

Ancestral State Estimation

Ancestral State Estimation

Many macroevolutionary questions require inferences on past states:

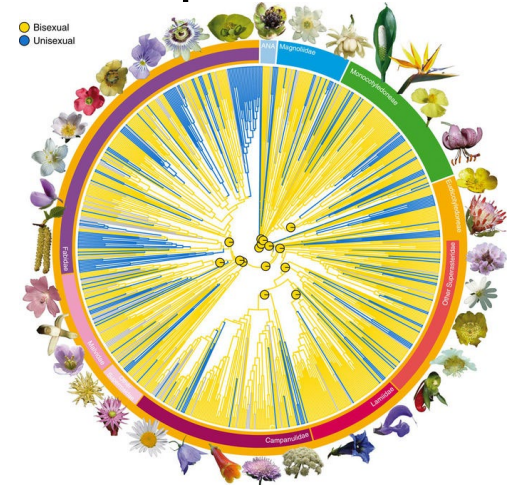
What was the structure of the earliest flower?

Did early turtles have genotypic or temperature-dependent sex determination?

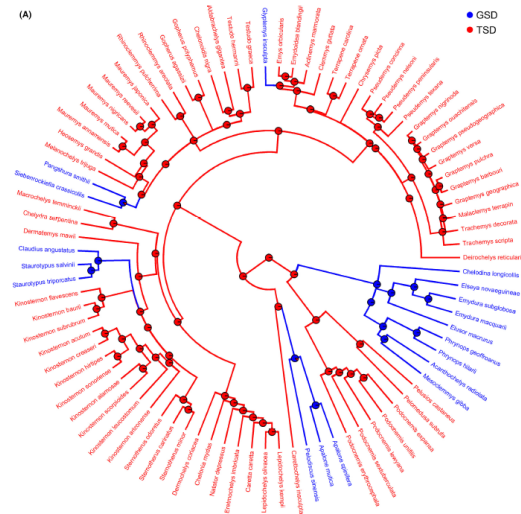
What was the ‘direction’ of trait change through time?

Addressing such questions requires estimating ancestral character states*

*Note: this is *NOT* ancestral state reconstruction. We are estimating past states, not reconstructing with certainty



Sauquet et al. *Nature Comm.* (2017)

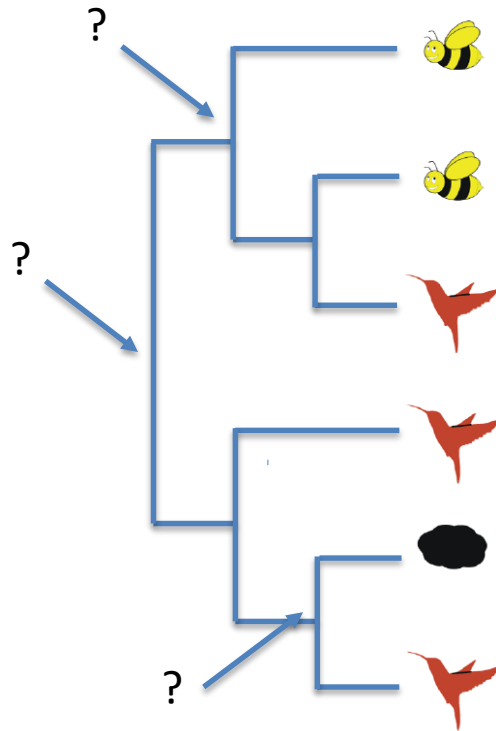


Sebath et al. *Ecol. & Evol.* (2016)

Ancestral State Estimation

The Goal

-Utilize phylogeny and extant data to infer past states



From Joy et al. *PLoS ONE*. (2016)

-Incorporating fossil information (when available) is important!

Outline

-Discrete Data Methods

- Parsimony
- Maximum Likelihood
- Bayesian Approaches (stochastic character mapping)

-Continuous Data Methods

- Maximum Likelihood
- Squared Change Parsimony
- Bayesian Approaches

-Incorporating Fossil Information

Discrete Data: Maximum Parsimony

- Maximum Parsimony

- Estimate ancestral states that minimize trait change on phylogeny

- We want the simplest model possible to explain the data

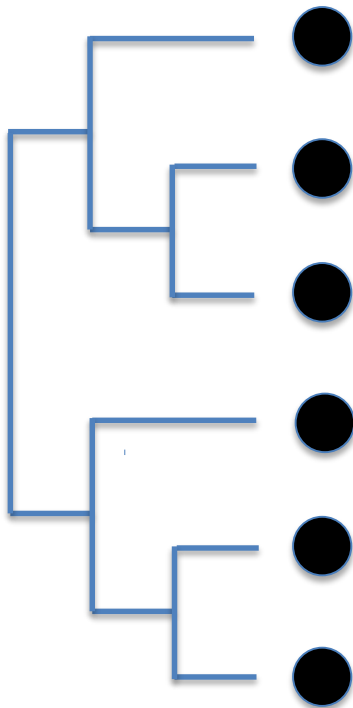
- Nodal values representing the fewest trait changes to ‘map extant trait values on the phylogeny embody this

- Thus, estimating ancestral states such that trait changes are minimized is a ‘parsimonious’ solution

Discrete Data: Maximum Parsimony

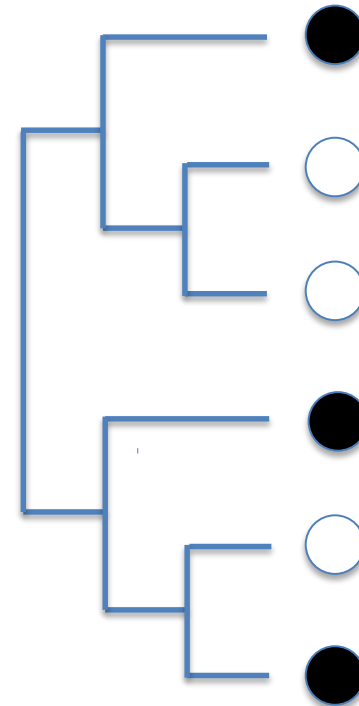
-Best embodied by simple example:

-Possible states: Black/ White



Ancestral Estimates?

● for all nodes



But what about now?

Need an algorithm

Discrete Data: Maximum Parsimony

-Parsimony algorithm

- ‘Double pass’: traverse tips \rightarrow root, then root \rightarrow tips
 - Requires rules to assign and evaluate nodal values

Wagner Parsimony (assumes equally weighted unordered characters)

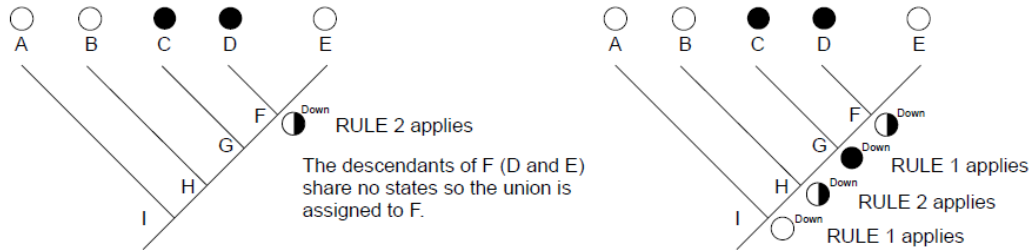
- 1: If descendant nodes share states, assign state to ancestor
- 2: If descendants do not share states, assign ancestor the union of states

Will result in both unambiguous and equivocal ancestor state estimates

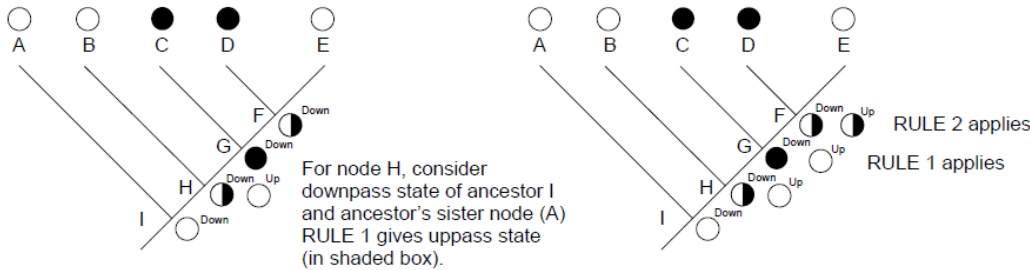
Discrete Data: Maximum Parsimony

-Example Wagner Parsimony

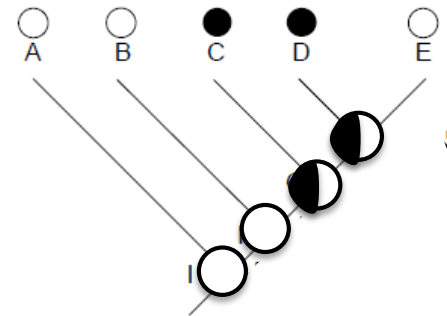
(1) **Downpass optimization:** proceed 'down' the tree towards the root, optimizing each ancestral node.



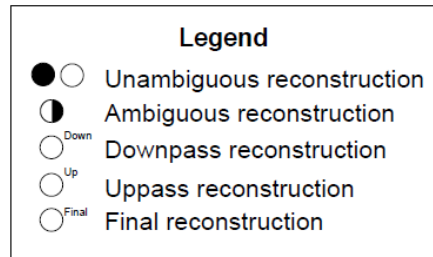
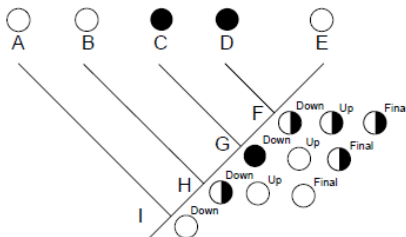
(2) **Uppass optimization:** proceed 'up' the tree away from the root, optimizing each ancestral node.



(3) **Final optimization**

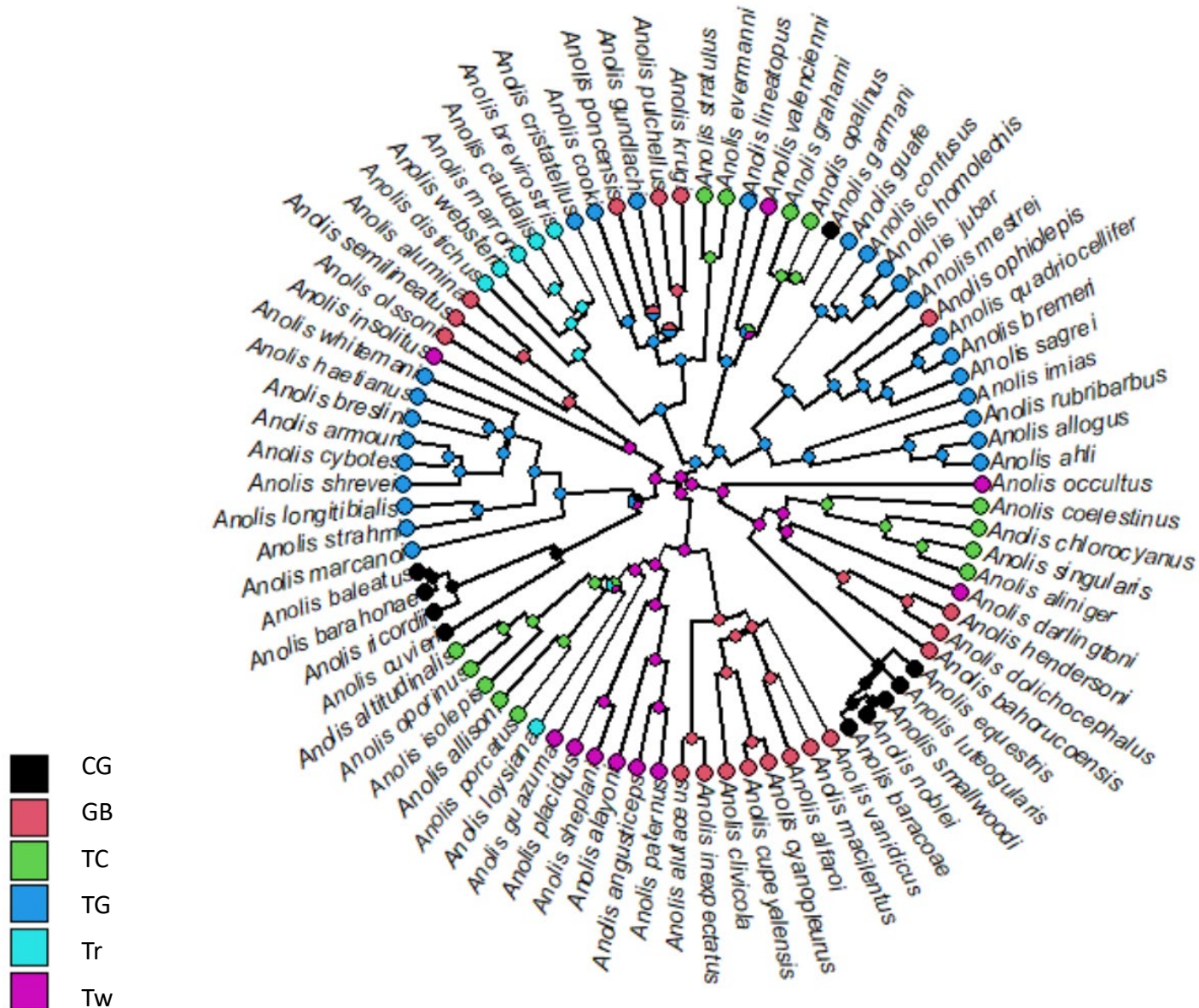


(3) **Final optimization**



Discrete Data: Maximum Parsimony

-Example: Habitat use in *Anolis* lizards



Discrete Data: Maximum Parsimony

- Conclusions

- Advantages

 - Quick, intuitive, and straight-forward

- Disadvantages

 - ‘Cost’ equal for all states (i.e., equal weighting of state changes)

 - Assumes equal transitions among states ($q_{A \rightarrow B} = q_{B \rightarrow A}$)

 - Branch length information ignored!!!

Upshot: Parsimony does not optimize (or even have!) an evolutionary model of trait change: it is a simple rule-set

Discrete Data: Parsimony to Likelihood

-Extensions

-Other parsimony algorithms have been proposed

-e.g., unequal state change weighting (e.g, $A \rightarrow B$ 3X cost of $B \rightarrow A$)

Better, but still ad-hoc

What weights should be used?

How to select them?

Branch lengths still ignored

NOTE: Such adjustments *try* to get at a formal model of trait evolution, so why not just use such a model?

That leads us to likelihood

Discrete Data: Maximum Likelihood

-Estimate ancestral states by maximizing the probability of the data given the phylogeny, a model of evolution, and anc. states:

$$\Pr(\mathbf{X} \mid \tau, \theta, \mathbf{A})$$

$$\mathcal{L}(\theta, \mathbf{A}) = \Pr(\mathbf{X} \mid \tau, \theta, \mathbf{A})$$

Need the data (\mathbf{X}), the phylogeny (τ), and a model (θ)

Simultaneously estimate model parameters and ancestral states via ML

Discrete Data: Maximum Likelihood

-A Markov process used as model for character state changes

-Recall transition rates for binary traits (Pagel 1994)*

$$\theta = \mathbf{Q}(t) = \begin{bmatrix} (1 - q_{01})dt & q_{01}dt \\ q_{10}dt & (1 - q_{10})dt \end{bmatrix}$$

Thus, for the tips data (X): $\mathcal{L}(\theta) = \Pr(X | \tau, \theta)$

-Now use \mathbf{Q} to estimate ancestor states ($A_1 - A_{N-1}$)

-Yields the likelihood of the data conditioned on the phylogeny, the model, and the ancestral states:

$$\mathcal{L}(\theta, A) = \Pr(X | \tau, \theta, A)$$

Use search algorithm to maximize $\mathcal{L}(\theta, A)$ (see Schluter et al. 1997; Pagel 1999)

*Method works for multi-state discrete, and also continuous characters (see below)

-Calculate probabilities of ancestral states (A) given the data (D), phylogeny (τ), and model (θ)

$$\Pr(A|X, \tau, \theta) = \frac{\Pr(X | \tau, \theta, A) \Pr(A | \tau, \theta)}{\Pr(X | \tau, \theta)}$$

Obtain posterior distributions for A & θ via MCMC

Both Empirical and Hierarchical Bayesian implementations exist

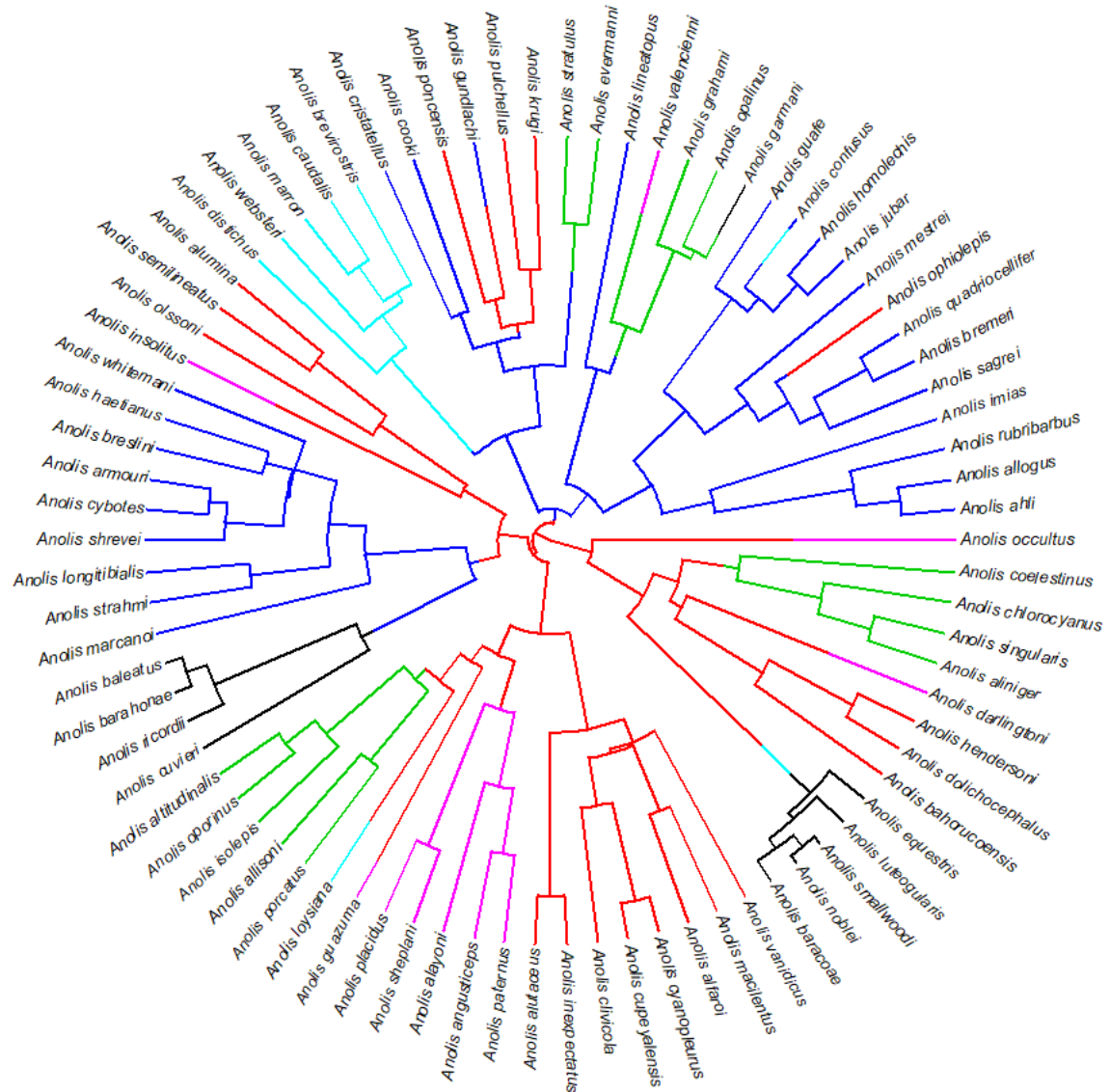
Discrete Data: Bayesian Stochastic Mapping

- Stochastic character mapping one approach
- Sample possible character histories from posterior under MCMC
- Summary across many runs provides ancestral estimates

Discrete Data: Bayesian Stochastic Mapping

-Example: Habitat use in *Anolis* lizards

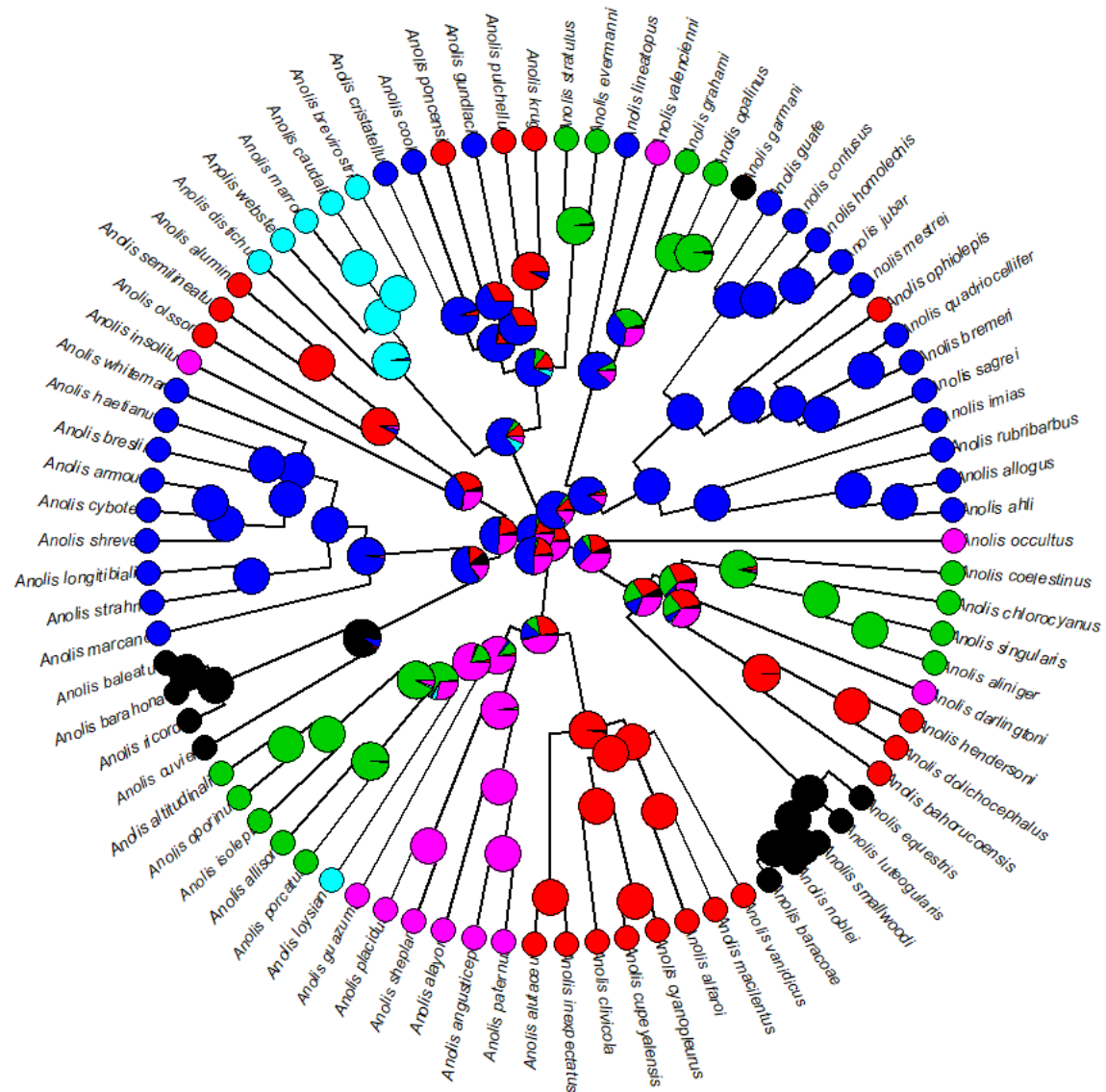
One run: Not overly useful



Discrete Data: Bayesian Stochastic Mapping

-Example: Habitat use in *Anolis* lizards

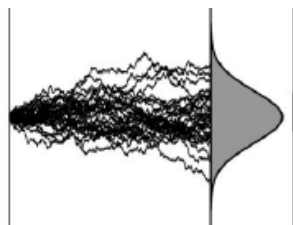
From 100 runs. MUCH more useful



Continuous Data: Maximum Likelihood

-Recall that Brownian motion embodies the Markov process for continuous traits

Instead of rate matrix \mathbf{Q} , we need a rate parameter σ^2 which describes instantaneous changes in X



$$dX(t) = \sigma^2(t)$$

Can formulate the likelihood conditioned on the phylogeny, the model, and the ancestral states:

$$\mathcal{L}(\theta, A) = \Pr(\mathbf{X} \mid \tau, \theta, A)$$

Use search algorithm to maximize $\mathcal{L}(\theta, A)$ (see Schluter et al. 1997)

Continuous Data: Squared Change Parsimony

-Another approach: find ancestral values that minimize the sum-of-squared trait changes along the branches of the phylogeny: squared change parsimony (SCP)

SS are weighted inversely by branch lengths

$$SS = \sum_{i=1}^{N-1} \frac{(x_i - x_{i'})^2}{v_i}$$

Where x_i & $x_{i'}$ are trait values for nodes at opposite end of branch with branch length: v_i

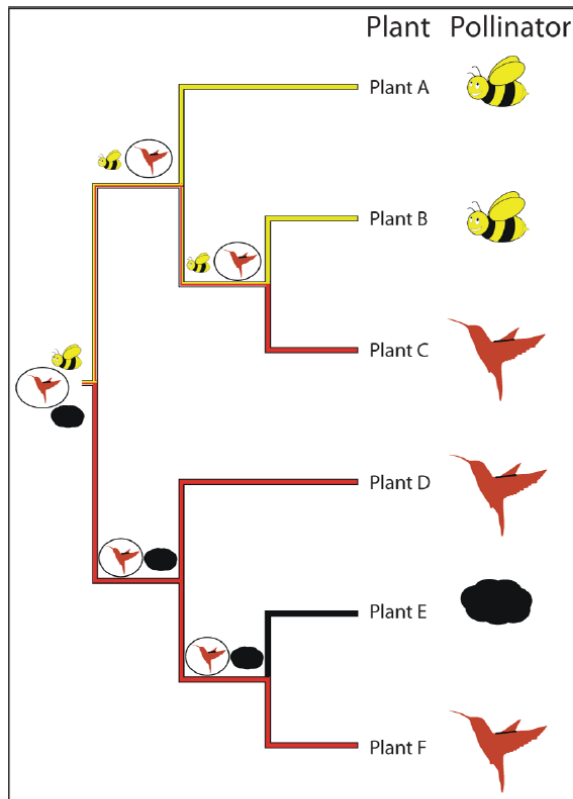
Also, GLS model could be used (Martins and Hansen 1997)

ML, GLS, and SCP should yield equivalent estimates

Ancestral States: Incorporating Fossils

-Incorporating fossil information in any of the above improves estimates

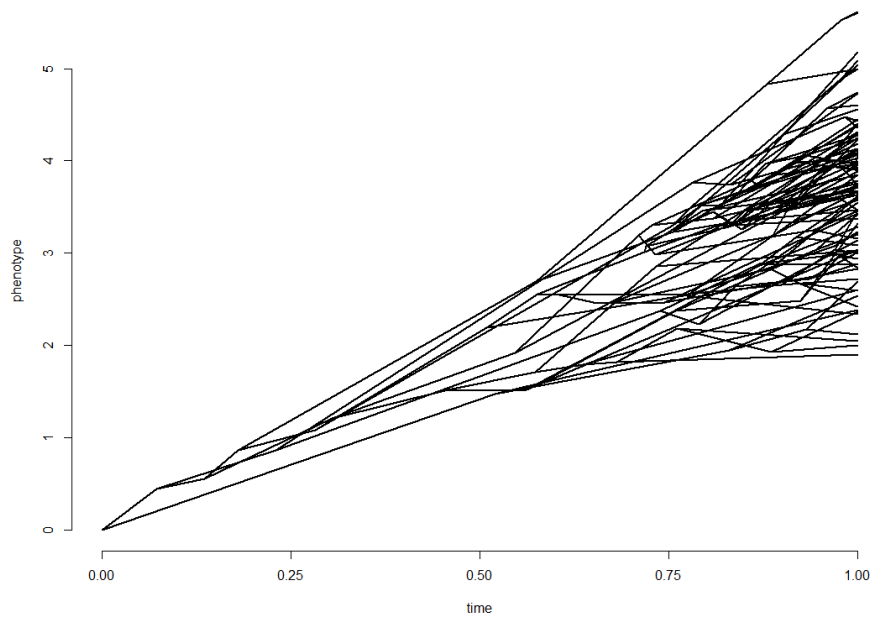
Example: Here MP anc. states equivocal *UNLESS* fossil information included



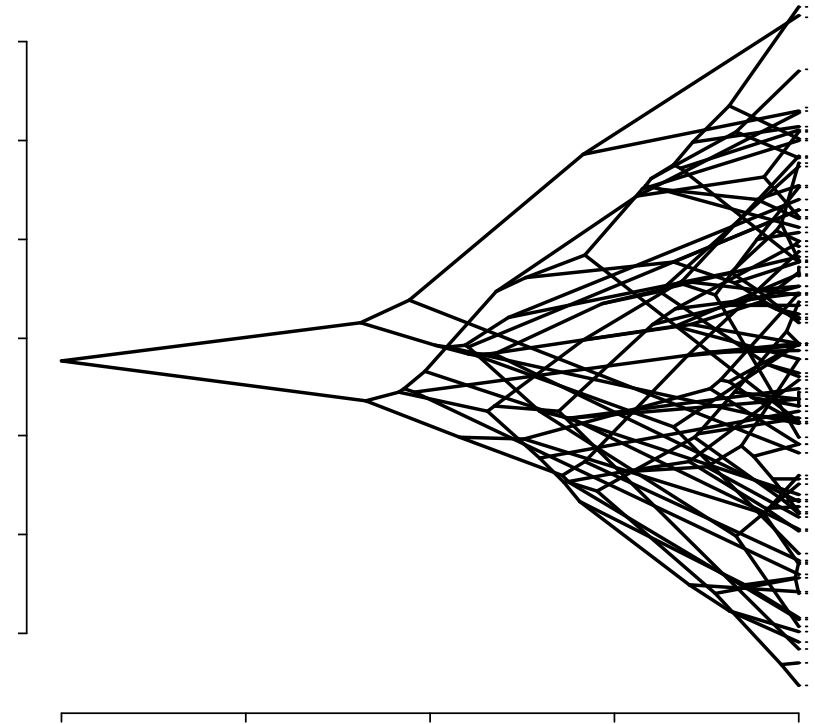
NOTE: Assigning 'hummingbird' as root state (through e.g., **FOSSIL KNOWLEDGE**) allows final nodal assignment: (see Joy et al. 2016)

Ancestral States: Challenges

-If one's actual data has a trend, ancestral states will be misleading



Data Simulated with Trend

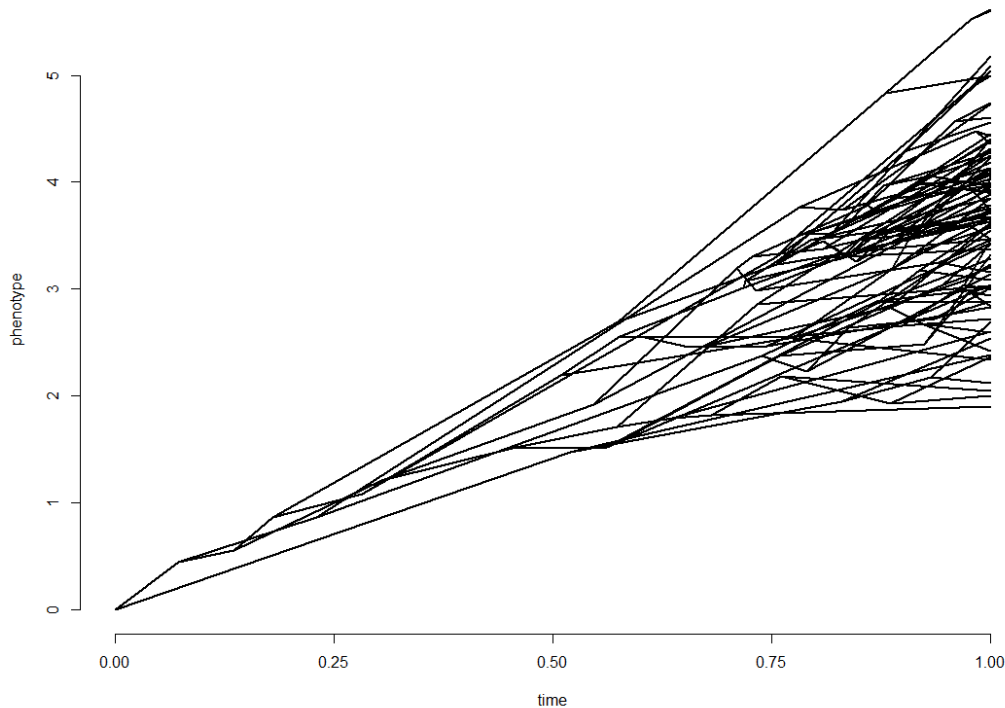


Estimated Ancestral States

Ancestral States: Incorporating Fossils

-Fossil information may be incorporated in ML and Bayesian approaches to improve estimation

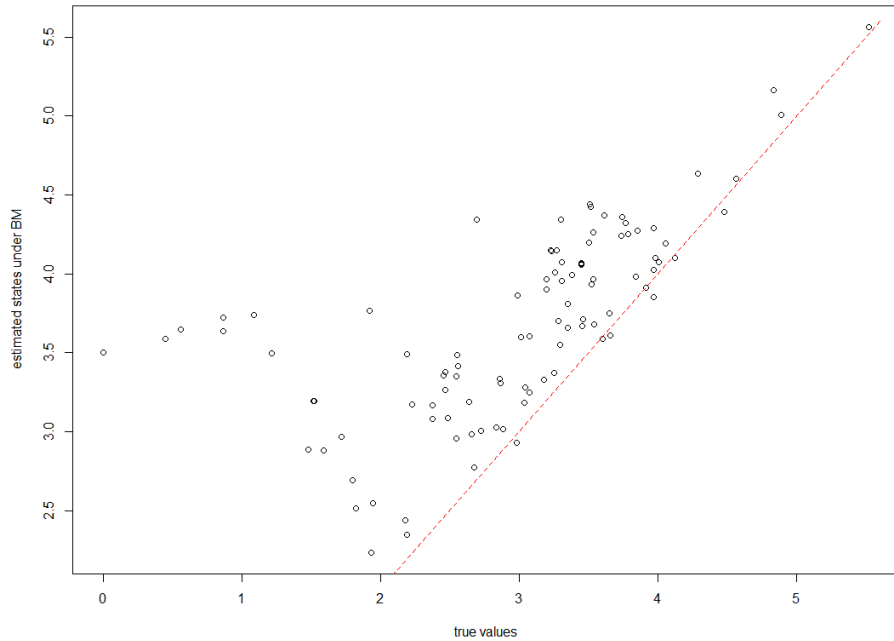
Simulated example (data simulated with trend):



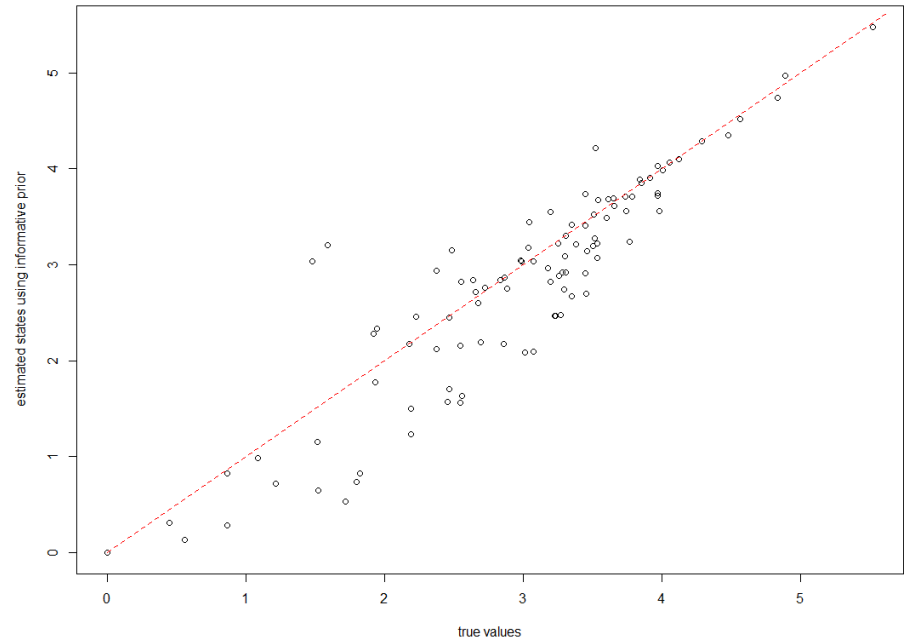
Ancestral States: Incorporating Fossils

-Correlation of actual vs. estimated ancestors with and without fossil information (Bayesian estimation fixing 2 nodes)

estimated without prior information



estimated using informative prior



Estimation MUCH better with fossils!

Conclusions

- ML and Bayesian methods for both discrete and continuous data

- Discrete Data Methods

 - Maximum Likelihood

 - Bayesian Approaches (stochastic character mapping)

- Continuous Data Methods

 - Maximum Likelihood = Squared Change Parsimony = GLS

 - Bayesian Approaches

- Incorporating fossil information can be critically important