Phylogenetics

Beyond topology

Rachel Warnock, Laura Mulvey
rachel.warnock@fau.de, laura.l.mulvey@fau.de

September 7, 2022

Analytical Paleobiology Workshop, Erlangen 2022

Today's objectives

Understand the goal of phylogenetic comparative methods.

Models of trait evolution:

- Brownian motion
- Ornstein-Uhlenbeck

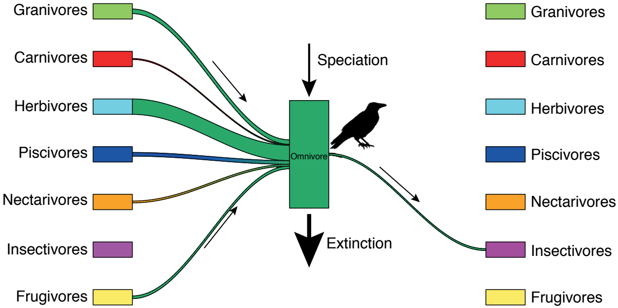
Onto the next adventure

Phylogenetic inference → estimating relationships between taxa

Phylogenetic comparative methods (PCMs) → to test hypotheses about characteristics or traits that drive the diversification of species

Phylogenetic comparative methods

Aim: to test hypotheses about characteristics or traits that drive the diversification of species.



• How did traits evolve through time? • What factors or traits drive speciation & extinction? • Why are some lineages more specious? • When and why do we see shifts in diversification? • What are the features of extinct ancestors?

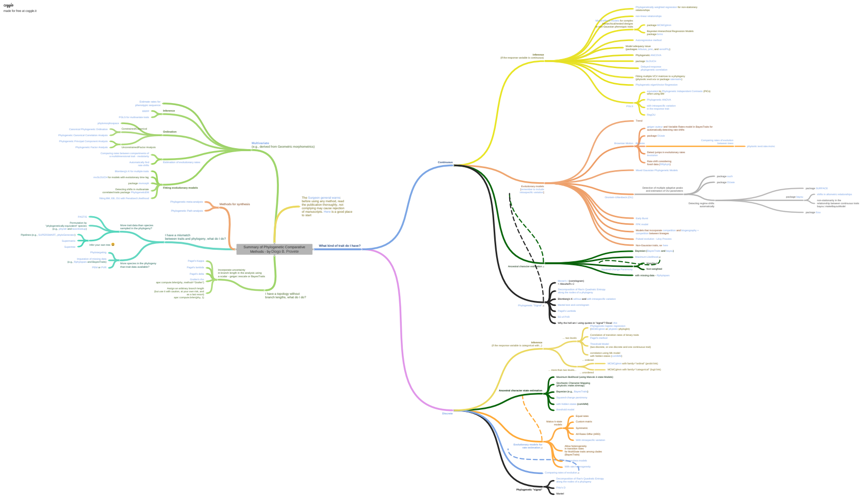
Cornwell & Nakagawa (2017) *Phylogenetic comparative methods*.

Phylogenetic comparative methods

There are two types of traits: discrete (e.g. biogeographic area) and continuous (e.g. body size).

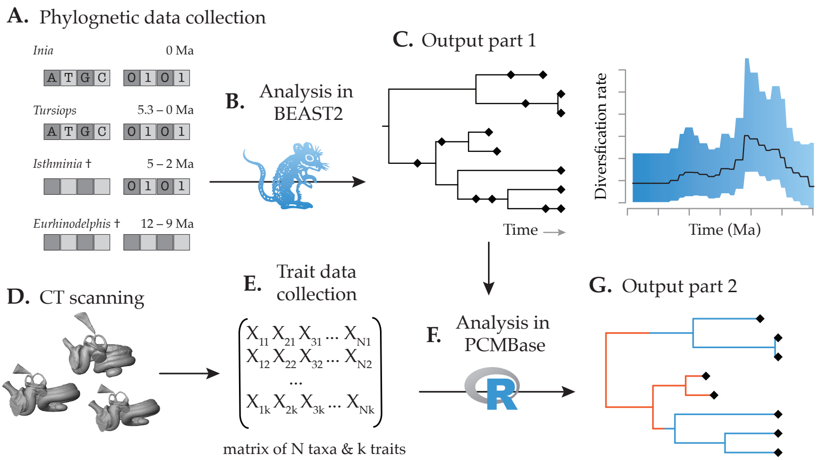
Next we’ll focus on continuous trait models.

There are now a huge number of PCMs



Explore Diogo Provete's infographic to learn more.

Where do phylogenetic comparative methods fit in?



Example project workflow combining phylogenetic inference and PCMs. It is becoming increasingly possible to do both simultaneously (and this might even be preferable).

Comparative methods

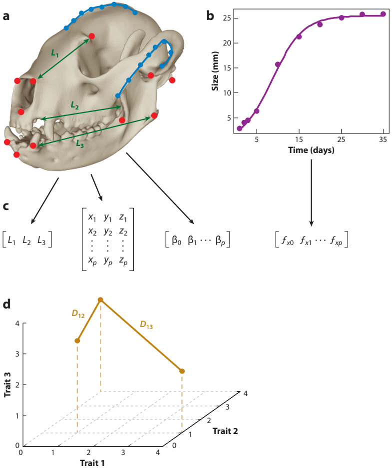


Photo Matt Artz Unsplash

Adams & Collyer (2019) *Annual Reviews Eco Evo Sys*. Phylogenetic comparative methods and the evolution of multivariate phenotypes

Phylogenetic non-independence

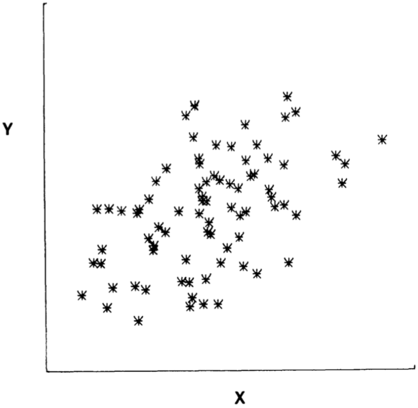


FIG. 6.—A typical data set that might be generated for the phylogeny in fig. 5 using the model of independent Brownian motion (normal increments) in each character.

Felsenstein (1985) *American Naturalist*

Phylogenetic non-independence

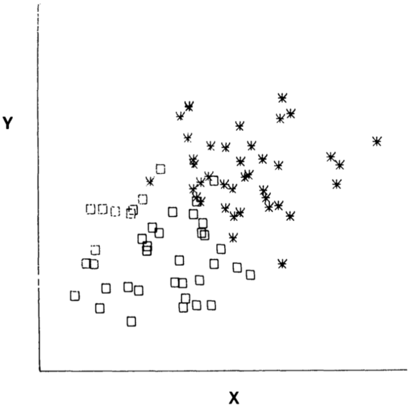
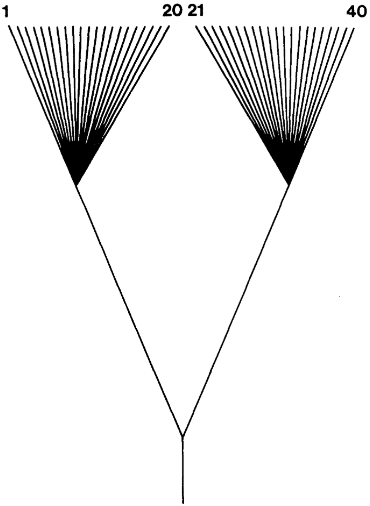


FIG. 7.—The same data set, with the points distinguished to show the members of the 2 monophyletic taxa. It can immediately be seen that the apparently significant relationship of fig. 6 is illusory.

Felsenstein (1985) *American Naturalist*

Phylogenetic non-independence



Felsenstein (1985) *American Naturalist*

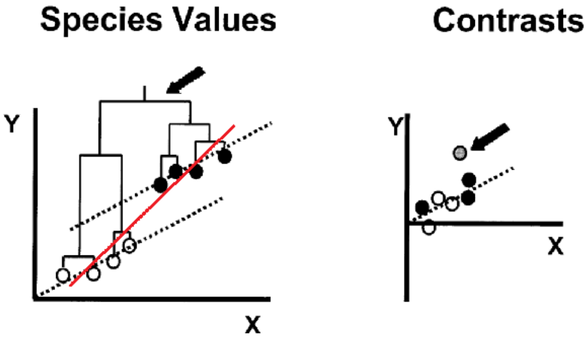
Taking phylogenetic non-independence into account

The problem: species in the same clade are more similar because of their shared evolutionary history.

Regression analysis assumes individual data points are statistically independent – this assumption is violated for species data.

Phylogenetic independent contrasts (PIC) and phylogenetic generalised least squares (PGLS) are two early approaches that take phylogenetic non-independent into account in the study of trait evolution.

Taking phylogenetic non-independence into account



The red line would be the slope of a standard least squares regression.

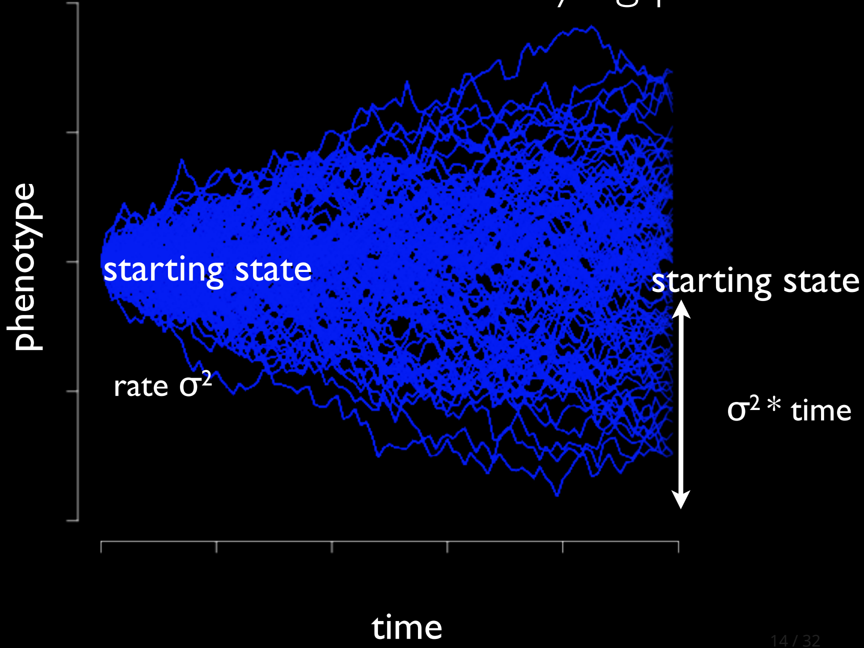
Image source: Laura Soul's PCM tutorial. Adapted from Nunn and Barton (2001).

Model's of continuous trait evolution applied to mammals

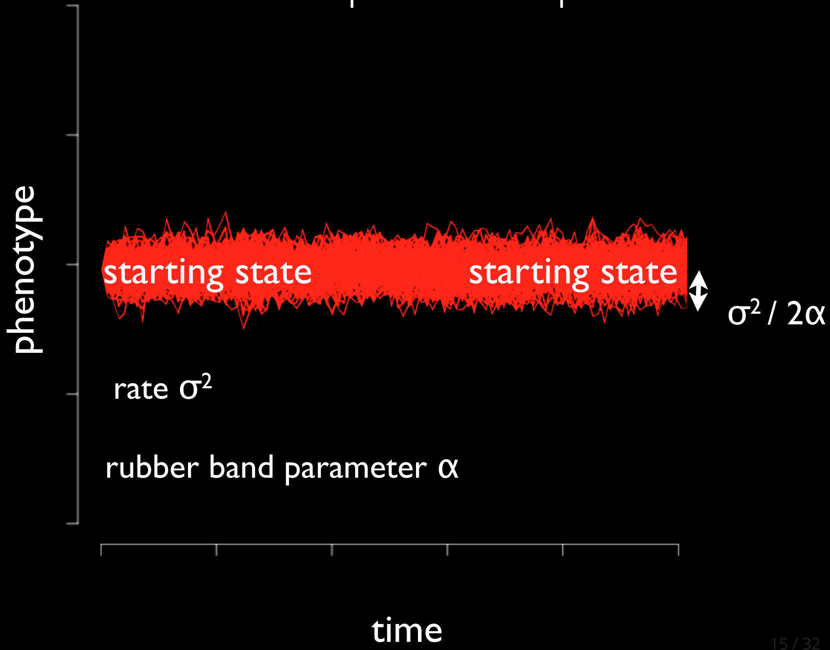
Slides borrowed from Graham Slater’s phylosemiar.

See also Slater (2013). *Methods Evolution & Ecology*.

Brownian motion is a diversifying process



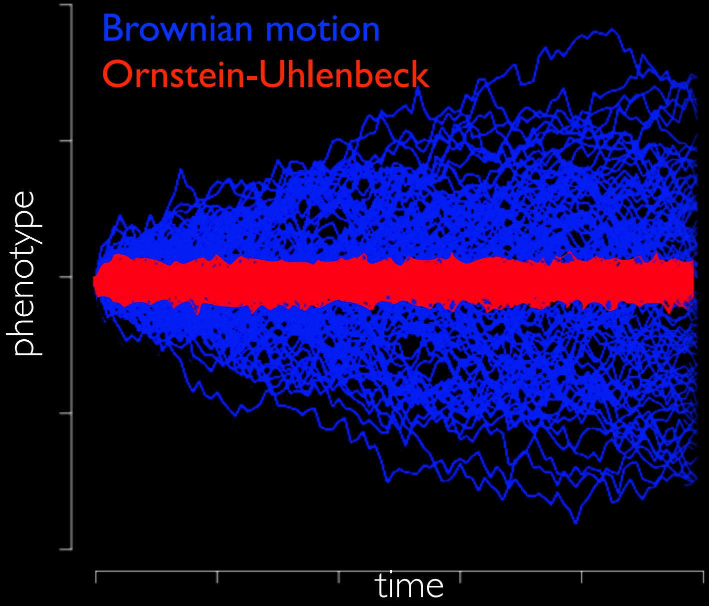
OU is an equilibrium process



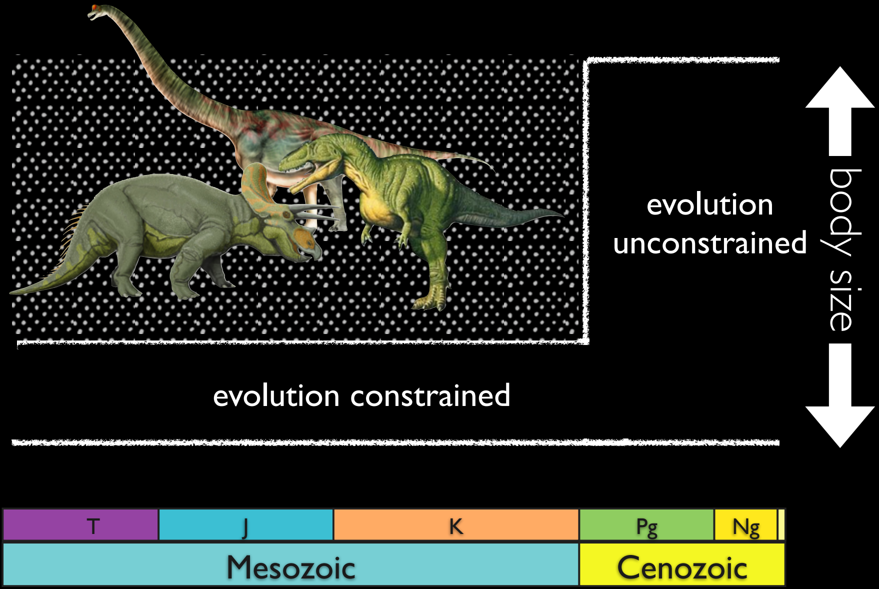
rate σ^2

rubber band parameter α

BM and OU simulated at the same rate give very different disparities



variation in mode



evolution unconstrained

evolution constrained

body size

3 paleo-motivated models for mammalian body size evolution

K-Pg Shift

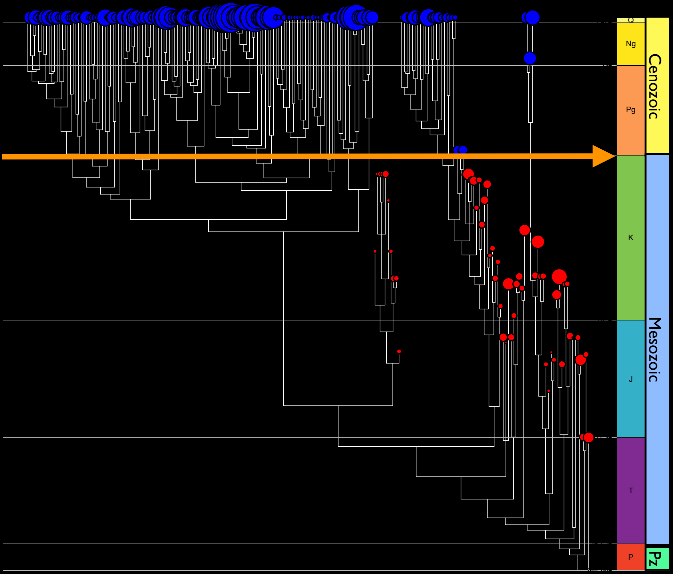
BM rate 1 BM rate 2

ecological release

Ornstein-Uhlenbeck BM

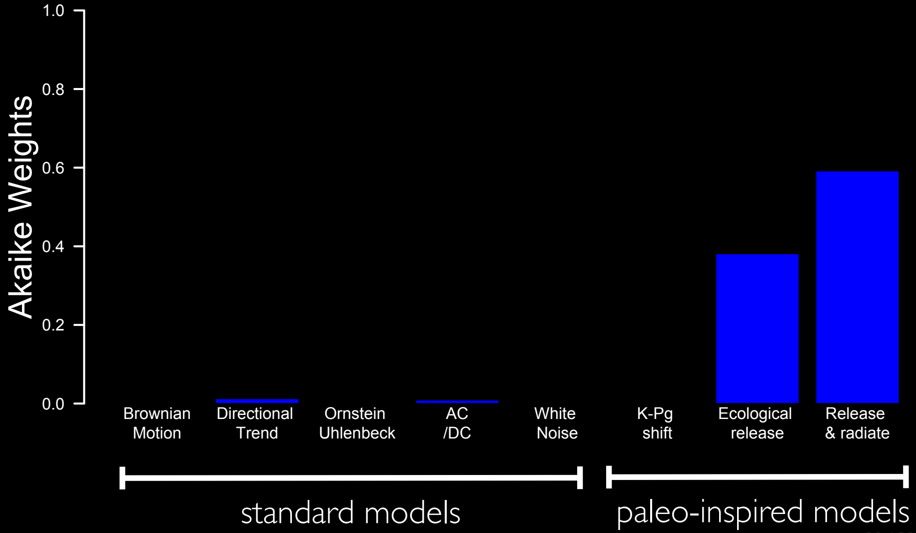
release and radiate

Ornstein-Uhlenbeck BM*



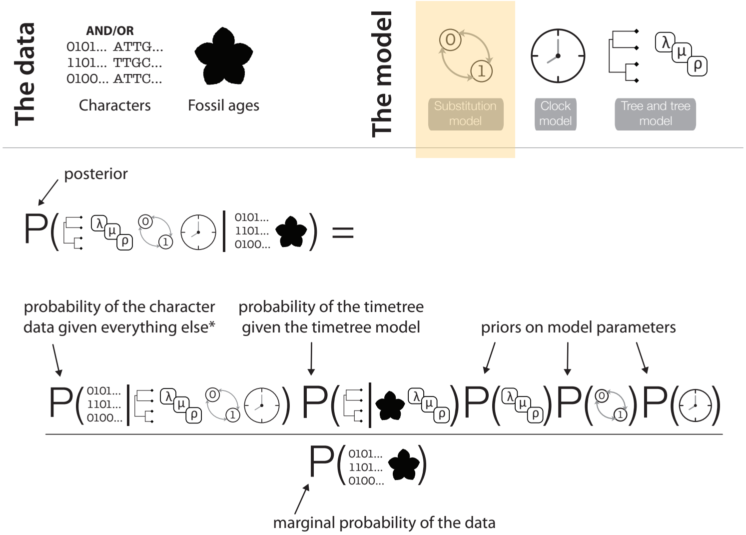
Slater (2013) *Methods Ecol. Evol.*

release & radiate fits best but ecological release is almost as good



Continuous trait models in Bayesian inference

Bayesian phylogenetic dating



The data

AND/OR
0101... ATTG...
1101... TTGC...
0100... ATTC...

Characters

Fossil ages

The model

Substitution model

Clock model

Tree and tree model

posterior

probability of the character data given everything else*

probability of the timetree given the timetree model

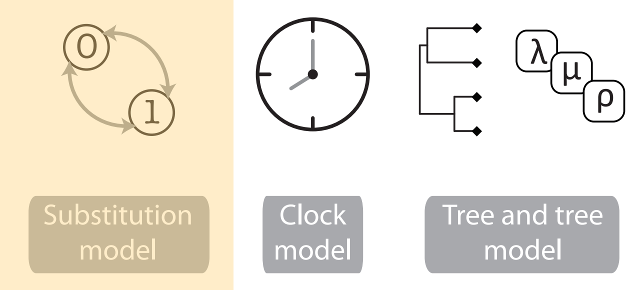
priors on model parameters

marginal probability of the data

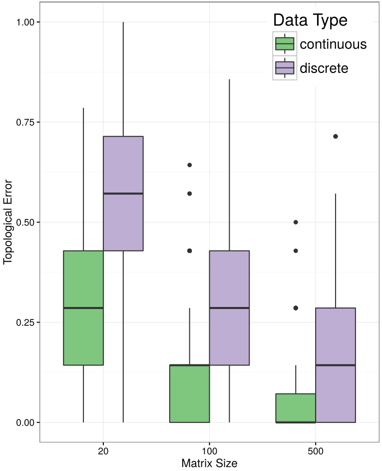
*the timetree, the parameters and the tripartite model

Recap: Bayesian phylogenetic dating requires three model components

- The **substitution model** ← describes how sites evolve over time.
- The **clock model** ← describes how evolutionary rates vary across the tree.
- The **tree model** ← describes how trees grow over time. Temporal evidence is included here.

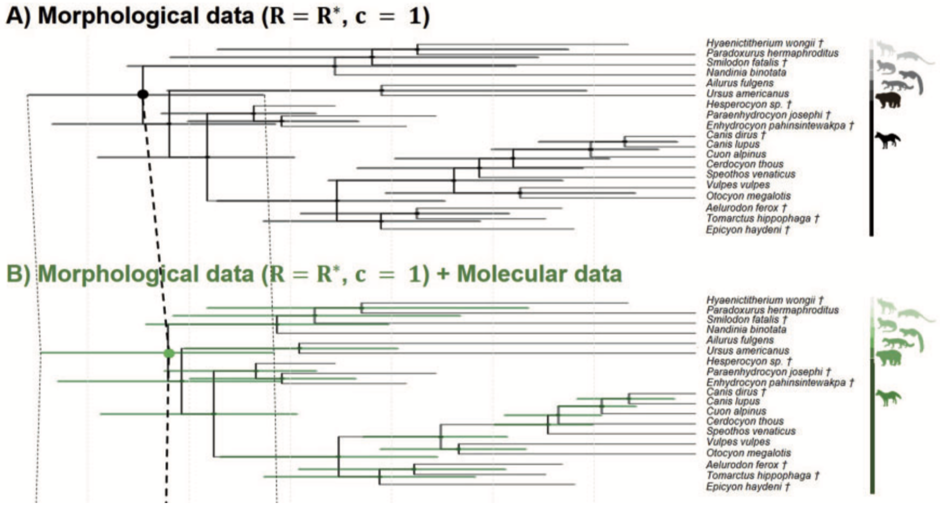


Continuous trait models can be used to build trees



Parins-Fukuchi (2018) *Sys Bio*. Use of continuous traits can improve morphological phylogenetics.

Continuous trait models can be used to build dated trees


A) Morphological data ($R = R^*$, $c = 1$)

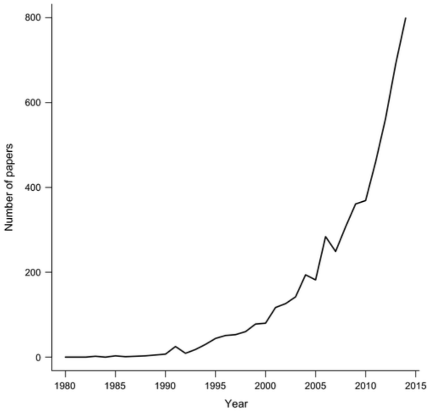
B) Morphological data ($R = R^*$, $c = 1$) + Molecular data

Álvarez et al. (2019) *Sys Bio*. Bayesian estimation of divergence times using correlated quantitative characters.

Summary of what we can do with continuous trait models

1. Test hypotheses about evolutionary process
2. Build trees
3. Date trees

The "dark side" of PCMs



Tree uncertainty, trait uncertainty, model uncertainty.

Limitations of methods are rarely addressed or well documented.

Cooper et al. (2016) *MEE*. Shedding light on the ”dark side” of phylogenetic comparative methods.

See also Cornwell & Nakagawa (2017) *Current Biology*. Phylogenetic comparative methods primer.

The "dark side" of PCMs

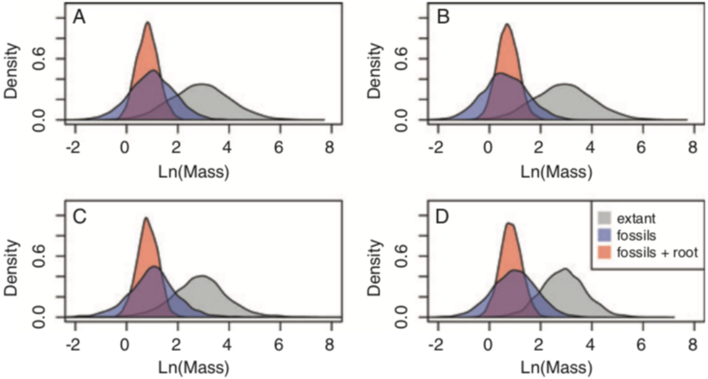
Make sure the method is appropriate for your question and data.

Try to understand the method assumption and limitations.

Don’t over-interpret your results.

Taking phylogenetic history into account often gives you more information than ignoring it.

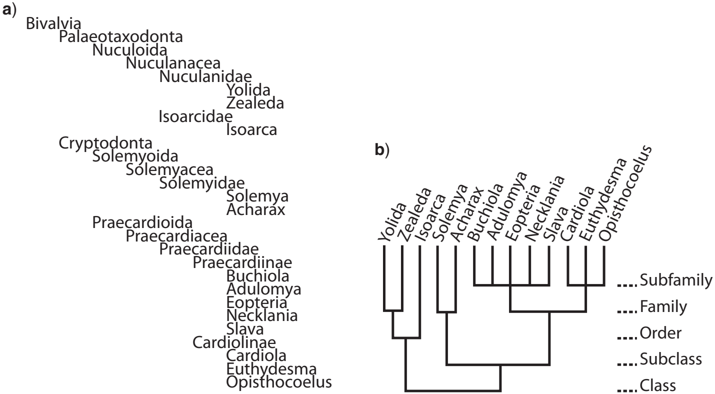
The bright side of PCMs?



Simulations show that incorporating fossil information improves our ability to distinguish among models of quantitative trait evolution using comparative data.

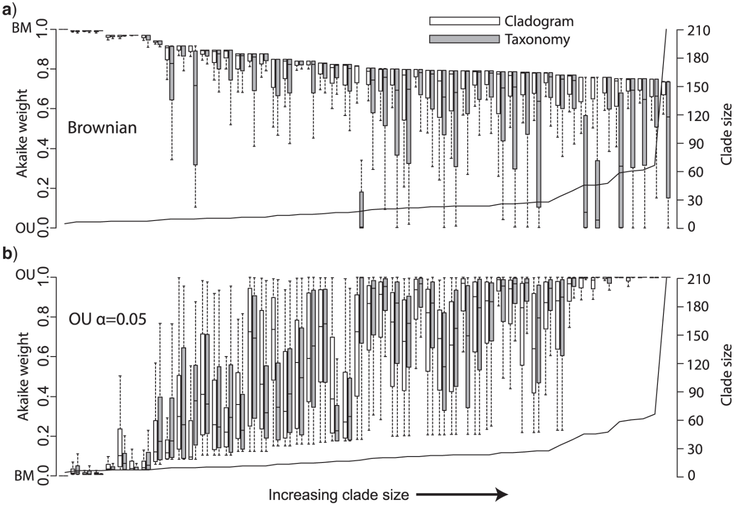
Slater et al. (2012). *Evolution*.

The bright side of PCMs?



Soul & Friedman (2015). *Sys Bio*. Taxonomy and phylogeny can yield comparable results in comparative palaeontological analyses.

The bright side of PCMs?



Soul & Friedman (2015). *Sys Bio*. Taxonomy and phylogeny can yield comparable results in comparative palaeontological analyses.

Take homes

Species as data points are not independent.

PCMs provide a statistical approach for studying the evolution of traits in a phylogenetic framework (among other things).

PCMs also have a dark side — they are statistical methods after all!

Further reading

The PCM community it very R centric!

Laura Soul & David Wright have an excellent introductory tutorial for doing PCMs using R.

Luke Harmon has a brilliant online book all about PCMs.

Check out Graham Slater’s phylosemiar to learn more about fossils and PCMs.