

# Model complexity and adequacy for phylogenetic comparative methods

---

Luke J Harmon  
University of Idaho





*We let mornings and evenings  
go by and just waited.*





# Model complexity

---

# Model adequacy

---

# Implementation

---

# Model complexity

---

rjMCMC

$(\alpha_1 \alpha_1 \alpha_1 \alpha_1 \alpha_1 \alpha_1 \alpha_1 \alpha_1 \alpha_1 \alpha_1)$   $k = 1$



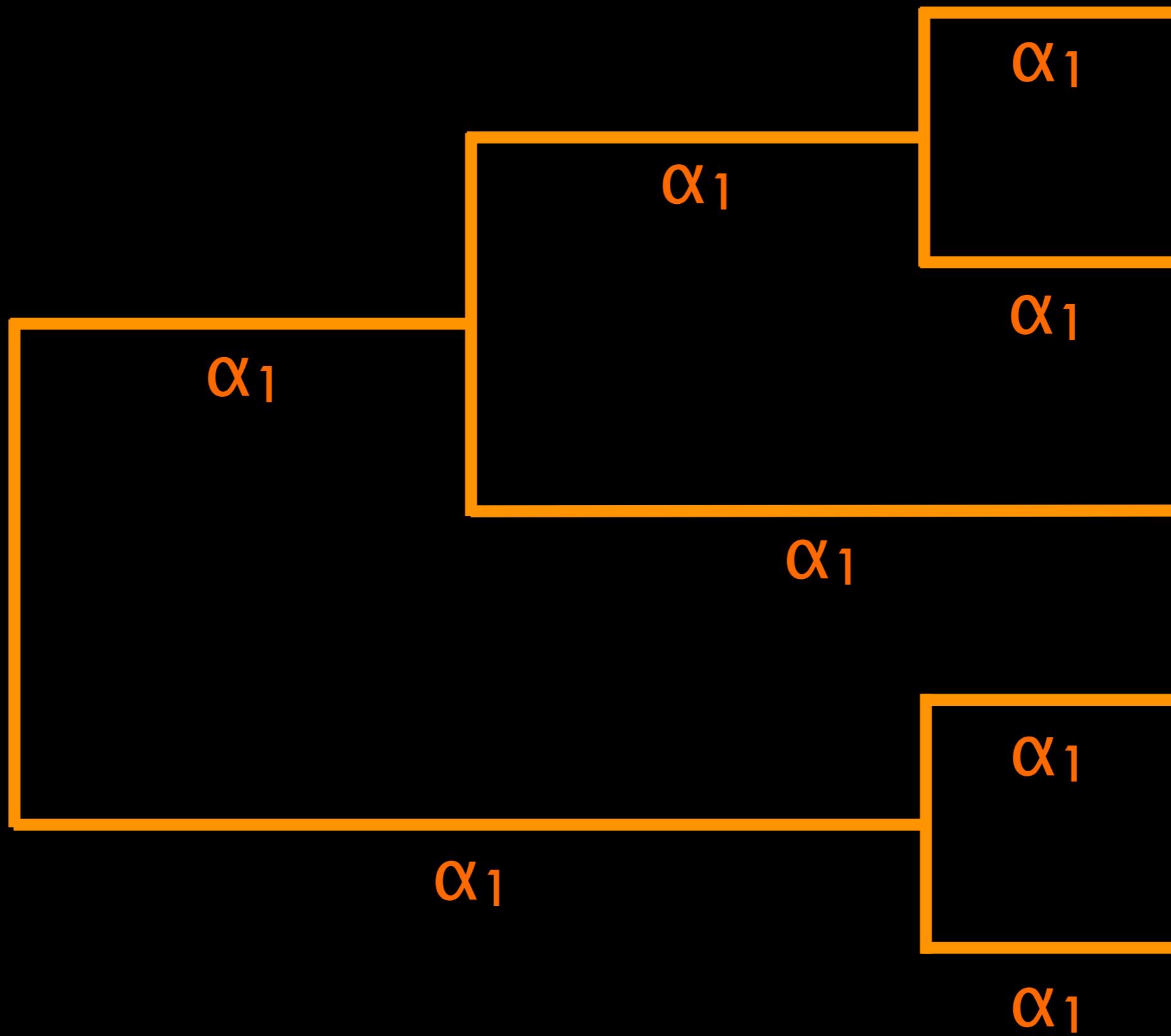
$(\alpha_1 \alpha_1 \alpha_1 \alpha_1 \alpha_1 \alpha_2 \alpha_2 \alpha_2 \alpha_2 \alpha_2)$   $k = 2$

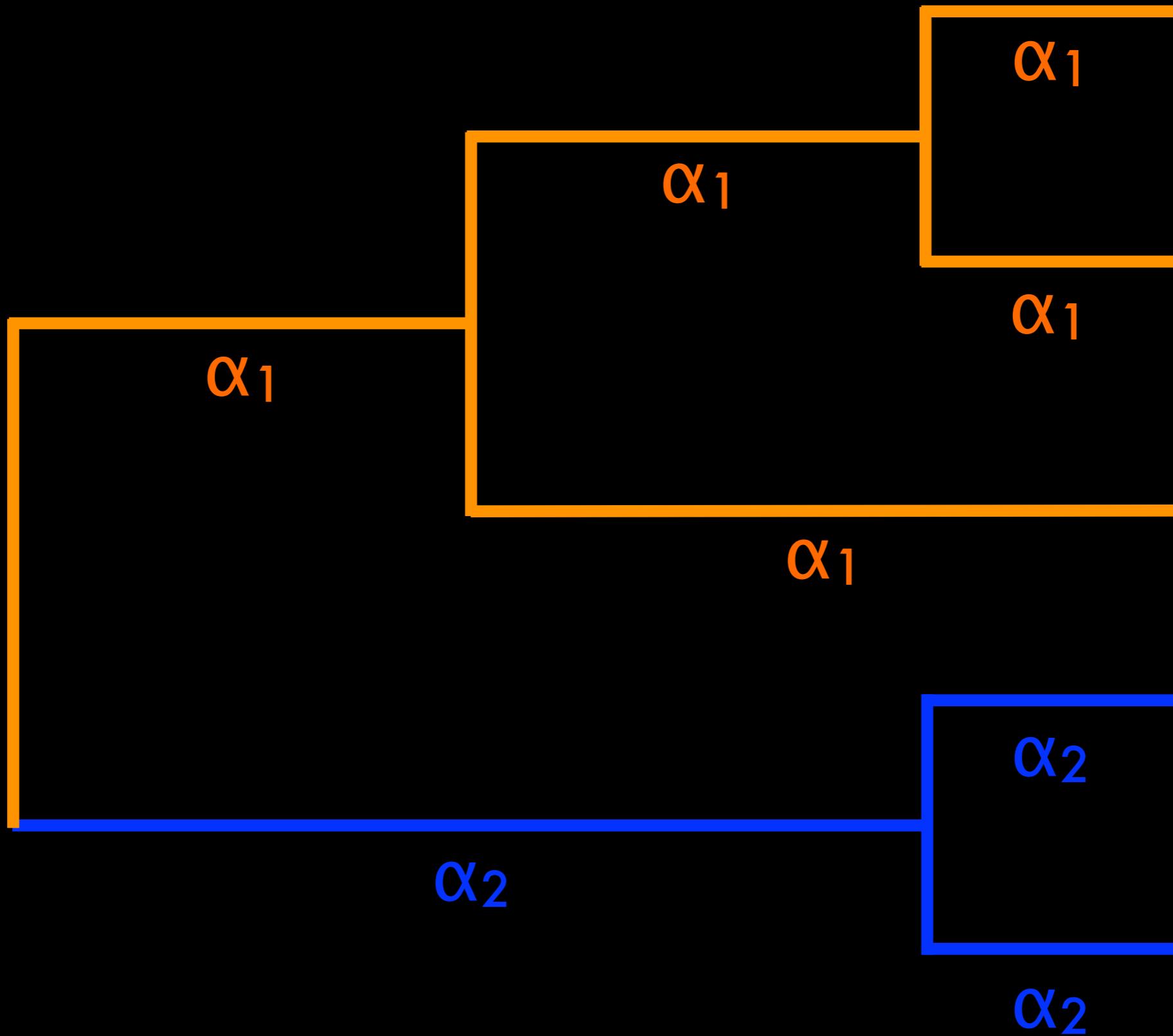


$(\alpha_3 \alpha_3 \alpha_3 \alpha_1 \alpha_1 \alpha_2 \alpha_2 \alpha_2 \alpha_2 \alpha_2)$   $k = 3$



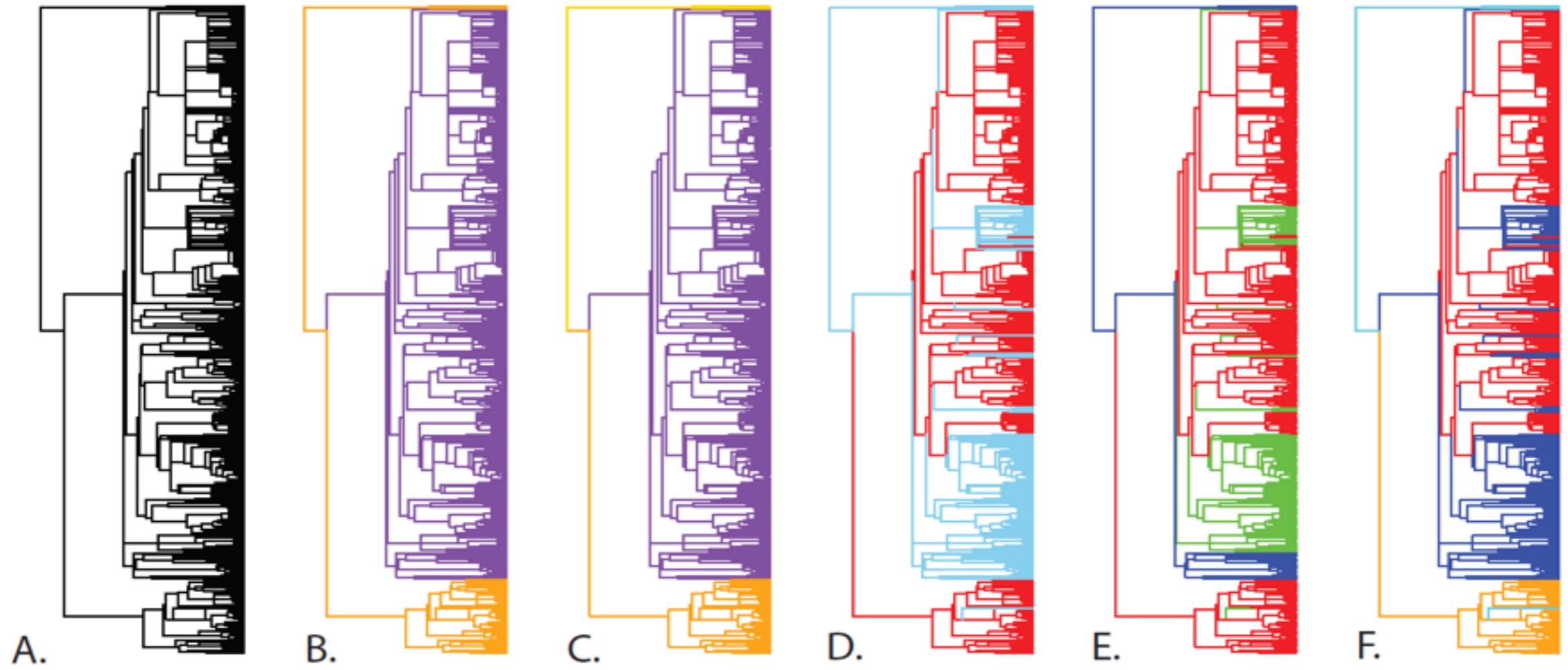
$(\alpha_3 \alpha_3 \alpha_3 \alpha_1 \alpha_1 \alpha_1 \alpha_1 \alpha_1 \alpha_1)$   $k = 2$





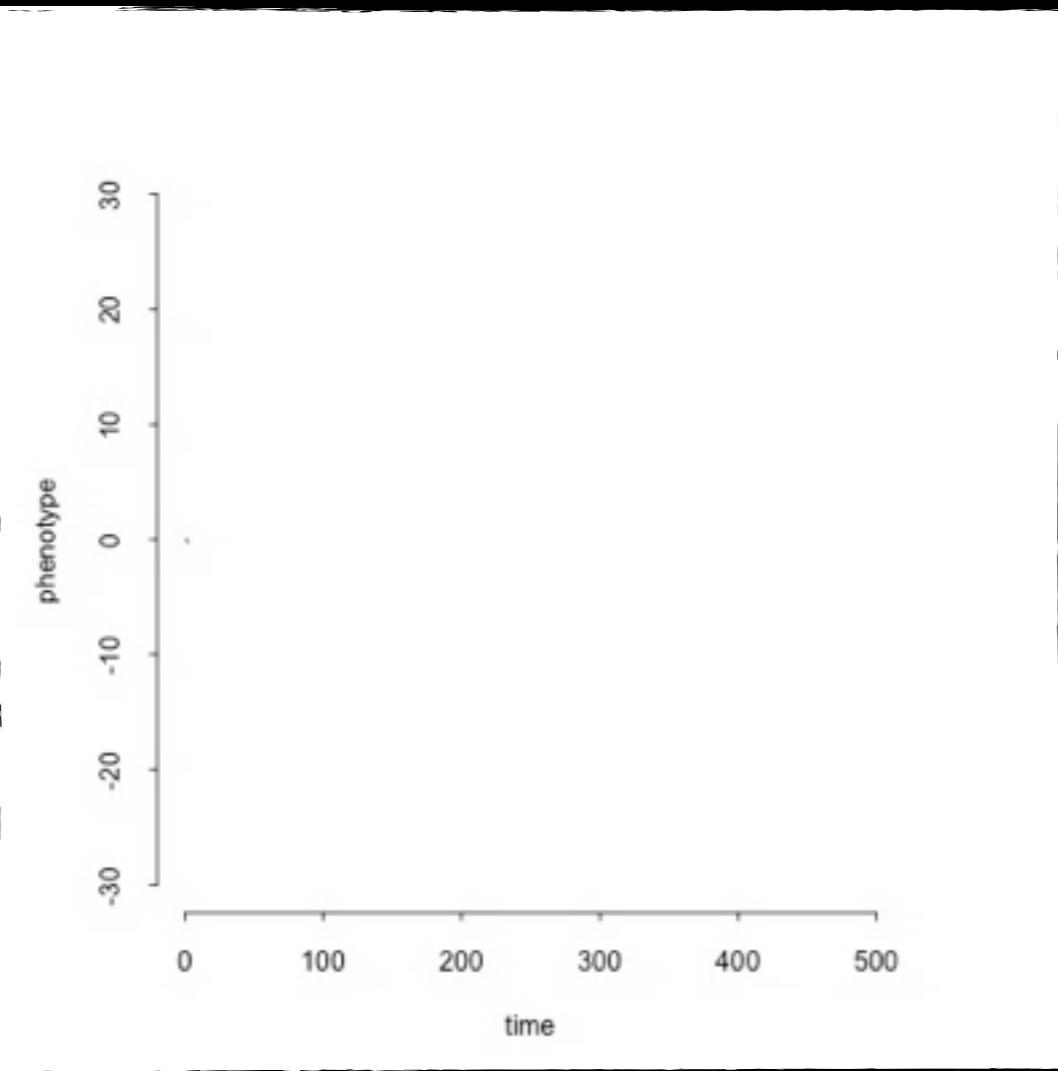
Through a series of **split** and **merge** moves

we can fit **complex models** to comparative data

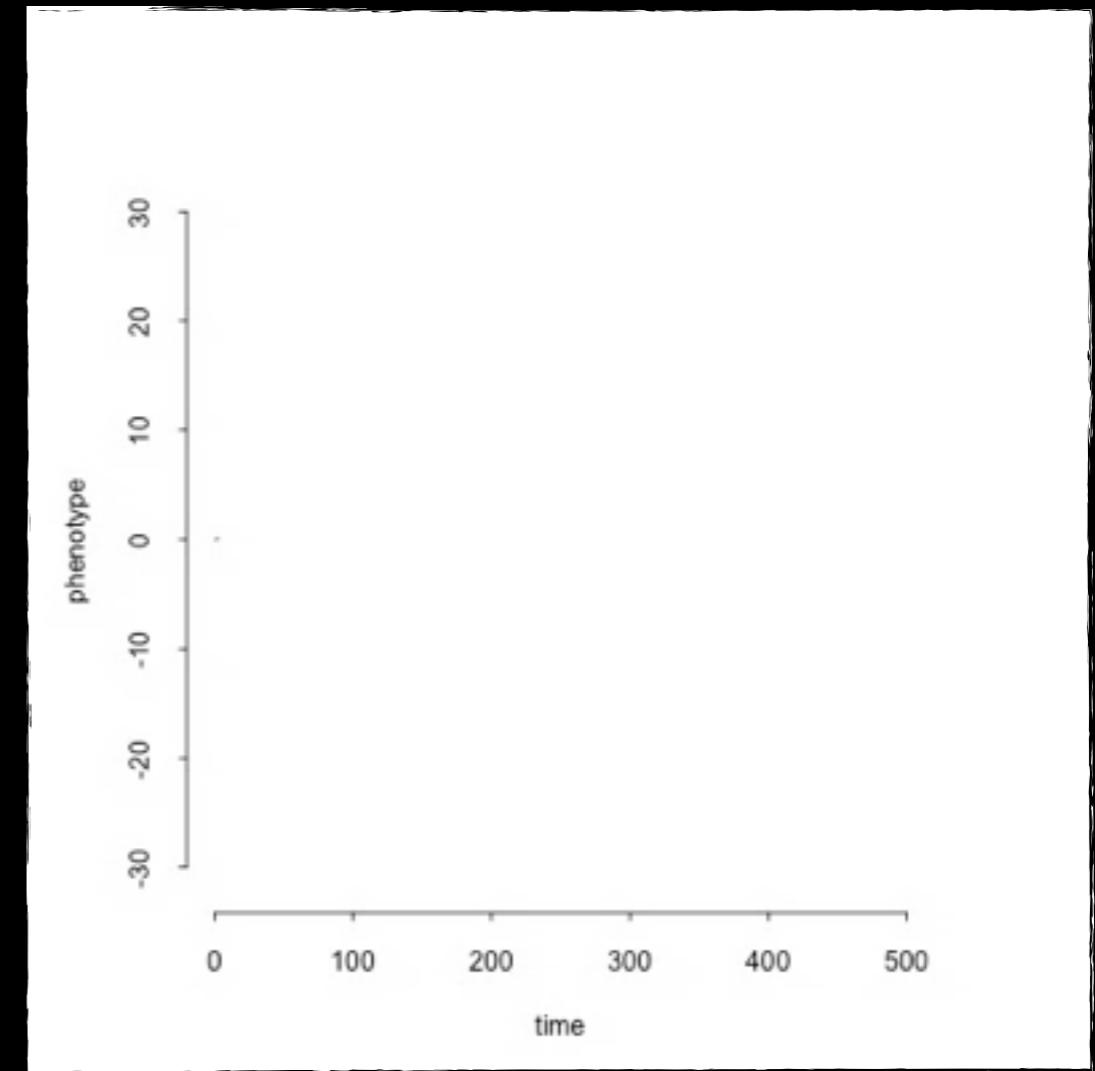


⊖

$$\sigma^2$$

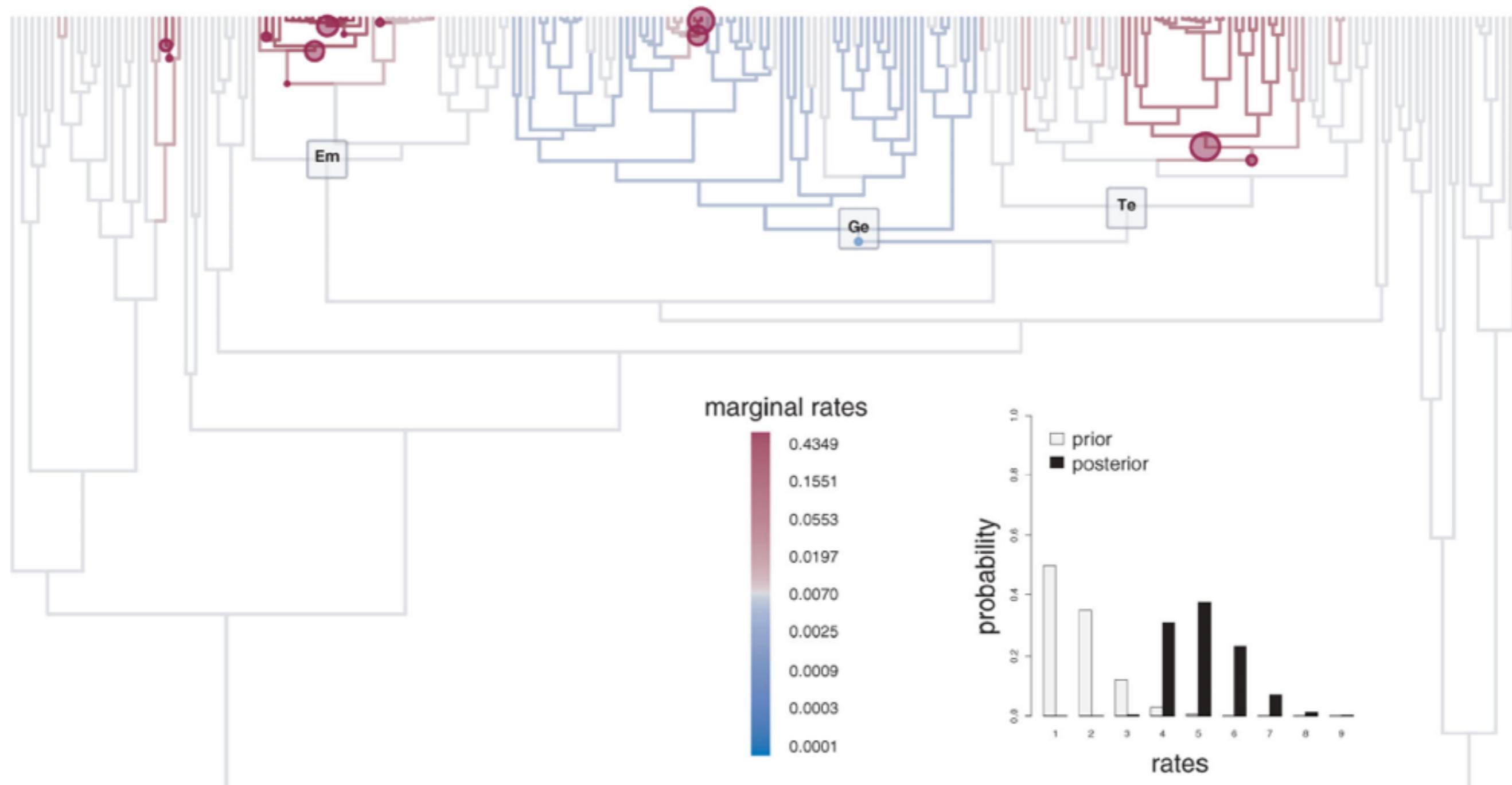


**single-rate diffusion**

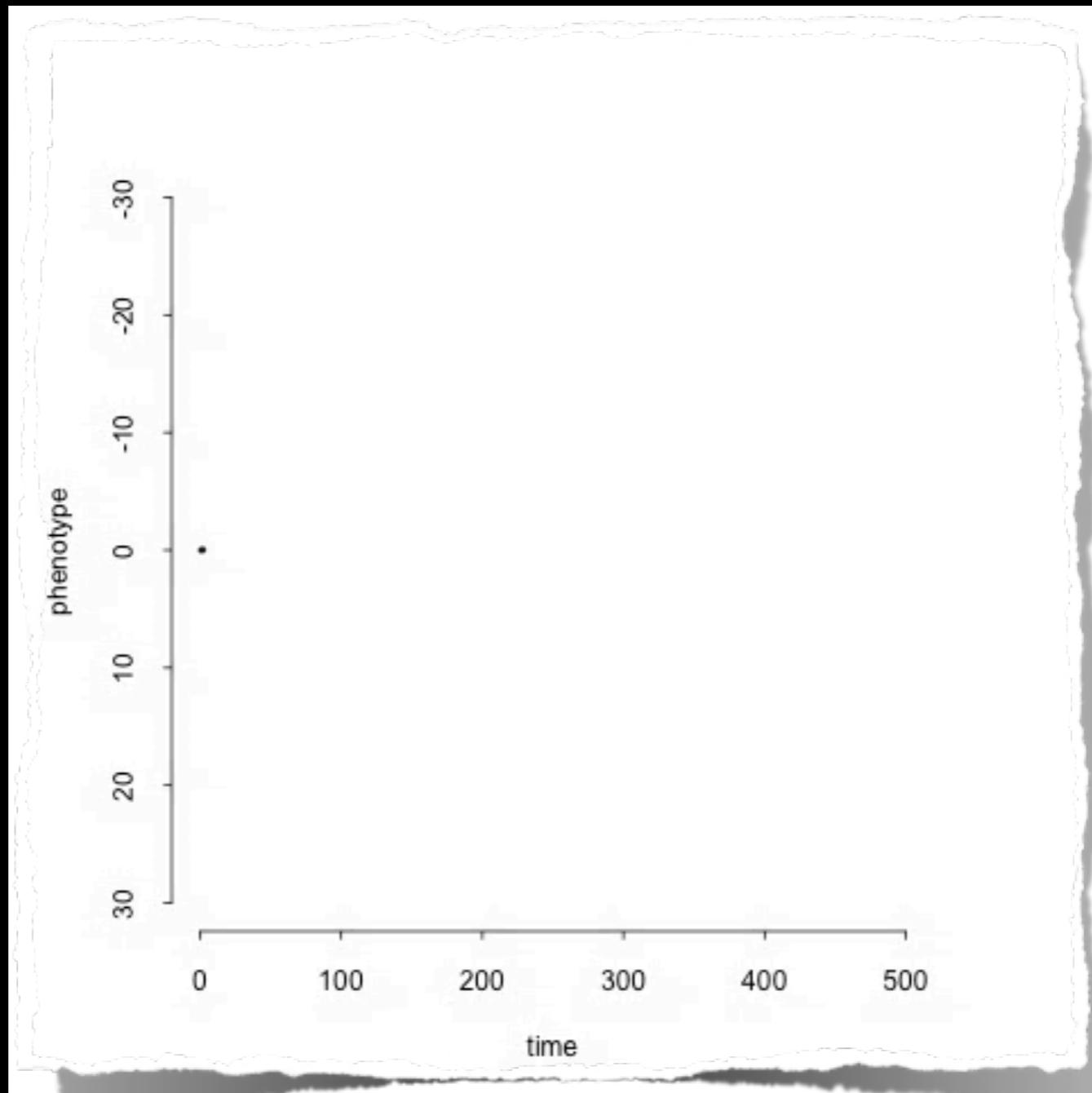


**multiple-rate diffusion**

Variation in evolutionary rates:  
auteur



# Jump diffusion



Simpson's "quantum evolution"

# Ornstein-Uhlenbeck Model

$\Theta$

$\sigma^2$

a

# **Simpsonian “Evolution by Jumps” in an Adaptive Radiation of *Anolis* Lizards**

Jonathan M. Eastman<sup>1</sup>, Daniel Wegmann<sup>2</sup>, Christoph Leuenberger<sup>3</sup>, and Luke J. Harmon<sup>1\*</sup>

<sup>1</sup> Department of Biological Sciences & Institute for Bioinformatics and Evolutionary Studies (IBEST),  
University of Idaho Moscow, ID 83843-3051

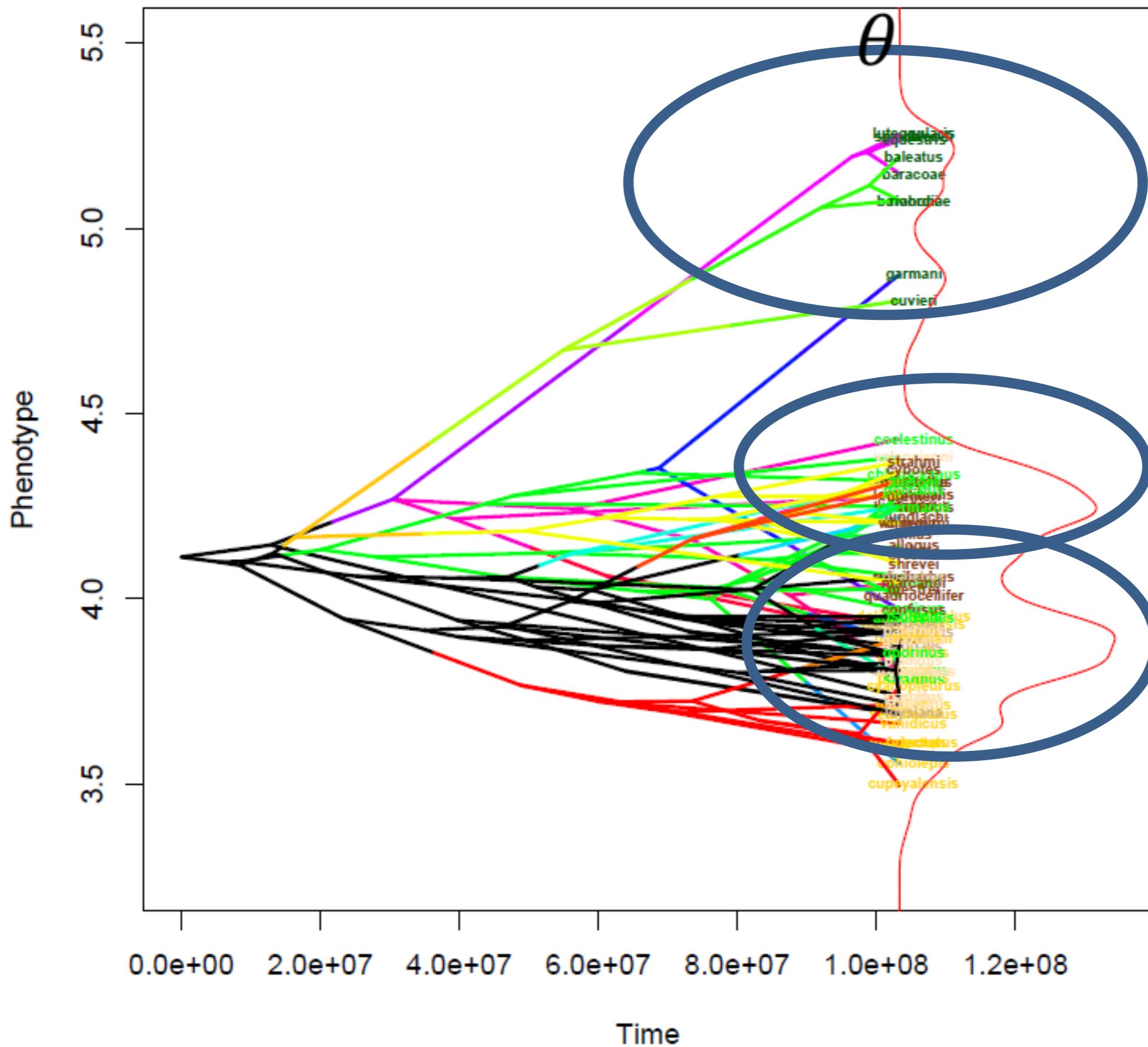
<sup>2</sup> Department of Biology, Fribourg University, Fribourg, Switzerland

<sup>3</sup> Department of Mathematics, Fribourg University, Fribourg, Switzerland

\* corresponding author: lukeh@uidaho.edu

Variation in evolutionary jumps

Variation in adaptive regimes:  
bayou



C [bamm-project.org/index.html](http://bamm-project.org/index.html)

WoS journals msOnline git construct nov14 agents CNN CNN WoS Facebook g



**Bayesian Analysis of  
Macroevolutionary Mixtures**

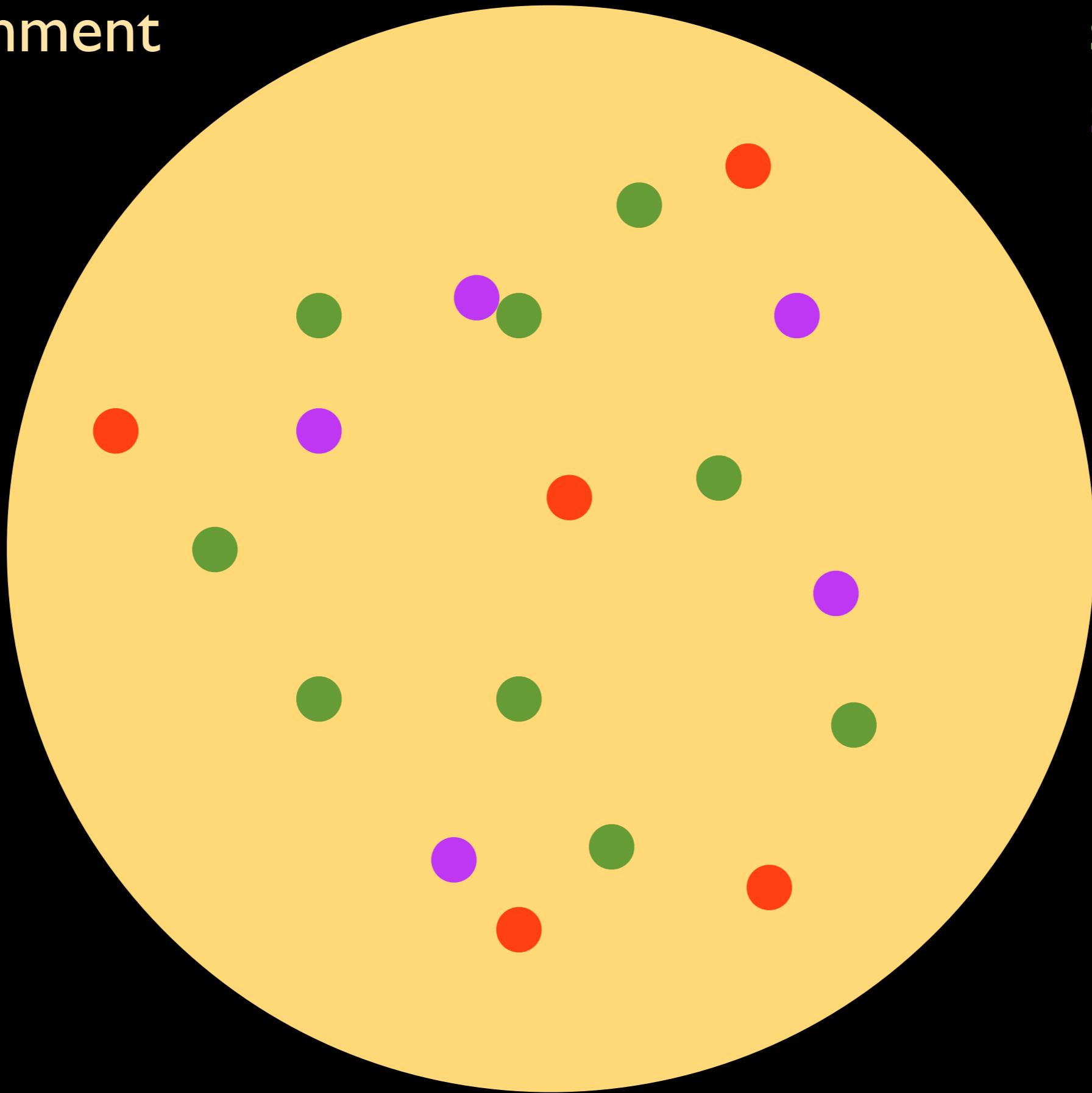
---

[Home](#) | [Download](#) | [Documentation](#) | [Contact Us](#)



environment

species A  
species B  
species C



# Model complexity

---

# Model adequacy

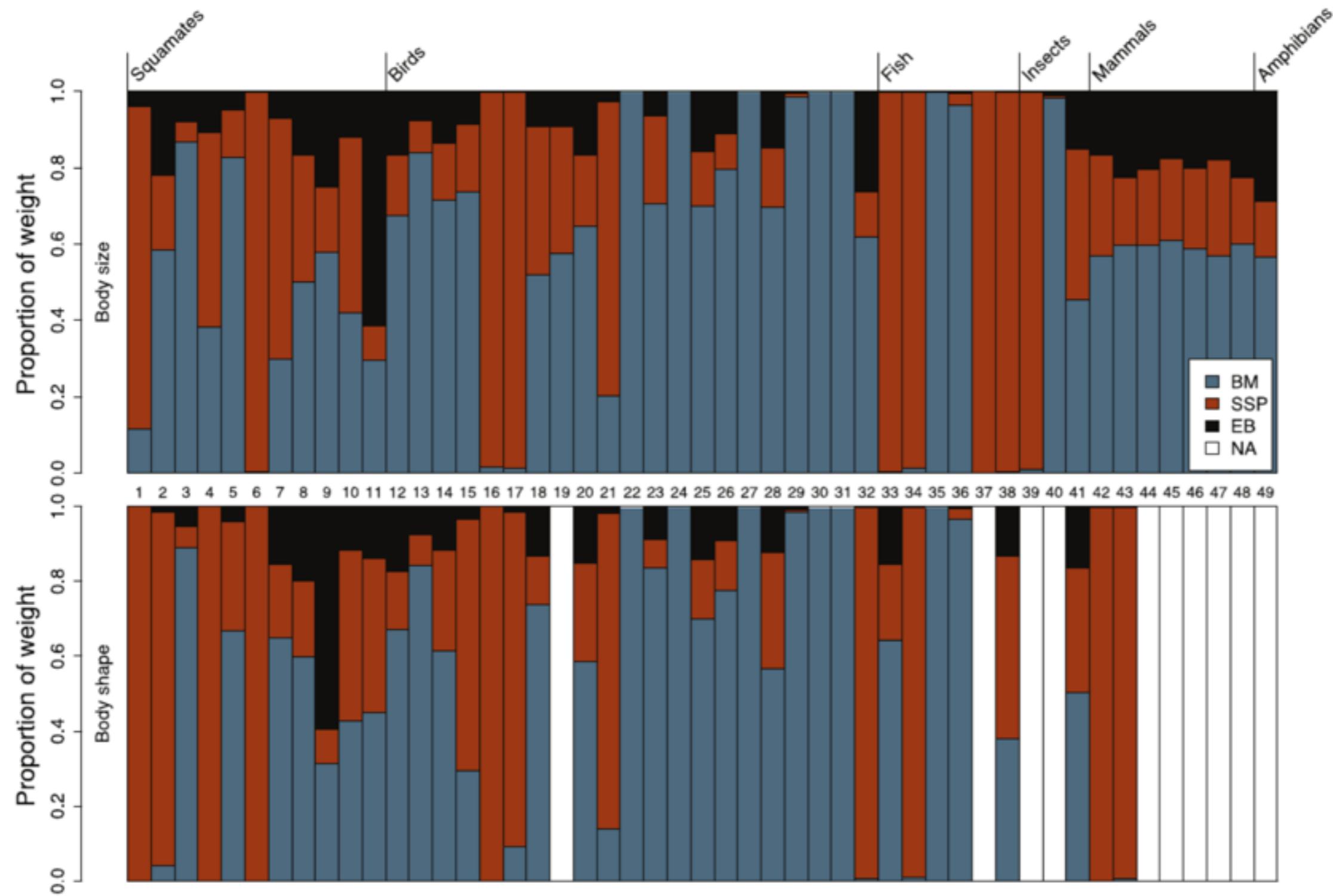
---

# Implementation

---

## Model adequacy

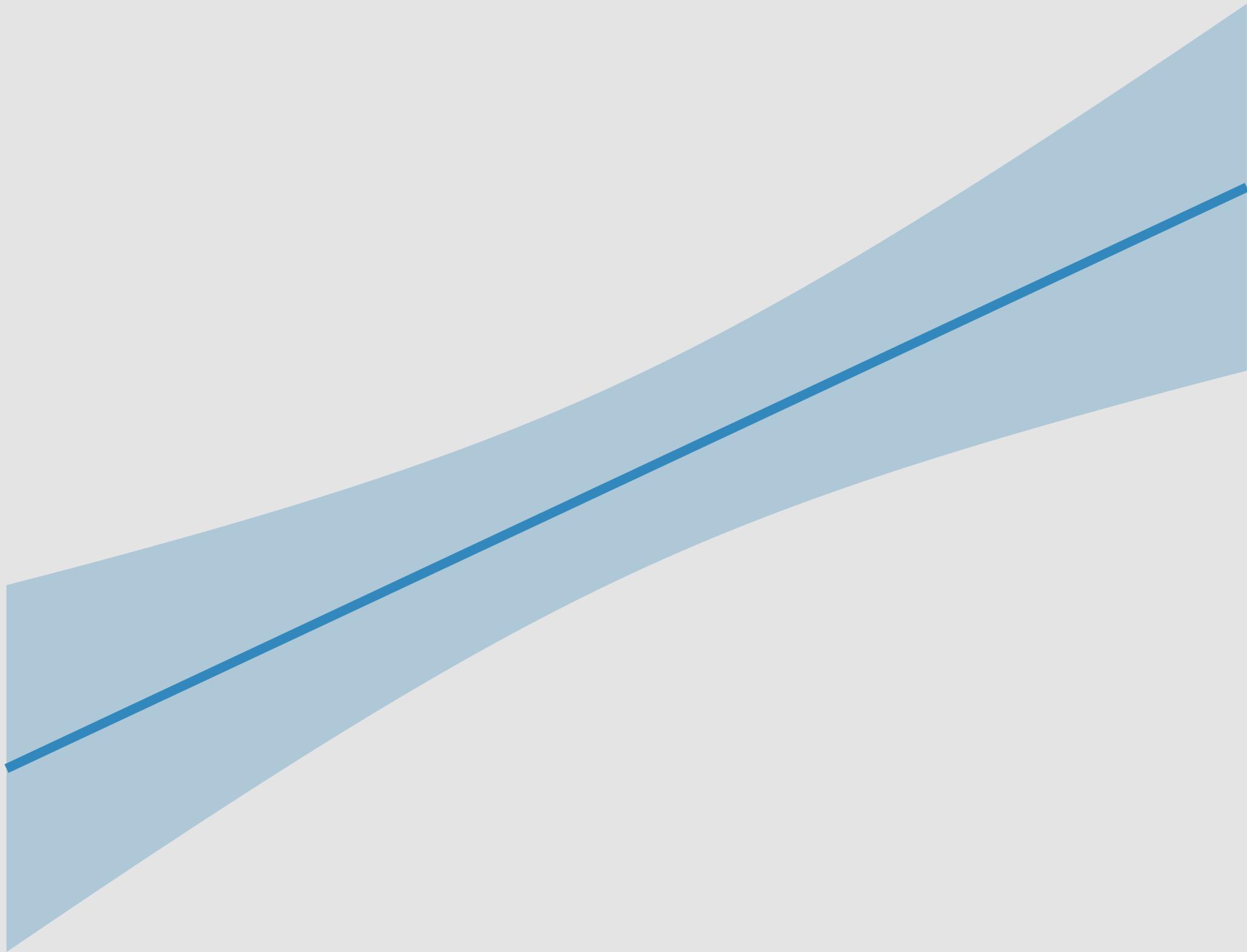
---



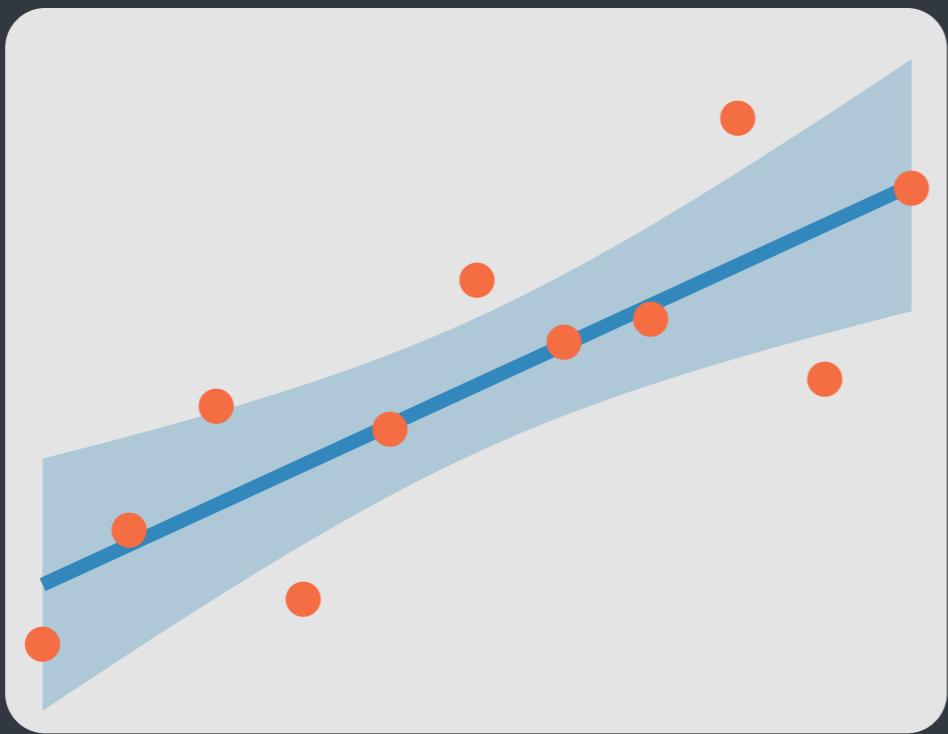
harmon et al. 2010

We can compare the relative fit  
of a set of candidate models...

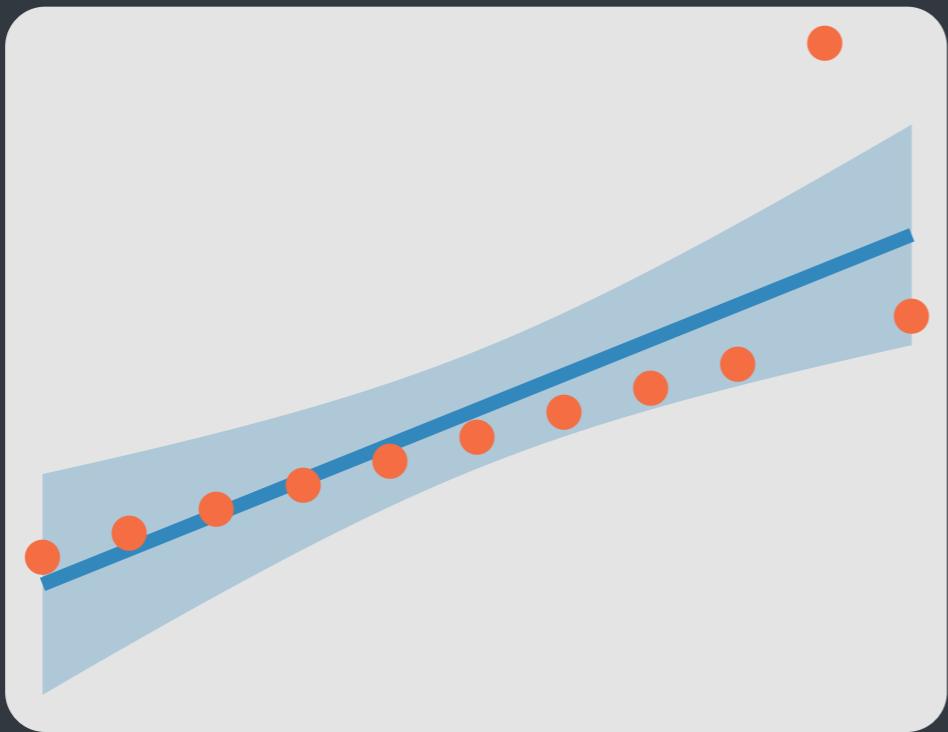
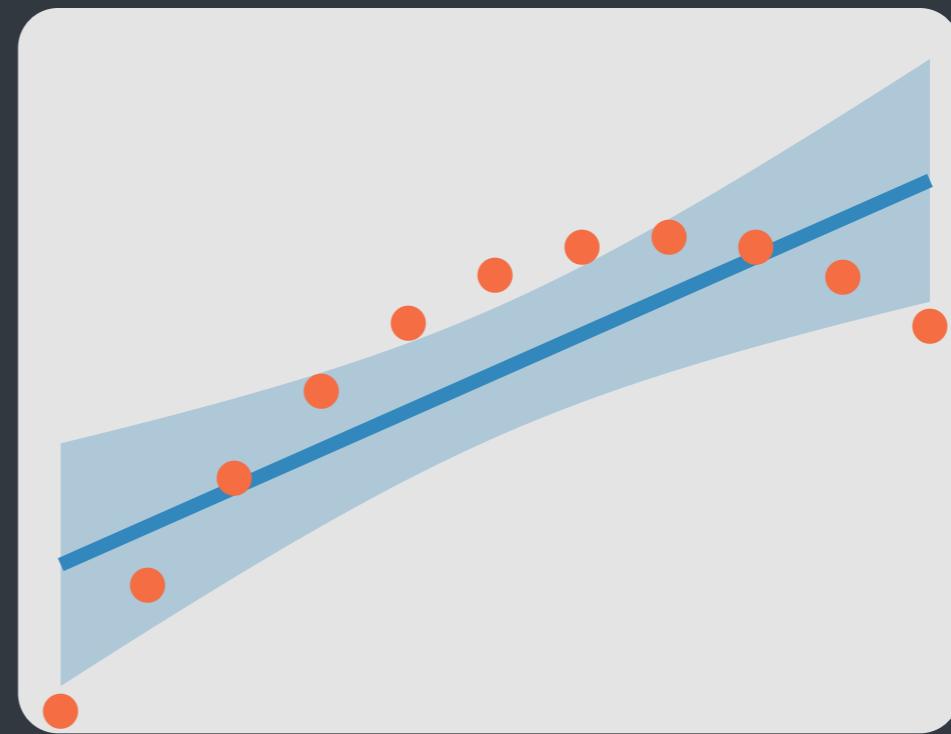
... but how do we know if the best of these models is actually any good?



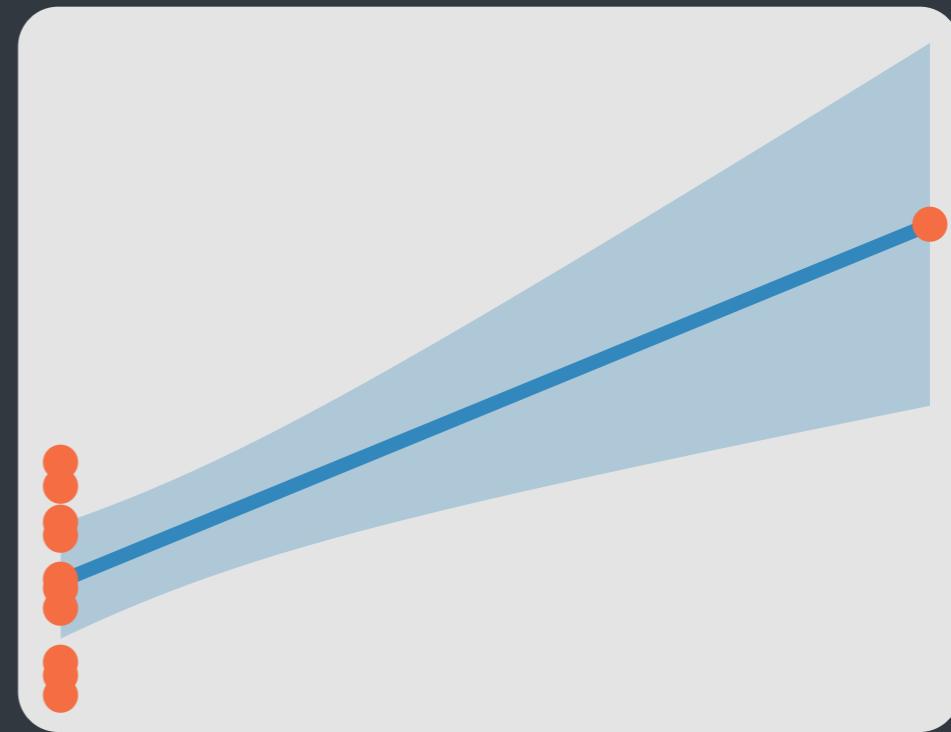
$R^2=0.67$  p=0.002



$R^2=0.67$  p=0.002



$R^2=0.67$  p=0.002



$R^2=0.67$  p=0.002

If we re-ran evolution, how likely are  
we to see a data set like ours

# SIMILAR

Model is likely adequate

# DIFFERENT

Model is likely inadequate

# How similar is similar

Problem:

No two datasets are exactly alike

# How similar is similar

**Problem:**

No two datasets are exactly alike

**Solution:**

Use test statistics to summarize data in meaningful ways

# How similar is similar

**Problem:**

No two datasets are exactly alike

**Solution:**

Use test statistics to summarize data in meaningful ways

**Problem:**

Species are not independent data points

# How similar is similar

**Problem:**

No two datasets are exactly alike

**Solution:**

Use test statistics to summarize data in meaningful ways

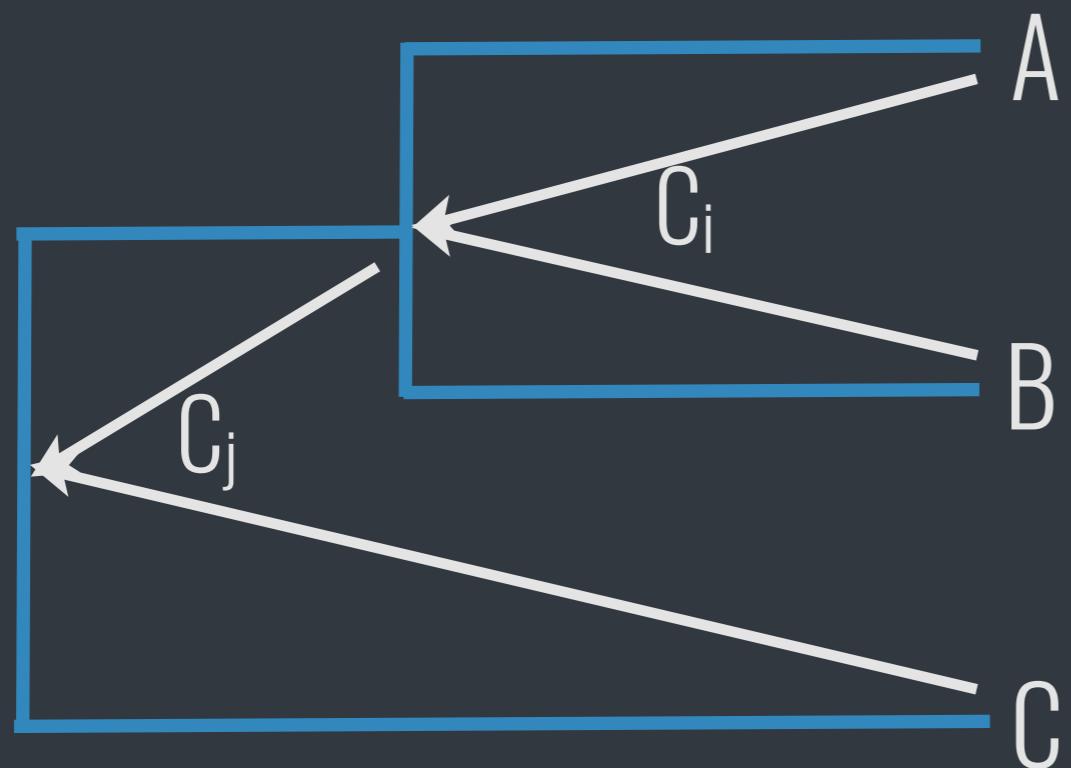
**Problem:**

Species are not independent data points

**Solution:**

Calculate test statistics on **contrasts** rather than the data

# Independent contrasts



$n-1$  contrasts for  $n$  tips

Under Brownian motion  
 $C \sim \text{Normal}(0, \sigma)$

# For non-Brownian models

**Problem:**

Contrasts will no longer be normally distributed

# For non-Brownian models

**Problem:**

Contrasts will no longer be normally distributed

**Solution:**

Use model parameters to standardize branch lengths by the expected (co)variance that will accumulate along them

# For non-Brownian models

**Problem:**

Contrasts will no longer be normally distributed

**Solution:**

Use model parameters to standardize branch lengths by the expected (co)variance that will accumulate along them

Refer to rescaled tree as a **unit tree**

# Test statistics

Slope of contrasts  
vs. expected  
variances

Slope of contrasts  
vs. ancestral  
state

Slope of contrasts  
vs. node height

Mean of squared  
contrasts

Coefficient of  
variation of  
contrasts

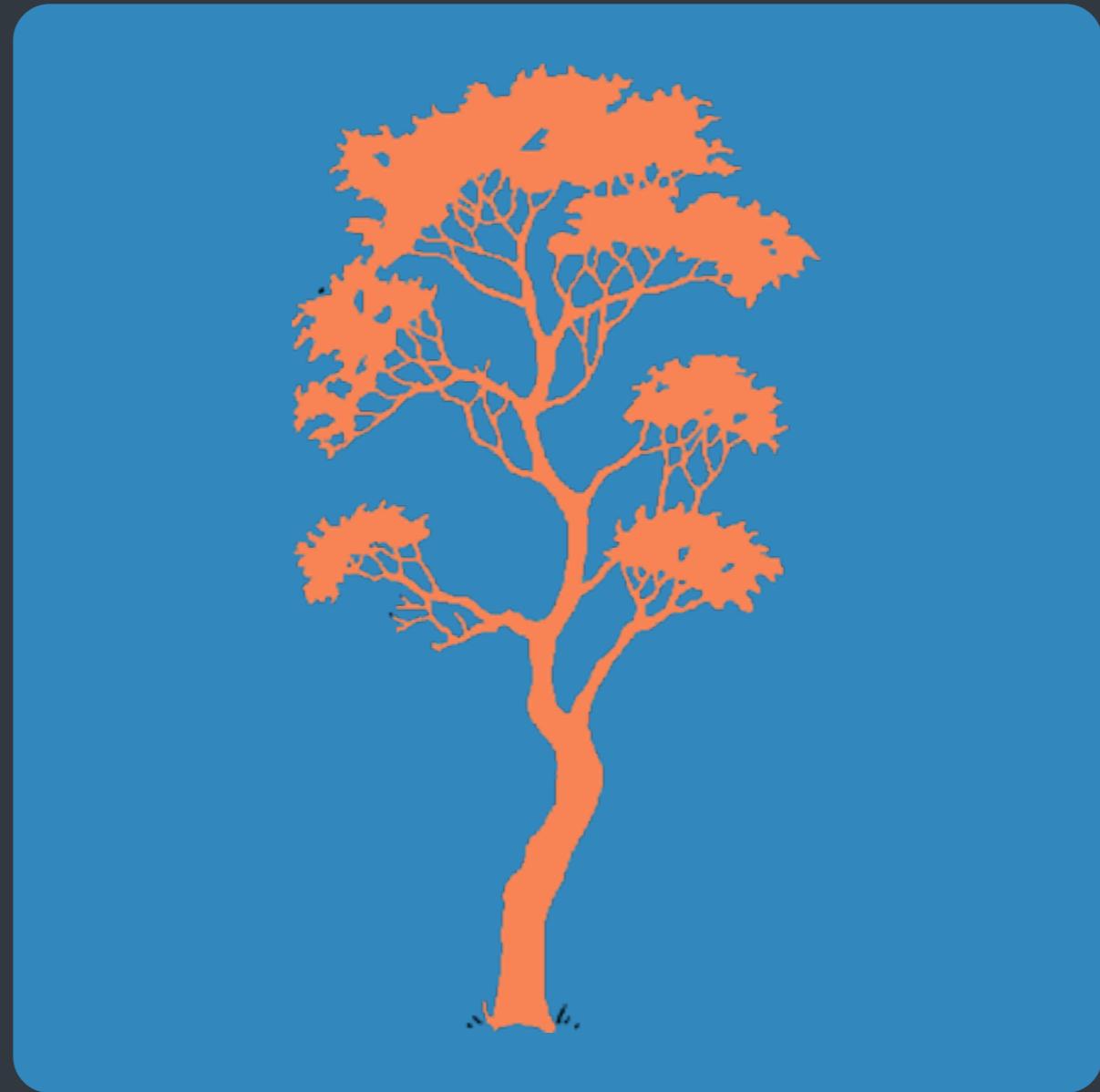
KS-Test for  
normality of  
contrasts

# Simulating datasets for comparison

Simulate a lot of new datasets on **unit tree**

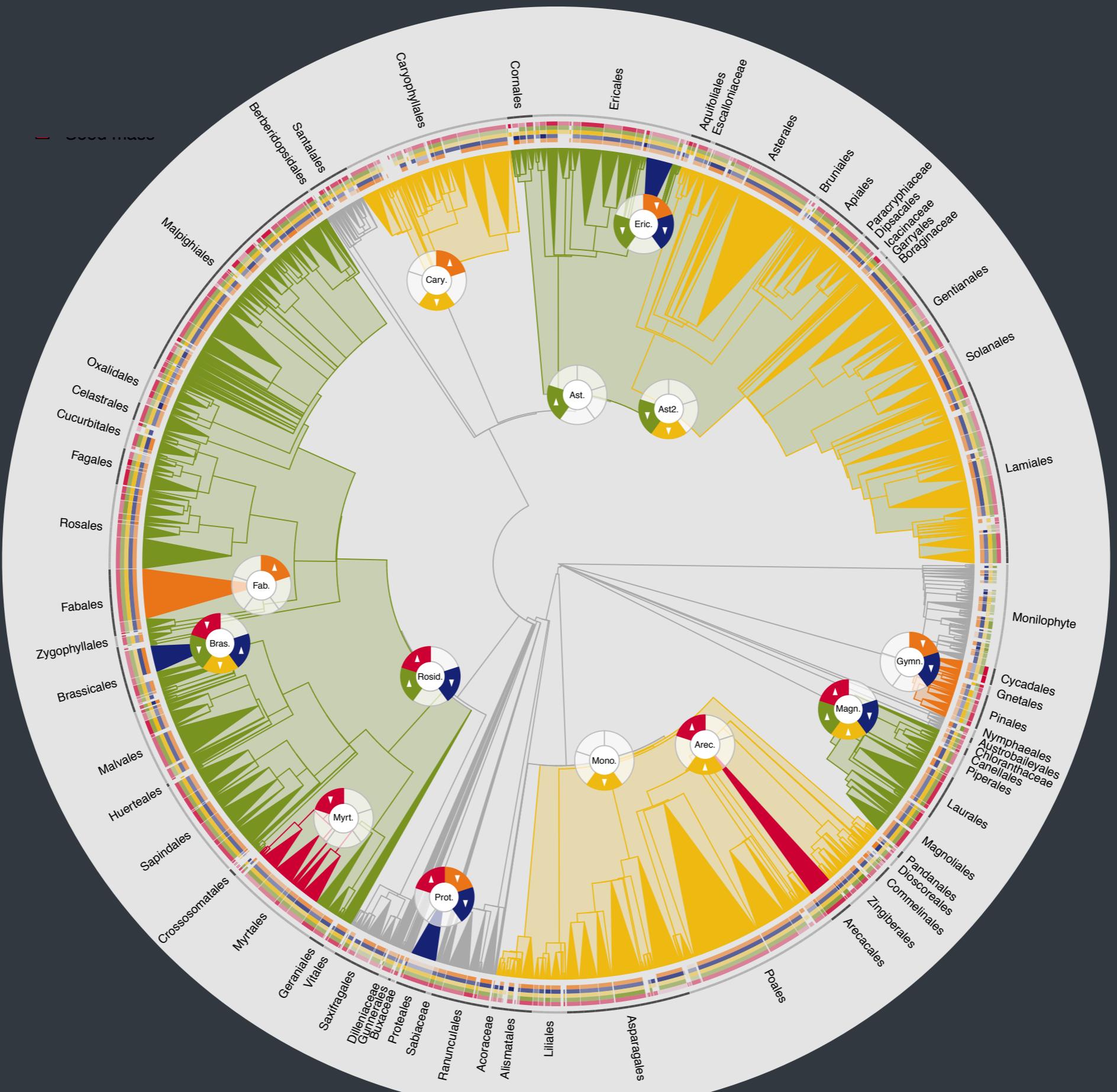
Use a BM model with a rate of **1**

Calculate test statistics on simulated dataset



# ARBUTUS

<https://github.com/mwpennell/arbutus>



# Specific Leaf Area

72 datasets (20 - 2,200 species)

# Seed mass

226 datasets (20 - 22,817 species)

# Leaf Nitrogen Content

39 datasets (20 - 936 species)

Zanne et al. 2014 Nature

Wright et al. 2004 Nature

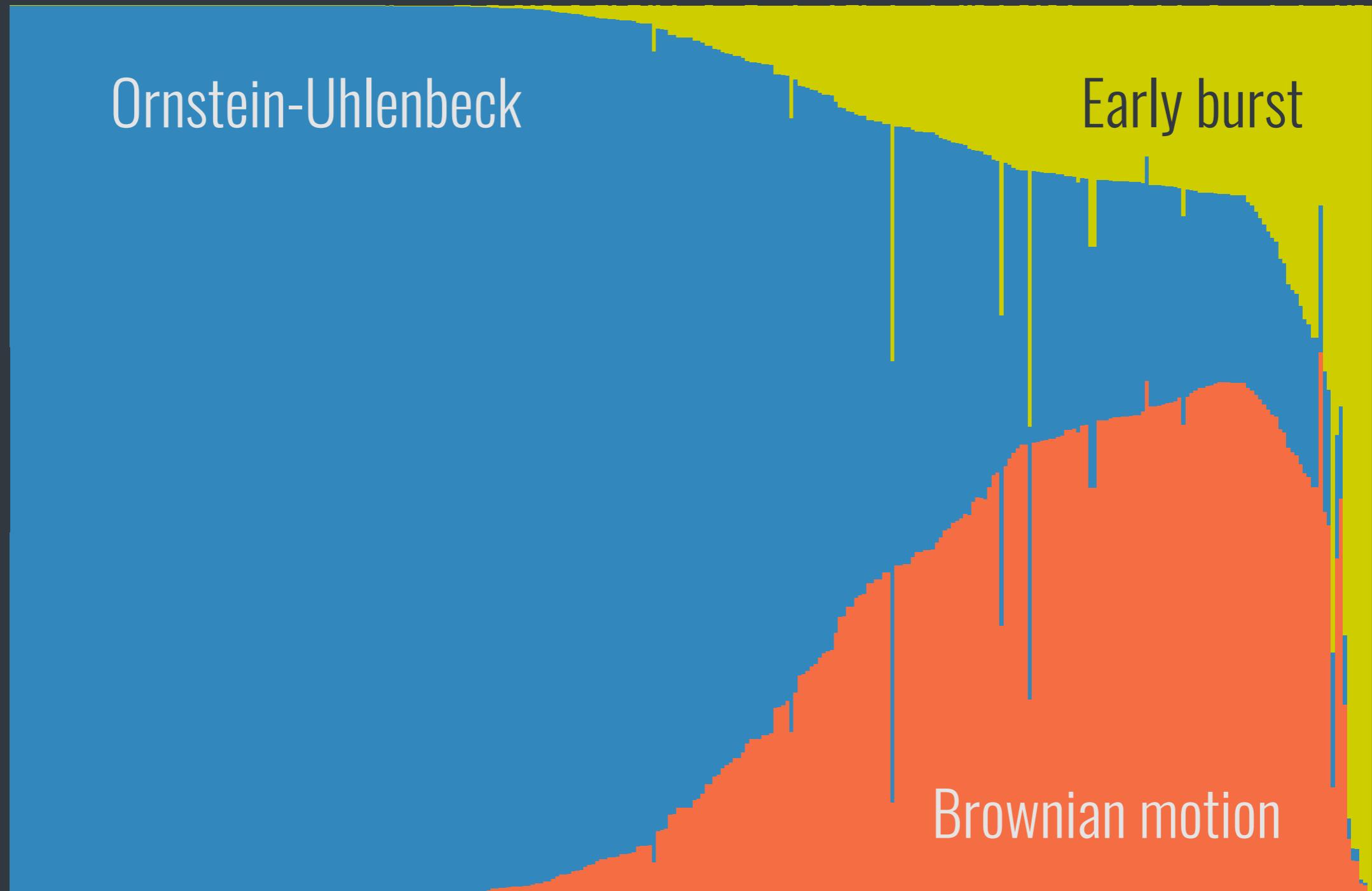
Kleyer et al. 2008 J Ecology

Kew Seed Information Database 2014

# Empirical analyses

1. Fit Brownian motion, Ornstein-Uhlenbeck and Early Burst to each dataset
2. Calculate relative support using AIC
3. Assess adequacy of best-fitting model

Model Support (AIC)



Dataset (1 - 337)

Pennell et al. 2015 Am Nat

# Specific Leaf Area

Model deviations detected in 32/72 datasets

# Seed mass

Model deviations detected in 153/226 datasets

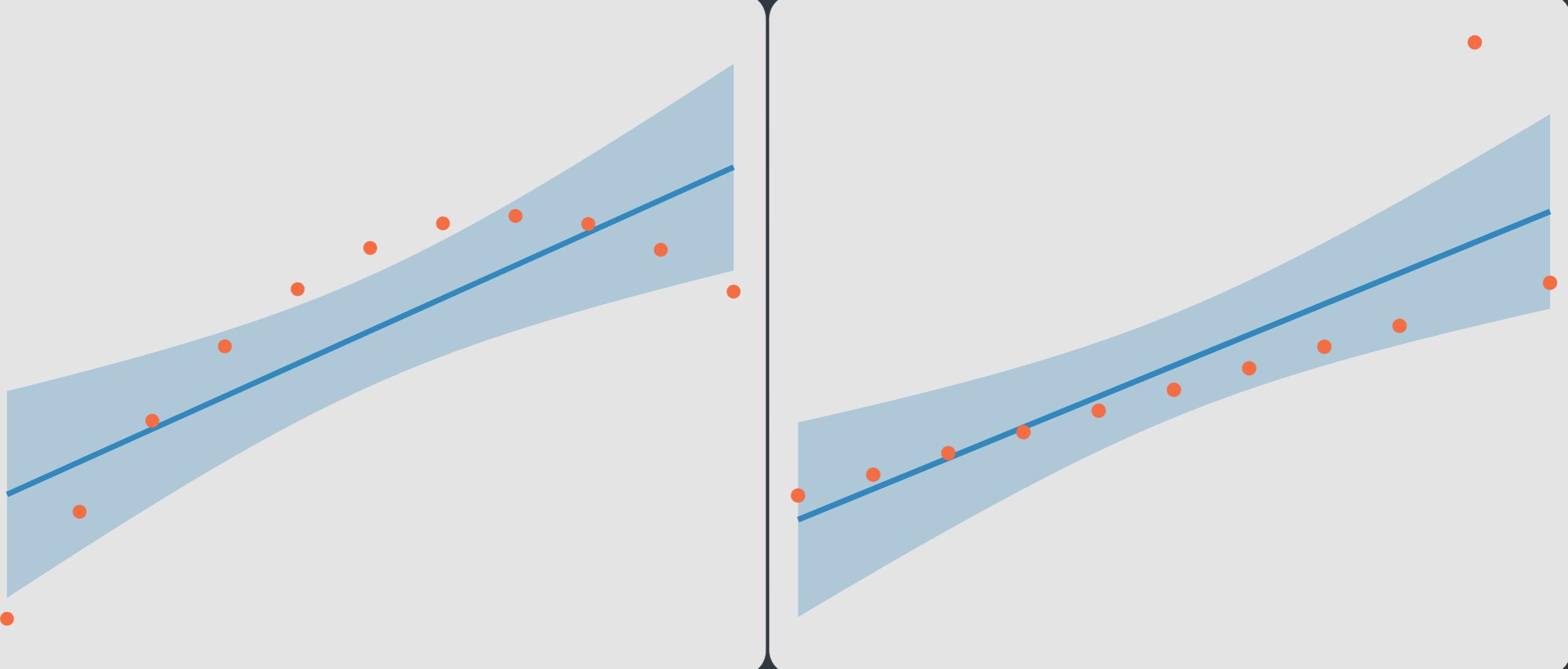
# Leaf Nitrogen Content

Model deviations detected in 19/39 datasets

Simple, commonly used models are  
often (woefully) inadequate

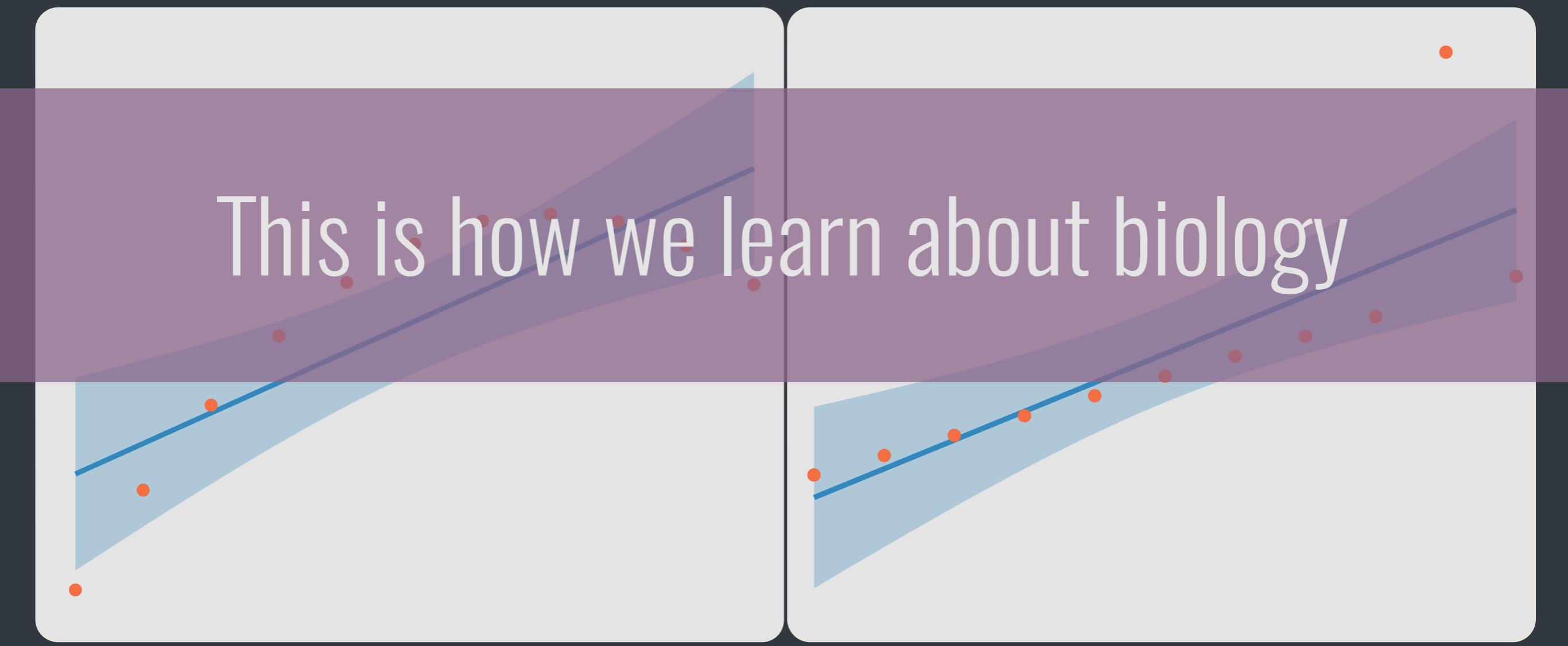
But we already knew that...

# We are (often) here

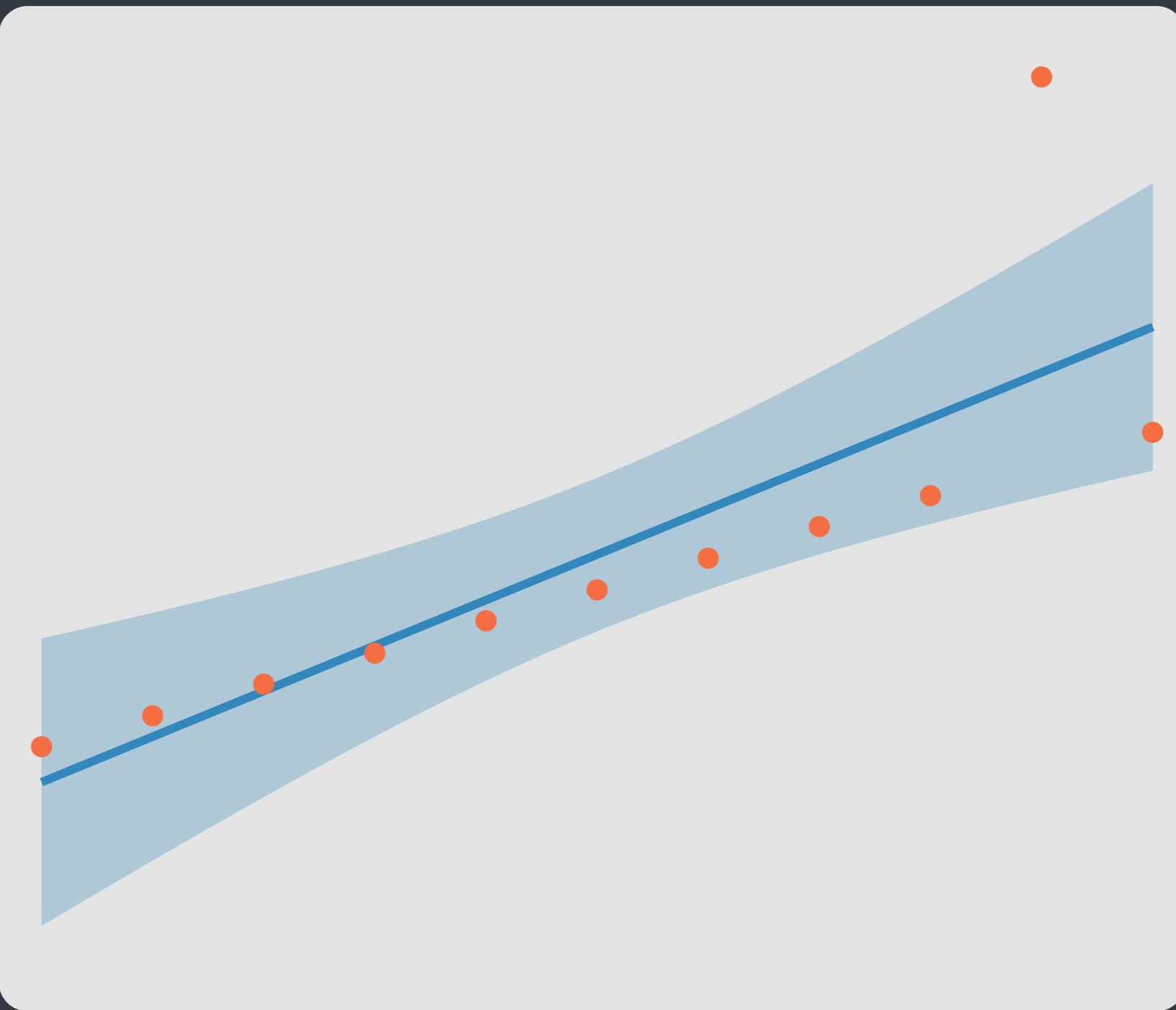


# We are (often) here

This is how we learn about biology



# Learn about our data



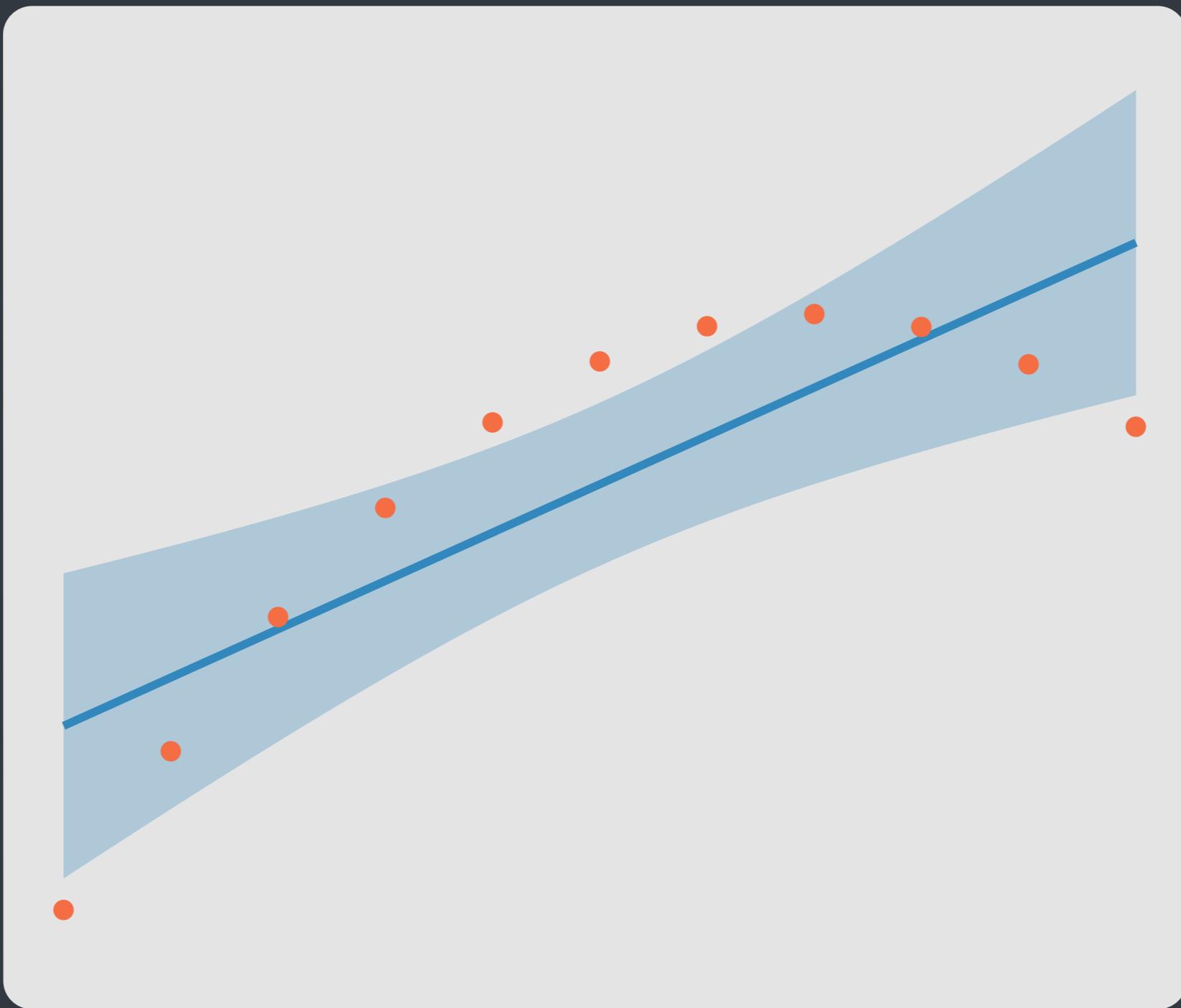
# Learn about our data

Phylogenetic error (topology and branch lengths)

Measurement error

Biologically interesting “outlier” species

# Learn about evolutionary processes



# Learn about evolutionary processes

Time heterogeneous models

Different models for different parts of the tree

Biologically motivated models

Understanding how and why a model fails  
can provide new biological insights

# Model complexity

---

# Model adequacy

---

# Implementation

---

# Implementation

---



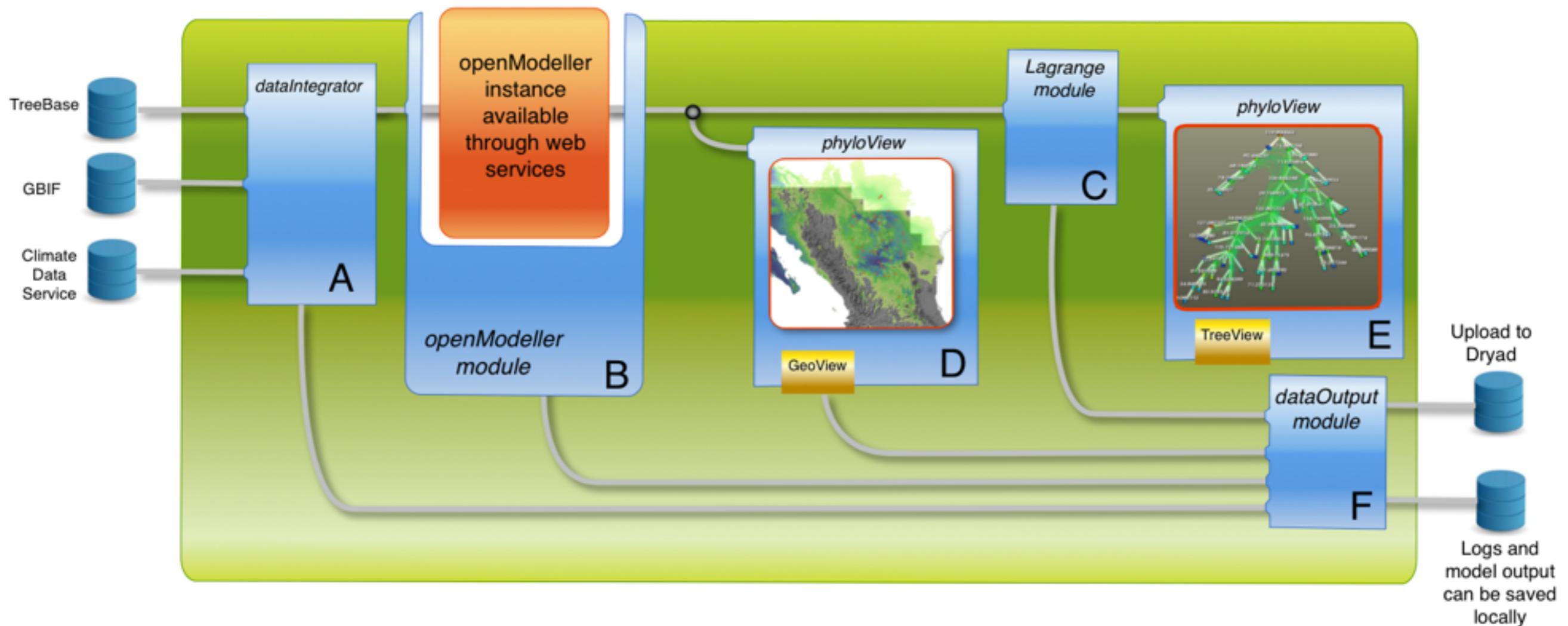
**ARBOR**  
revolutionary workflows

Arbor: big picture

Comparative  
methods through  
visual workflows

Girder user management

Romanesco workflow  
engine



Comparative  
methods through  
visual workflows

Data exchange  
via  
web services

Open Tree of  
Life

Phenomics



Dryad

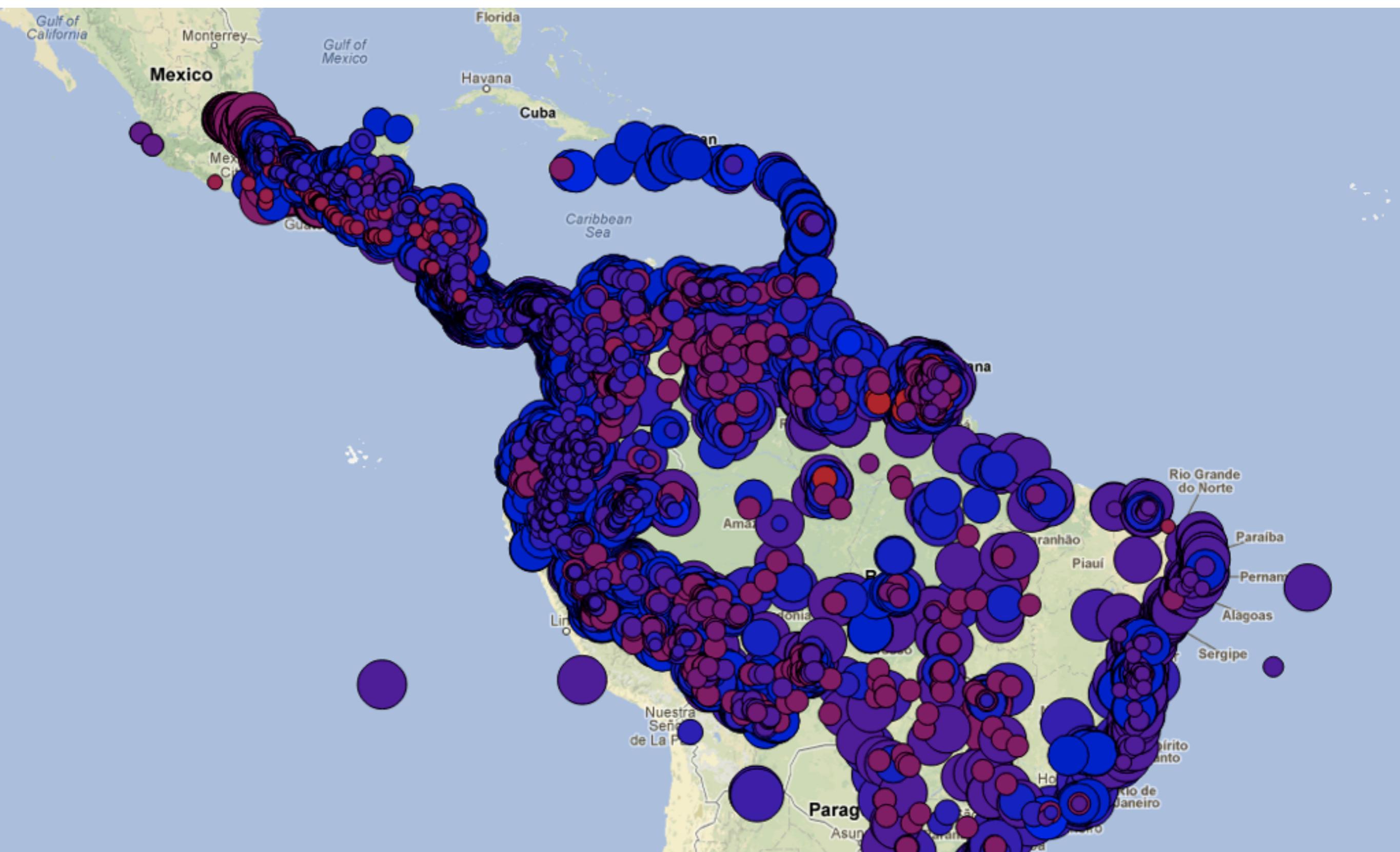
GBIF / BISON

MongoDB database

Comparative  
methods through  
visual workflows

Data exchange  
via  
web services

Implement novel  
analysis workflows

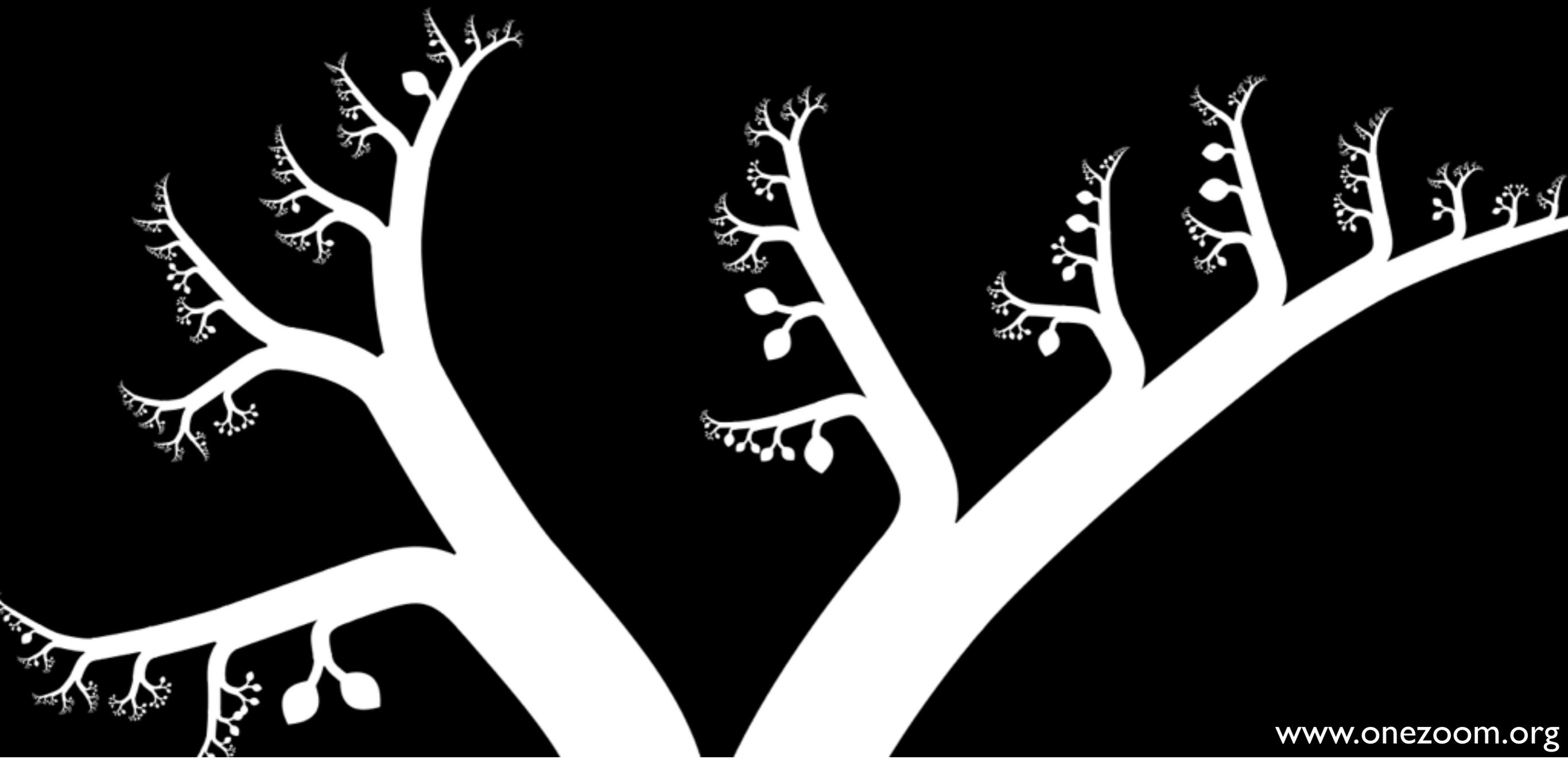


Comparative  
methods through  
visual workflows

Data exchange  
via  
web services

Implement novel  
analysis workflows

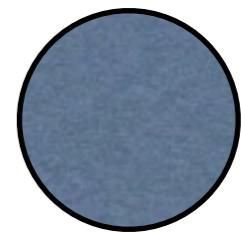
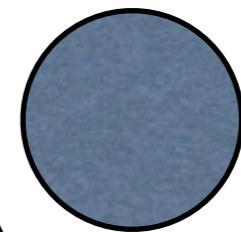
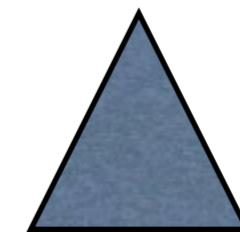
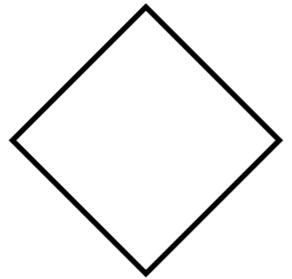
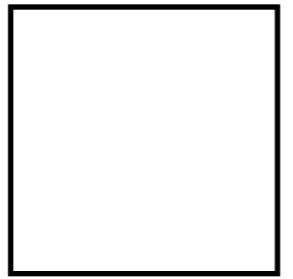
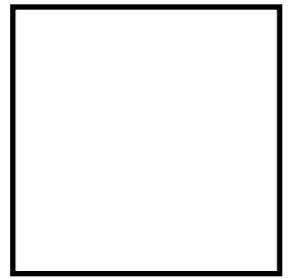
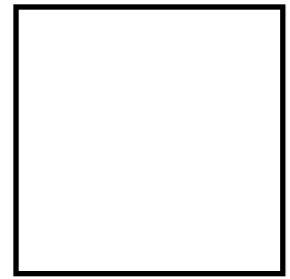
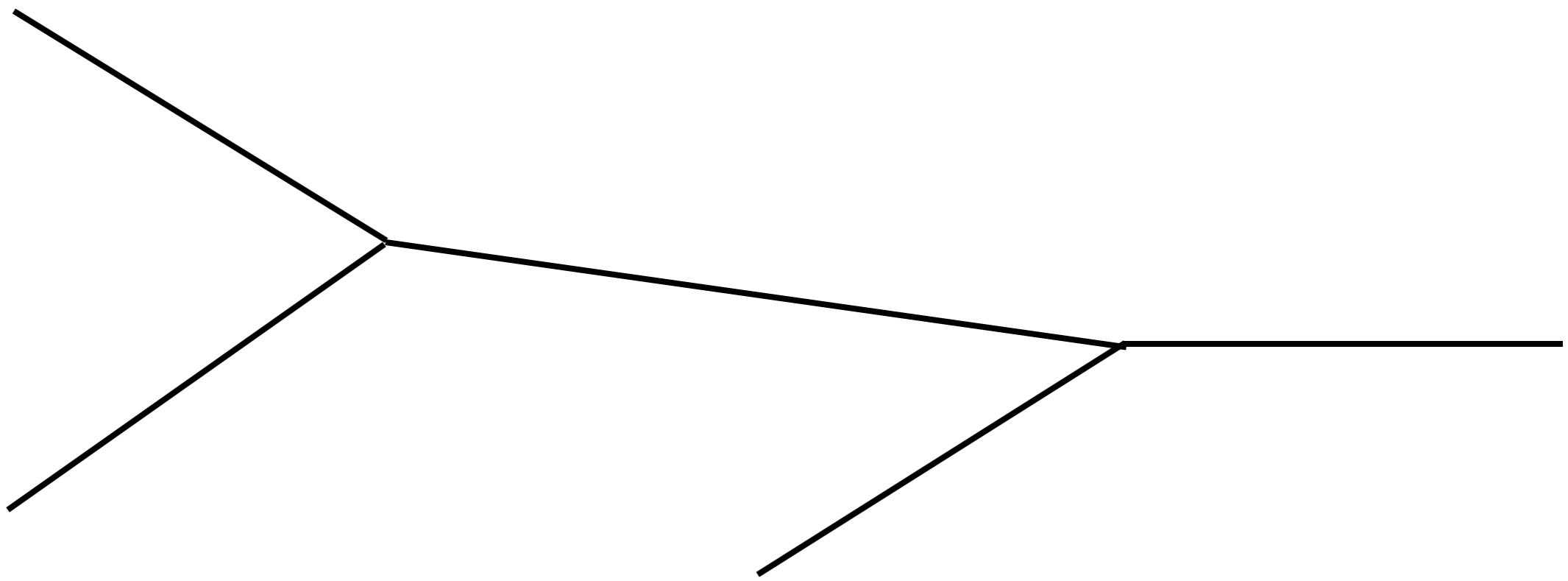
Scale to  
very large trees

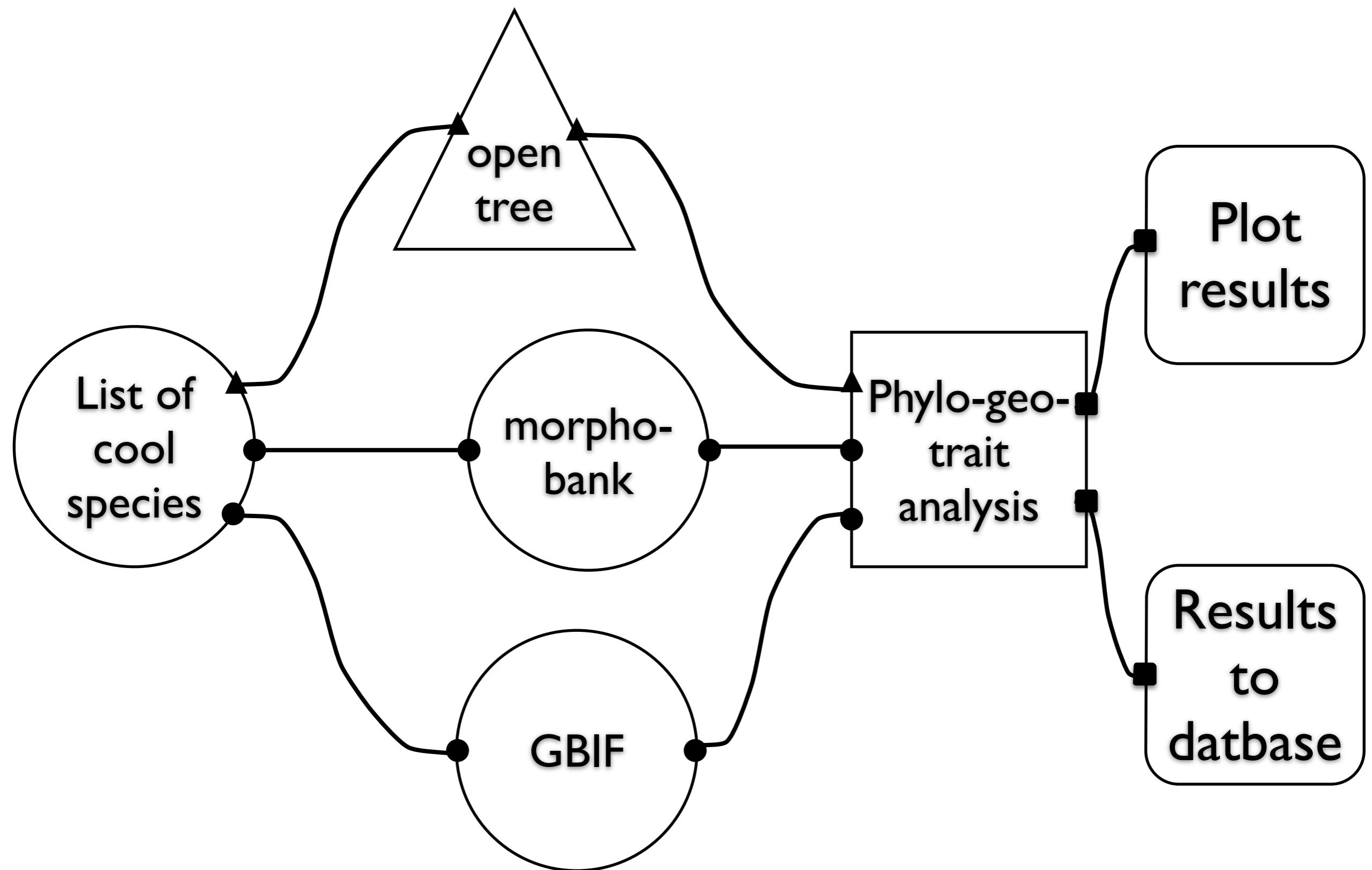


[www.onezoom.org](http://www.onezoom.org)



arbor





<https://arbor.kitware.com>

[arborworkflows / arborworkflows.github.com](#) [Star 0](#) [Fork 0](#)

# Arbor Demonstration Videos

curtislisle edited this page 15 hours ago · 1 revision

Here are some short demonstration videos to illustrate some of the features of the Arbor web interface:

[An Introduction to the Arbor Interface](#)

[Making a New Analysis with the Arbor Interface](#)

[How to Create and Execute a multistep workflow in Arbor](#)

▼ Pages 6

[Arbor Wiki Homepage](#)

[Arbor Demonstration Videos](#)

[Background Papers](#)

[Description of Table Operations in Arbor](#)

[Home](#)

[Open Tree Integration Analysis in Arbor](#)

Clone this wiki locally

<https://github.com/arborworkf...>



[Clone in Desktop](#)



Use the panel below to visualize, analyze, or manage your data.

Logged in as Girder Admin [Log Out](#)

Data Management

Analysis

Visualization

Browse or drop files

Name

Public

New collection

Default

OpenTree

Table Operations

Save or download data

table2.csv (Table Operations) ▾

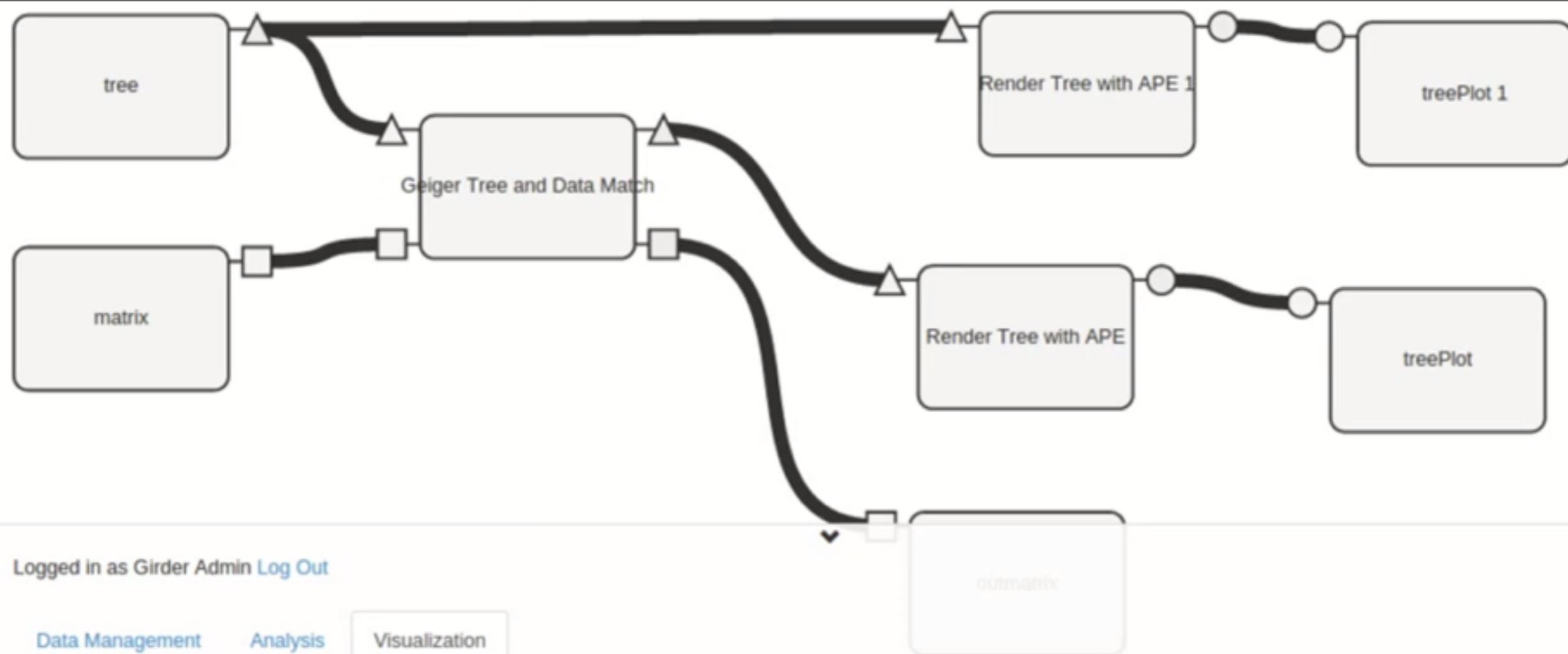
Quick view

geospiza\_characters.csv

csv

Download





Logged in as Girder Admin [Log Out](#)

Data Management

Analysis

Visualization

Select visualization

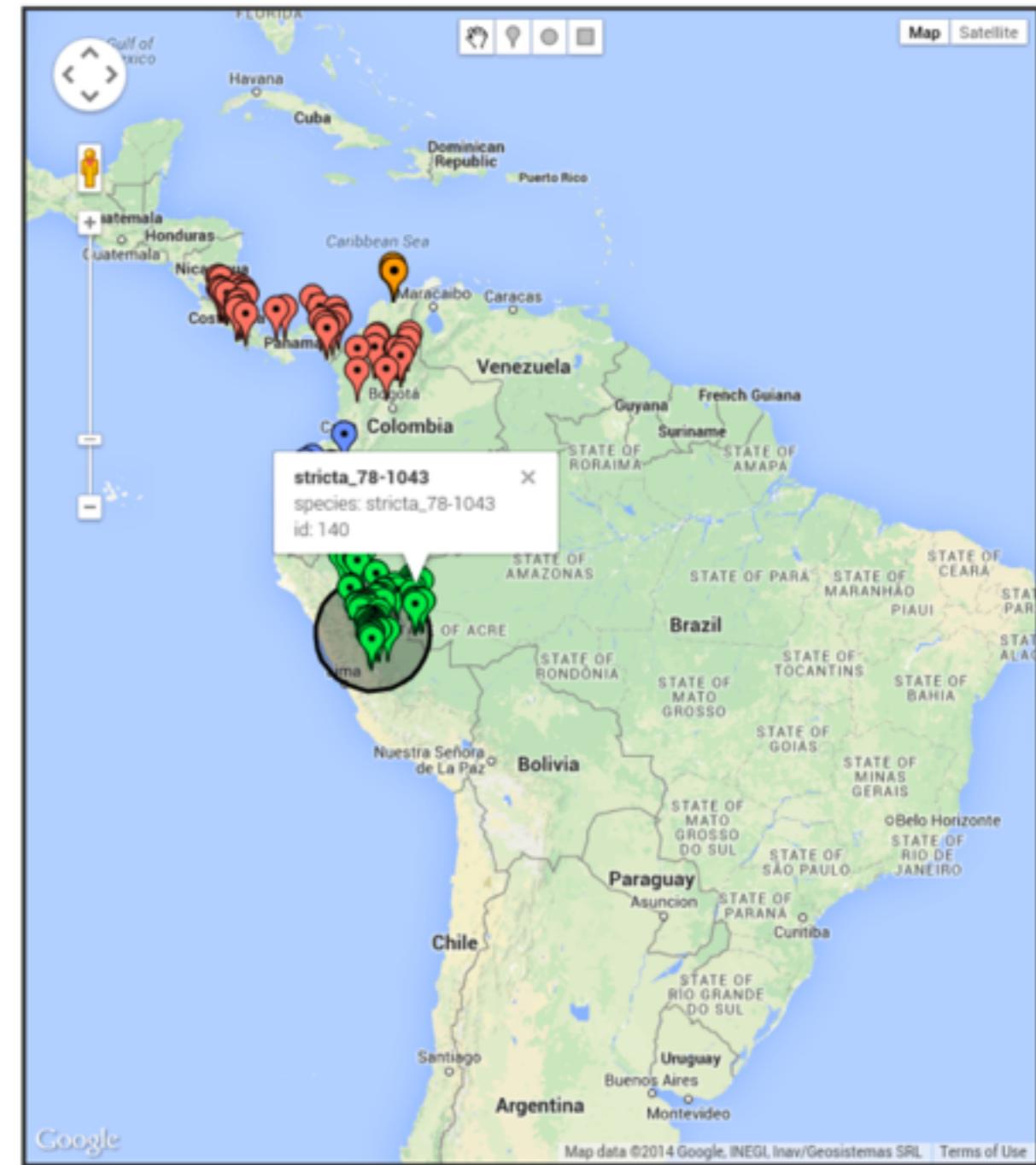
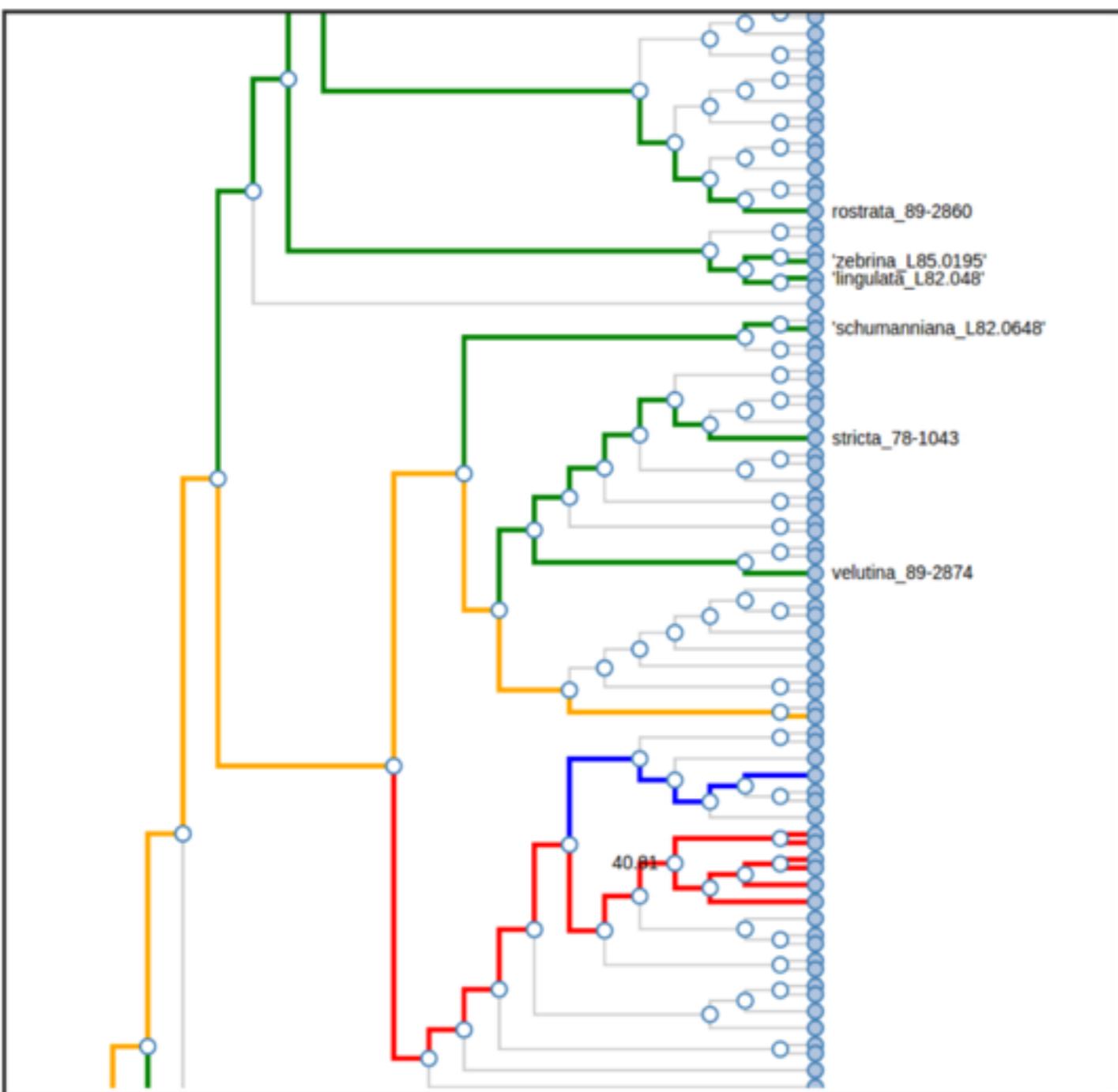
image

Update

data

Tree and Matrix Match Workflow treePlot

Arbor Collection: Default ▾ Dataset fullHeliconia-loc-ids.nested.json

[Empty Tree](#) [Full Tree](#)  Show Names  Enable Clade Selection [Increase Depth](#) [Decrease Depth](#) [Increase Height](#) [Decrease Height](#) [Clear Highlights](#) [Clear Map Markers](#)

Luke HARMON  
Curtis LISLE  
Chelsea SPECHT  
Jorge SOBERON  
Bob THACKER  
Charlie HUGHES  
Wes TURNER  
Jeff BAUMES





(Slide from 2012)

# What can we learn from a tree?

Luke J. Harmon  
University of Idaho



“I call it the **law of the instrument**, and it may be formulated as follows: Give a small boy a hammer, and he will find that everything he encounters needs pounding.”

---

Kaplan, 1964

“What else is a man to do when he has an idea...  
but ride it as hard as he can, and leave it to  
others to hold it back within proper limits?”

---

Kaplan, 1964

# Acknowledgements

- Jon Eastman, Josef Uyeda, Matt Pennell, Scott Nuismer, many other collaborators
- University of Idaho
- Funding from US National Science Foundation
- Visit us: [www.onezoom.org](http://www.onezoom.org);  
[www.arborworkflows.com](http://www.arborworkflows.com)