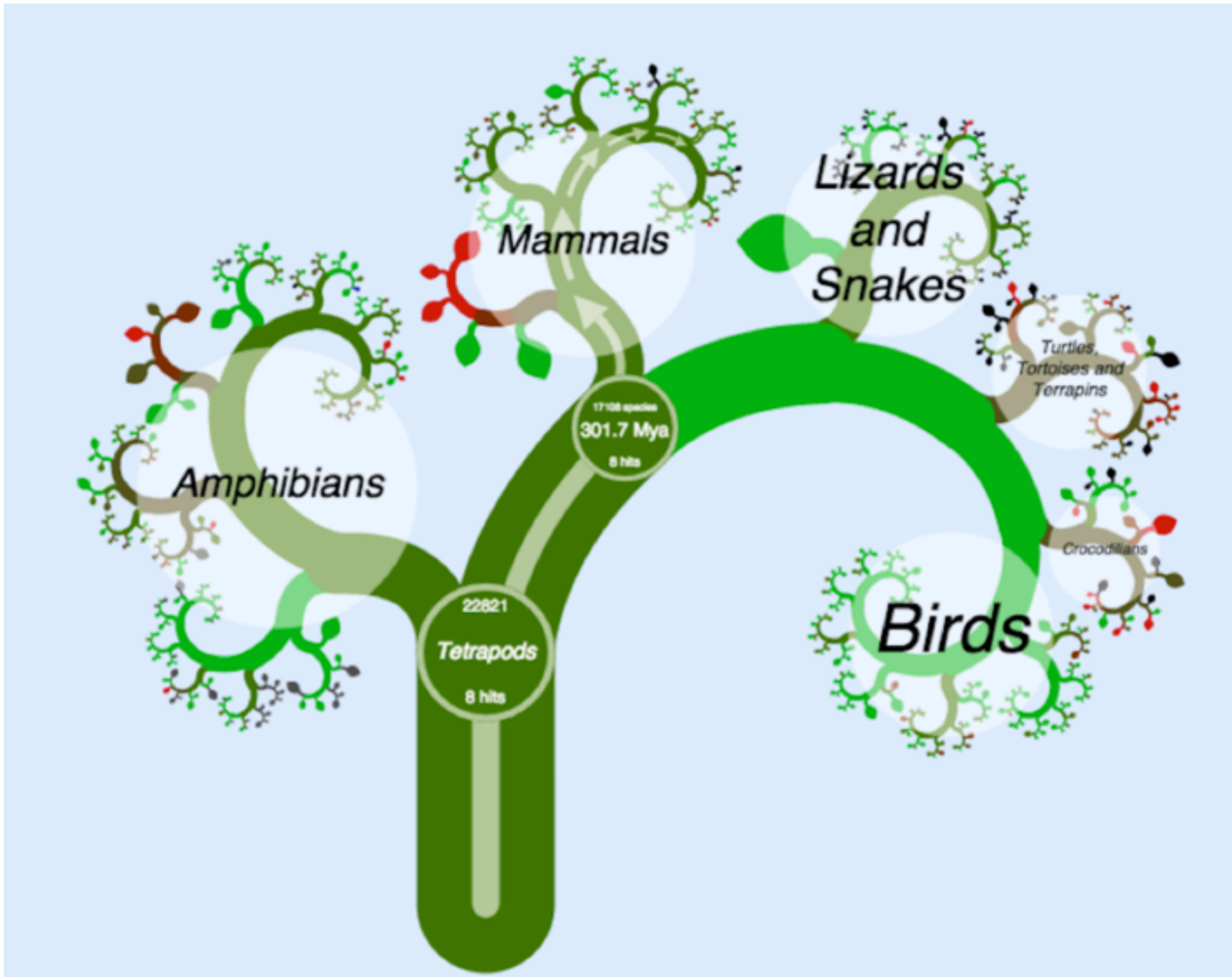
Applications of SSA: Phylogenetics

Shan Shan
Department of Mathematics
Duke University

SLICERMORPH

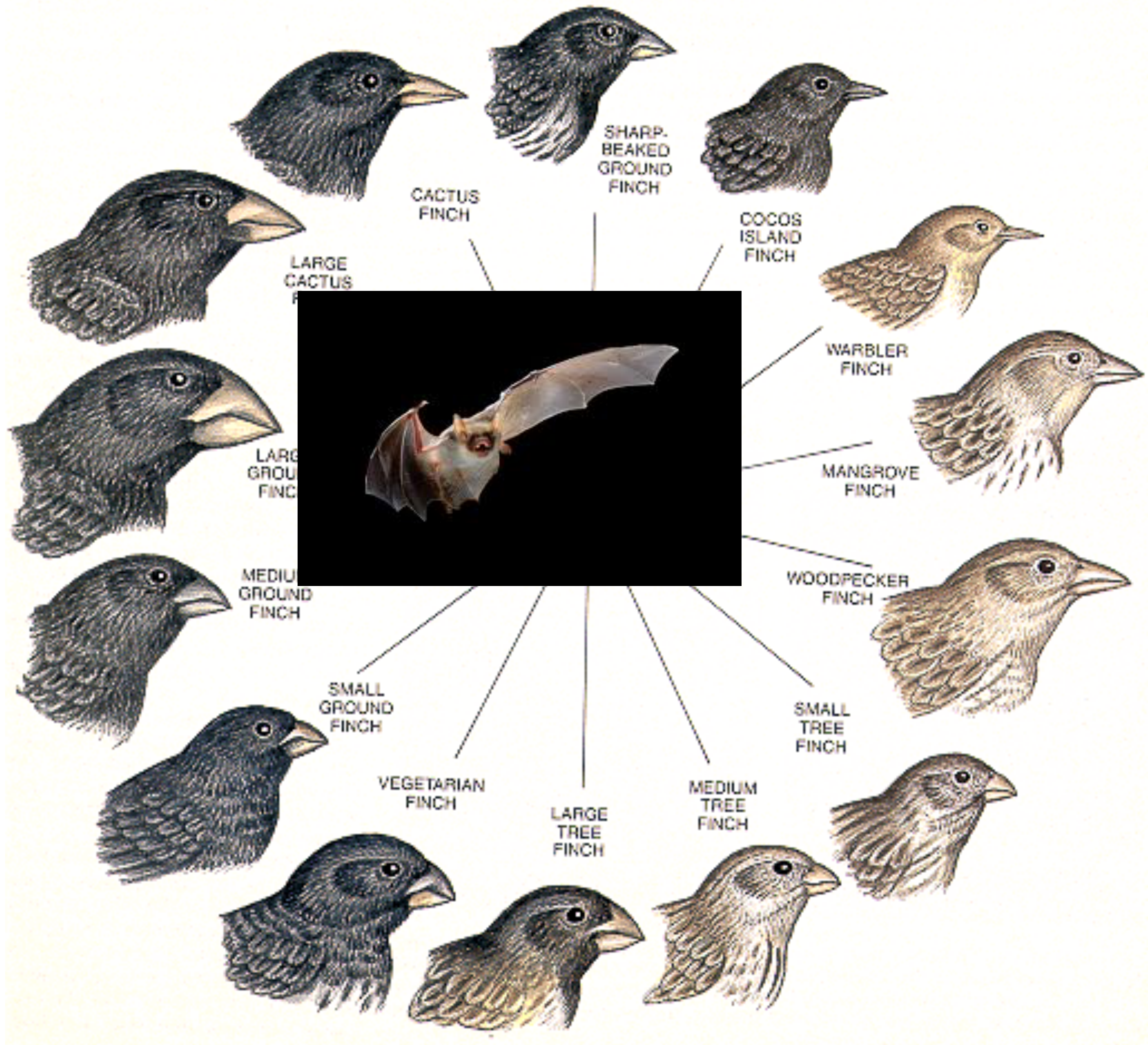
Evolution: tree of life



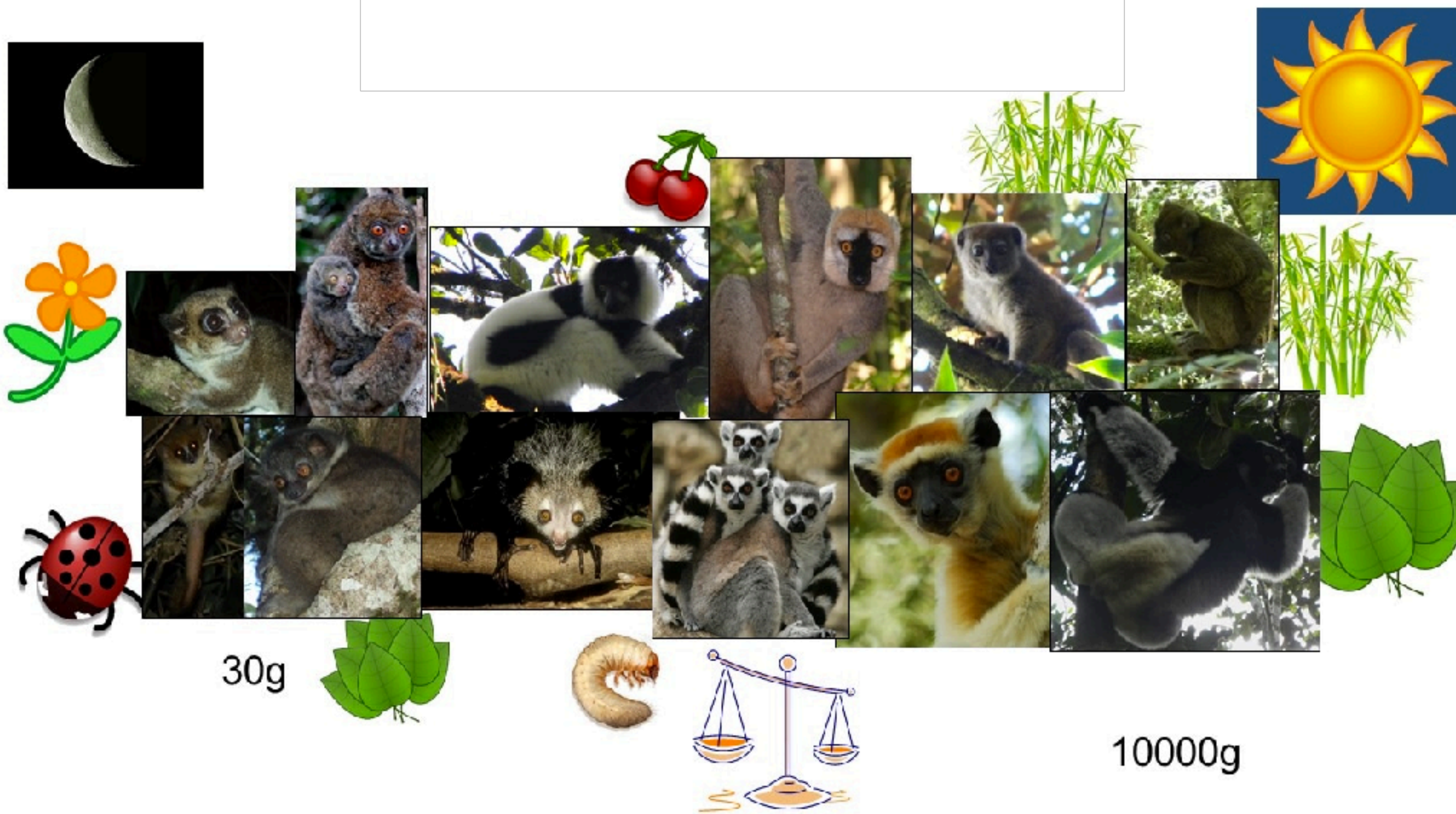
Patterns of evolution

- Divergent evolution
  Similar species begin to evolve to look extremely different from one another

Divergent evolution - adaptive radiation



Divergent evolution - adaptive radiation

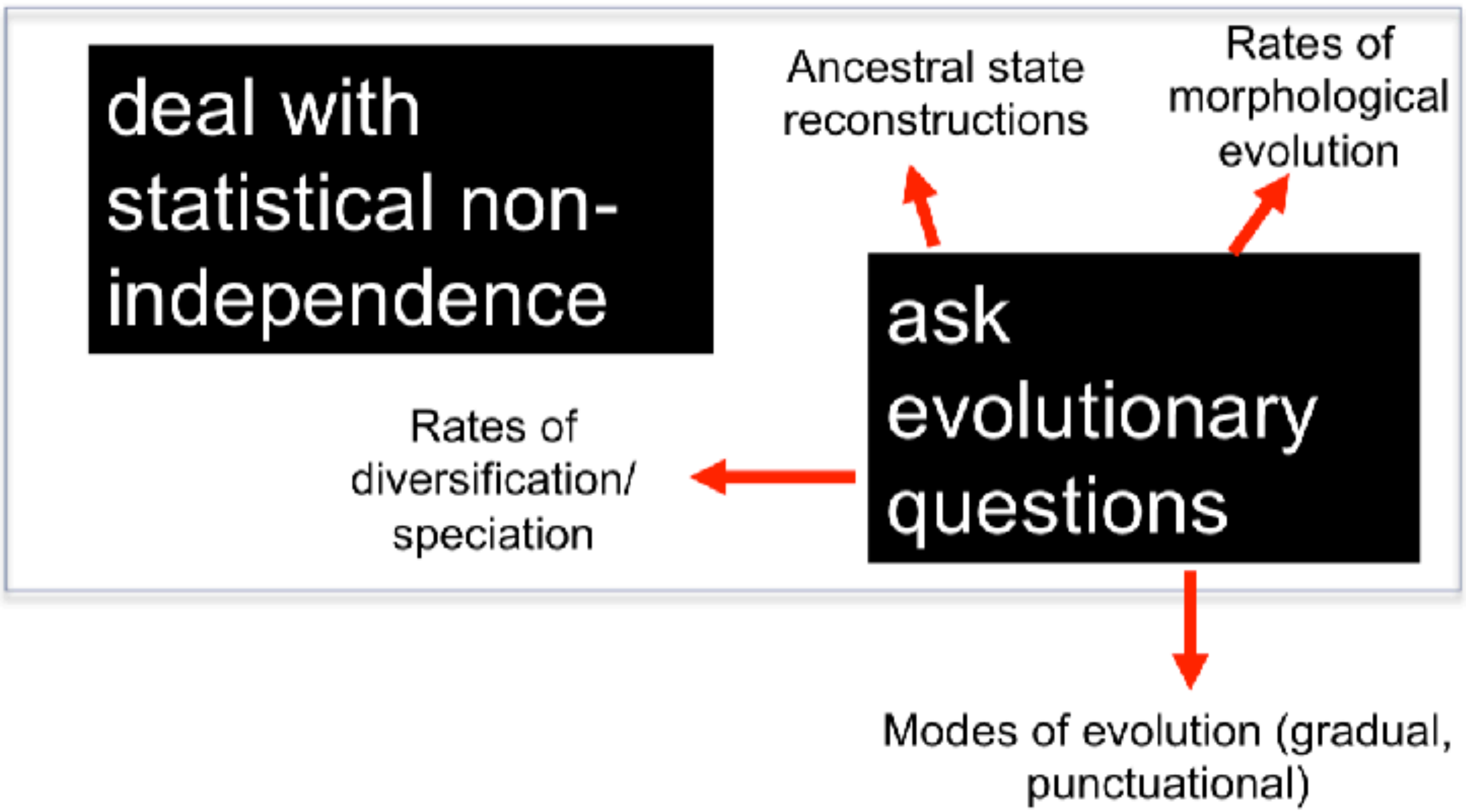


Patterns of evolution

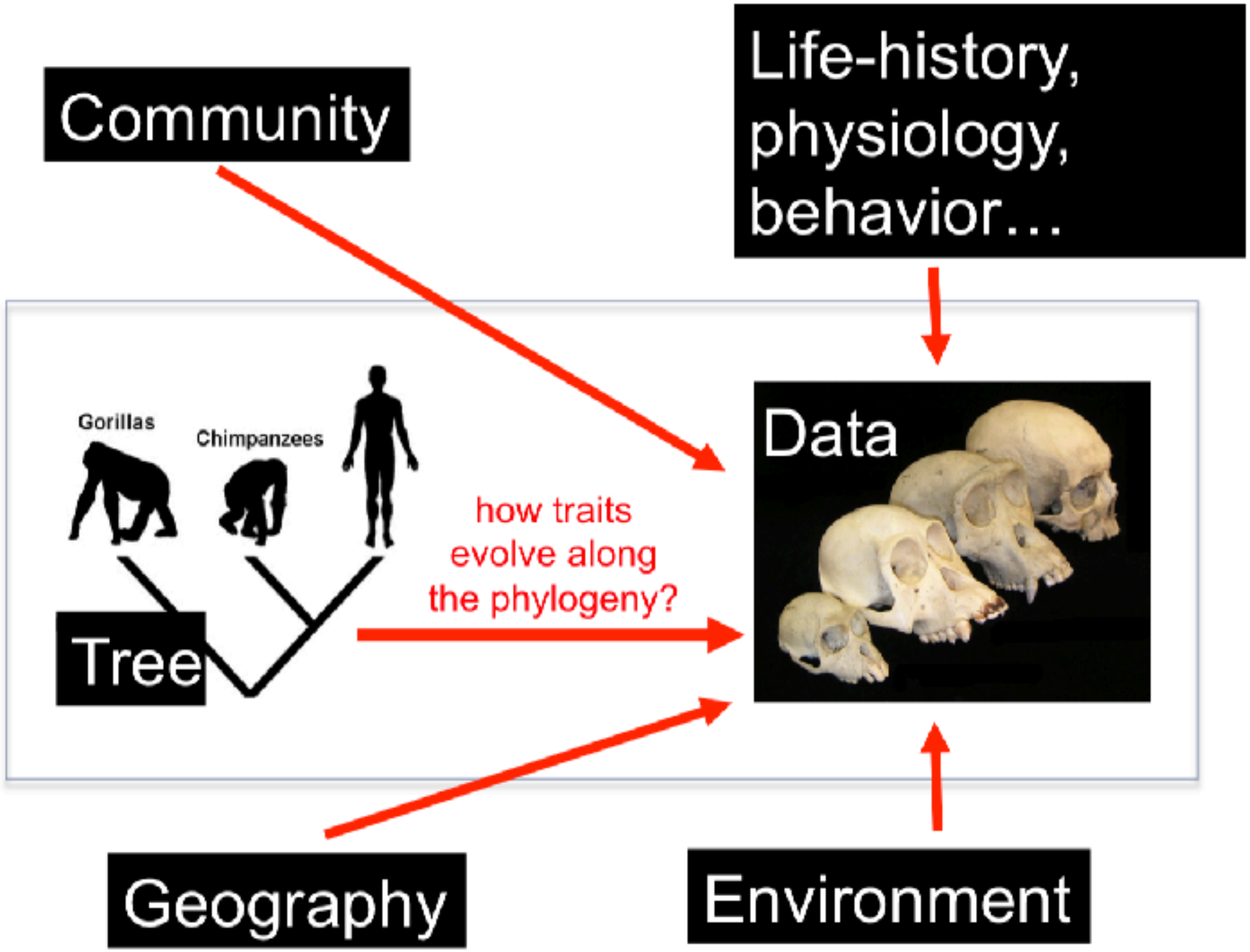
- Divergent evolution
  Similar species begin to evolve to look extremely different from one another
- Convergent evolution
  Distantly related organism evolve to look more similar

Convergent evolution - adaptive radiation

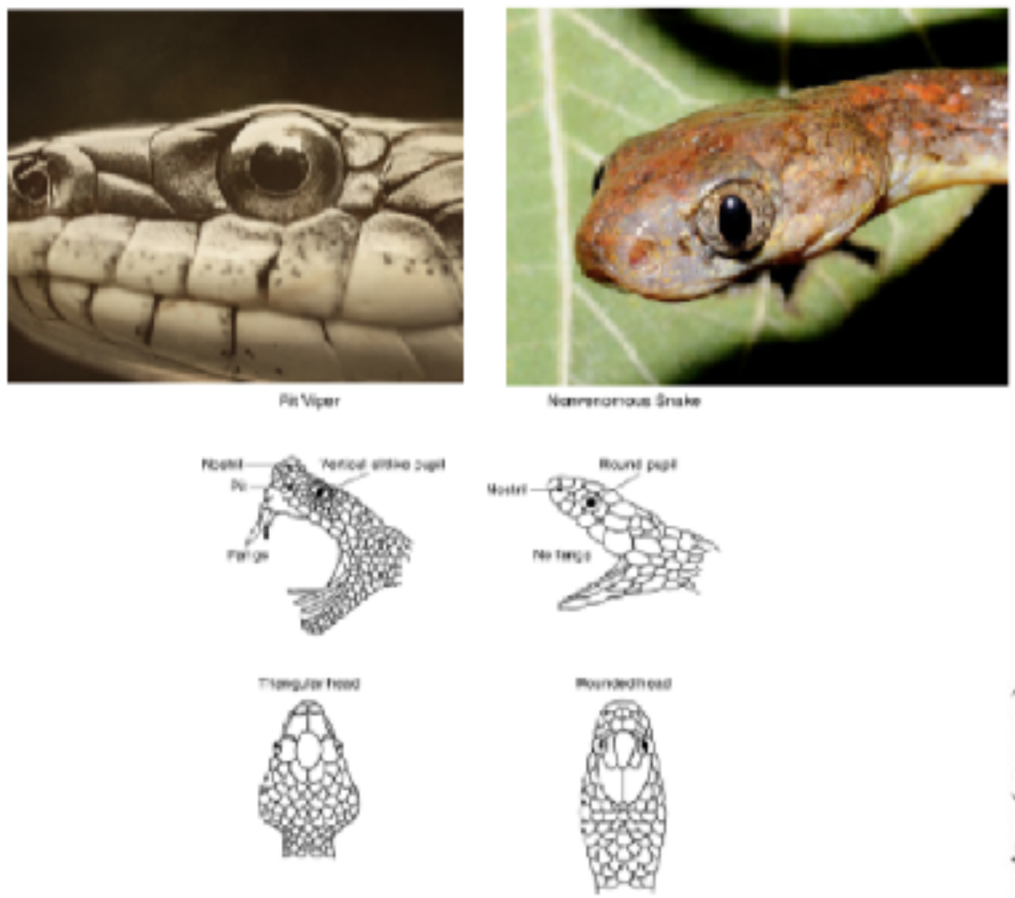
Phylogenetic methods in evolution studies



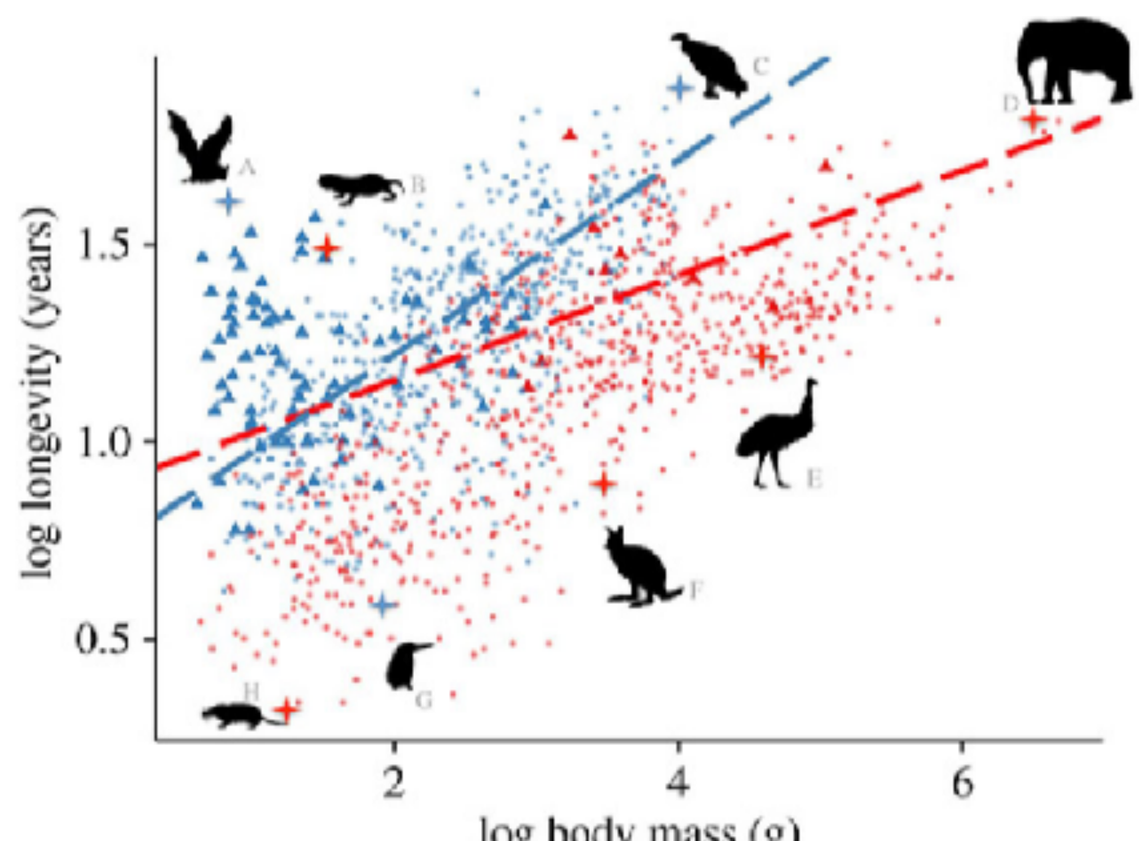
Evolution with shape data



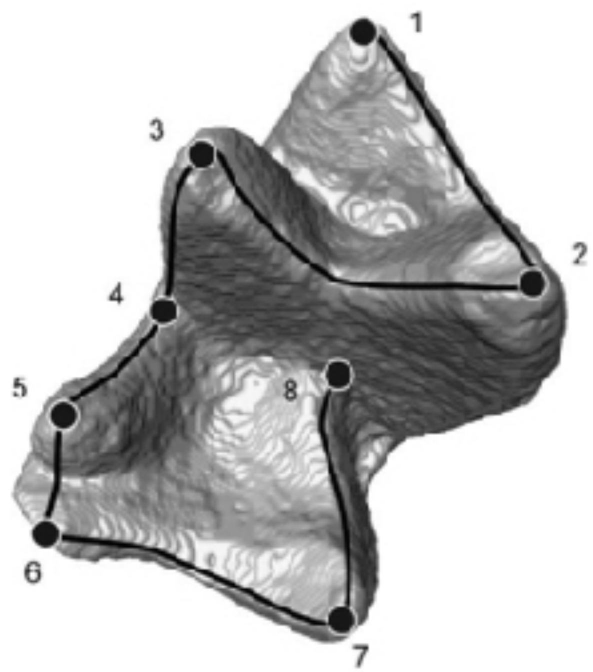
Feature is **one number**

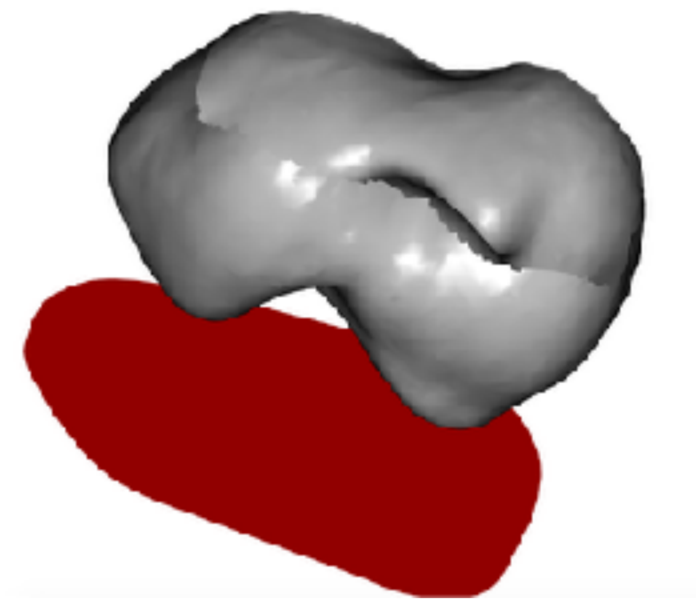


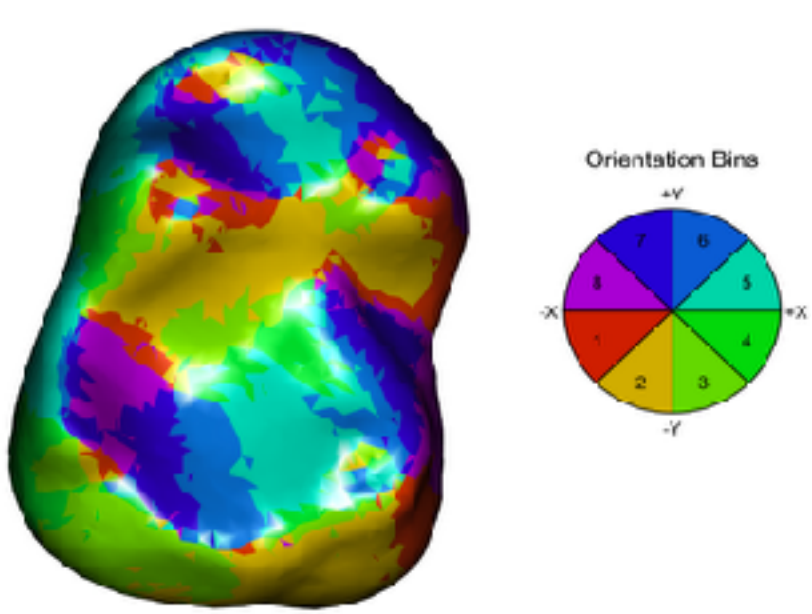
Pupil shape in snakes



Examples of features on shapes







Models for Phenotypic Evolution

- Brownian Motion (**BM**)
- Ornstein-Uhlenbeck (**OU**)
- Early Burst (**EB**)

Why these three?

- Brownian Motion (**BM**) is assumed by almost all phylogenetic comparative methods
- Ornstein-Uhlenbeck (**OU**) may capture the importance of constraints on evolution
- Early Burst (**EB**) corresponded to one idea of adaptive radiation

Brownian Motion (BM)

$$dY_{(t)} = \sigma dB_{(t)}$$

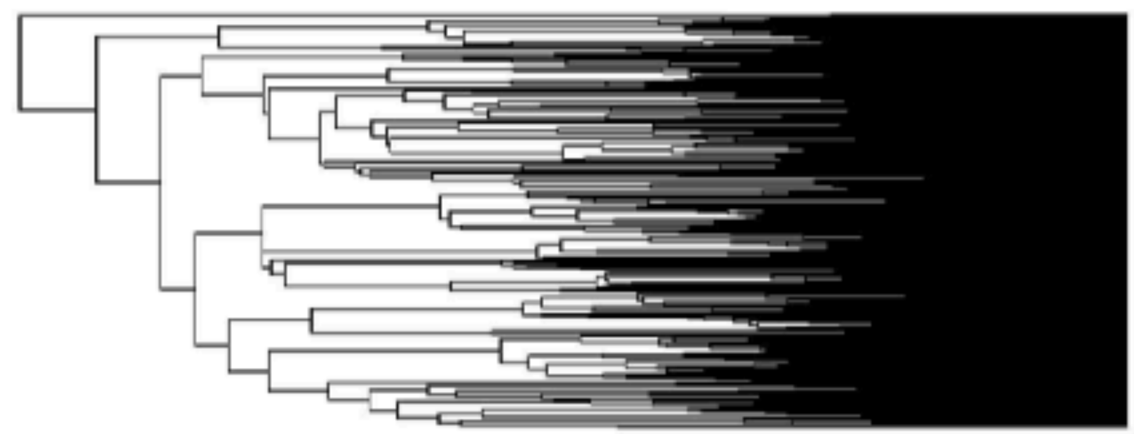
Shape features

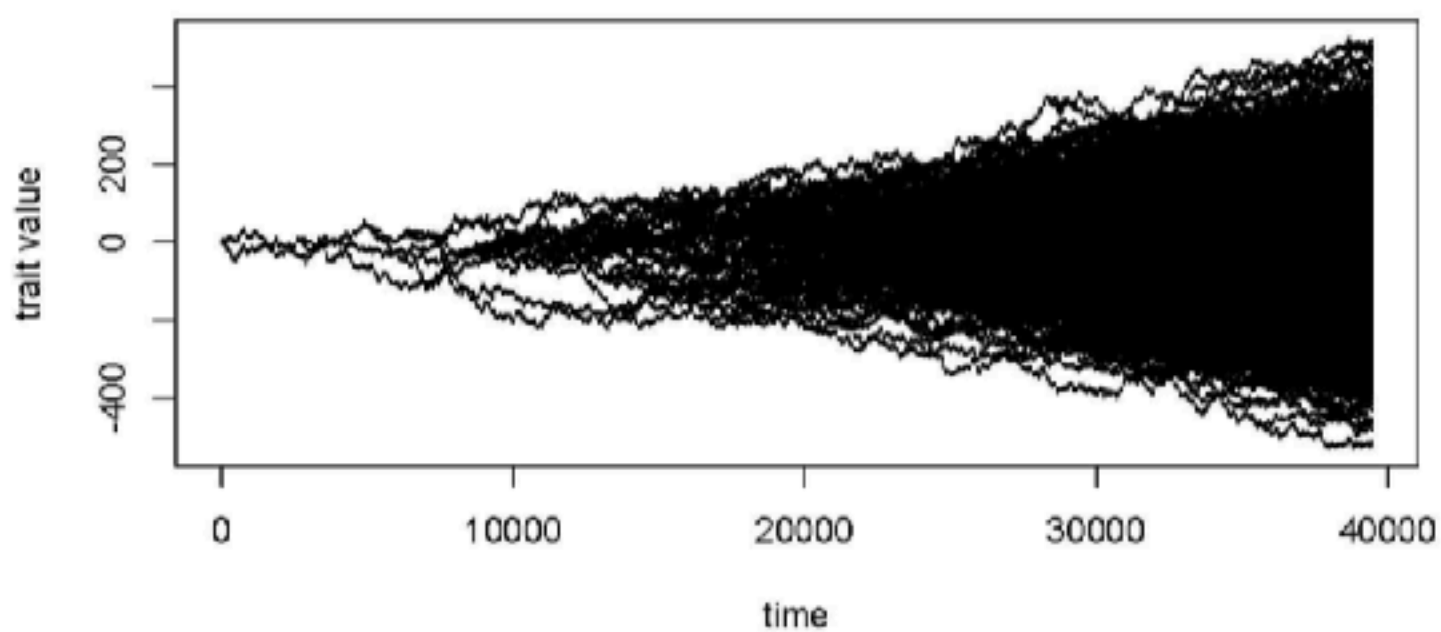
Brownian motion

Brownian motion models a constant rate evolution

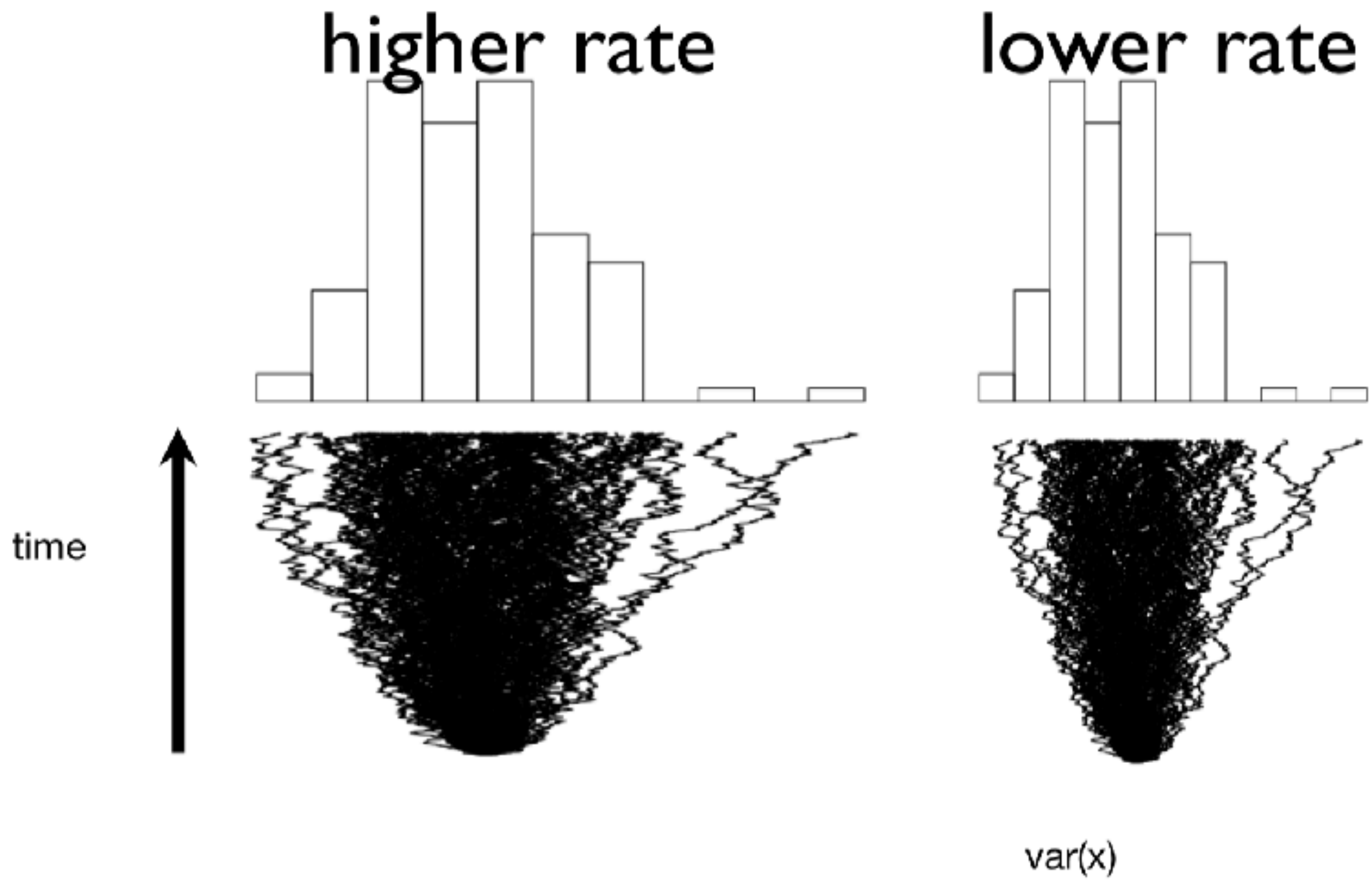
One parameter: rate (σ)

Brownian Motion



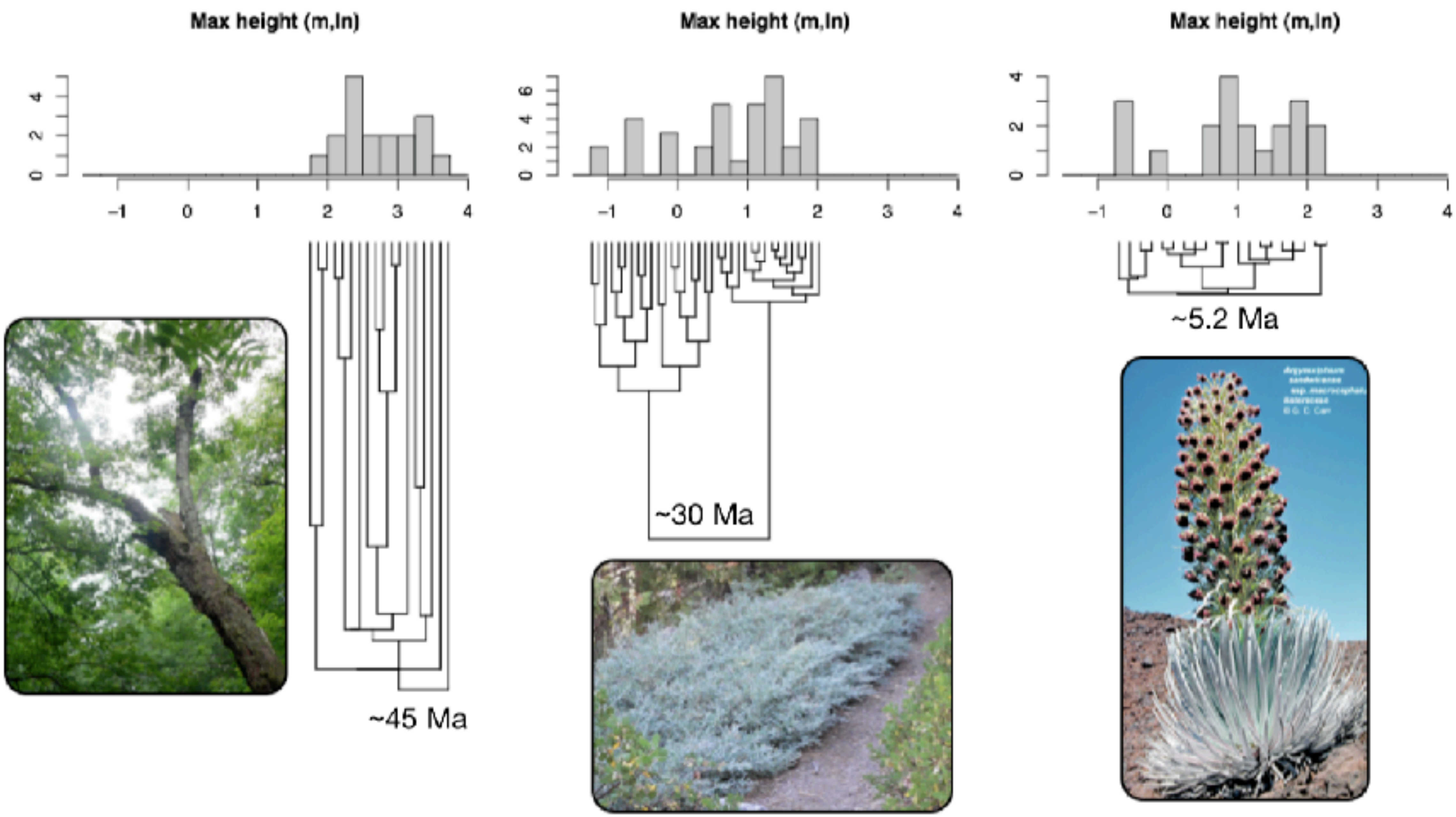


Rates of Brownian Motion



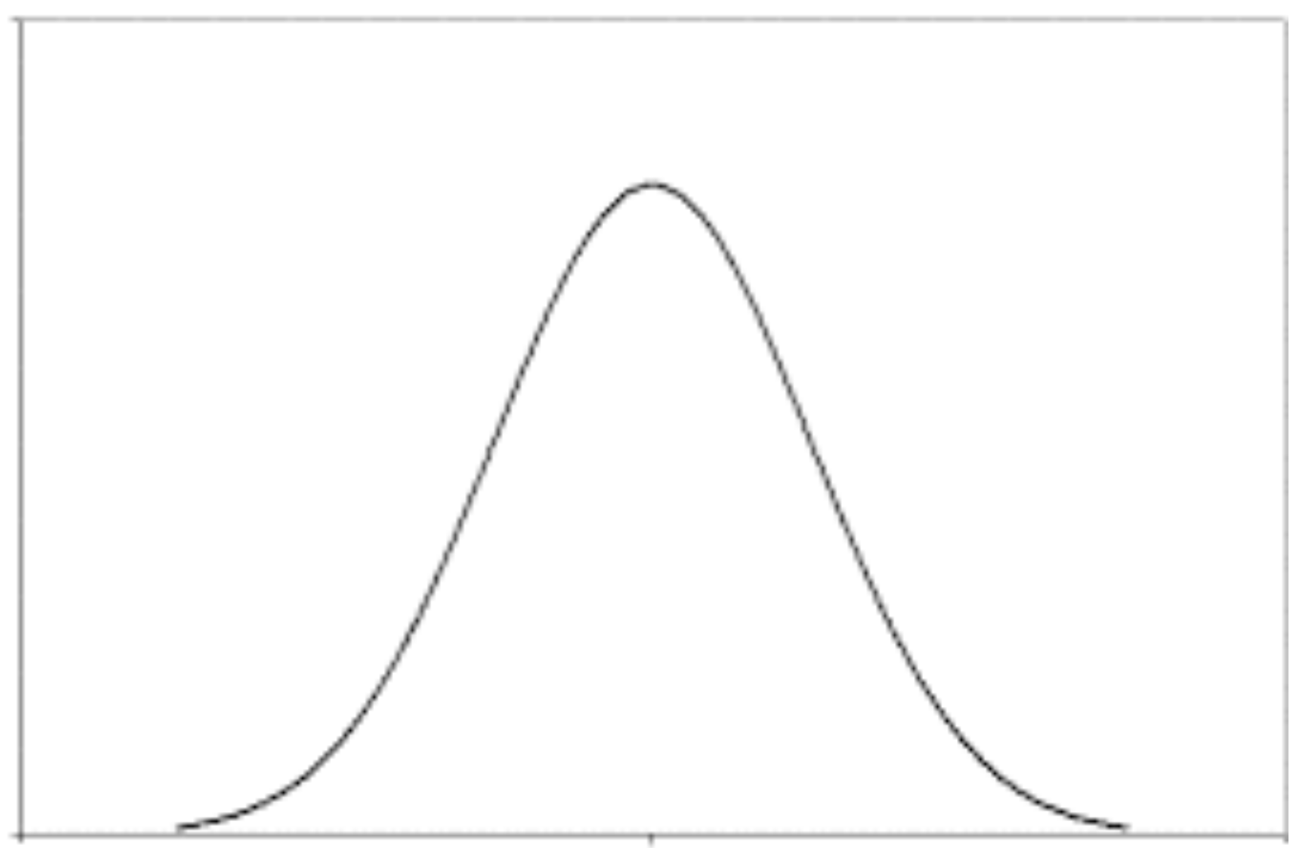
Diversification of heights in trees

rate = 0.015 felsens   0.10 felsens   0.83 felsens

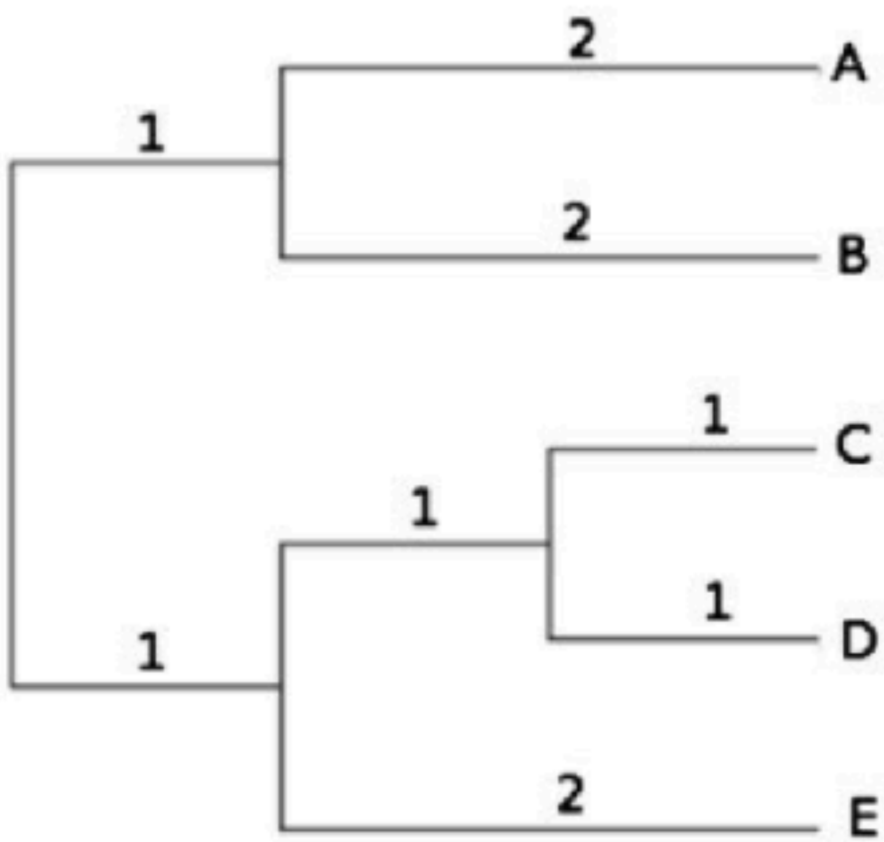


Brownian Motion on a Single Feature

$$Y \sim \mathcal{N}(0, \sigma^2 t)$$



Variance-Covariance Matrix



	A	B	C	D	E
A	3	1	0	0	0
B	1	3	0	0	0
C	0	0	3	2	1
D	0	0	2	3	1
E	0	0	1	1	3

Ornstein - Uhlenbeck (OU)

$$dY_{(t)} = \alpha\left(\theta - Y_{(t)}\right)dt + \sigma dB_{(t)}$$

Change towards optimum

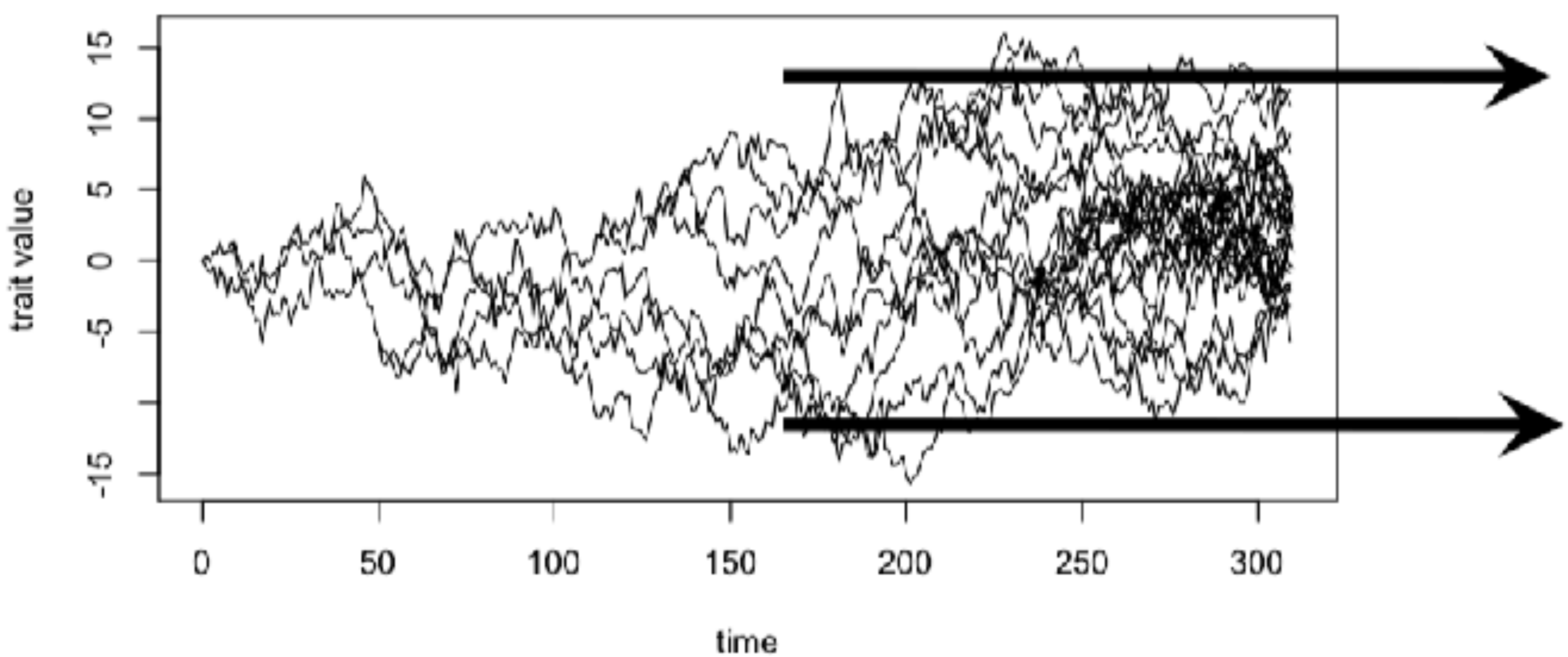
Brownian motion

Evolution has a tendency to move towards some medial value

Three parameters: rate (σ), optimal value (θ), pull towards optimal (α)

Ornstein - Uhlenbeck (OU)

$$dY_{(t)} = \alpha\left(\theta - Y_{(t)}\right)dt + \sigma dB_{(t)}$$



Early Burst (EB)

$$dY_{(t)} = \gamma(t)\ dB_{(t)}$$

$$\gamma^2(t) = \sigma^2 e^{rt}, \quad r < 0$$

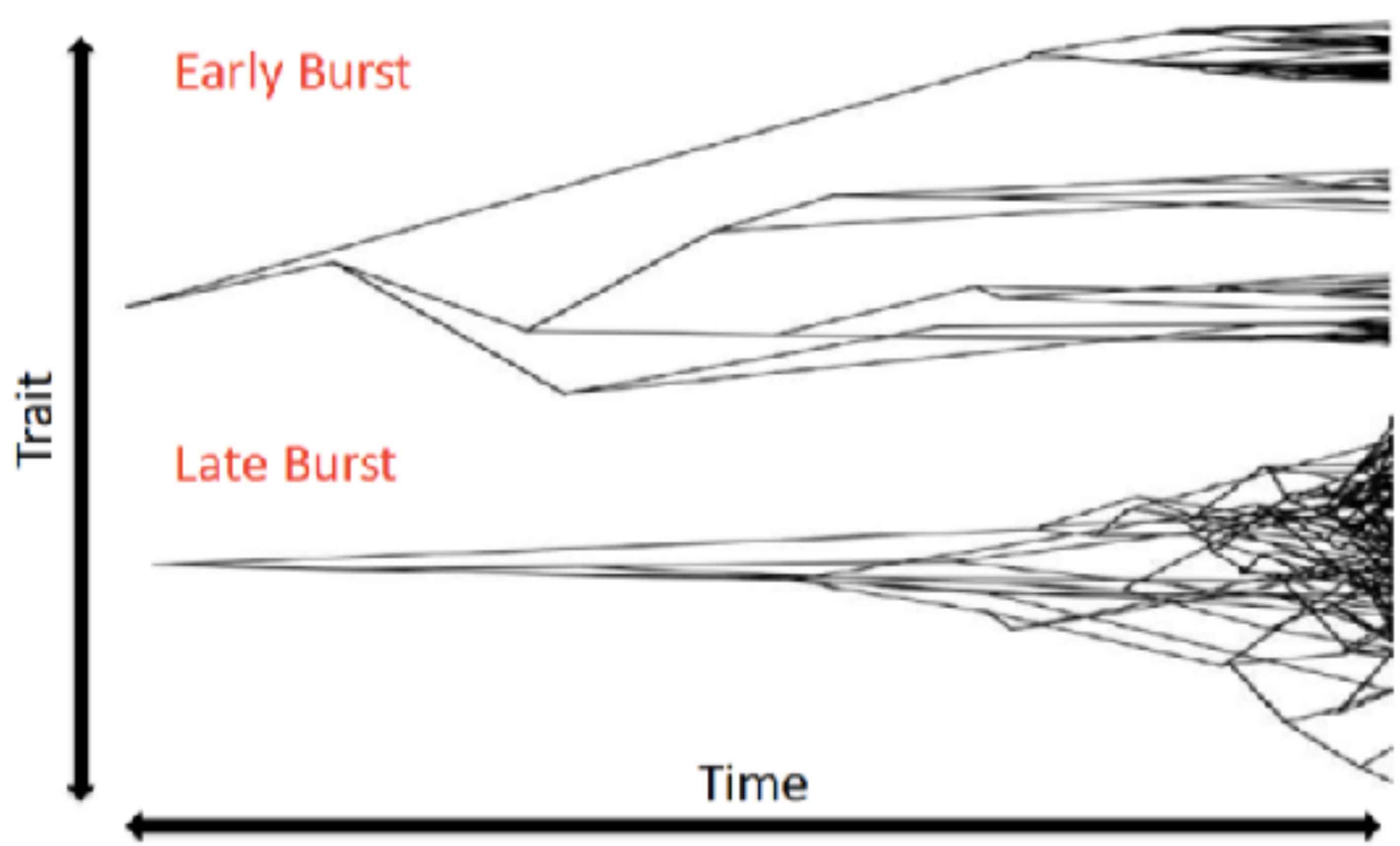
Rate of evolution is changing

Early burst models the rate of evolution is slowing through time

Two parameters: starting rate (σ) and rate change (r)

Early Burst (EB)

$$dY_{(t)} = \gamma(t)\ dB_{(t)}$$
$$\gamma^2(t) = \sigma^2 e^{rt}, \quad r < 0$$



Variance-Covariance matrix of OU and EB

$$V_{ij} = \frac{\sigma^2}{\alpha} e^{-2\alpha(T-s_{ij})}(1 - e^{-2\alpha s_{ij}})$$

$$V_{ij} = \int_0^{s_{ij}} \sigma_0^2 \, e^{rt} \, dt = \sigma_0^2 \, \frac{e^{r s_{ij}}}{r}$$

Model selection with Maximum likelihood

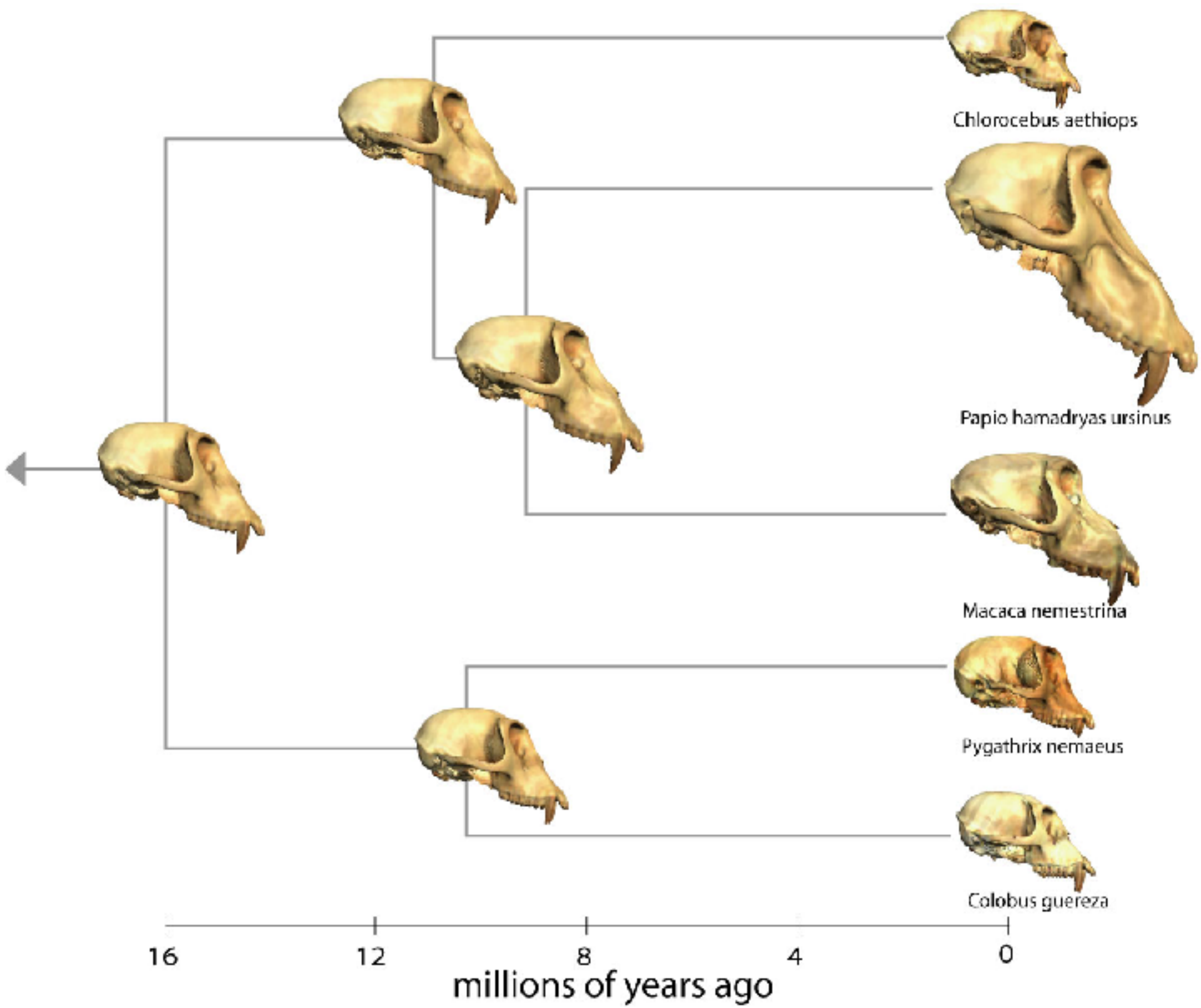
Brief discussion on maximum likelihood

possible software to achieve this

Ancestral state reconstruction

- Given
  - the tree
  - the shape feature
  - the evolution model
- Goal is to obtain estimates for shape feature at the ancestral nodes or along the branches of phylogeny

Reconstruct surface at an ancestral state



Evolution model with the entire shape

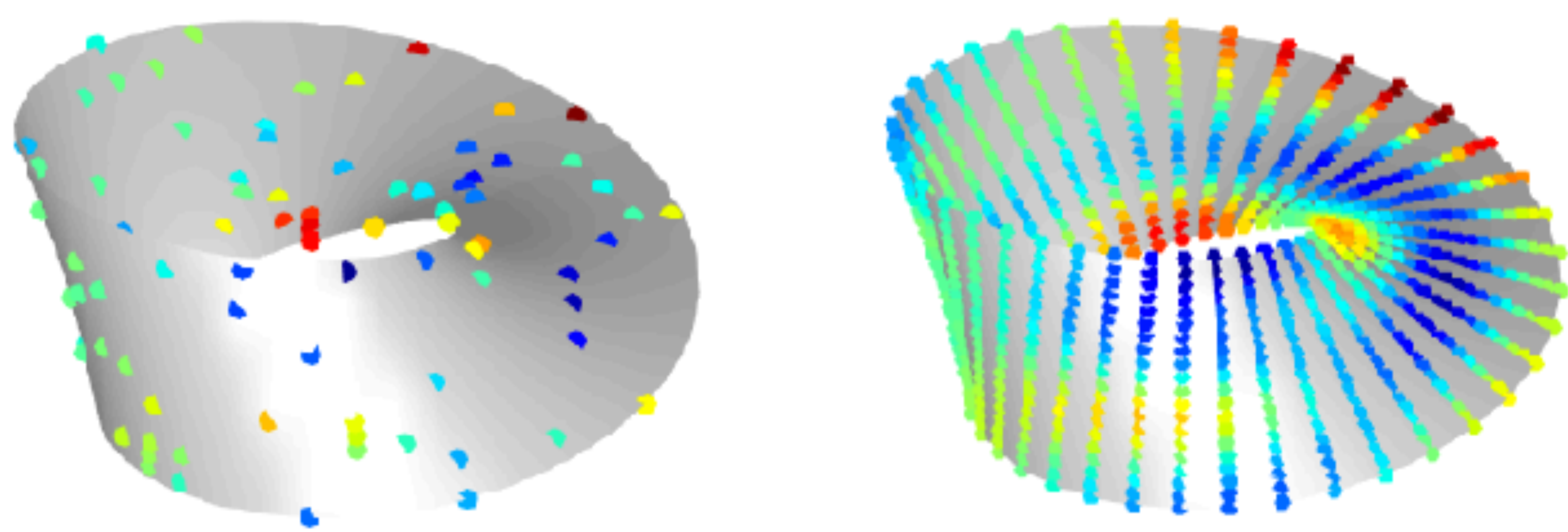
Gaussian process

Definition $f \sim \mathcal{GP}(m, k)$ is a *Gaussian process*

$$\textit{mean } m : \mathcal{X} \to \mathbb{R}$$
$$\textit{covariance } k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$$

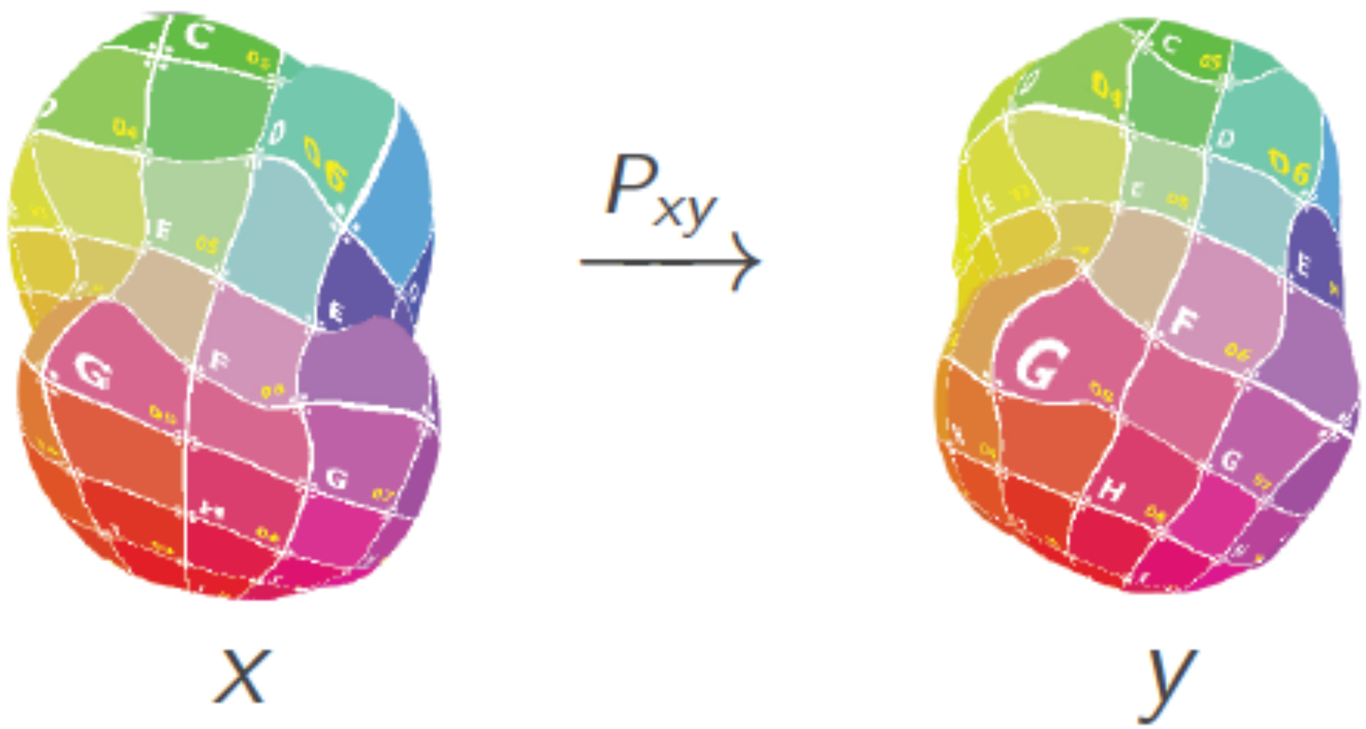
if any finite realization $\{f(x_i)\}_{i=1}^n$ is jointly normal.

Prediction



(a) Observed (b) Predicted

Correspondence maps again



Gaussian process on correspondence maps

Shape evolution on the data points level

$$y_i = f(e_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2).$$
$$f \sim \mathcal{GP}(0, k)$$

Define k as a product (evolution $\times$ geometry).
Let $e_i = (z_i, u_i), e_j = (z_j, u_j)$ by local trivialization.

$$k(e_i, e_j) = k(z_i, u_i; z_j, u_j) = k_{\text{evol}}(z_i, z_j) \cdot k_{\text{fibre}}(u_i, u_j)$$

where

$$k_{\text{fibre}}(u_i, u_j) = \exp\left(-\frac{d^2(P_{z_i z_j}(u_i), u_j)}{t_F}\right)$$

Model selection with the entire shape

Our parameter space is

$$\theta = \{\theta_{\text{evol}}, \theta_{\text{fibre}}\}$$

Goal is to compute posterior

$$p(\theta|\mathcal{D}) = \frac{p(\theta)p(\mathcal{D}|\theta)}{\int_{\theta'} p(\theta')p(\mathcal{D}|\theta')d\theta'}$$

a probabilistic analogue to the projected kernel method:
θ_{evol} contains information about the evolutionary process.
Compute the marginal posterior on the evolution parameter

$$p(\theta_{\text{evol}}|\mathcal{D}) = \int p(\theta_{\text{evol}}|\theta^*, \mathcal{D})p(\theta^*|\mathcal{D})d\theta^*,$$

Compute with MCMC with a discrete prior on θ.

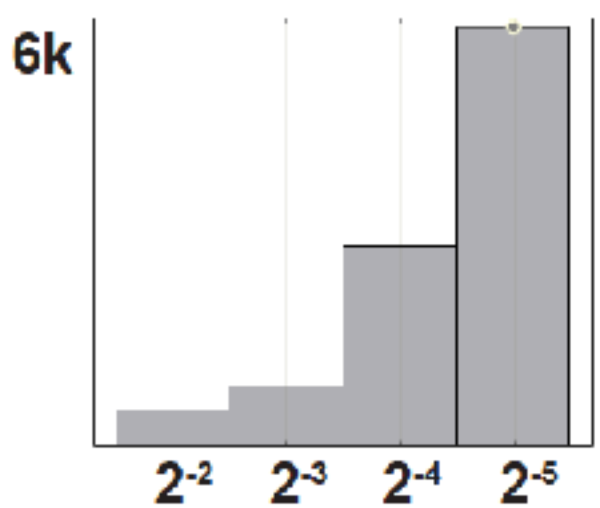
Lemurs of Madagascar

Recall EB model.

$$k_{\text{evol}}(z_i, z_j) = \tau^2 \left(\frac{e^{rt_{ij}} - 1}{r} \right), \quad r < 0.$$

r close to 0 returning BM.

We apply GP on fibre bundle to 40 teeth of 40 lemurs.



Posterior of $-r$

Our approach suggests a BM model for lemurs of Madagascar.

Future work