maximum likelihood, instead of only two fits required for information criteria.

R package `pmc` can use it for Monte Carlo simulation of Geiger and ouch models

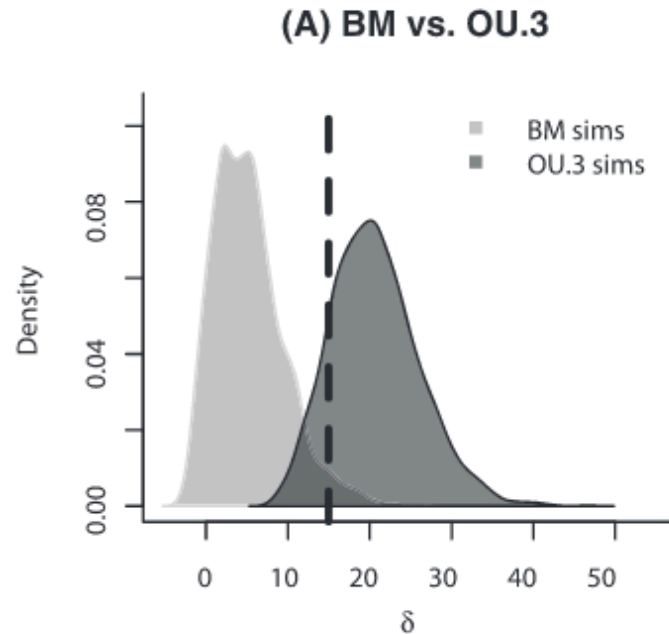
2. Do we have the power to distinguish between different evolutionary models?



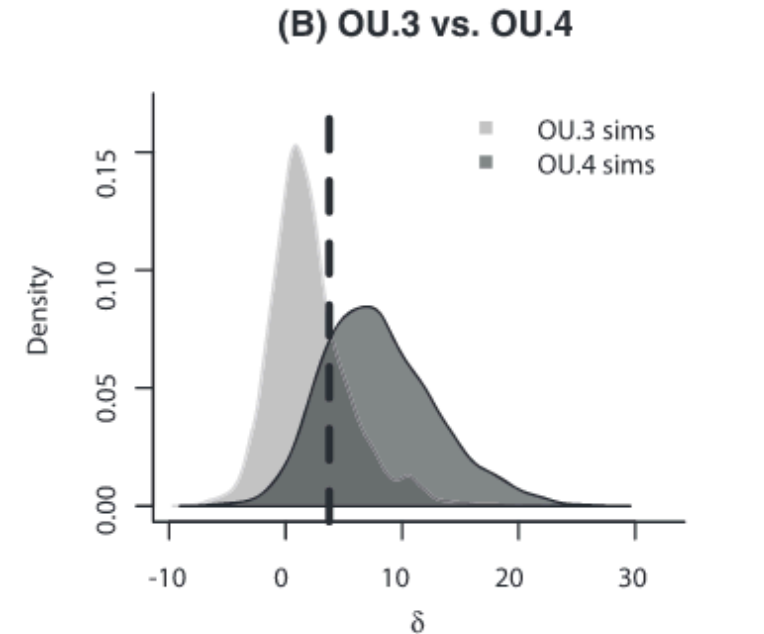
Fitted 5 models,
estimated the
parameter values
→ simulated under
all of them → fitted
all models →
estimated pairwise
likelihood ratio

2. Do we have the power to distinguish between different evolutionary models?

Fitted 5 models,
estimated the
parameter values
→ simulated under
all of them → fitted
all models →
estimated pairwise
likelihood ratio



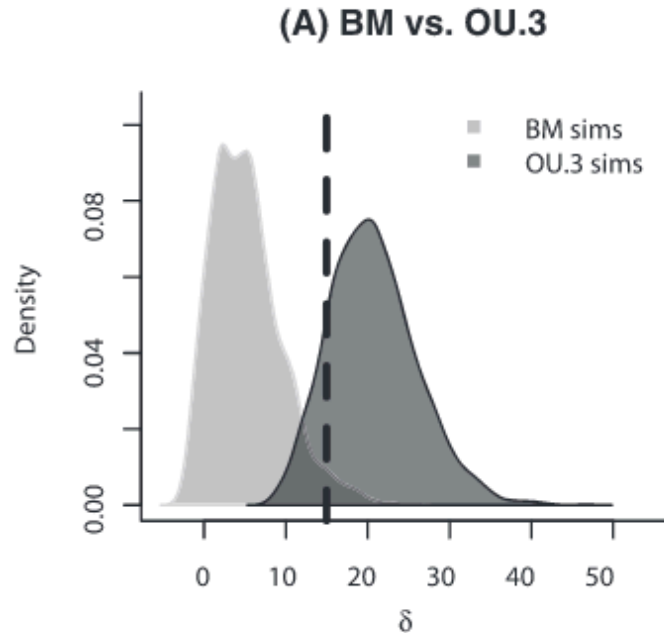
$\text{Log}(L(\text{OU3})) - \text{Log}(L(\text{BM}))$



$\text{Log}(L(\text{OU4})) - \text{Log}(L(\text{OU3}))$

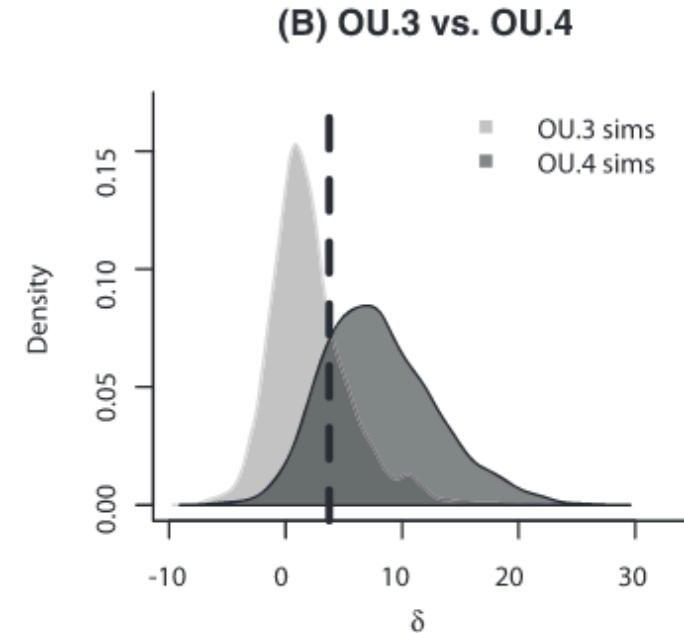
2. Do we have the power to distinguish between different evolutionary models?

Fitted 5 models,
estimated the
parameter values
→ simulated under
all of them → fitted
all models →
estimated pairwise
likelihood ratio



→ $\text{Log}(L(\text{OU3})) - \text{Log}(L(\text{BM}))$

Observed likelihood ratio is
much more likely under
OU.3. = we have the power!



→ $\text{Log}(L(\text{OU4})) - \text{Log}(L(\text{OU3}))$

Observed likelihood
ratio is ~equally likely =
we don't have the power

2. Do we have the power to distinguish between different evolutionary models?



simulationexercise_SSBworkshop.R

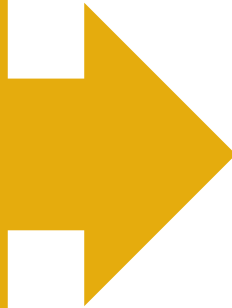
- 1) Example simulation to estimate confidence intervals around Pagel's λ
- 2) Example using ouwie.boot to estimate confidence intervals around OU multi-peak parameters
- 3) **Example simulation to estimate power and uncertainty in discrete trait model fitting**

Five questions to ask

1. What is the uncertainty associated with the estimated parameter values?
2. Do we have the power to distinguish between different evolutionary models?
3. Can we correctly reject the null hypothesis?
 - What is the correct null hypothesis?
4. Do our models adequately reflect biological reality?
5. Do our methods adequately control for phylogenetic pseudo-replication?

3. Can we correctly reject the null hypothesis?

1. Simulate data under null model

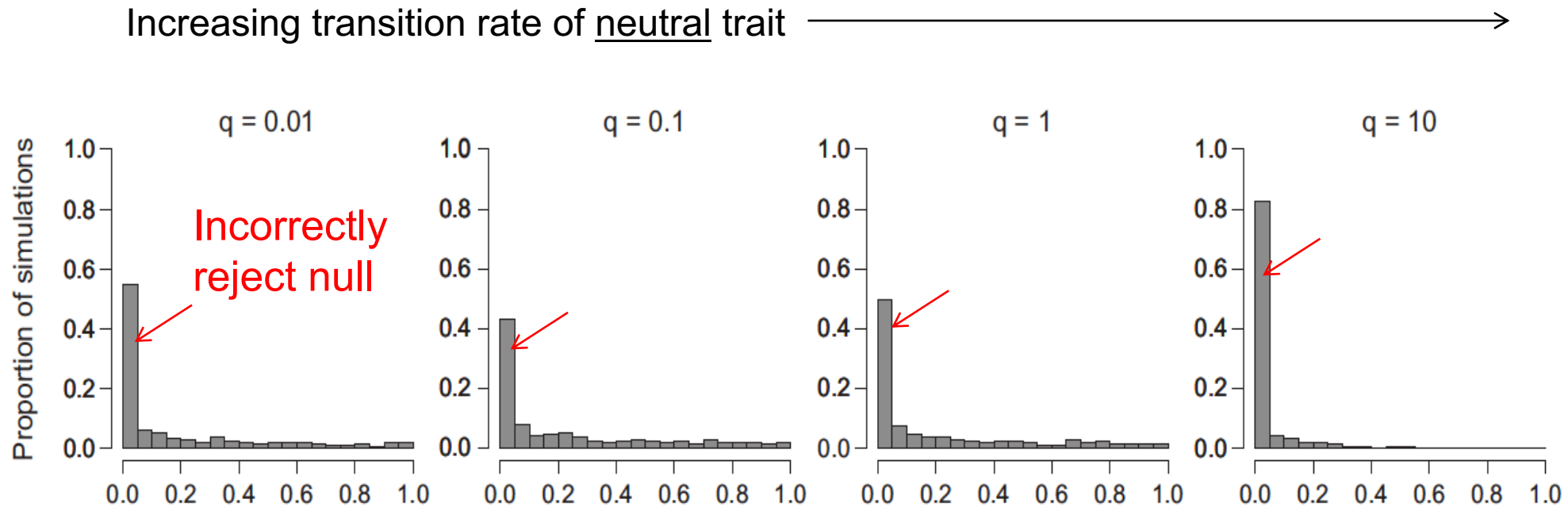


2. Fit null and alternative models to simulated data and estimate test statistic/ model fit

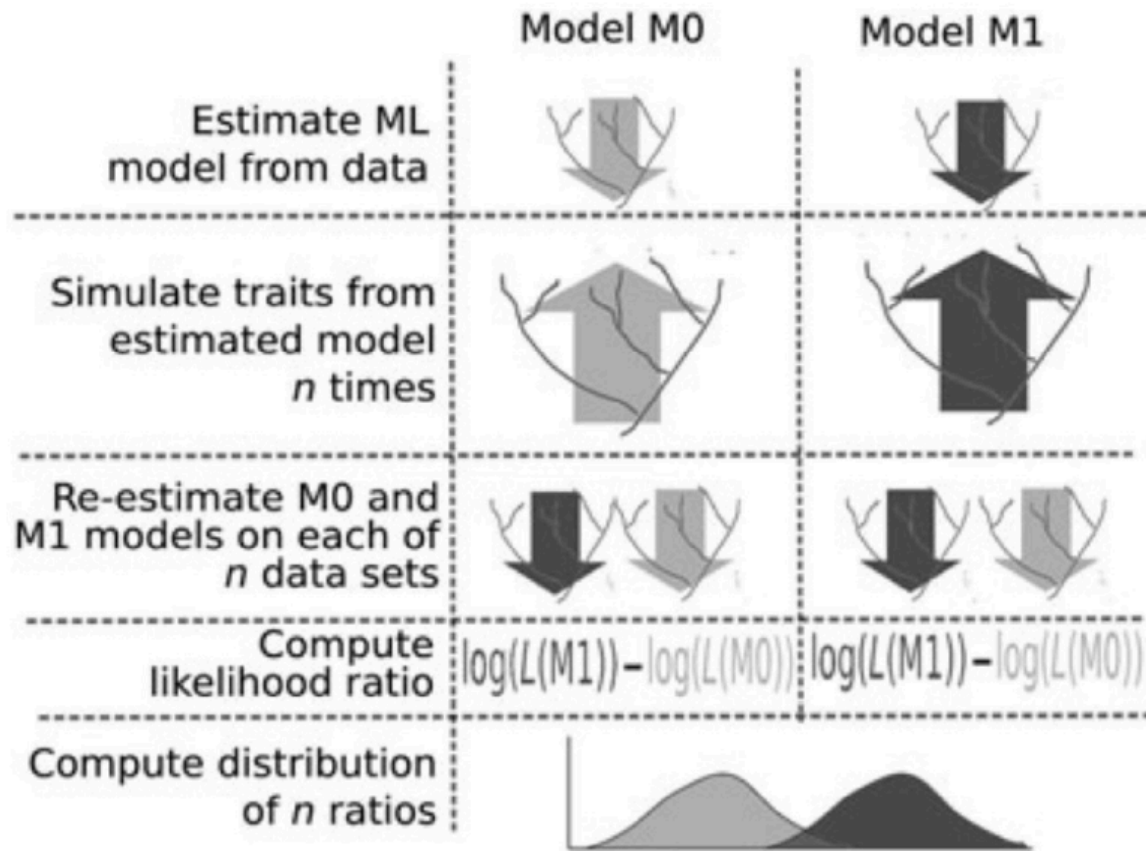
Ideally 100% of the time (or at least 95%) you don't reject the null hypothesis

3. Can we correctly reject the null hypothesis?

BiSSE Null simulations of independent binary trait evolution and lineage diversification



3. Can we correctly reject the null hypothesis?



One general solution is simulation

→ Phylometrics R package for *SSE type questions (Hua & Bronham 2016)

3. Can we correctly reject the null hypothesis? What is the correct null model?

When *SSE models were implemented the fit of simple rate (null) models were compared to models that allowed rate variation dependent on trait of interest. The correct null should have been to rate variation unrelated to trait of interest (Rabosky & Goldberg, 2015).

→ Hidden-rate SSE models (Beaulieu & O'Meara 2016 hisse R package)

3. Can we correctly reject the null hypothesis? **What is the correct null model?**

When *SSE models were implemented the fit of simple rate (null) models were compared to models that allowed rate variation dependent on trait of interest. The correct null should have been to rate variation unrelated to trait of interest (Rabosky & Goldberg, 2015).

→ Hidden-rate SSE models (Beaulieu & O'Meara 2016 hisse R package)

Are the models adequate? Can they explain the observed data?

3. Can we correctly reject the null hypothesis? **What is the correct null model?**

Not just *SSE models – the correct null model to determine relationship between a discrete character and rates of continuous-character evolution should be rate variation unrelated to trait of interest!

May & Moore 2019 *bioRxiv*, Venditti & Pagel 2011

↓
MuSSCRat in RevBayes

↓
BayesTraits v3

Are the models adequate? Can they explain the observed data?

Five questions to ask

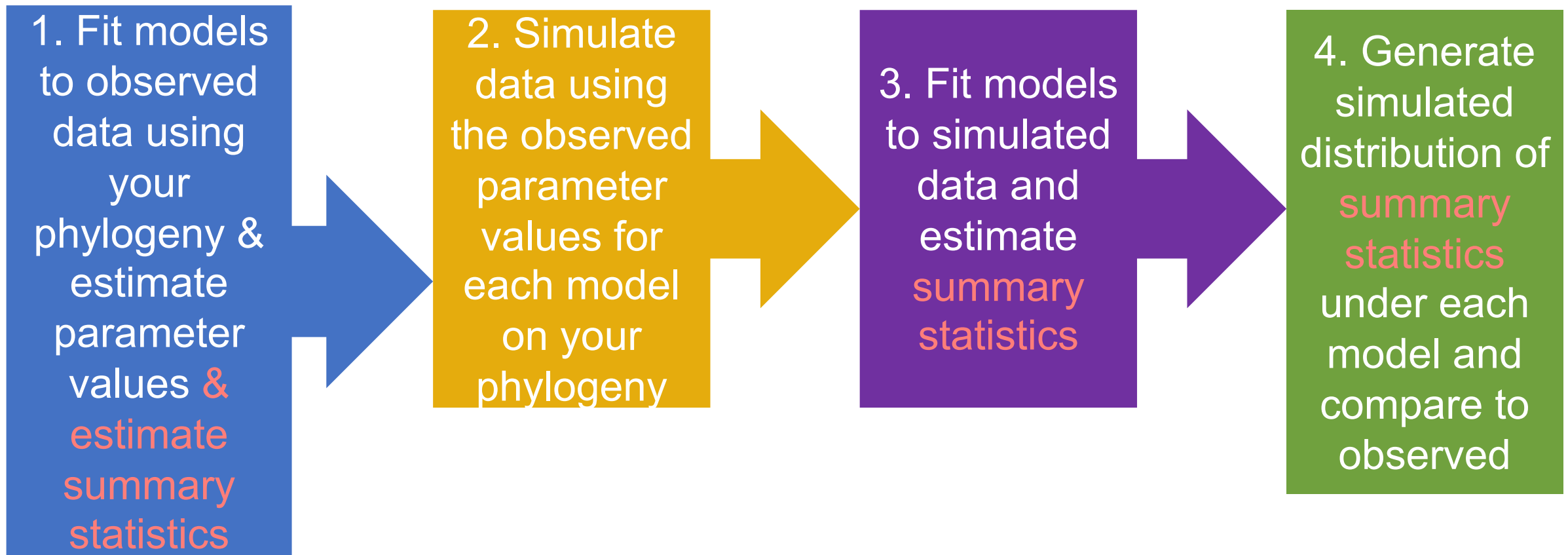
1. What is the uncertainty associated with the estimated parameter values?
2. Do we have the power to distinguish between different evolutionary models?
3. Can we correctly reject the null hypothesis?
 - What is the correct null hypothesis?
4. Do our models adequately reflect biological reality?
5. Do our methods adequately control for phylogenetic pseudo-replication?

4. Do our models adequately reflect biological reality?

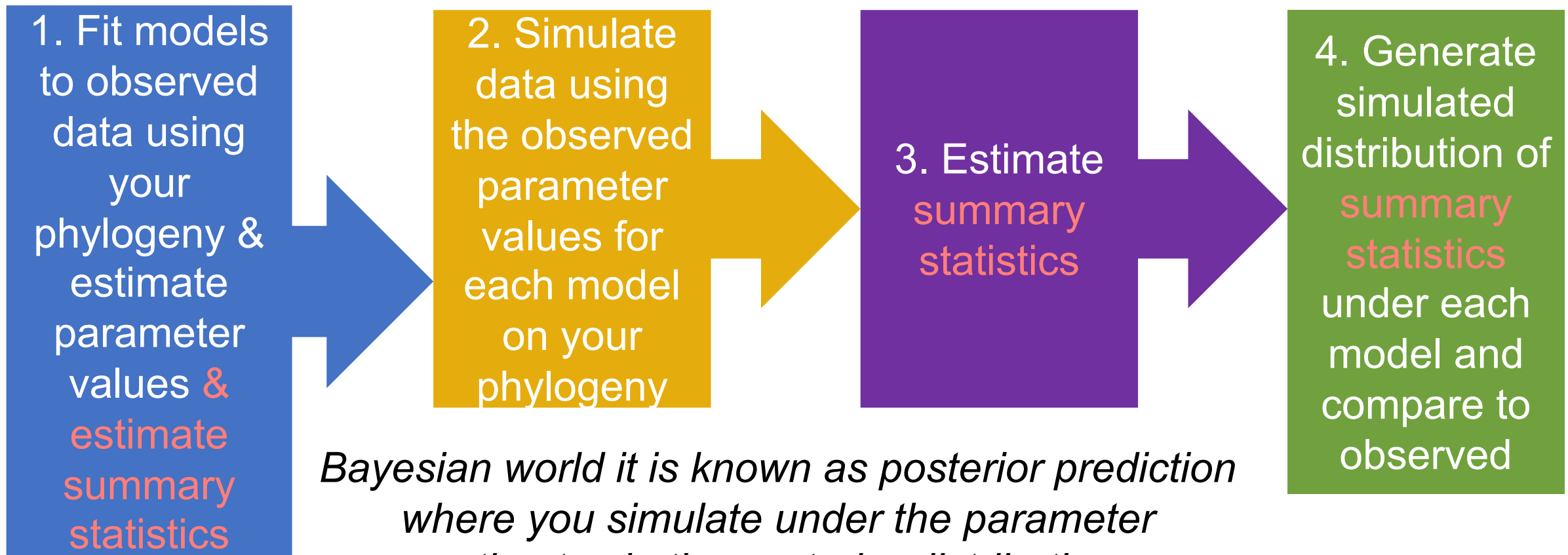
Just because one or two models fit the data better than the others in the model set it doesn't mean they approximate* reality!

* Evolution is complex so the model is always wrong but it can be a useful approximation

4. Do our models adequately reflect biological reality?

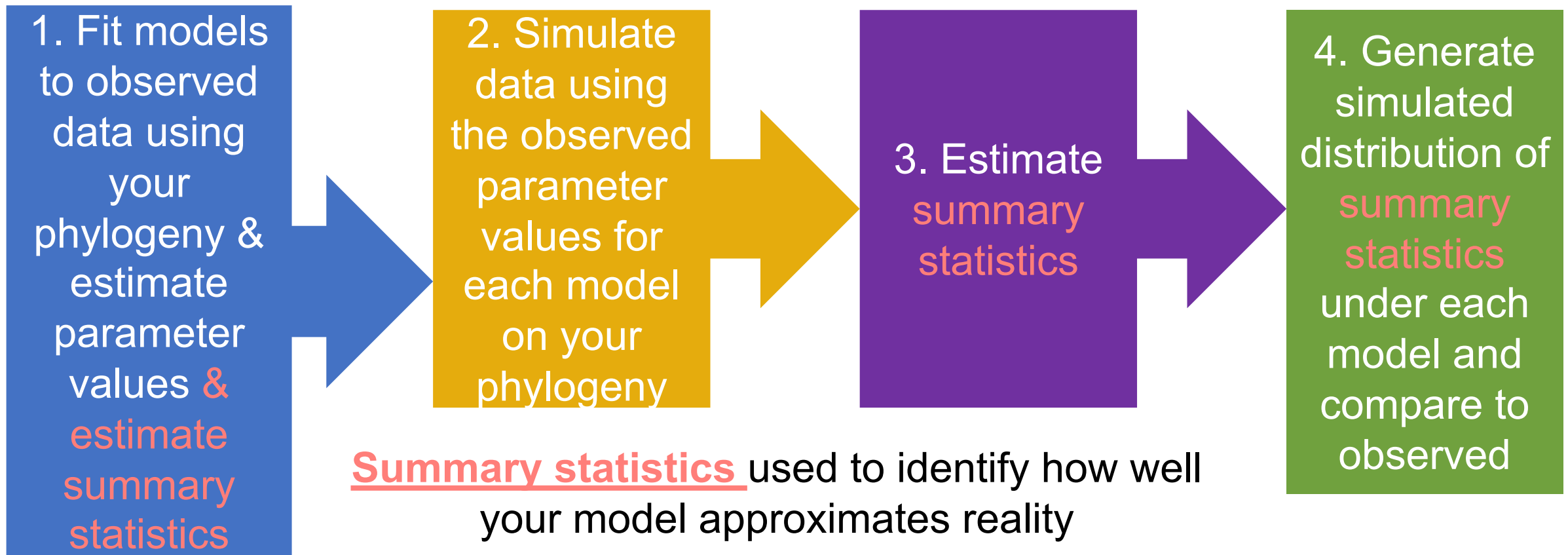


4. Do our models adequately reflect biological reality?



Bayesian world it is known as posterior prediction where you simulate under the parameter estimates in the posterior distribution.

4. Do our models adequately reflect biological reality?



4. Do our models adequately reflect biological reality?

How to identify summary statistics to appropriately quantify the observed data?

4. Do our models adequately reflect biological reality?

How to identify summary statistics to appropriately quantify the observed data?

Phylogenetics: Brown 2014

- Robinson-Foulds distances
- Statistical entropy (i.e., the information gain) between the marginal prior and posterior distributions of tree topologies.

Continuous Trait evolution: Pennell et al. 2015

- Calculate Independent contrasts and estimate 6 test statistics that indicate different violations of BM

Discrete Trait evolution?

4. Do our models adequately reflect biological reality?

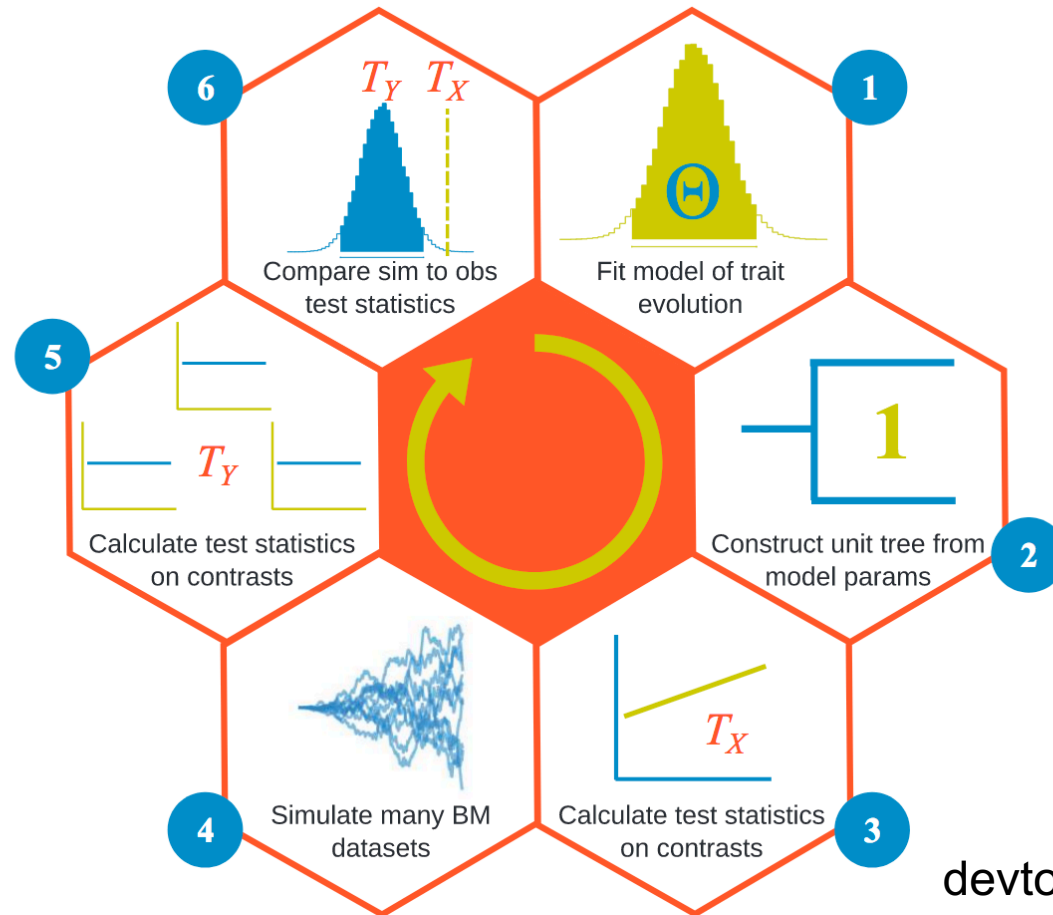
Table 1: Description of test statistics used to assess model adequacy

Test statistic	Description
M_{SIG}	The mean of the squared contrasts. This is equivalent to the restricted maximum likelihood estimator of the Brownian motion rate parameter σ^2 (Garland et al. 1992; Rohlf 2001). M_{SIG} is a metric of overall rate. Violations detected by M_{SIG} indicate whether the overall rate of trait evolution is over- or underestimated.
C_{VAR}	The coefficient of variation (standard deviation/mean) of the absolute value of the contrasts. If C_{VAR} calculated from the observed contrasts is greater than that calculated from the simulated contrasts, it suggests that we are not properly accounting for rate heterogeneity across the phylogeny. If C_{VAR} from the observed is smaller, it suggests that contrasts are even more than the model assumes. We use the coefficient of variation rather than the variance because the mean and variance of contrasts can be highly correlated.
S_{VAR}	The slope of a linear model fitted to the absolute value of the contrasts against their expected variances (following Garland et al. 1992). Each (standardized) contrast has an expected variance proportional to the sum of the branch lengths connecting the node at which it is computed to its daughter lineages (Felsenstein 1985). Under a model of Brownian motion, we expect no relationship between the contrasts and their variances. We use it to test whether contrasts are larger or smaller than we expect based on their branch lengths. If, for example, more evolution occurred per unit time on short branches than long branches, we would observe a negative slope. If S_{VAR} calculated from the observed data deviates substantially from the expectations, a likely explanation is branch length error in the phylogenetic tree.
S_{ASR}	The slope of a linear model fitted to the absolute value of the contrasts against the ancestral state inferred at the corresponding node. We estimated the ancestral state using the least squares method suggested by Felsenstein (1985) for the calculation of contrasts. (We note that this is not technically an ancestral state reconstruction [see Felsenstein 1985]; it is more properly thought of as a weighted average value for each node.) We used this statistic to evaluate whether there is variation in rates relative to the trait value. For example, do larger organisms evolve proportionally faster than smaller ones?
S_{HGT}	The slope of a linear model fitted to the absolute value of the contrasts against node depth (after Purvis and Rambaut 1995). This is used to capture variation relative to time. It is alternatively known as the “node-height test” and has been used to detect early bursts of trait evolution during adaptive radiations (for uses and modifications of this test see Freckleton and Harvey 2006; Slater and Pennell 2014).
D_{CDF}	The D statistic obtained from a Kolmogorov-Smirnov test from comparing the distribution of contrasts to that of a normal distribution with mean 0 and standard deviation equal to the root of the mean of squared contrasts (the expected distribution of the contrasts under Brownian motion; see Felsenstein 1985; Rohlf 2001). We chose this to capture deviations from normality. For example, if traits evolved via a “jump-diffusion”-type process (Landis et al. 2013) in which there were occasional bursts of rapid phenotypic evolution (Pennell et al. 2013), the tip data would no longer be multivariate normal owing to a few contrasts throughout the tree being much larger than the rest (i.e., the distribution of contrasts would have heavy tails).

The mean of the squared contrasts = metric of overall rate. Violations detected indicate whether the overall rate of trait evolution is over- or underestimated.

The slope of a linear model fitted to the absolute value of the contrasts against node depth = rate variation relative to time. Violations can detect early or late bursts of trait evolution

4. Do our models adequately reflect biological reality?



Transforming or
rescaling tree under
model parameters

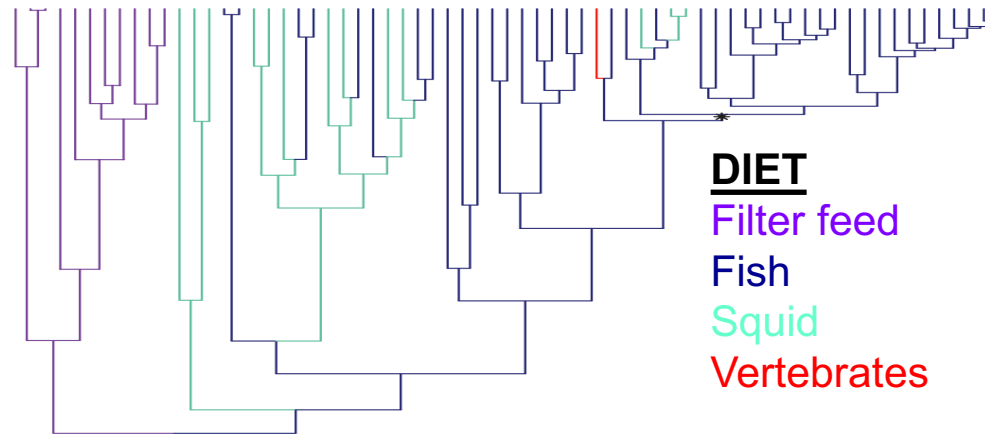
R package arbutus

```
devtools::install_github("mwpenne/arbustus")
```

Pennell et al. 2015

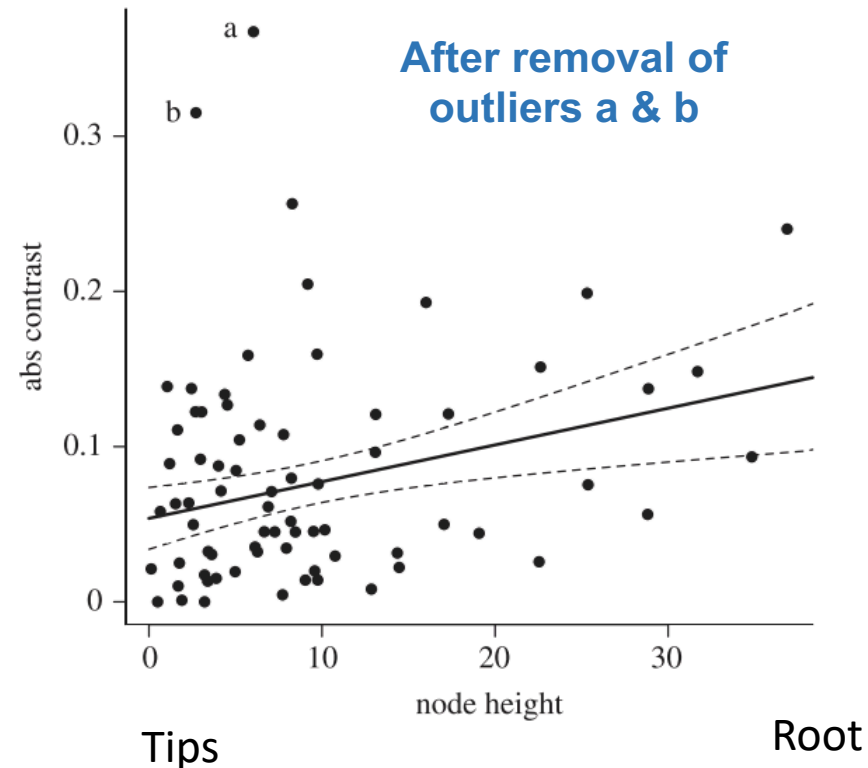
4. Do our models adequately reflect biological reality?

Example: Early burst of Cetacean body-size evolution or shift in Mysticetes?



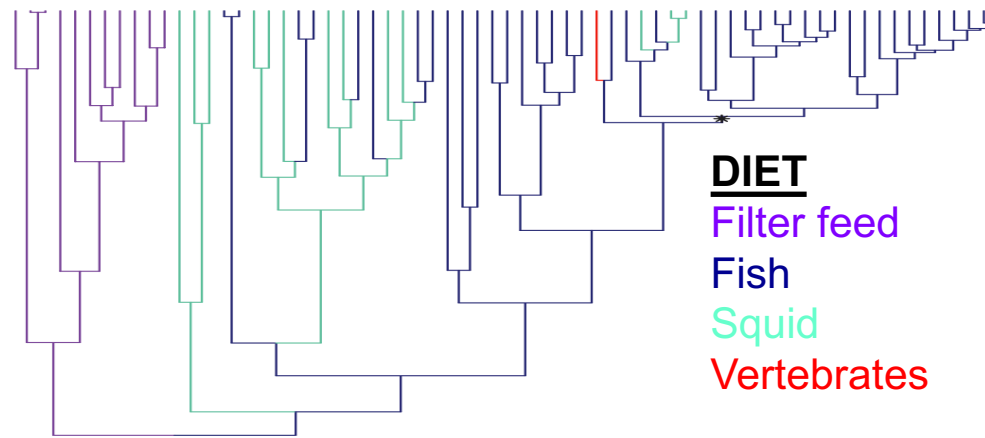
	AICc
BM	149.7
EB	150.4

Slater, Price et al. 2010
Venditti et al. 2011



4. Do our models adequately reflect biological reality?

Example: Early burst of Cetacean body-size evolution or shift in Mysticetes?

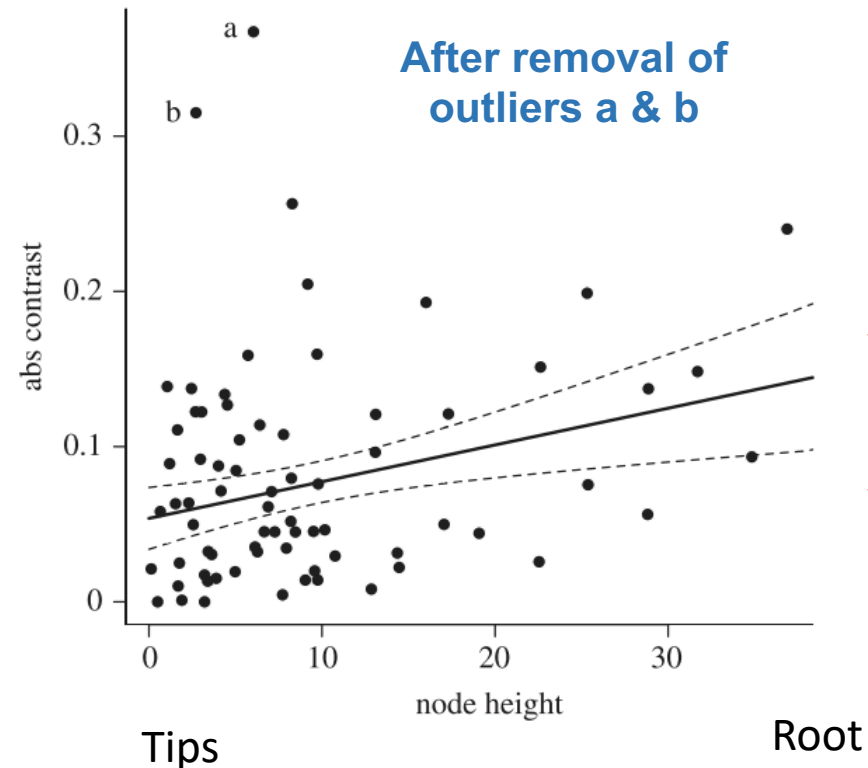


AICc

BM 149.7

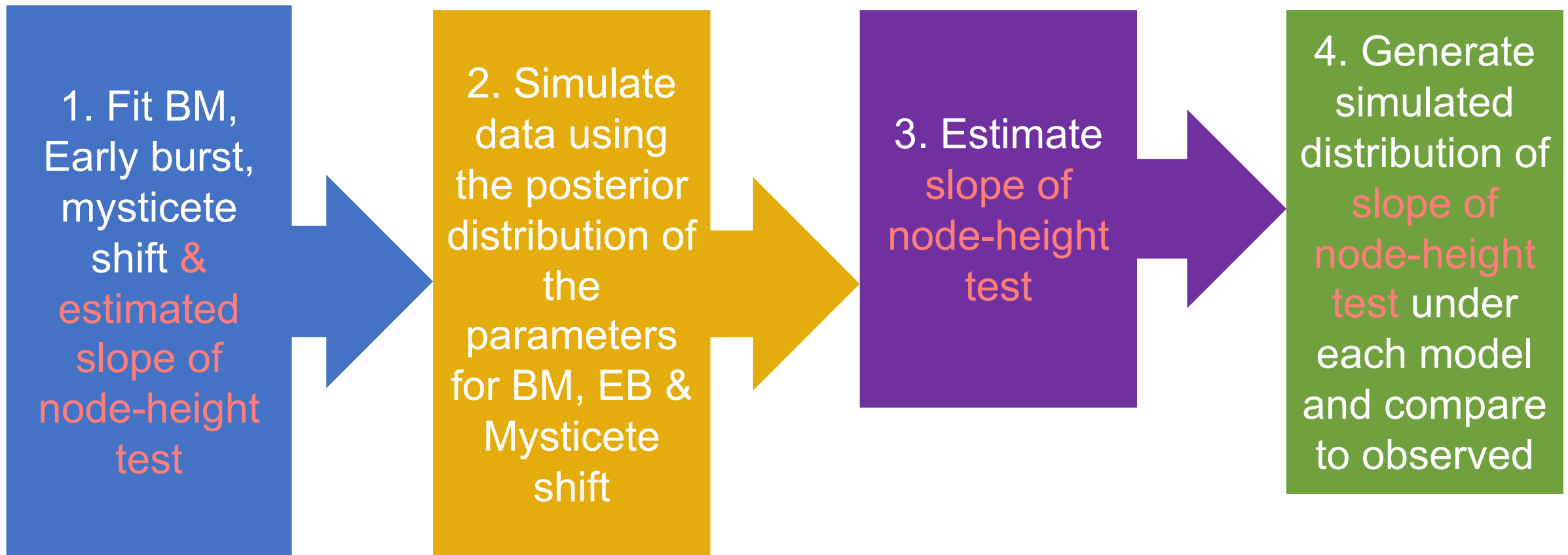
EB 150.4

Slater, Price et al. 2010
Venditti et al. 2011

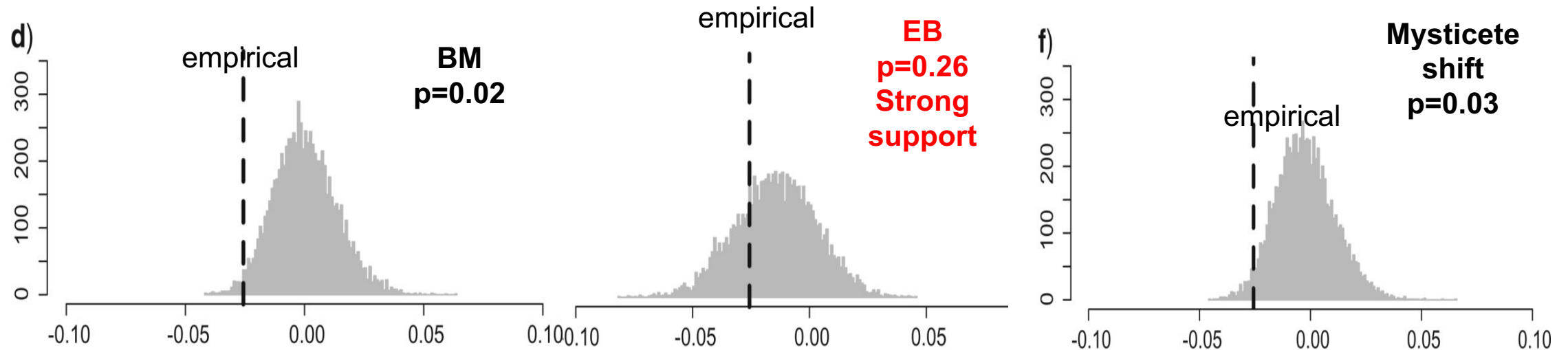


Summary
statistic –
Slope of
Log
absolute
value of
contrasts
vs. node
height

4. Do our models adequately reflect biological reality?



4. Do our models adequately reflect biological reality?



Node height test regression coefficient
(using robust regression to help with outliers)

4. Do our models adequately reflect biological reality?



simulationexercise_SSBworkshop.R

- 1) Example simulation to estimate confidence intervals around Pagel's λ
- 2) Example using `ouwie.boot` to estimate confidence intervals around OU multi-peak parameters
- 3) Example simulation to estimate power and uncertainty in discrete trait model fitting
- 4) **Example of using independent contrast summary statistics in `arbutus` to estimate model adequacy of models implemented in `fitContinuous`**

Five questions to ask

1. What is the uncertainty associated with the estimated parameter values?
2. Do we have the power to distinguish between different evolutionary models?
3. Can we correctly reject the null hypothesis?
 - What is the correct null hypothesis?
4. Do our models adequately reflect biological reality?
5. Do our methods adequately control for phylogenetic pseudo-replication?

5. Do our methods adequately control for phylogenetic pseudo-replication?

Maddison & Fitzjohn 2014

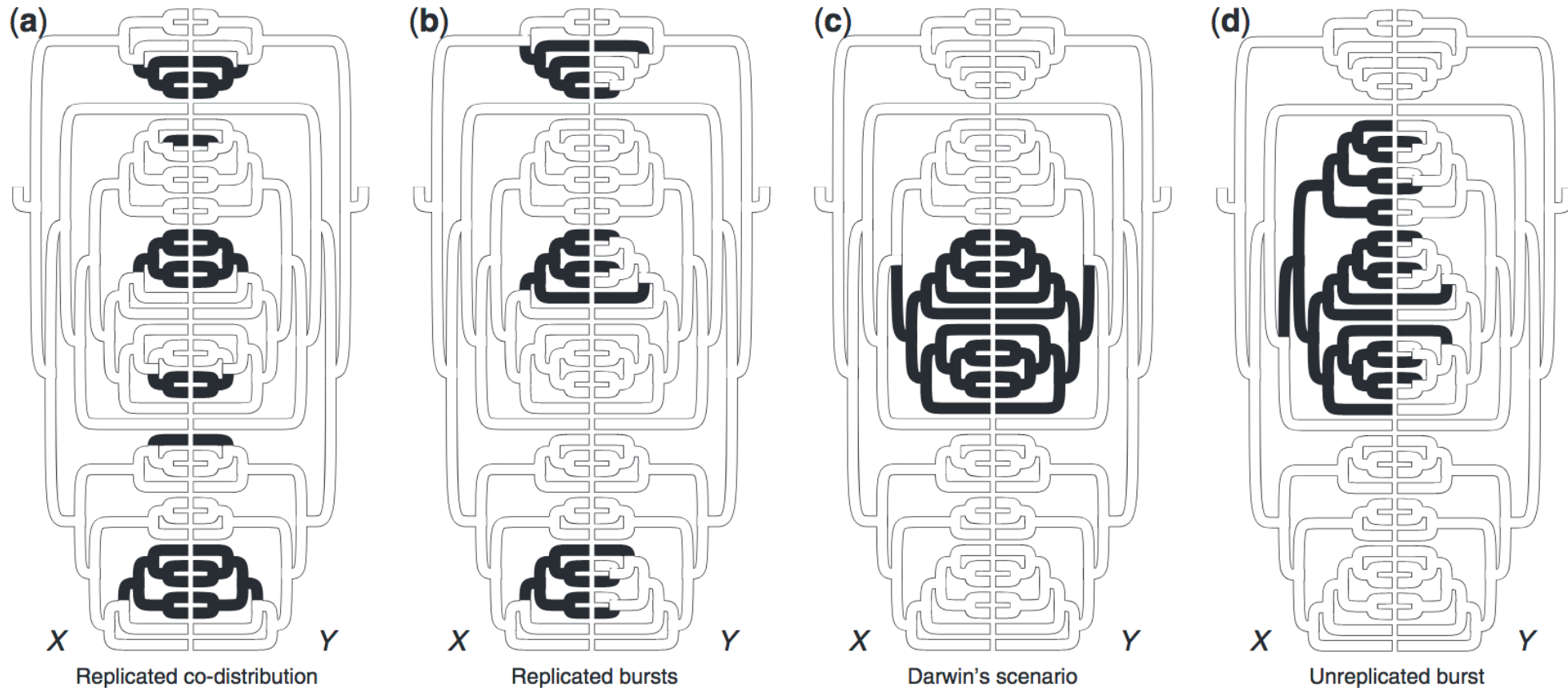
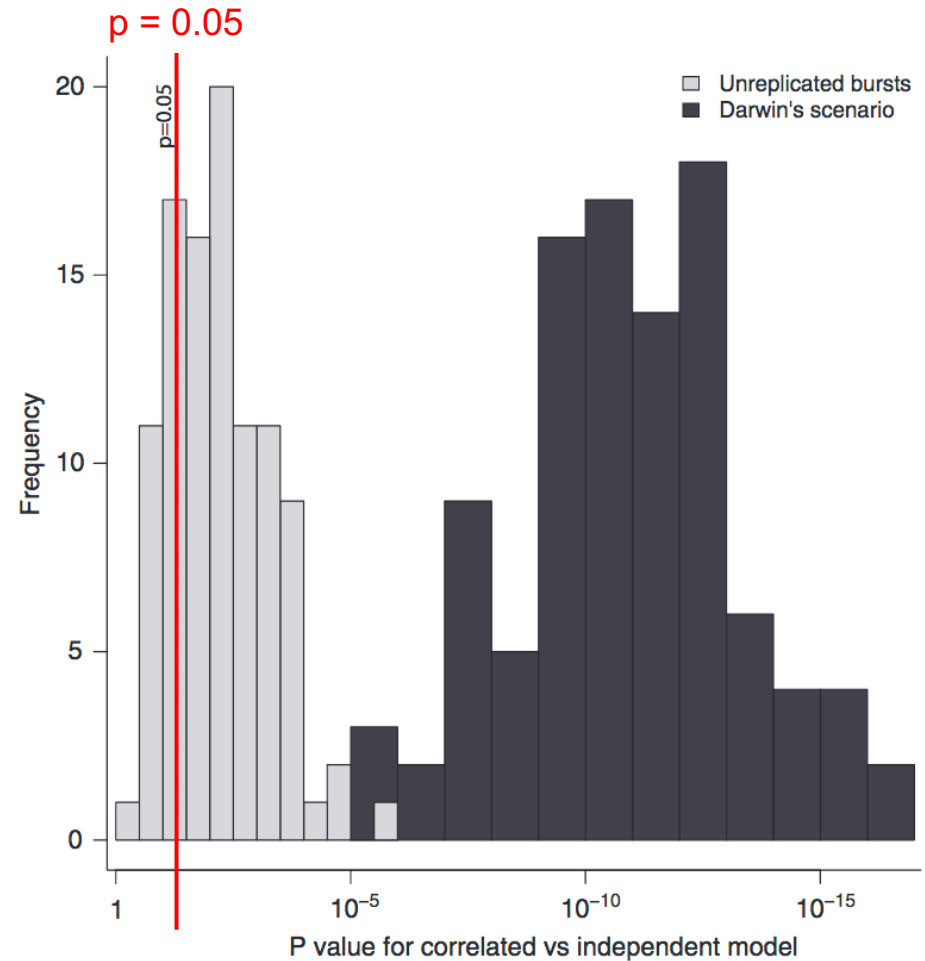


FIGURE 1. Four scenarios for the evolution of states of characters X and Y. In each, the same phylogeny is mirrored to show X at left and Y at right. State 0 = white; State 1 = black. (a) Replicated co-distribution. (b) Replicated bursts. (c) Darwin's scenario. (d) Unreplicated burst. Panels (a) and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

5. Do our methods adequately control for phylogenetic pseudo-replication?

Correlated discrete trait evolution Null simulations under Darwin's scenario or unreplicated burst → fitted models of independent (null) or dependent evolution (Pagel 1994).



5. Do our methods adequately control for phylogenetic pseudo-replication?

The problem of unique events extends beyond discrete trait correlations

Potential solutions

- Maddison & Fitzjohn 2014 suggests independent contrasts (Felsenstein 1985) with outlier detection (Jones & Purvis 1997) and pairwise comparisons (Read & Nee 1995).
- Uyeda et al. 2018 propose combining hypothesis testing with methods that have no *a priori* expectations e.g. fit shifts in BM rate, OU peak or Speciation/Extinction rate

Five questions to ask

1. What is the uncertainty associated with the estimated parameter values?
2. Do we have the power to distinguish between different evolutionary models?
3. Can we correctly reject the null hypothesis?
 - What is the correct null hypothesis?
4. Do our models adequately reflect biological reality?
5. Do our methods adequately control for phylogenetic pseudo-replication?

The future is bright when you think carefully about:

BIOLOGY – what are your expectations & predictions?

MODELS – interpretation of parameter estimates and what they can tell you given your DATA & PHYLOGENY

Cited References

- Beaulieu, J.M. and O'Meara, B.C., 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic biology*, 65(4), pp.583-601.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- Hua, X. and Bromham, L., 2016. Phylometrics: an R package for detecting macroevolutionary patterns, using phylogenetic metrics and backward tree simulation. *Methods in Ecology and Evolution*, 7(7), pp.806-810.
- Jones K.E., Purvis A. 1997. An optimum body size for mammals? Comparative evidence from bats. *Functional Ecology*. 11:751–756.
- May, M.R. and Moore, B.R., 2019. A Bayesian Approach for Inferring the Impact of a Discrete Character on Rates of Continuous-Character Evolution in the Presence of Background-Rate Variation. *bioRxiv*, p.576207.
- Pennell, Matthew W., Richard G. FitzJohn, William K. Cornwell, and Luke J. Harmon. "Model adequacy and the macroevolution of angiosperm functional traits." *The American Naturalist* 186, no. 2 (2015): E33-E50.
- Read, A.F. and Nee, S., 1995. Inference from binary comparative data. *Journal of Theoretical Biology*, 173(1), pp.99-108.
- Slater, G.J., Price, S.A., Santini, F. and Alfaro, M.E., 2010. Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society B: Biological Sciences*, 277(1697), pp.3097-3104.
- Slater, G.J. and Pennell, M.W., 2013. Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. *Systematic Biology*, 63(3), pp.293-308.
- Uyeda, J.C., Zenil-Ferguson, R. and Pennell, M.W., 2018. Rethinking phylogenetic comparative methods. *Systematic Biology*, 67(6), pp.1091-1109.
- Venditti, C., Meade, A. and Pagel, M., 2011. Multiple routes to mammalian diversity. *Nature*, 479(7373), p.393.

SIX questions to ask

1. What is the uncertainty associated with the estimated parameter values?
2. Do we have the power to distinguish between different evolutionary models?
3. Can we correctly reject the null hypothesis?
 - What is the correct null hypothesis?
4. Do our models adequately reflect biological reality?
5. Do our methods adequately control for phylogenetic pseudo-replication?
6. Do I understand the models that I am fitting to my data?

5. Do I understand the models that I am fitting to my data?

If you have questions and are struggling to intuit the answer you can often simulate to gain insights or to check your intuition

- How do α and σ^2 interact to generate tip variance?
- When are α and σ^2 identifiable?
- What is the effect of extinction on parameter estimation?
- Can you distinguish accelerating evolution from OU?