

Is your phylogeny informative?

Measuring the power of comparative methods

Carl Boettiger
University of California, Davis
SICB 2011

Fell free to share live



1985

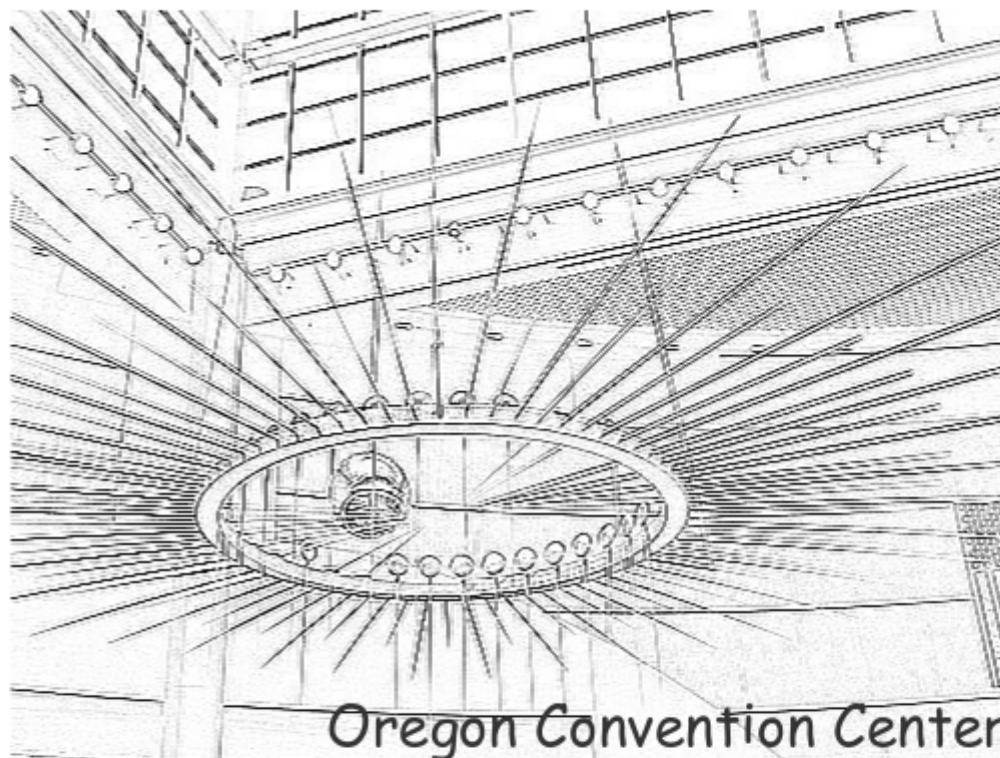




Brownian Motion

Grafen (1989)
Garland et. al. (1993)
Freckleton & Harvey (1996)
Schluter et. al. (1997)
Hansen (1997)
Blomberg et. al. (2002)
Butler & King (2004)
O'Meara et. al. (2006)





"Has the pendulum swung too far?"

What if



Natural selection has erased the
effect of shared ancestry?

Two Questions

Right Model?

Sufficient Data?

Currently we worry that we don't have the right models.

We should worry that we don't have the power to know if we have the right models.

Right Model?

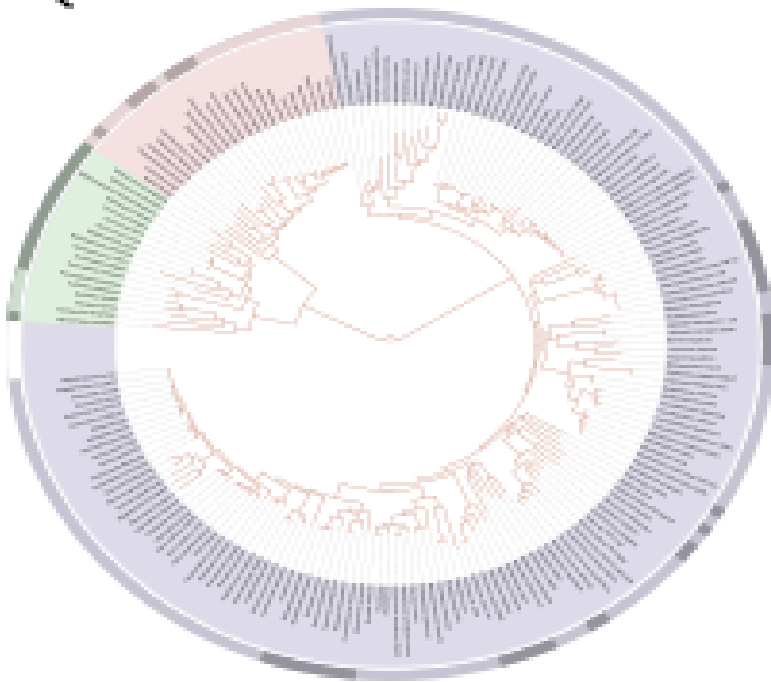
Have we crammed enough
biology into the tree?



(diagram shows tree
under weight of so
many models)

Sufficient Data?

Is the data-set big enough?



... informative about
the things we want to
estimate?



A model of "phylogenetic signal"

Are closely related species really more similar?

λ

"No signal"

0

"strong signal"

1

(Brownian motion)



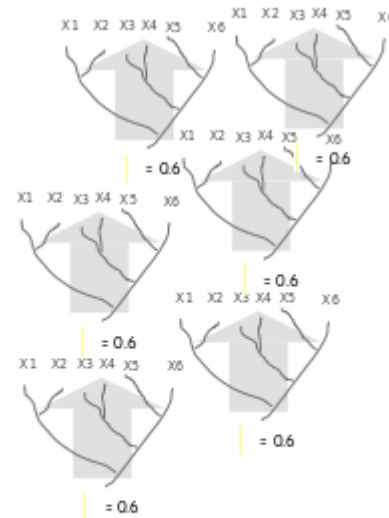
A simulation-based test of this method

The Space of Possible Outcomes

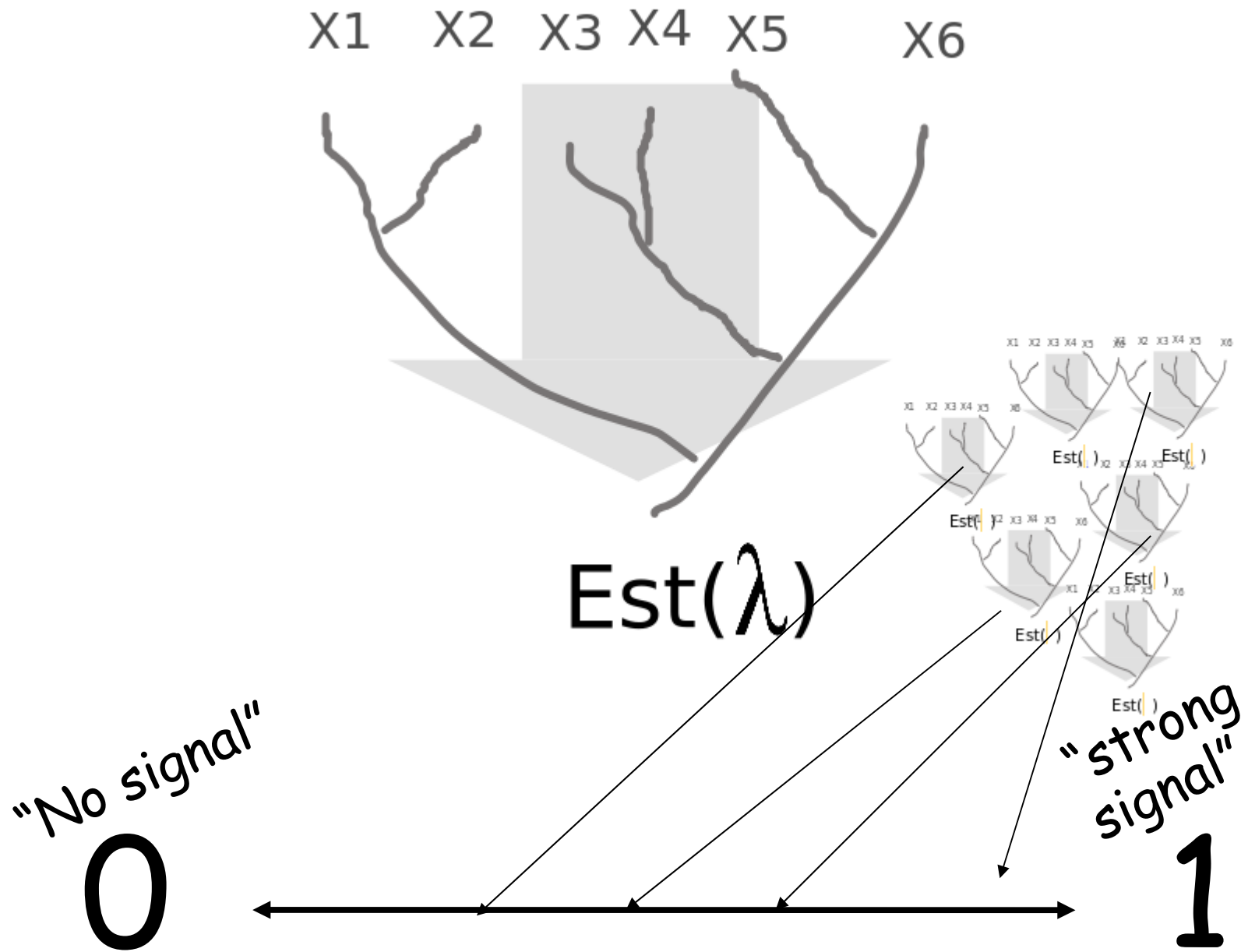
X1 X2 X3 X4 X5 X6



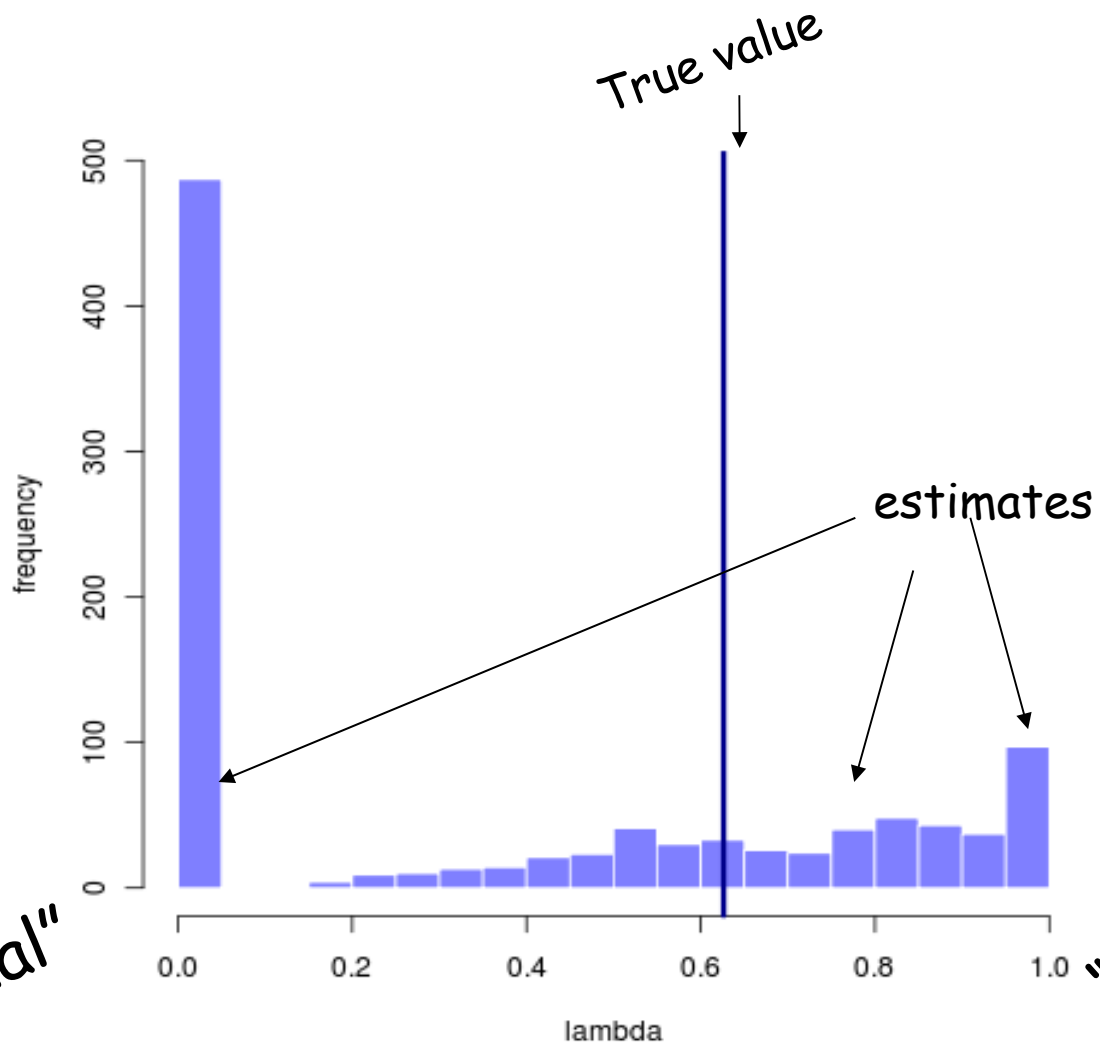
$$\lambda = 0.6$$



A simulation-based test of this method



A simulation-based test of this method



"No signal"

0

"strong
signal"

1

(Brownian
motion)

What went wrong?

Was supposed to address the
problem of adequate data...

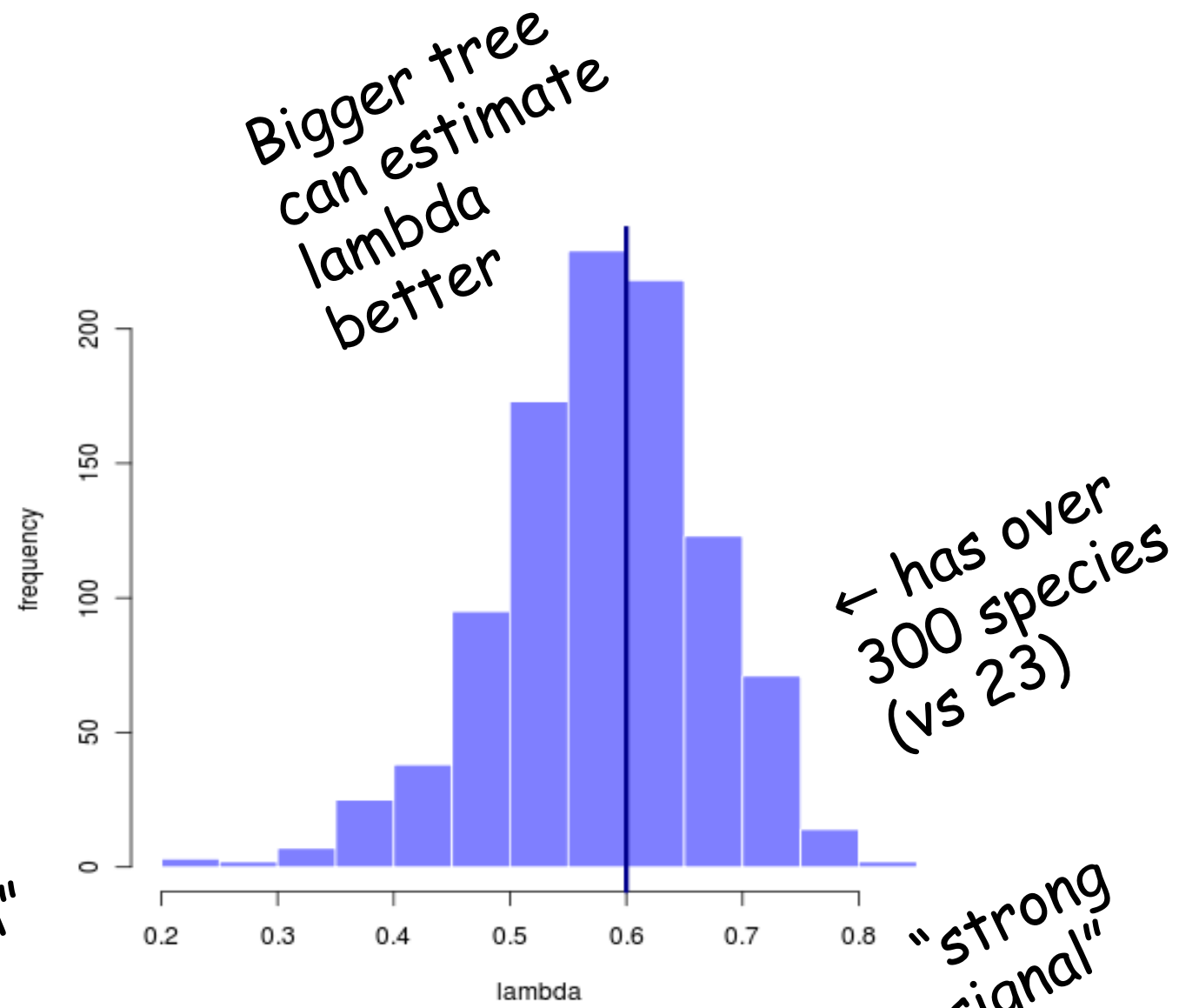
Instead, it's just fallen prey
to the problem

Insufficient signal to
estimate signal



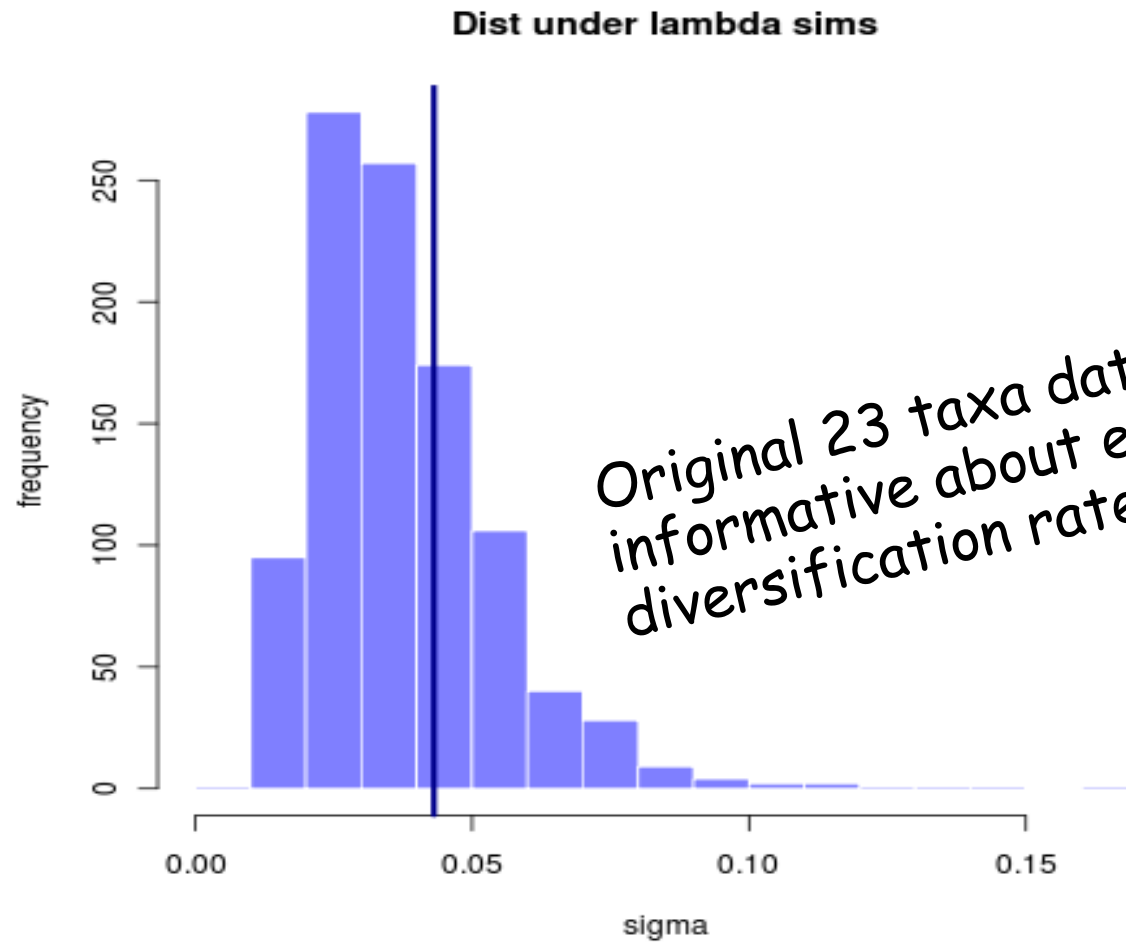
"Has the pendulum swung too far?"

"No signal"
0



"strong signal"
1

Not just size that matters....



Is the phylogeny informative?

Depends on the question

Depends on the amount of data

So as we add more complex models, do we have the data to justify them?

How would we know?

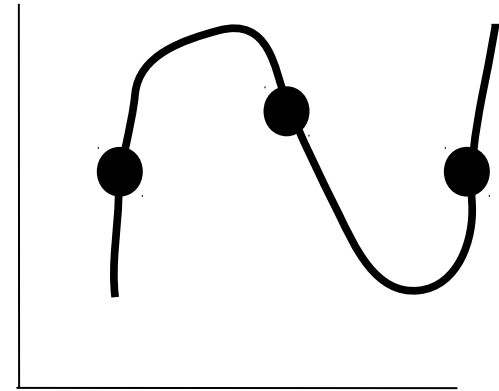
Grafen (1989)
Garland et. al. (1993)
Freckleton & Harvey (1996)
Schluter et. al. (1997)
Hansen (1997)
Blomberg et. al. (2002)
Butler & King (2004)
O'Meara et. al. (2006)



Process vs Pattern

Models represent *mechanisms*

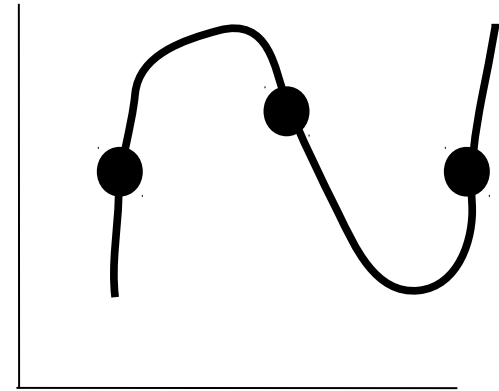
More complex models *fit the data better*



Process vs Pattern

Models represent *mechanisms*

More complex models *fit the data better*

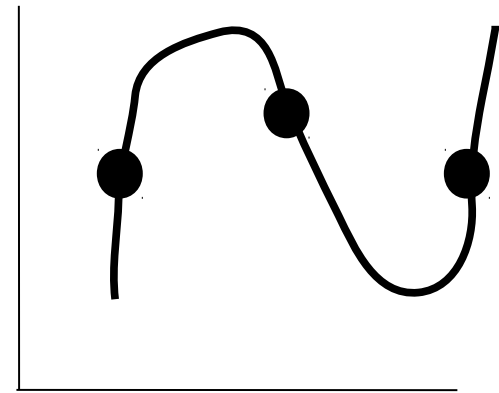


Can we tell when we've
matched the mechanism and
not just the pattern?

Process vs Pattern

Models represent *mechanisms*

More complex models *fit the data better*

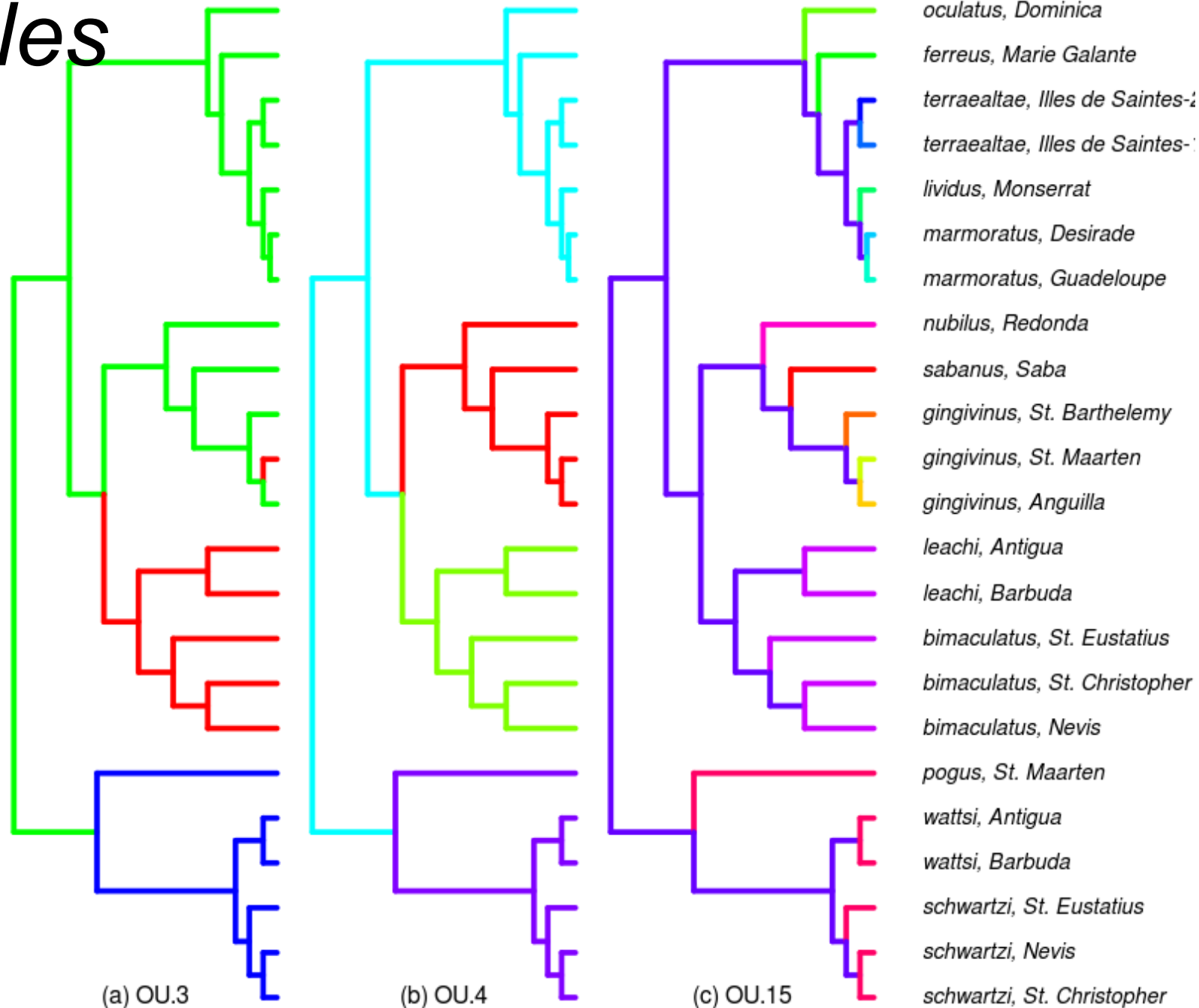


Can we tell when we've
matched the *mechanism* and
not just the *pattern*?

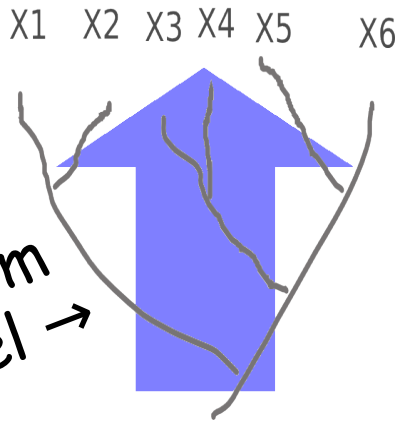
Yes!

Some possible models of evolution for *Anoles*

Nature Precedings : doi:10.1038/npre.2011.6453.1 : Posted 23 Sep 2011

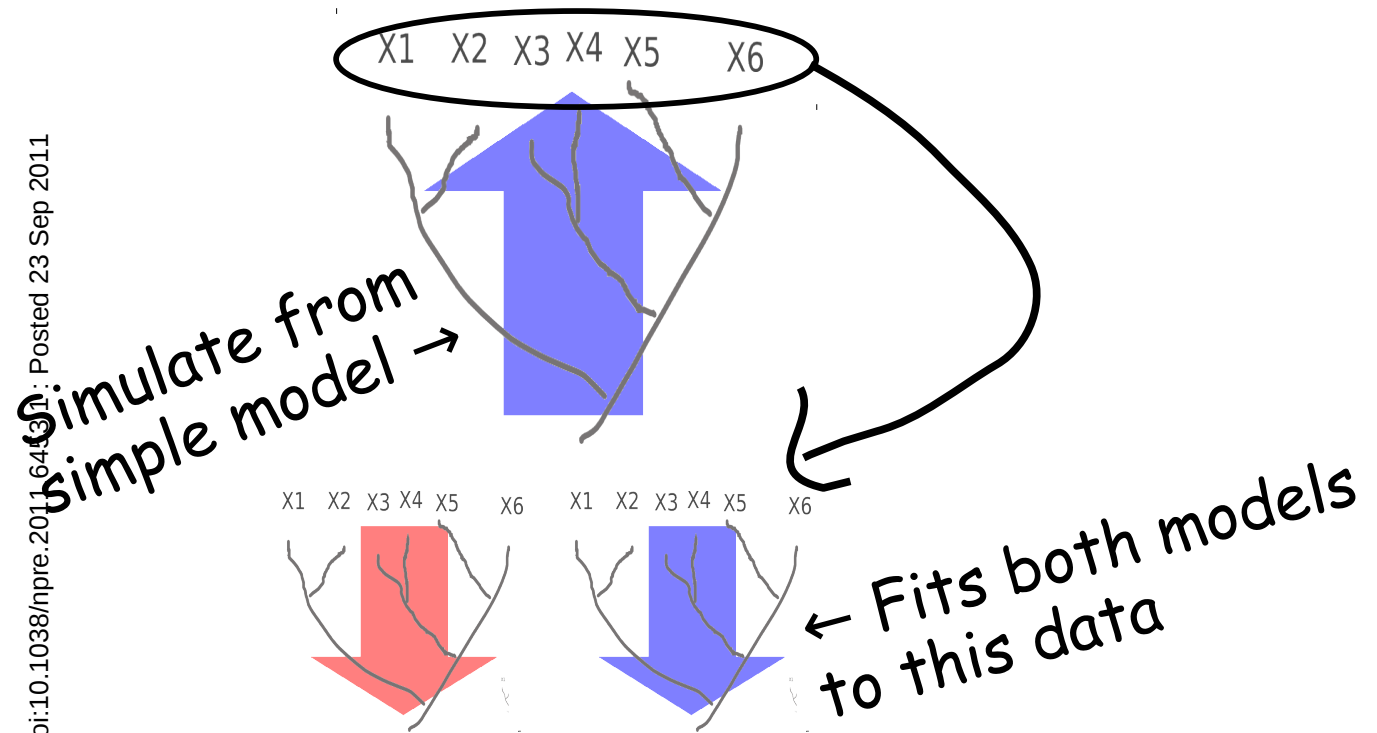


Phylogenetic Monte Carlo

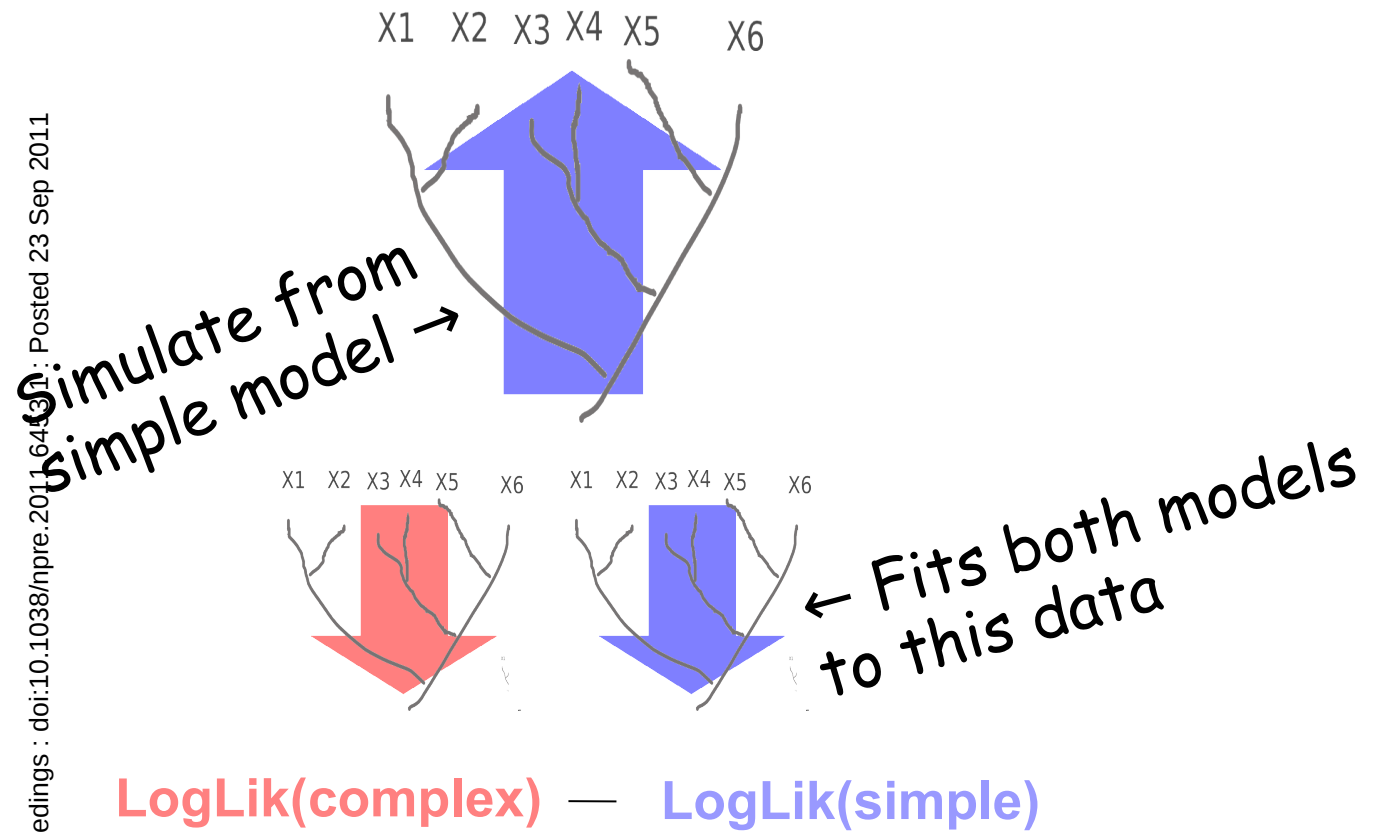


Simulate from
simple model →

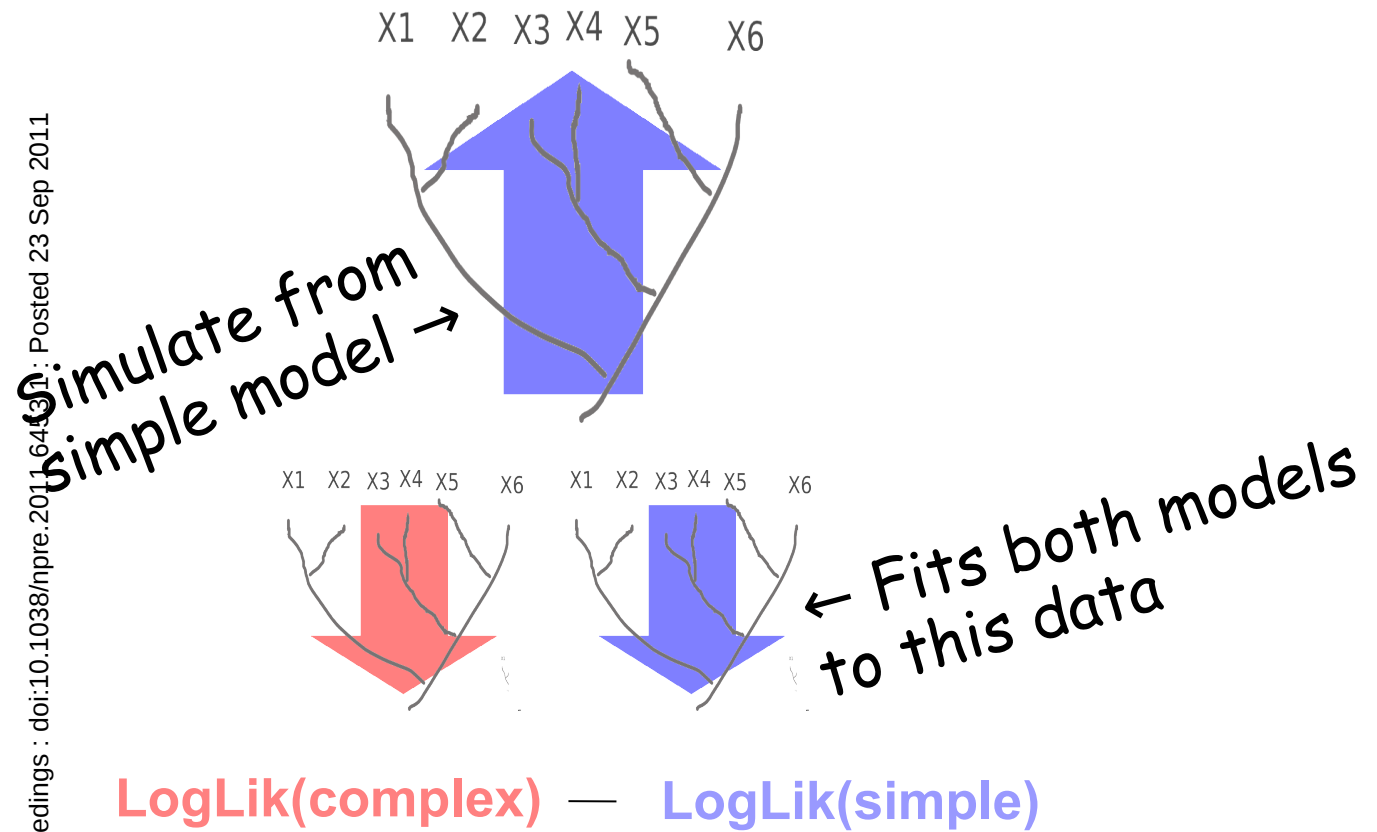
Phylogenetic Monte Carlo



Phylogenetic Monte Carlo



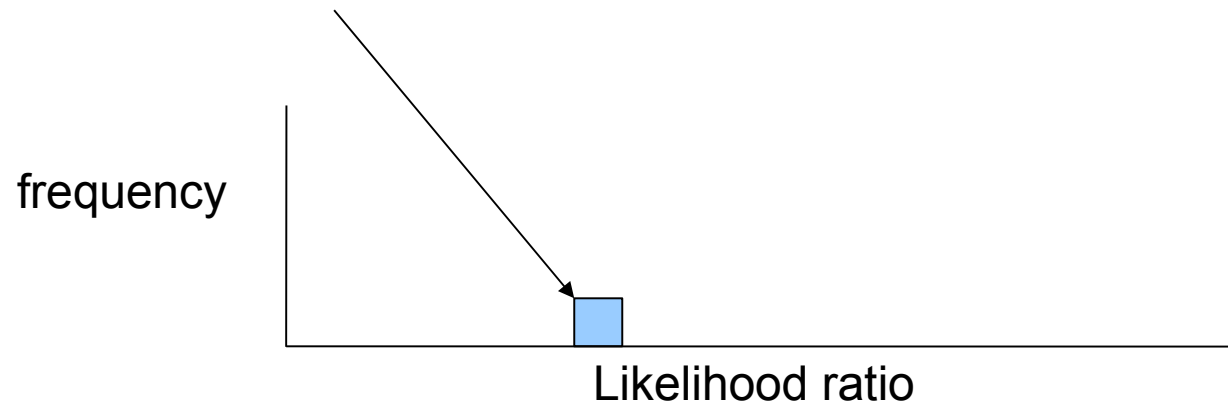
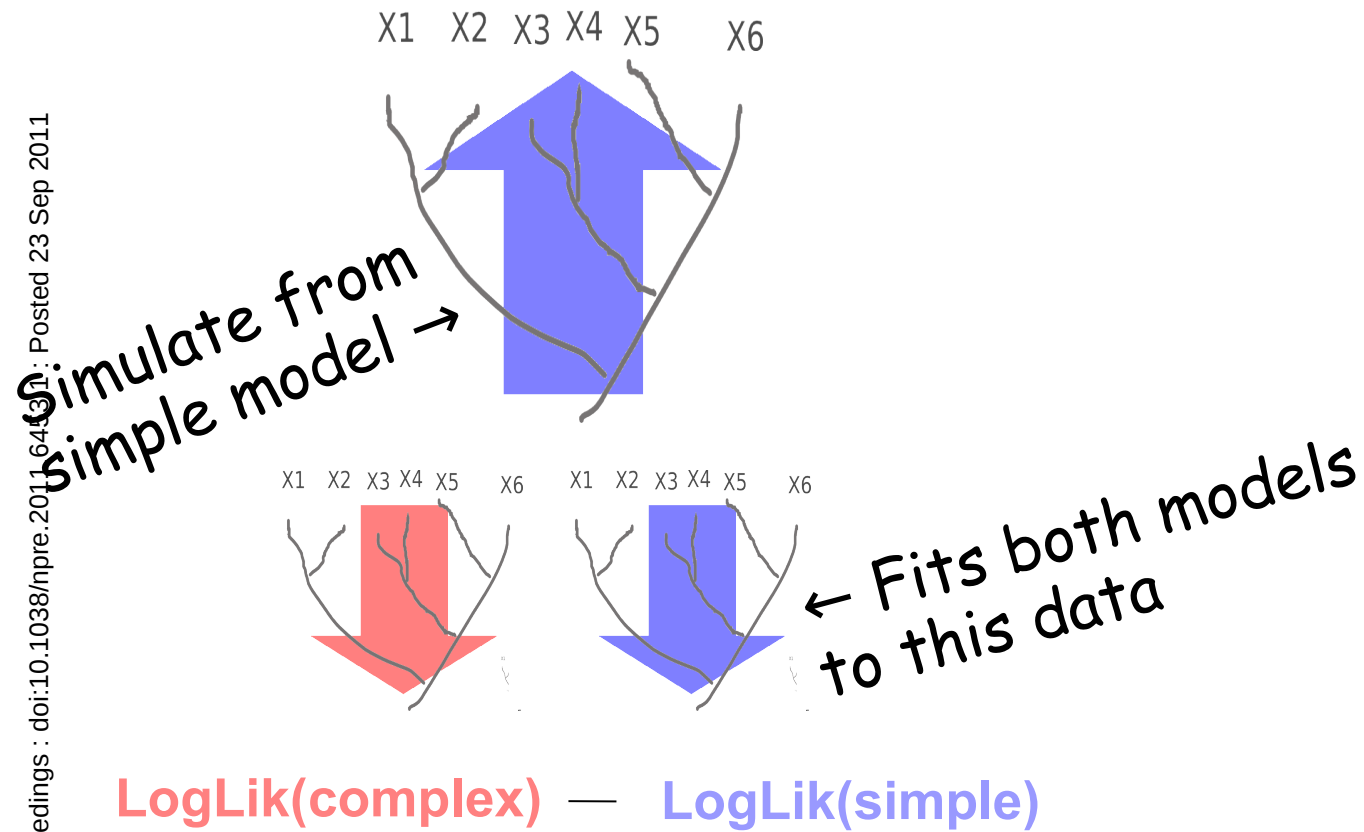
Phylogenetic Monte Carlo



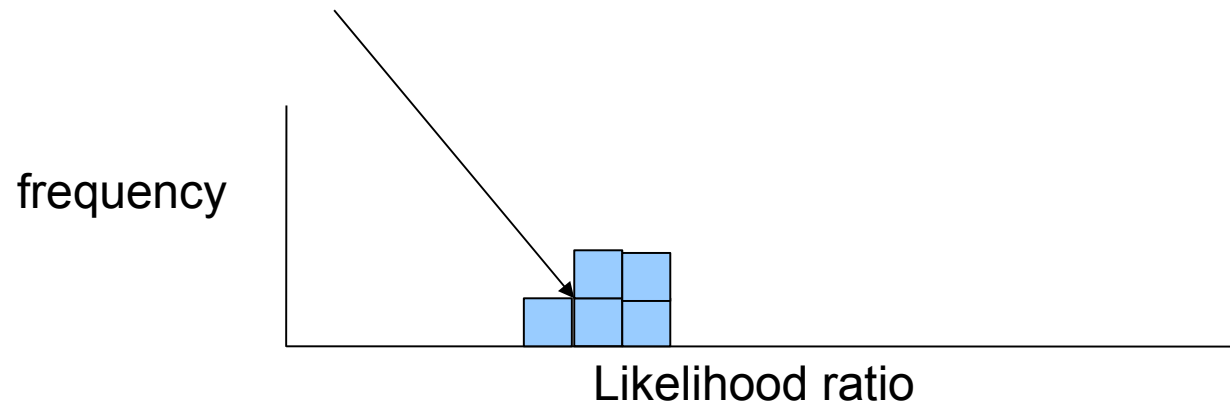
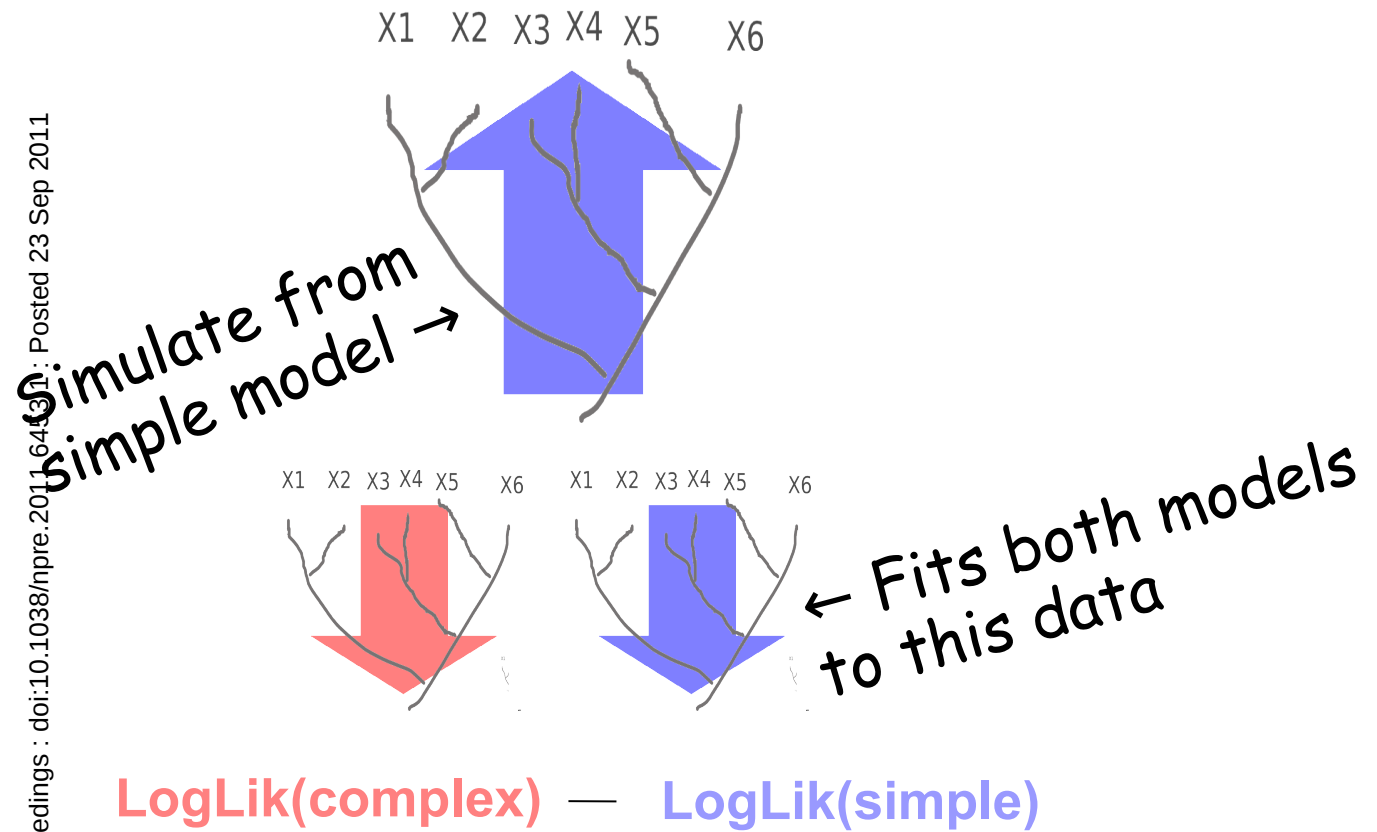
frequency

Likelihood ratio

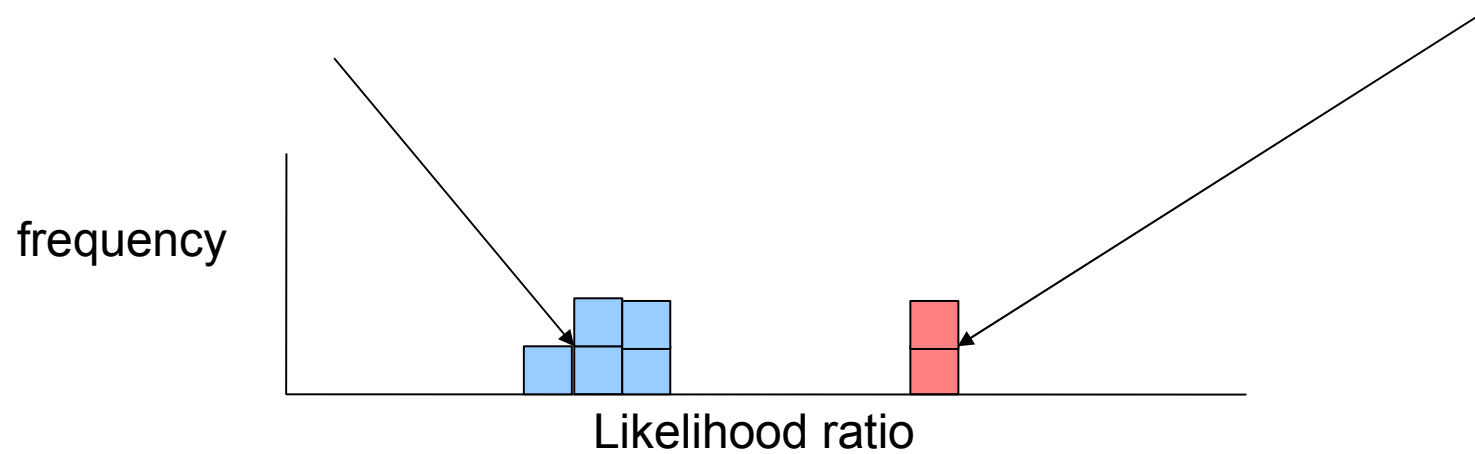
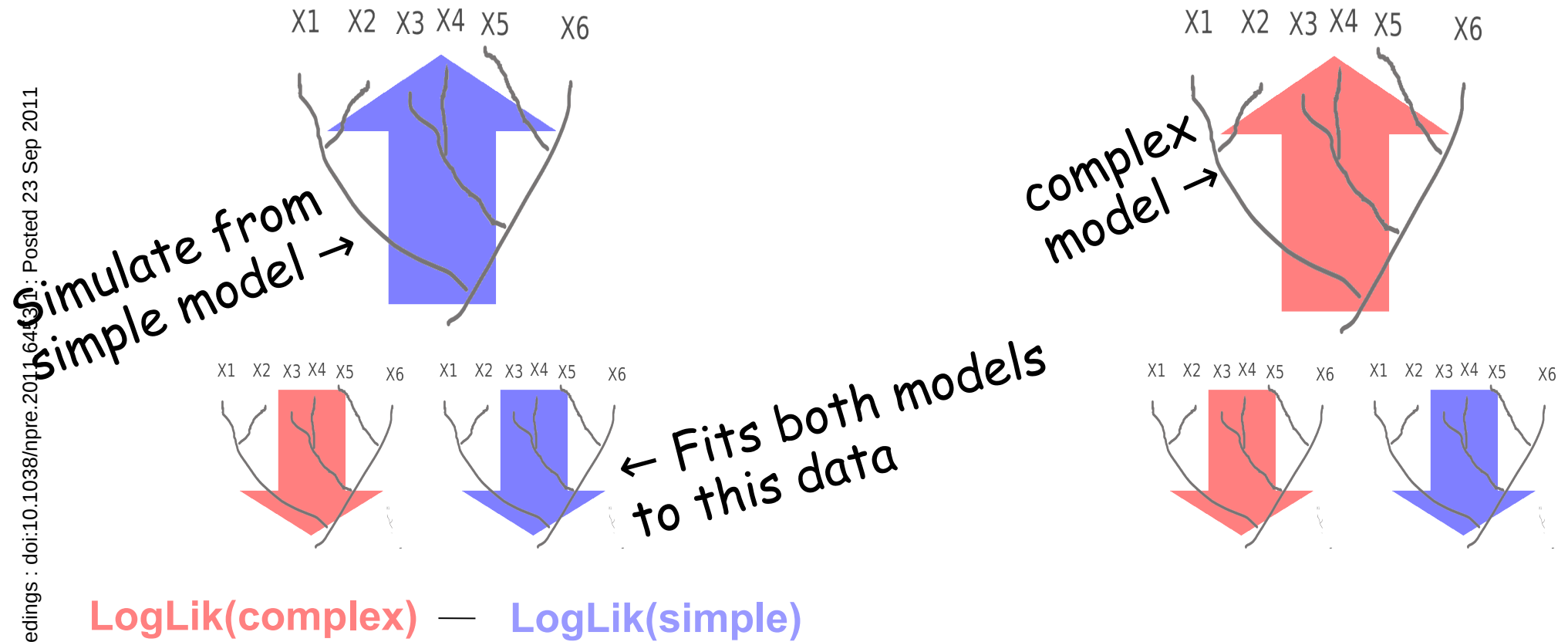
Phylogenetic Monte Carlo



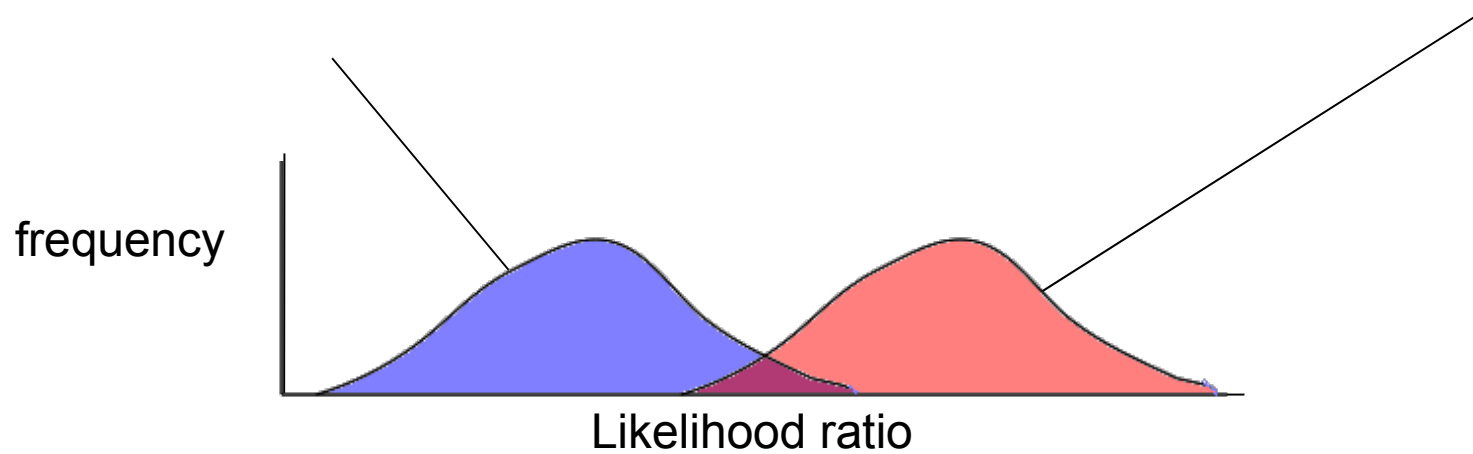
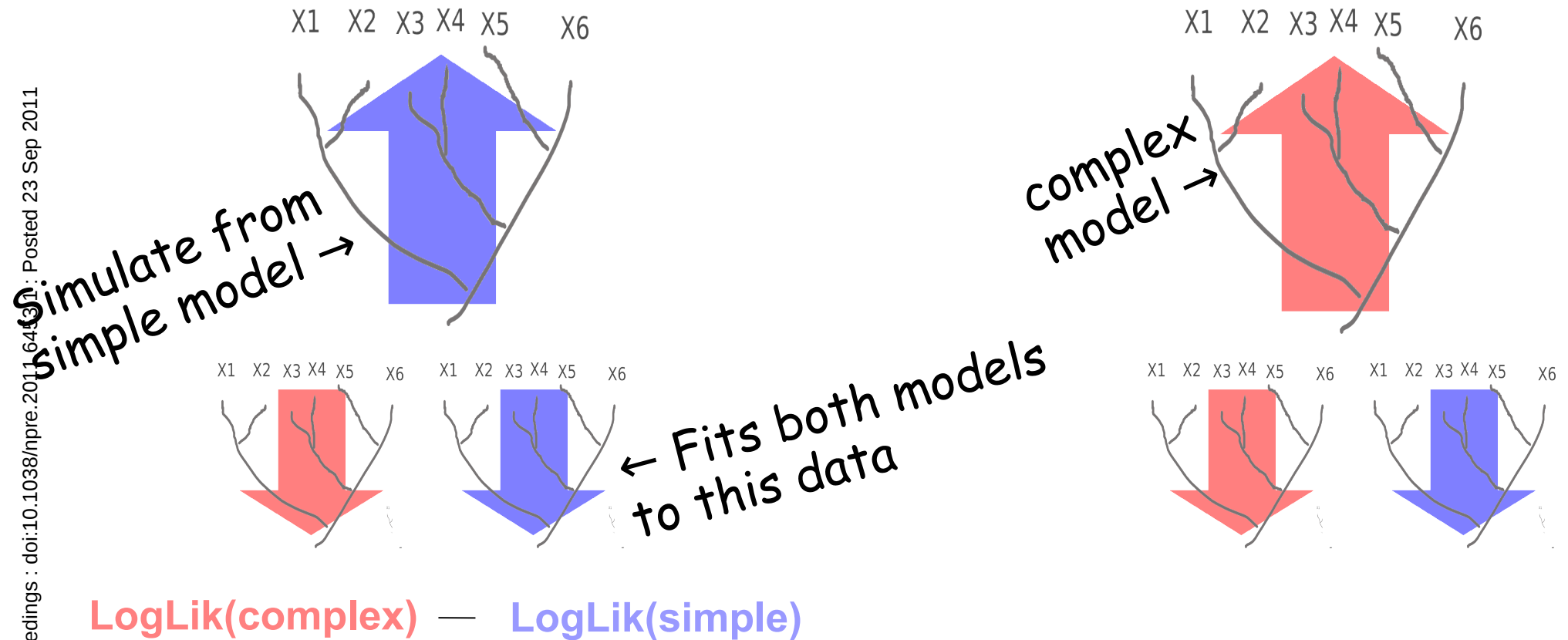
Phylogenetic Monte Carlo



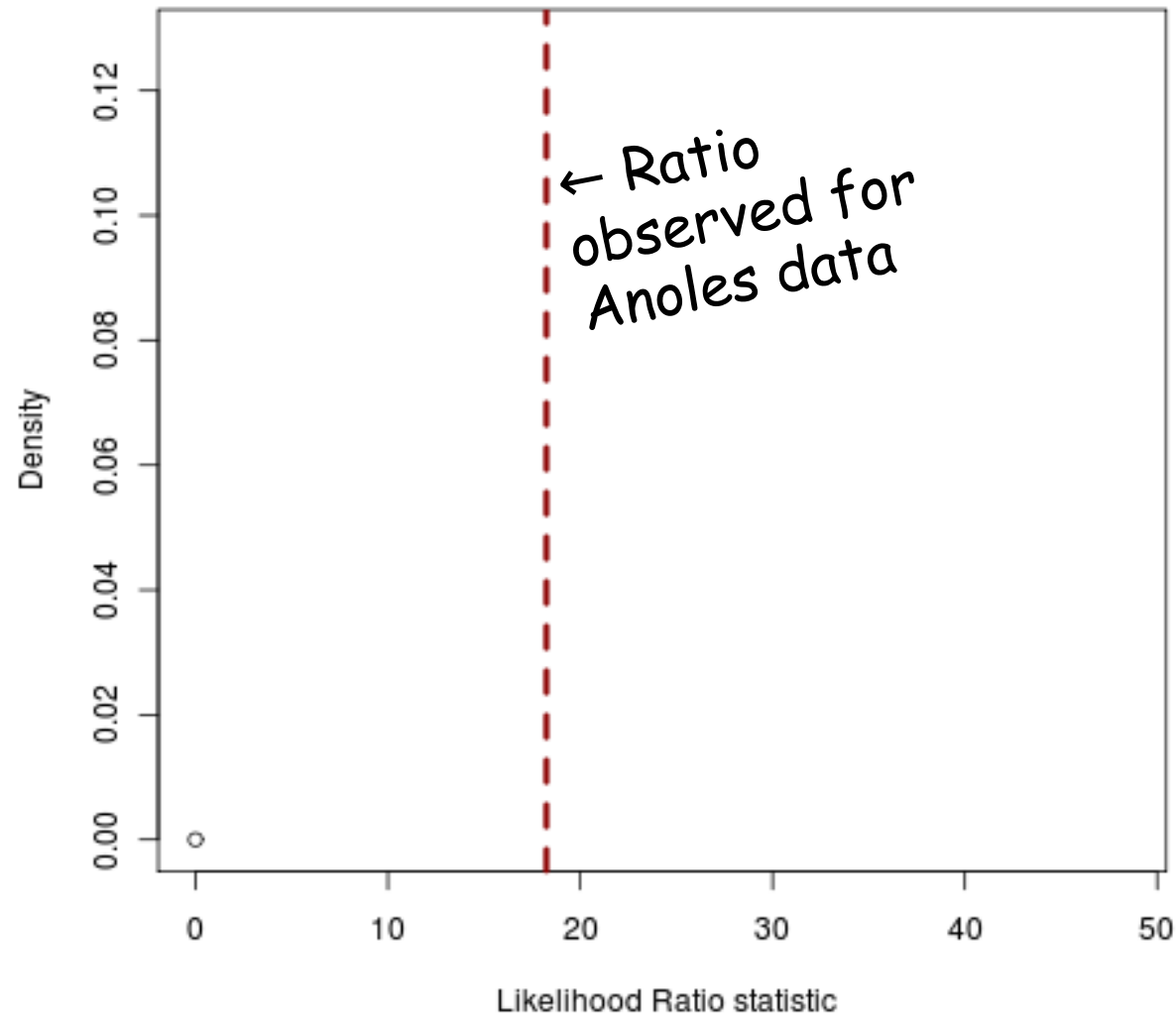
Phylogenetic Monte Carlo



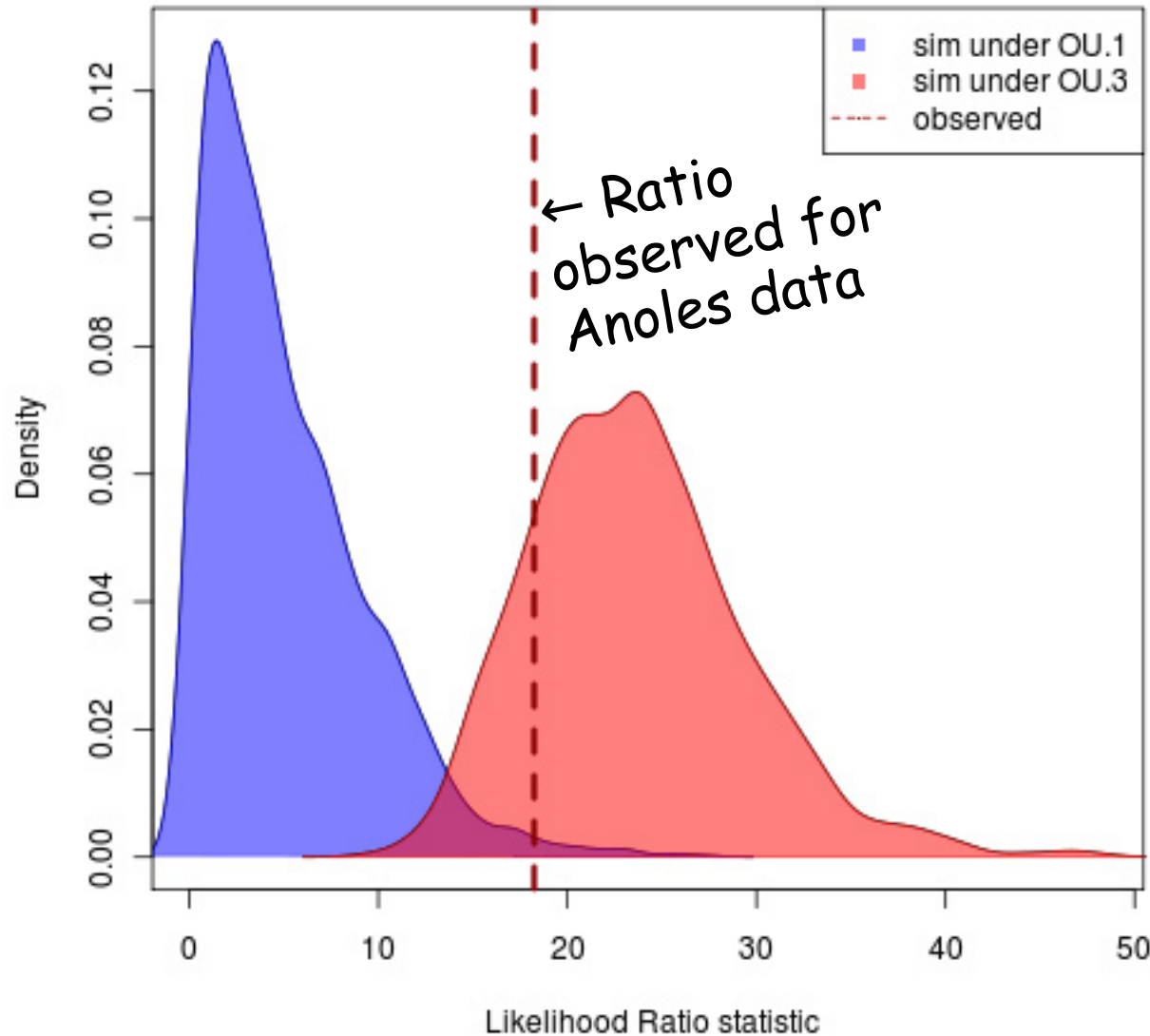
Phylogenetic Monte Carlo



A reliable way to justify the complex model



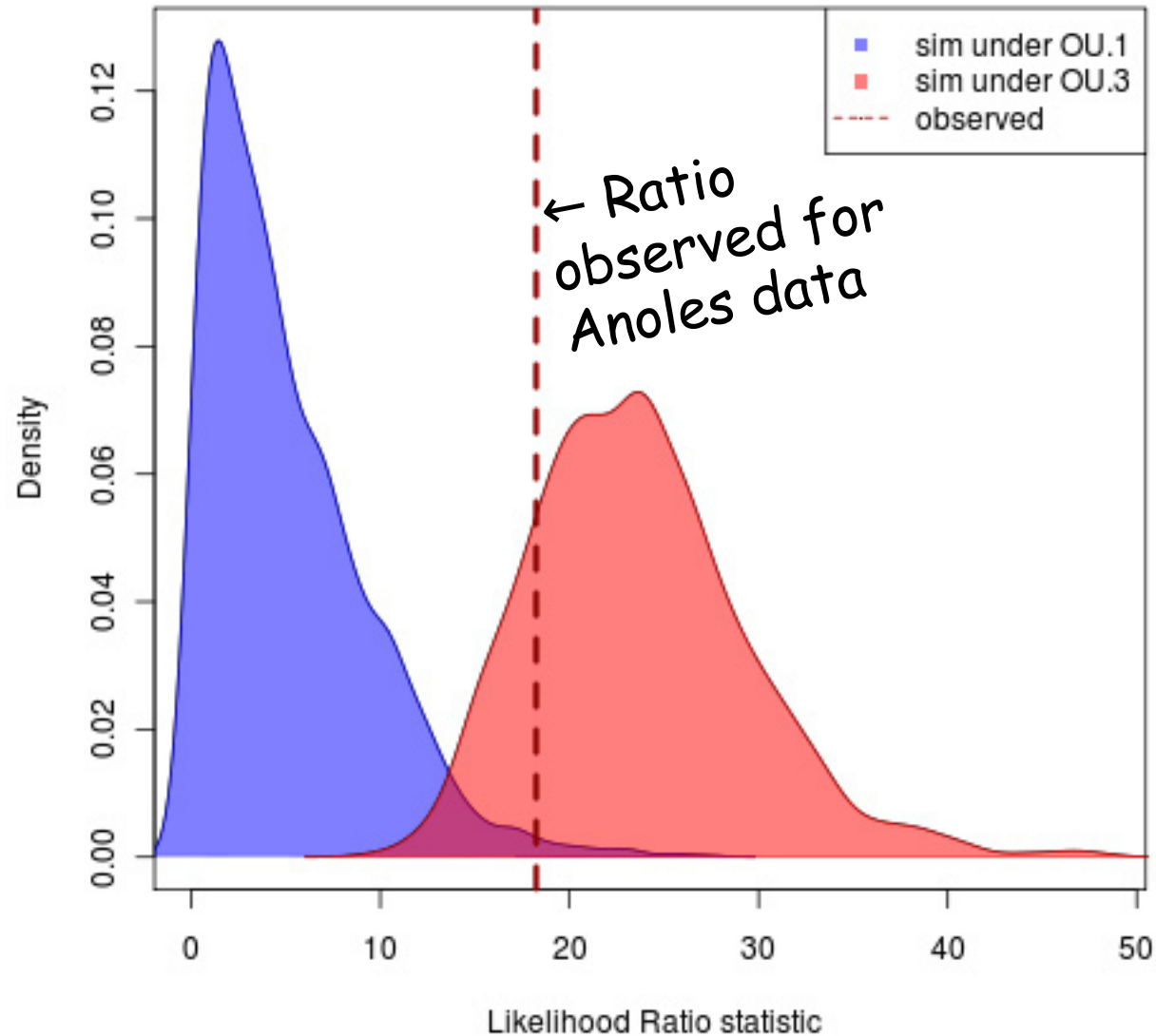
A reliable way to justify the complex model



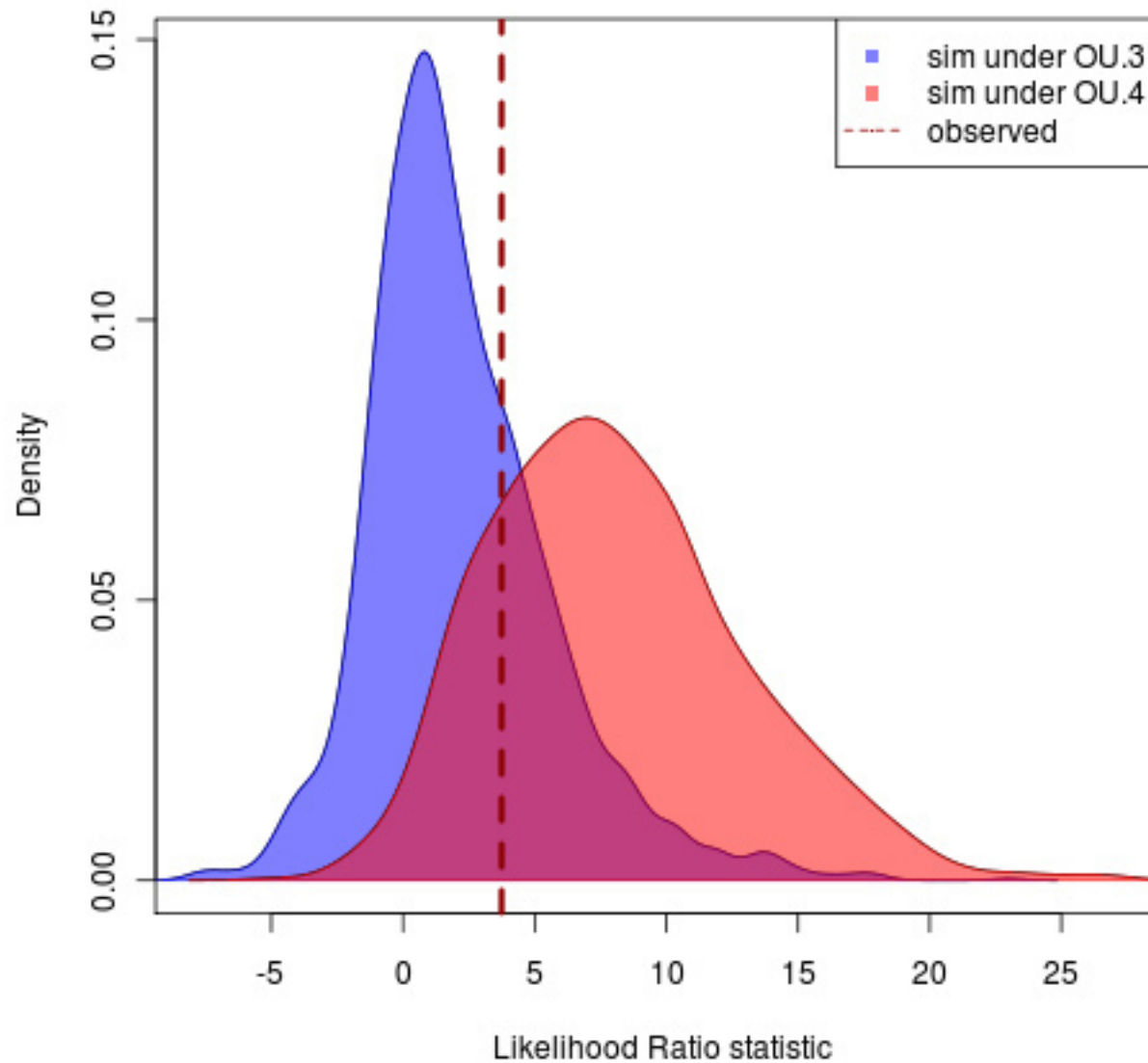
Prob of false positive: 0.0115

Power at 95% confidence: 0.99

A reliable way to justify the complex model

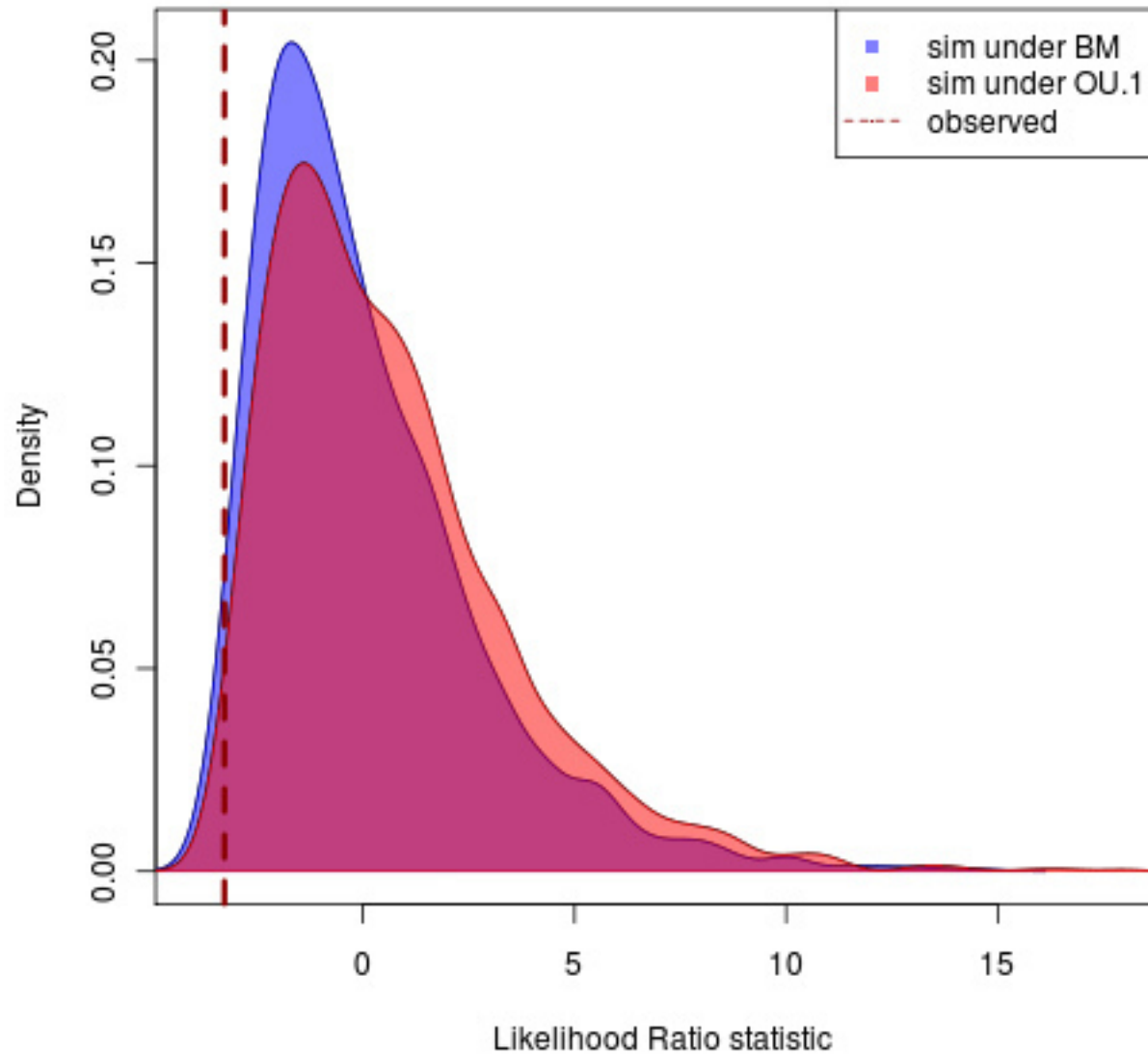


Greater overlap indicates less information,
less power to distinguish

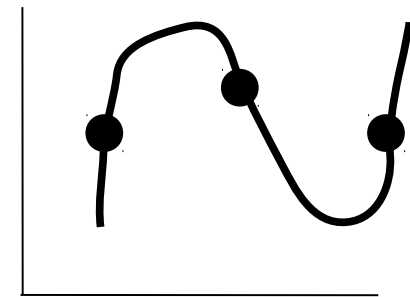
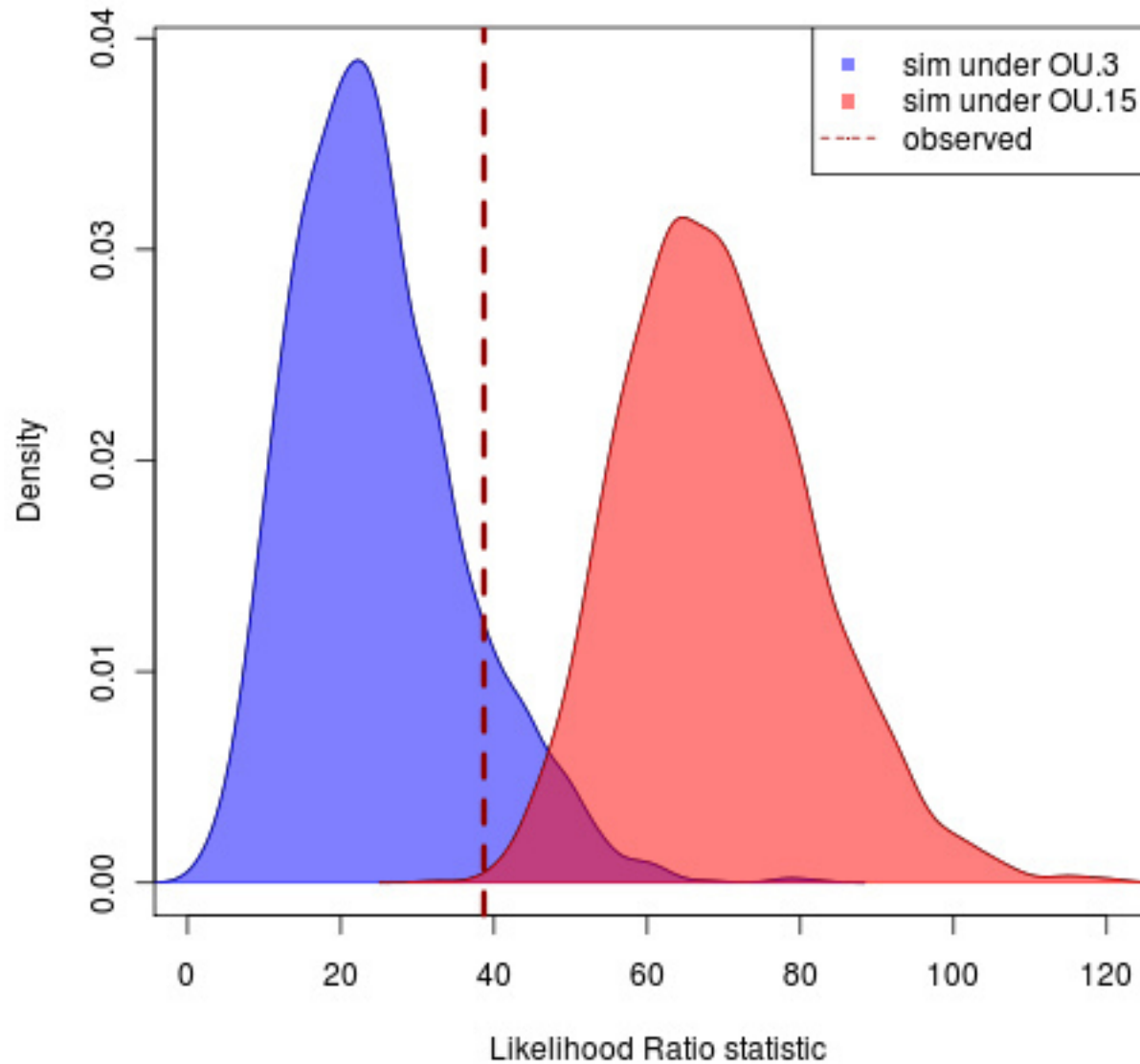


No trouble with
non-nested
models

And we can identify cases where the tree really has no information at all



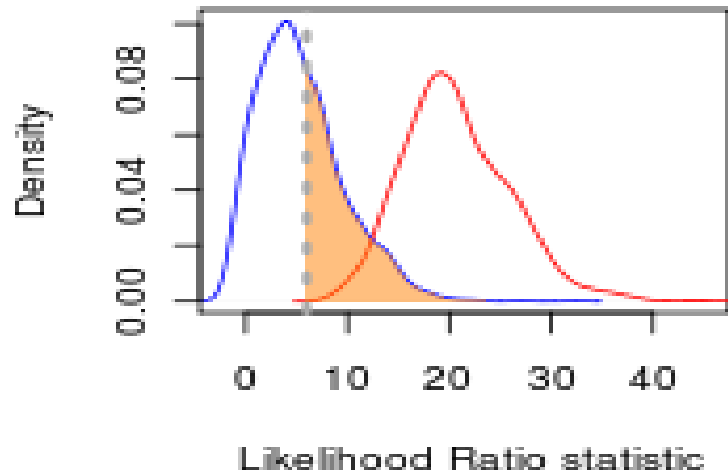
Preferring the simpler model isn't always a lack of power



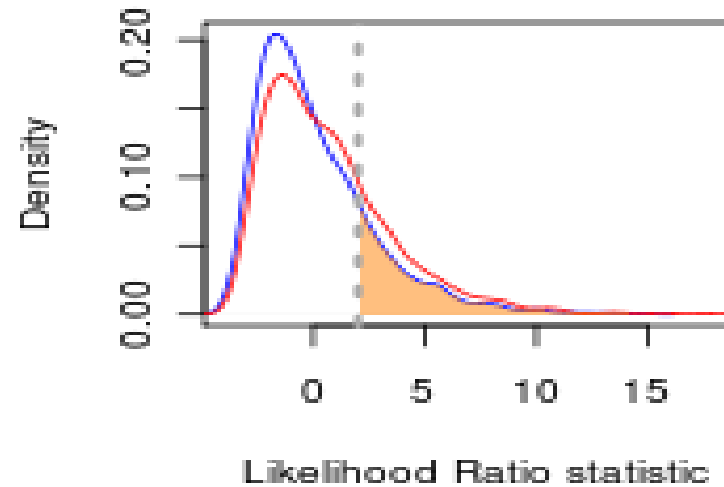
Avoids
over-fitting

Compared to what we already do?

BM vs OU.3

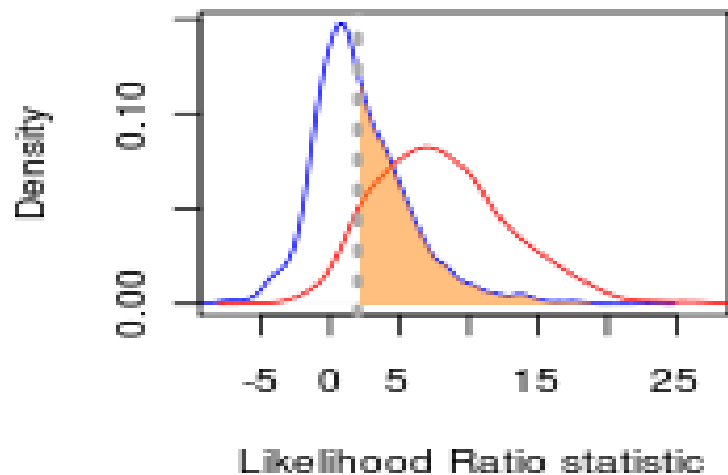


BM vs OU.1

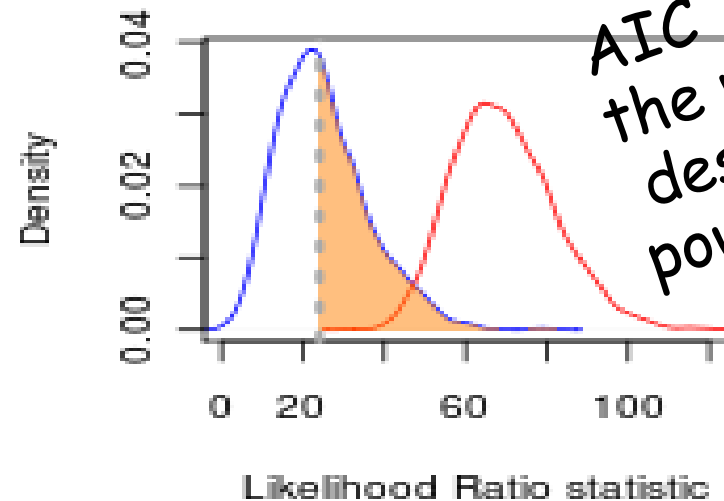


AIC has no
notion of
power

OU.3 vs OU.4



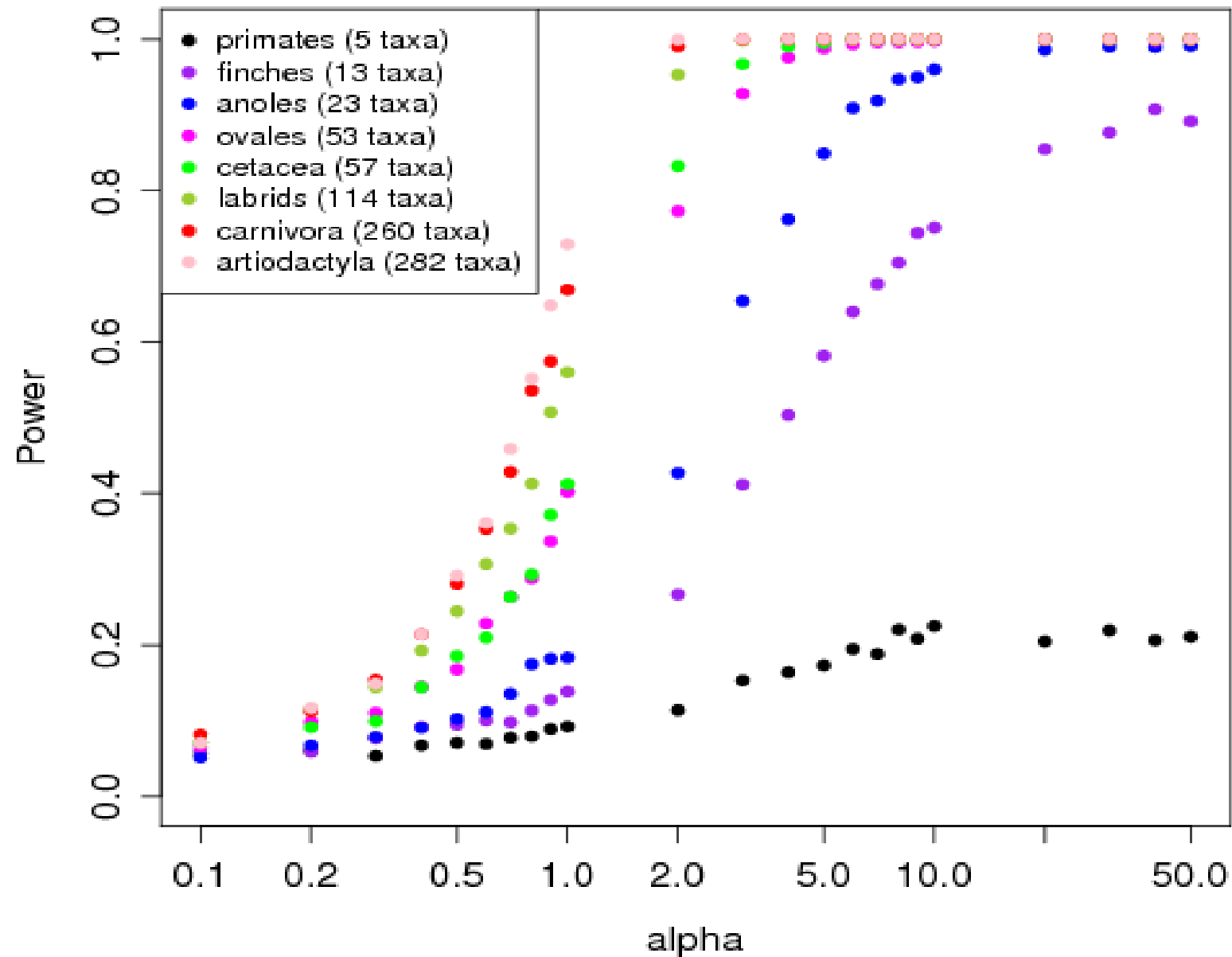
OU.3 vs OU.15



AIC chooses
the wrong model
despite good
power!

About as good as flipping a coin...

Larger trees can identify more subtle differences between models



Conclusions

- We've been worried about the right model. First we must worry if we have enough data to tell.
- Without this, we will often choose wrong models.
- Future lies in big trees!
- PMC: an R package for Phylogenetic Monte Carlo

<https://github.com/cboettig/Comparative-Phylogenetics/>

Acknowledgements

- Graham Coop
- Peter Ralph
- Wainwright Lab



