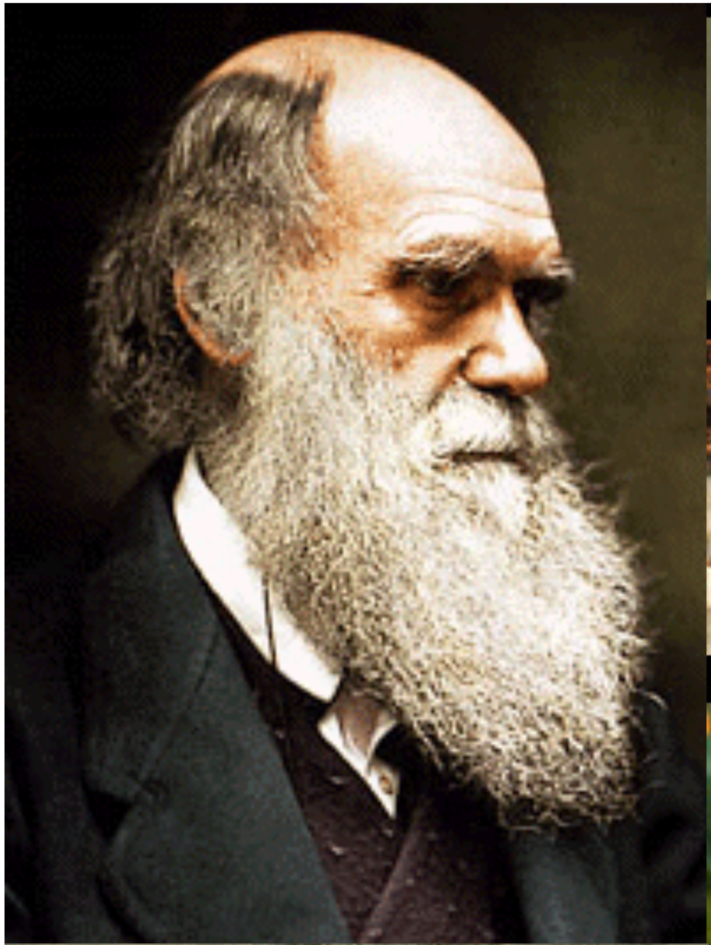
Trait evolution and ancestral state reconstruction

László Zsolt Garamszegi
Estación Biológica de Doñana-CSIC, Seville, Spain



small ground finch

medium ground finch

large ground finch

sharp-beaked ground finch

cactus finch

large cactus finch

small tree finch

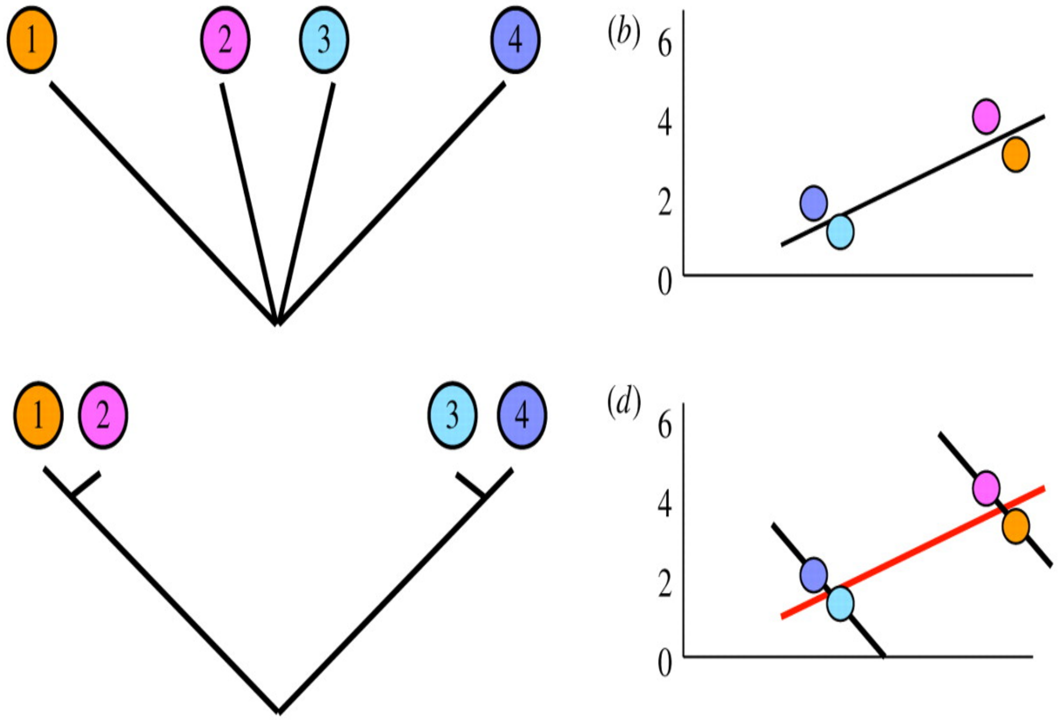
large tree finch?

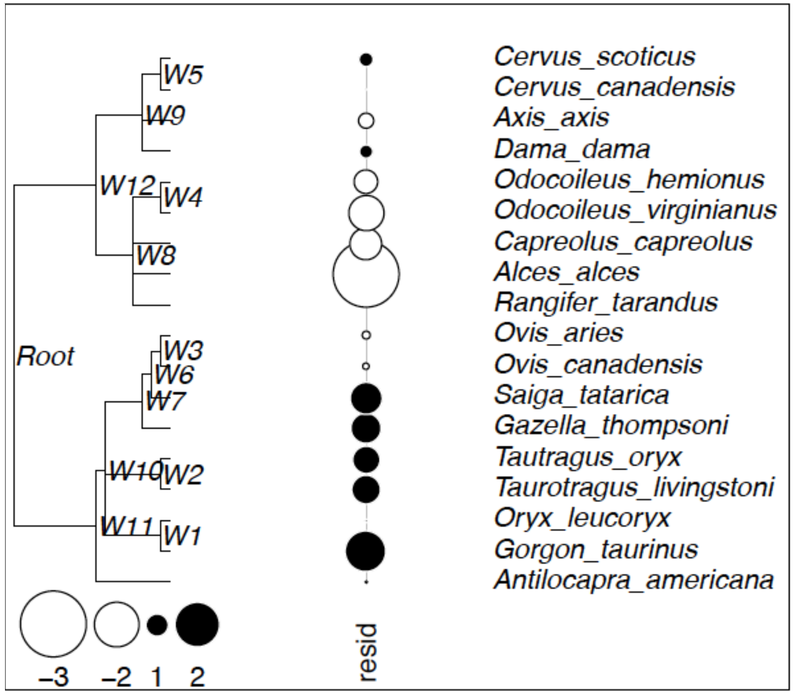
vegetarian finch

woodpecker finch

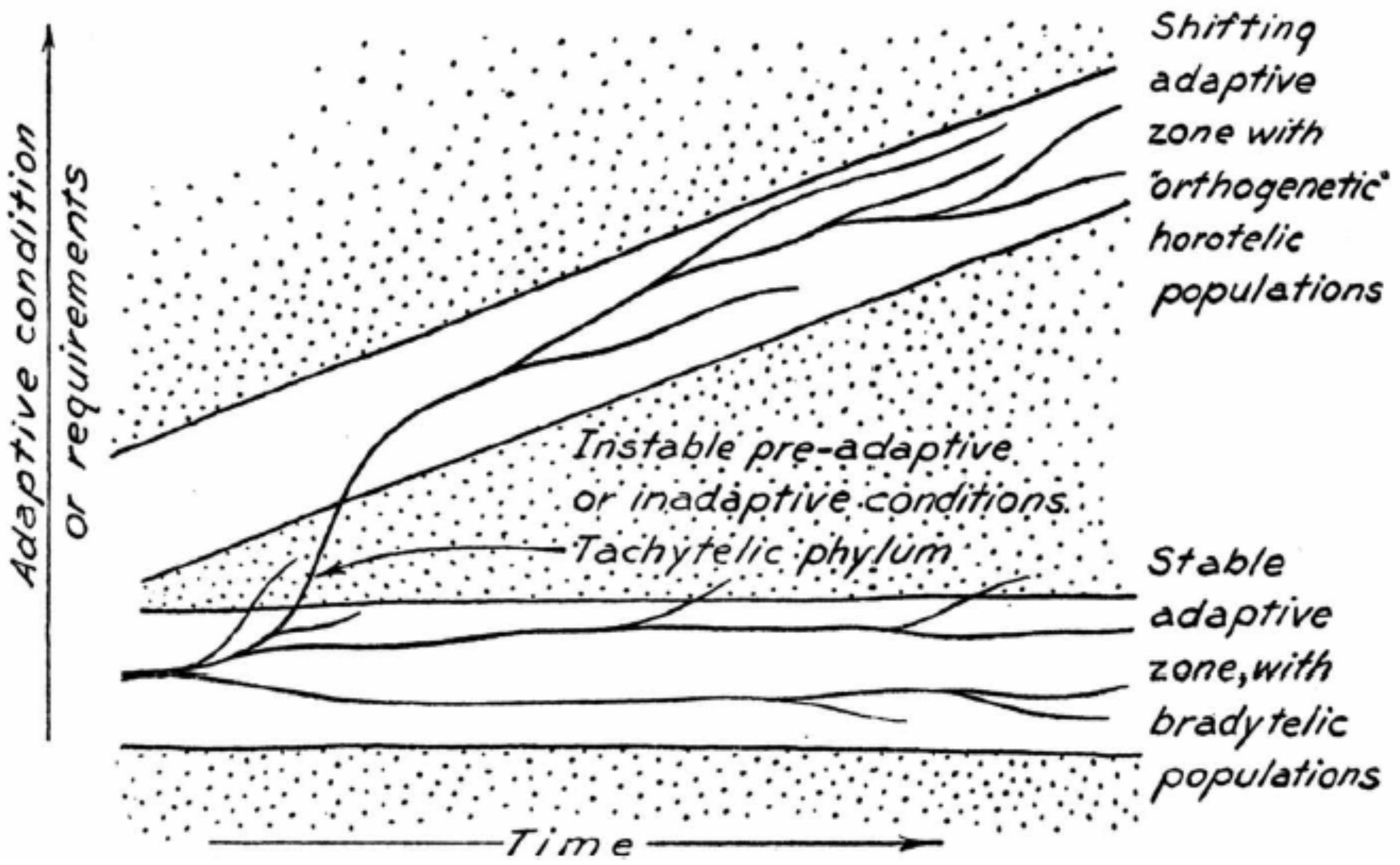
warbler finch

Interspecific data are not independent



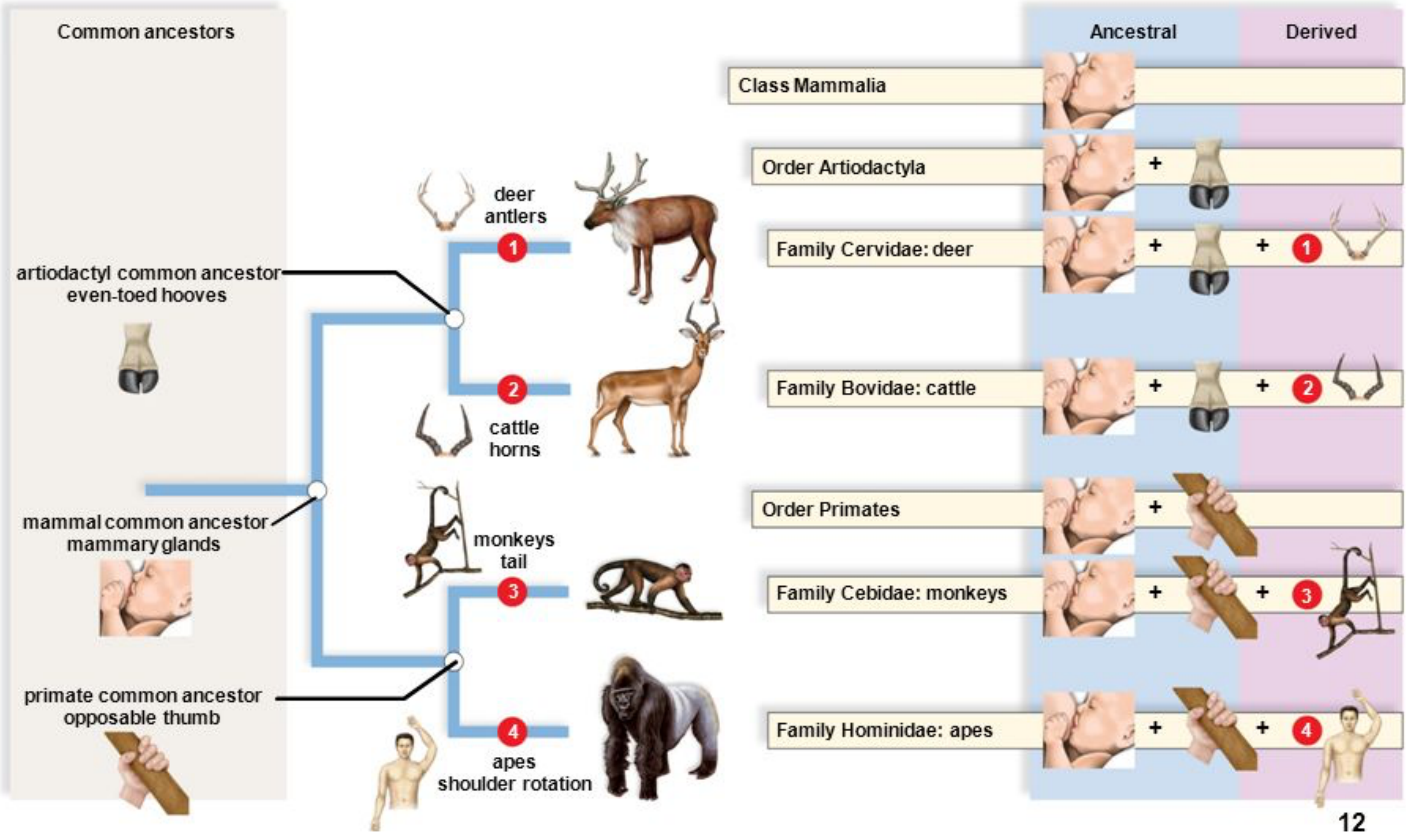


Felsenstein 1985



Simpson 1946

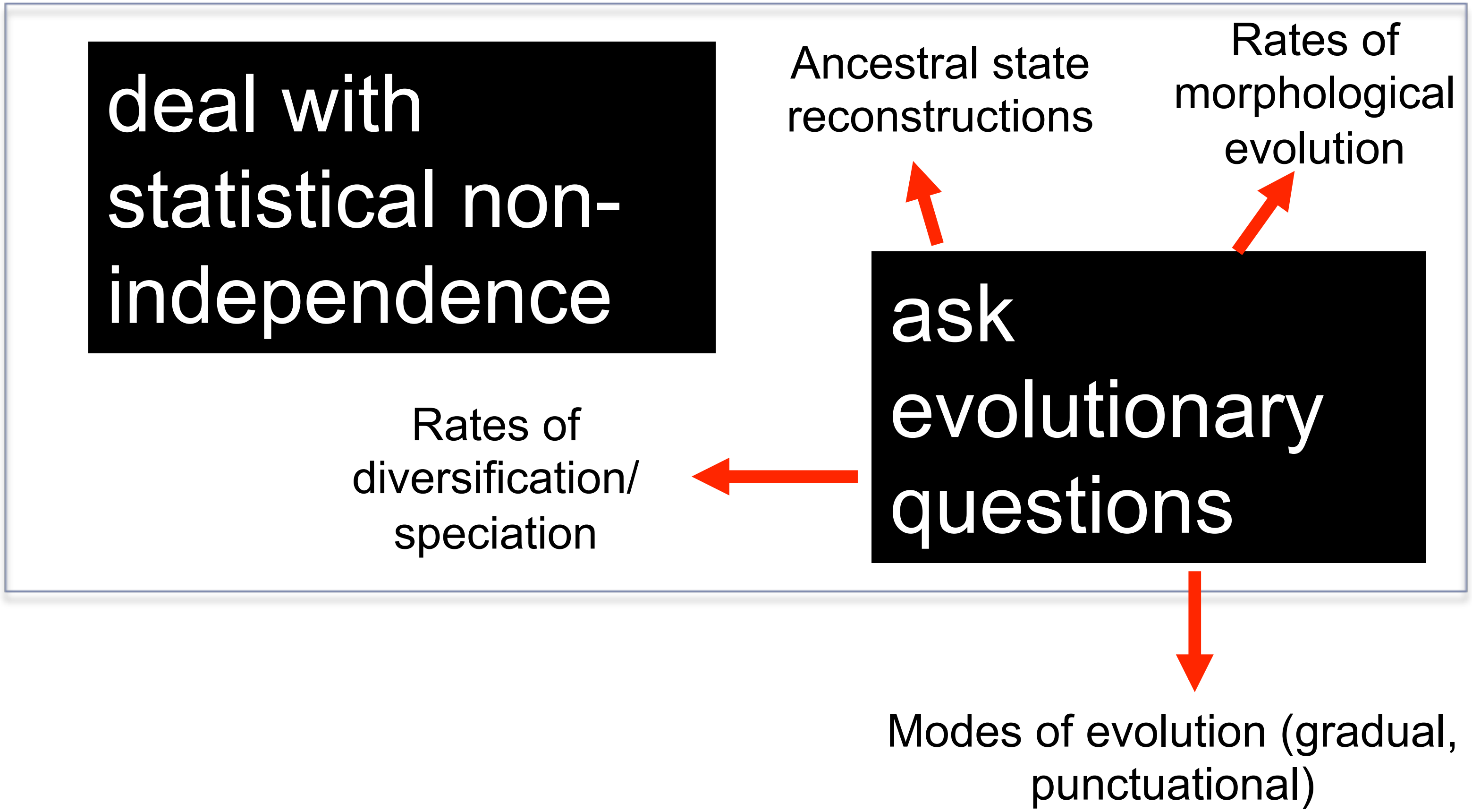
The relationship between phylogeny, classification and traits



Axis of biodiversity

		Taxonomic	Phenotypic
Metric	Non-Phylo	Number of Species	Disparity (variance)
	Phylogenetically corrected	Diversification rate	Rate of trait Evolution

Phylogenetic methods can be used to



deal with statistical non-independence

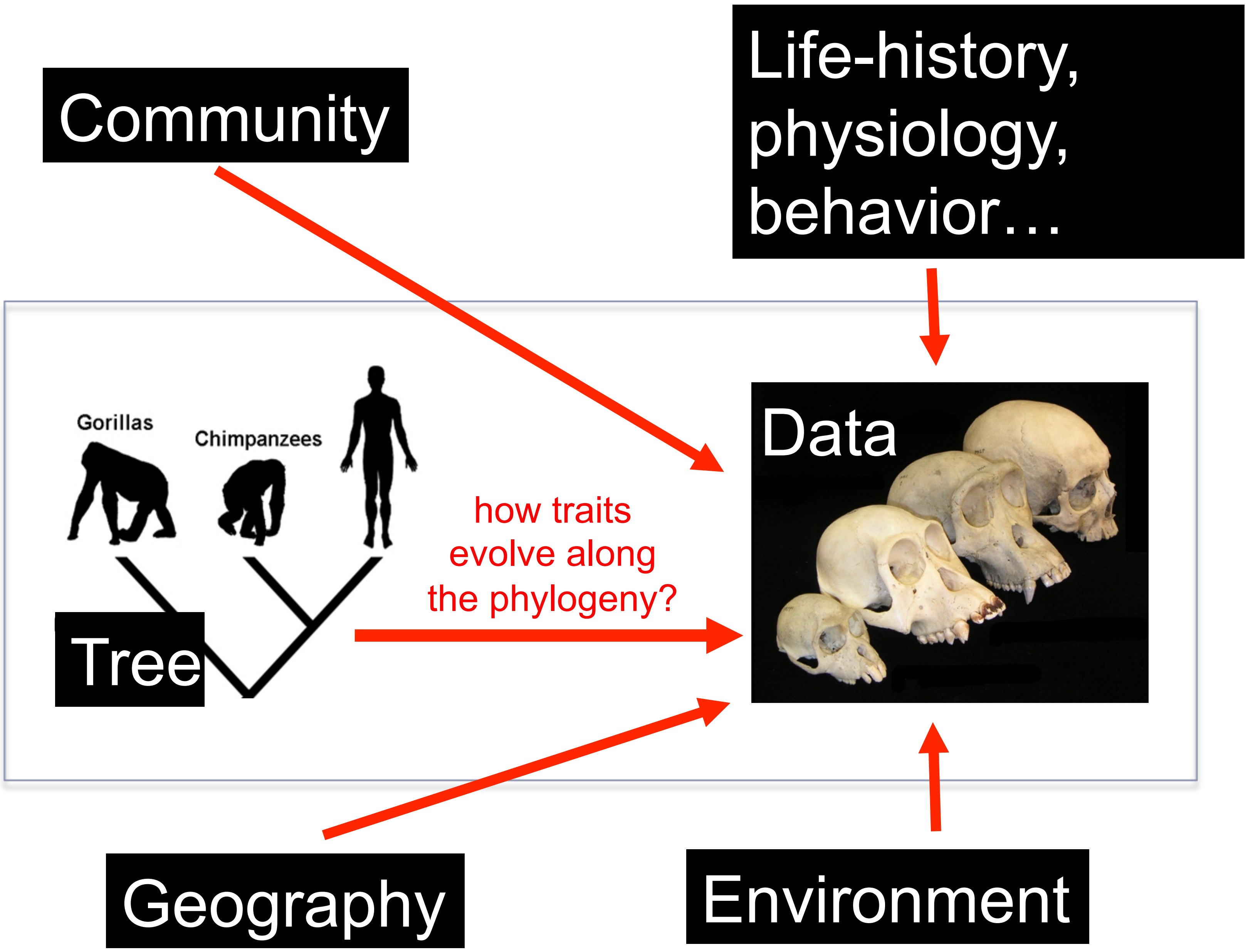
ask evolutionary questions

Ancestral state reconstructions

Rates of morphological evolution

Rates of diversification/ speciation

Modes of evolution (gradual, punctuational)



Community

Life-history,
physiology,
behavior…

Tree

how traits
evolve along
the phylogeny?

Data

Geography

Environment

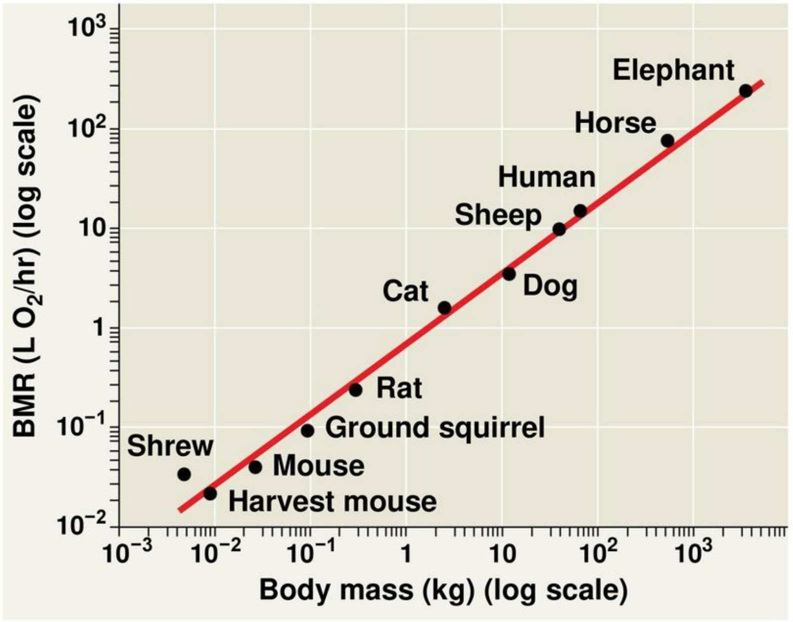
Outline

- Definitions and some assumptions
- Models of evolution
  - Continuous
    - Brownian, Early Burst, Ornstein-Uhlenbeck, Trend
  - Discrete
    - Mk model, extended Mk models (SYM, ARD), threshold model
  - Phylogenetic signal
- Ancestral-state reconstructions
  - Parsimony
  - Maximum-likelihood
  - Stochastic mapping

Definitions: what is trait?

- Heritable and reliable species-specific characteristics
    - morphology
    - behavior
    - physiology
    - life-history
    - gene sequence
- Continuous vs. discrete

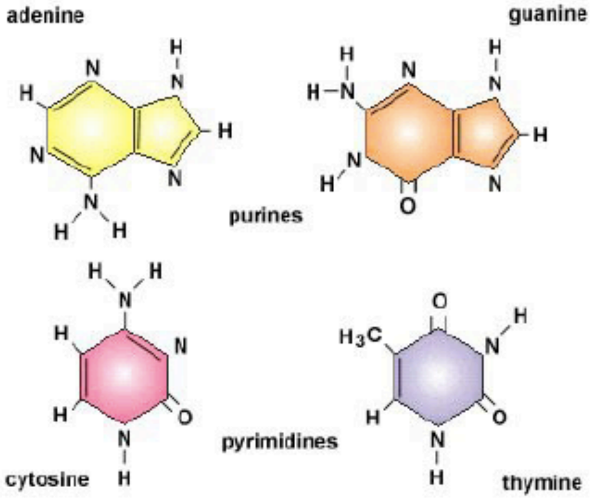
Figure 40.20a



(a) Relationship of basal metabolic rate (BMR) to body size for various mammals

Definitions: what is trait?

- Discretely-coded traits
  - Intrinsically discrete traits




Wings


© Kim A. Cabrera

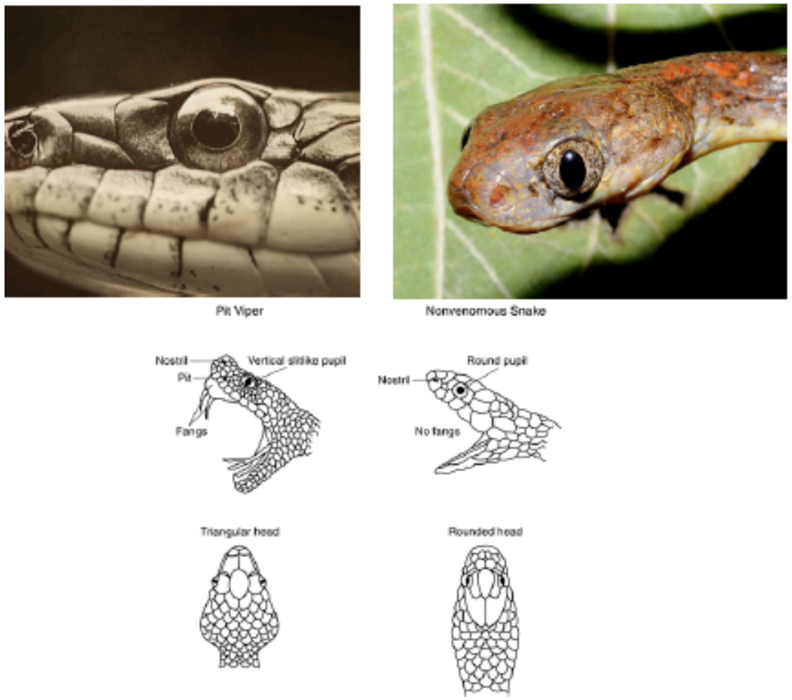
No wings


Aquatic

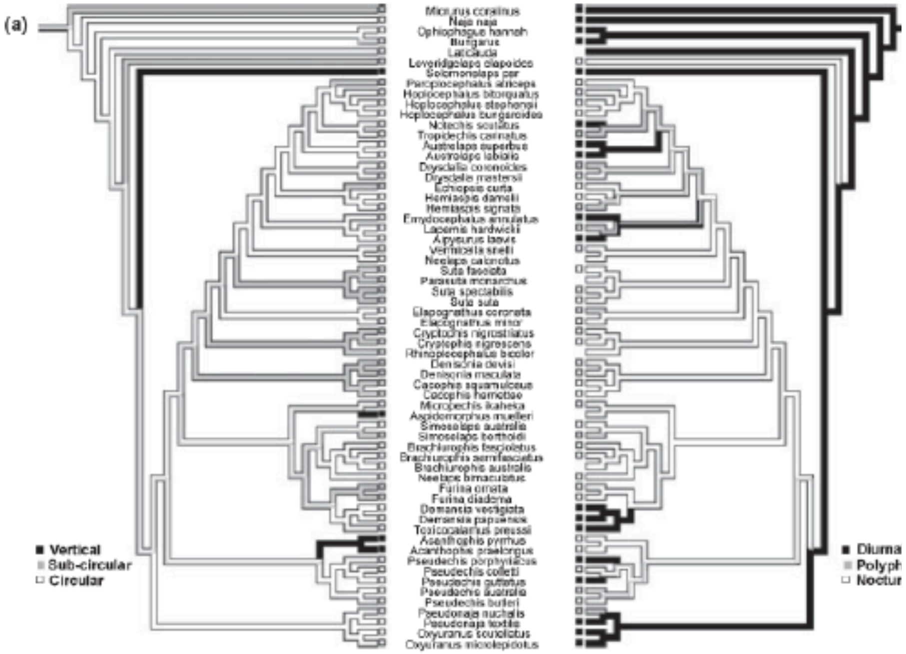

Terrestrial

Definitions: what is trait?

- Discretely-coded traits
  - Discretize continuous traits



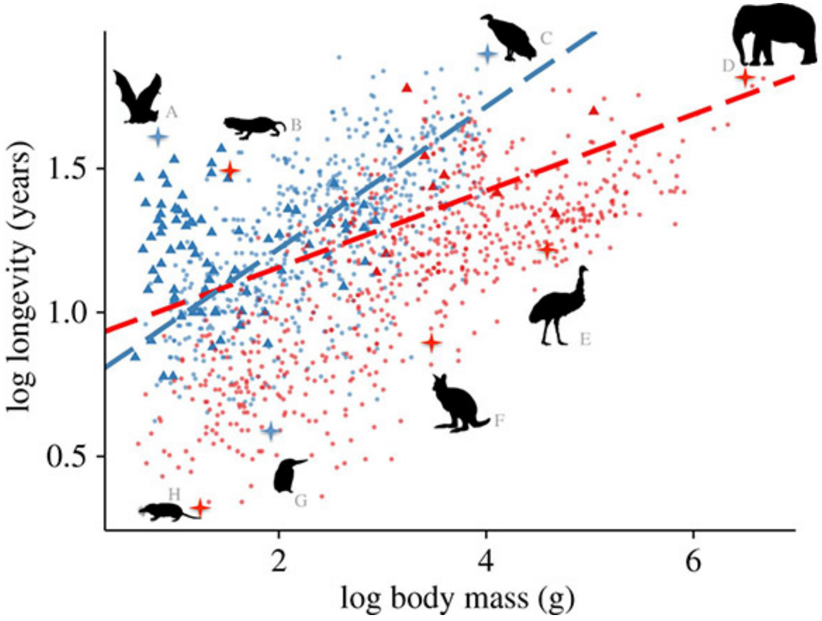
Pupil shape in snakes

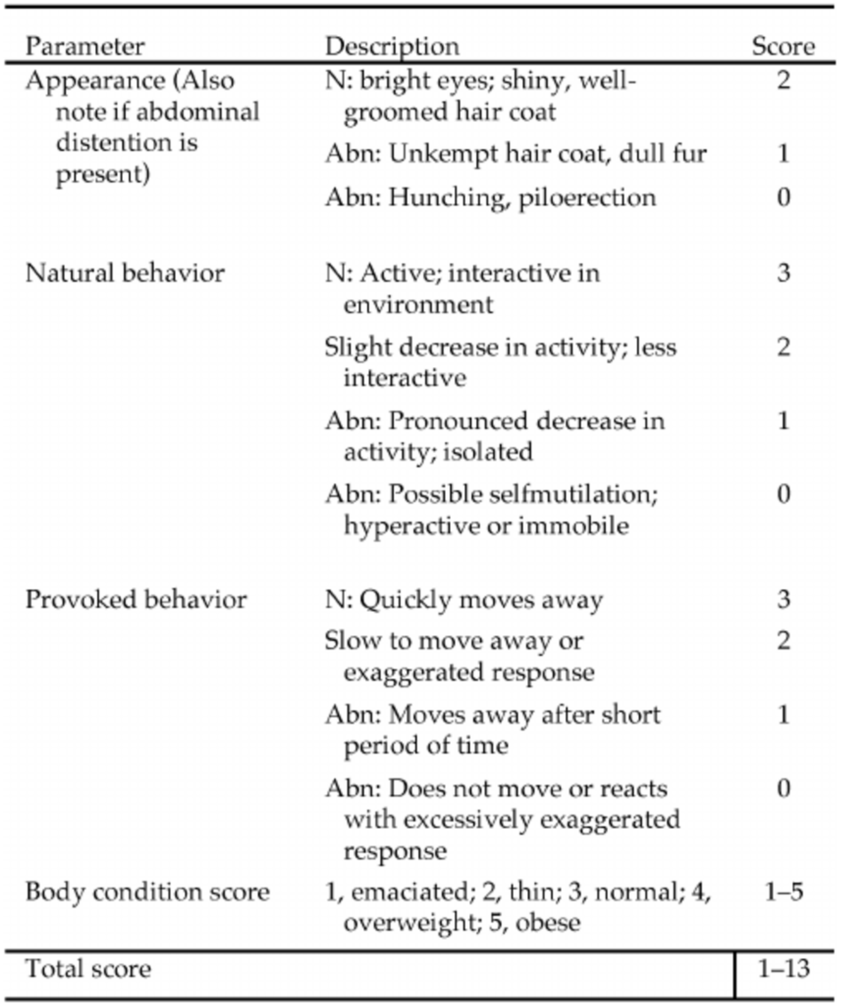


Brischoux et al. 2010

Definitions: what is trait?

- Continuous traits
  - Ordinal
  - Interval



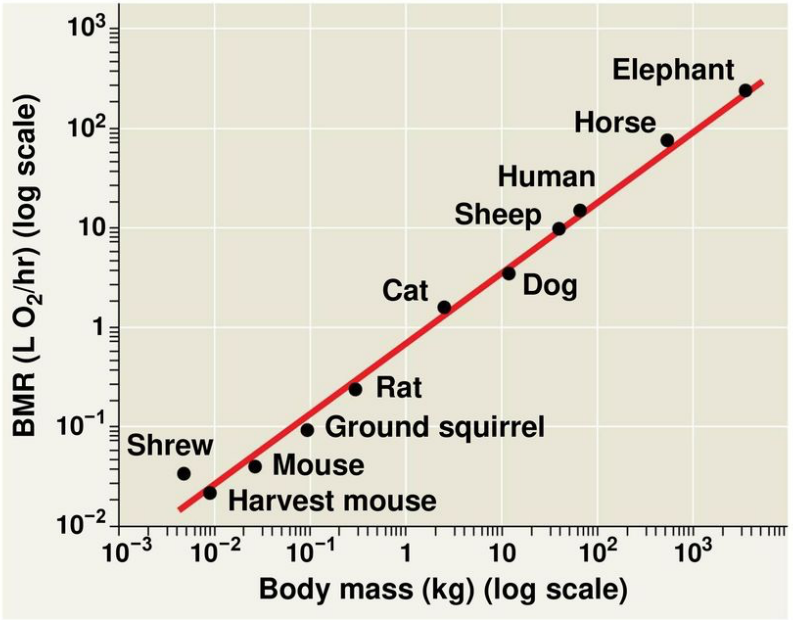


Parameter	Description	Score
Appearance (Also note if abdominal distention is present)	N: bright eyes; shiny, well-groomed hair coat	2
	Abn: Unkempt hair coat, dull fur	1
	Abn: Hunching, piloerection	0
Natural behavior	N: Active; interactive in environment	3
	Slight decrease in activity; less interactive	2
	Abn: Pronounced decrease in activity; isolated	1
	Abn: Possible selfmutilation; hyperactive or immobile	0
Provoked behavior	N: Quickly moves away	3
	Slow to move away or exaggerated response	2
	Abn: Moves away after short period of time	1
	Abn: Does not move or reacts with excessively exaggerated response	0
Body condition score	1, emaciated; 2, thin; 3, normal; 4, overweight; 5, obese	1–5
Total score		1–13

Definitions: what is trait?

- Heritable and reliable species-specific characteristics
  - morphology
  - behavior
  - physiology
  - life-history
  - gene sequence
- Continuous vs. discrete
- Often measured with error
  - Within-species variance
  - Use of proxies
  - Apple / orange problem
- Original vs. log transformed scale

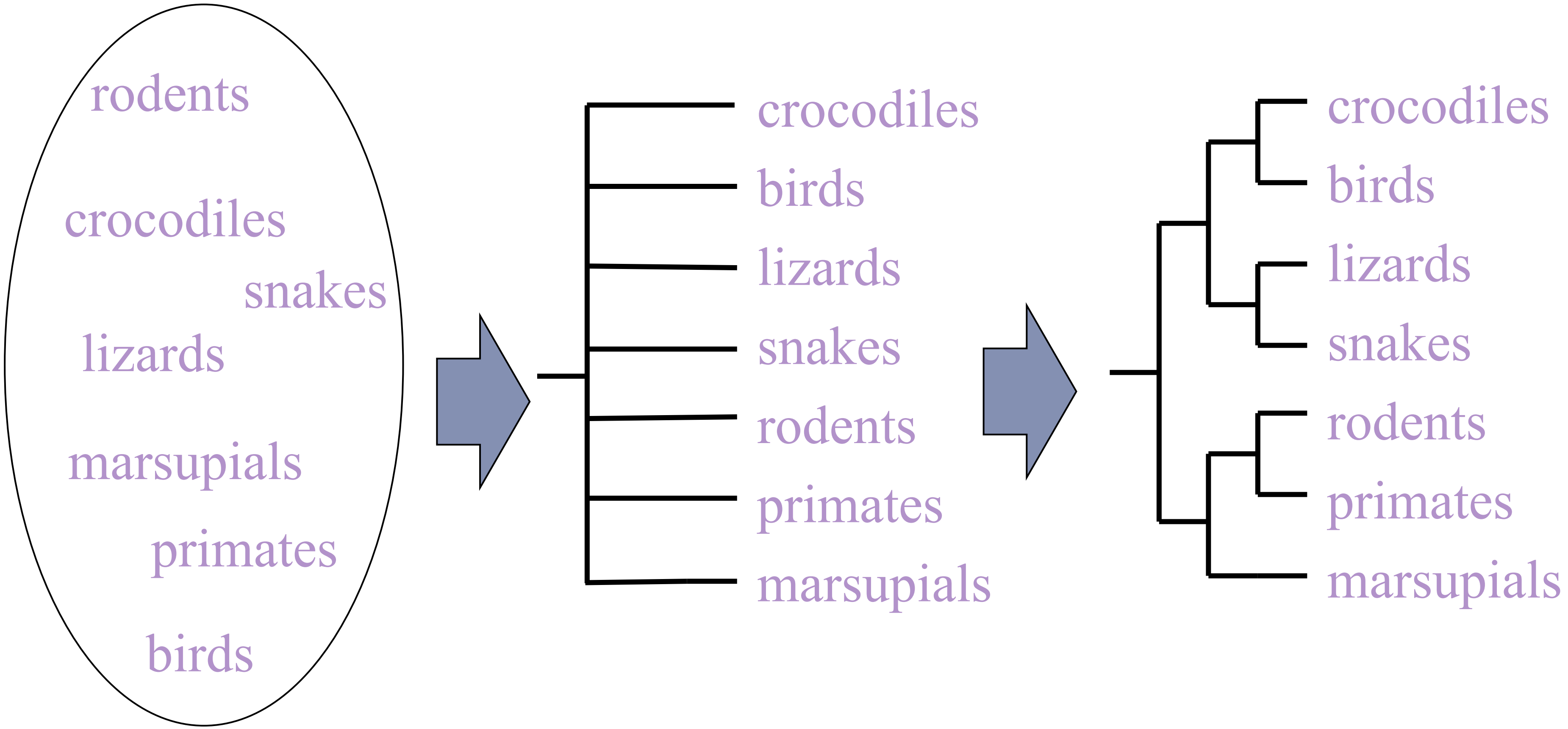
Figure 40.20a



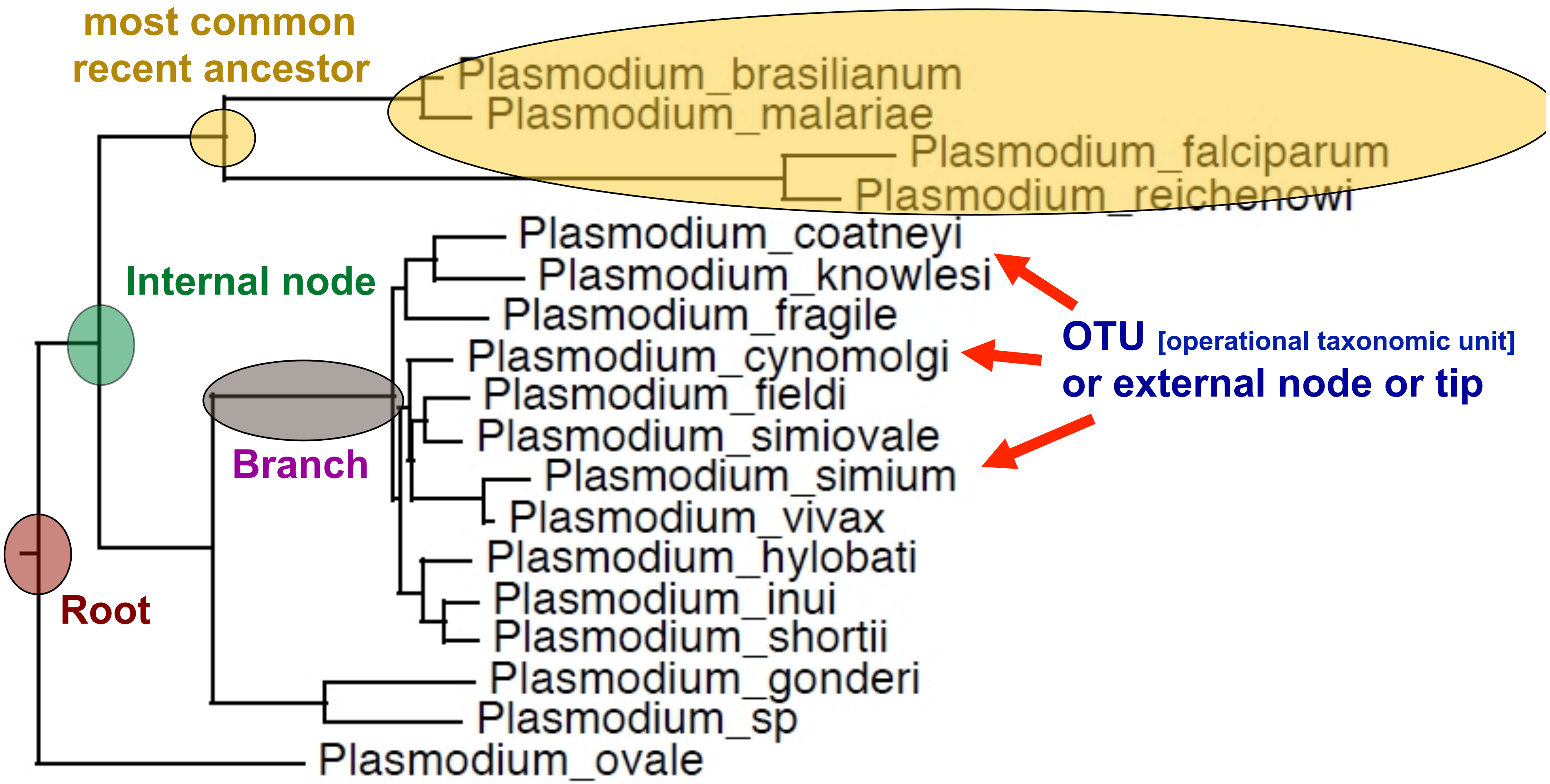
(a) Relationship of basal metabolic rate (BMR) to body size for various mammals

Definitions: what is tree?

A phylogenetic tree is the hierarchical classification of taxa that reflects their evolutionary relationships



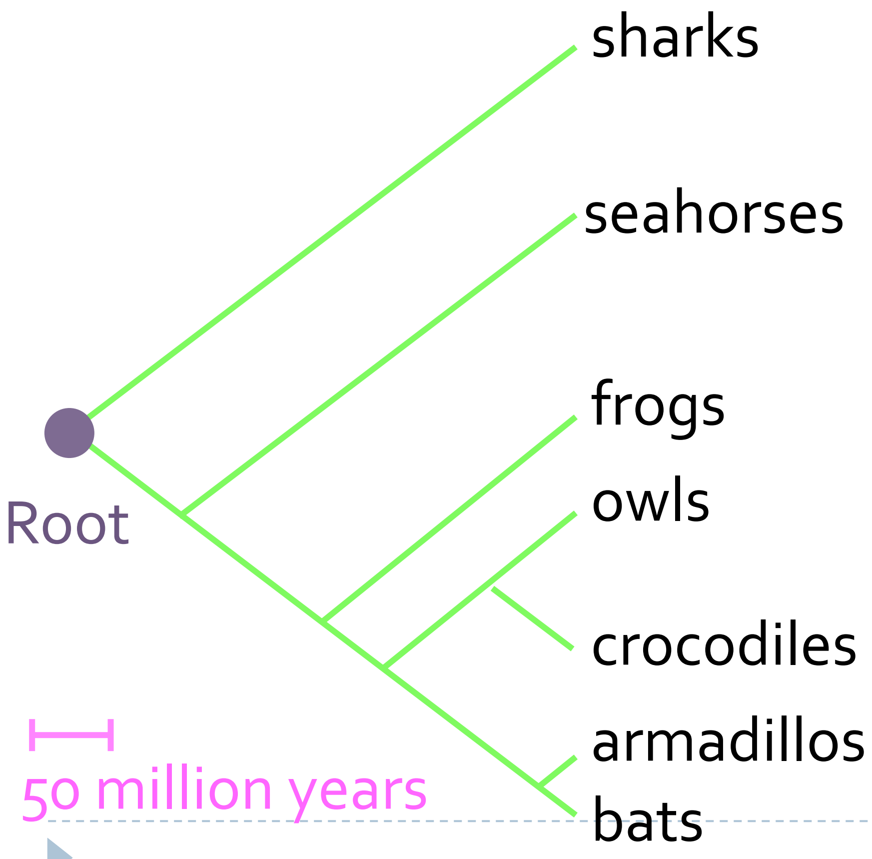
Terminology



Phylogram of primate-infecting malaria

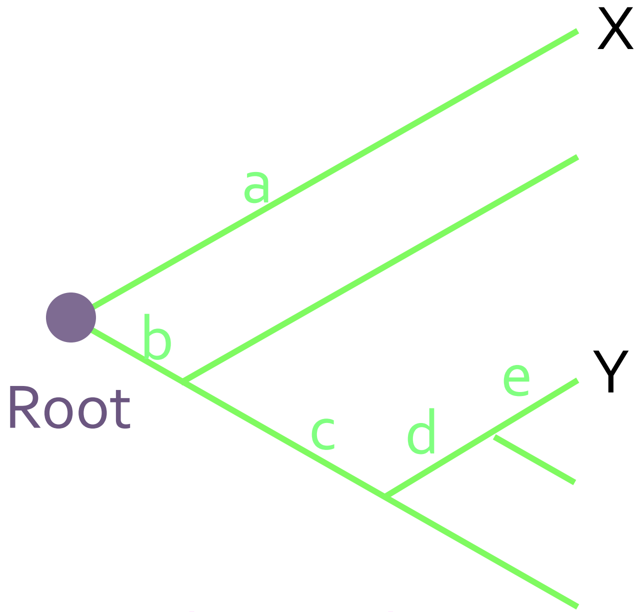
Time scaled phylogenies are ultrametric

Evolutionary trees measure ***time***.



Ultrametricity

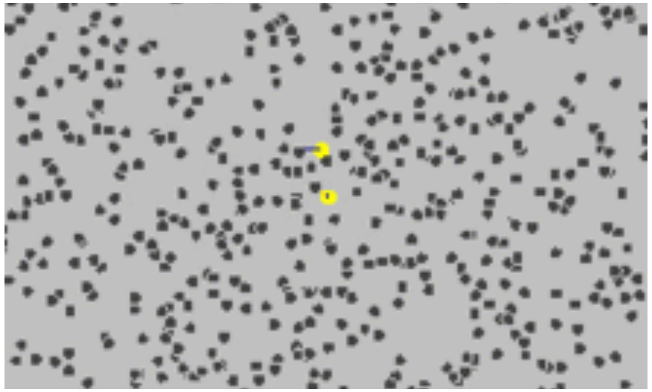
All tips are an equal distance from the root.



$a = b + c + d + e$

Evolution of continuous traits

Brownian motion



- Robert Brown (1773 – 1858)

$$dX_{(t)} = \sigma^2 * t$$

t = the step over which BM occurs

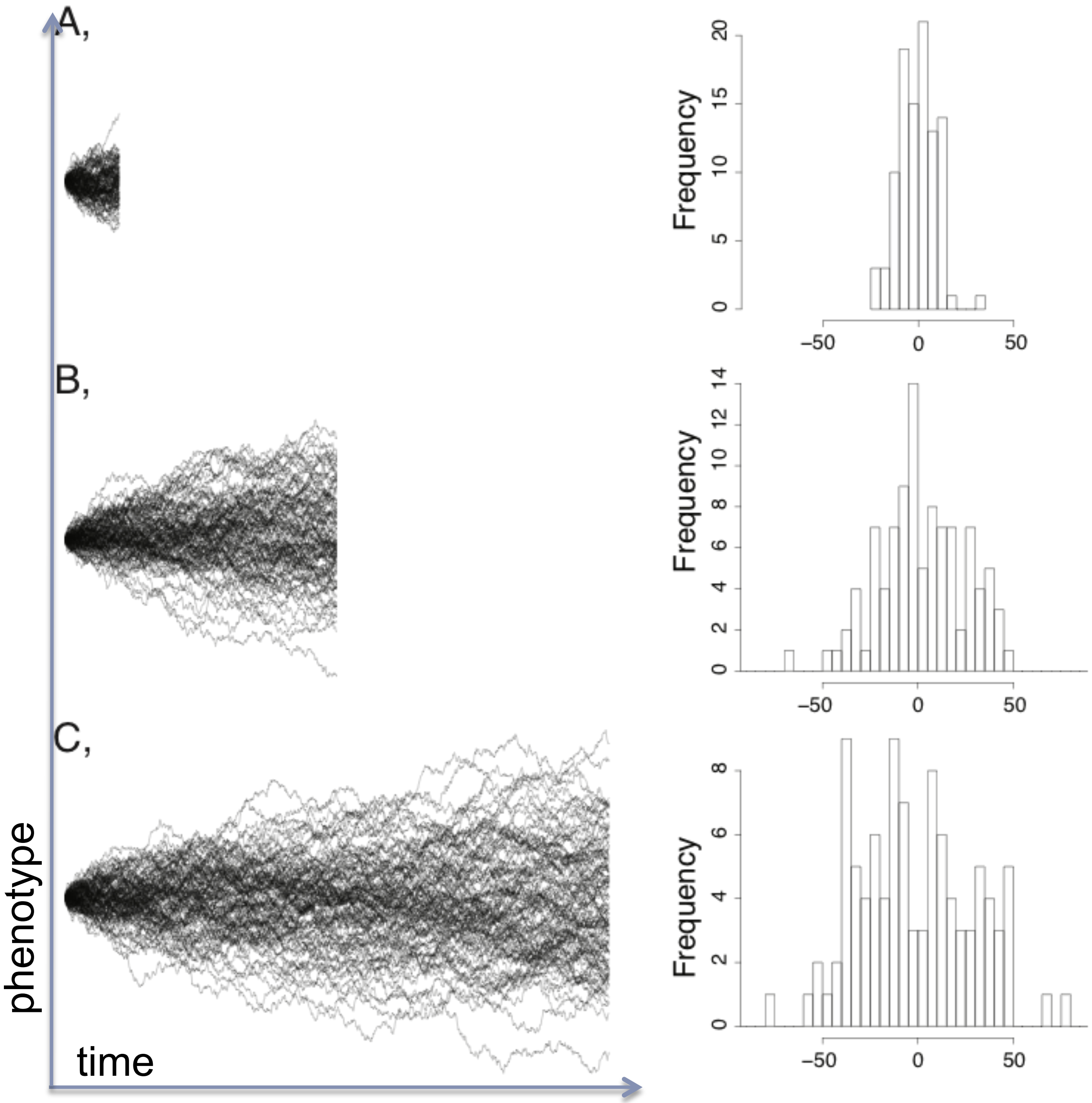
σ^2 = Brownian rate

stochastic, "random walk": changes of movements occur randomly and independently, in both direction and distance, at any time interval

$$E[\bar{z}(t)] = \bar{z}(0)$$

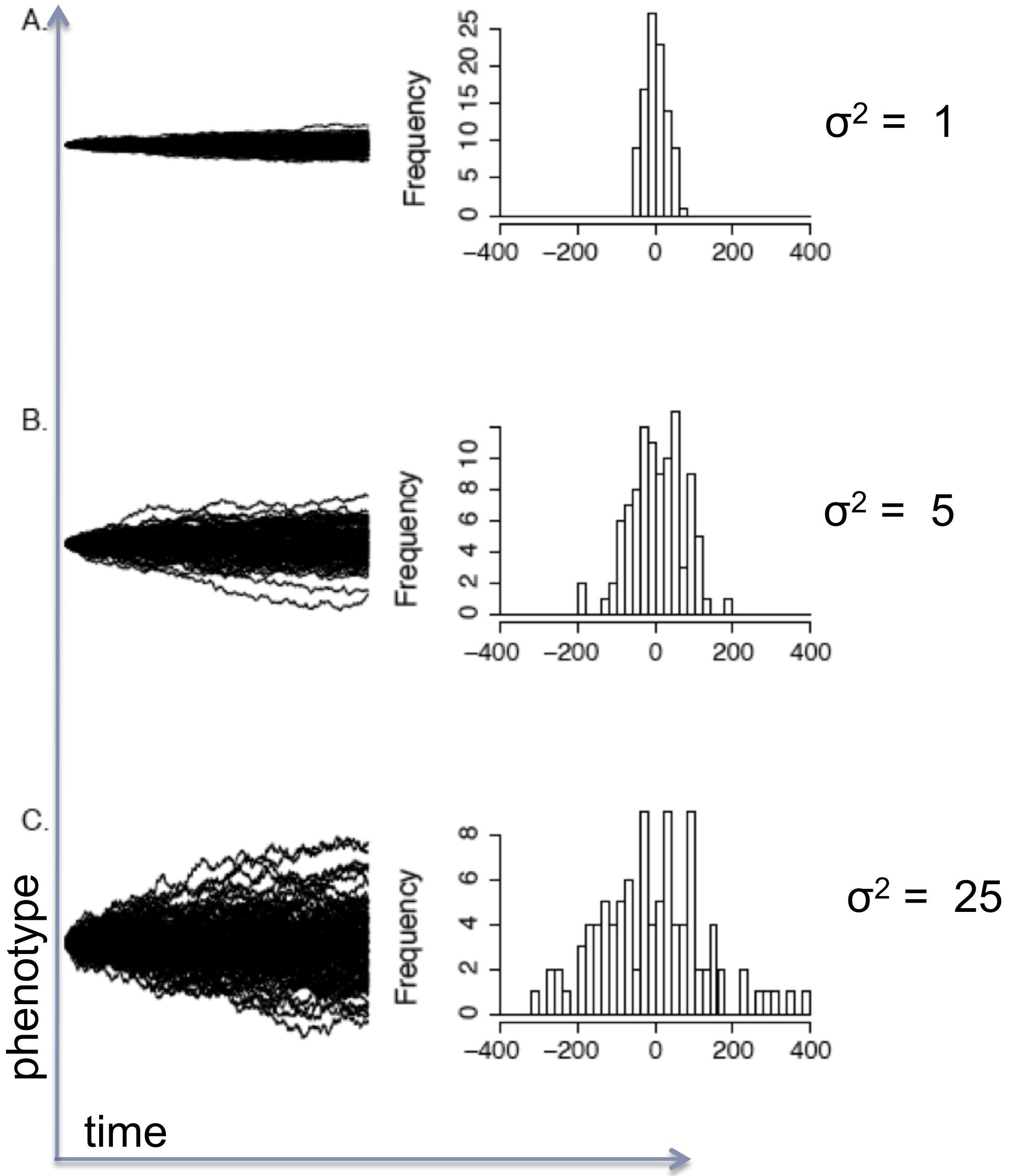
$$\bar{z}(t) \sim N(\bar{z}(0), \sigma^2 t)$$

Effect of time



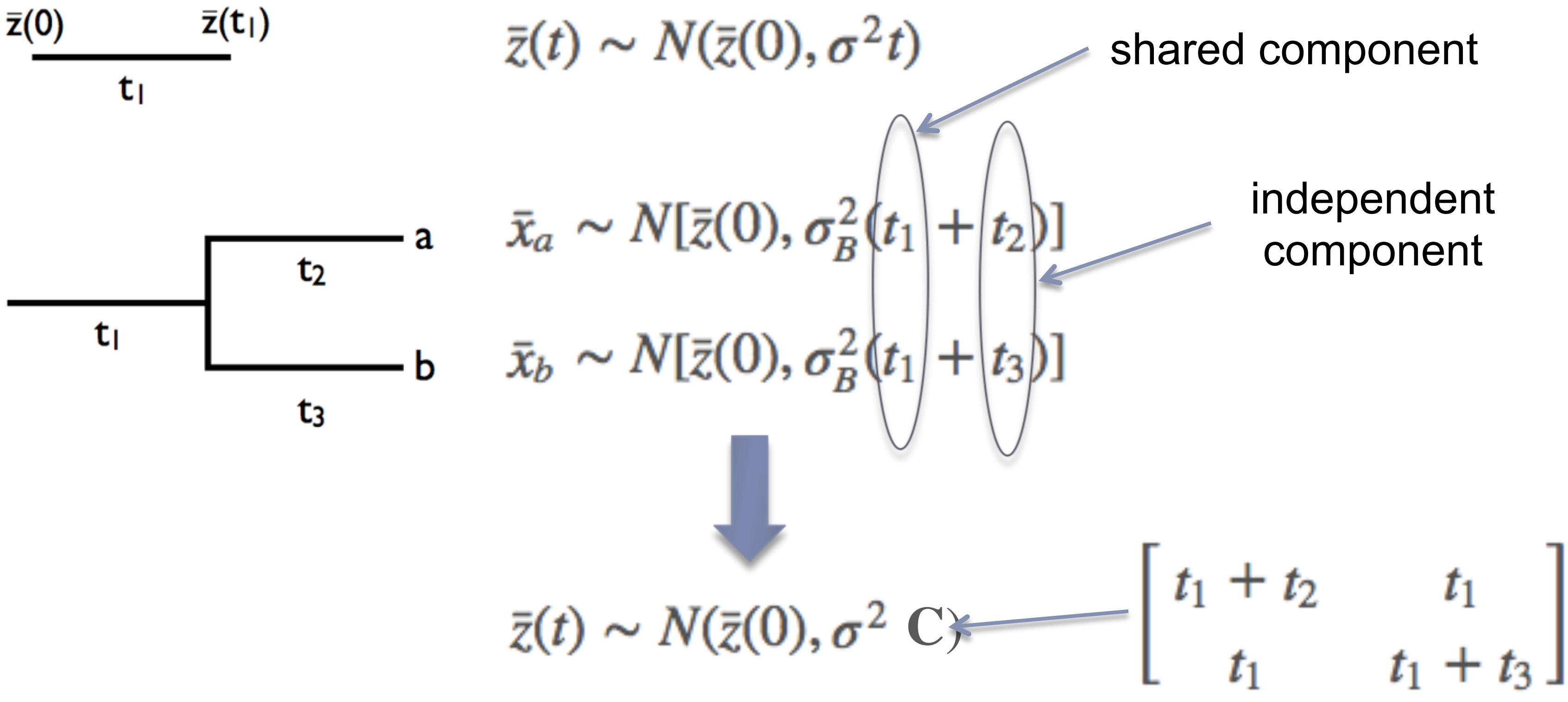
Harmon 2018: Phylogenetic Comparative Methods learning from trees

Effect of rate



Harmon 2018: Phylogenetic Comparative Methods learning from trees

Brownian motion on a phylogeny



$$\bar{z}(t) \sim N(\bar{z}(0), \sigma^2 t)$$

shared component

independent component

$$\bar{x}_a \sim N[\bar{z}(0), \sigma_B^2(t_1 + t_2)]$$

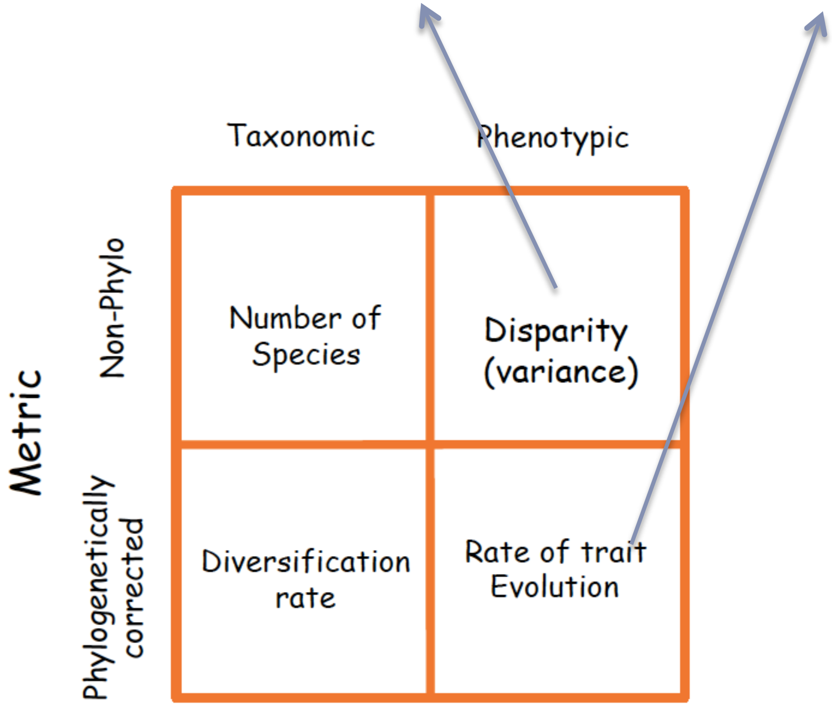
$$\bar{x}_b \sim N[\bar{z}(0), \sigma_B^2(t_1 + t_3)]$$

$$\bar{z}(t) \sim N(\bar{z}(0), \sigma^2 \mathbf{C}) \qquad \begin{bmatrix} t_1 + t_2 & t_1 \\ t_1 & t_1 + t_3 \end{bmatrix}$$

Harmon 2018: Phylogenetic Comparative Methods learning from trees

Brownian motion on a phylogeny

$$E(\text{disparity}) = \sigma^2\left[\frac{1}{N}\operatorname{tr}(\mathbf{C}) - \frac{1}{N^2}\mathbf{1}'\mathbf{C}\mathbf{1}\right]$$



Metric		Taxonomic	Phenotypic
	Non-Phylo	Number of Species	Disparity (variance)
	Phylogenetically corrected	Diversification rate	Rate of trait Evolution

O'Meara et al. 2006

Brownian motion on a phylogeny

rate of evolution

$$\mathrm{E}(\text{disparity}) = \sigma^2\left[\frac{1}{N}\mathrm{tr}(\mathbf{C}) - \frac{1}{N^2}\mathbf{1}'\mathbf{C}\mathbf{1}\right]$$

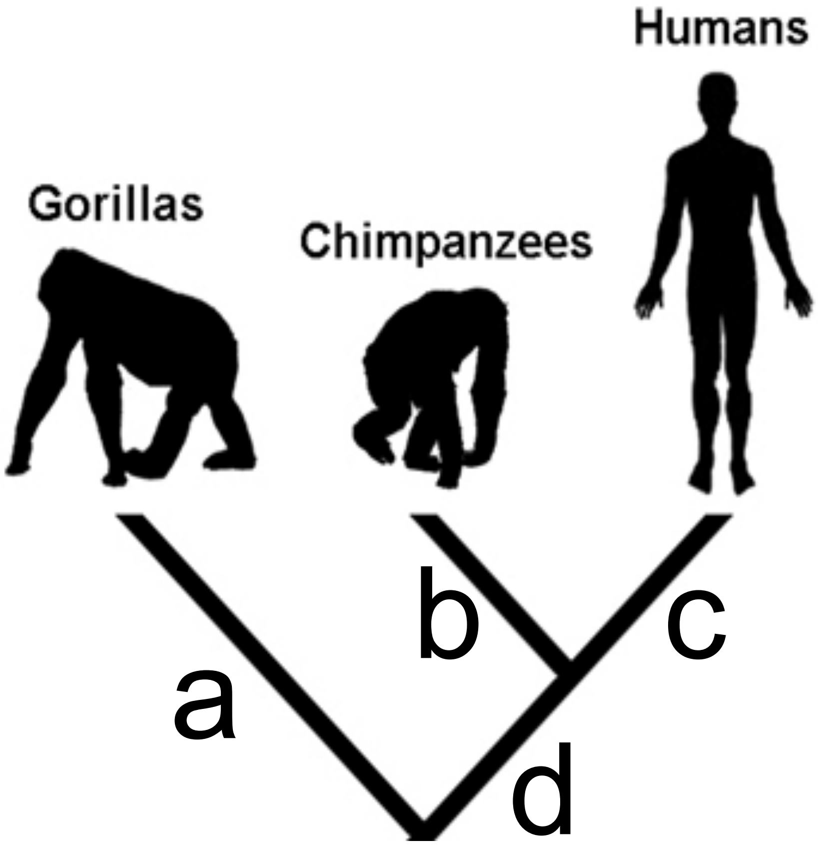
Expected trait
variance under BM

mean total
evolutionary history
(average distance
from tips to the root)

mean of **non-independent**
evolutionary history
(average amount of
shared-distance)

O'Meara et al. 2006

C: Variance-covariance matrix

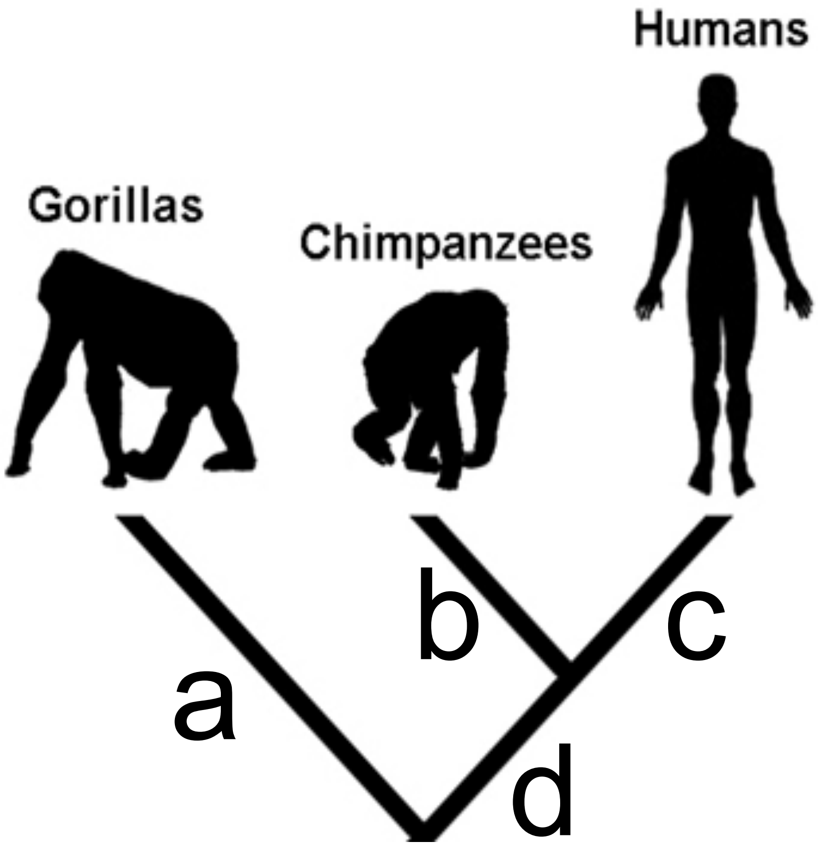


c = b

c+d = a

b+d = a

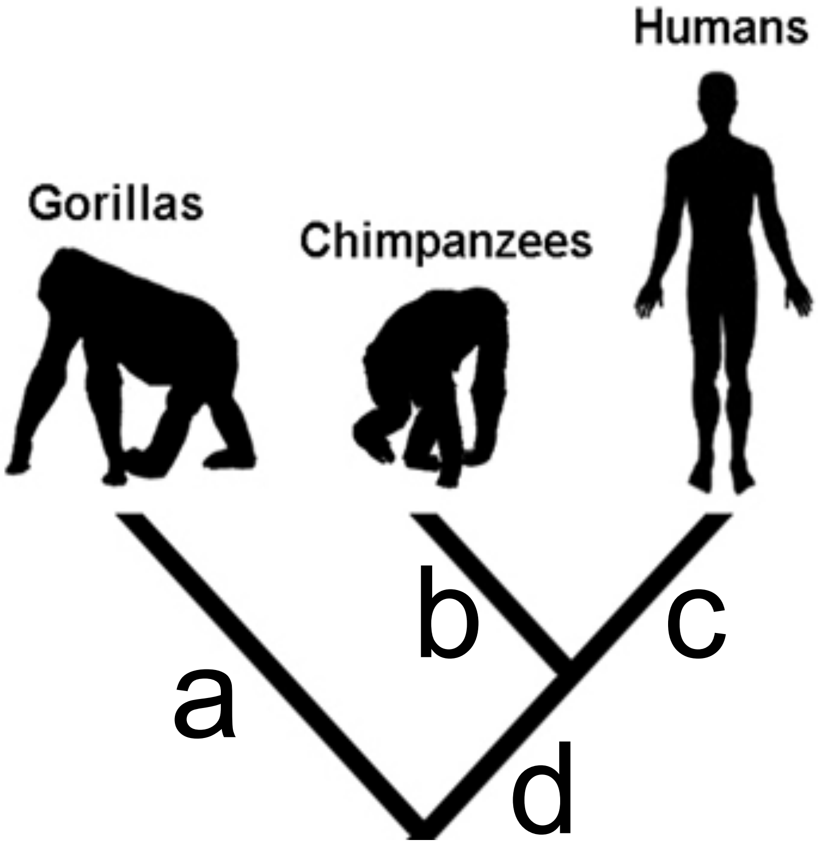
C: Variance-covariance matrix



Diagonal = tree length

	H	**C**	**G**
H	🟨		
C		🟨	
G			🟨

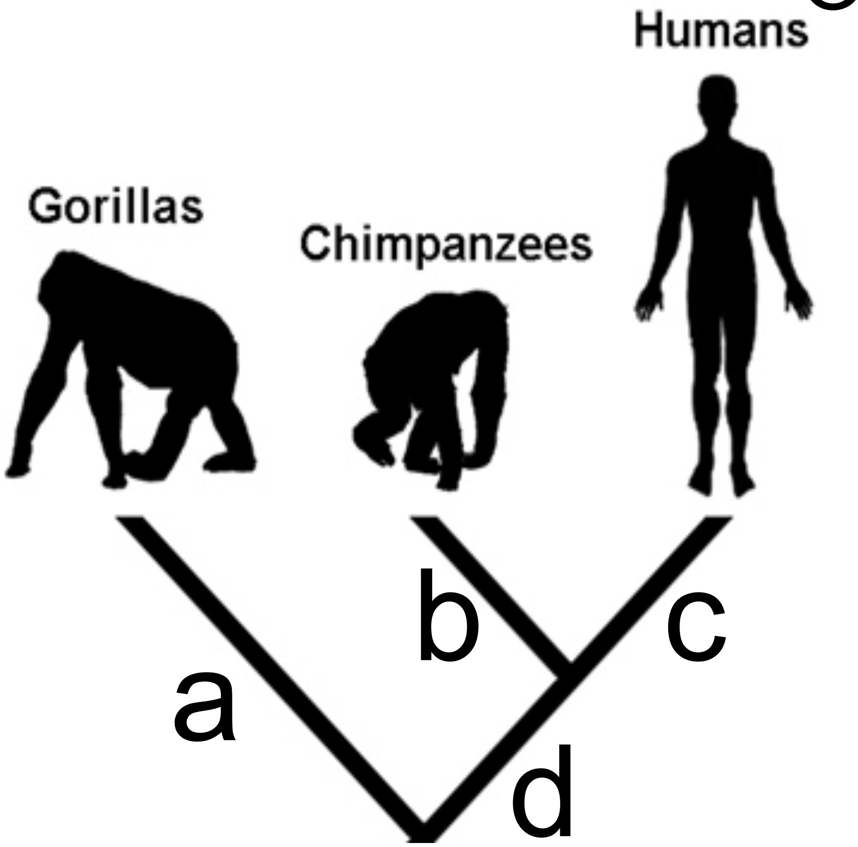
C: Variance-covariance matrix



Diagonal = tree length

	H	**C**	**G**
H	d+c		
C		d+b	
G			a

C: Variance-covariance matrix

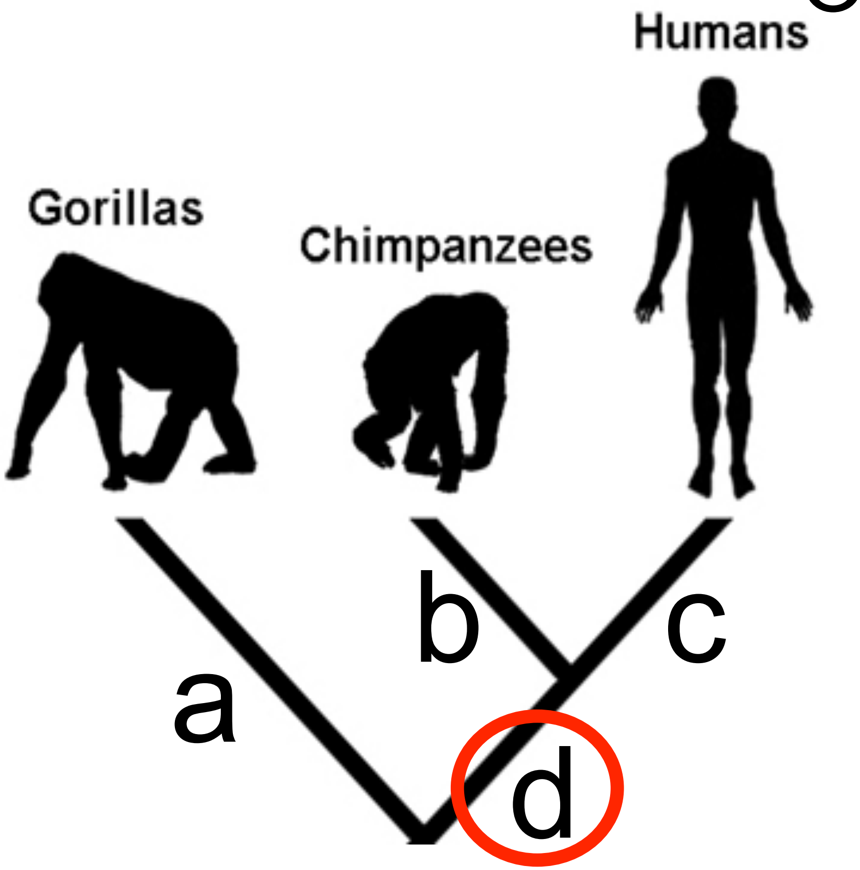


Off-diagonal = shared branch length for each pair

	H	**C**	**G**
H	d+c		
C		d+b	
G			a

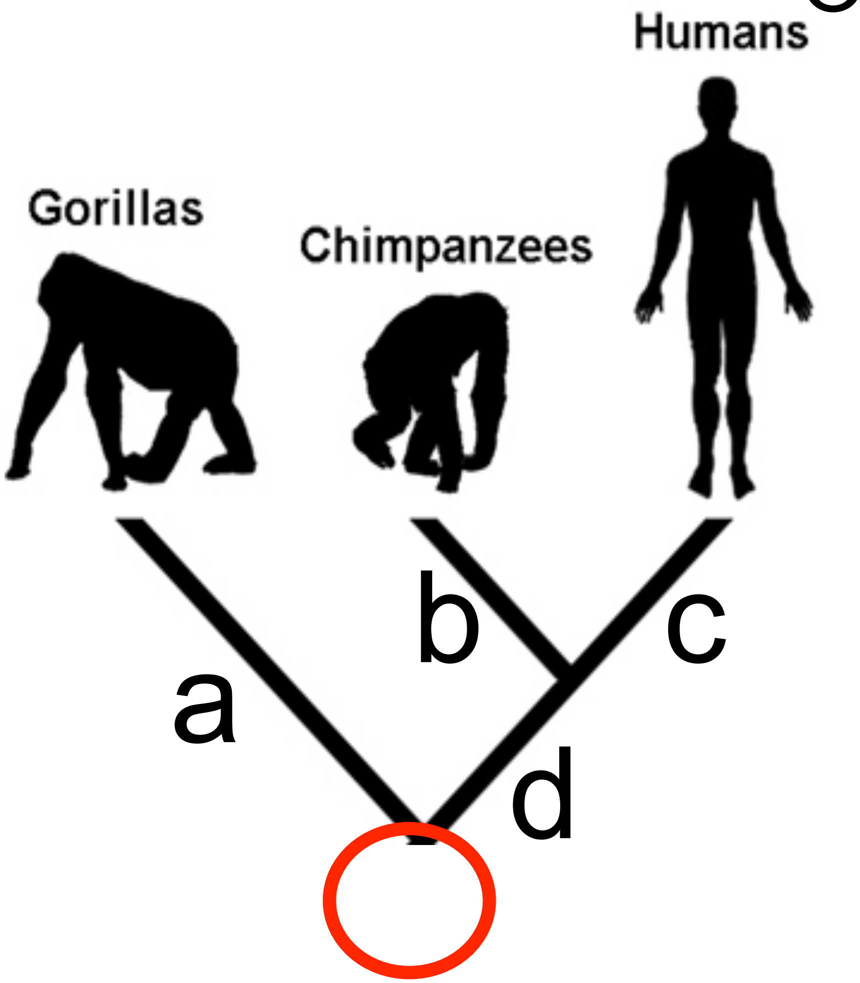
C: Variance-covariance matrix

Off-diagonal = shared branch length for each pair



	H	C	G
H	d+c	d	
C	d	d+b	
G			a

C: Variance-covariance matrix

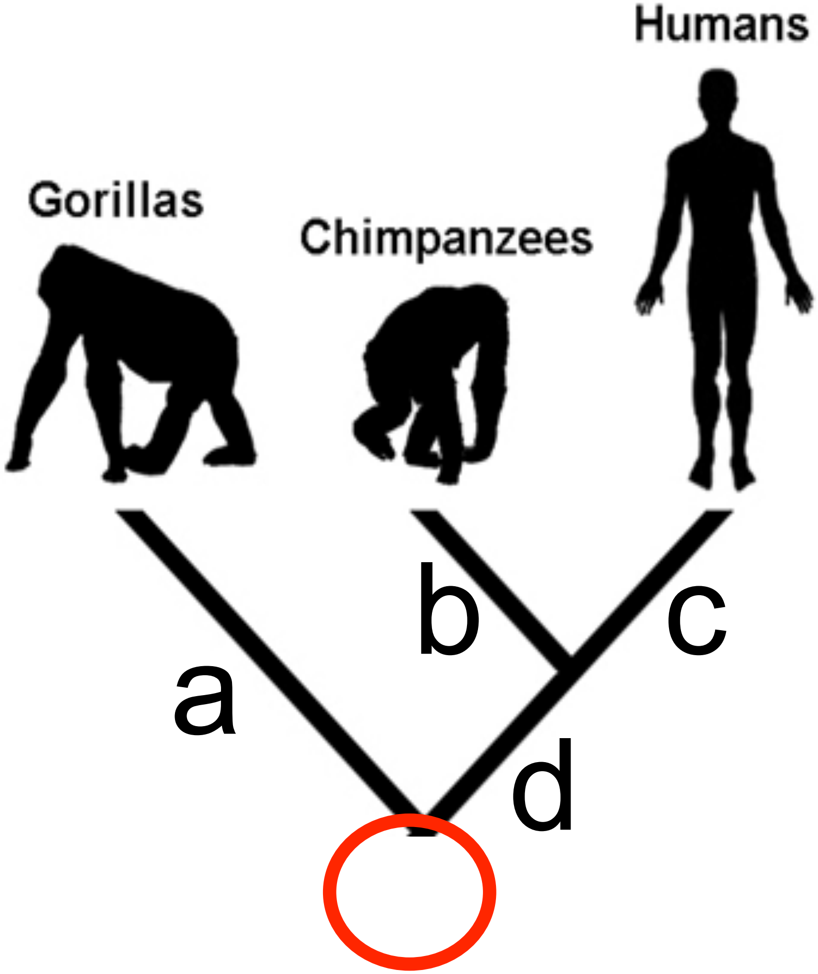


Off-diagonal = shared branch length for each pair

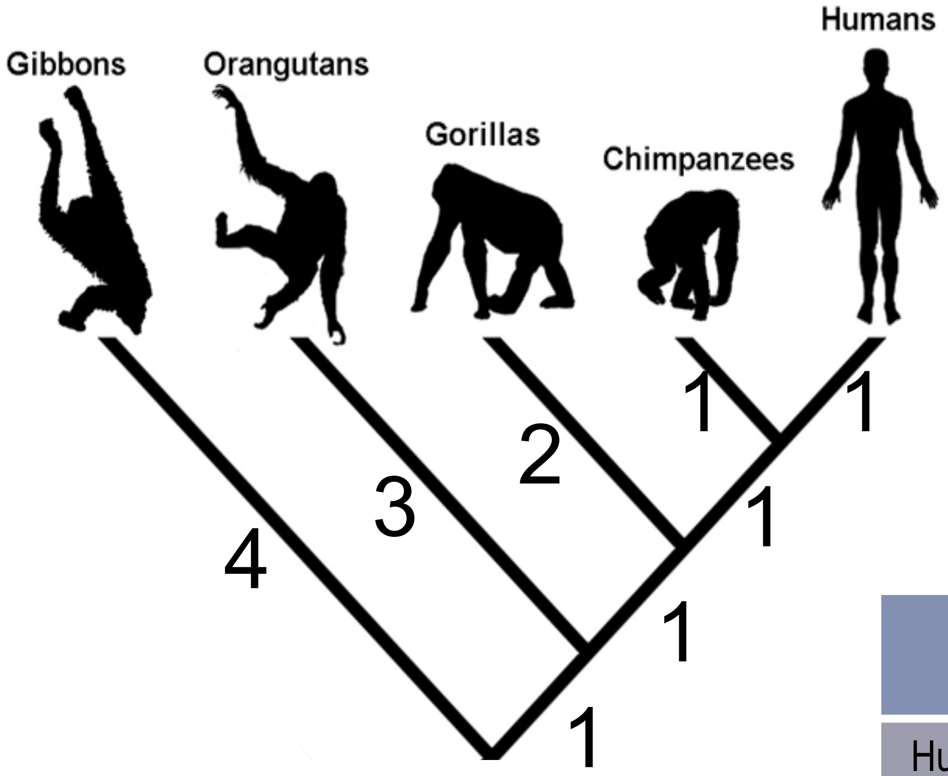
	H	C	G
H	d+c	d	0
C	d	d+b	
G	0		a

C: Variance-covariance matrix

Off-diagonal = shared branch length for each pair



	H	C	G
H	d+c	d	0
C	d	d+b	0
G	0	0	a



	Human	Pan	Gorilla	Pongo	Gibbon
Human	**4**	3	2	1	0
Pan	3	**4**	2	1	0
Gorilla	2	2	**4**	1	0
Pongo	1	1	1	**4**	0
Gibbon	0	0	0	0	**4**

$$\mathrm{E(disparity)} = \sigma^2\left[\boxed{\frac{1}{N}\mathrm{tr}(\mathbf{C})} - \frac{1}{N^2}\mathbf{1}'\mathbf{C}\mathbf{1}\right]$$

[4 -]

average distance from tips to the root (tree length)

	Human	Pan	Gorilla	Pongo	Gibbon
Human	**4**	3	2	1	0
Pan	3	**4**	2	1	0
Gorilla	2	2	**4**	1	0
Pongo	1	1	1	**4**	0
Gibbon	0	0	0	0	**4**

$$E(\text{disparity}) = \sigma^2\left[\frac{1}{N}\mathrm{tr}(\mathbf{C}) - \frac{1}{N^2}\mathbf{1}'\mathbf{C}\mathbf{1}\right]$$

$$[\;4 \quad - \quad 1.6\,]$$

average amount of shared-distance (average entry of **C**)

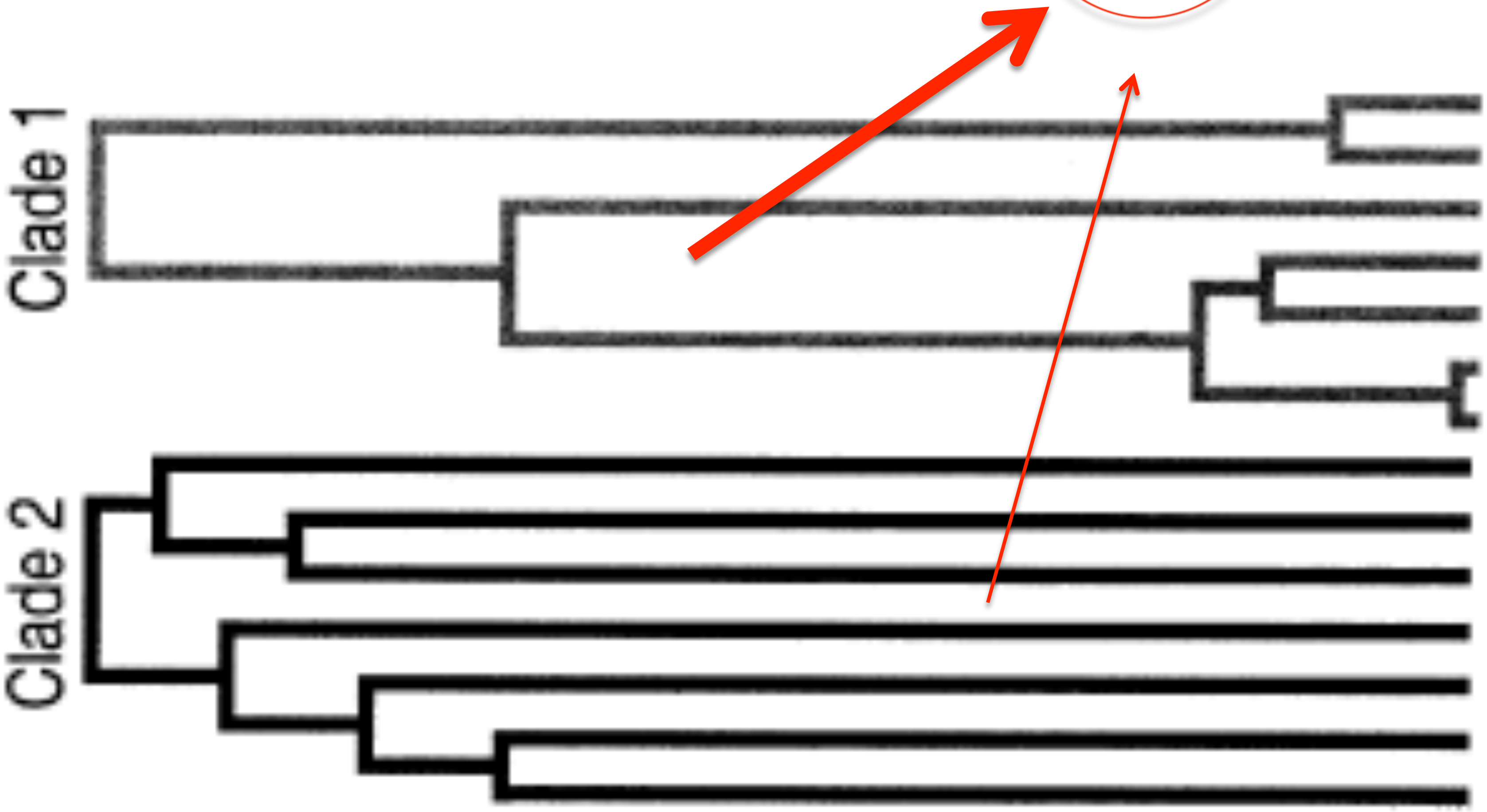
	Human	Pan	Gorilla	Pongo	Gibbon
Human	**4**	3	2	1	0
Pan	3	**4**	2	1	0
Gorilla	2	2	**4**	1	0
Pongo	1	1	1	**4**	0
Gibbon	0	0	0	0	**4**

$$E(\text{disparity}) = \sigma^2\left[\frac{1}{N}\operatorname{tr}(\mathbf{C}) - \frac{1}{N^2}\mathbf{1}'\mathbf{C}\mathbf{1}\right]$$

$$2.4 = [\ 4 \quad - \quad 1.6]$$

	Human	Pan	Gorilla	Pongo	Gibbon
Human	**4**	3	2	1	0
Pan	3	**4**	2	1	0
Gorilla	2	2	**4**	1	0
Pongo	1	1	1	**4**	0
Gibbon	0	0	0	0	**4**

$$E(\text{disparity}) = \sigma^2\left[\frac{1}{N}\mathrm{tr}(\mathbf{C}) - \boxed{\frac{1}{N^2}\mathbf{1}'\mathbf{C}\mathbf{1}}\right]$$



O’Meara et al. 2006

Alternatives to Brownian motion

- Variable rates over the tree
- Declining rates through time (Early Burst, EB/AC)
- Accelerating rates through time (Late Burst, LB/DC)
- A single stable adaptive peak (Ornstein-Uhlenbeck, OU)
- Variable adaptive peaks (Ornstein-Uhlenbeck, OU)
- Trends in the mean trait value (BM with a trend)
- Mixtures of the above, and more

Alternatives to Brownian motion

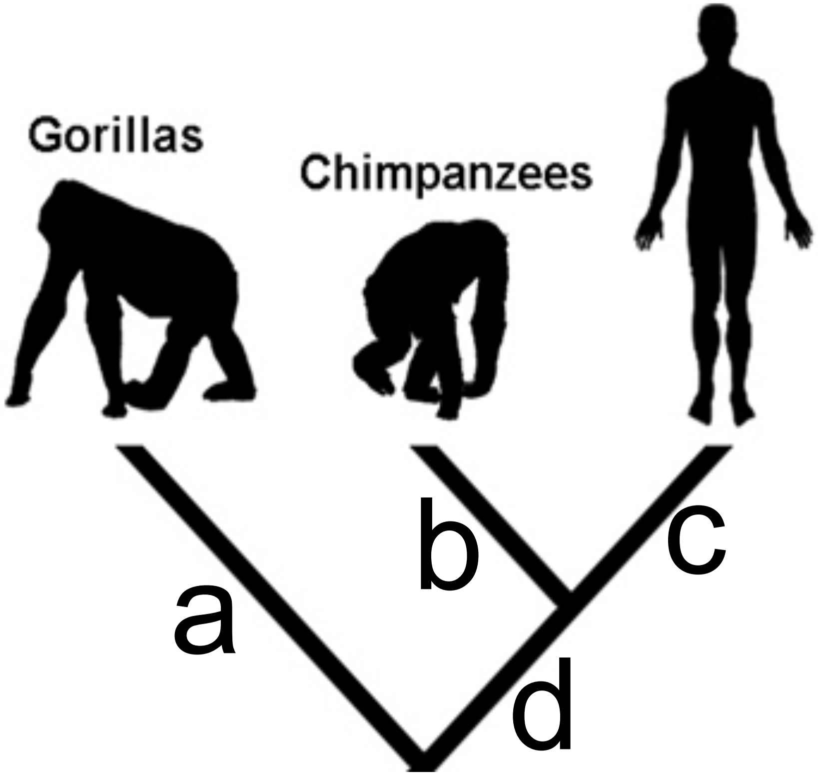
- Variable rates over the tree
- Declining rates through time (Early Burst, EB/AC)
- Accelerating rates through time (Late Burst, LB/DC)
- A single stable adaptive peak (Ornstein-Uhlenbeck, OU)
- Variable adaptive peaks (Ornstein-Uhlenbeck, OU)
- Trends in the mean trait value (BM with a trend)
- Mixtures of the above, and more

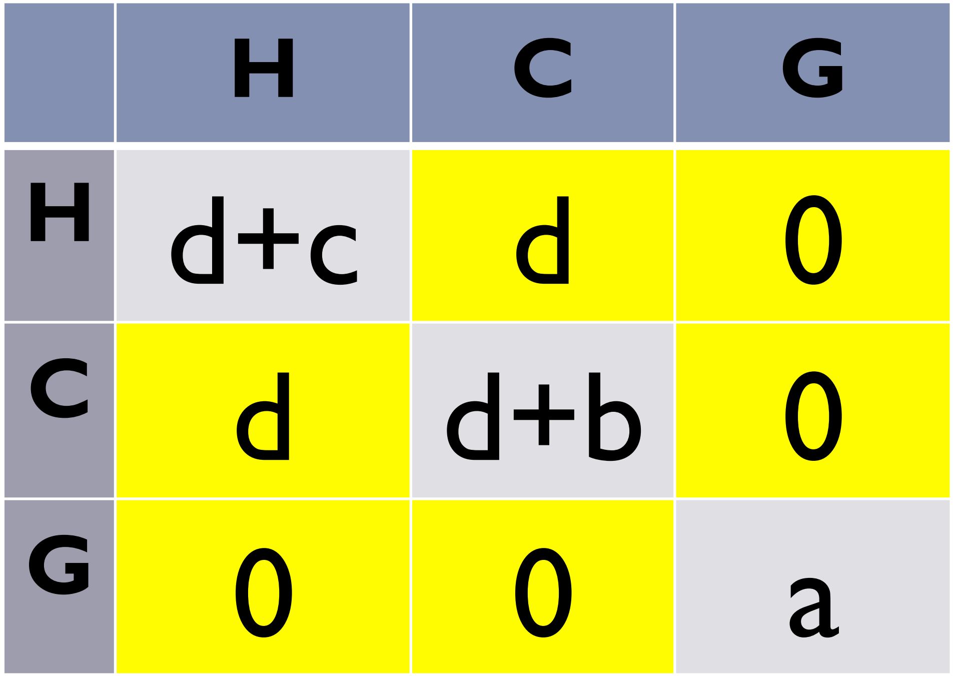
Tree transformations: altering rate of evolution

- Longer branches, higher rate
- Pagel's transformations
- Alteration of the **C** matrix

Harmon 2018: Phylogenetic Comparative Methods learning from trees

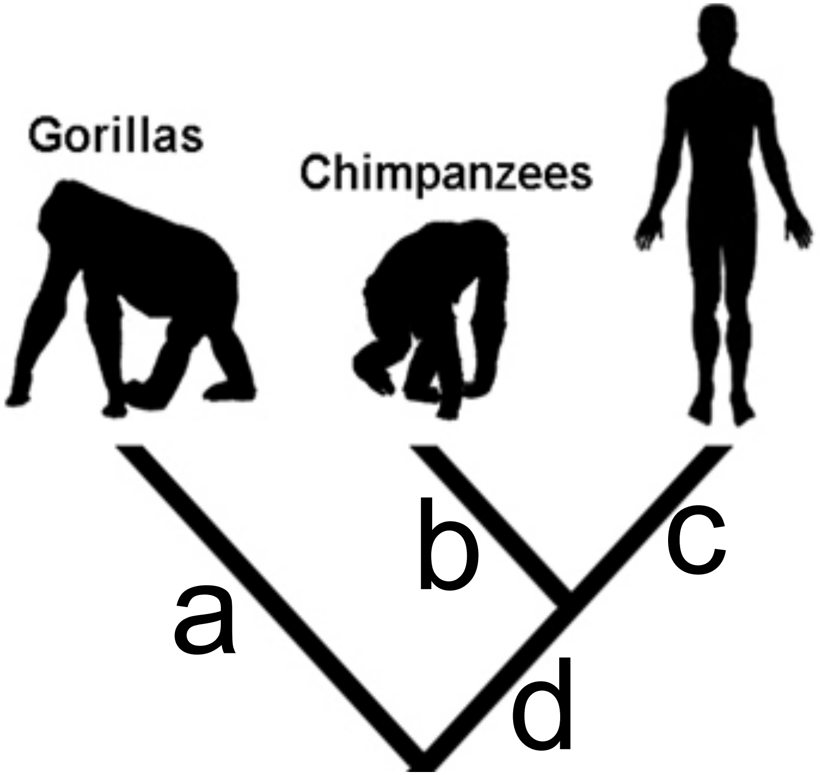
Lambda: multiplying off-diagonals





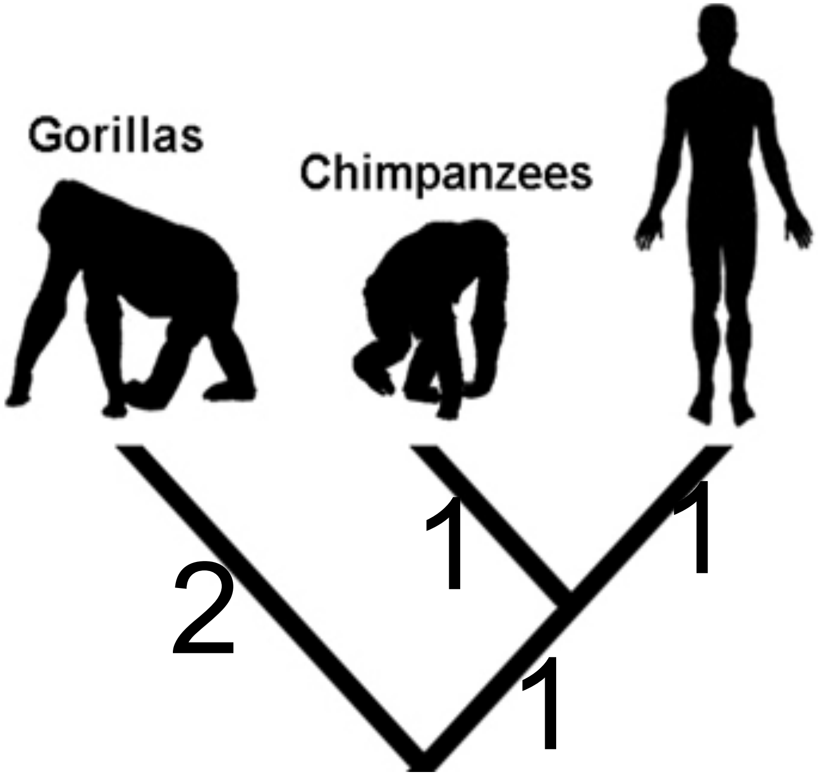
	H	C	G
H	d+c	d	0
C	d	d+b	0
G	0	0	a

Lambda: multiplying off-diagonals



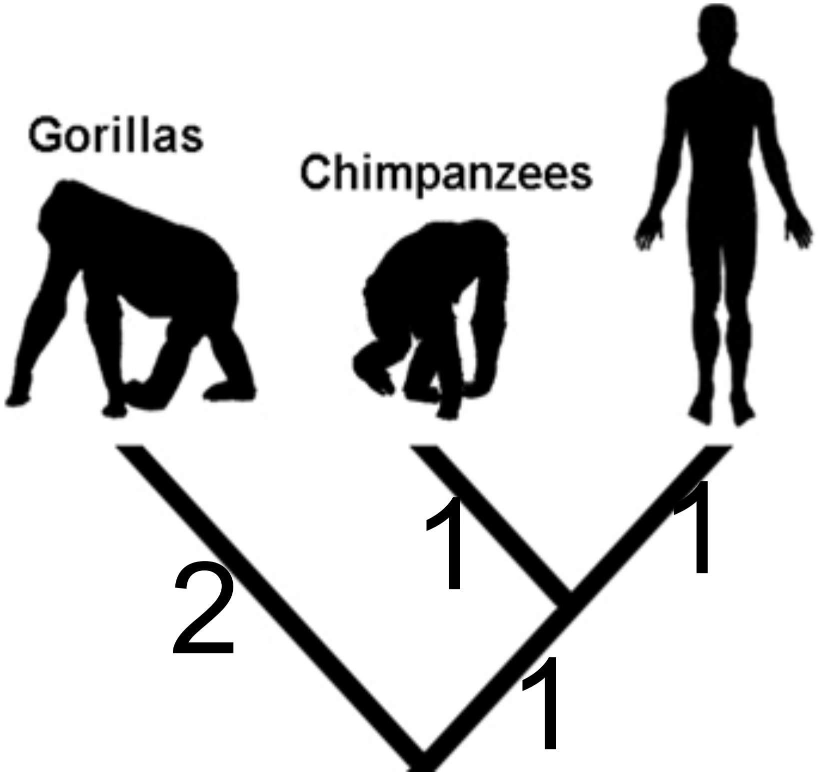
	H	C	G
H	d+c	λd	λ0
C	λd	d+b	λ0
G	λ0	λ0	a

Lambda: multiplying off-diagonals



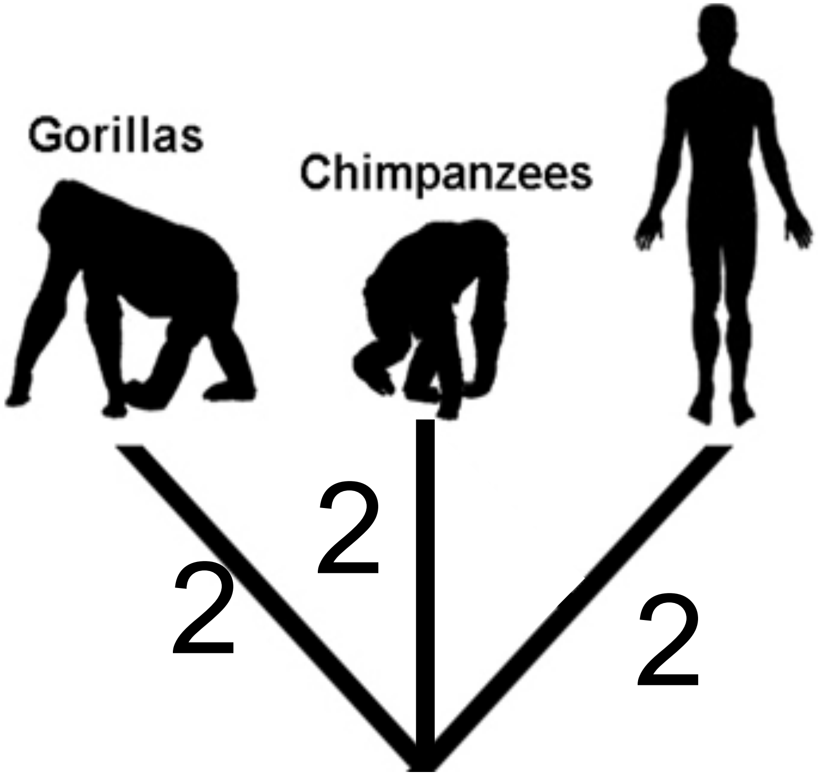
	H	C	G
H	2	$\lambda 1$	$\lambda 0$
C	$\lambda 1$	2	$\lambda 0$
G	$\lambda 0$	$\lambda 0$	2

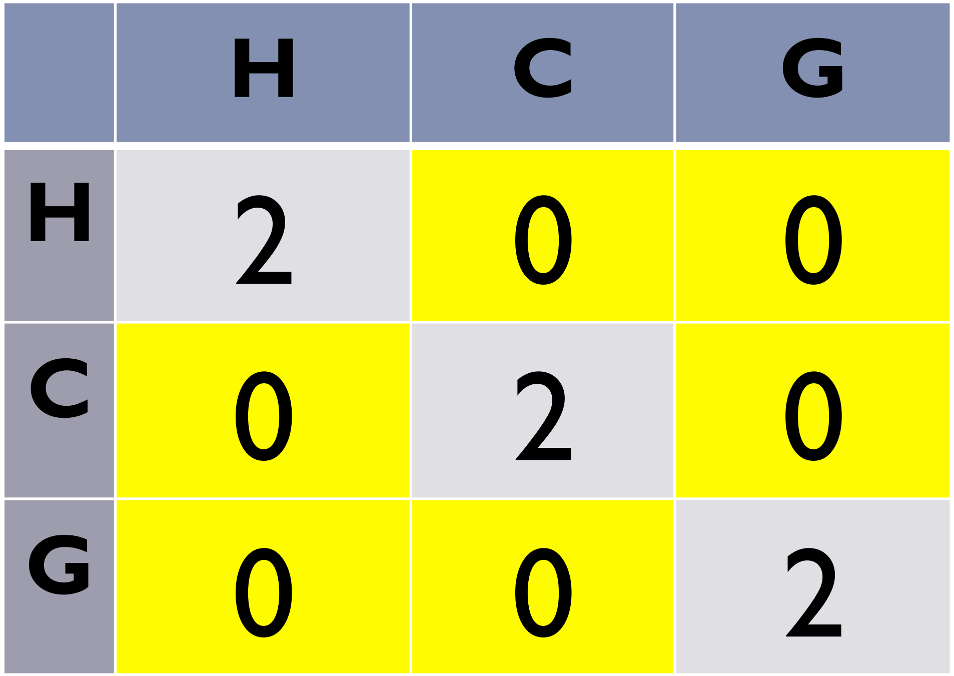
Lambda: multiplying off-diagonals



	H	C	G
H	2	λ	0
C	λ	2	0
G	0	0	2

Lambda: multiplying off-diagonals

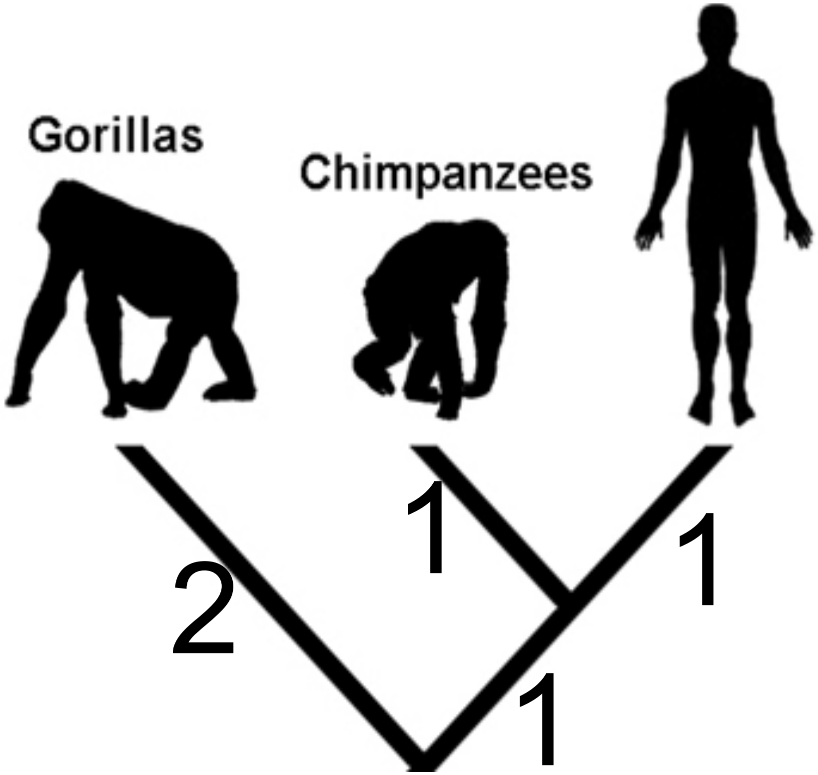




	H	C	G
H	2	0	0
C	0	2	0
G	0	0	2

$\lambda = 0$

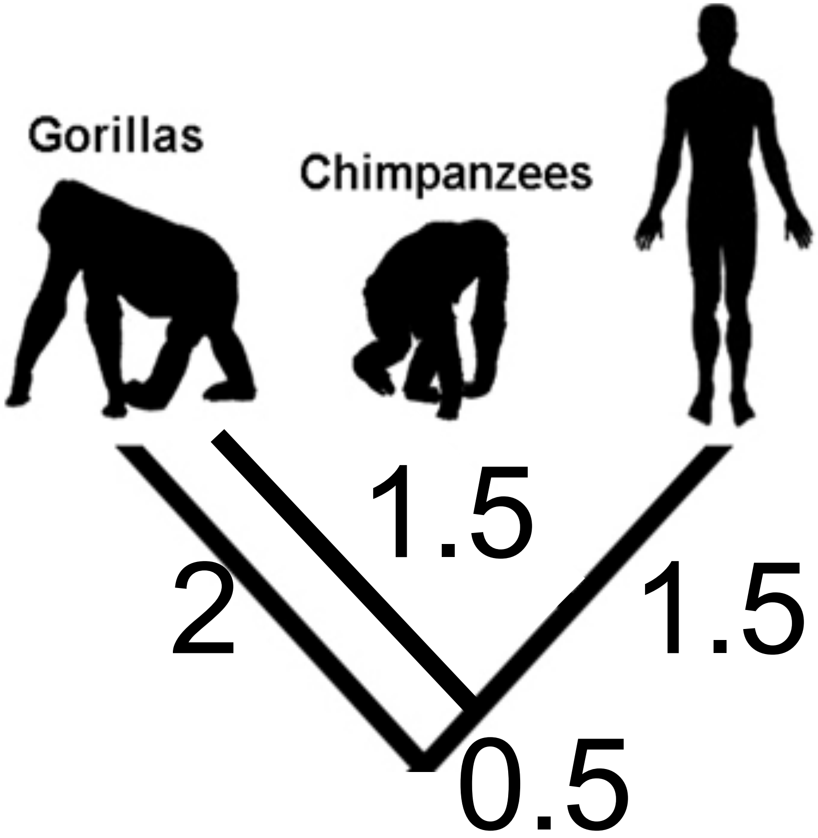
Lambda: multiplying off-diagonals



	H	C	G
H	2	1	0
C	1	2	0
G	0	0	2

$\lambda = 1$

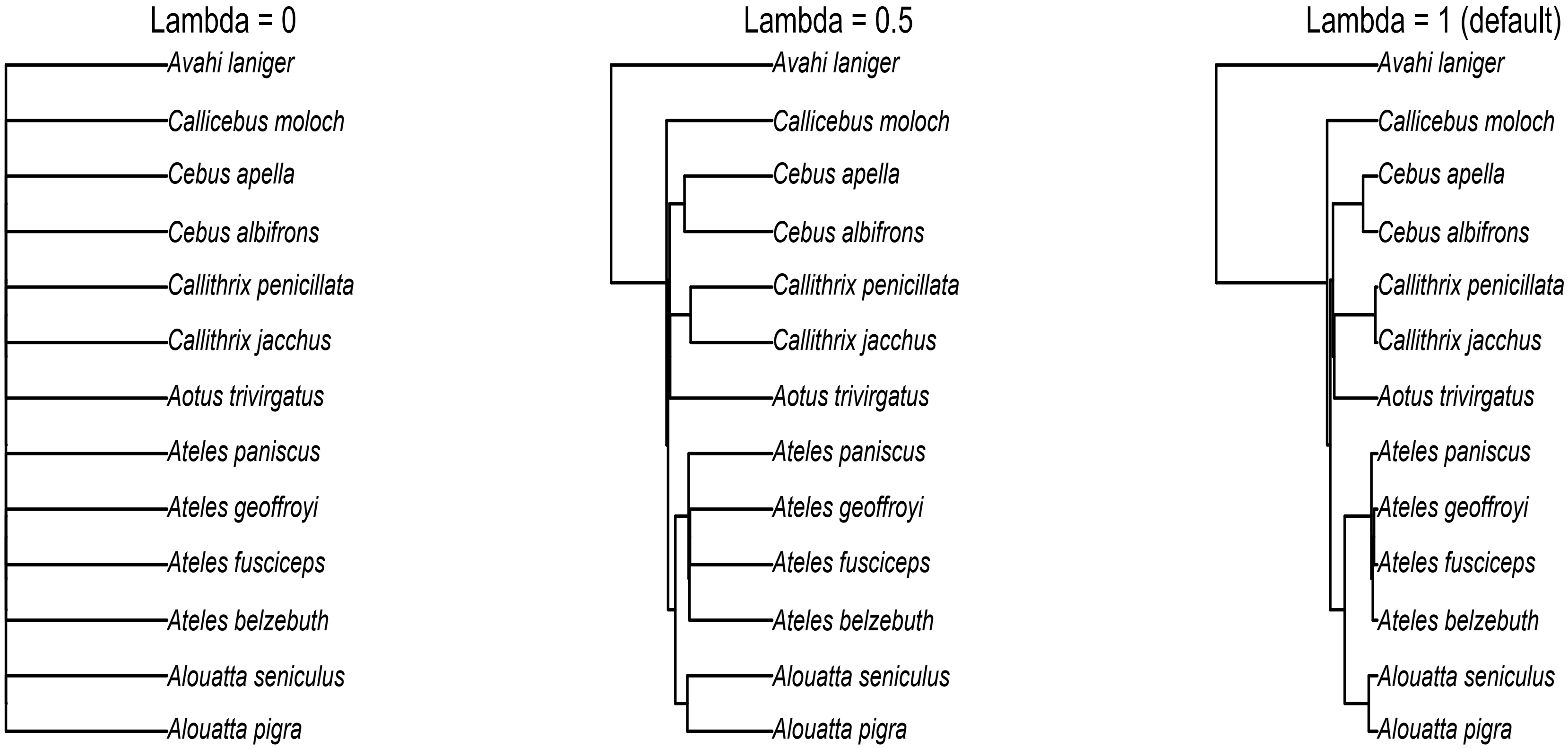
Lambda: multiplying off-diagonals



	H	C	G
H	2	0.5	0
C	0.5	2	0
G	0	0	2

$\lambda = 0.5$

Lambda: multiplying off-diagonals



Lambda: multiplying off-diagonals

- Shortens the internal branches relative to the tips
- λ = 0: no relationship between trait and phylogeny = star phylogeny
- λ = 1: trait values are as expected under Brownian motion = phylogeny is unchanged
- Measure of PHYLOGENETIC SIGNAL

Phylogenetic signal

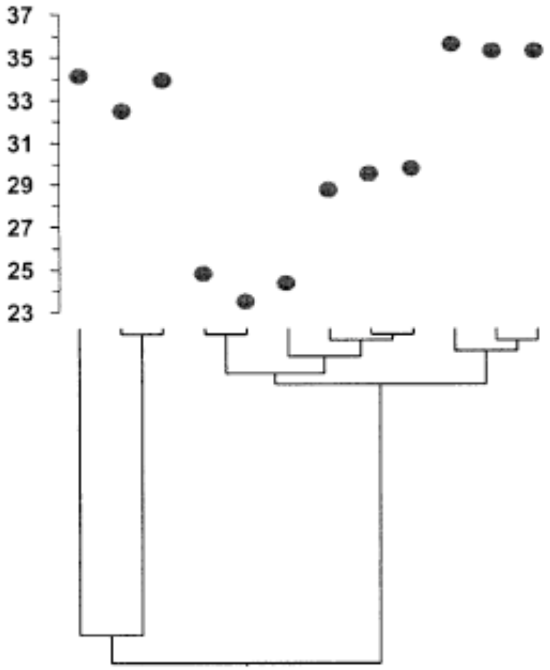
statistical non-independence among species trait values due to their phylogenetic relatedness OR

the tendency for related species to resemble each other more than expected by chance

Note this is a pattern not a process

Phylogenetic signal

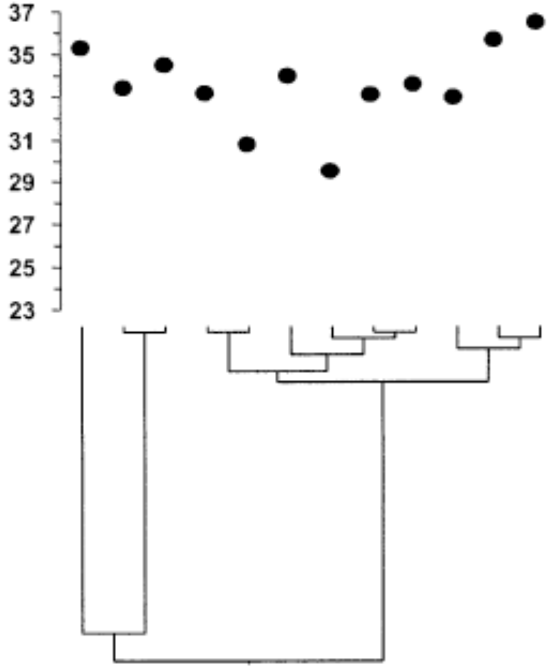
- Theory vs. real world > phylogenetic signal is an empirical issue
  - Convergent evolution (distantly related species are similar)
  - Character displacement (closely related species are dissimilar)
- Phylogenetic signal in the data can be lower than expected



Preferred Body Temperatures (°C) of 12 species of Australian skinks

Randomization Test for Phylogenetic Signal, $P < 0.001$

$K = 0.453$

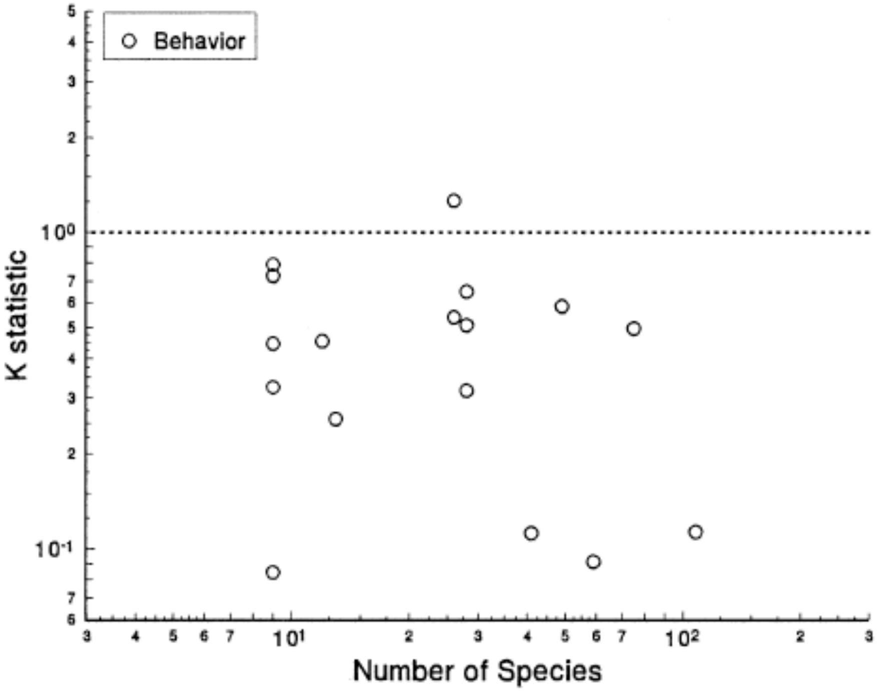


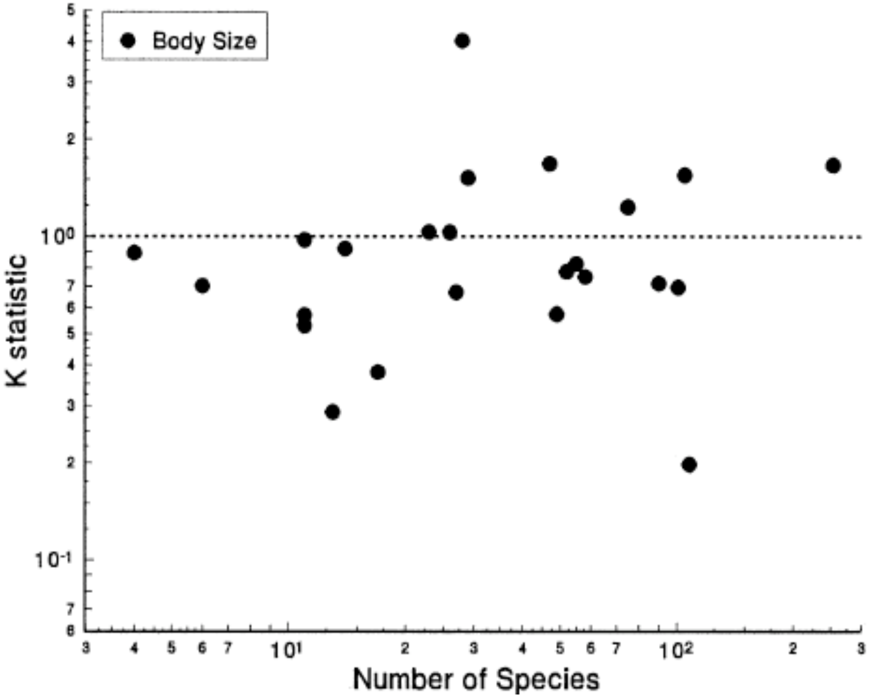
Optimal Body Temperatures for Sprinting

Randomization Test for Phylogenetic Signal, $P = 0.167$

$K = 0.101$

Quantifying phylogenetic signal





e.g. Freckleton et al. 2002; Bllomberg et al. 2003

Phylogenetic signal

Many measures have been suggested. The two most popular are:

1) Pagel's λ
2) Blomberg's *K* (K not kappa)

Phylogenetic signal: Blomberg' s *K*

$$K = \text{observed} \frac{\mathrm{MSE}_0}{\mathrm{MSE}} \Big/ \text{expected} \frac{\mathrm{MSE}_0}{\mathrm{MSE}}$$

(Very simply) MSE = mean squared error

= variance in trait

Phylogenetic signal: Blomberg’s *K*

Variance of the tip data relative to phylogenetic mean

$$\frac{\mathrm{MSE}_0}{\mathrm{MSE}} = \mathrm{LARGE}$$

Variance of the tip data relative to phylogeny → LOW – if tree explains variation in the data well

Phylogenetic signal: Blomberg’s *K*

Variance of the tip data relative to phylogenetic mean

$$\frac{\mathrm{MSE}_0}{\mathrm{MSE}} = \text{SMALL}$$

Variance of the tip data relative to phylogeny

HIGH – if tree doesn’t explain variation in the data well

Phylogenetic signal: Blomberg’s *K*

$$\frac{\mathrm{MSE_0}}{\mathrm{MSE}}$$ is different for every tree (it depends on tree size and shape)

Therefore we divide observed value by the expected value under Brownian motion so we can compare trees

$$\mathrm{K} = \text{ observed}\frac{\mathrm{MSE_0}}{\mathrm{MSE}}\Big/\text{expected}\frac{\mathrm{MSE_0}}{\mathrm{MSE}}$$

Phylogenetic signal: Blomberg's *K*

- K = 1: trait values are as expected under BM (= λ = 1)
- K > 1: trait values more similar than expected under BM
- K = 0: no relationship between phylogeny and trait (= λ = 0)

Blomberg's *K*: Summary

1) Ratio of variance in trait relative to phylogenetic mean and variance in trait relative to phylogeny
2) K = 0: no relationship between trait and phylogeny
3) K = 1: trait values are as expected under Brownian motion
4) K > 1: trait values more similar than expected under Brownian motion

Phylogenetic signal: λ versus *K**

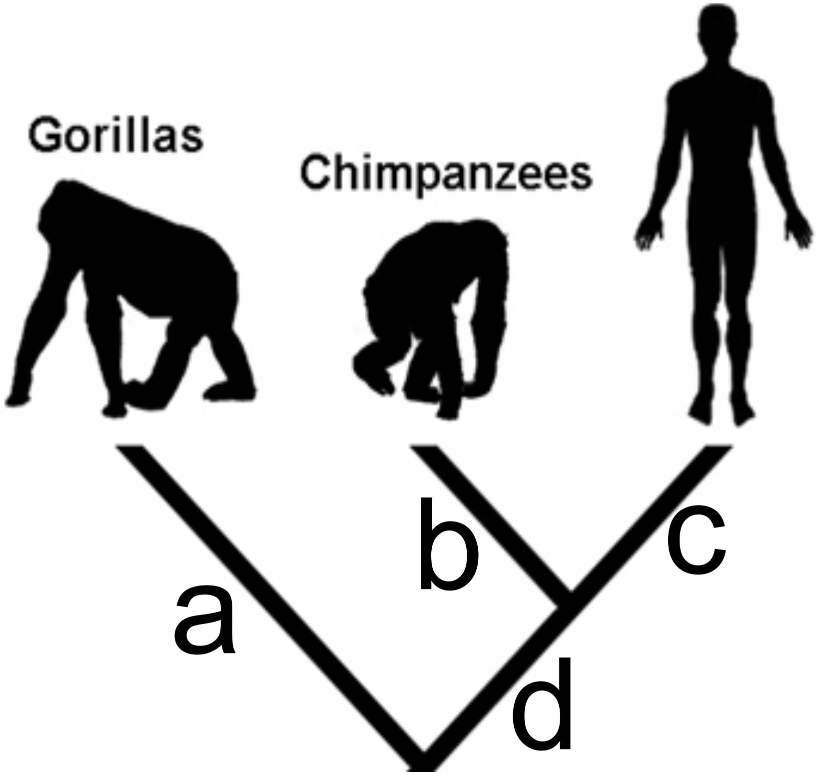
1. Ranges from 0 to just above 1 (though most functions in R fix lambda to be <=1)
2. The maximum possible value is set by the tree in question

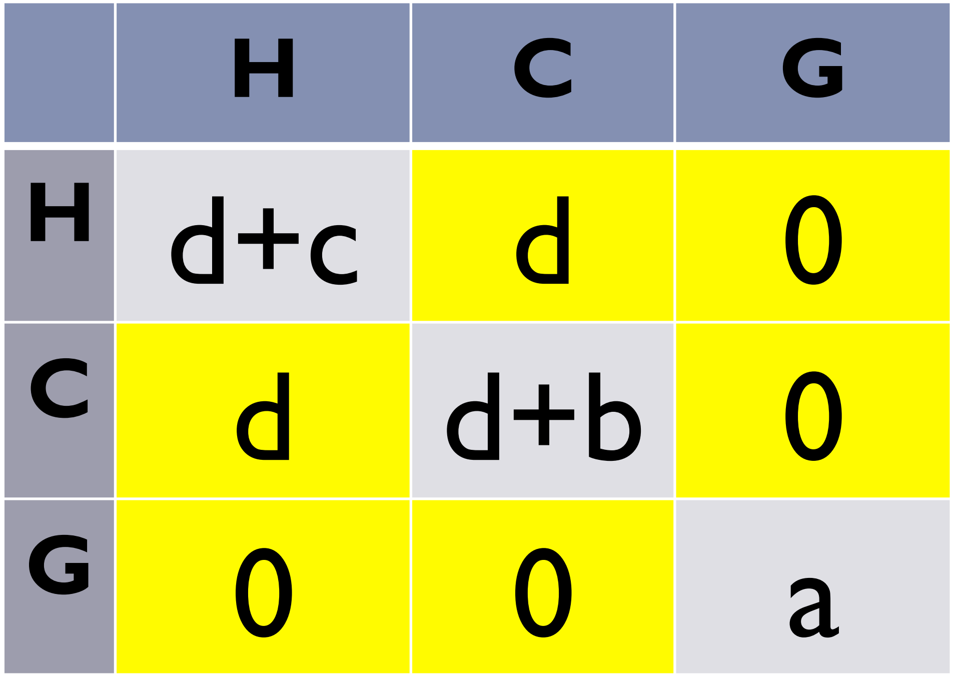
1. Ranges from 0 to some trait dependent maximum
2. Useful for looking at phylogenetic signal in traits showing a lot of conservatism (PNC: see Losos 2008, Cooper et al 2010)

Other methods

Nested analysis of variance	(Harvey & Pagel 1991)
Autocorrelation coefficient (ρ)	(Cheverud et al. 1985, Gittleman & Kot 1990 see also Grafen 1990)
R^2	(Cheverud et al. 1985, (Gittleman & Kot 1990)
Moran's *I*	(Gittleman & Kot 1990)
Randomization for discrete characters	(Maddison & Slatkin 1991)
Quantitative covergence index (QCI)	(Ackerly & Donoghue 1998)
Fritz and Purvis' D	(Fitz & Purvis 2010)
…	

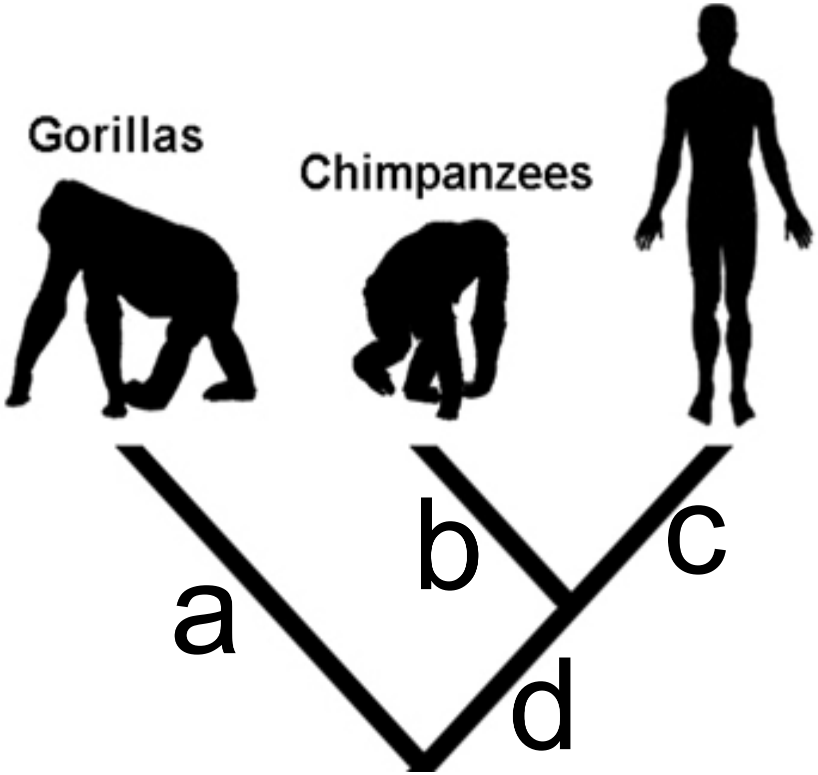
Delta: elements are raised to a power δ





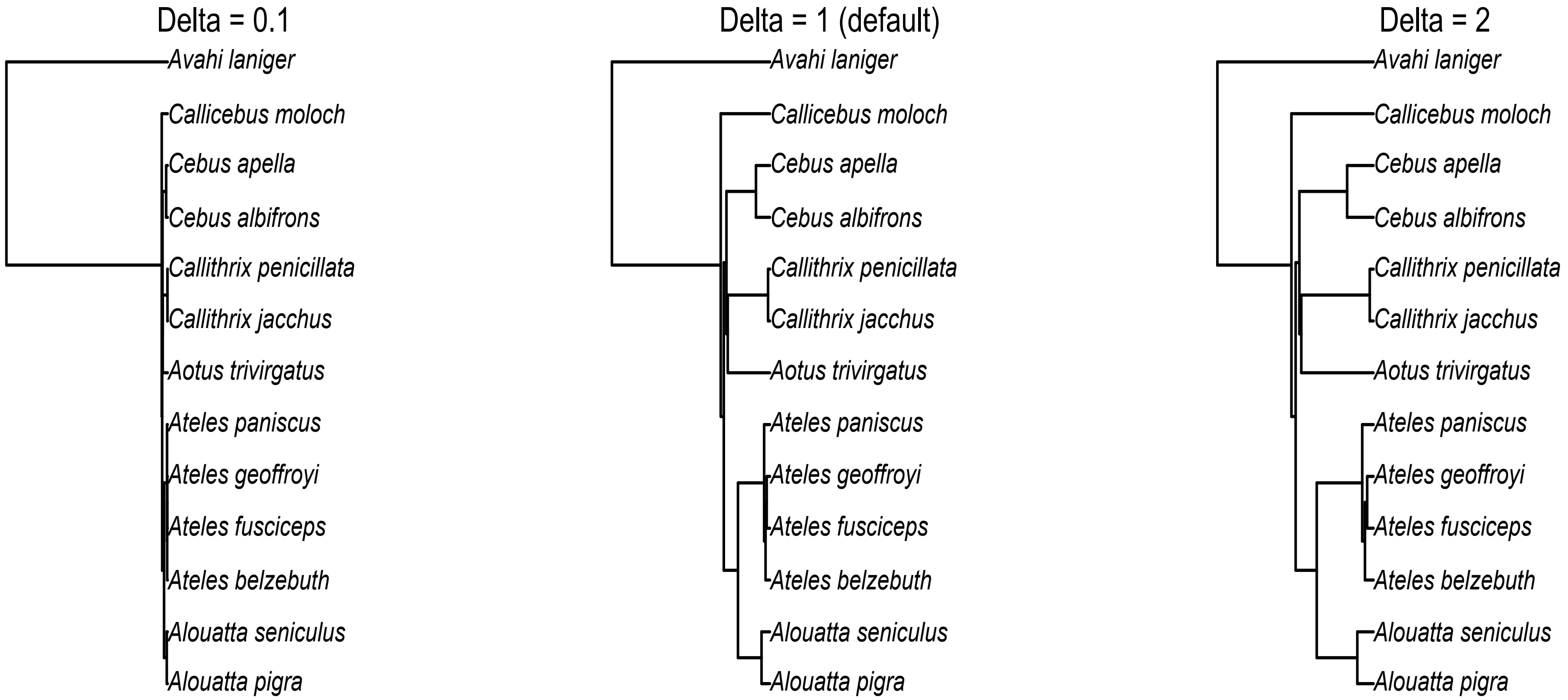
	H	C	G
H	d+c	d	0
C	d	d+b	0
G	0	0	a

Delta: elements are raised to a power δ



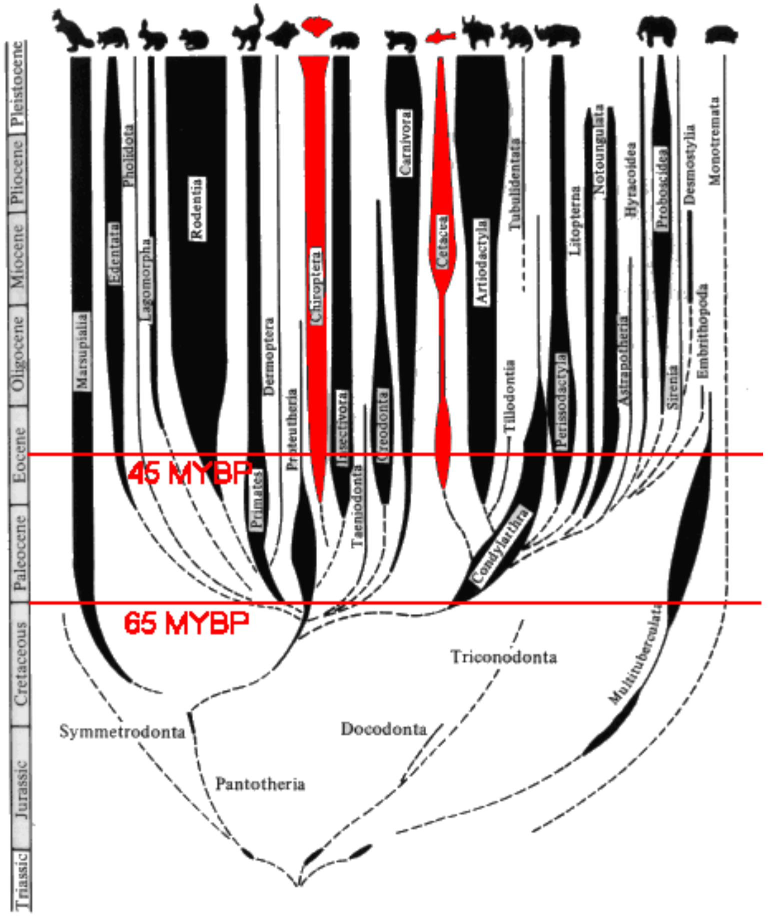
	H	C	G
H	$(d+c)^{\delta}$	$(d)^{\delta}$	$(0)^{\delta}$
C	$(d)^{\delta}$	$(d+b)^{\delta}$	$(0)^{\delta}$
G	$(0)^{\delta}$	$(0)^{\delta}$	$(a)^{\delta}$

Delta: elements are raised to a power δ

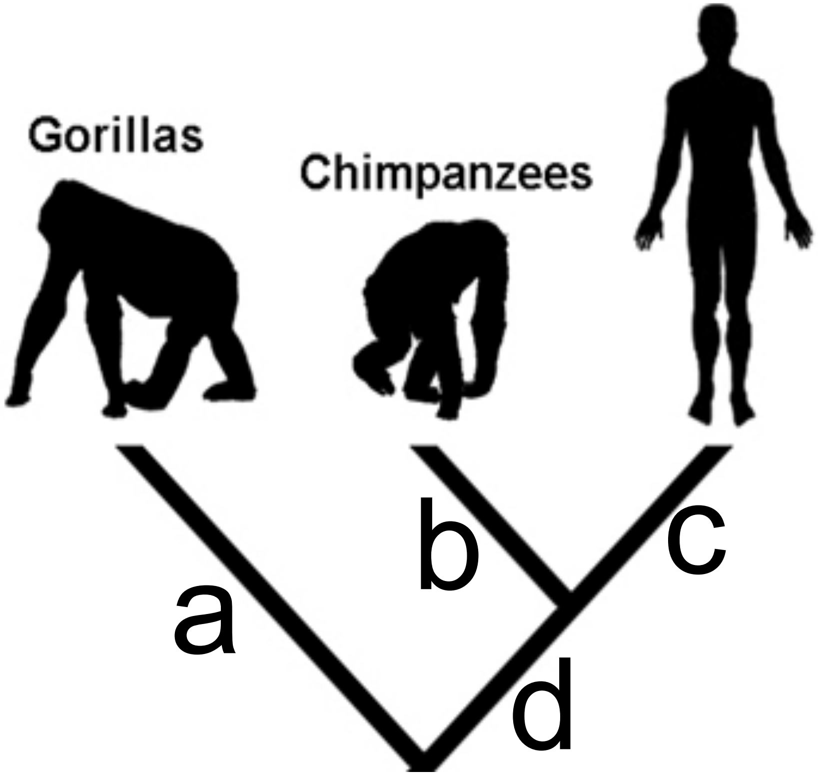


Delta: elements are raised to a power δ

- Scales overall path lengths in the phylogeny (node height)
- Can be used to test for accelerated evolution versus adaptive radiation
  - $\delta < 1$ shorter paths (earlier evolution in the phylogeny) contribute disproportionately to trait evolution (adaptive radiation)
  - $\delta > 1$ longer paths contribute more to trait evolution (accelerated evolution)
- Delta is a parameter that detects differential rates of evolution over time and re-scales the phylogeny to a basis in which the rate of evolution is constant

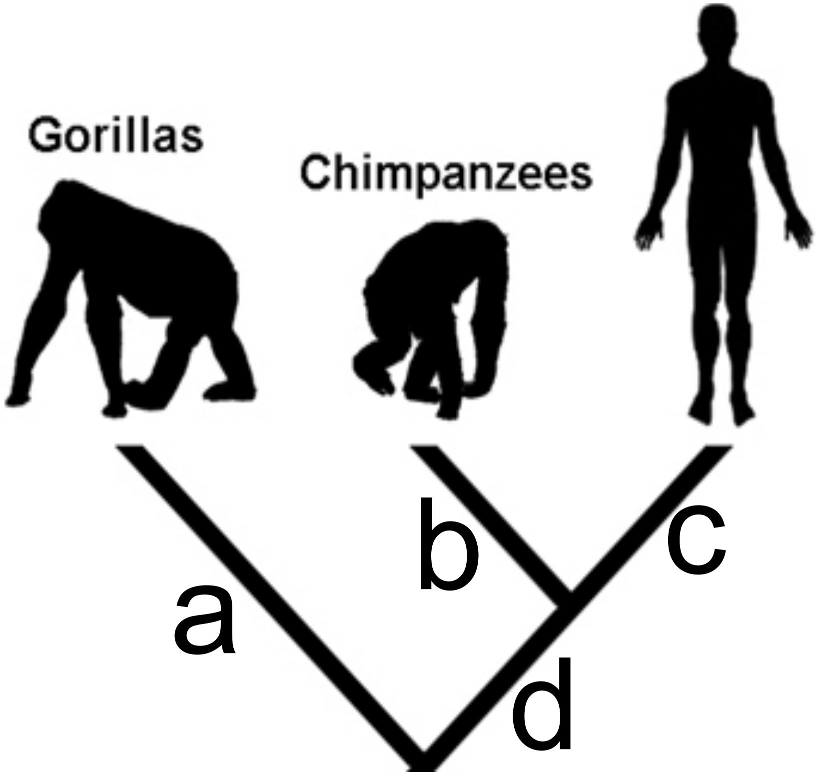


Kappa: branch length are raised to a power κ



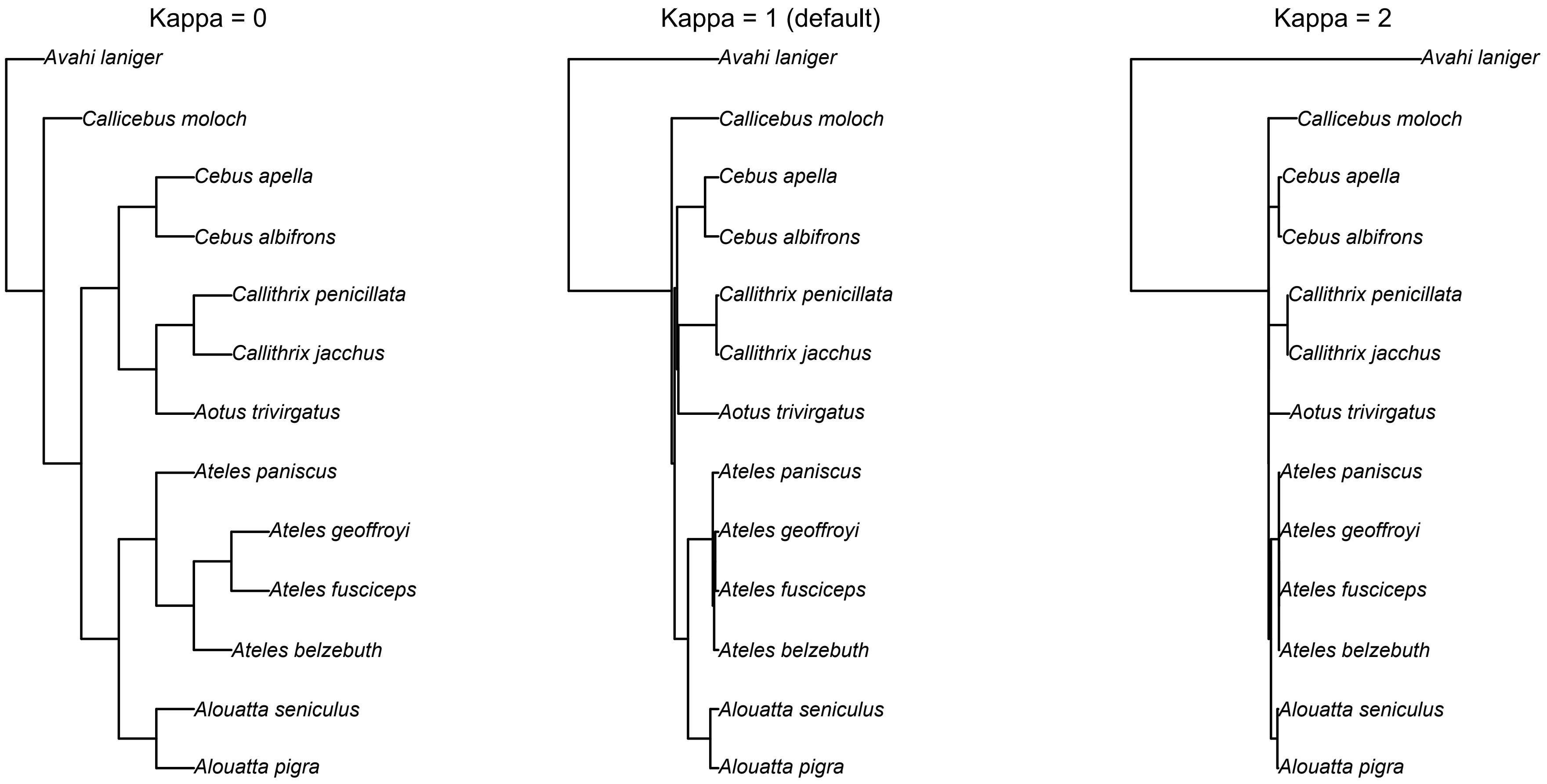
	H	C	G
H	d+c	d	0
C	d	d+b	0
G	0	0	a

Kappa: branch length are raised to a power κ



	H	C	G
H	$d^{\kappa}+c^{\kappa}$	d^{κ}	0^{κ}
C	d^{κ}	$d^{\kappa}+b^{\kappa}$	0^{κ}
G	0^{κ}	0^{κ}	a^{κ}

Kappa: branch length are raised to a power κ



Kappa: branch length are raised to a power κ

- Differentially stretches or compresses individual phylogenetic branch lengths
- Can be used to test for a punctuational versus gradual mode of trait evolution
  - κ < 1 compresses longer branches more than shorter ones
  - κ > 1 stretches longer branches more than shorter ones
  - κ ~ 0 evolution is independent on branch length (punctuational evolution)
  - κ ~ 1 gradual evolution
- Captures patterns of "speciational" change in tree
  - character change is more or less concentrated at speciation events

Alternatives to Brownian motion

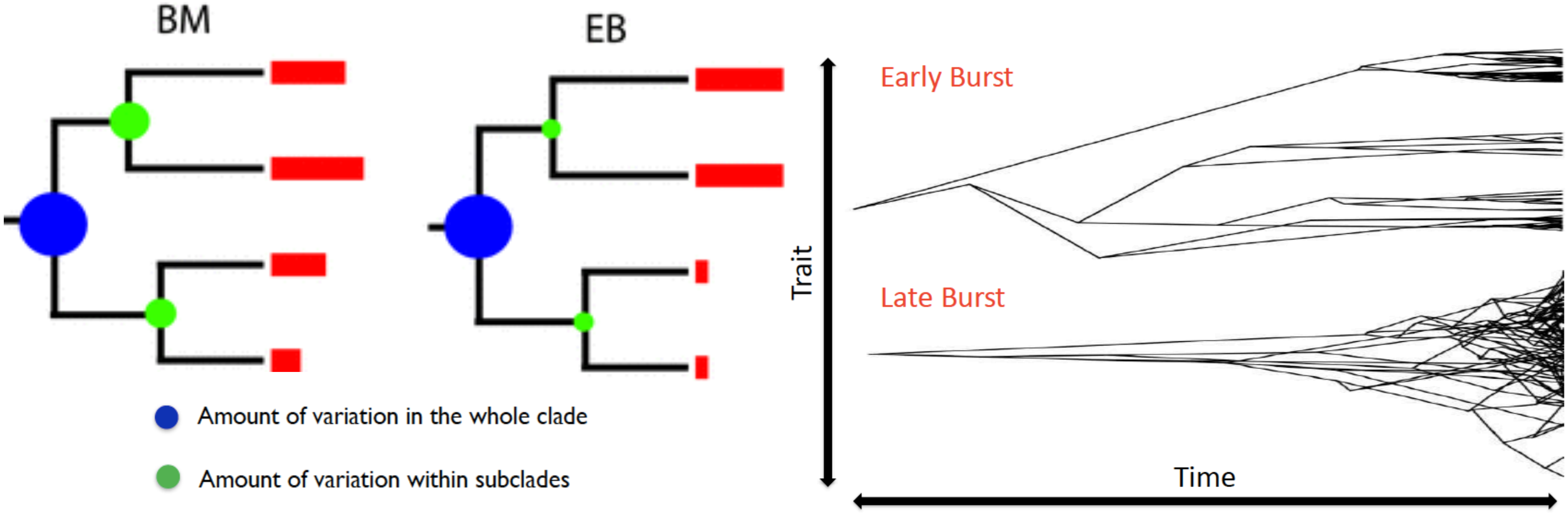
- Variable rates over the tree
- Declining rates through time (Early Burst, EB/AC)
- Accelerating rates through time (Late Burst, LB/DC)
- A single stable adaptive peak (Ornstein-Uhlenbeck, OU)
- Variable adaptive peaks (Ornstein-Uhlenbeck, OU)
- Trends in the mean trait value (BM with a trend)
- Mixtures of the above, and more

Alternatives to Brownian motion

- Variable rates over the tree
- Declining rates through time (Early Burst, EB/AC)
- Acceler B/DC)
- A singl beck, OU)
- Variabl OU)
- Trends nd)
- Mixtures of the above, and more

Early Burst (EB/AC) - Late Burst (LB/DC)

- EB: BM with a declining rate parameter, most of the phenotypic divergence occurs early in the phylogeny
- LB: BM with an accelerating rate parameter, most of the phenotypic divergence occurs late in the phylogeny



Harmon et al. 2010

Early Burst (EB/AC) - Late Burst (LB/DC)

starting parameter

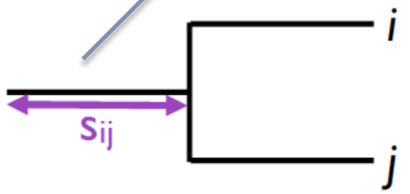
$$\mu_i(t) = \bar{z}_0$$

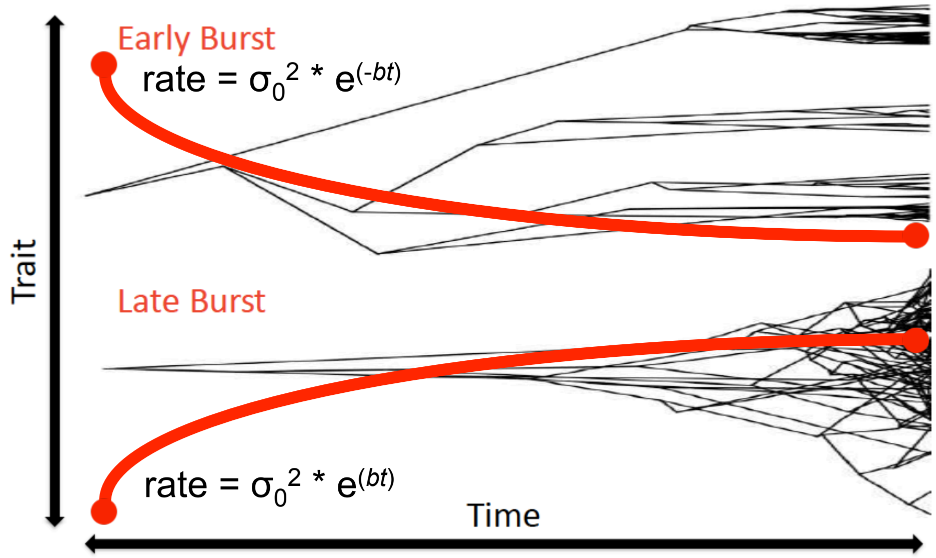
$$\sigma^2(t) = \sigma_0^2 e^{bt}$$

parameter describing the pattern of rate change through time

initial value for the net rate parameter

$$V_{ij}(t) = \sigma_0^2 \frac{e^{b s_{ij}}-1}{b}$$



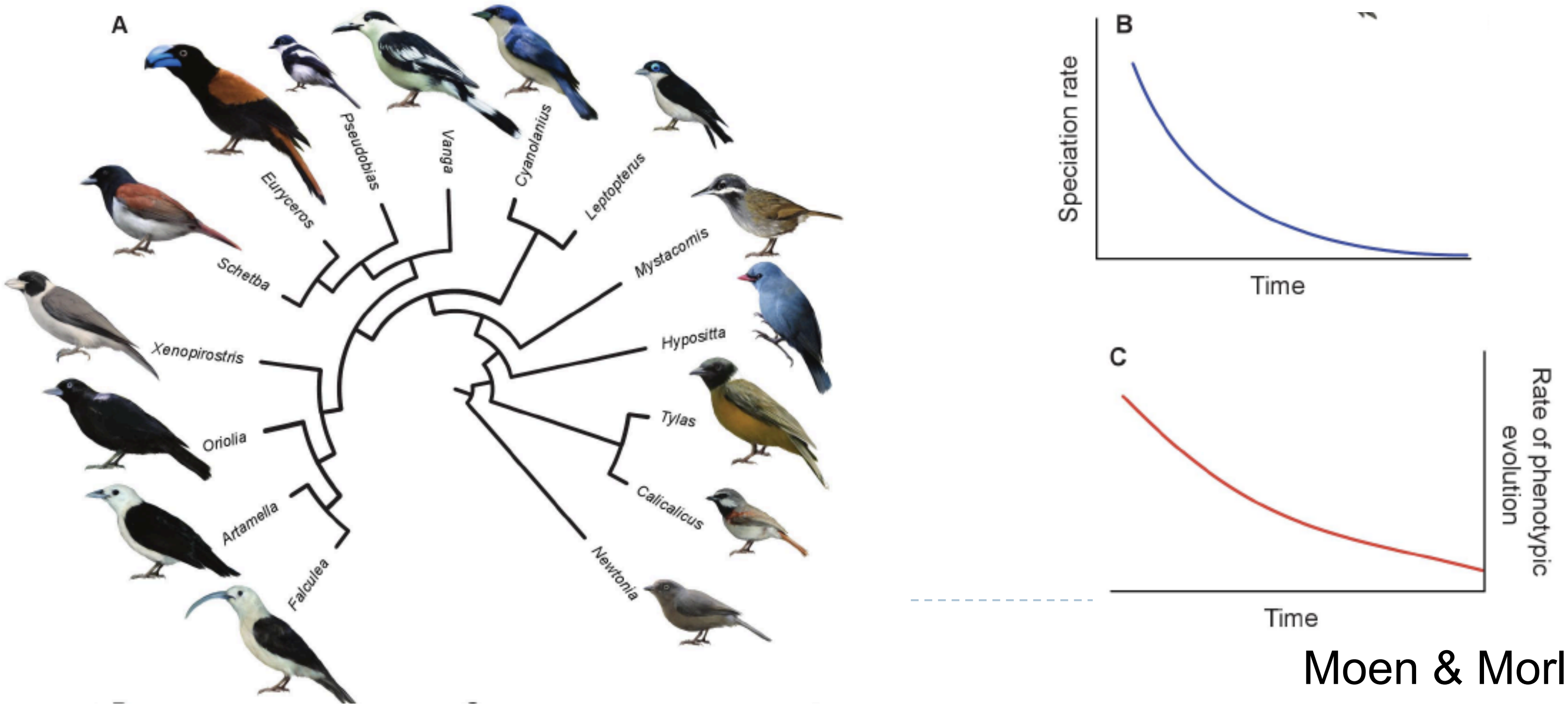


Harmon et al. 2010

Early Burst (EB/AC) - Late Burst (LB/DC)

- **Consistent with the adaptive radiation hypothesis**
  - Clades entering into new niches should diversify quickly
  - Rates slow down as the niches fill

The adaptive radiation of the bird clade *Vanginae*

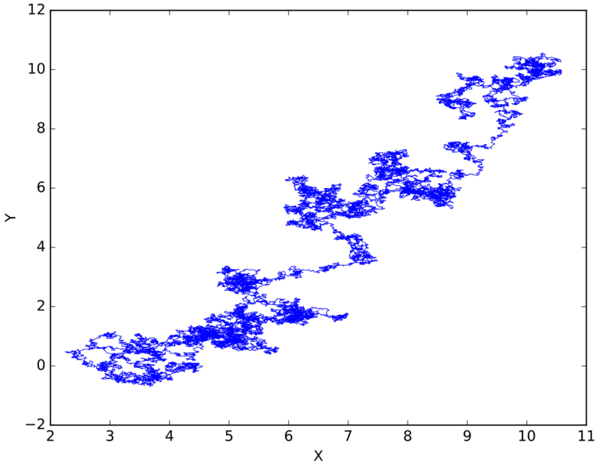


Moen & Morlon 2014

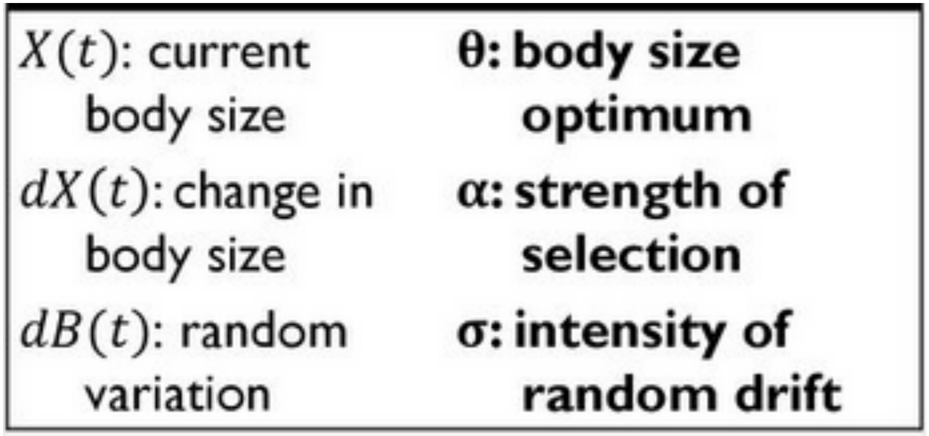
Alternatives to Brownian motion

- Variable rates over the tree
- Declining rates through time (Early Burst, EB/AC)
- Accelerating rates through time (Late Burst, LB/DC)
- A single stable adaptive peak (Ornstein-Uhlenbeck, OU)
- Variable adaptive peaks (Ornstein-Uhlenbeck, OU)
- Trends in the mean trait value (BM with a trend)
- Mixtures of the above, and more

The Ornstein-Uhlenbeck process



$$dX_{(t)} = \alpha\,[\Theta - X_{(t)}]dt + \sigma\, dB_{(t)}$$



$X(t)$: current body size

$dX(t)$: change in body size

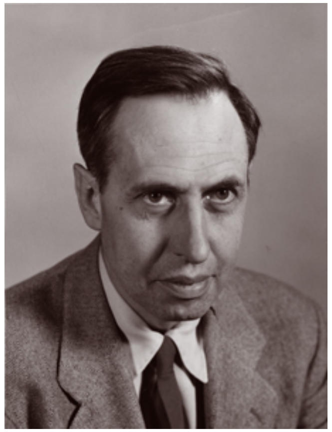
$dB(t)$: random variation

θ: body size optimum

α: strength of selection

σ: intensity of random drift

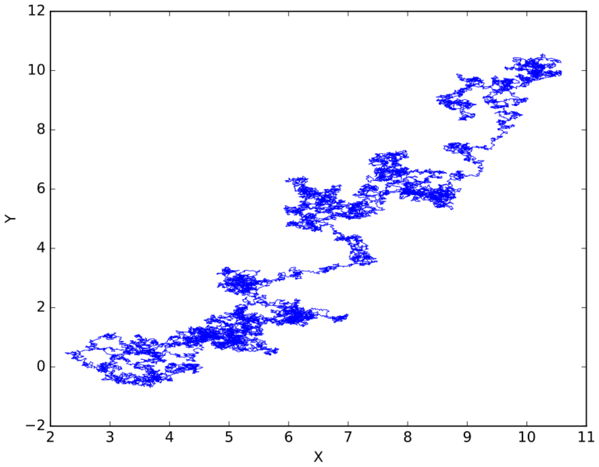
Bayesian walk under the influence of friction: tendency to move back towards a central location (rubber band effect)



- Leonard Ornstein (1880 – 1941)

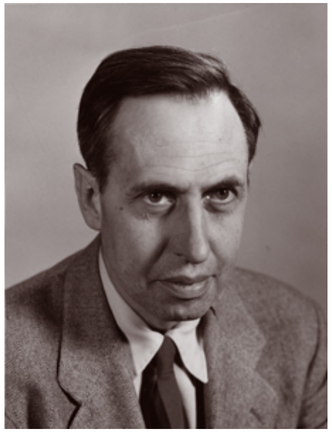
George Uhlenbeck (1900 – 1988)

The Ornstein-Uhlenbeck process



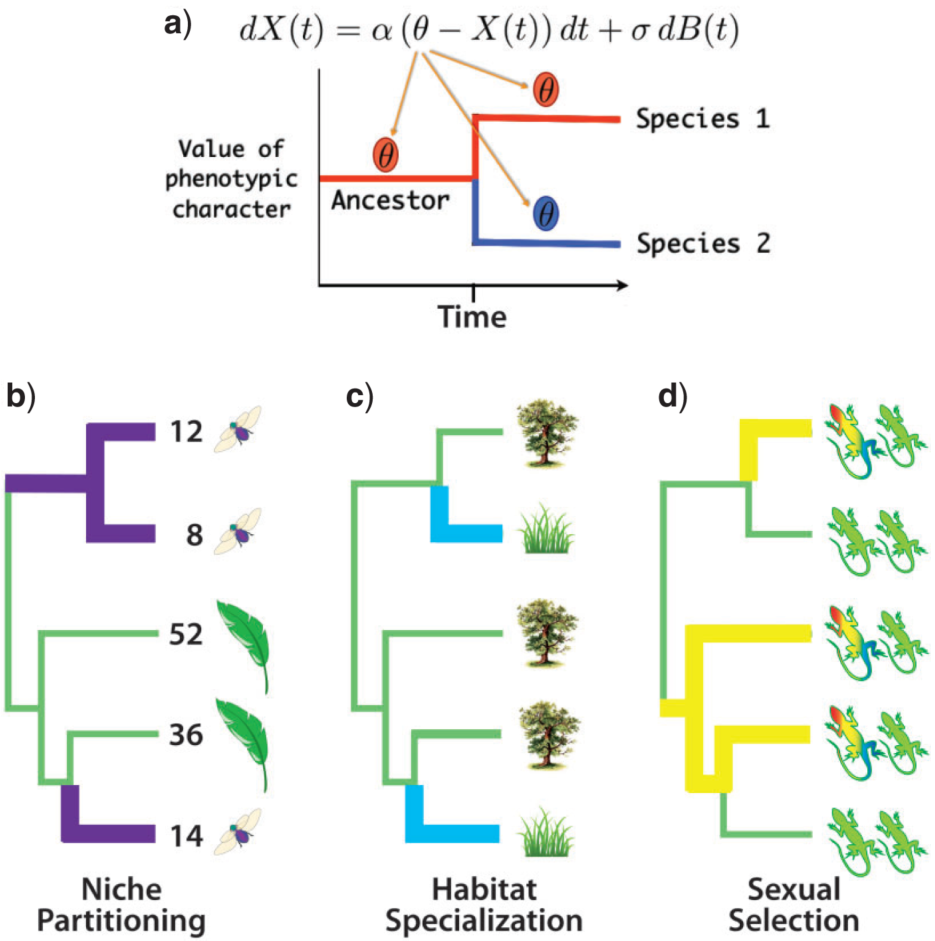
$$dX_{(t)} = \underbrace{\boxed{\alpha\,[\Theta - X_{(t)}]dt}}_{\text{change towards an optimum}} + \underbrace{\boxed{\sigma\, dB_{(t)}}}_{\text{Brownian motion}}$$

Bayesian walk under the influence of friction: tendency to move back towards a central location (rubber band effect)



- Leonard Ornstein (1880 – 1941)

George Uhlenbeck (1900 – 1988)

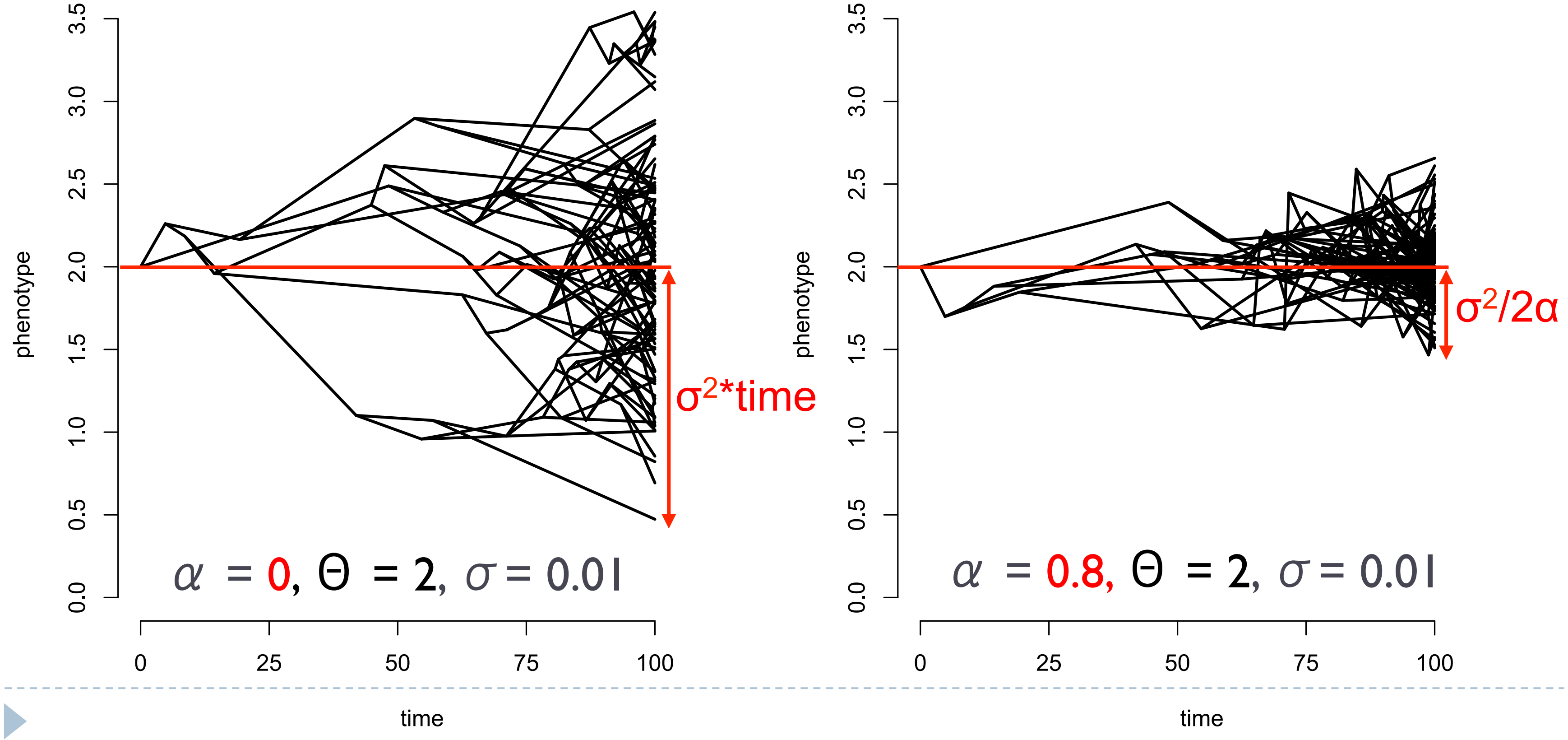


From: Detecting Adaptive Evolution in Phylogenetic Comparative Analysis Using the Ornstein–Uhlenbeck Model

Syst Biol. 2015;64(6):953-968. doi:10.1093/sysbio/syv043

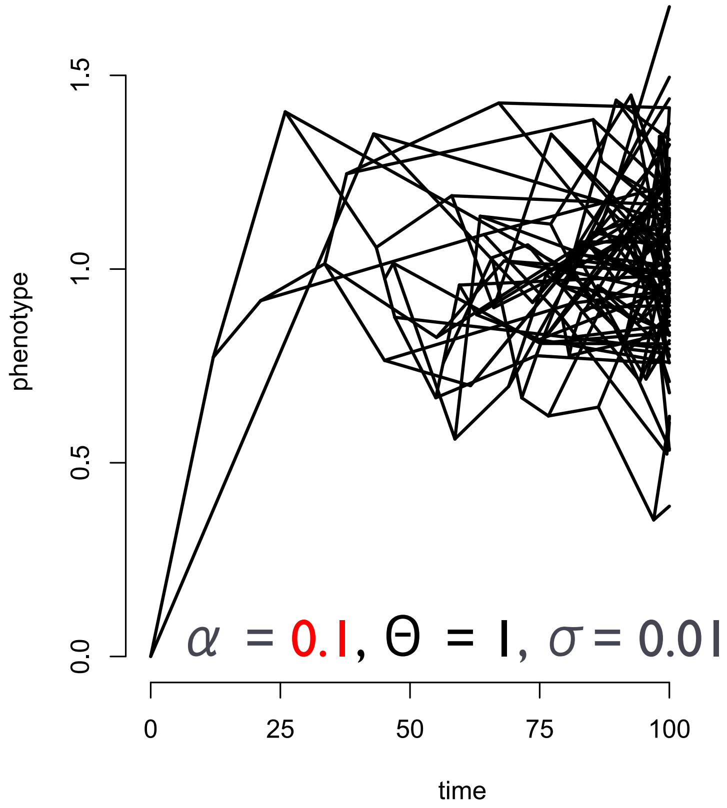
$dX_{(t)} = \alpha [\Theta - X_{(t)}] dt + \sigma dB_{(t)}$

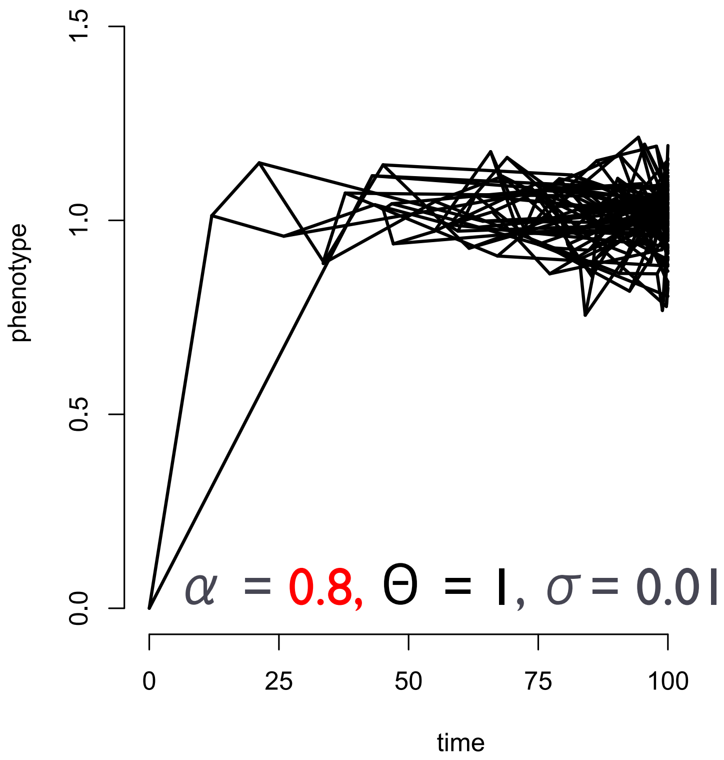
if α = 0, it defines a diversifying process (BM), if α > 0 it becomes an equilibrium process (OU)



$$dX_{(t)} = \alpha\,[\,\Theta - X_{(t)}\,]\,dt + \sigma\, dB_{(t)}$$

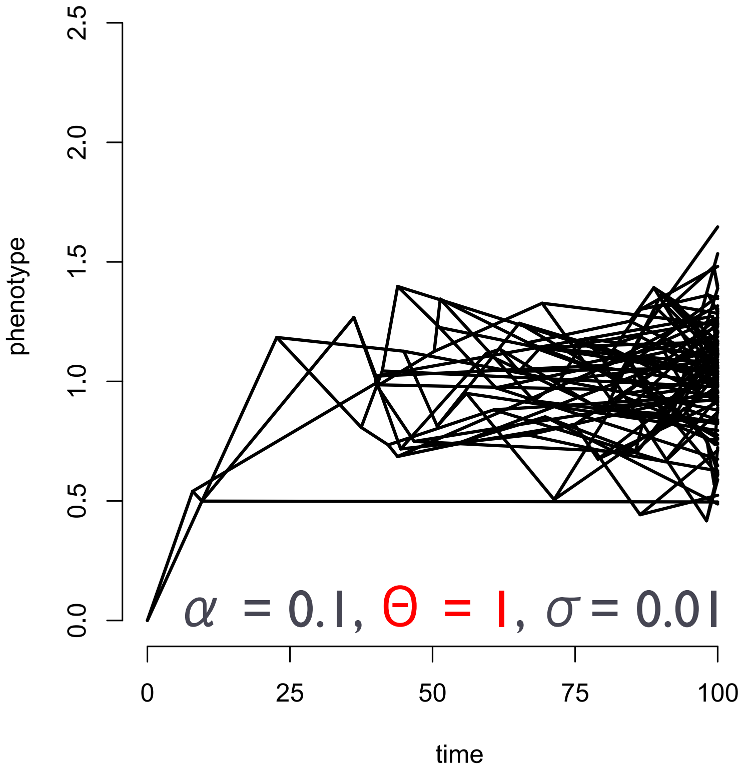
The higher the attraction parameter α the more quickly the optima is reached and the lower the variance

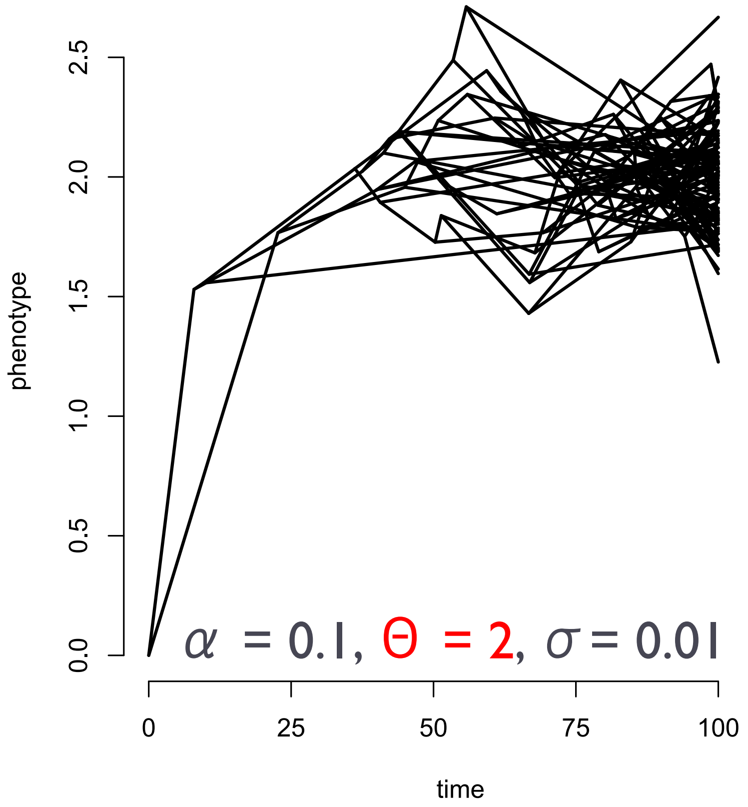




$dX_{(t)} = \alpha [\Theta - X_{(t)}] dt + \sigma dB_{(t)}$

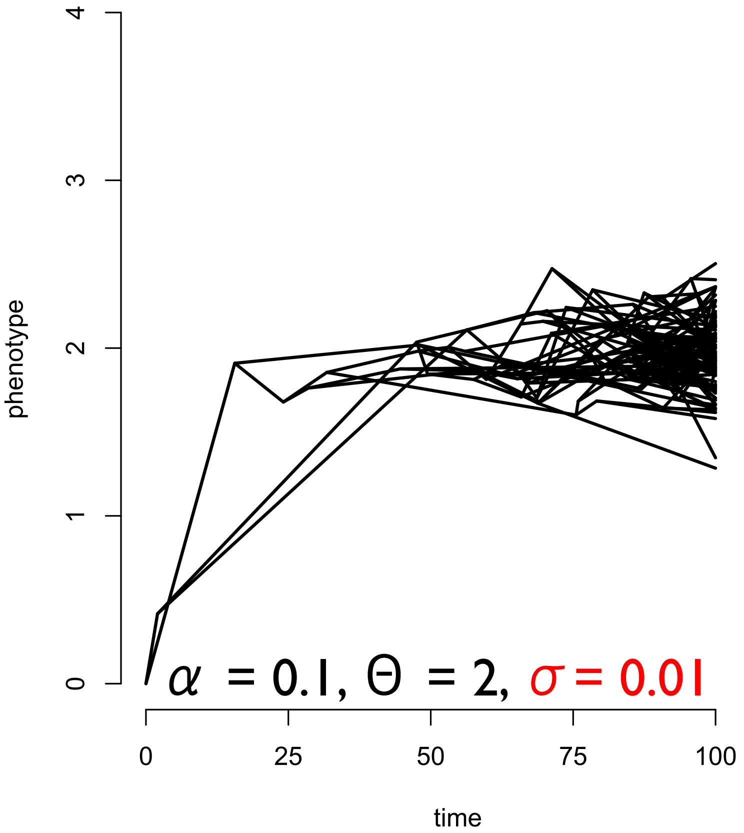
The higher the optimal value Θ the greater the trait value

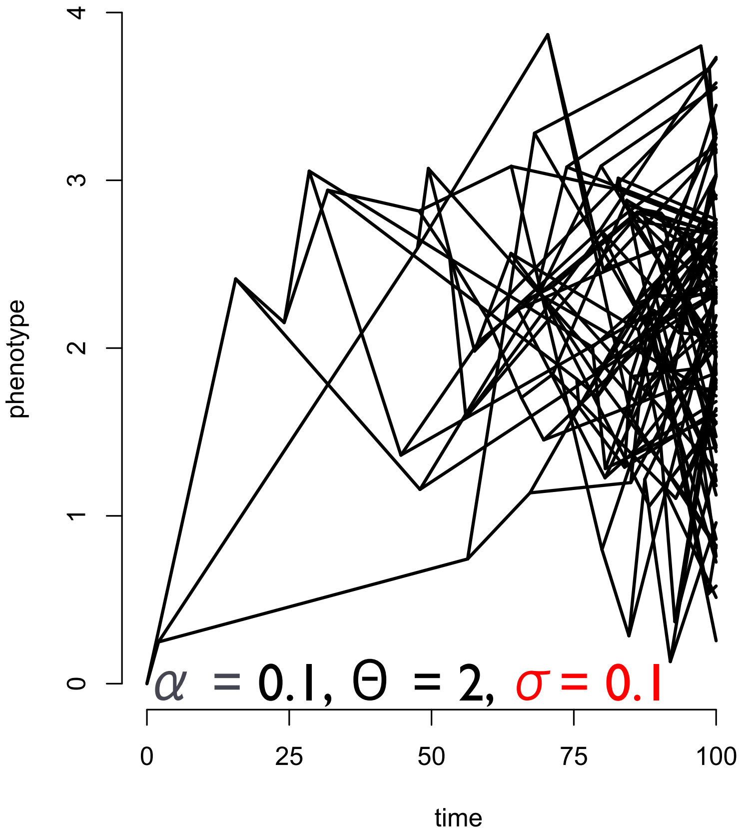




$$dX_{(t)} = \alpha\,[\Theta - X_{(t)}]\,dt + \sigma\, dB_{(t)}$$

The higher the rate parameter σ the greater the variance



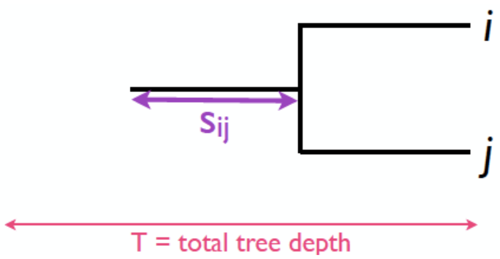


The Ornstein-Uhlenbeck process

ancestral state

$$\mu_i(t) = \theta + e^{-\alpha T_i}(\bar{z}_0 - \theta)$$

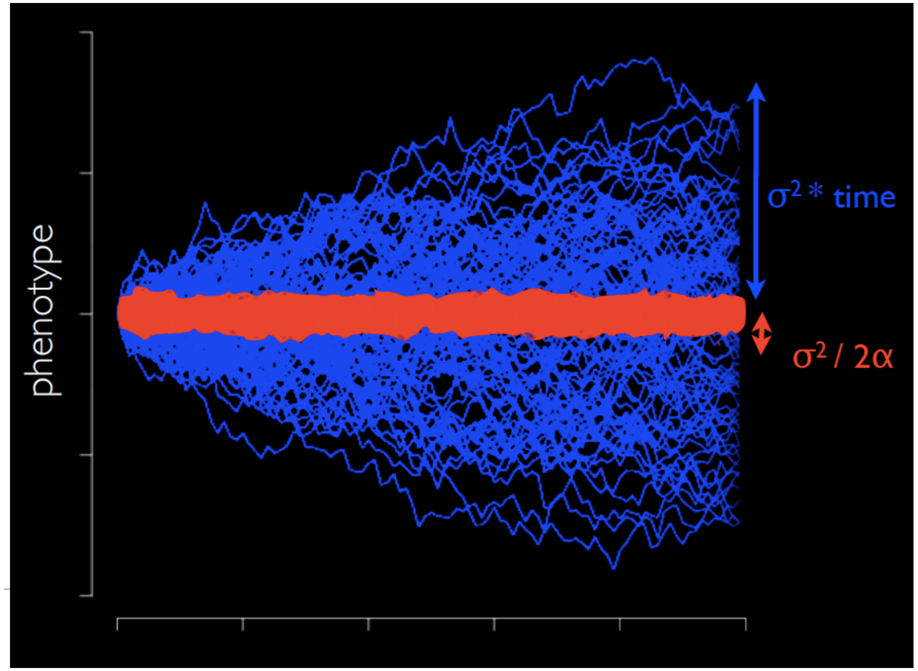
trait optimum



rate

$$V_{ij}(t) = \frac{\sigma^2}{2\alpha}\left(e^{-2\alpha(T_i - s_{ij})} - e^{-2\alpha T_i}\right)$$

strength of selection



Alternatives to Brownian motion

- Variable rates over the tree
- Declining rates through time (Early Burst, EB/AC)
- Accelerating rates through time (Late Burst, LB/DC)
- A single stable adaptive peak (Ornstein-Uhlenbeck, OU)
- Variable adaptive peaks (Ornstein-Uhlenbeck, OU)
- Trends in the mean trait value (BM with a trend)
- Mixtures of the above, and more

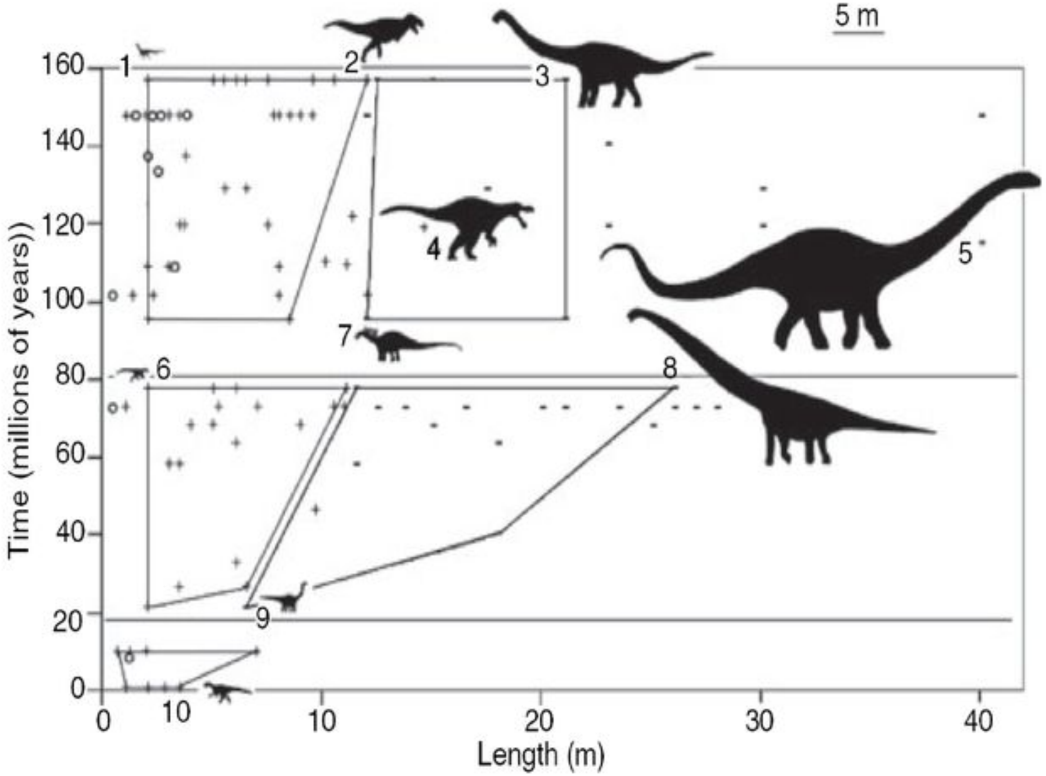
BM with trend

$$dX_{(t)} = \sigma \, dB_{(t)}$$

rate

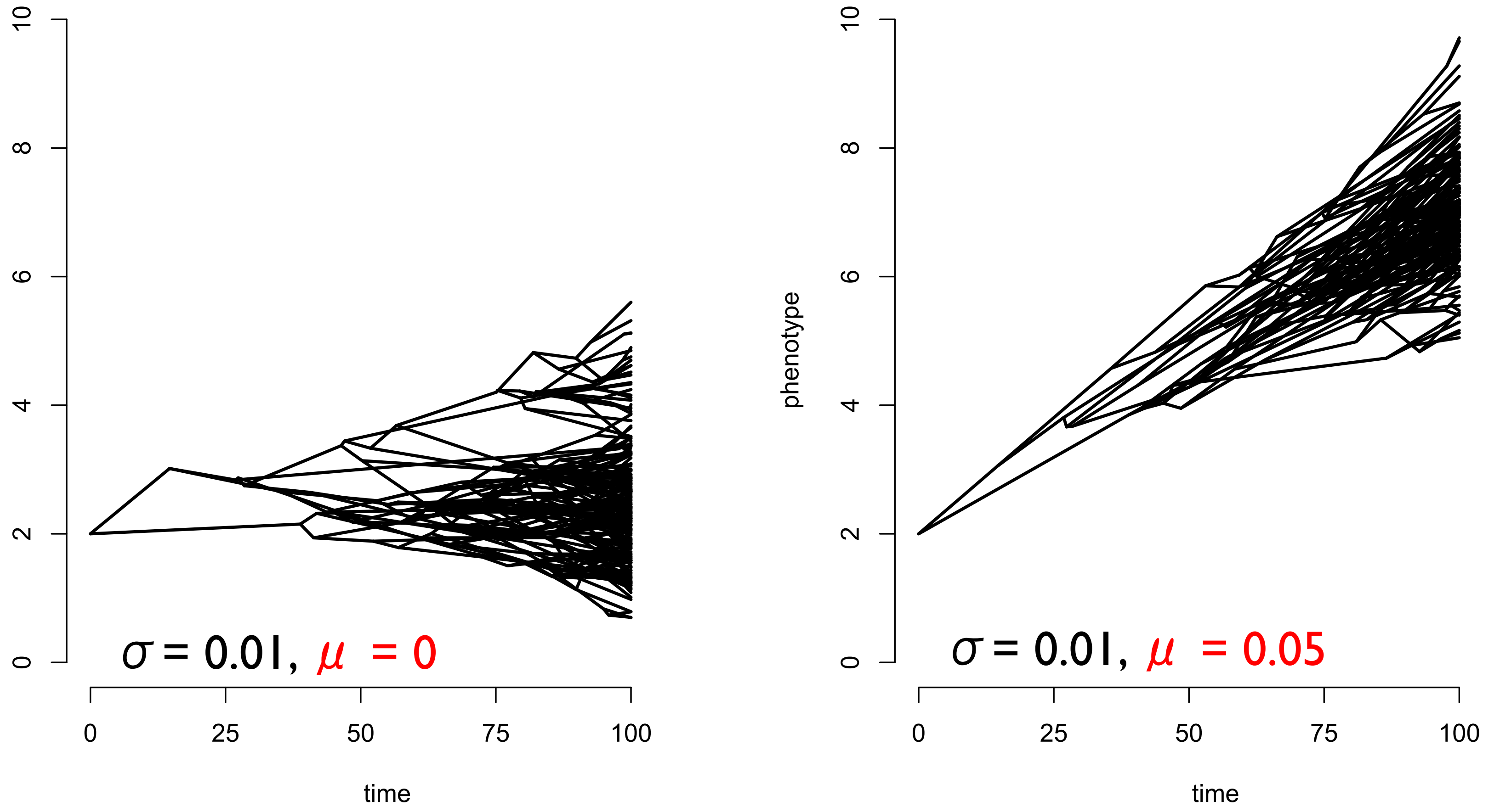
normal distribution
where mean = t * μ
μ > 0 : increase
μ < 0 : decrease
μ = 0 : BM without trend

Cope's rule



- Hone et al. 2005

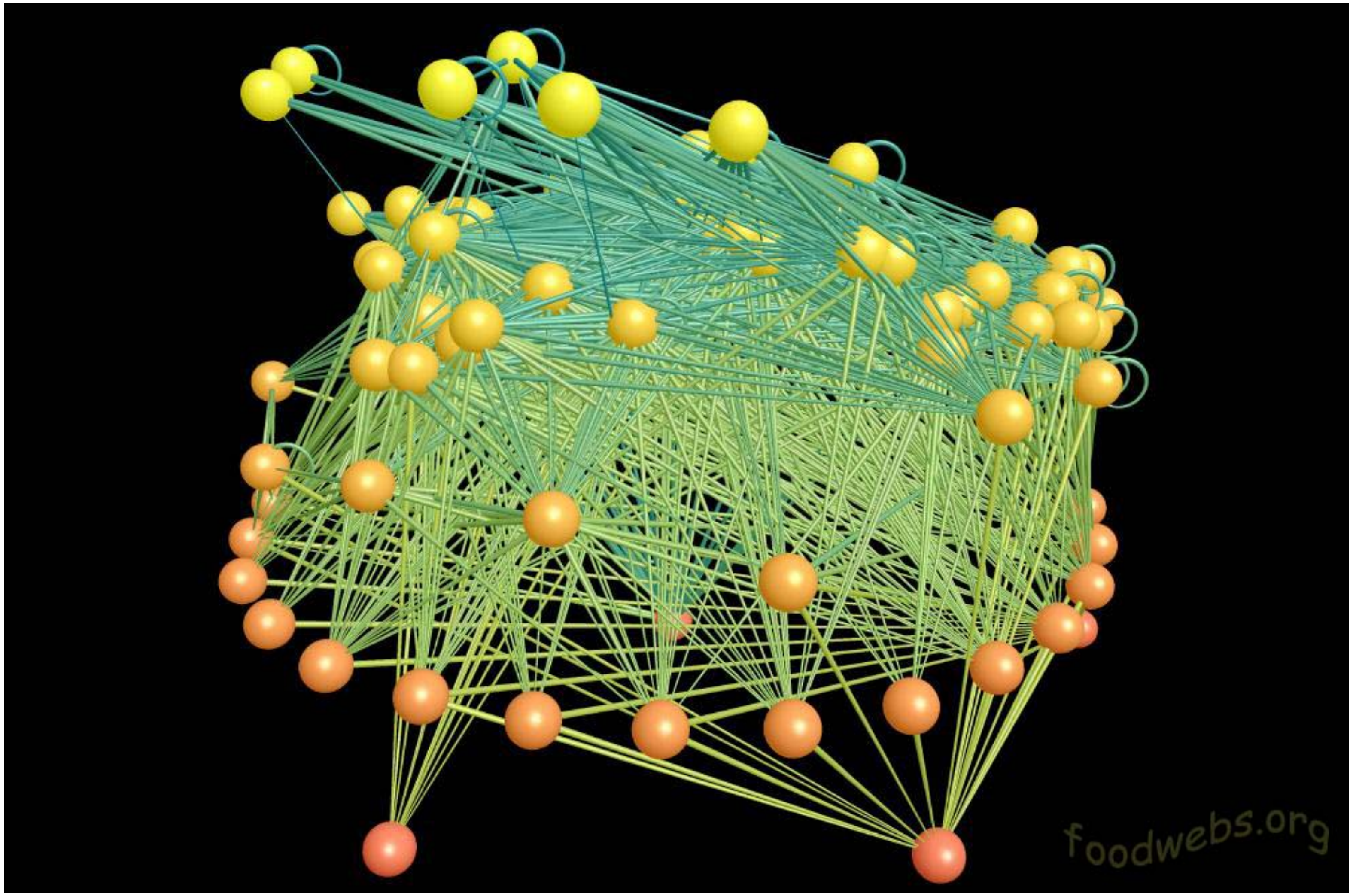
$dX_{(t)} = \sigma \, dB_{(t)}$



Alternatives to Brownian motion

- Variable rates over the tree
- Declining rates through time (Early Burst, EB/AC)
- Accelerating rates through time (Late Burst, LB/DC)
- A single stable adaptive peak (Ornstein-Uhlenbeck, OU)
- Variable adaptive peaks (Ornstein-Uhlenbeck, OU)
- Trends in the mean trait value (BM with a trend)
- Mixtures of the above, and more

How to choose from so many models?



Model selection methods

Empirical

Adjusted R-squared
Bootstrap
Cross-validation
Generalized cross-validation
k-fold crossvalidation
leave-one-out crossvalidation
Jacknife
Linear regression
Shibata's model selector
signal-to-noise ratio
test set validation

Theoretical

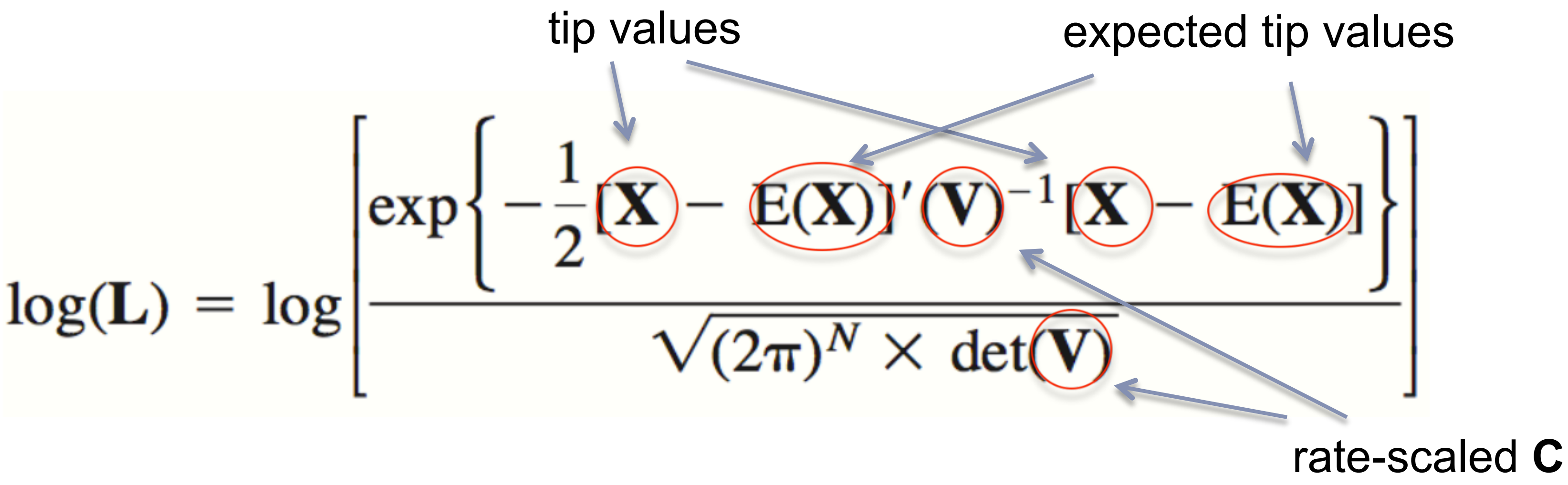
Akaike information criterion (AIC, AICc, QAIC)
Bayesian information criterion (BIC)
CP (Mallow's Cp)
Deviance information criterion (DIC)
Focused Information criterion (FIC)
Final prediction error (FPE)
Geweke and Meese criterion
Generalized prediction error (GPE)
Hannan and Quinn criterion (HQ)
Kullback information criterion (KIC, KICc)
Minimum description length (MDL)
Minimum message length (MML)
Predicted squared error (PSE)
Predicted Residual Sum of. Squares criterion
Schwarz information criterion (SIC)
Structural risk minimization (SRM)
Takeuchi's information criterion (TIC)
VC-dimension

Model selection methods

- Model comparison
  - evaluate multiple hypotheses in competition with one another
    - nested models
      - likelihood ratio tests ("old tool")
    - non-nested models
      - model comparison based on information theory (IT)

Likelihood, Maximum likelihood and likelihood ratio

- Likelihood: probability of obtaining the observed data under a given hypothesis (model and its parameters)
  - Pr(D|H) (but not Pr(H_0|D)!)
- The multivariate normal likelihood for BM



tip values

expected tip values

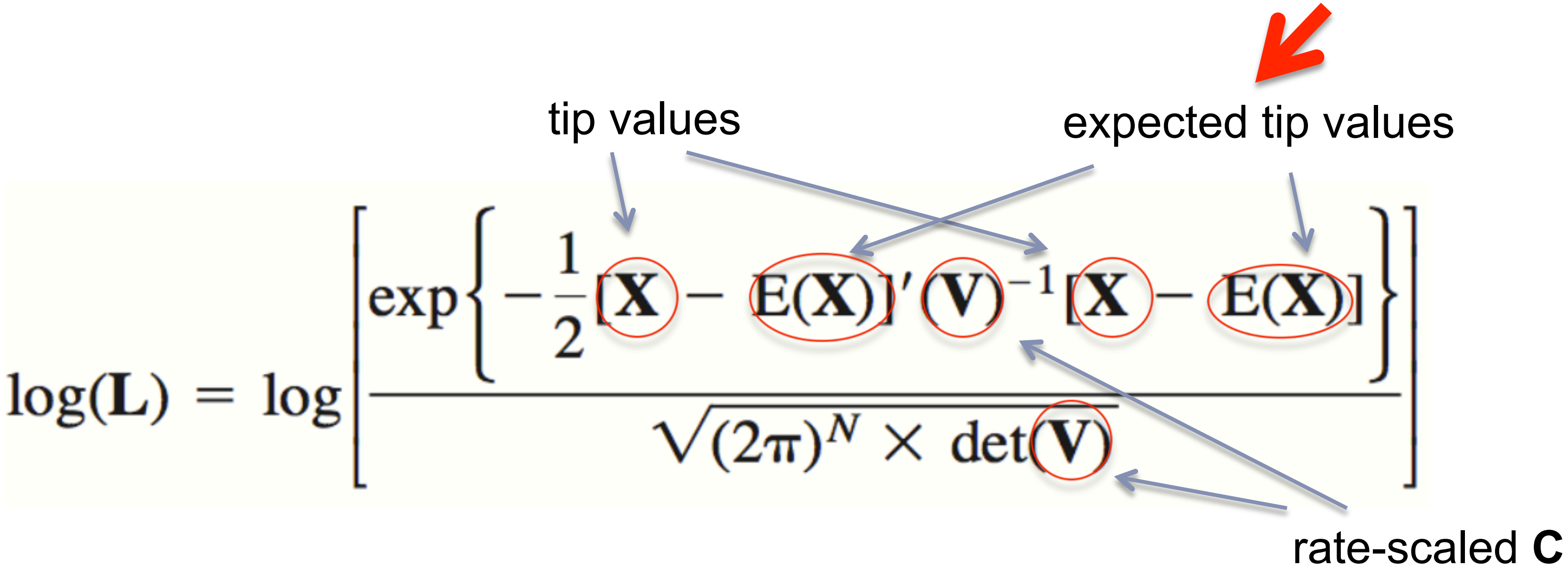
$$\log(\mathbf{L}) = \log\left[\frac{\exp\left\{-\frac{1}{2}[\mathbf{X}-\mathrm{E}(\mathbf{X})]'(\mathbf{V})^{-1}[\mathbf{X}-\mathrm{E}(\mathbf{X})]\right\}}{\sqrt{(2\pi)^N \times \det(\mathbf{V})}}\right]$$

rate-scaled **C**

Likelihood, Maximum likelihood and likelihood ratio

- Likelihood: probability of obtaining the observed data under a given hypothesis (model and its parameters)
  - Pr(D|H) (but not Pr(H_0|D)!)
- The multivariate normal likelihood for BM

model and its parameters: (starting value, strenght and direction of trend..)



tip values

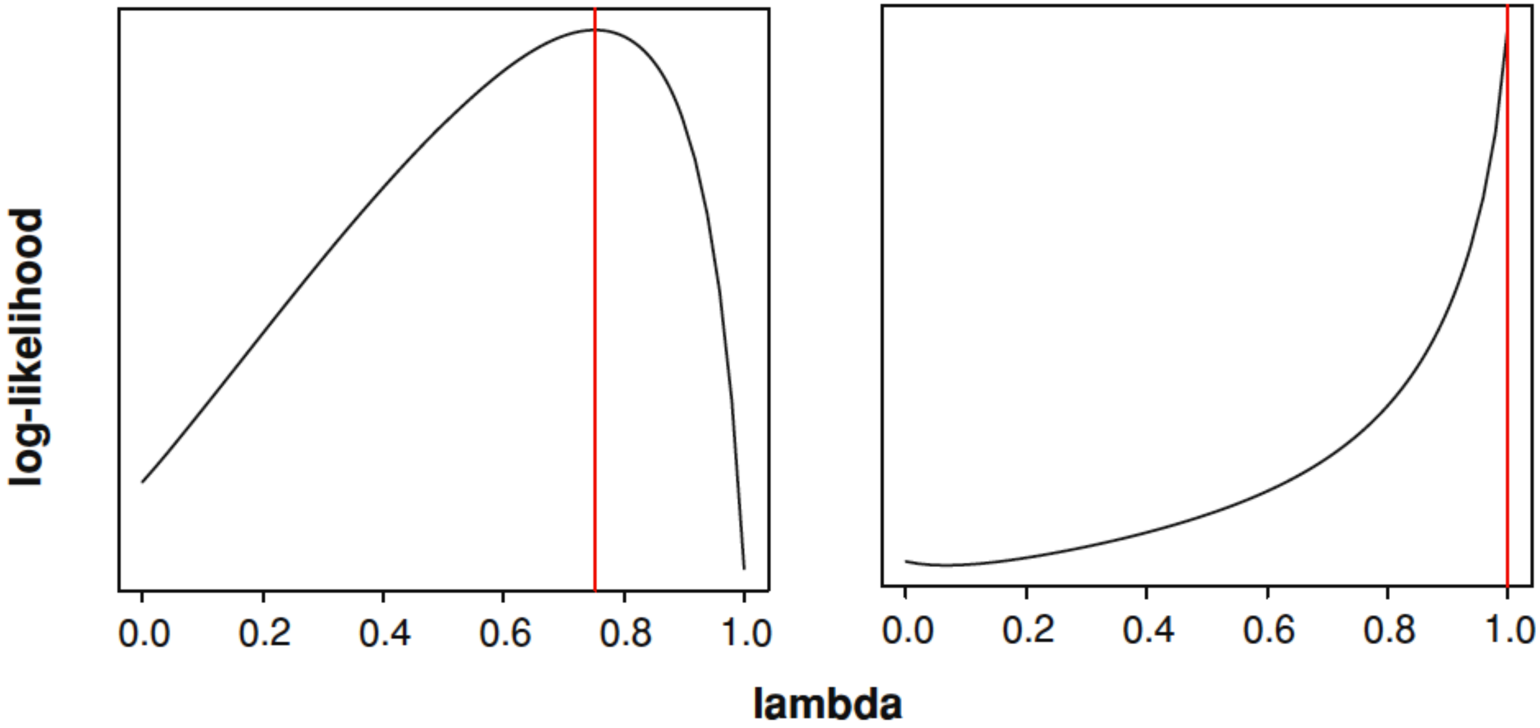
expected tip values

$$\log(\mathbf{L}) = \log\left[\frac{\exp\left\{-\frac{1}{2}[\mathbf{X}-\mathrm{E}(\mathbf{X})]'(\mathbf{V})^{-1}[\mathbf{X}-\mathrm{E}(\mathbf{X})]\right\}}{\sqrt{(2\pi)^N \times \det(\mathbf{V})}}\right]$$

rate-scaled **C**

Likelihood, Maximum likelihood and likelihood ratio

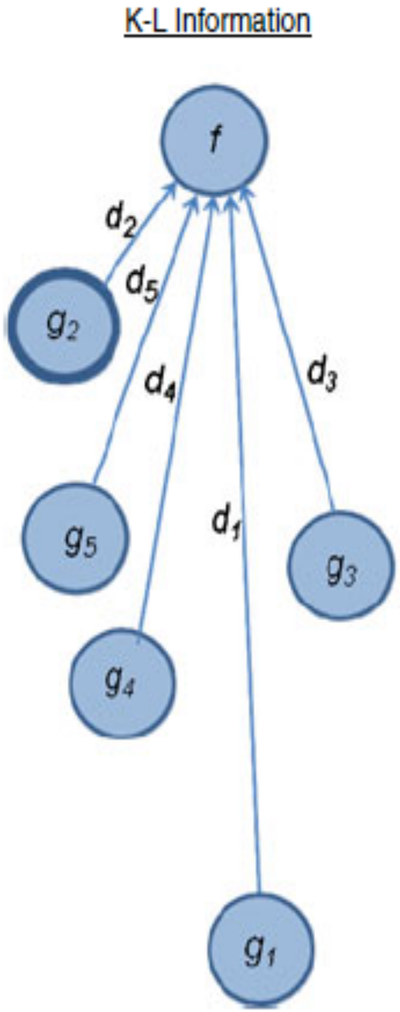
- Maximum likelihood: the value of one or more parameters for a given model, which maximizes the likelihood

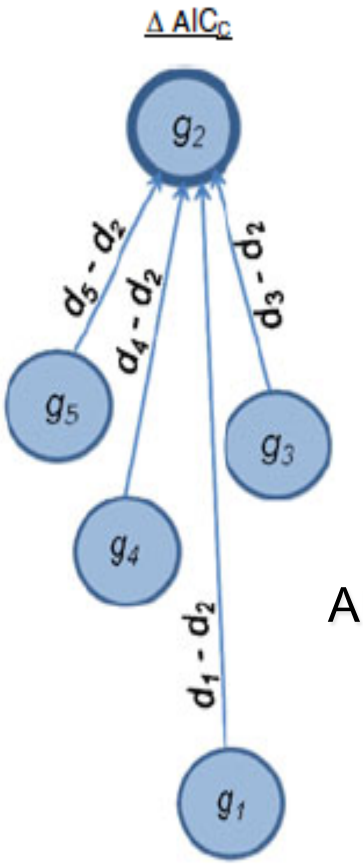


Likelihood, Maximum likelihood and likelihood ratio

- Likelihood ratio: model fit of one model relative to another
- Likelihood ratio test (LRT): a statistical test of the goodness-of-fit between two models
  - LRT = $2 * [\ln(L_1) - \ln(L_2)]$
  - approximates a chi-square distribution
  - with df = nr. of parameters differing between models

Information theoretic approach





fit
(likelihood)

parsimony
(nr. of parameters)

$$\mathrm{AIC} = -2\ln(L) + 2k$$

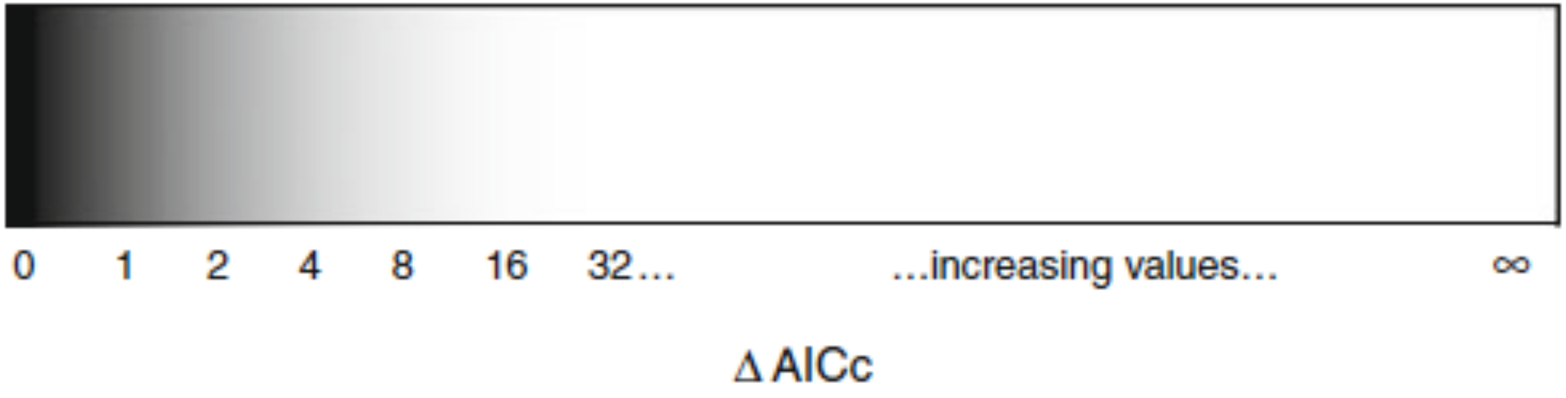
Akaike's Information Criterion

$$\mathrm{AICc} = \mathrm{AIC} + (2K(K+1))/(n-K-1).$$

Akaike 1974

Information theoretic approach

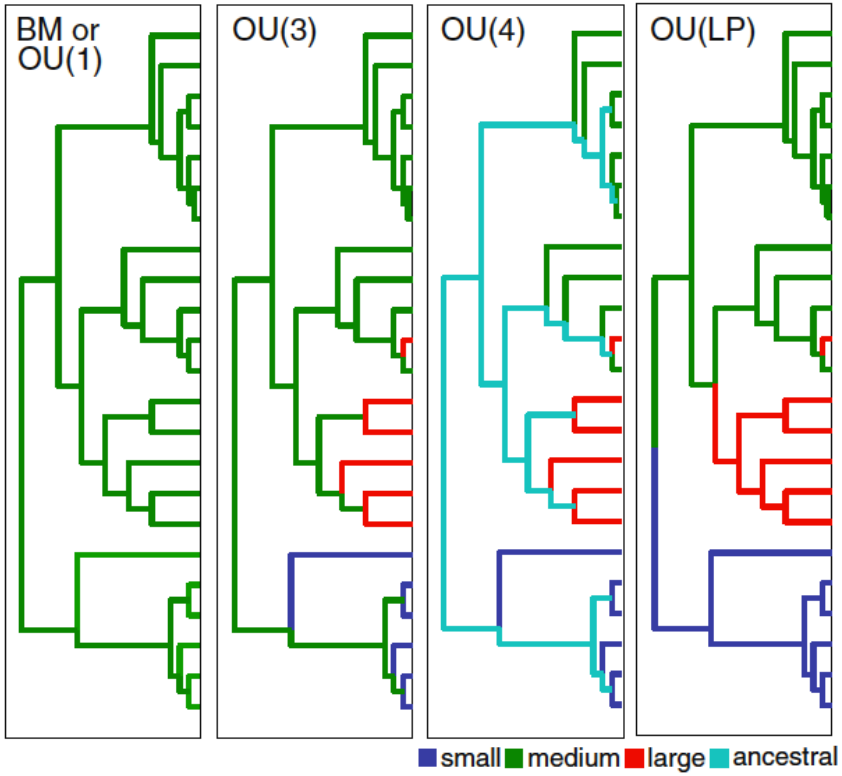
Candidate models	AIC	ΔAIC	Akaike weight
BM	-51.49	0.00	0.867
BM with κ = 0	-47.72	3.77	0.132
BM with trend	-37.59	13.90	0.001
EB	-32.95	18.54	0.000
OU	-32.06	19.43	0.000



Information theoretic approach

- More than one models are selected
- Δ values, model likelihoods, model weights and evidence ratios (instead of *P* values)
  - *hypothesis* H_4 *is 22 times more likely than* H_2
  - *the probability of* H_4 *is 0.78, while the probability of* H_2 *is 0.015*
  - *~~significant~~, ~~strong~~, ~~robust~~….*
- Model averaging
- Uncertainty is inherent to biological data

Evolution of body size in *Anolis* lizards



oculatus	22.3	Dominica
ferreus	28.4	Marie Galante
terraealtae	21.5	Illes de Saintes-2
terraealtae	21.3	Illes de Saintes-1
lividus	18.7	Monserrat
marmoratus	19.9	Desirade
marmoratus	18.9	Guadeloupe
nubilus	21.1	Redonda
sabanus	18.3	Saba
gingivinus	19.7	St. Barthelemy
gingivinus	19.6	St. Maarten
gingivinus	18.8	Anguilla
leachi-a	28.8	Antigua
leachi-b	28.6	Barbuda
bimaculatus	23.6	St. Eustatius
bimaculatus	27.9	St. Christopher
bimaculatus	27.1	Nevis
pogus	13.5	St. Maarten
wattsi	14.9	Antigua
wattsi	14.5	Barbuda
schwartzi	14.3	St. Eustatius
schwartzi	14.2	Nevis
schwartzi	14.3	St. Christopher

	BM	**OU(1)**	**OU(3)**	**OU(4)**	**OU(LP)**
$-2\ln\mathcal{L}$	−34.66	−34.66	−40.21	−47.22	−49.69
AIC	−30.66	−26.66	−28.21	−33.22	−37.69
LR		0	5.55	12.56	15.03
P		1	0.24	0.028	0.0046

Butler and King 2004

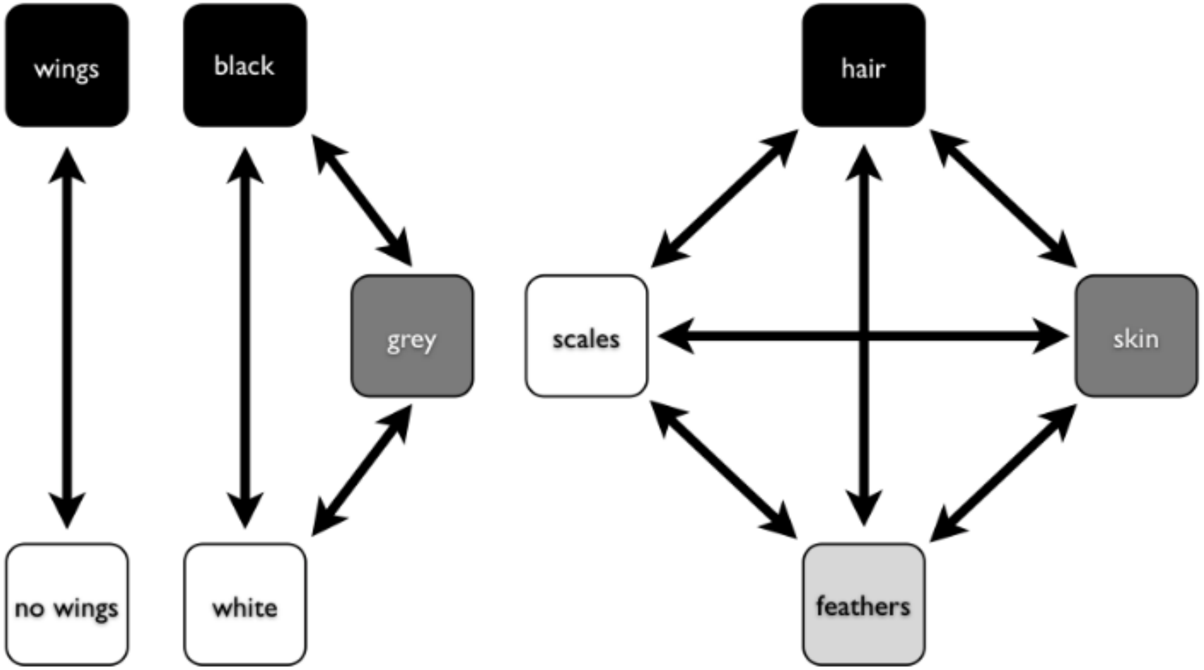
Evolution of discrete traits

Discrete trait models

- The *Mk* model
- The extended *Mk* model
- Models accommodating changes in the rate of evolution
- The threshold model

The *Mk* model

- Evolutionary changes between $k > 1$ states of a character
- Markov process: change depends on current state only
- Every state is equally likely



Harmon 2018: Phylogenetic Comparative Methods learning from trees

The *Mk* model

- Instantaneous rate of change parameter: q
  - number of changes of character over t when $t \sim 0$
  - $q_{12} = q_{21}, q_{12} = q_{13} \cdots$
- Transition matrix, **Q**

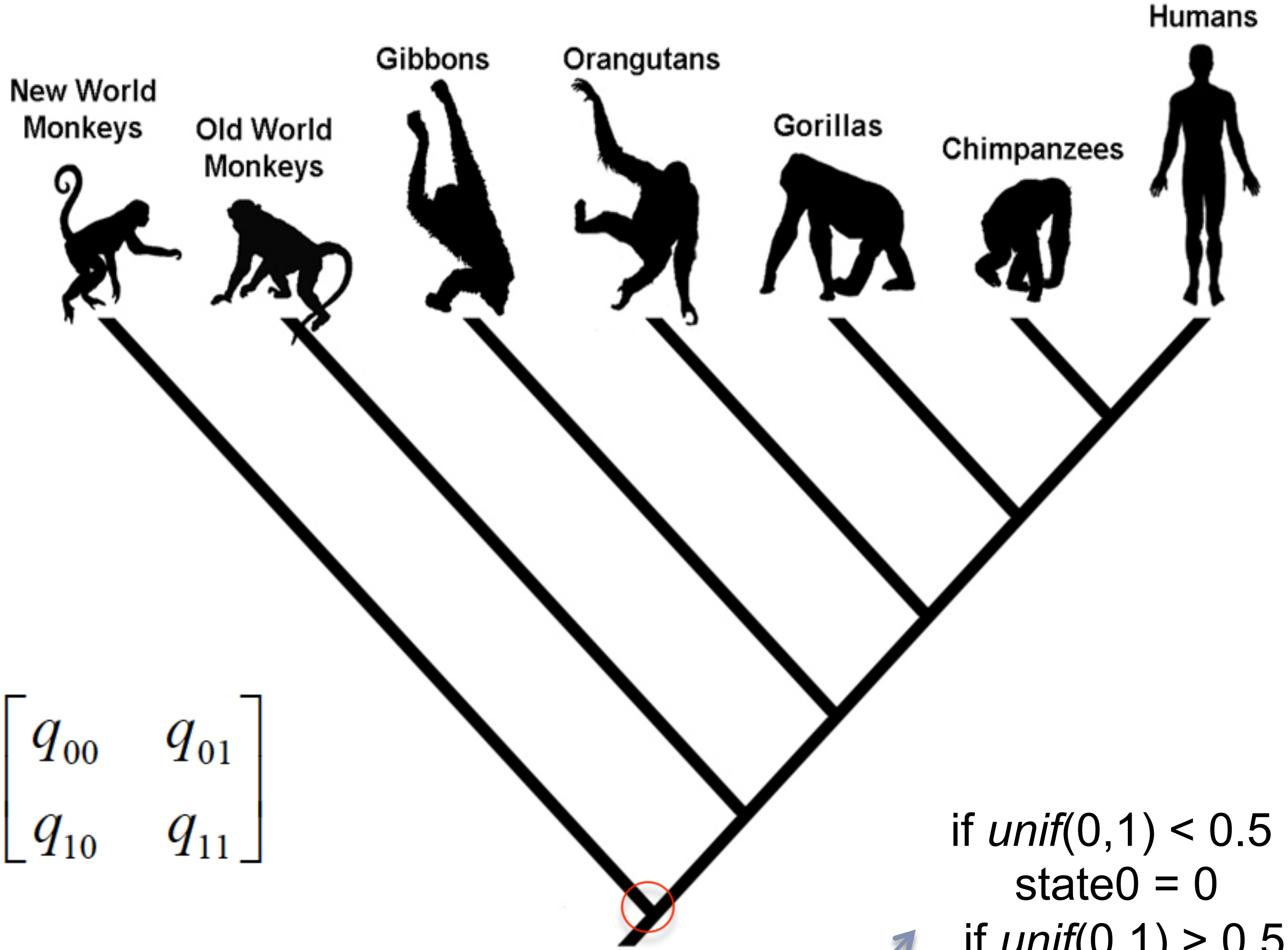
$$\mathbf{Q} = \begin{bmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} -d_1 & q_{12} & \cdots & q_{1k} \\ q_{21} & -d_2 & \cdots & q_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ q_{k1} & qk2 & \cdots & -d_k \end{bmatrix}$$

$$0 \rightarrow \quad d_1 = \sum_{i=2}^{k} q_{1i}$$

- Probability distributions of traits after t
  - $P(t) = e^{Qt}$

The *Mk* model

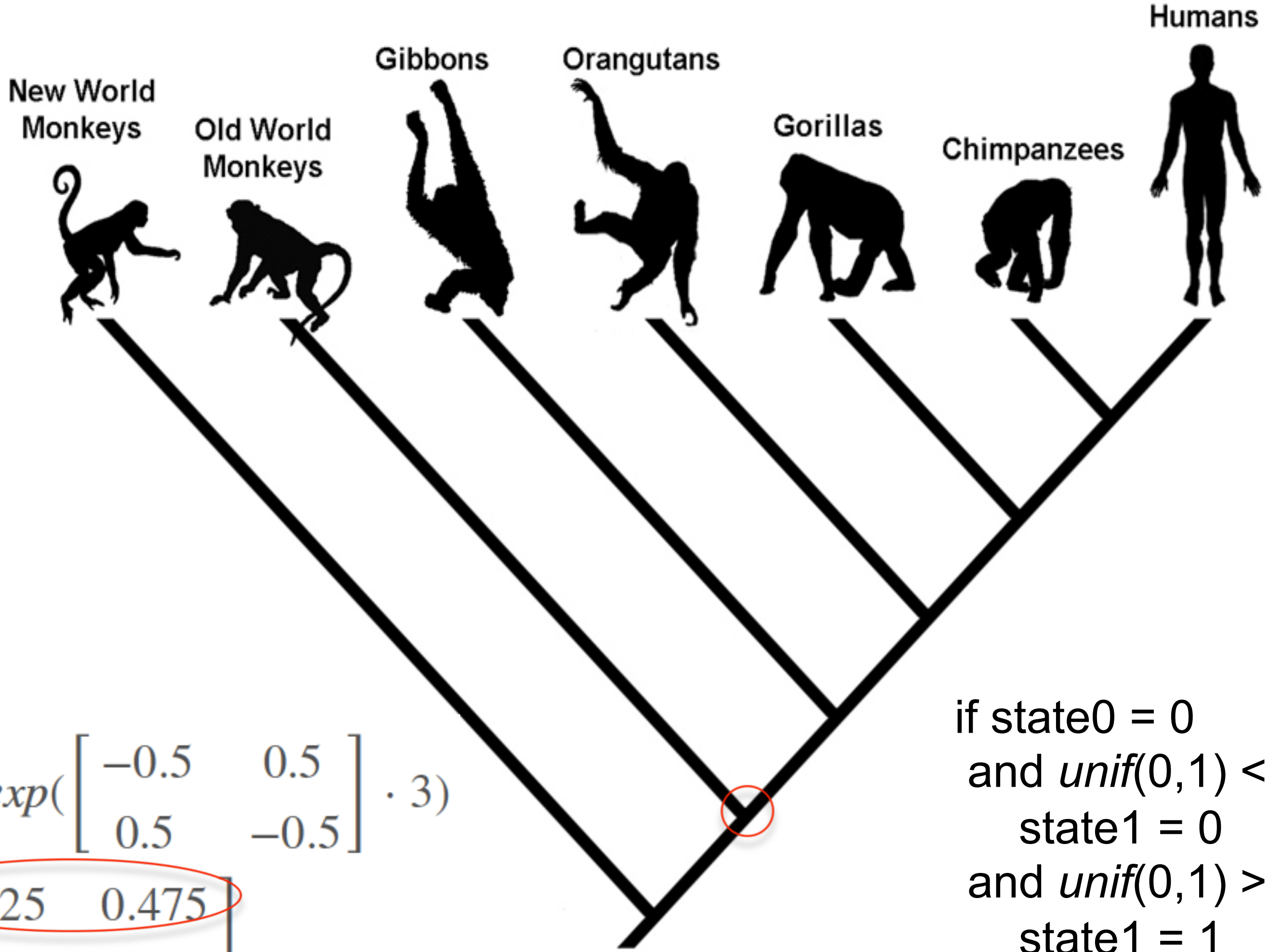


$$\mathbf{Q} = \begin{bmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{bmatrix}$$

if *unif*(0,1) < 0.5
state0 = 0
if *unif*(0,1) > 0.5
state0 = 1

$q_{12} = q_{21} \rightarrow p_{state0} = 0.5$

The *Mk* model

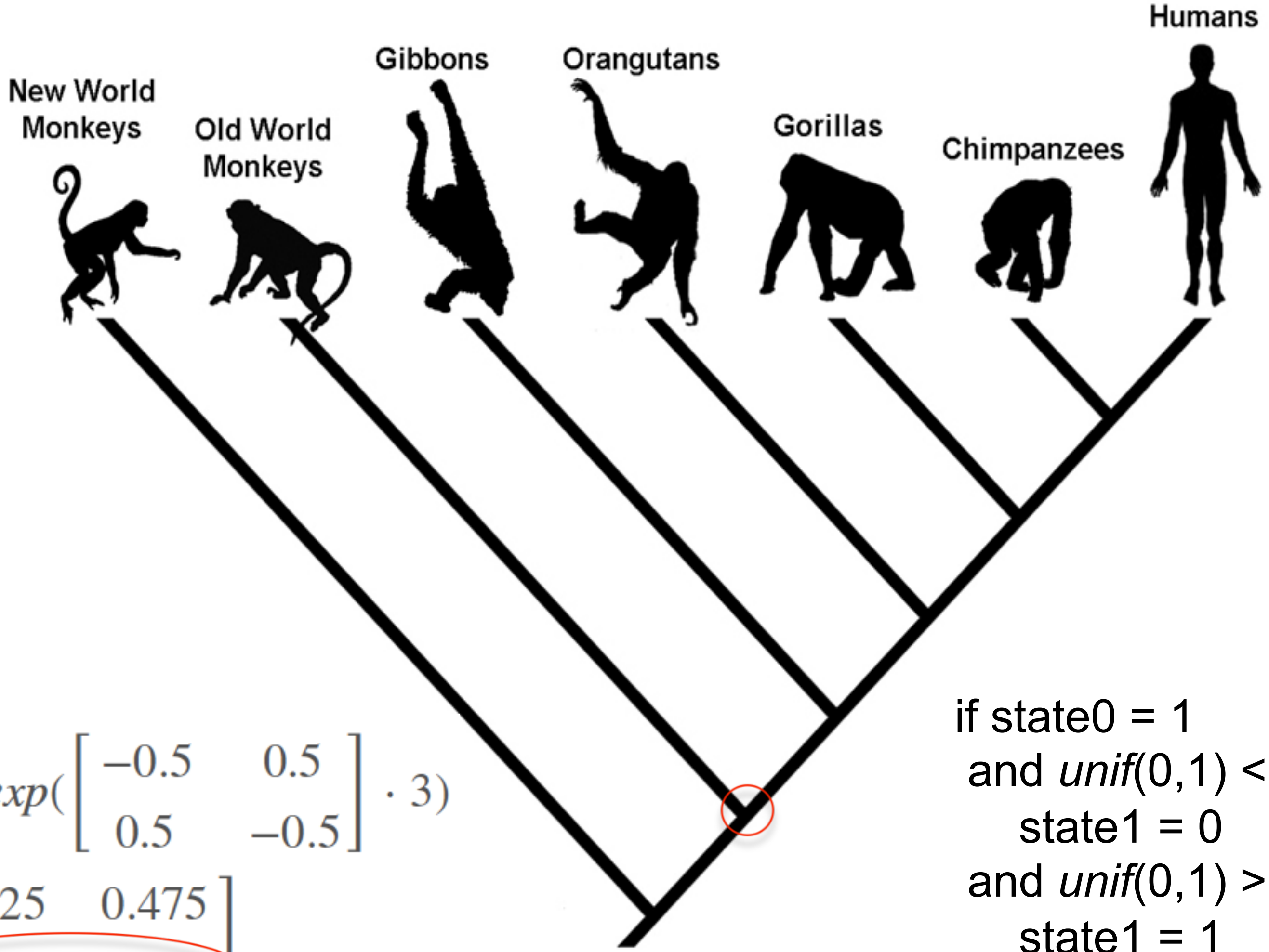


$$\mathbf{P}(t) = e^{\mathbf{Q}t} = exp(\begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix} \cdot 3)$$

$$= \begin{bmatrix} 0.525 & 0.475 \\ 0.475 & 0.525 \end{bmatrix}$$

if state0 = 0
 and *unif*(0,1) < 0.525
  state1 = 0
 and *unif*(0,1) > 0.525
  state1 = 1

The *Mk* model

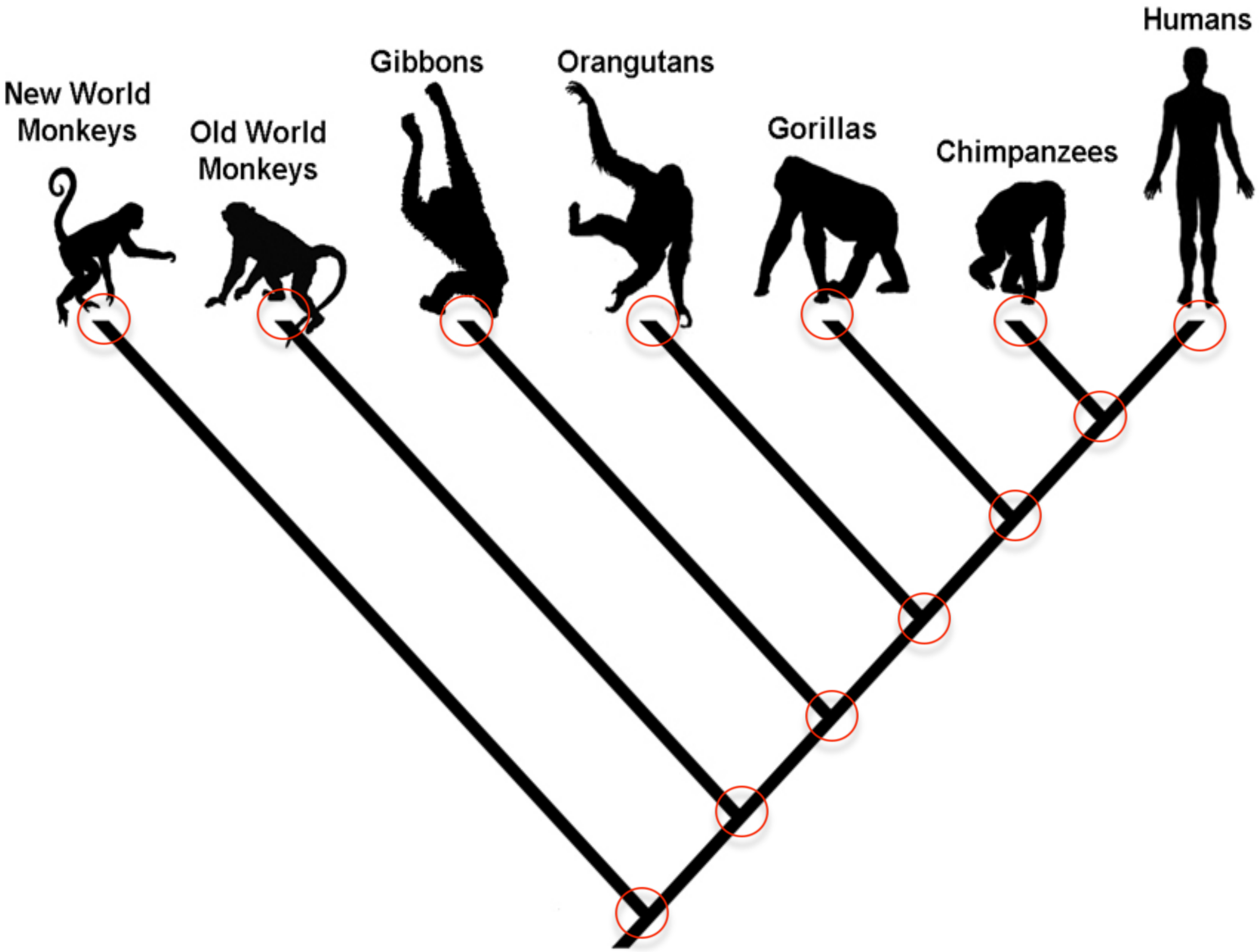


$$\mathbf{P}(t) = e^{\mathbf{Q}t} = exp(\begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix} \cdot 3)$$

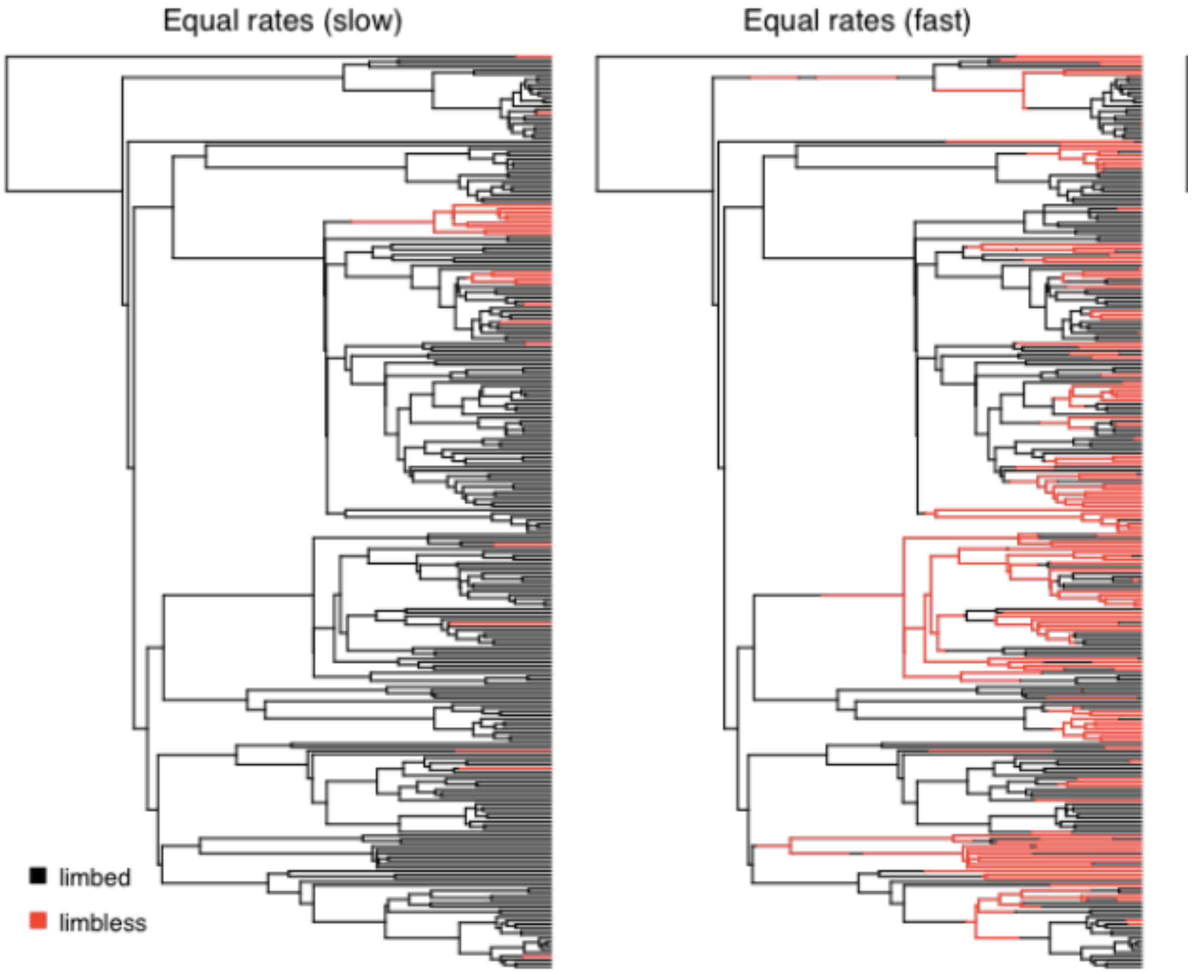
$$= \begin{bmatrix} 0.525 & 0.475 \\ 0.475 & 0.525 \end{bmatrix}$$

if state0 = 1
and *unif*(0,1) < 0.475
state1 = 0
and *unif*(0,1) > 0.475
state1 = 1

The *Mk* model


New World Monkeys
Old World Monkeys
Gibbons
Orangutans
Gorillas
Chimpanzees
Humans

The *Mk* model



Harmon 2018: Phylogenetic Comparative Methods learning from trees

Discrete trait models

- The *Mk* model
- **The extended *Mk* model**
- Models accommodating changes in the rate of evolution
- The threshold model

Discrete trait models

- The *Mk* model
- The extended *Mk* model
  - SYM
  - ARD
- Models accommodating changes in the rate of evolution
- The threshold model

The extended *Mk* model

- The *Mk* model assumes: $q_{12} = q_{21}$, $q_{12} = q_{13} \ldots$
  - Memoryless: a character that changes state from 0 -> 1 has an equal probability of reverting back
  - Homogeneous: same rate among all states

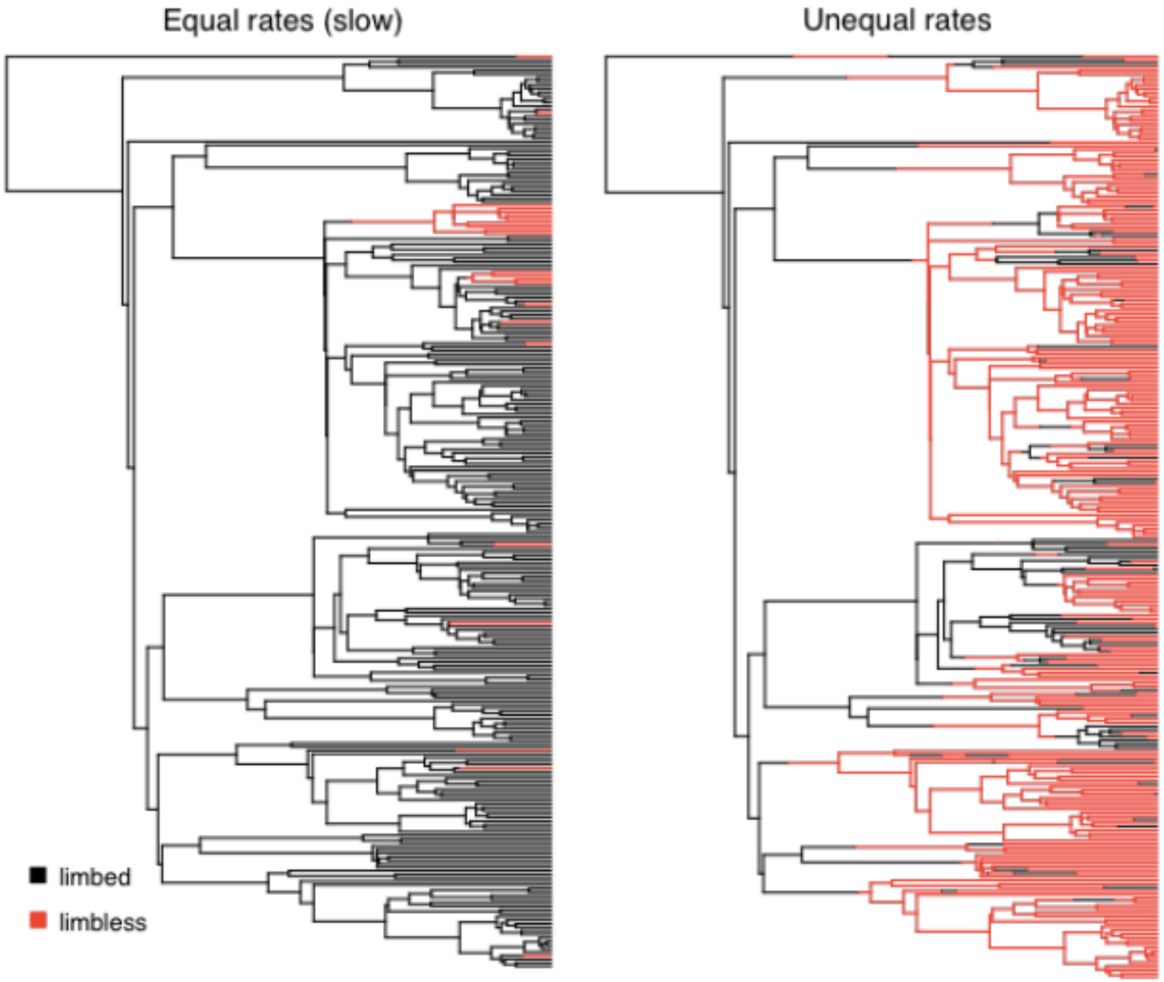
$$\begin{bmatrix} -q & q \\ q & -q \end{bmatrix}$$

- SYM only assumes: $q_{12} = q_{21}$
- ARD: all rates can be different

$$\begin{bmatrix} -q_1 & q_1 \\ q_2 & -q_2 \end{bmatrix}$$

- more parameters
- **Q** and **P**(t) matrices can be redefined
- can lead to different states at the nodes and the tips

The extended *Mk* model



Harmon 2018: Phylogenetic Comparative Methods learning from trees

Discrete trait models

- The *Mk* model
- The extended *Mk* model
- Models accommodating changes in the rate of evolution
- The threshold model

Discrete trait models

- The *Mk* model
- The extended *Mk* model
- Models accommodating changes in the rate of evolution
  - *Mk* is more suitable for sequence data (protein, DNA)
    - extensions exists (e.g. adding heterogeneity across sites)
    - incorporate characters evolving under a shared model
  - Morphological character evolution
    - shared model across characters is unjustified
    - each character require specific parameters
- The threshold model

Discrete trait models

- The *Mk* model
- The extended *Mk* model
- Models accommodating changes in the rate of evolution
  - Pagel's model
  - Other *Mk* models that allows parameters vary across clades and/or time
- The threshold model

Tree transformations: altering rate of evolution

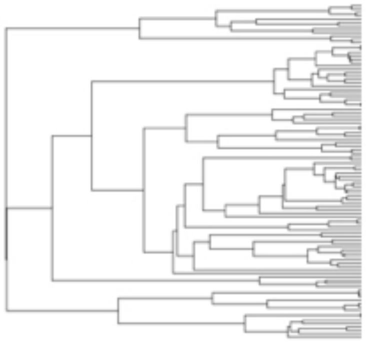
- Longer branches, higher rate
- Pagel's transformations
- Alteration of the **C** matrix

Harmon 2018: Phylogenetic Comparative Methods learning from trees

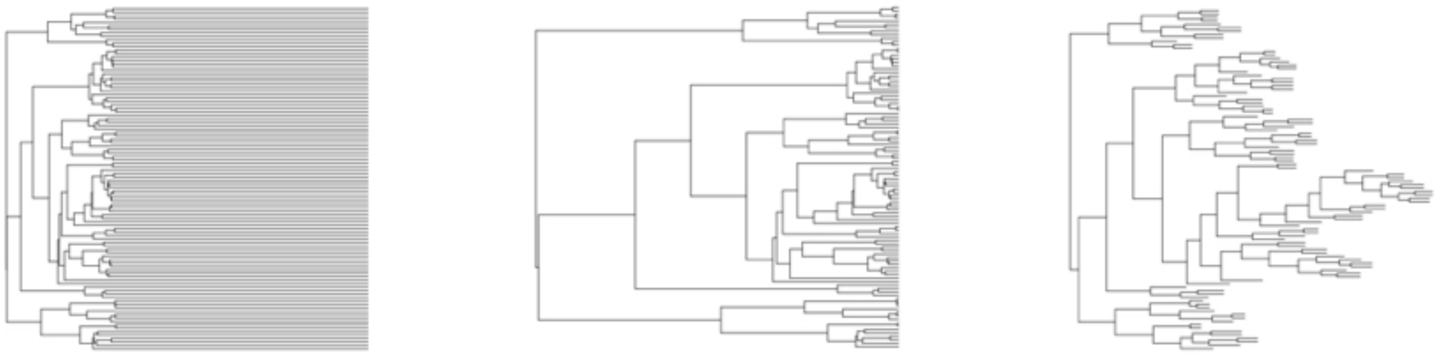
Tree transformations: altering rate of evolution

- Longer branches, higher rate
- Pagel's transformations
- Alter~~ation of the **C**~~ matrix

Starting tree



Tree Transformations

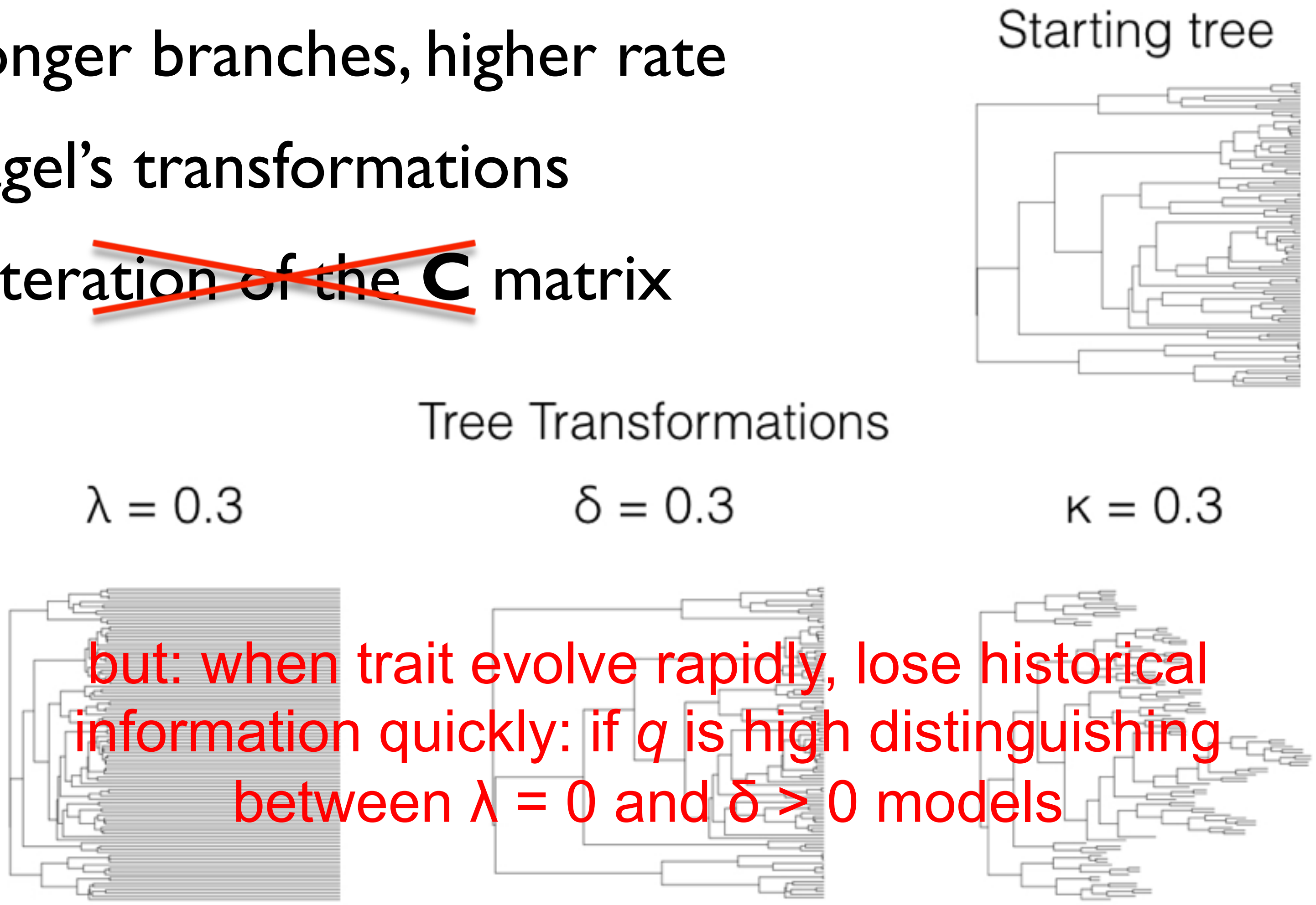


$\lambda = 0.3$ $\delta = 0.3$ $\kappa = 0.3$

Harmon 2018: Phylogenetic Comparative Methods learning from trees

Tree transformations: altering rate of evolution

- Longer branches, higher rate
- Pagel's transformations
- ~~Alteration of the **C**~~ matrix



but: when trait evolve rapidly, lose historical information quickly: if *q* is high distinguishing between λ = 0 and δ > 0 models

Harmon 2018: Phylogenetic Comparative Methods learning from trees

Discrete trait models

- The *Mk* model
- The extended *Mk* model
- Models accommodating changes in the rate of evolution
  - Pagel's model
  - Other *Mk* models that allows parameters vary across clades and/or time
    - rate of evolution varies between clades (multi-rate discrete models)
    - different **Q** matrix for different branches
    - rate parameters in **Q** varies with time

Discrete trait models

- The *Mk* model
- The extended *Mk* model
- Models accommodating changes in the rate of evolution
- The threshold model

Discrete trait models

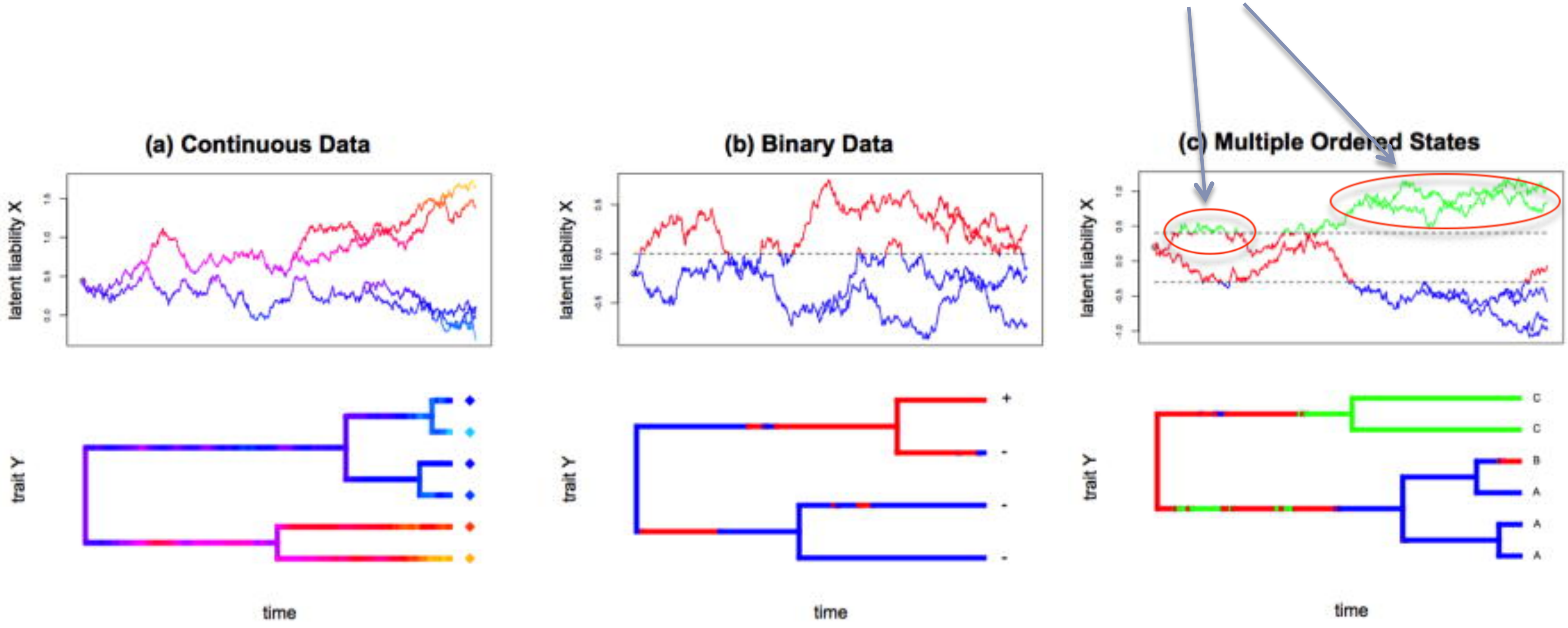
- The *Mk* model
- The extended *Mk* model
- Models accommodating changes in the rate of evolution
- The threshold model
  - the effective rate of change depends on the amount of time that a lineage has been in that state (while *Mk* is memoryless)
  - more realistic for some biological characters
  - allows variation in in transition rates without more parameters

The threshold model

- liability: the value of the discrete phenotype is determined by a latent continuous trait, if it crosses a fixed threshold value, the character changes state (Wright 1934)
  - unobserved, unmeasured (e.g. hormone) with multivarate normal distribution
  - can follow a BM (or OU) motion model of evolution (Felsenstein 2005, 2012)
  - proxy for the complex, multilocus genetic changes that are likely to underlie a shift in a discretely measured ecological trait (Revell 2013).

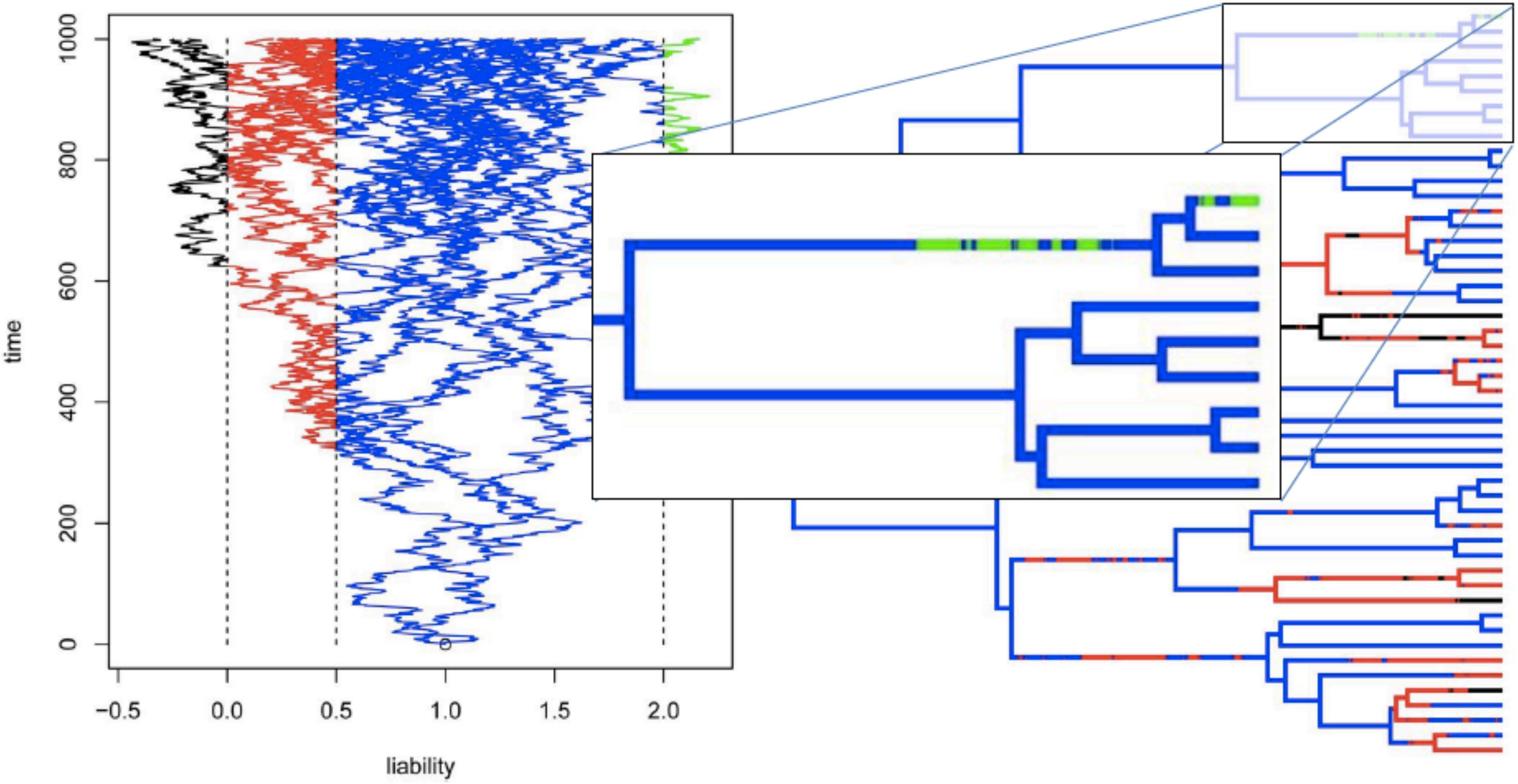
The threshold model

if a character changed state recently from *A* -> *C*, it is much more likely to change back immediately (when near the threshold) than far in the future.



Cybis et al 2015

The threshold model

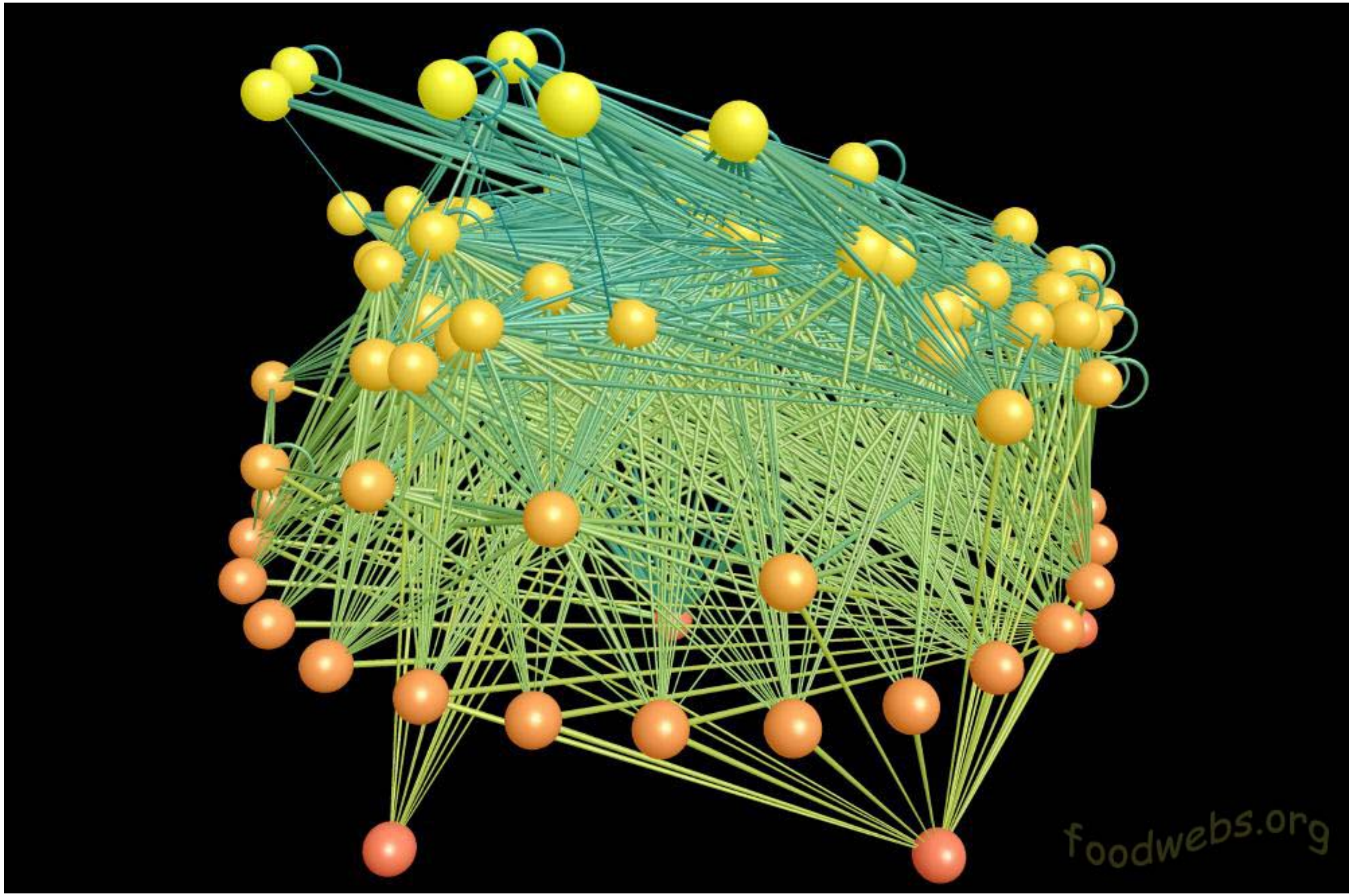


Revell 2013

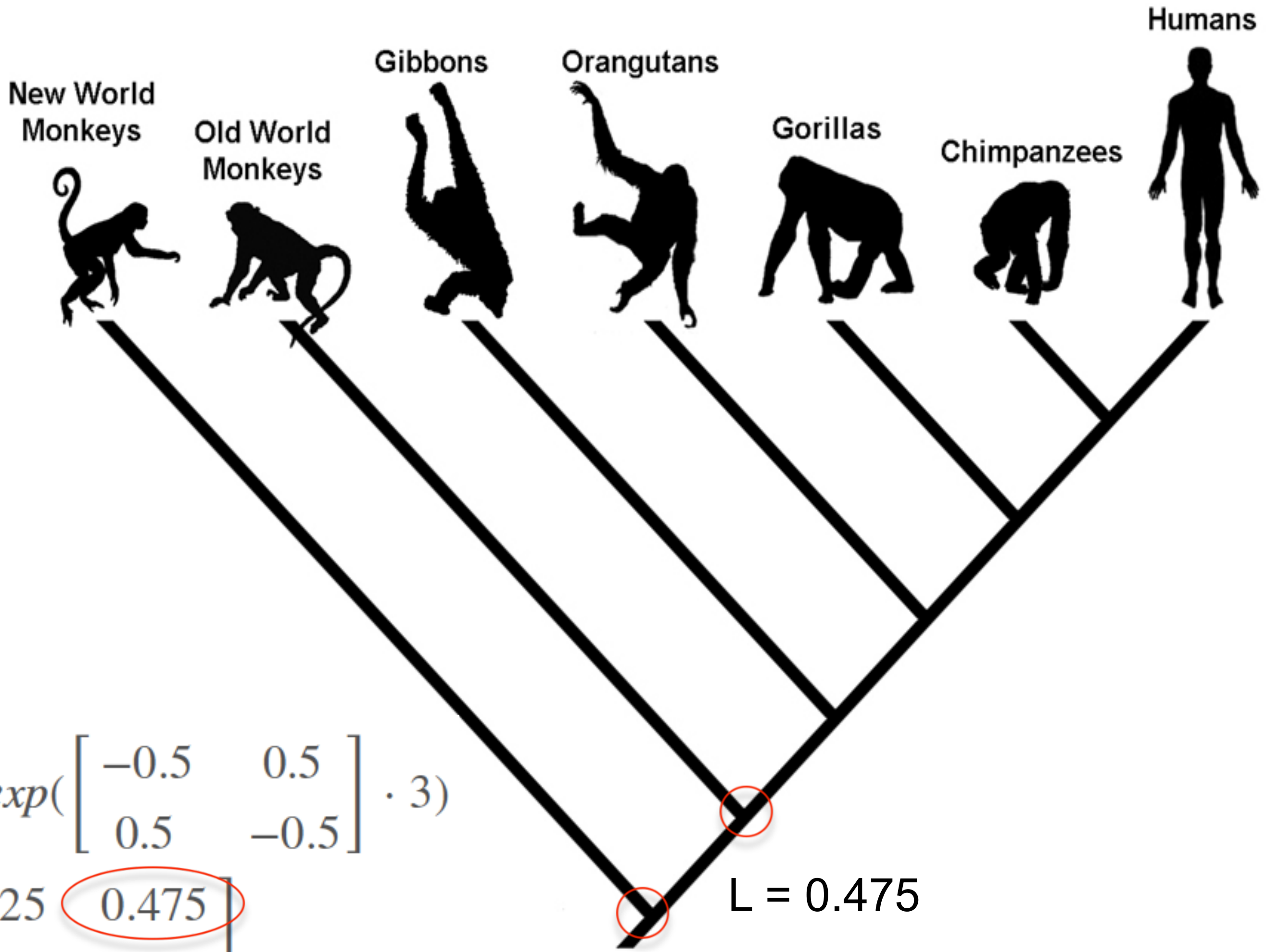
Discrete trait models

- The *Mk* model
- The extended *Mk* model
  - SYM
  - ARD
- Models accommodating changes in the rate of evolution
  - Pagel's model
  - Other *Mk* models that allows parameters vary across clades and/or time
- The threshold model

How to choose from so many models?



What is the likelihood for a change 0->1?

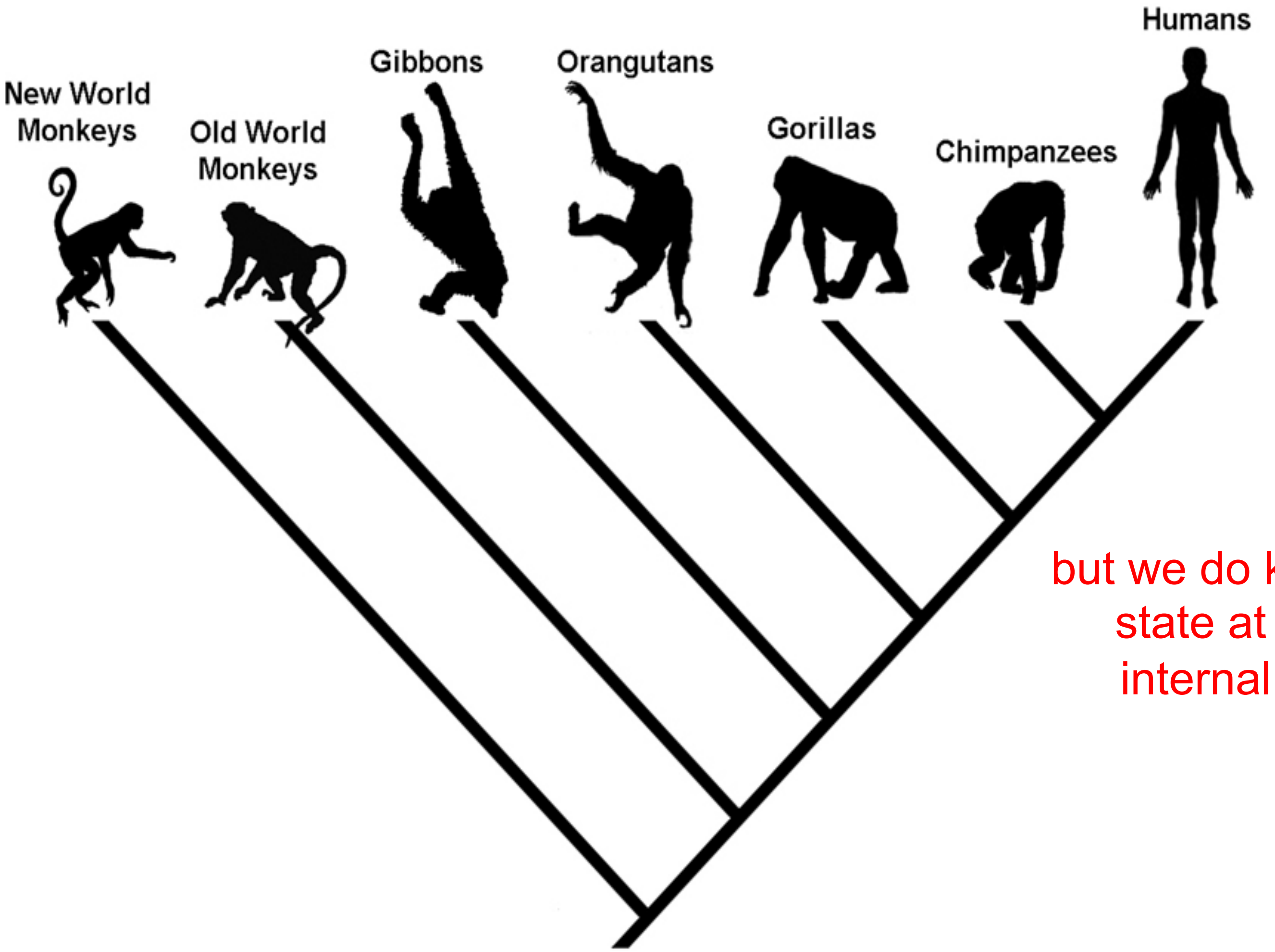


L = 0.475

$$\mathbf{P}(t) = e^{\mathbf{Q}t} = exp(\begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix} \cdot 3)$$

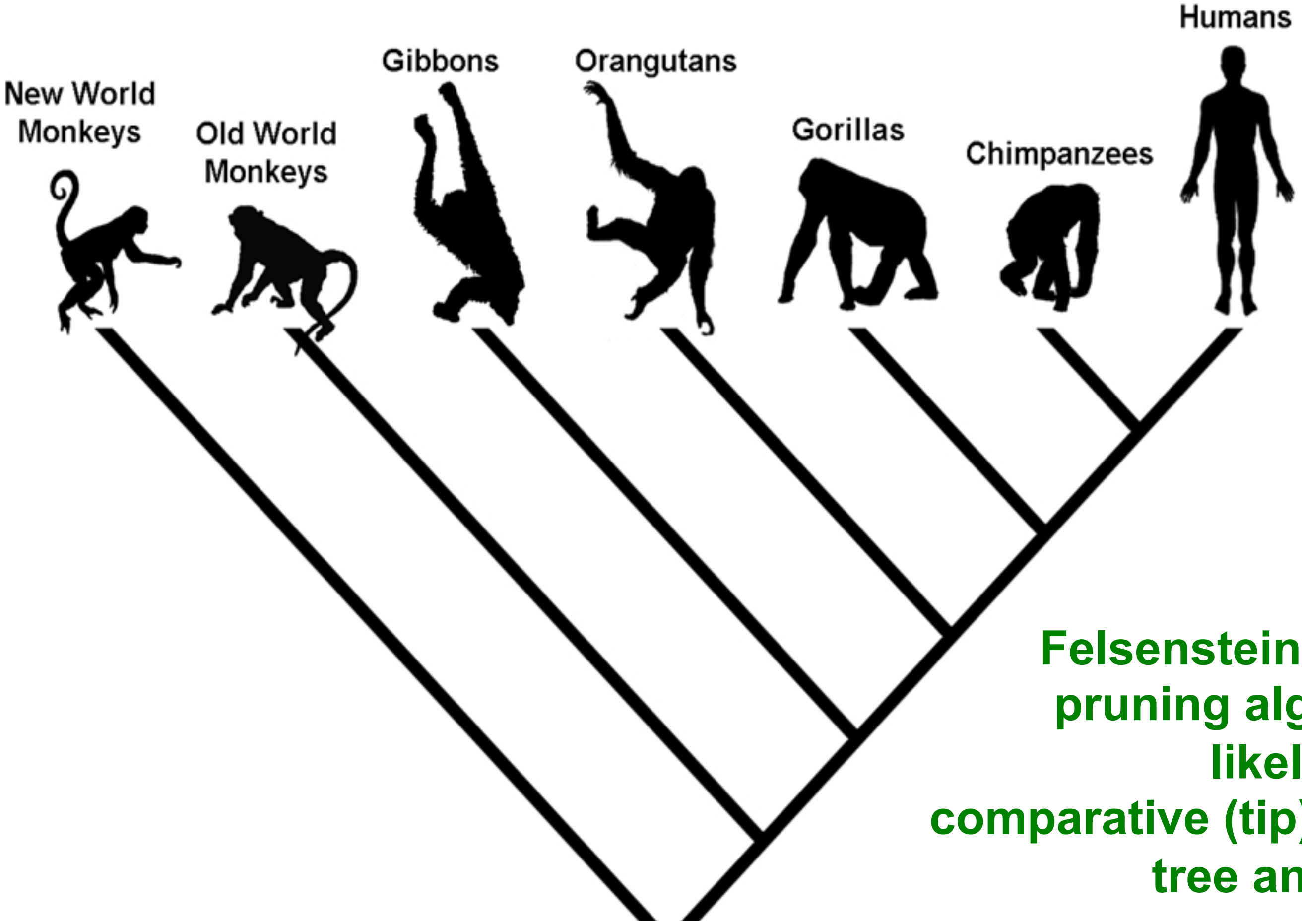
$$= \begin{bmatrix} 0.525 & \boxed{0.475} \\ 0.475 & 0.525 \end{bmatrix}$$

What is the likelihood for observed tip data given the tree and the model?



but we do know the state at root and internal braches

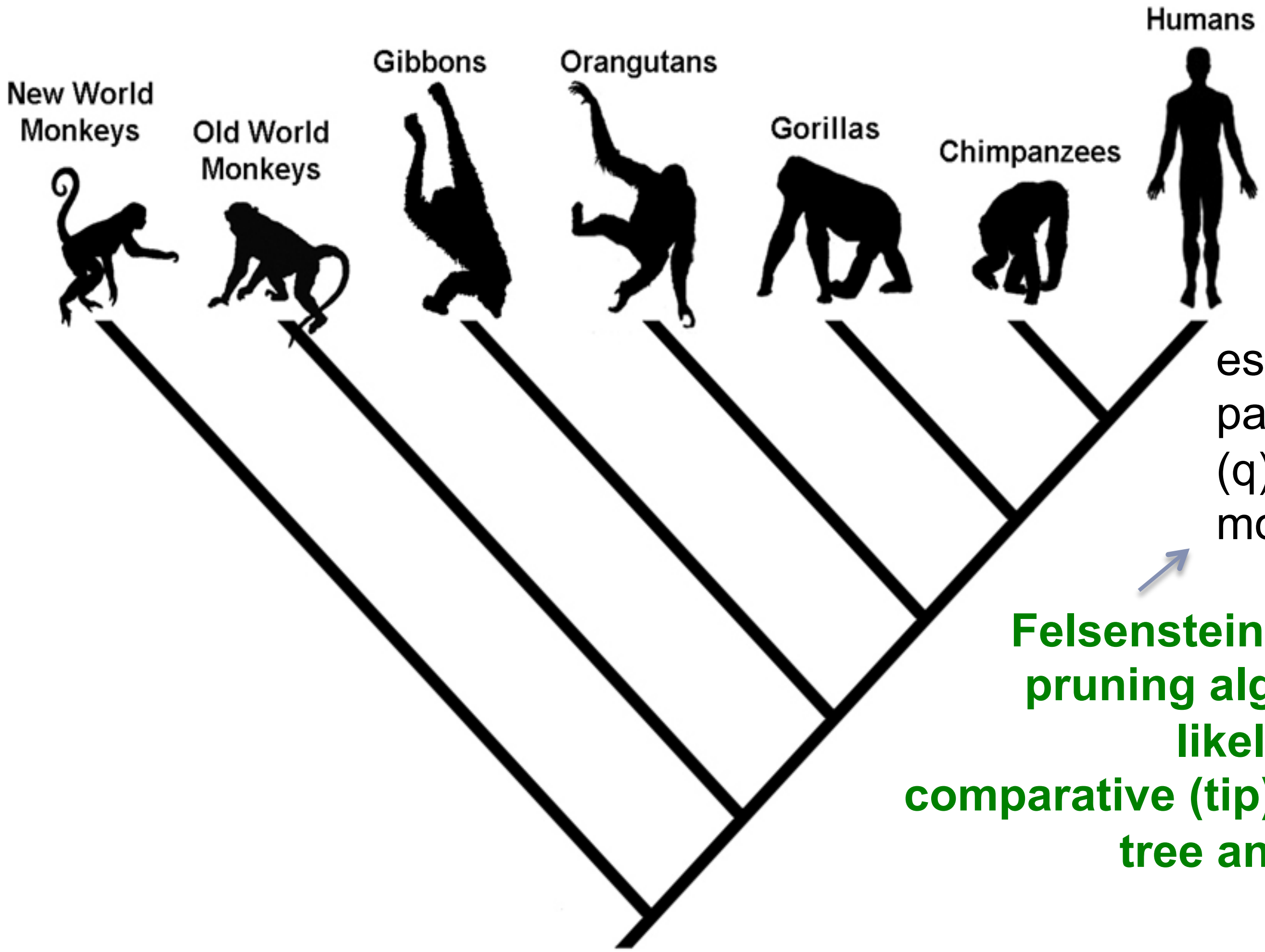
What is the likelihood for observed tip data given the tree and the model?



Felsenstein's (1973) pruning algorithm : likelihood of comparative (tip) given a tree and model

$$L_x = \sum_{S_x \in \Omega} P(S_x) \left(\sum_{S_y \in \Omega} P(S_y | S_x, t_{xy}) L_y \sum_{S_z \in \Omega} P(S_z | S_x, t_{xz}) L_z \right)$$

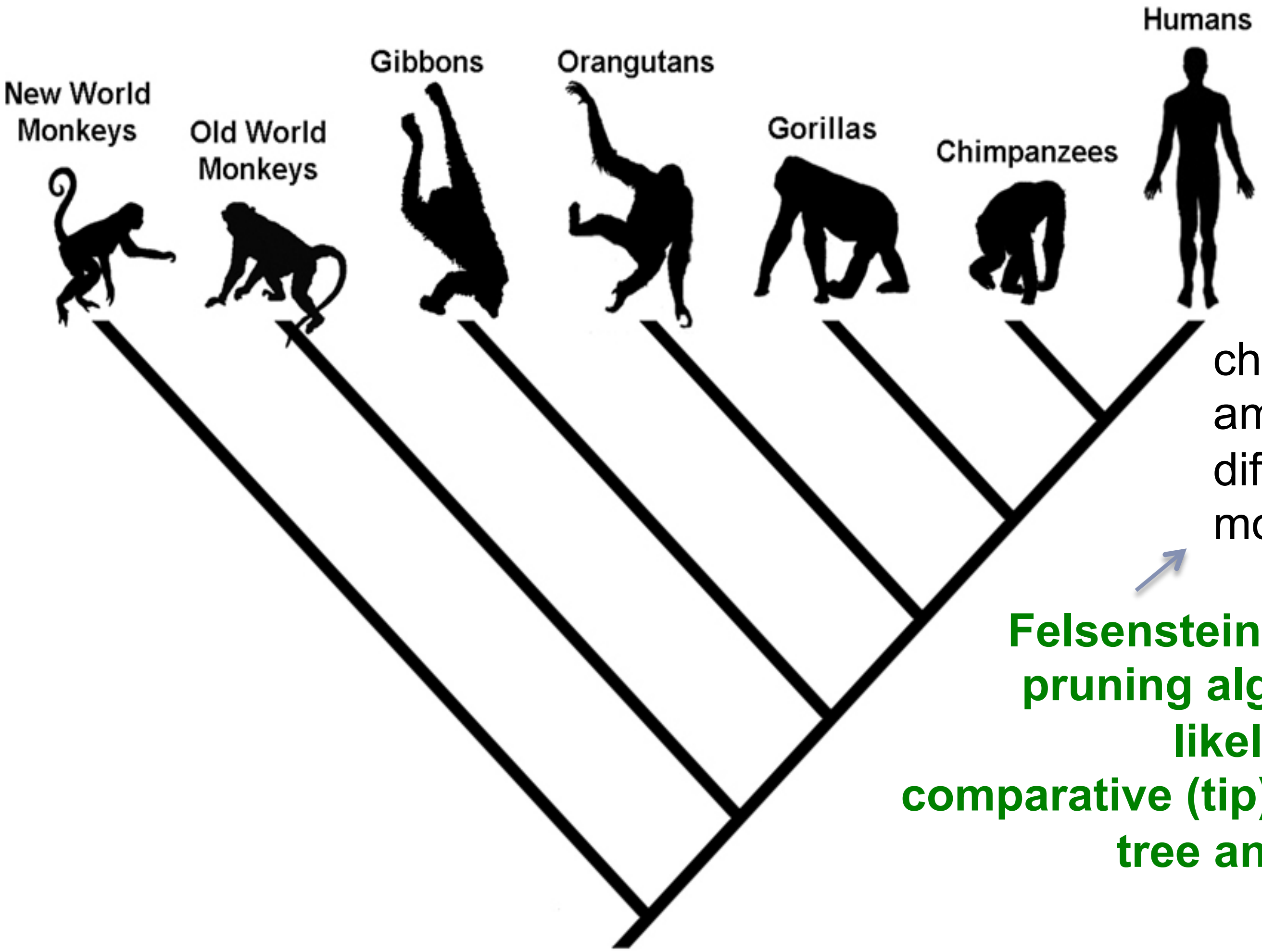
What is the likelihood for observed tip data given the tree and the model?



estimate parameters (q) of the Mk model

Felsenstein's (1973) pruning algorithm : likelihood of comparative (tip) given a tree and model

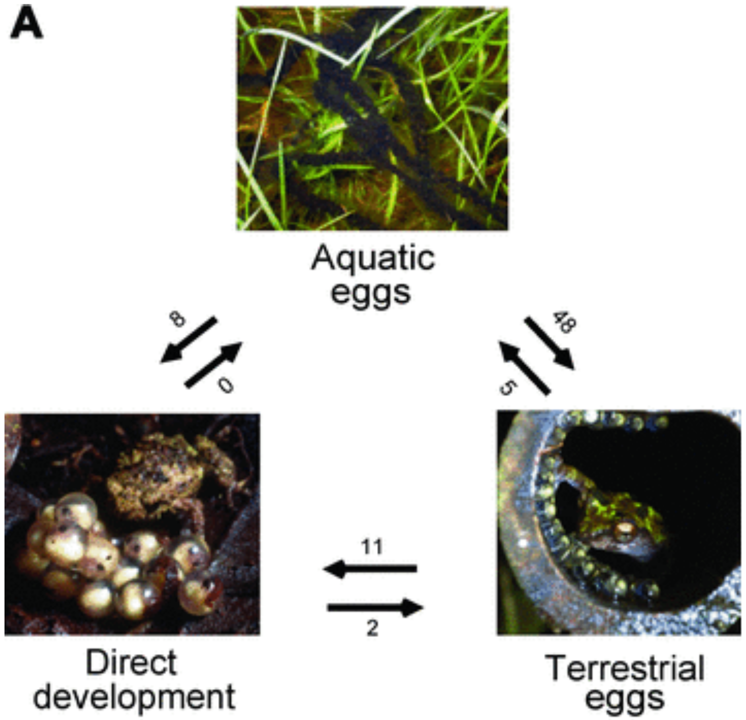
What is the likelihood for observed tip data given the tree and the model?



choosing among different models

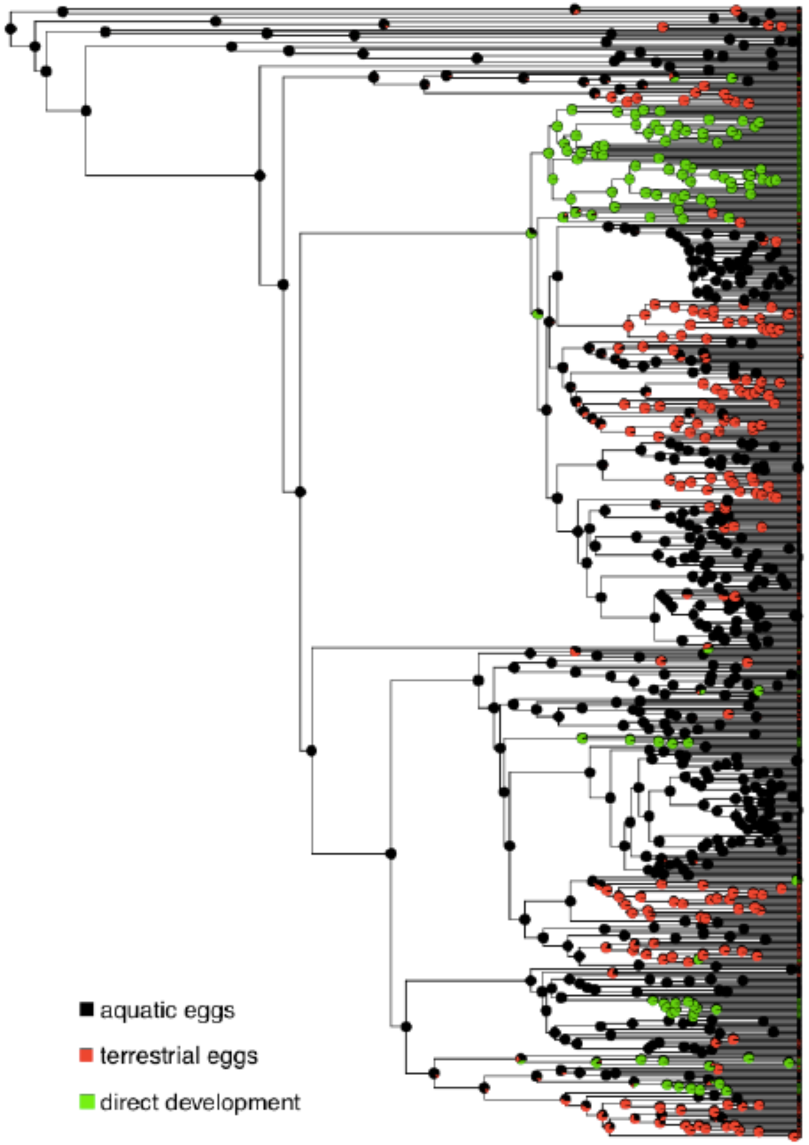
Felsenstein's (1973) pruning algorithm : likelihood of comparative (tip) given a tree and model

Evolution of reproductive modes in frogs

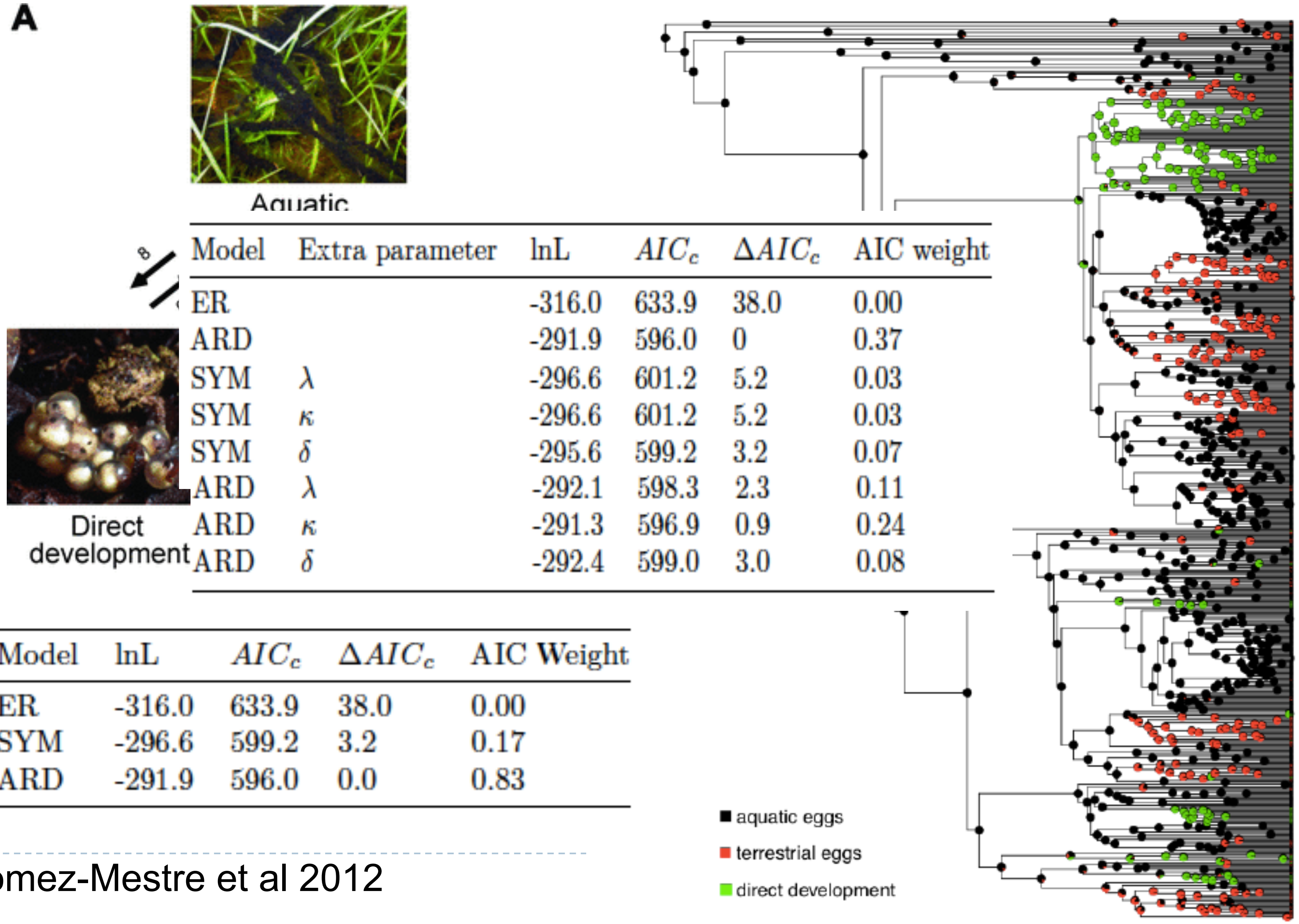


Model	lnL	AIC_c	ΔAIC_c	AIC Weight
ER	-316.0	633.9	38.0	0.00
SYM	-296.6	599.2	3.2	0.17
ARD	-291.9	596.0	0.0	0.83

- Gomez-Mestre et al 2012



Evolution of reproductive modes in frogs



Model	Extra parameter	lnL	AIC_c	ΔAIC_c	AIC weight
ER		-316.0	633.9	38.0	0.00
ARD		-291.9	596.0	0	0.37
SYM	λ	-296.6	601.2	5.2	0.03
SYM	κ	-296.6	601.2	5.2	0.03
SYM	δ	-295.6	599.2	3.2	0.07
ARD	λ	-292.1	598.3	2.3	0.11
ARD	κ	-291.3	596.9	0.9	0.24
ARD	δ	-292.4	599.0	3.0	0.08

Model	lnL	AIC_c	ΔAIC_c	AIC Weight
ER	-316.0	633.9	38.0	0.00
SYM	-296.6	599.2	3.2	0.17
ARD	-291.9	596.0	0.0	0.83

- Gomez-Mestre et al 2012

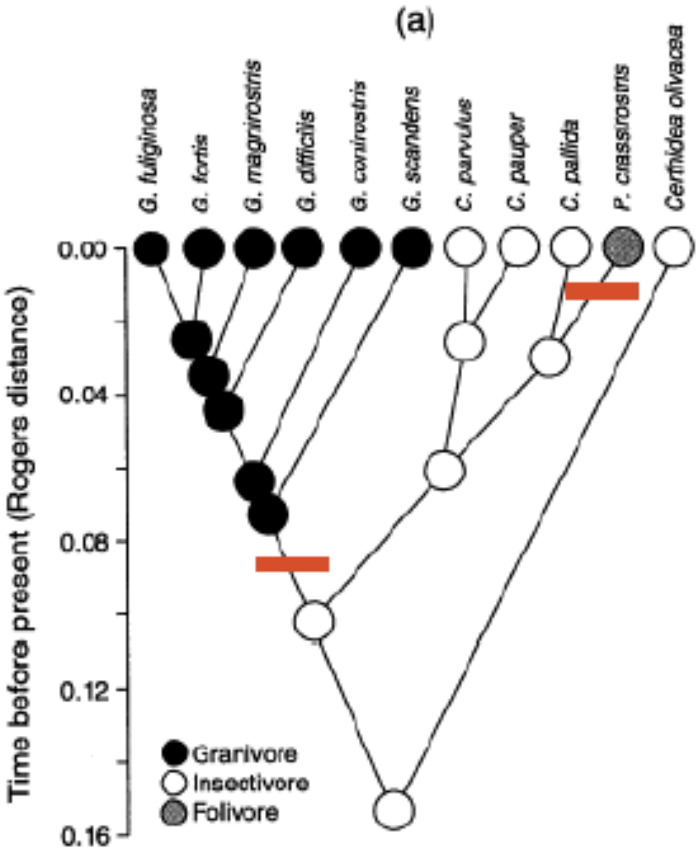
Ancestral state reconstructions

Ancestral state estimation

- Given
  - the tree
  - the character (continuous vs. discrete)
  - the model of evolution
- one can provide estimates for character states at the nodes or along the branches of the phylogeny
- these are associated with uncertainty
- different approaches exists
  - provide very nice graphs, but hard to check if they are true
  - just to name some of them…

Parsimony

> “All things being equal, the simplest solution tends to be the best one.”
>
> **William of Ockham**



Maximum Parsimony

● Granivore

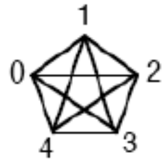
○ Insectivore

◍ Folivore

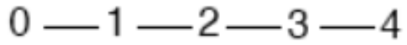
2 Changes

Parsimony

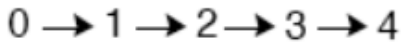
- Unordered (Fitch)



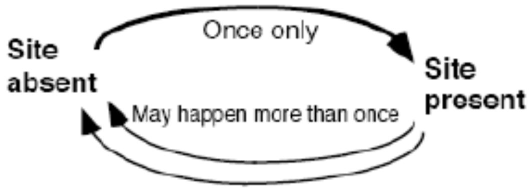
- Ordered (Wagner)



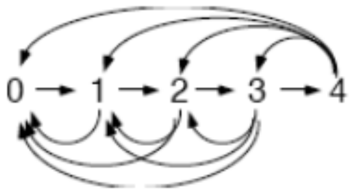
- Irreversible (Camin-Sokal)



- Dollo



- Step matrix

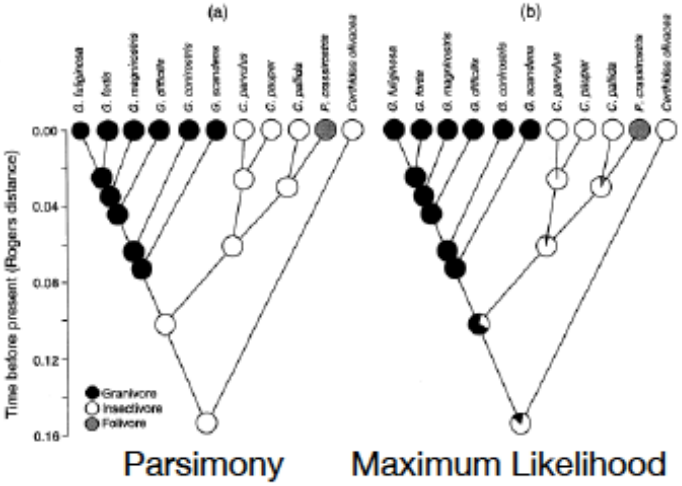


Parsimony

- Limitations:
  - Does not care about branch lengths (one change per branch regardless of how long)
  - Performs poorly with rapidly evolving traits, favors divergence toward the tips of the tree
    - the parsimony reconstruction will only accurately reflect the evolutionary process for our character when **Q** is very small
  - Does not provide errors, and does not say anything about less supported models

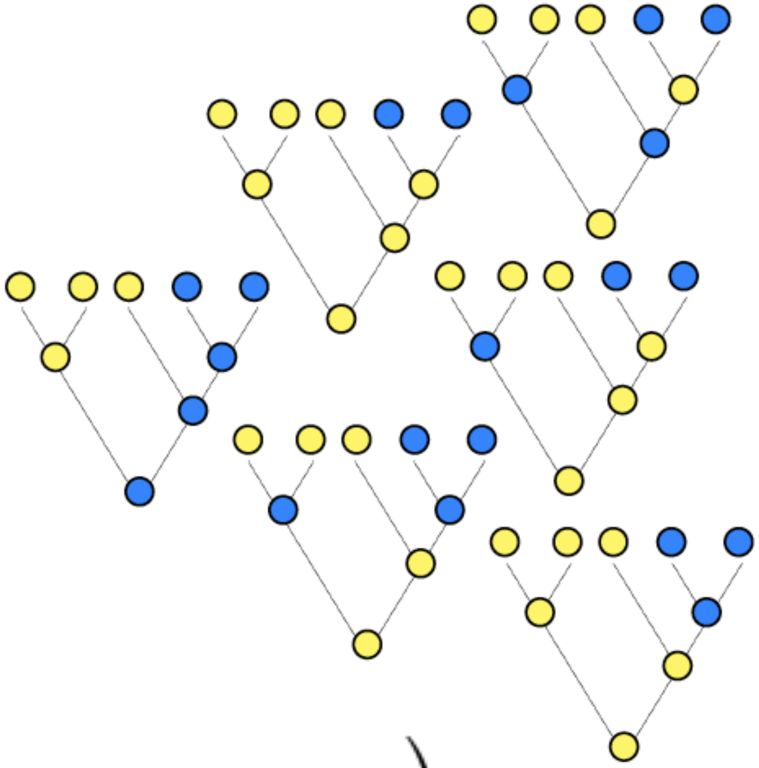
Maximum likelihood

- Uses the *Mk* model for evolution of discrete traits
- Uses maximum likelihood
  - to estimate rates
  - to reconstruct ancestral states in a form of probability
- Incorporate branch lengths
- Works well with fast rates
- Confidence/error around estimates
- It has its own limitations
  - requires a model
  - local optima problem for non-convex surfaces



Felsenstein’s (1973) pruning algorithm

nested sum of transition probabilities along the hierarchical structure of the tree



likelihood of subtree rooted at node x

$$L_x = \sum_{S_x\in\Omega} P(S_x)\left(\sum_{S_y\in\Omega} P(S_y|S_x,t_{xy})L_y \sum_{S_z\in\Omega} P(S_z|S_x,t_{xz})L_z\right)$$

character state at node x

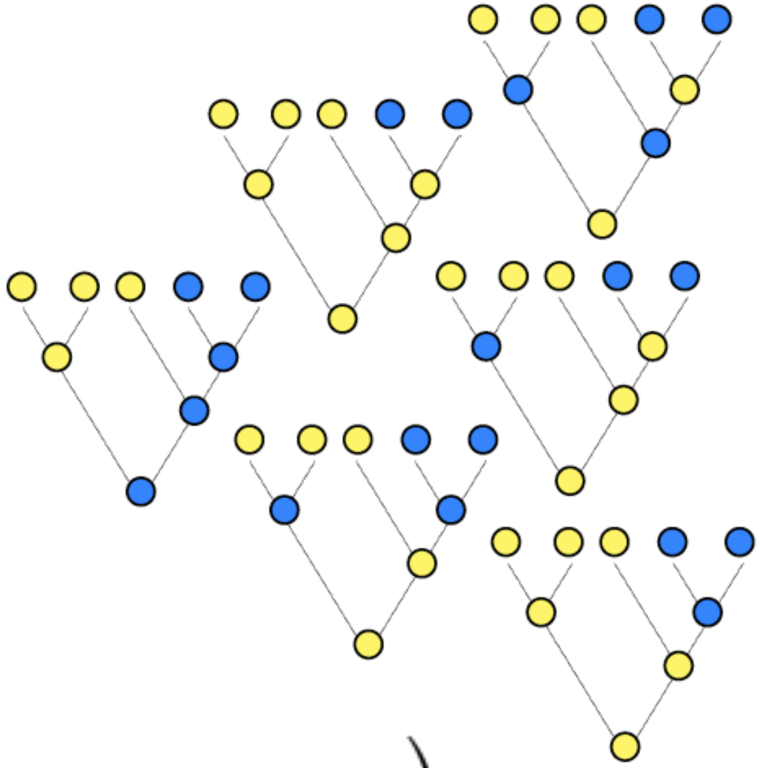
branch lenght

possible character states

Felsenstein’s (1973) pruning algorithm

nested sum of transition probabilities along the hierarchical structure of the tree

find the assignment to S_x for all x internal nodes that maximizes the likelihood of the observed data for a given tree.



likelihood of subtree rooted at node x

$$L_x = \sum_{S_x \in \Omega} P(S_x) \left(\sum_{S_y \in \Omega} P(S_y | S_x, t_{xy}) L_y \sum_{S_z \in \Omega} P(S_z | S_x, t_{xz}) L_z \right)$$

character state at node x

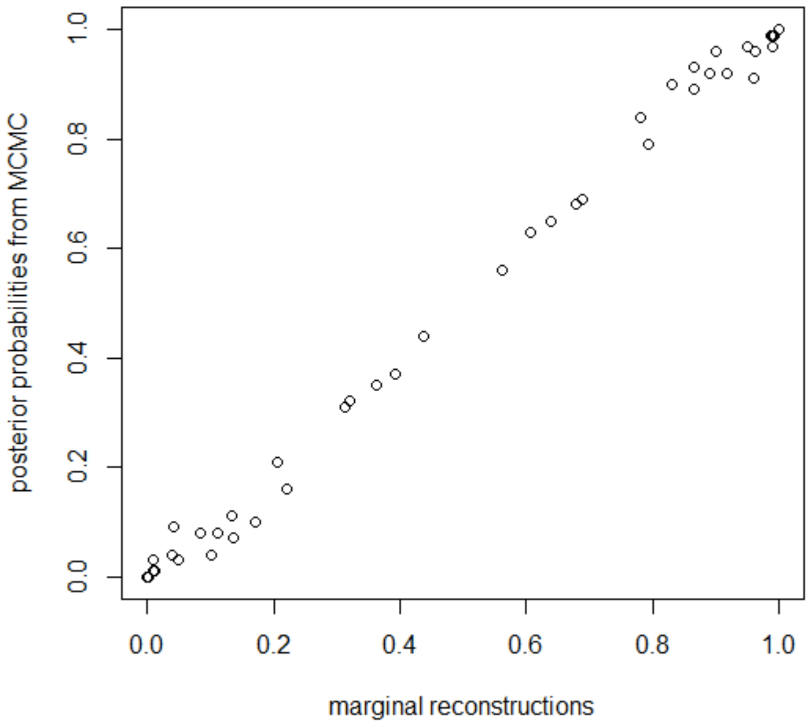
branch lenght

possible character states

Maximum likelihood

- Joint reconstruction: finding the set of character states at all nodes that (jointly) maximize the likelihood

- Marginal reconstruction: finding the state at the current node that maximizes the likelihood independently of the reconstruction of all other ancestral states



http://blog.phytools.org/2015/05/about-how-acemarginaltrue-does-not.html

Stochastic character mapping

- Sampling character histories in direct proportion to their posterior probability under a model
  - sample a transition matrix Q
  - sample ancestral states
  - simulate character histories along all the edges of the tree conditioned on **Q** and node states

Bollback 2006

Stochastic character mapping

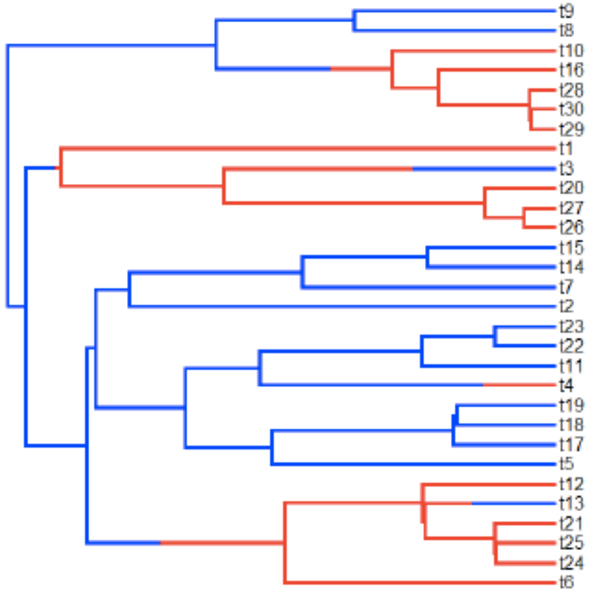
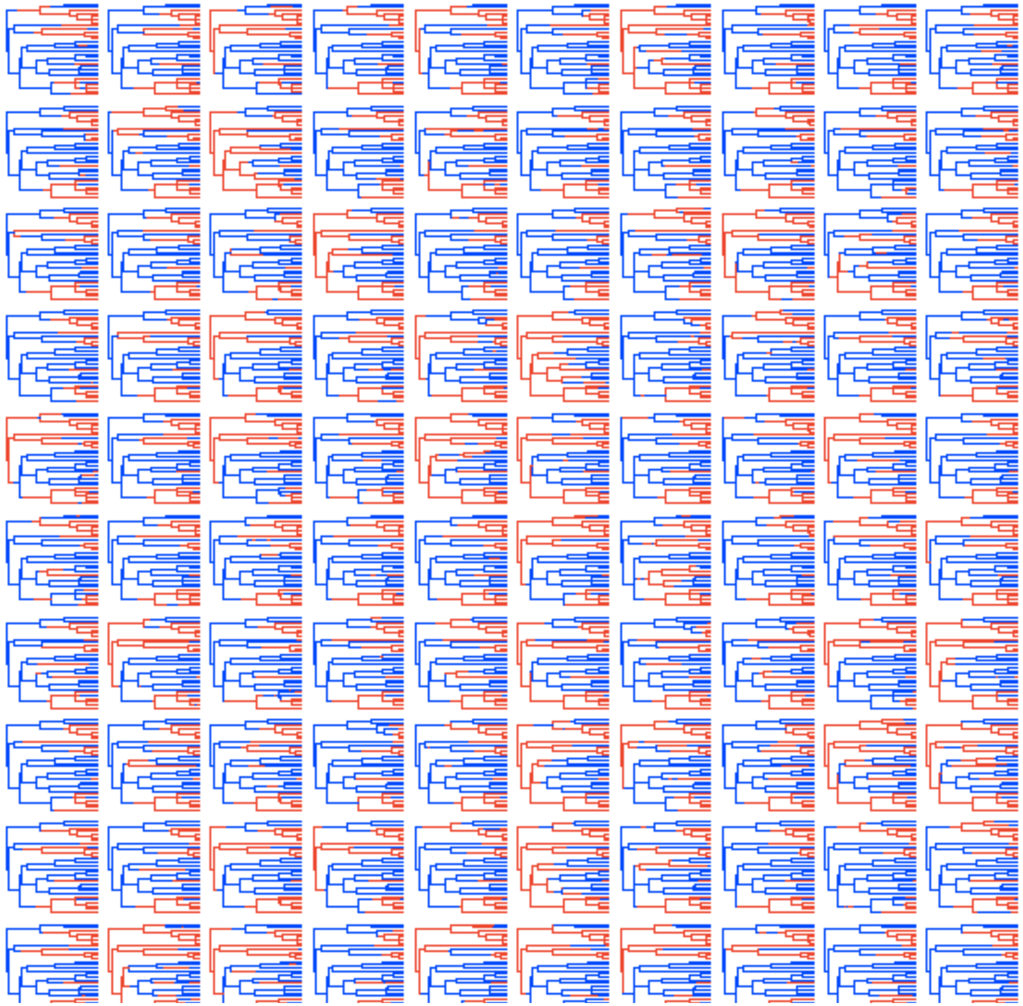


Figure. True history (above) & sample of stochastic character maps from the empirical Bayes posterior distribution (right).



Revell 2012

Stochastic character mapping

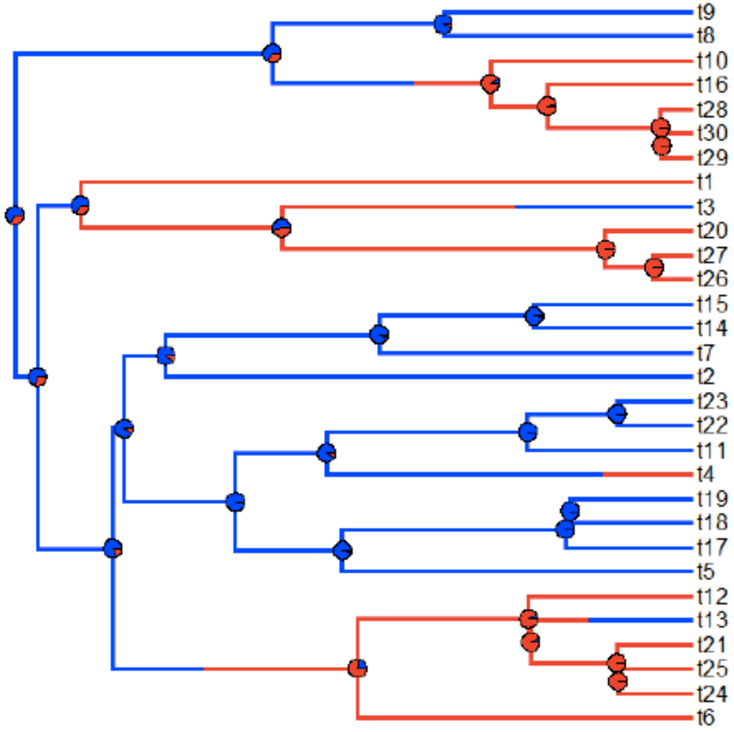


Figure. True history with posterior probabilities from stochastic mapping.

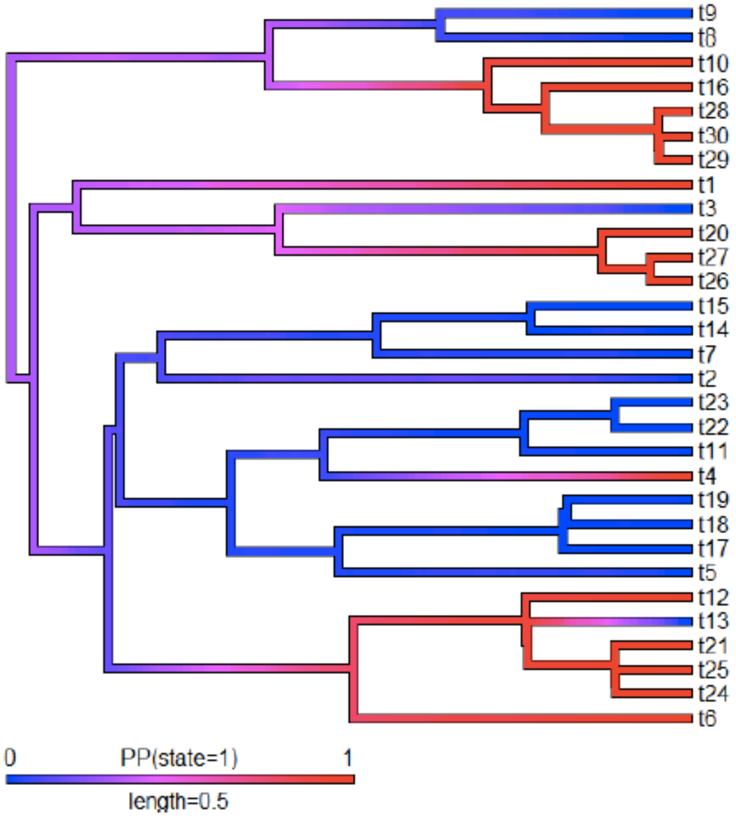


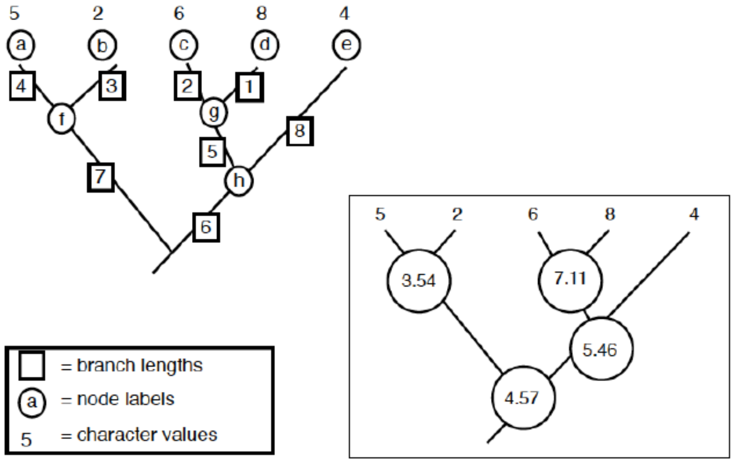
Figure. Posterior density map from stochastic mapping.

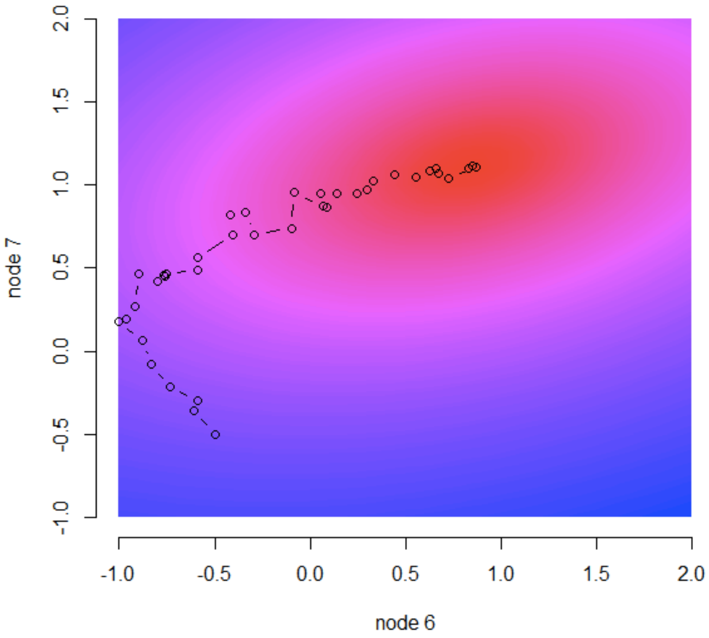
Revell 2012

Ancestral state estimation of continuous characters

- Maximum likelihood: need to find set of ancestral states that maximize the probability of data & tree

$$L(\mathbf{a},a_0,\sigma^2 \mid \mathbf{T},\mathbf{x}) = \frac{\exp[-\tfrac{1}{2}([\mathbf{x},\mathbf{a}]-a_0\mathbf{1})'(\sigma^2\mathbf{T})^{-1}([\mathbf{x},\mathbf{a}]-a_0\mathbf{1})]}{\sqrt{(2\pi)^{n+m-1}\left|\sigma^2\mathbf{T}\right|}}$$

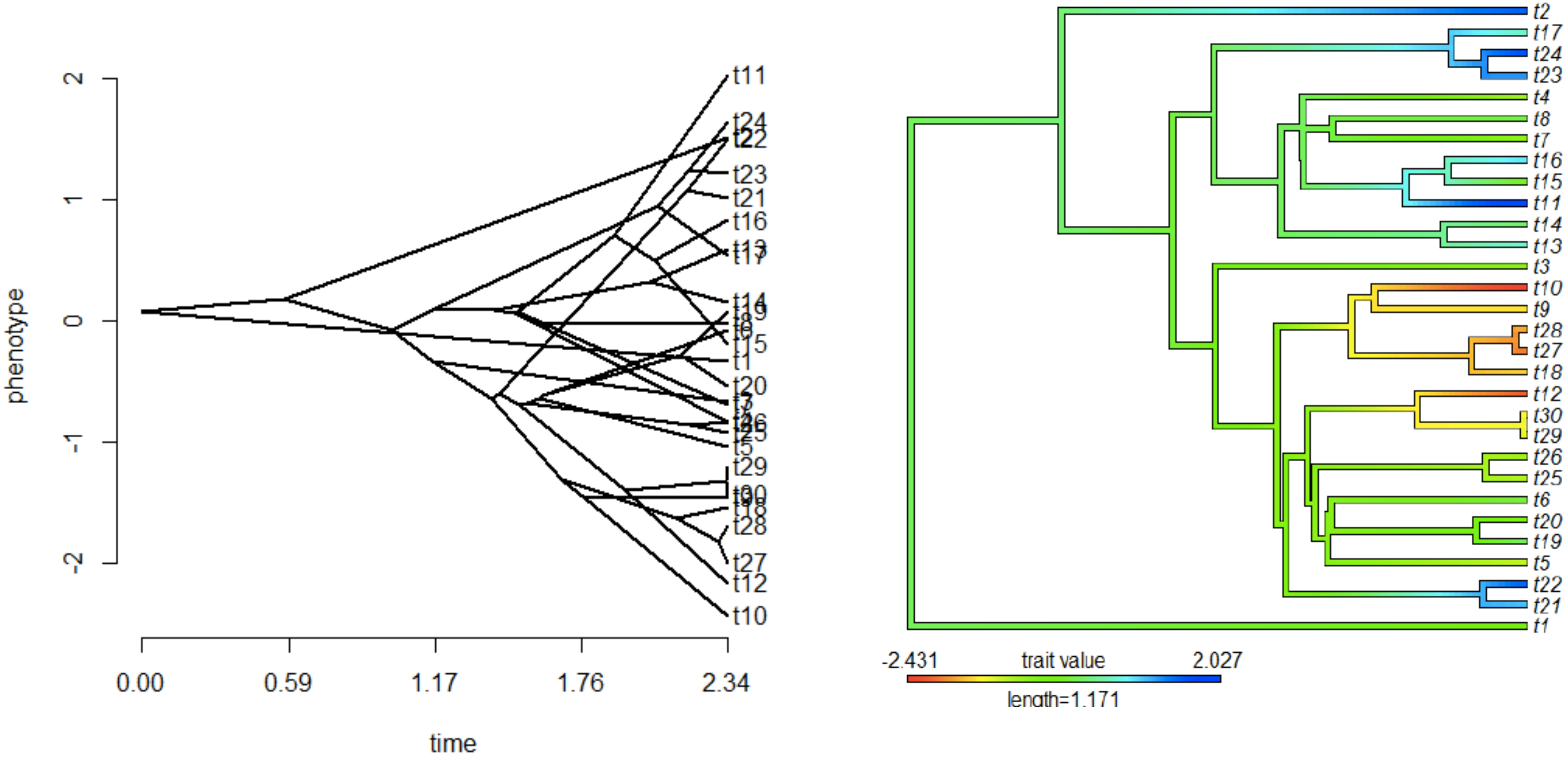




- http://www.phytools.org/eqg2015/asr.html

Revell 2012

Ancestral state estimation of continuous characters

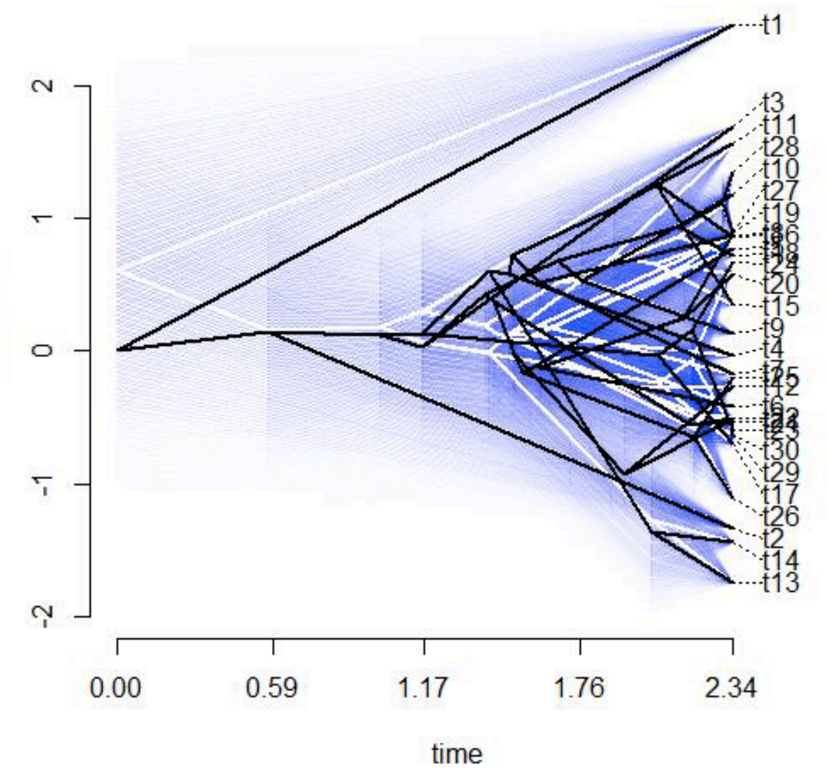


http://www.phytools.org/eqg2015/asr.html

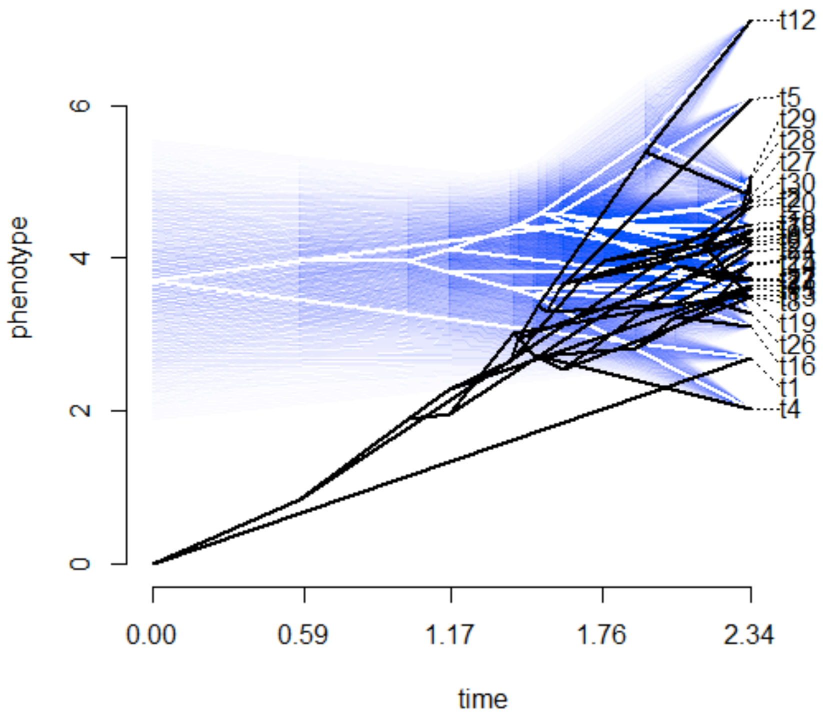
Revell 2012

Ancestral state estimation of continuous characters

Uncertainty



Bias



http://www.phytools.org/eqg2015/asr.html

Revell 2012