Aula 2

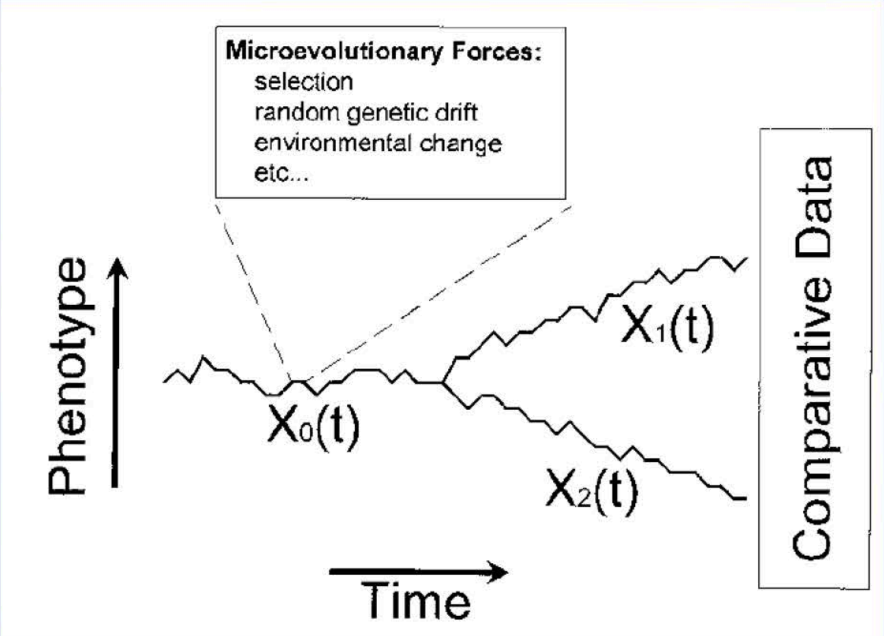
Modelos de evolução fenotípica de caracteres contínuos e categóricos

PGLS

Prática

- Introdução ao R
  - Importação e manipulação de dados
  - Importação e manipulação de ≠ formatos de filogenias
- Timetree.org
- Phyndr
- Phylotargeting
- rotl
- Rphylopars

EVOLUTIONARY MODELS



Mechanisms (selection, drift, mutations…)

Evolutionary models

Interspecific data

The analytical core of comparative analysis

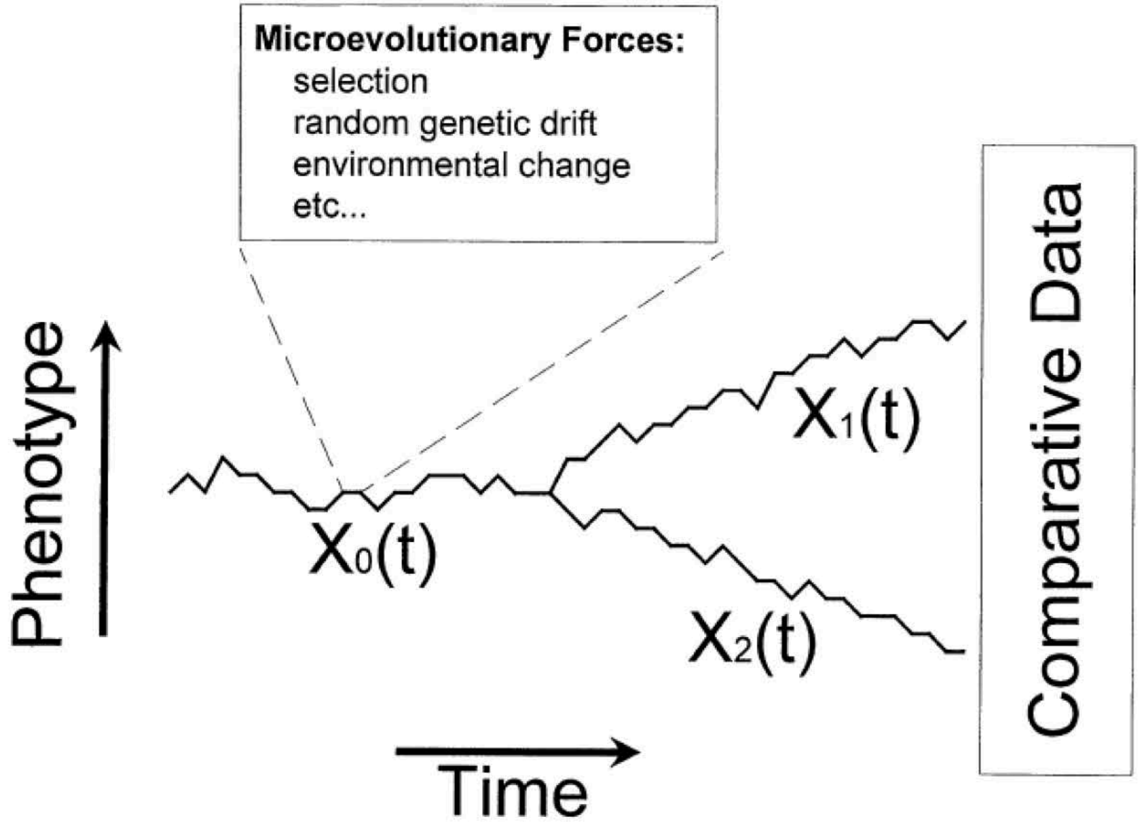
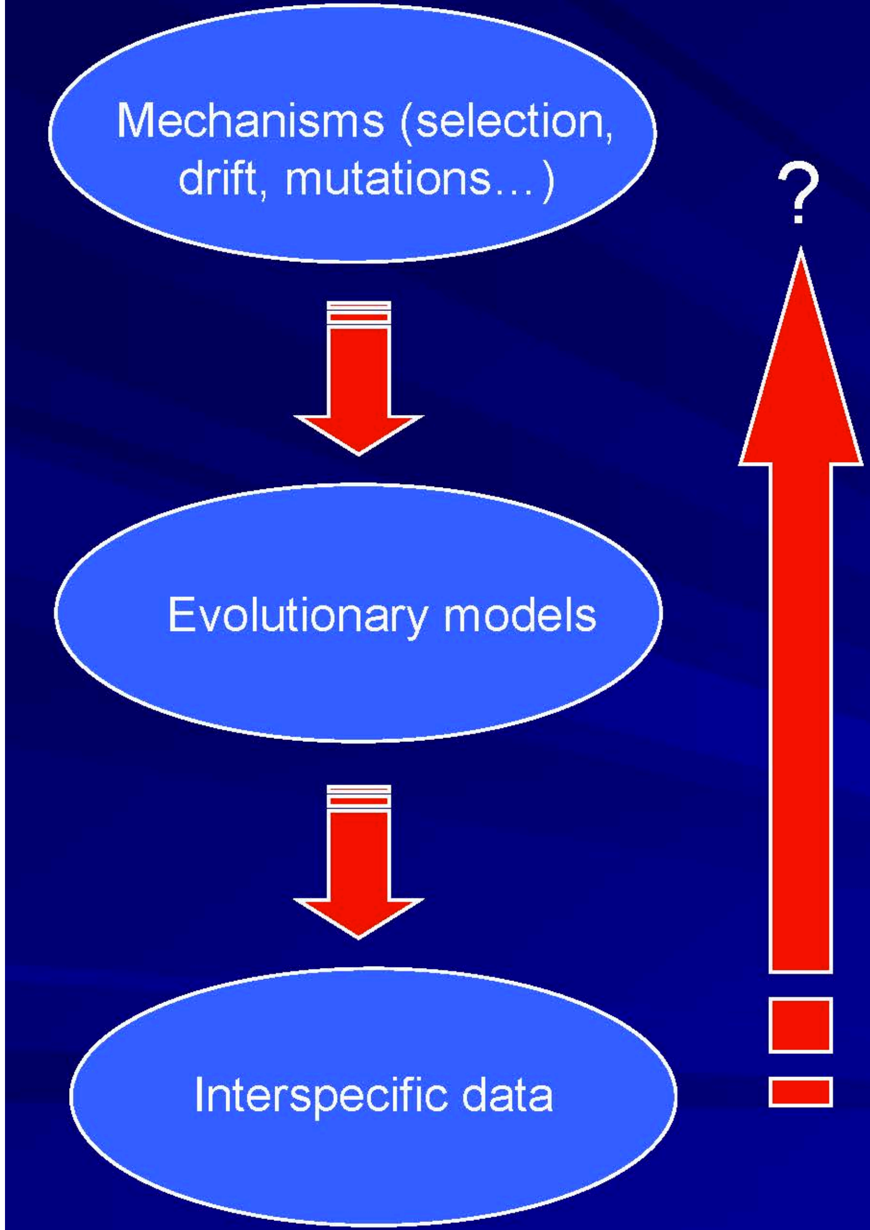


FIG. 1. An illustration of our basic model for translating between microevolutionary forces acting on a short time scale and the macroevolutionary pattern observed in a set of comparative data.

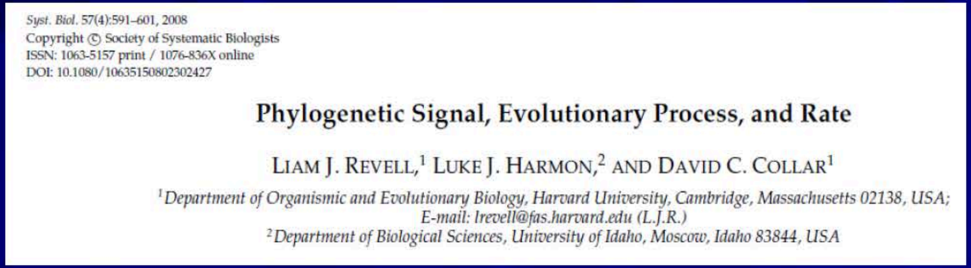


Mechanisms (selection, drift, mutations…)

Evolutionary models

Interspecific data

?



Syst. Biol. 57(4):591–601, 2008
Copyright © Society of Systematic Biologists
ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150802302427

Phylogenetic Signal, Evolutionary Process, and Rate

LIAM J. REVELL,[1] LUKE J. HARMON,[2] AND DAVID C. COLLAR[1]
[1]*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA;*
E-mail: lrevell@fas.harvard.edu (L.J.R.)
[2]*Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844, USA*

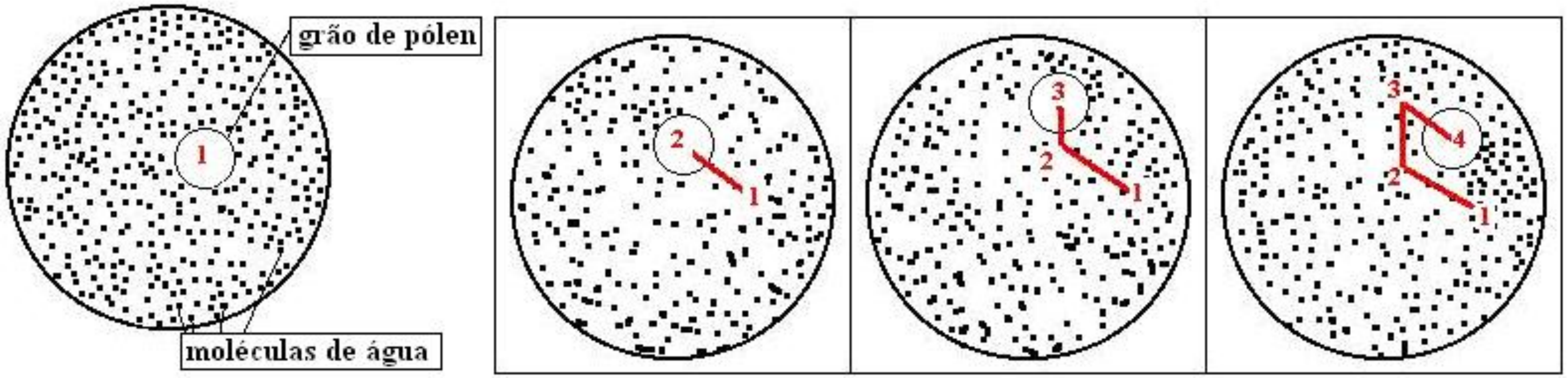
The path from evolutionary mechanisms (selection, drift, mutation and so on) to interspecific variation is a conceptual idea, but it may be hard (or even impossible) to reverse it and actually recover such processes from empirical data…

Modelos de evolução fenotípica

- Movimento Browniano (BM) => Muitos métodos usam este modelo. Tratabilidade analítica
- Early Burst (EB) => Incorpora a idéia de Radiação adaptativa
- Trend => Incorpora a idéia de tendência vista em fósseis
- Ornstein-Uhlenbeck (OU) => Incorpora a idéia de restrições em evolução

Movimento Browniano

- O que é o Movimento Browniano?
- Quando caracteres evoluem de forma parecida com o Movimento Browniano?
- Simulando BM em filogenias



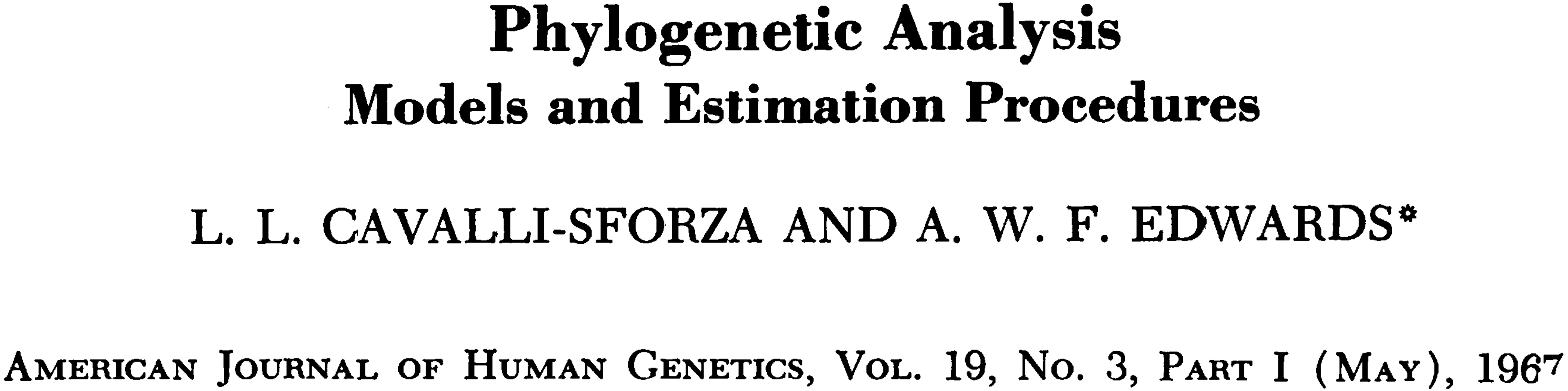
Robert Brown, botânico

Movimento Browniano

- Modelo para a evolução de caracteres contínuos
  - Seguindo um random walk

- Estados de caracter mudam continuamente ao longo do tempo
  - Processo estocástico de tempo contínuo

- Depois de um tempo e múltiplas simulações, o estado do caracter seguem uma distribuição normal



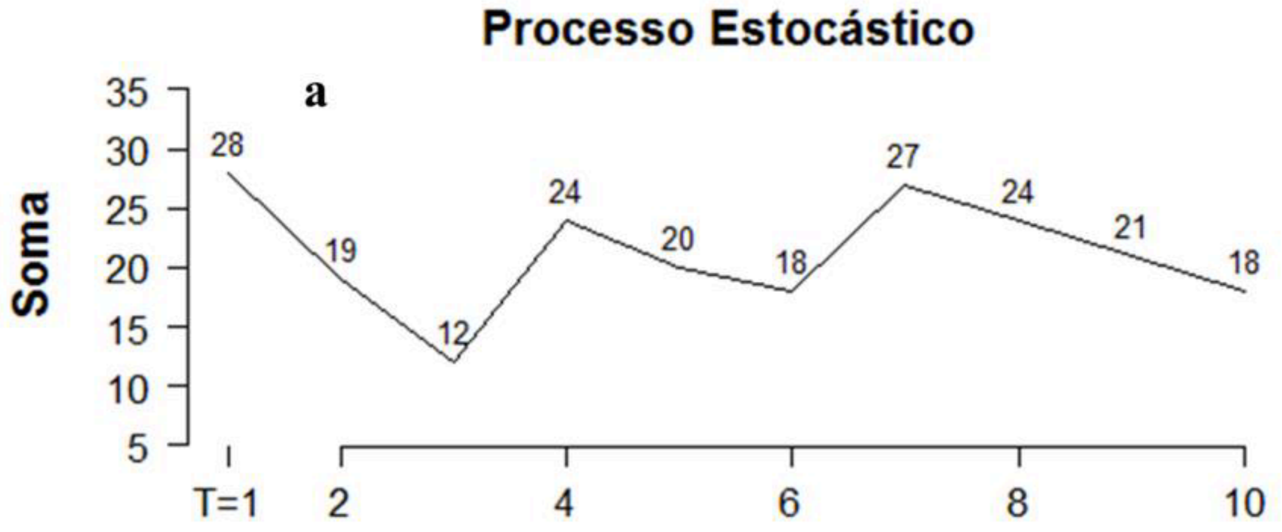
Phylogenetic Analysis

Models and Estimation Procedures

L. L. CAVALLI-SFORZA AND A. W. F. EDWARDS*

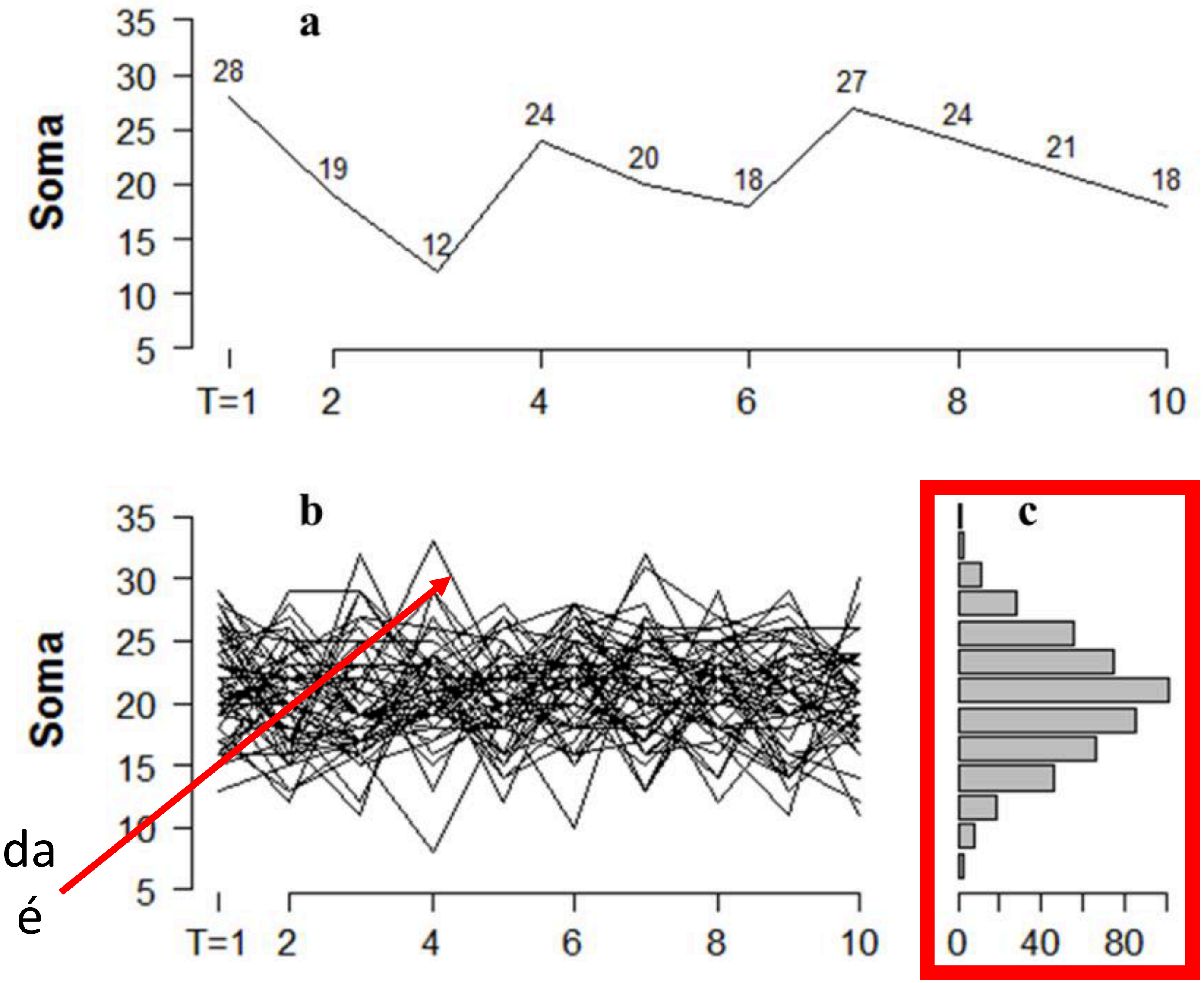
AMERICAN JOURNAL OF HUMAN GENETICS, VOL. 19, NO. 3, PART I (MAY), 1967

...mas whattahell é um processo estocástico de tempo contínuo?



Soma dos resultados de 2 dados (6-36)

...mas whattahell é um processo estocástico de tempo contínuo?



O resultado de cada jogada individual é imprevisível

Mas a distribuição de probabilidade de um conjunto de curvas é conhecida (aprox. distribuição normal) e permite predizer eventos futuros

Movimento Browniano

- Dois parâmetros => valor inicial (θ) e taxa de mudança ($\boldsymbol{\sigma}^2$)
- Três fatos descrevem um BM (Processo de Wiener)
  - W(t) é o valor do caracter no tempo t. Logo:
  - E[W(t)]=W(0)
  - Incrementos (passos sucessivos) são independentes (”pouca memória”)
  - W(t)~N(W(0), $\boldsymbol{\sigma}^2$t)
  - Traduzindo: o valor do caracter W no tempo t é aproximado pela distribuição normal com média 0 (*standard normal distribution*) e variância igual à taxa de mudança no tempo t

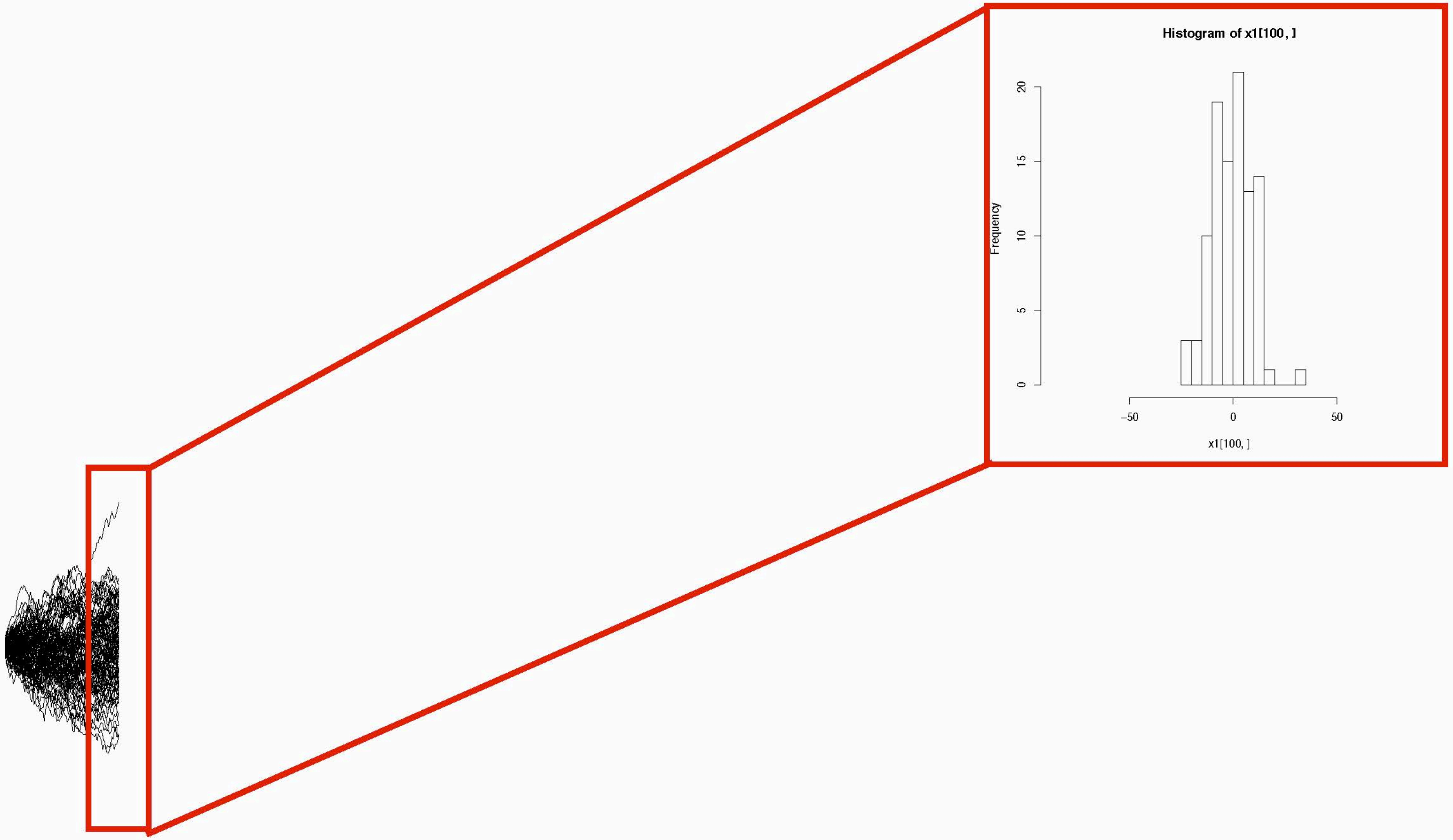
$$dX(t)=\sigma dB(t)$$

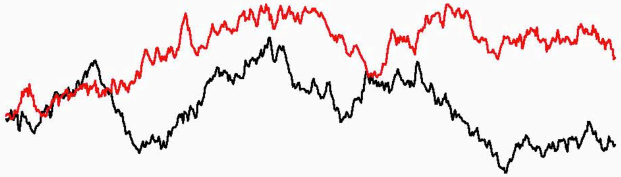
$$dX(t)=\sigma dB(t)$$

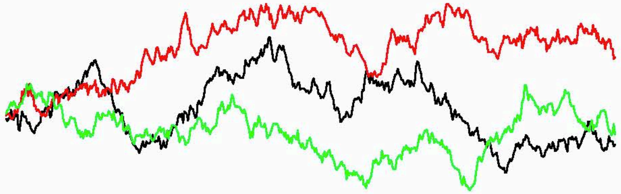
Quantidade de mudança no caracter X entre o tempo do tempo t para $t + dt$

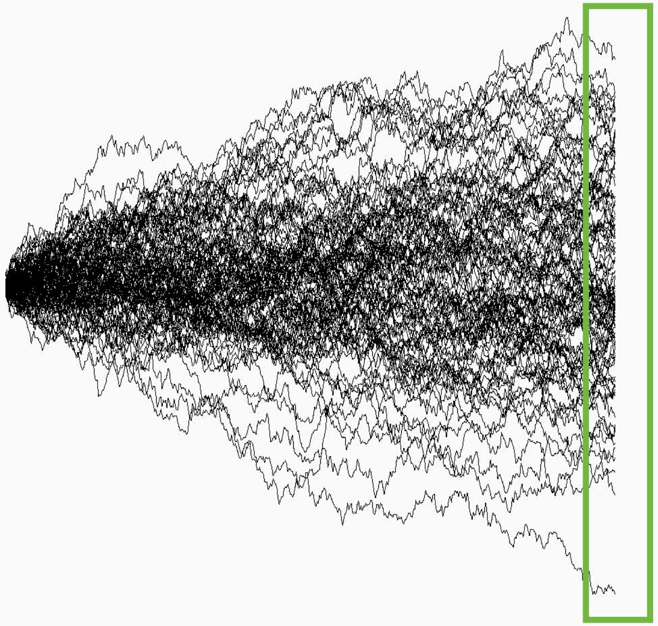
$$\mathbf{dX(t)} = \sigma dB(t)$$

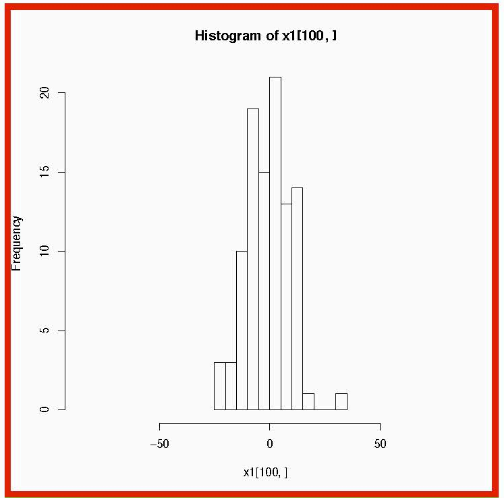
Descreve o ”random walk” na variável, que se aproxima de uma distribuição normal com média 0 e variância *dt*

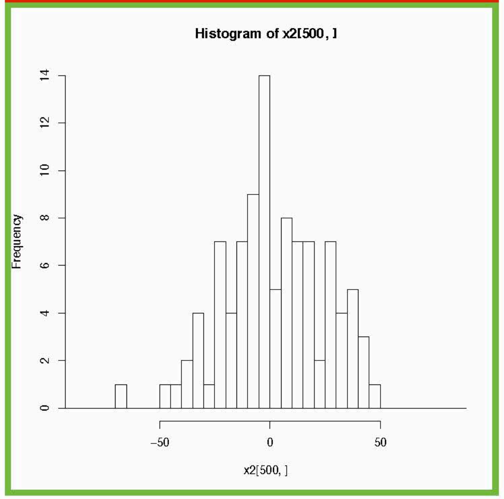

Histogram of x1[100,]
Frequency
0
5
10
15
20
−50
0
50
x1[100,]

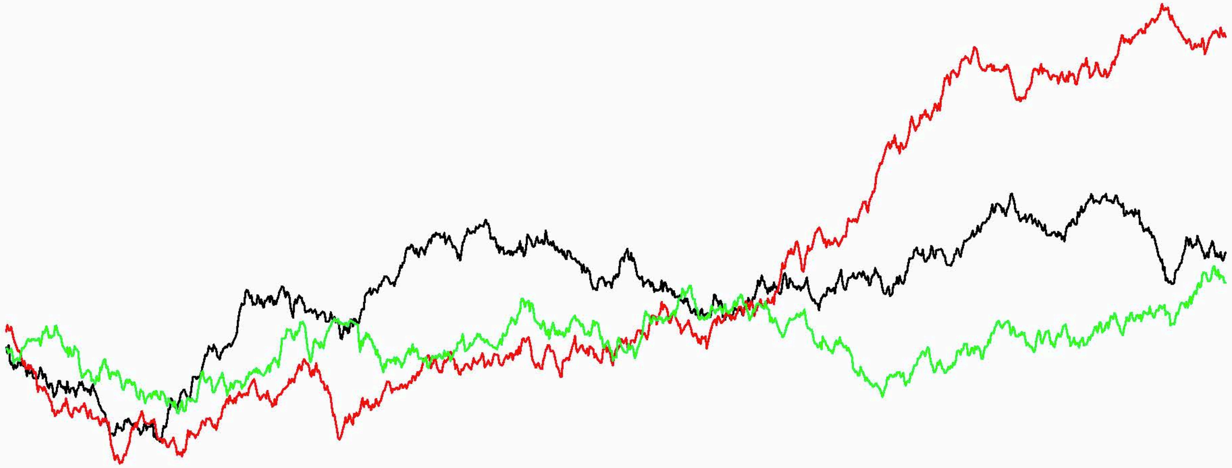


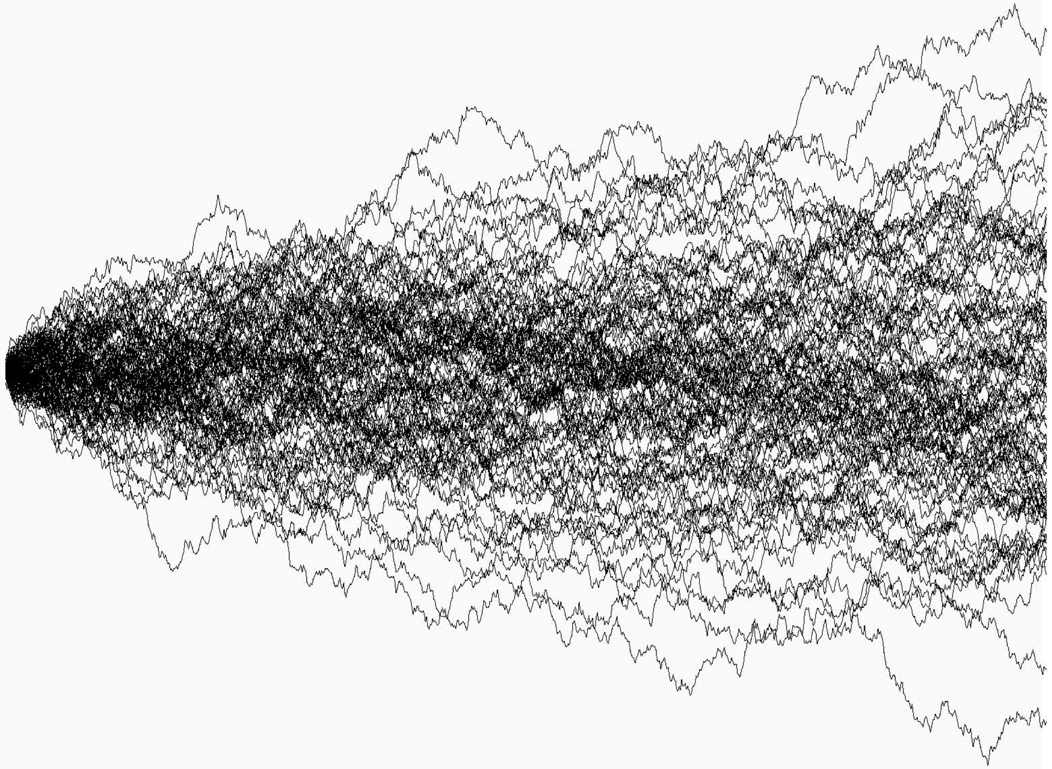


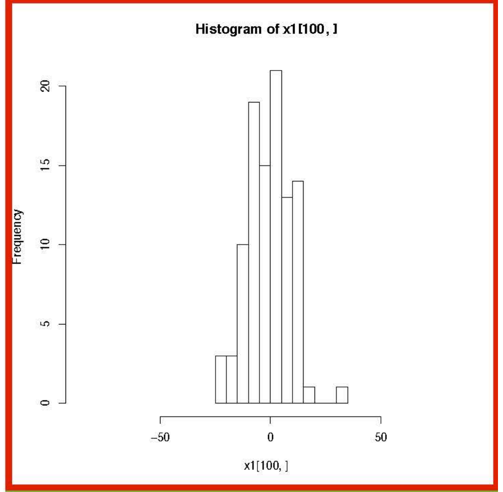


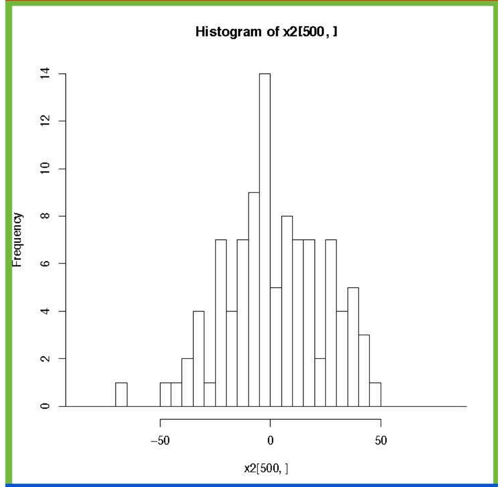

Histogram of x1[100,]

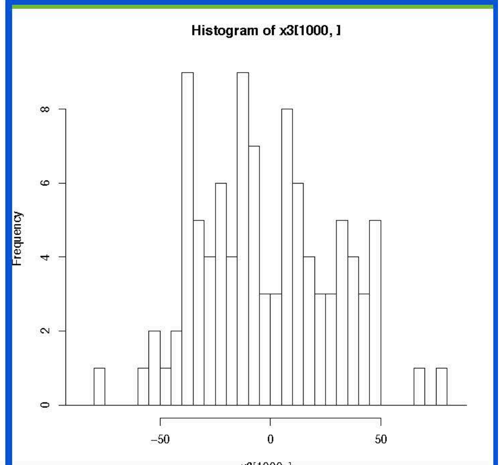

Histogram of x2[500,]

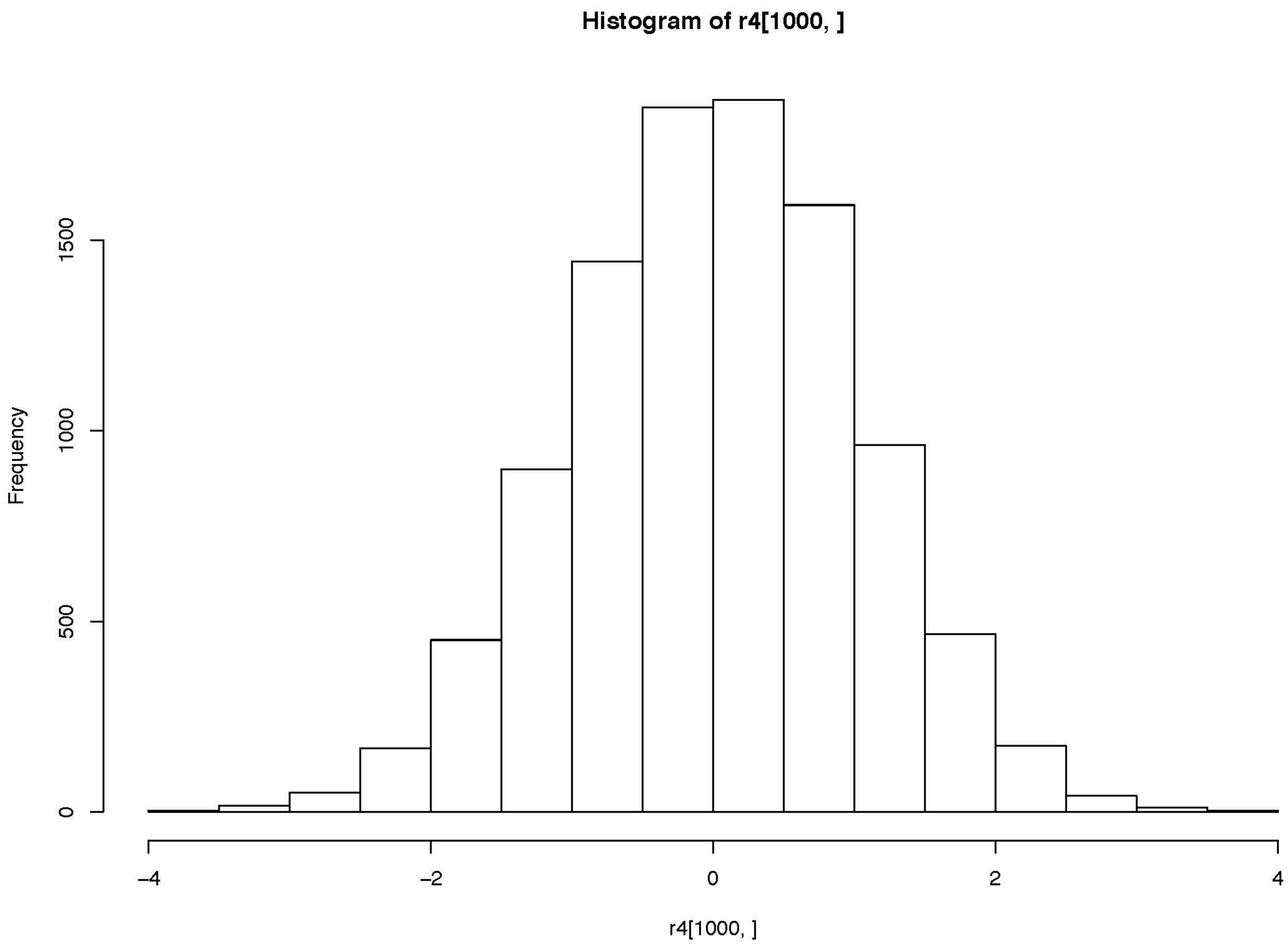


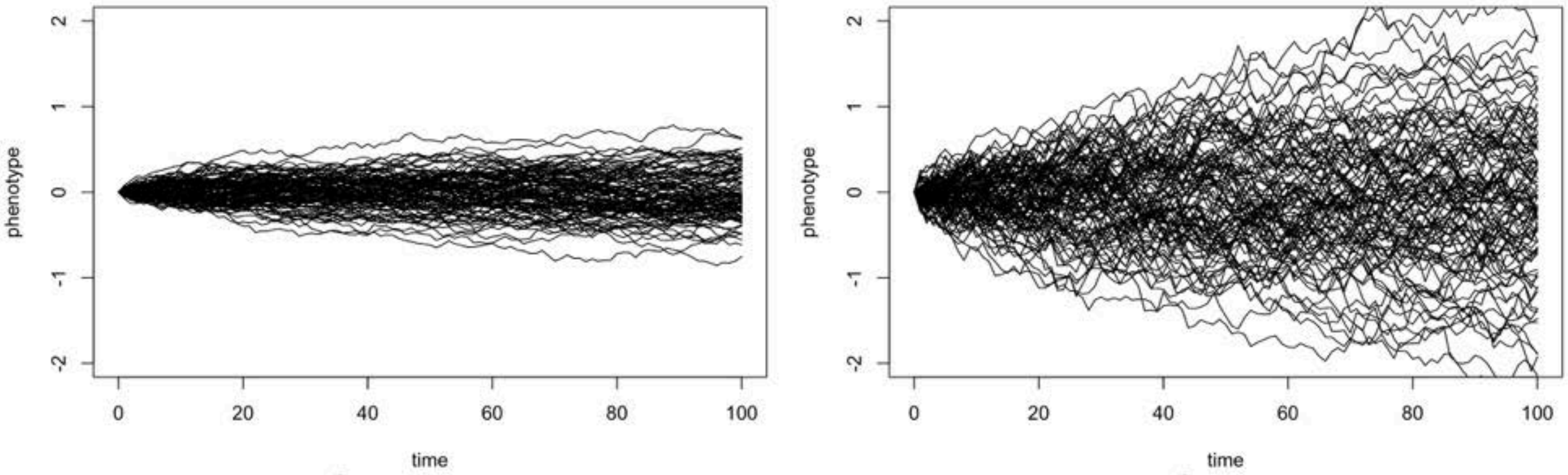



Histogram of x1[100,]
Frequency
0
5
10
15
20
−50
0
50
x1[100,]

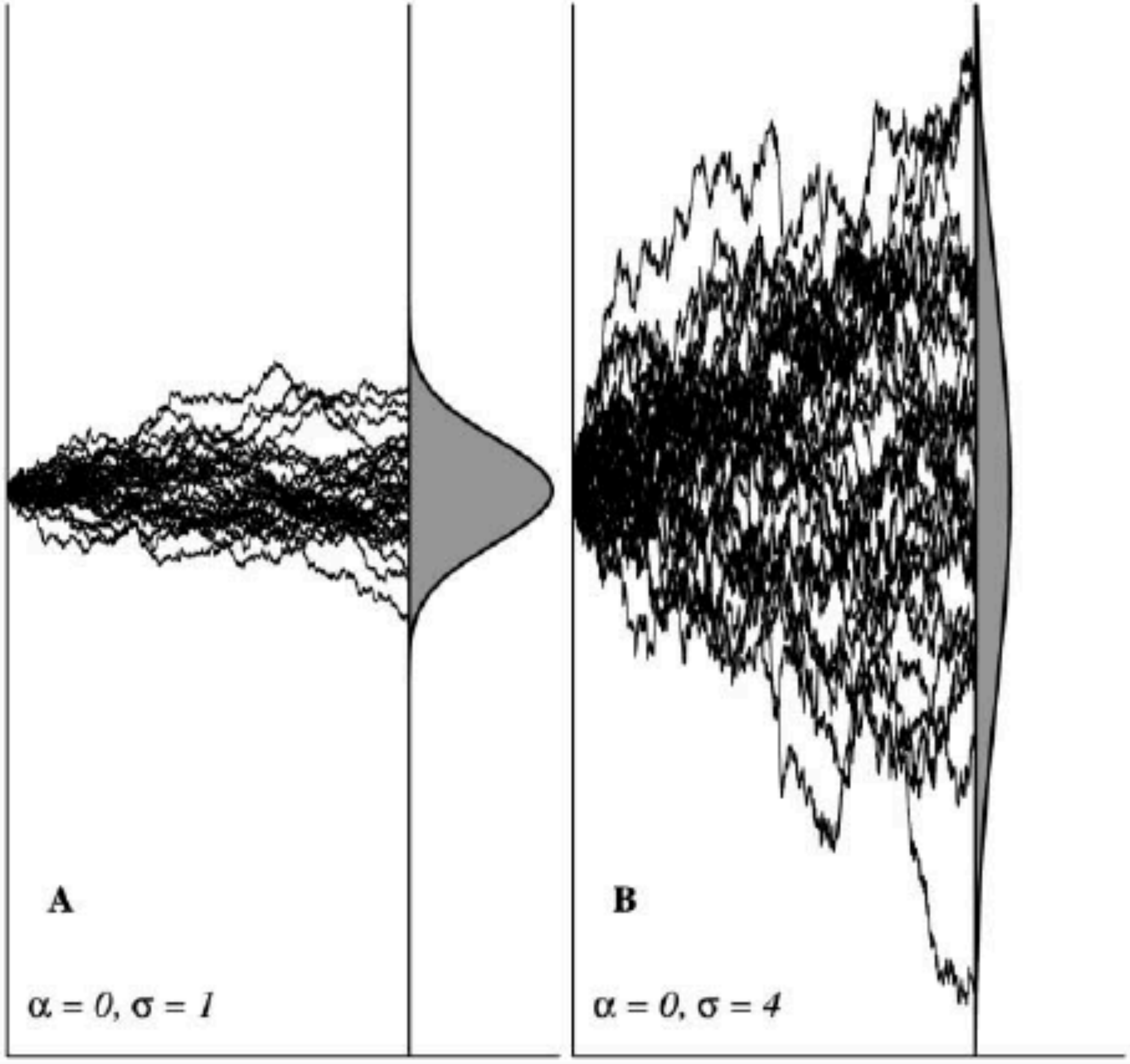

Histogram of x2[500,]
Frequency
0
2
4
6
8
10
12
14
−50
0
50
x2[500,]


Histogram of x3[1000,]
Frequency
0
2
4
6
8
−50
0
50


Histogram of r4[1000,]
Frequency
0
500
1000
1500
−4
−2
0
2
4
r4[1000,]

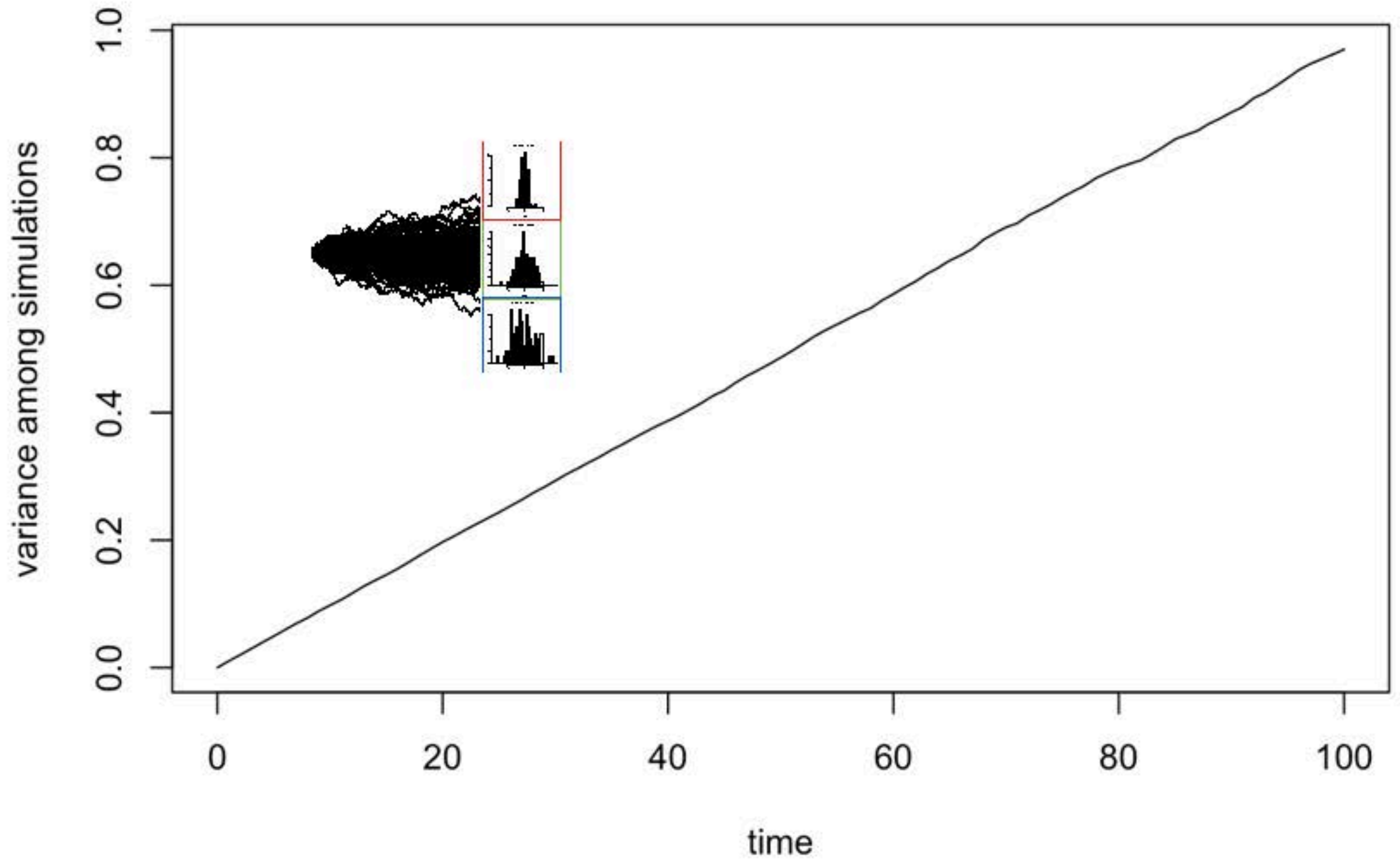


$\sigma^2 = 0.01$ $\sigma^2 = 0.1$



Butler & King 2004

variance increases with time



$$\mathrm{Var}(X) = \sigma^2 * \mathit{time}$$

Quando a evolução é descrita por BM?

- Deriva genética
- Mudança pontuada aleatória
- Quando a seleção é fraca em relação à escala de tempo considerada
- Quando seleção muda de direção aleatoriamente ao longo do tempo

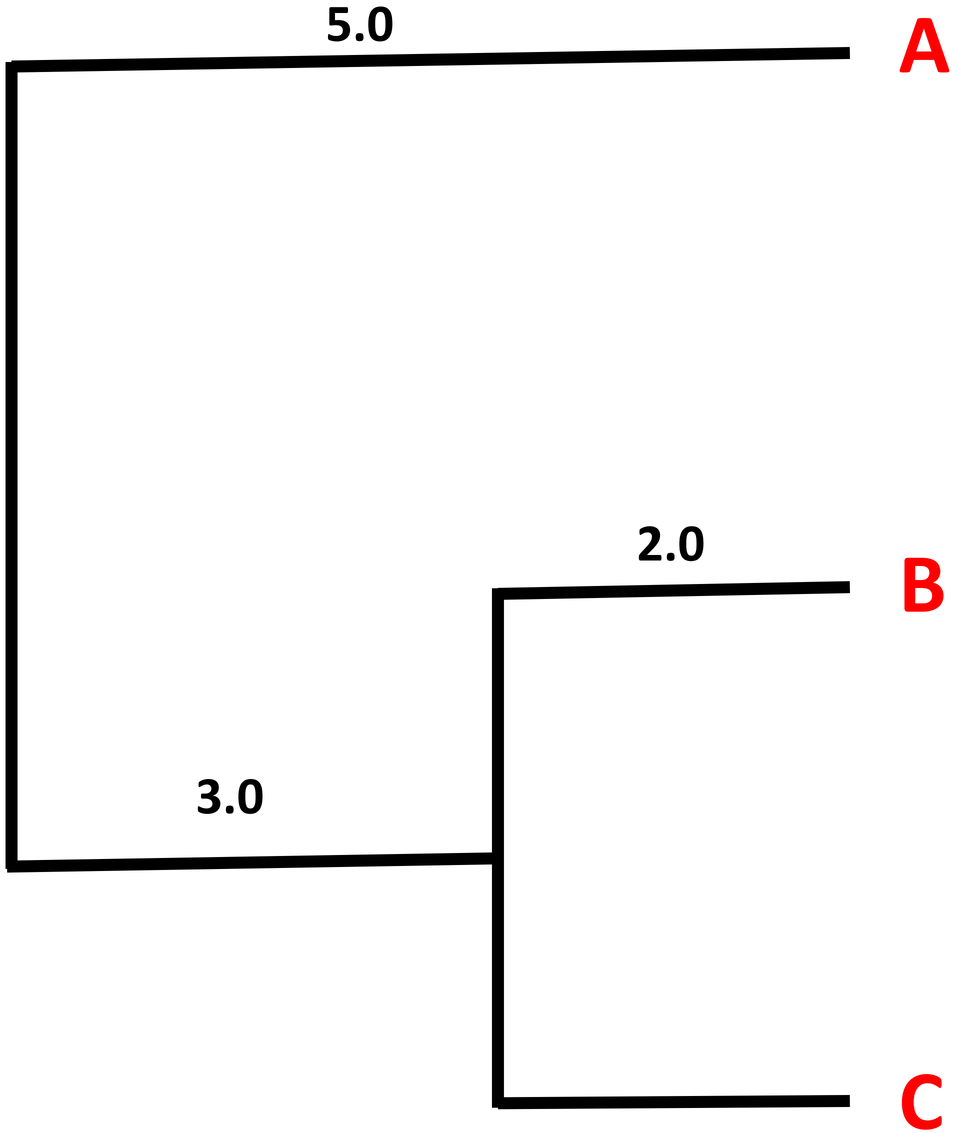
Como a filogenia influencia a evolução de fenótipos sob um processo BM?

O parentesco comum aumenta ou diminui a similaridade entre espécies em termos de fenótipo?

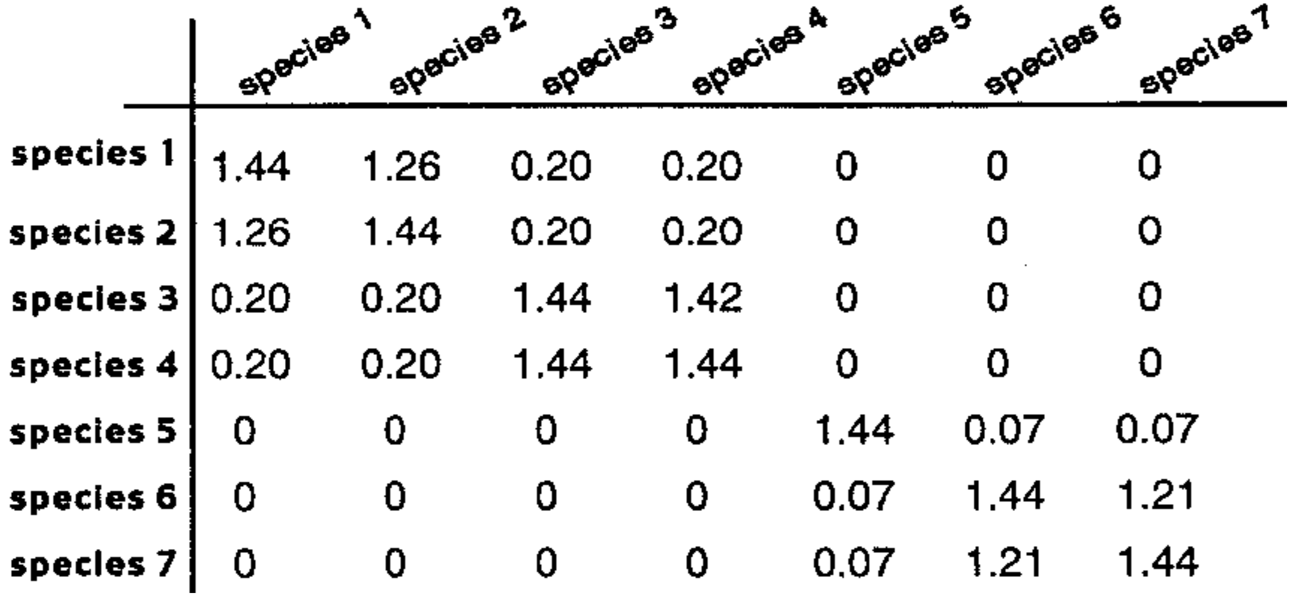
... Mas antes, vamos entender o que é uma matriz de distância filogenética (patrística) e de variância-covariância

Matriz de distância patrística

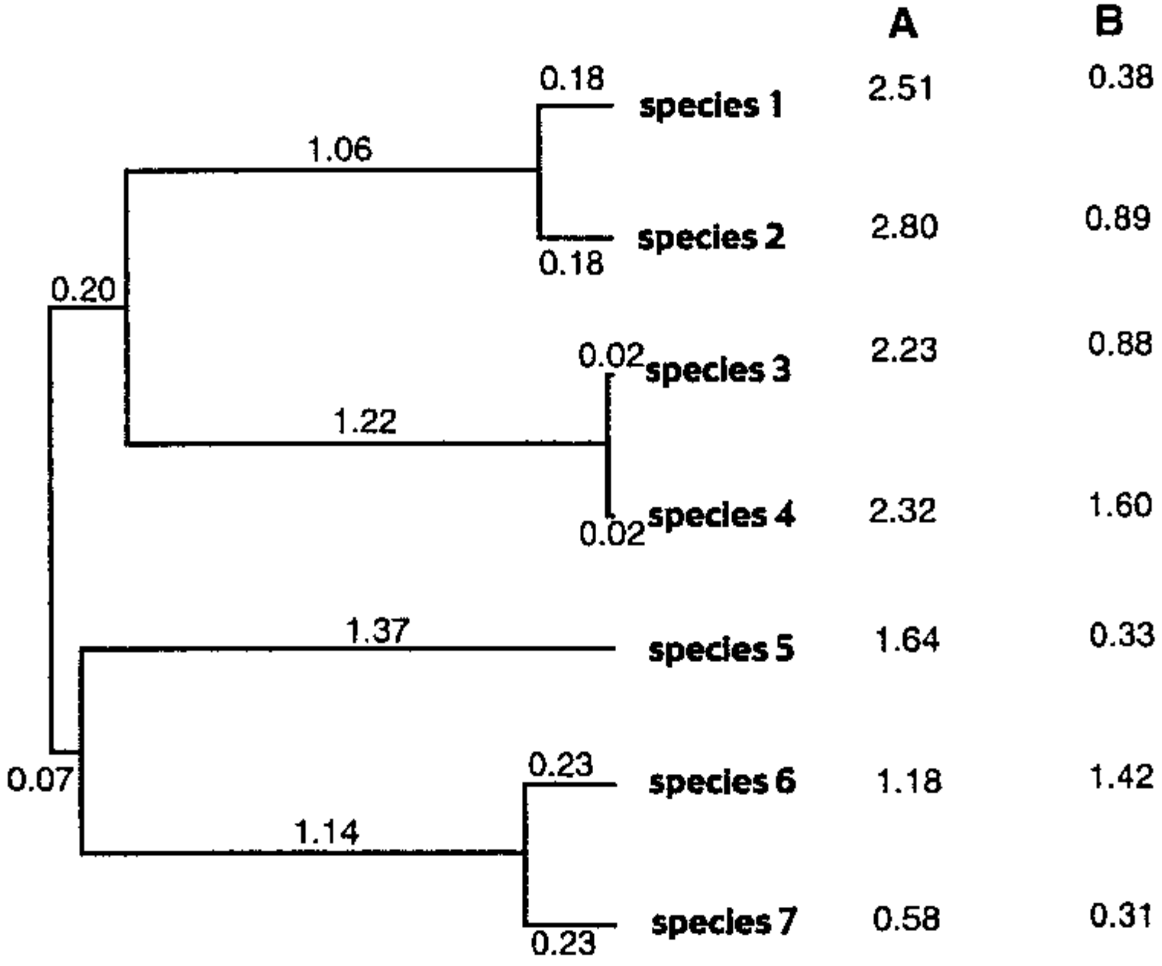
	A	B	C
A	0	10	10
B	10	0	4
C	10	4	0



Matriz de variância-covariância



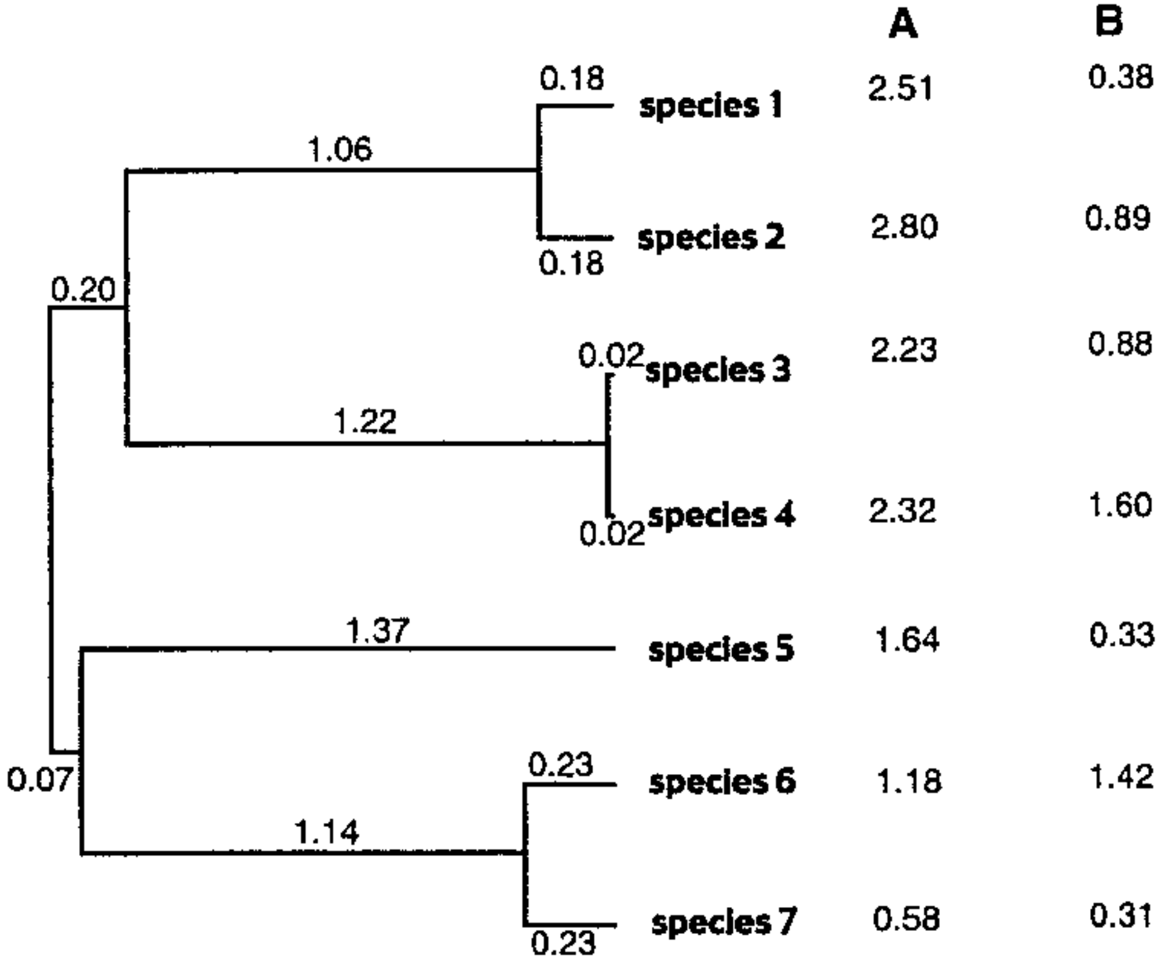
	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	1.44	1.26	0.20	0.20	0	0	0
species 2	1.26	1.44	0.20	0.20	0	0	0
species 3	0.20	0.20	1.44	1.42	0	0	0
species 4	0.20	0.20	1.44	1.44	0	0	0
species 5	0	0	0	0	1.44	0.07	0.07
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44



Freckleton et al. 2002

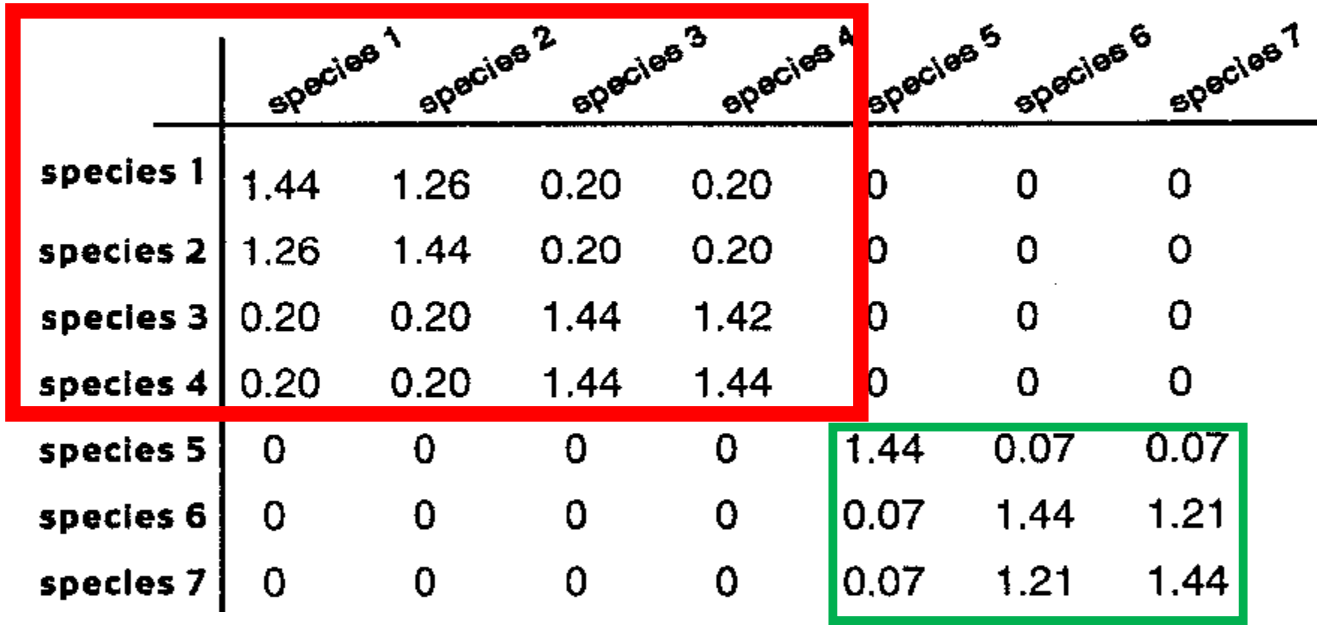
Matriz de variância-covariância

	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	var	cov	cov	cov	cov	cov	cov
species 2	cov	var	cov	cov	cov	cov	cov
species 3	cov	cov	var	cov	cov	cov	cov
species 4	cov	cov	cov	var	cov	cov	cov
species 5	cov	cov	cov	cov	var	cov	cov
species 6	cov	cov	cov	cov	cov	var	cov
species 7	cov	cov	cov	cov	cov	cov	var

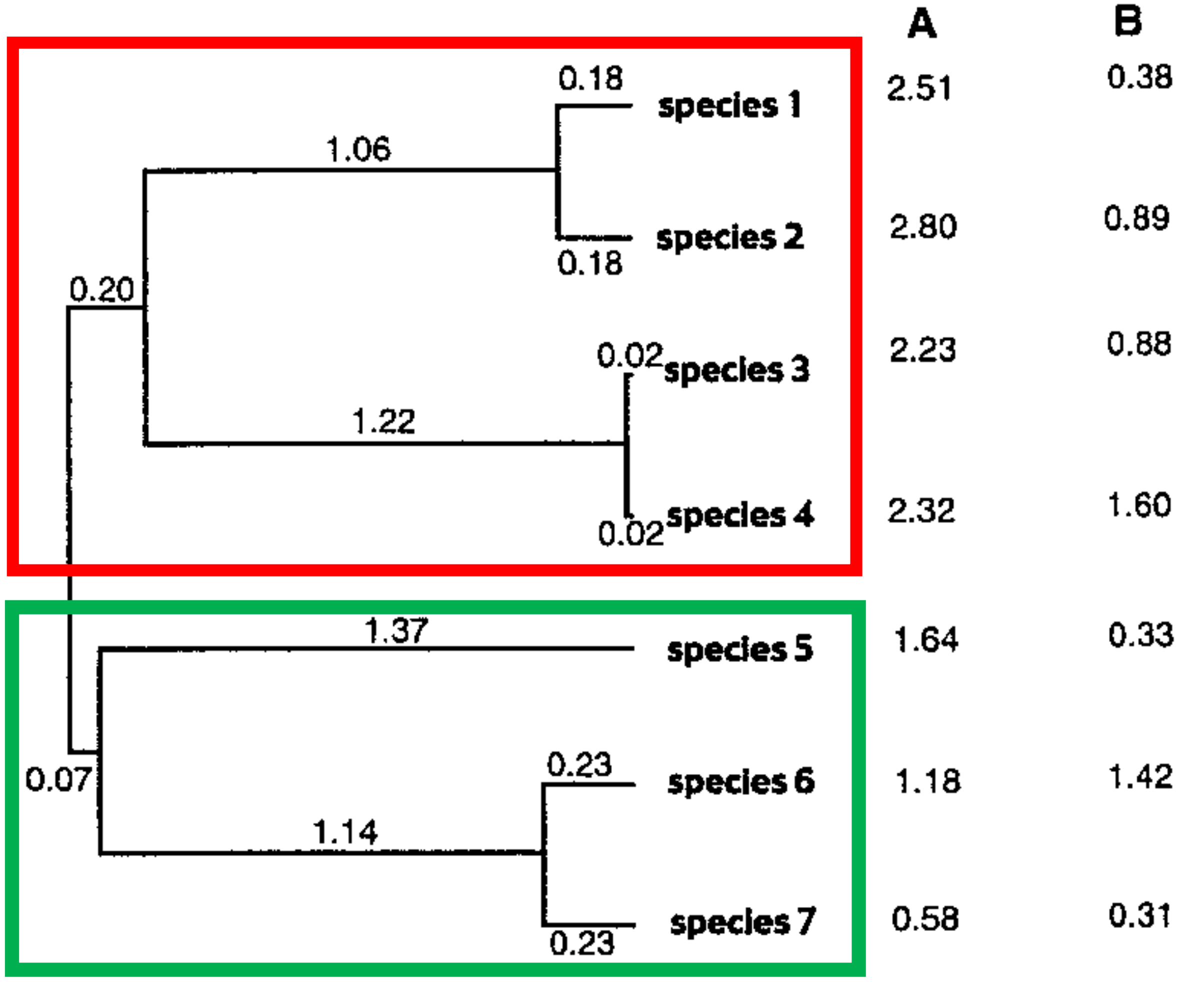


Freckleton et al. 2002

Matriz de variância-covariância

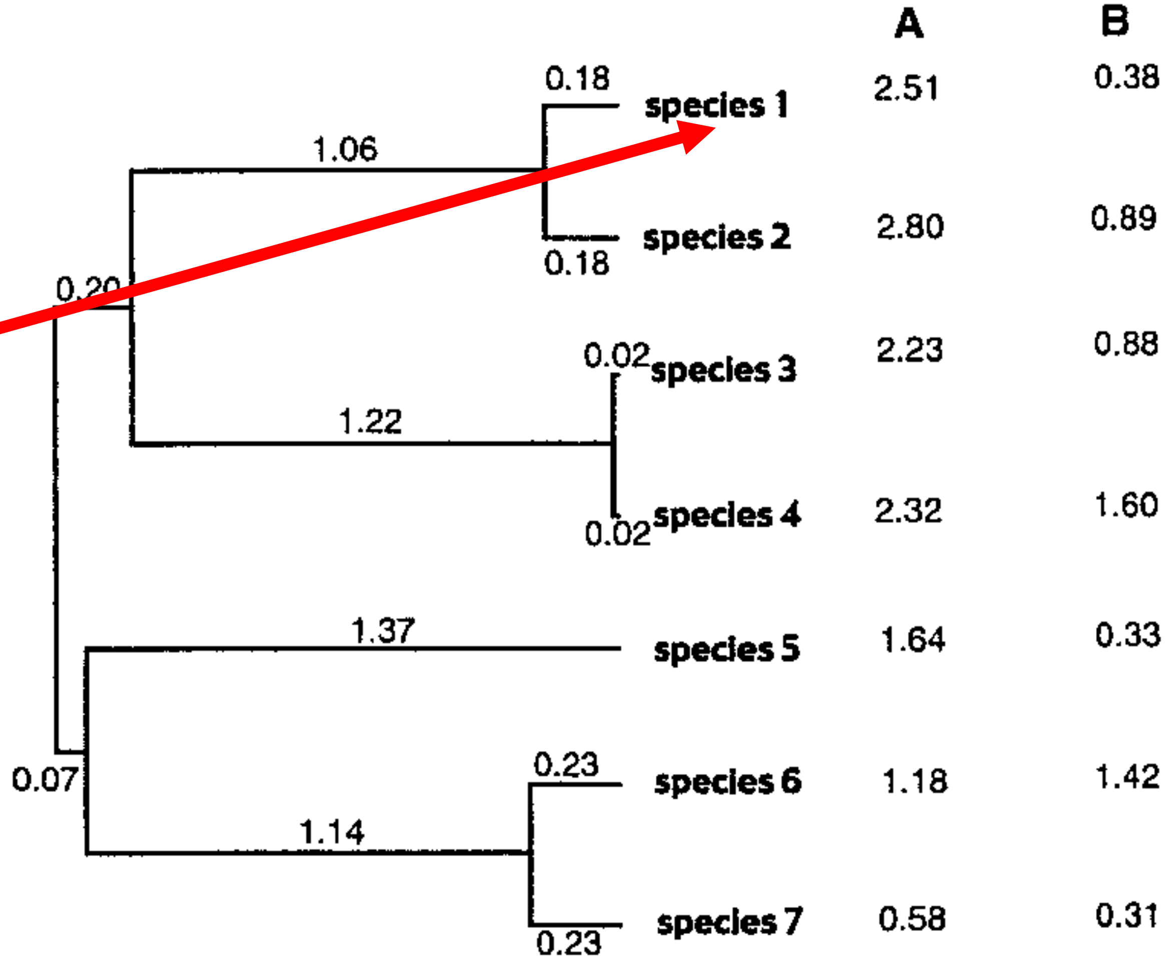


	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	1.44	1.26	0.20	0.20	0	0	0
species 2	1.26	1.44	0.20	0.20	0	0	0
species 3	0.20	0.20	1.44	1.42	0	0	0
species 4	0.20	0.20	1.44	1.44	0	0	0
species 5	0	0	0	0	1.44	0.07	0.07
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44



Freckleton et al. 2002

Matriz de variância-covariância



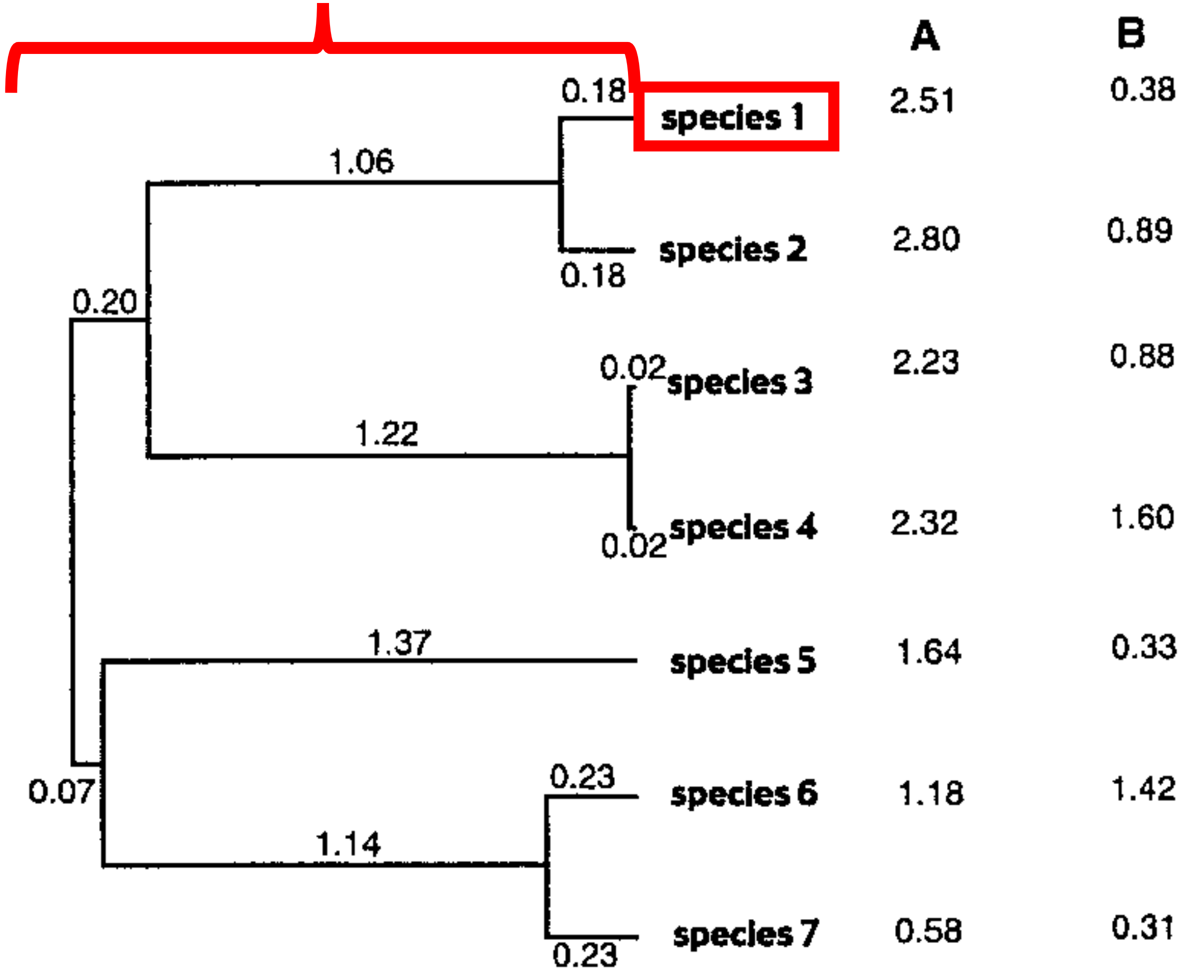
	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	1.44	1.26	0.20	0.20	0	0	0
species 2	1.26	1.44	0.20	0.20	0	0	0
species 3	0.20	0.20	1.44	1.42	0	0	0
species 4	0.20	0.20	1.44	1.44	0	0	0
species 5	0	0	0	0	1.44	0.07	0.07
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44

Freckleton et al. 2002

Matriz de variância-covariância

var(species1)=0.18+1.06+0.20=1.44

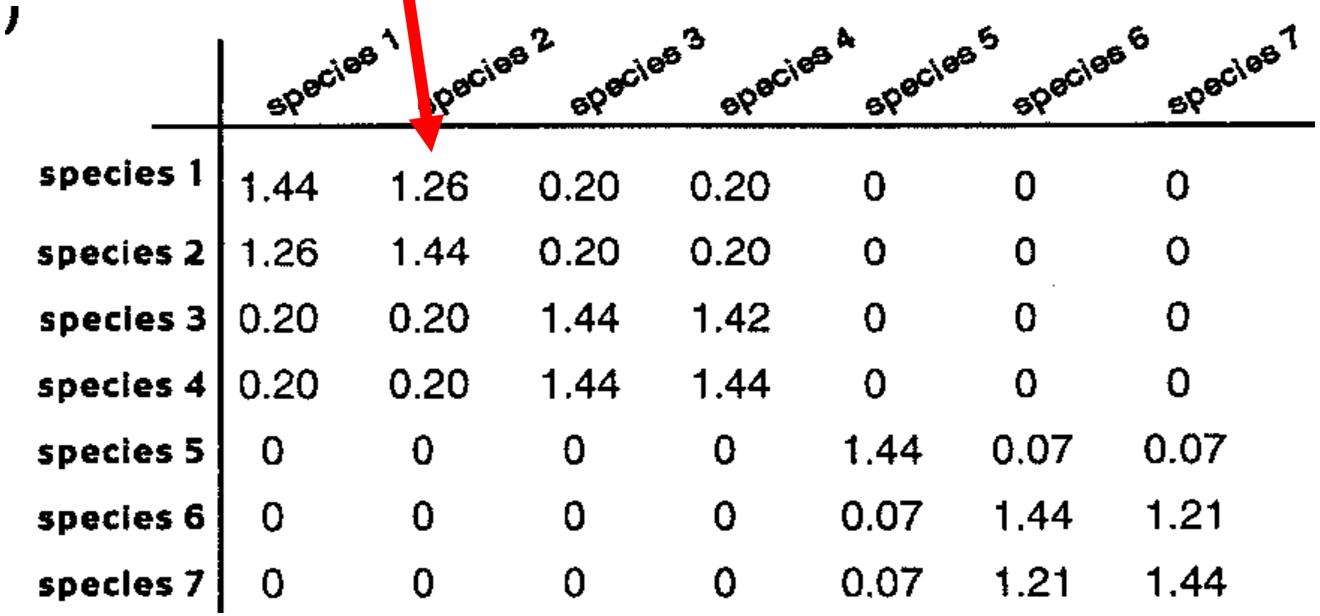
	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	1.44	1.26	0.20	0.20	0	0	0
species 2	1.26	1.44	0.20	0.20	0	0	0
species 3	0.20	0.20	1.44	1.42	0	0	0
species 4	0.20	0.20	1.44	1.44	0	0	0
species 5	0	0	0	0	1.44	0.07	0.07
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44



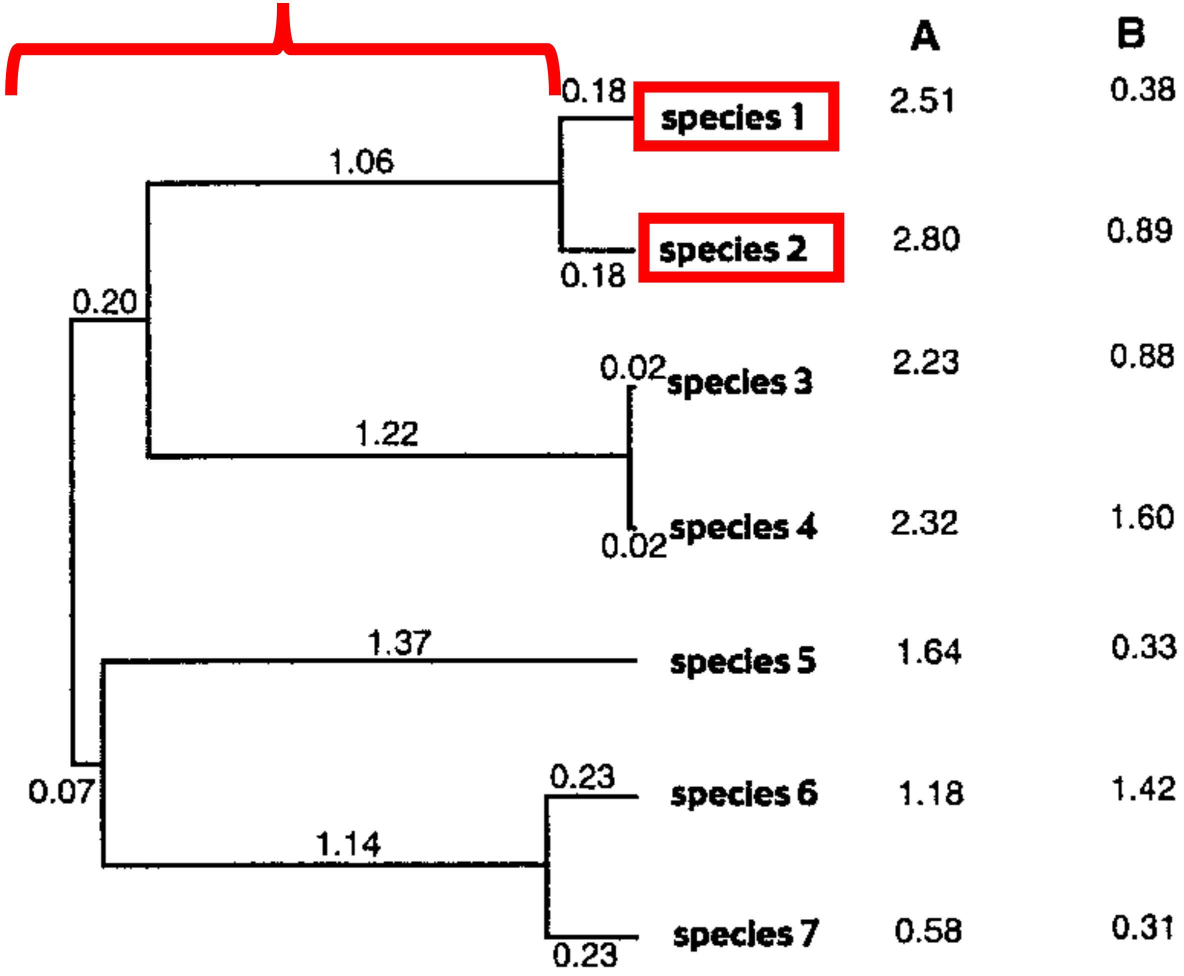
Freckleton et al. 2002

Matriz de variância-covariância

cov(1,2)=1.06+0.20=1.26



	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	1.44	1.26	0.20	0.20	0	0	0
species 2	1.26	1.44	0.20	0.20	0	0	0
species 3	0.20	0.20	1.44	1.42	0	0	0
species 4	0.20	0.20	1.44	1.44	0	0	0
species 5	0	0	0	0	1.44	0.07	0.07
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44

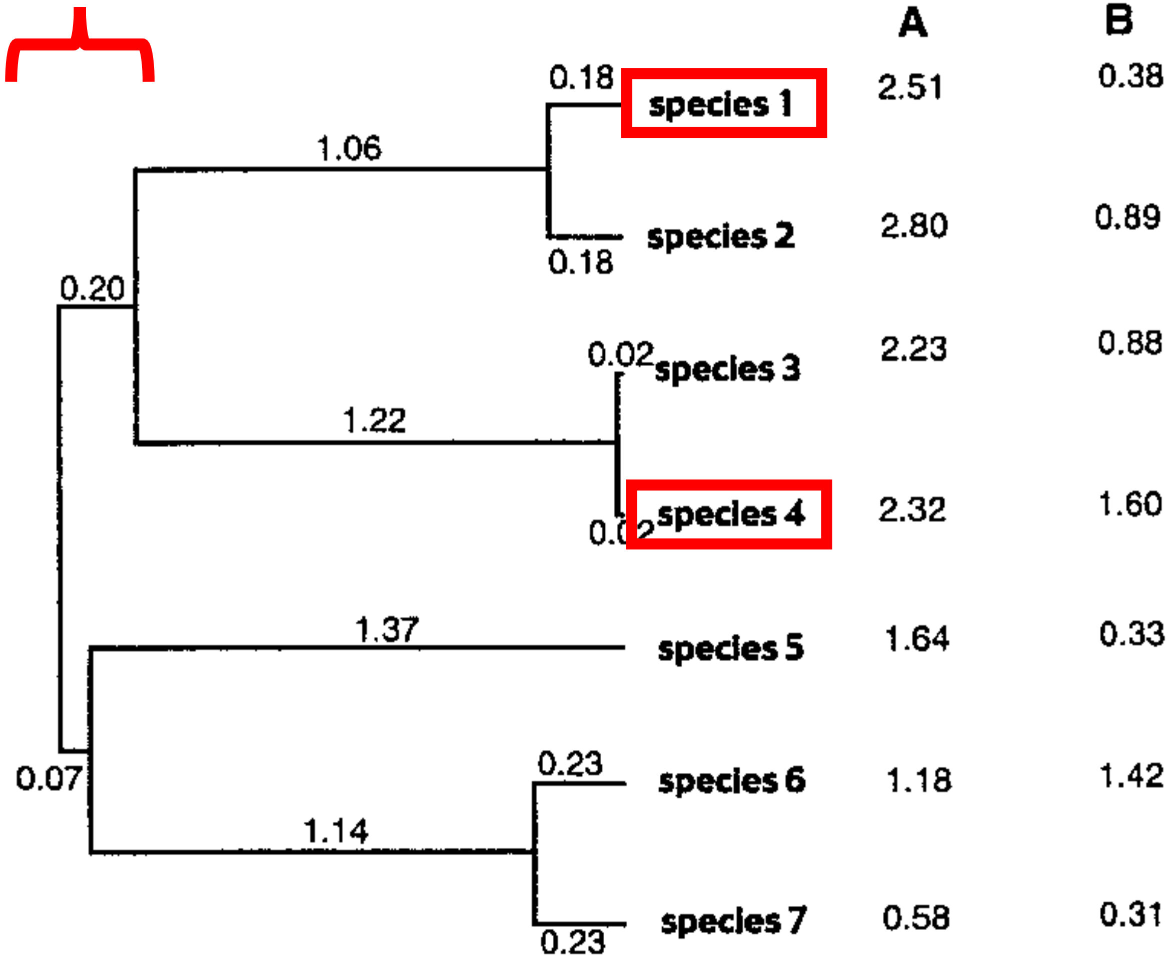


Freckleton et al. 2002

Matriz de variância-covariância

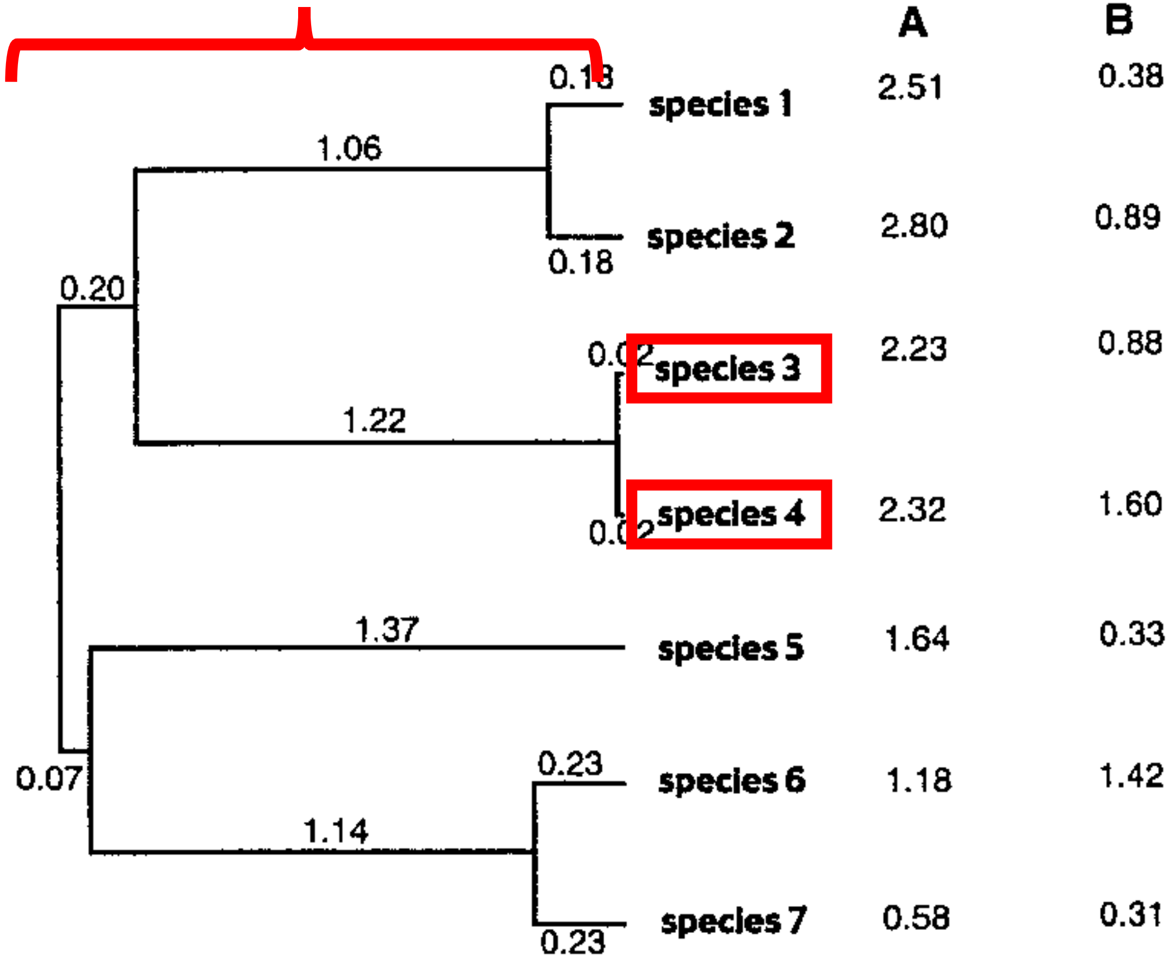
cov(1,4)=0.20

	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	1.44	1.26	0.20	0.20	0	0	0
species 2	1.26	1.44	0.20	0.20	0	0	0
species 3	0.20	0.20	1.44	1.42	0	0	0
species 4	0.20	0.20	1.44	1.44	0	0	0
species 5	0	0	0	0	1.44	0.07	0.07
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44

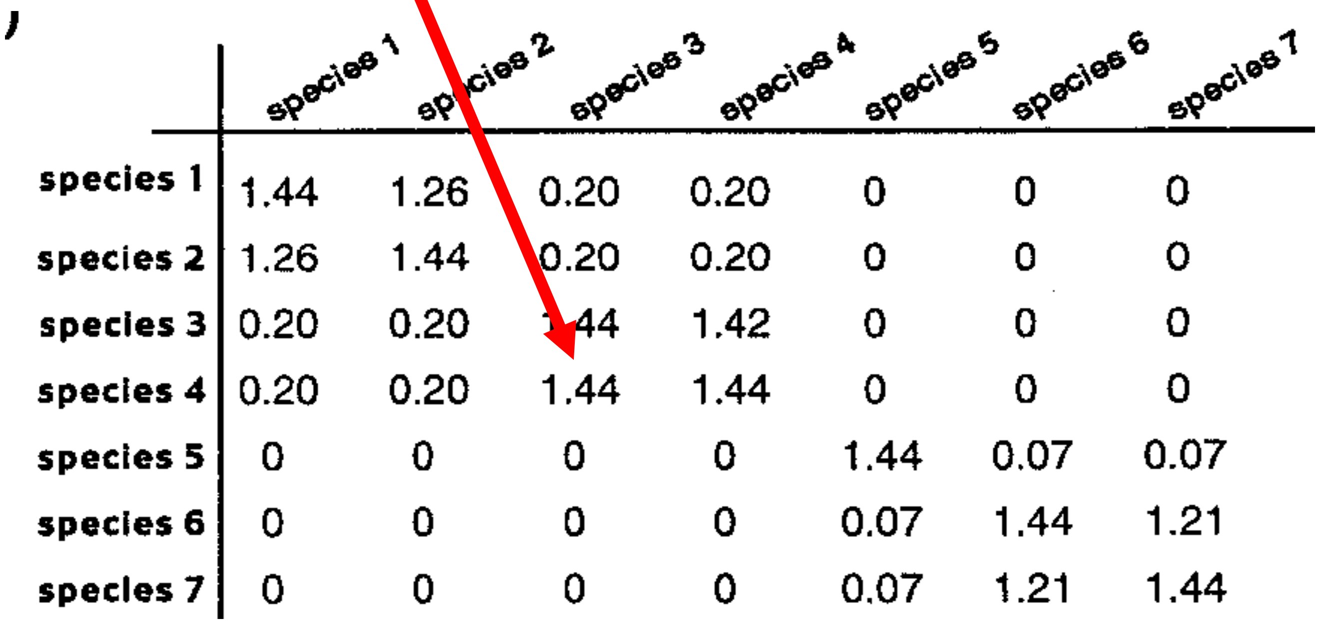


Freckleton et al. 2002

Matriz de variância-covariância



cov(3,4)=0.20+1.22+0.02=1.44

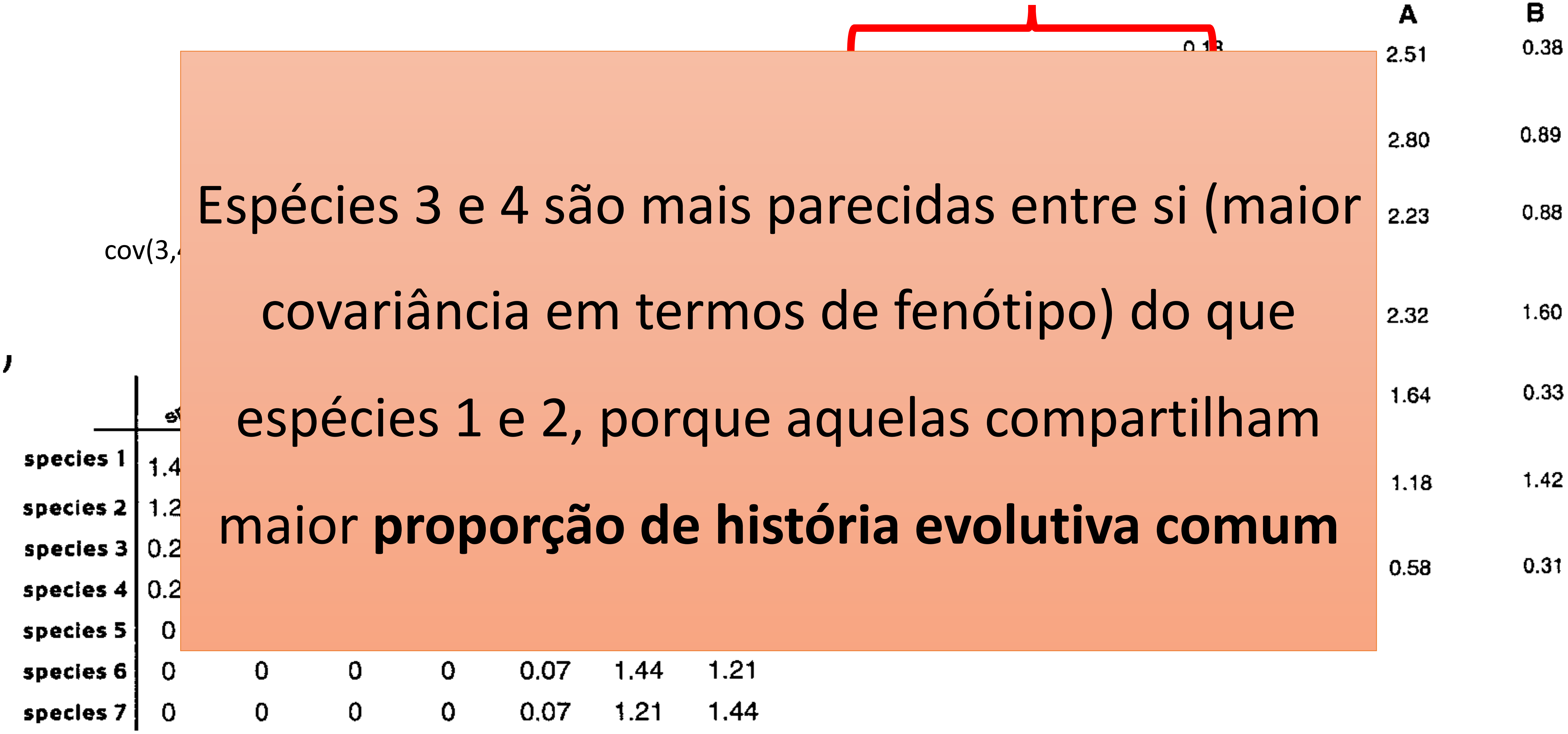


	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	1.44	1.26	0.20	0.20	0	0	0
species 2	1.26	1.44	0.20	0.20	0	0	0
species 3	0.20	0.20	1.44	1.42	0	0	0
species 4	0.20	0.20	1.44	1.44	0	0	0
species 5	0	0	0	0	1.44	0.07	0.07
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44

Freckleton et al. 2002

Matriz de variância-covariância

Espécies 3 e 4 são mais parecidas entre si (maior covariância em termos de fenótipo) do que espécies 1 e 2, porque aquelas compartilham maior **proporção de história evolutiva comum**



species 1	1.4						
species 2	1.2						
species 3	0.2						
species 4	0.2						
species 5	0						
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44

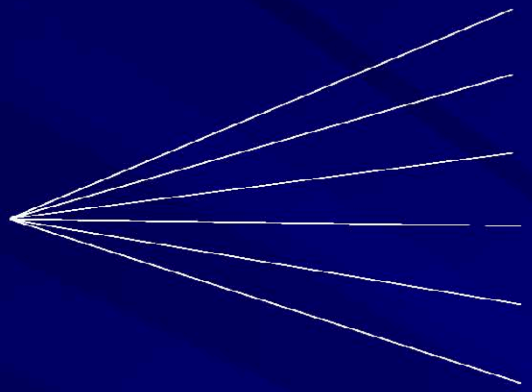
Freckleton et al. 2002

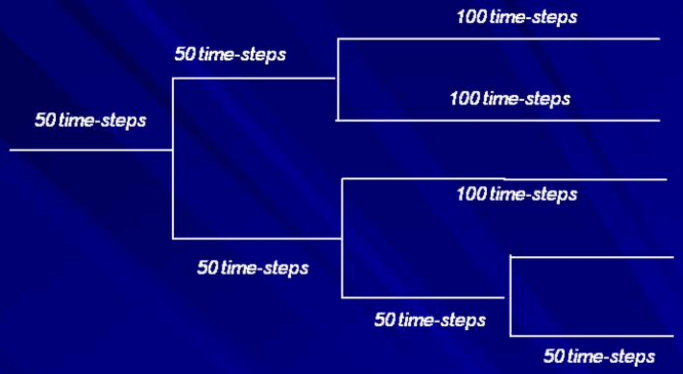
A mesma filogenia pode gerar diferentes matrizes VCV

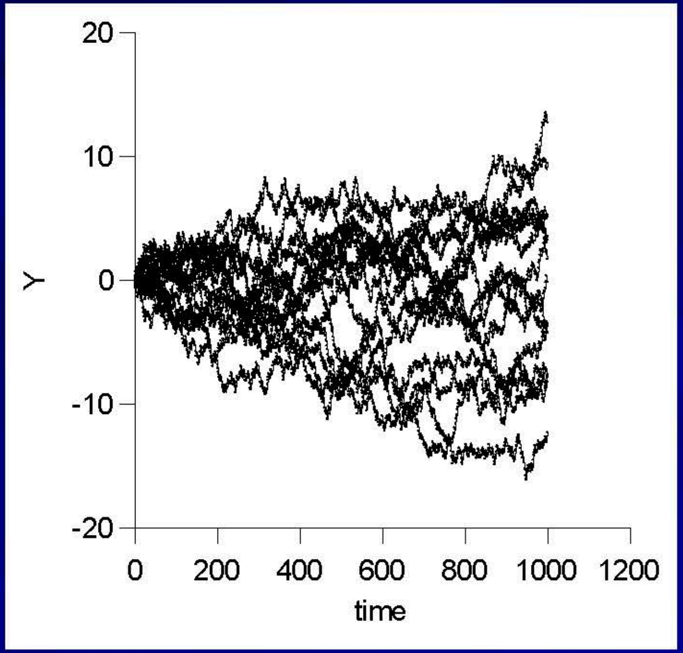
Modelos evolutivos

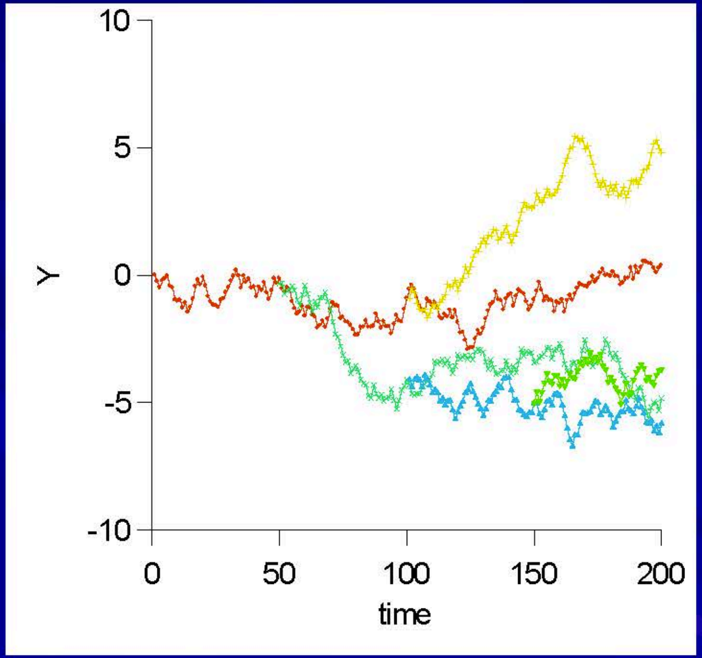
Como a filogenia influencia a evolução de fenótipos sob um processo BM?

O parentesco comum aumenta ou diminui a similaridade entre espécies em termos de fenótipo?







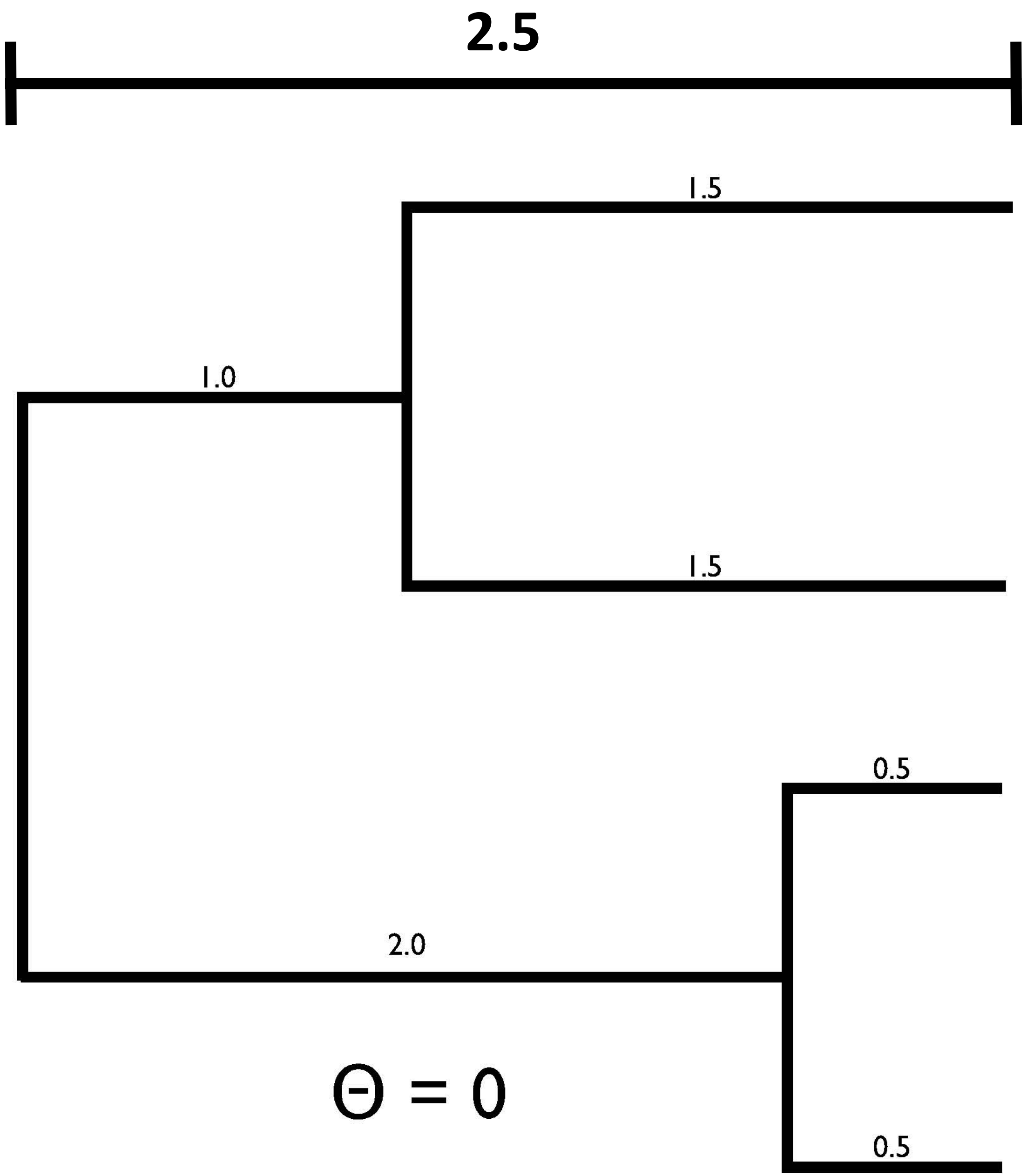


Here we assumed that species are INDEPENDENT (the started all at the root)

Here species are PHYLOGENETICALLY STRUCTURED

Simulando um processo BM

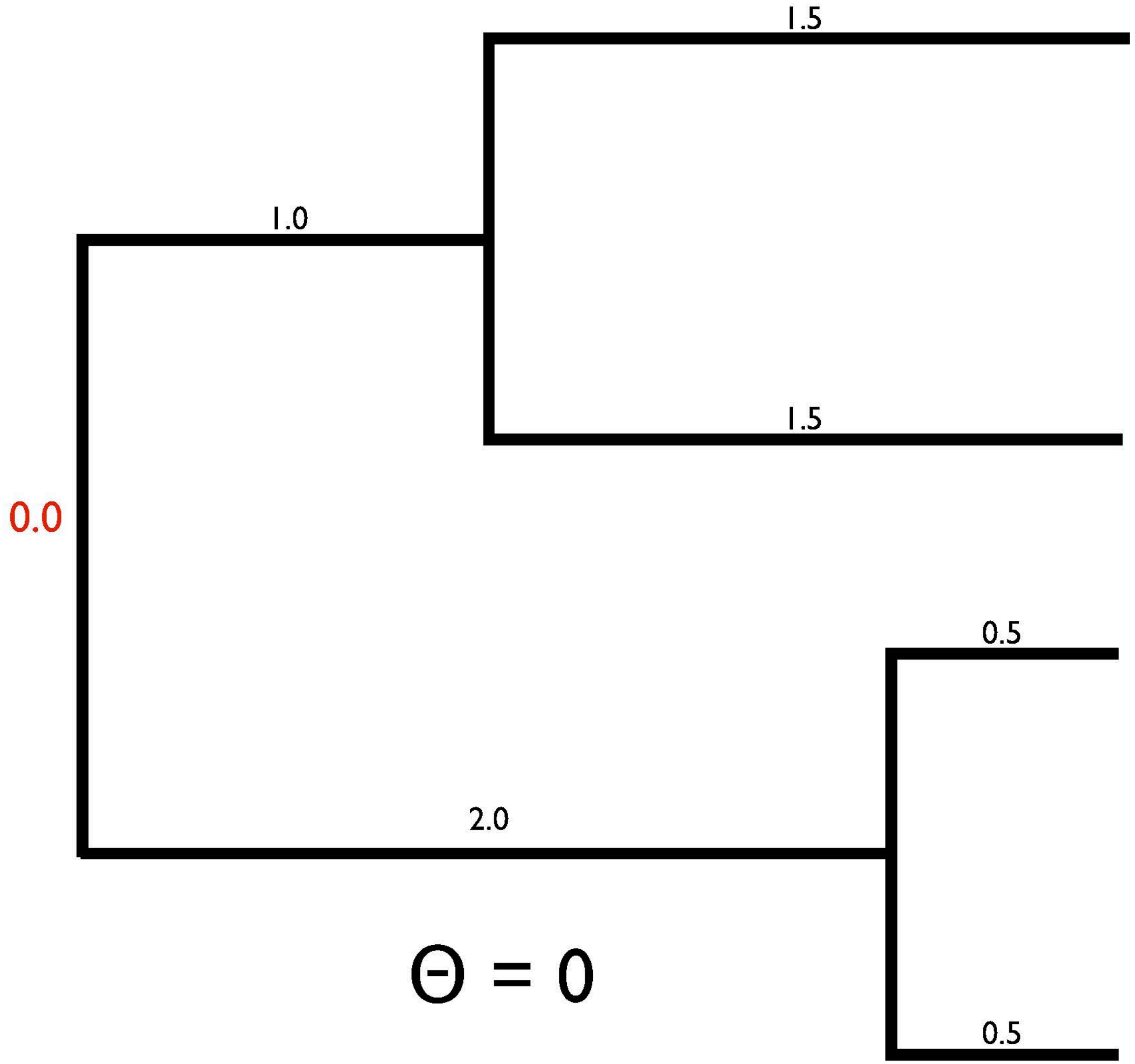
- Sortear valores de uma distribuição normal
- Variância depende do $\boldsymbol{\sigma}^2$ e t
- Valores dos ramos adjacentes são adicionados da raíz para os tips da filogenia



1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values

$\Theta = 0$

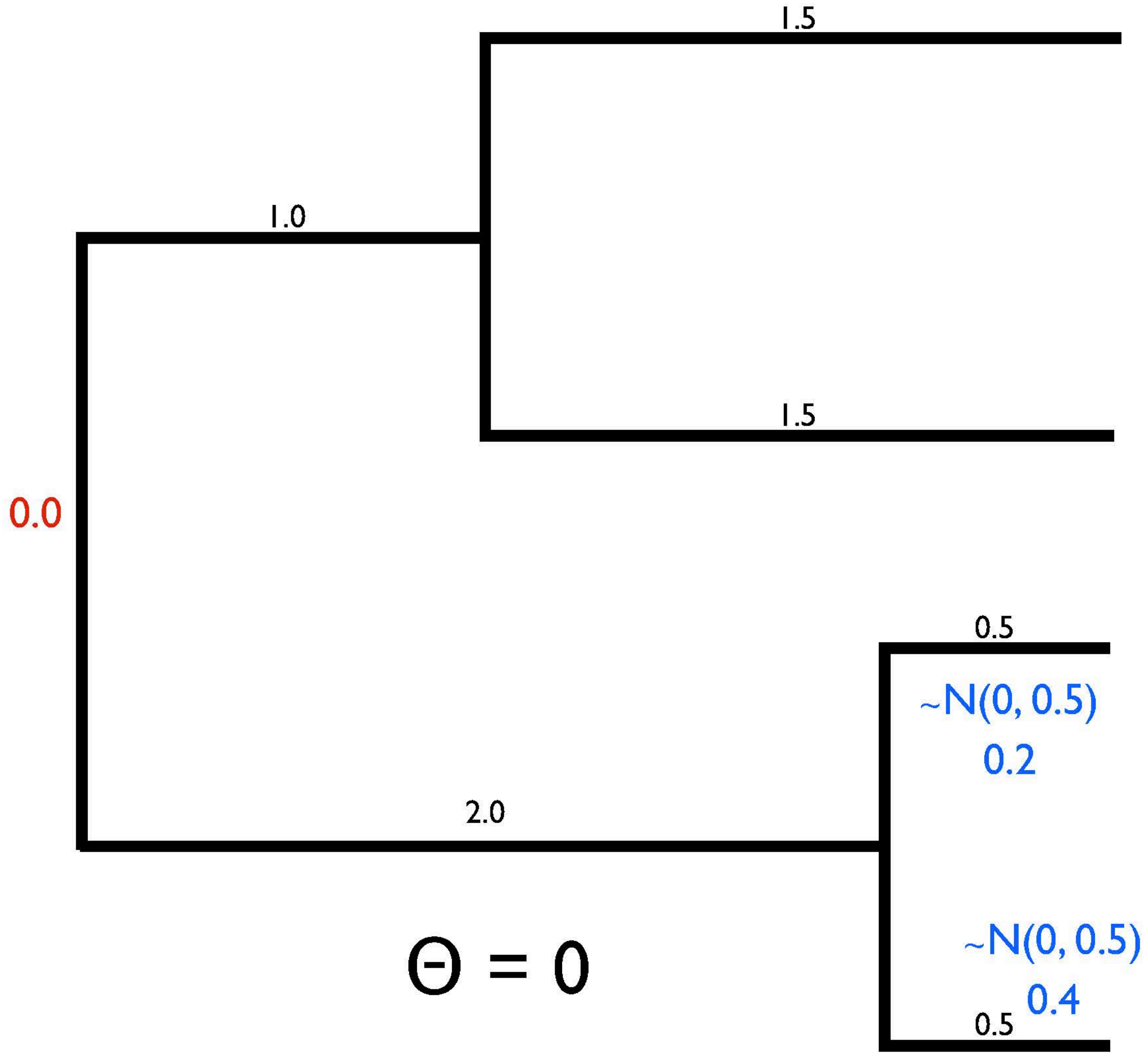
$\sigma^2 = 1.0$



$\Theta = 0$

$\sigma^2 = 1.0$

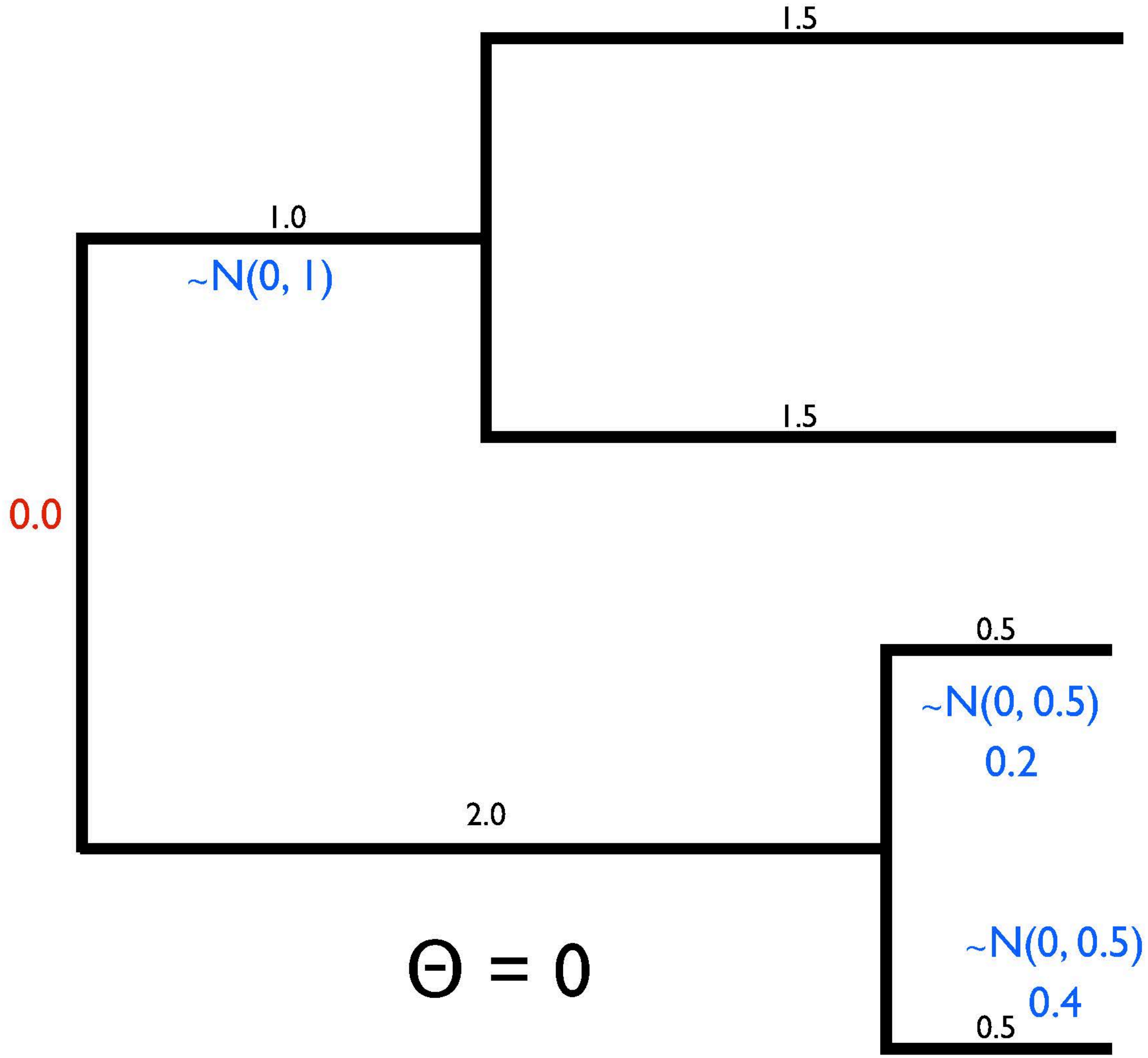
1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values



1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values

$\Theta = 0$

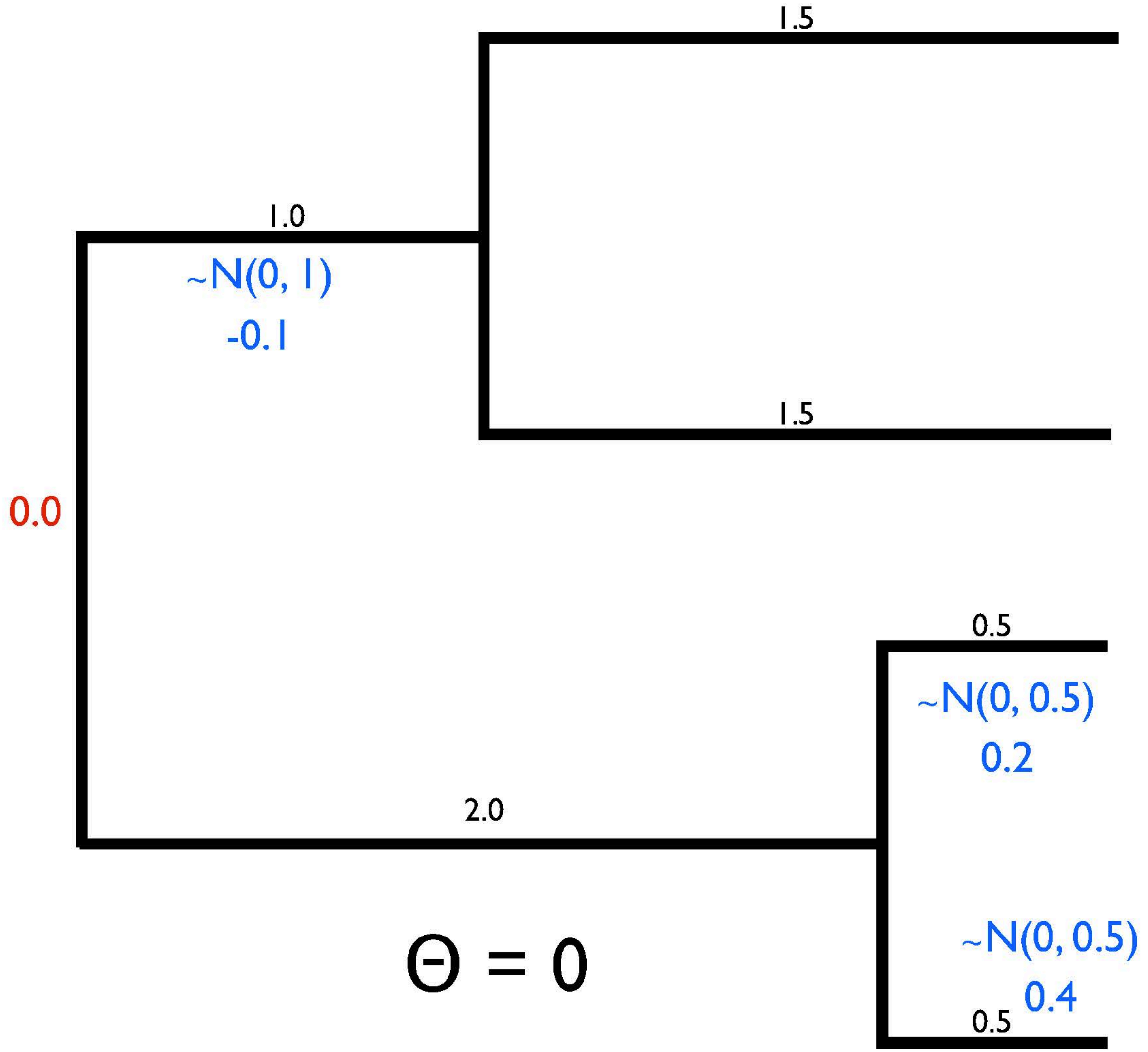
$\sigma^2 = 1.0$



$\Theta = 0$

$\sigma^2 = 1.0$

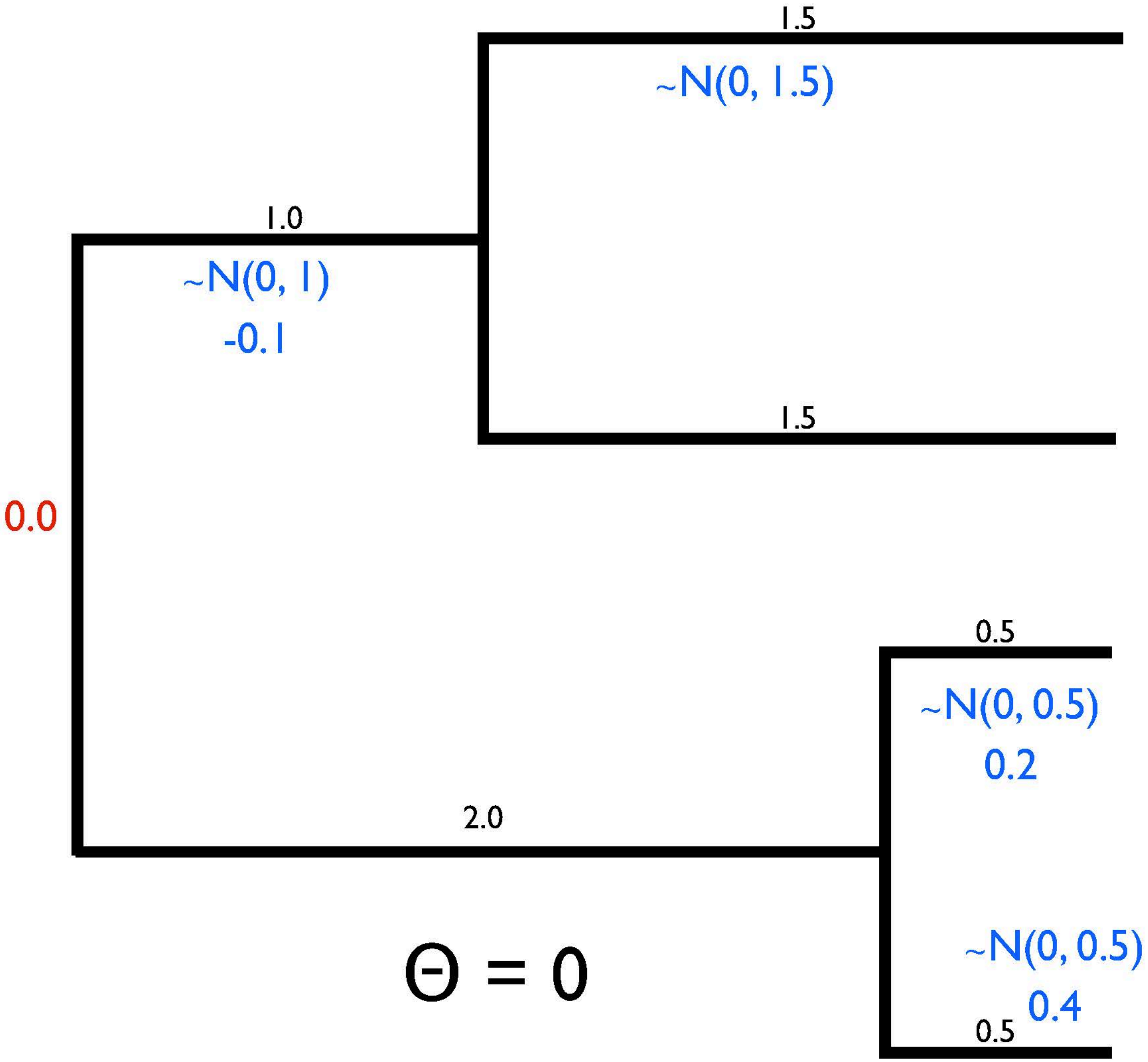
1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values



$\Theta = 0$

$\sigma^2 = 1.0$

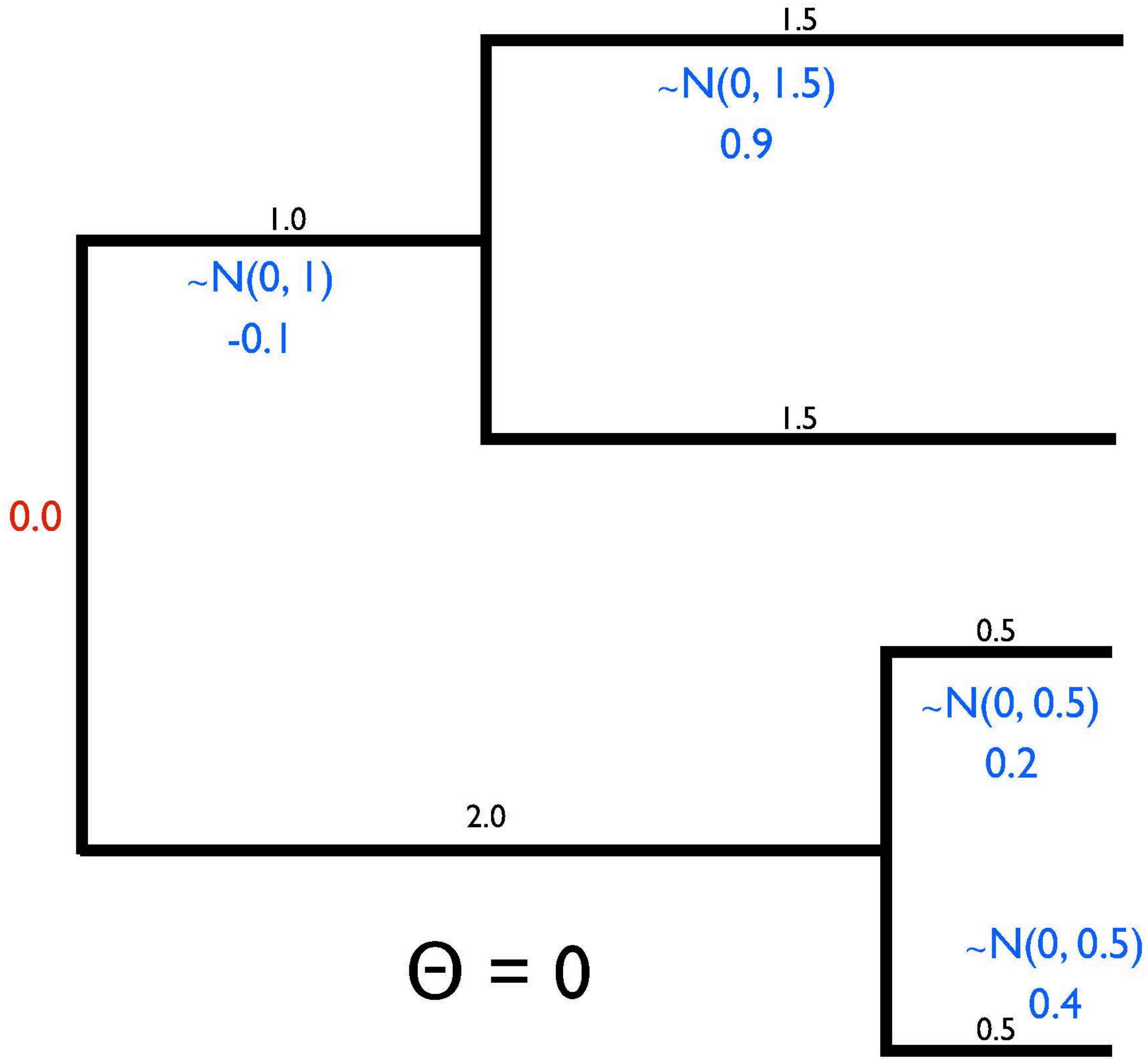
1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values



$\Theta = 0$

$\sigma^2 = 1.0$

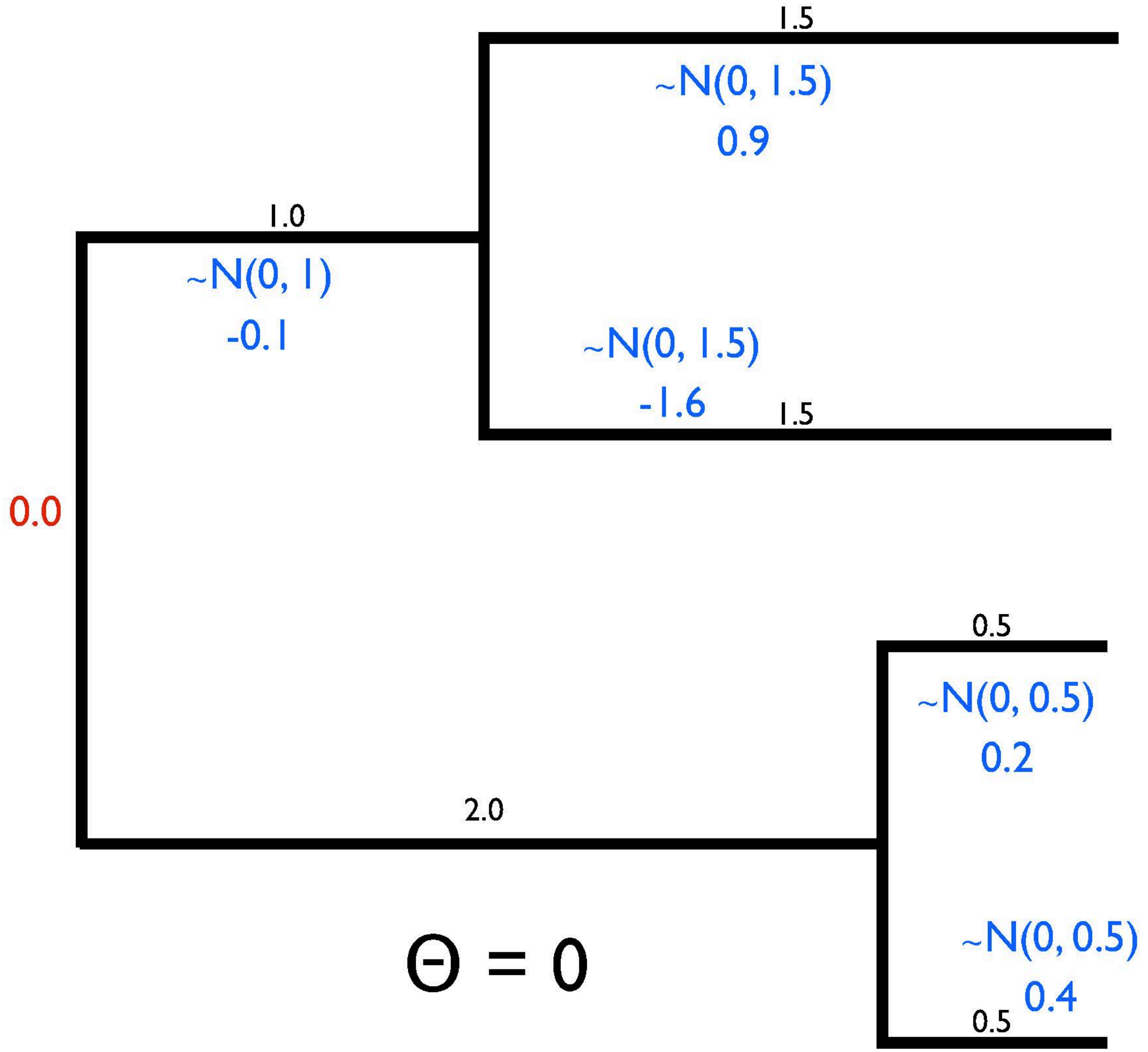
1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values



$\Theta = 0$

$\sigma^2 = 1.0$

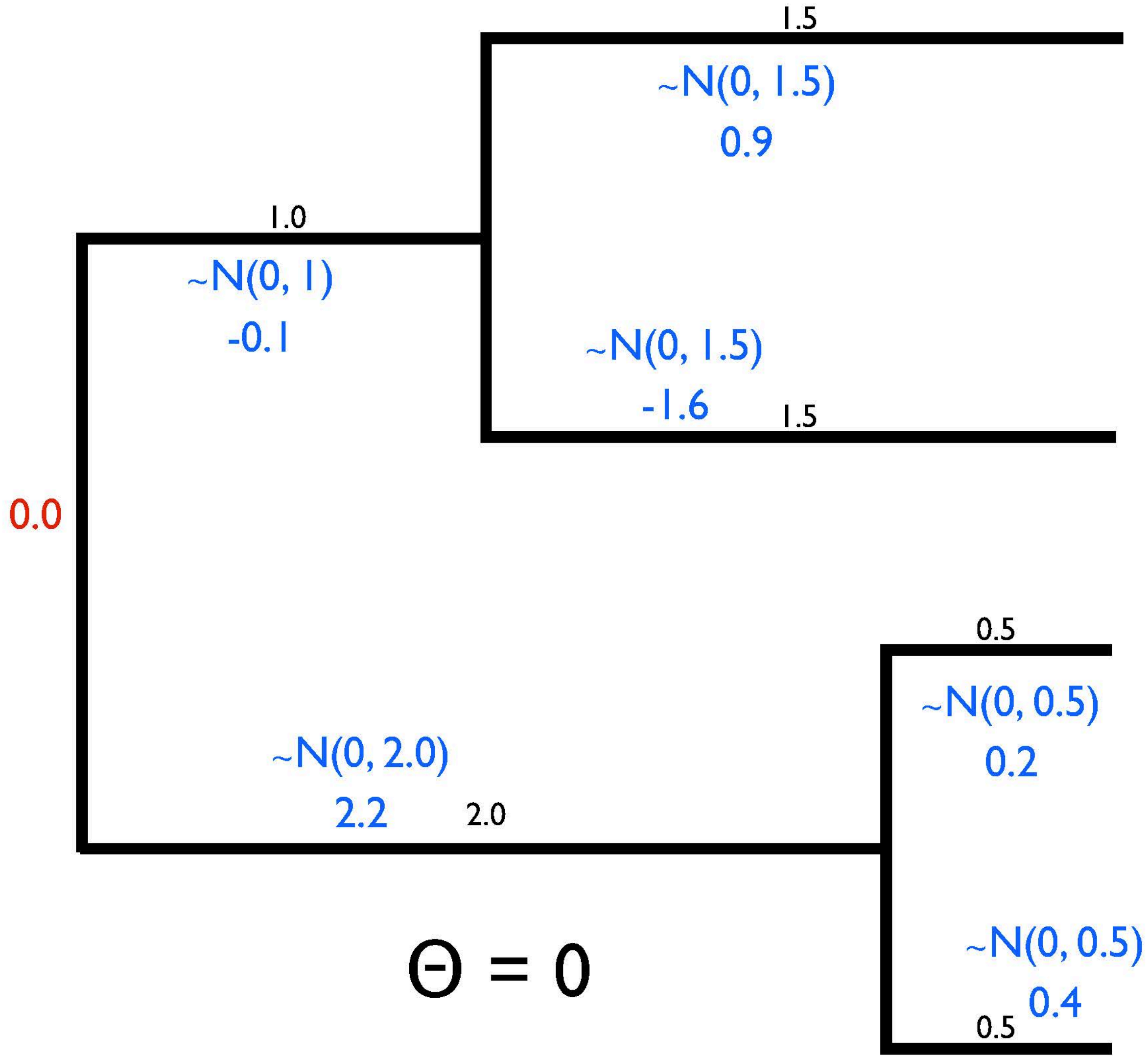
1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values



$\Theta = 0$

$\sigma^2 = 1.0$

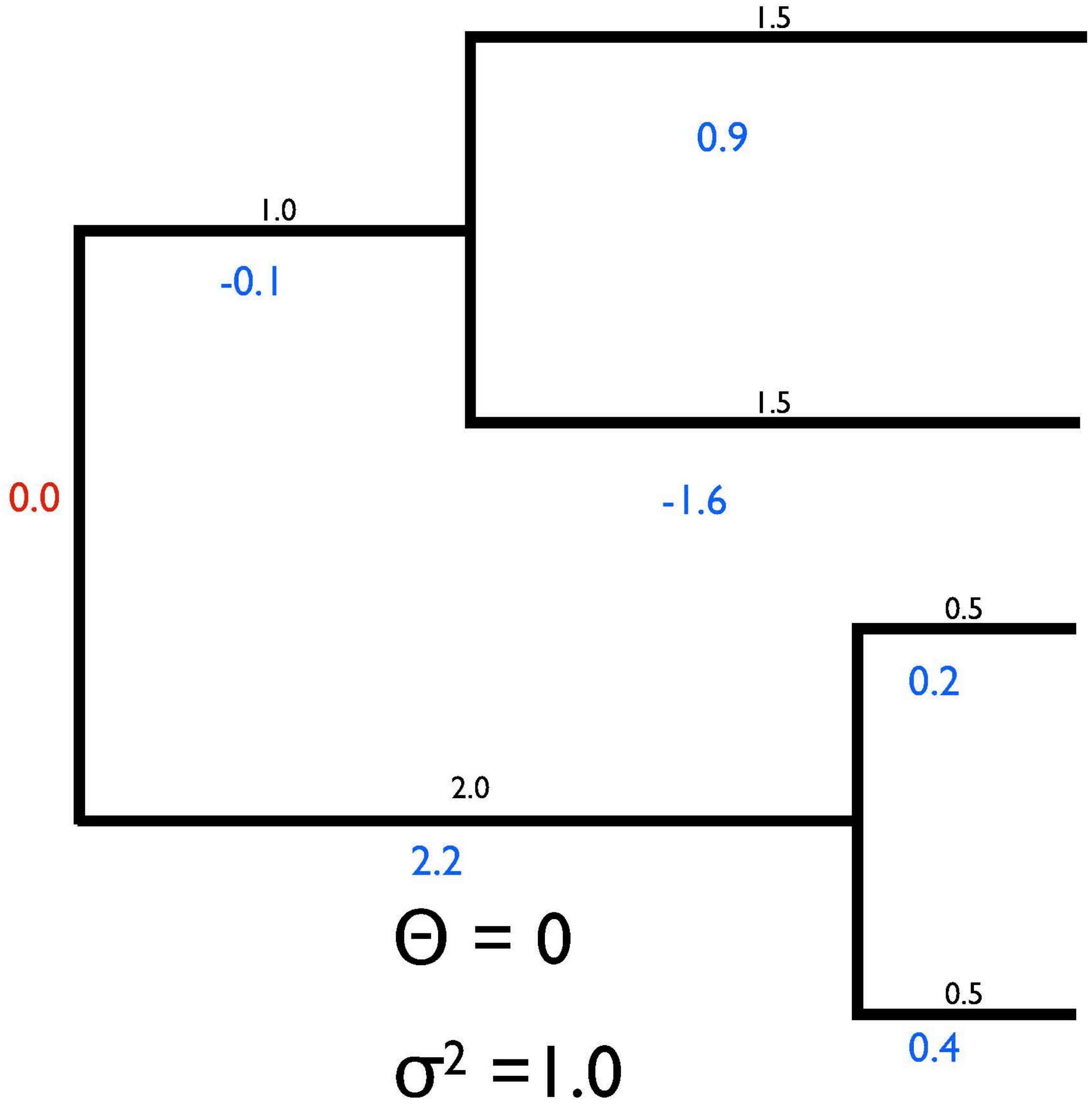
1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values



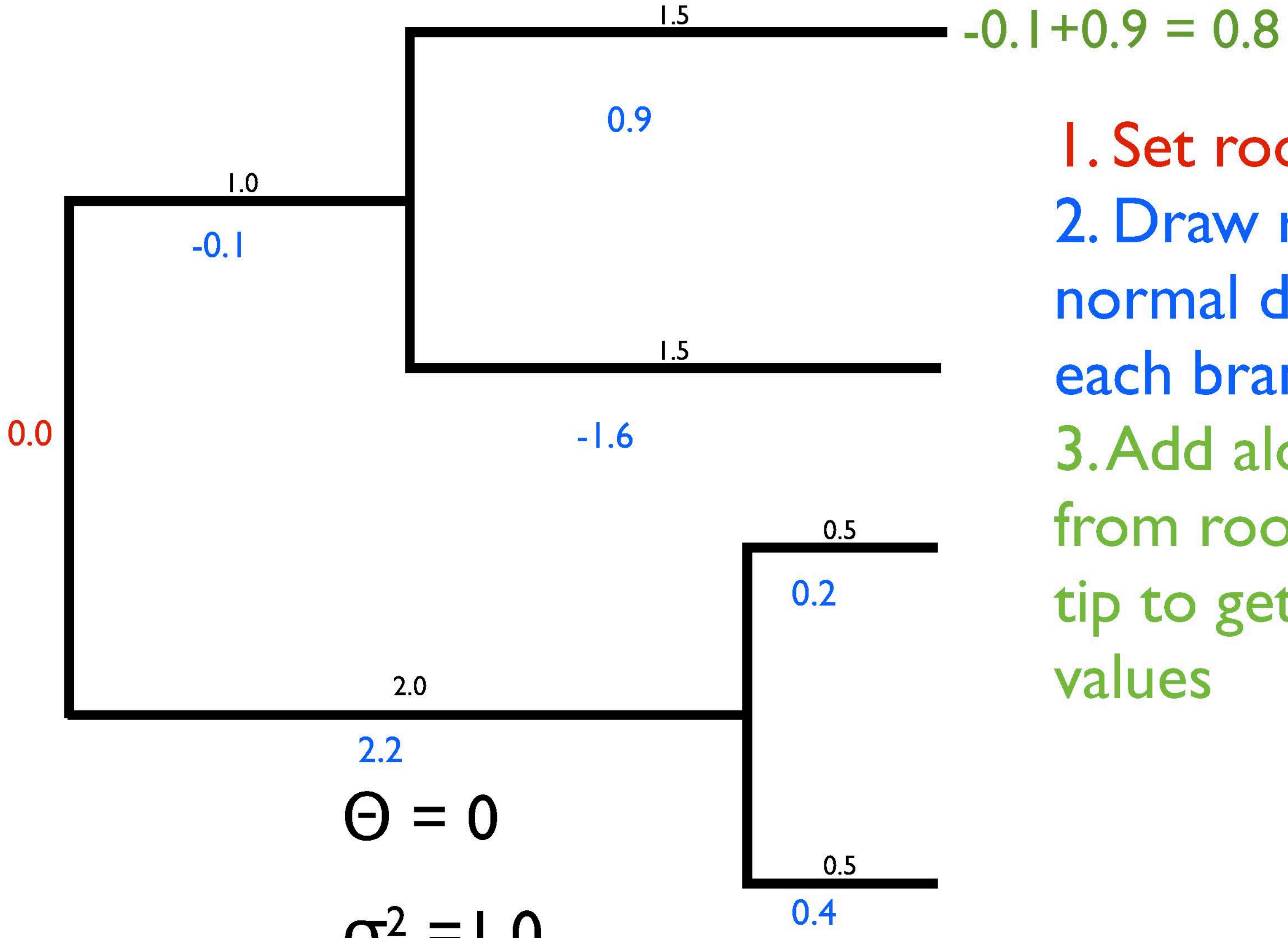
$\Theta = 0$

$\sigma^2 = 1.0$

1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values



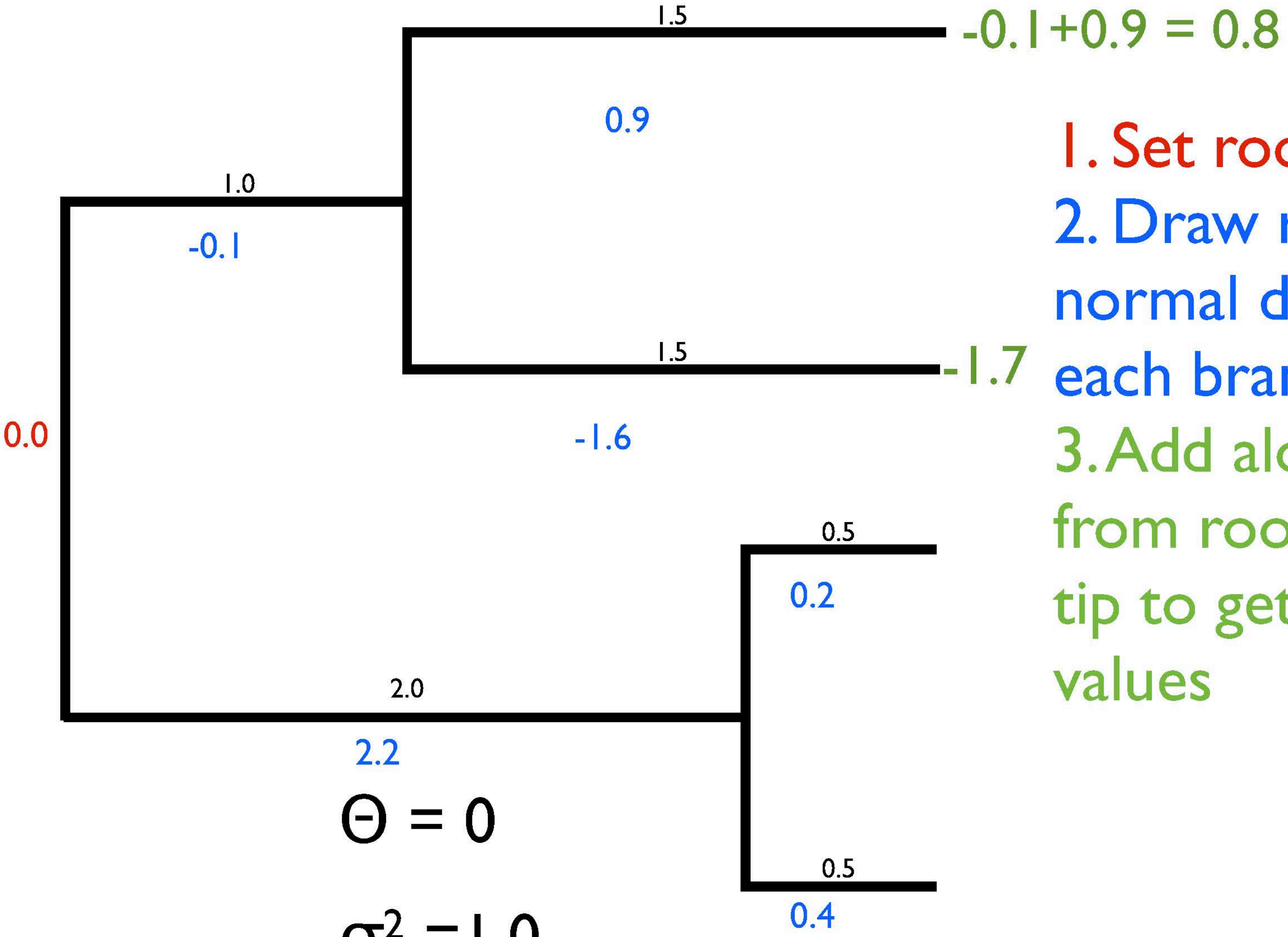
1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values



1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values

$\Theta = 0$

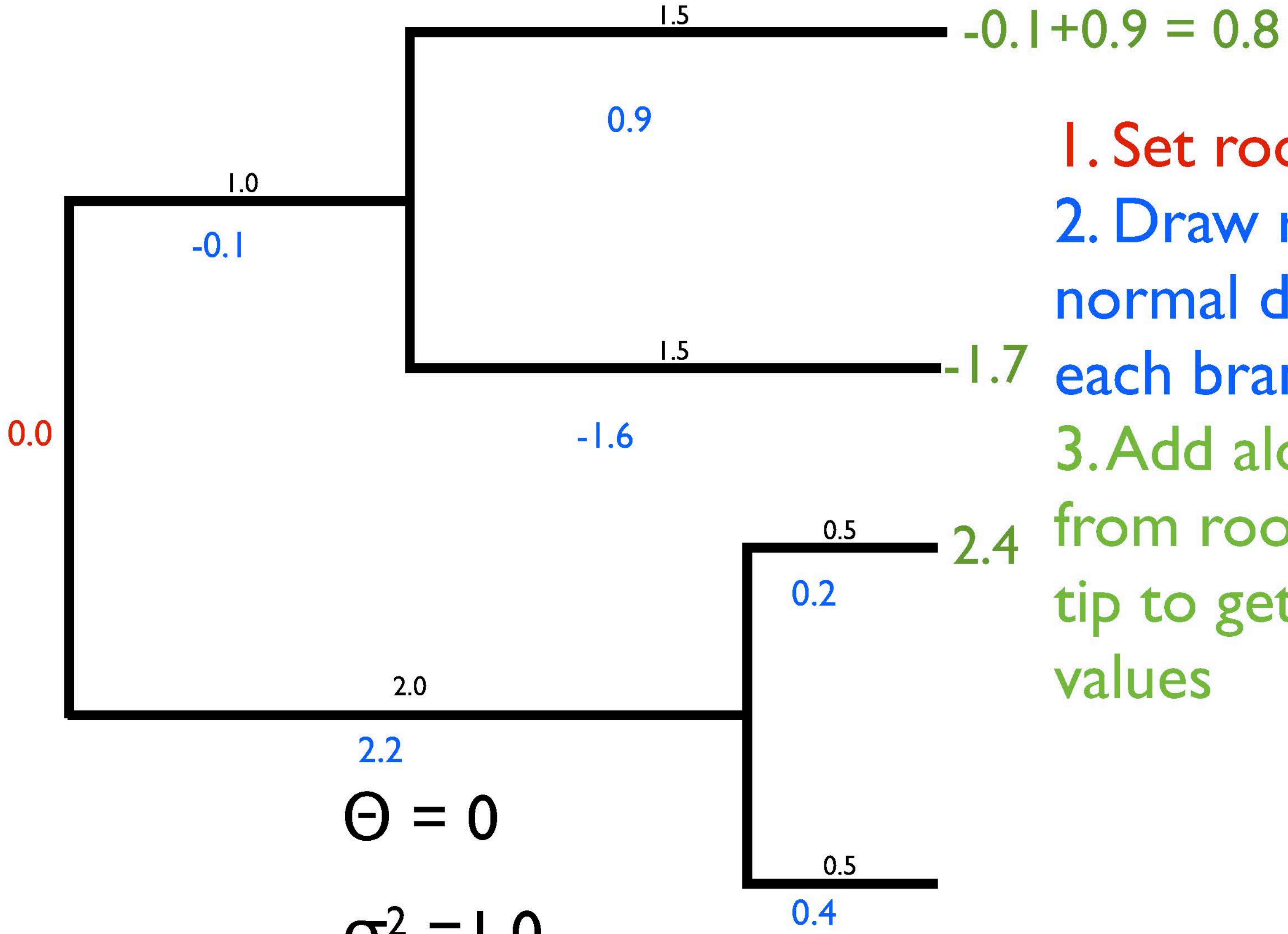
$\sigma^2 = 1.0$



$\Theta = 0$

$\sigma^2 = 1.0$

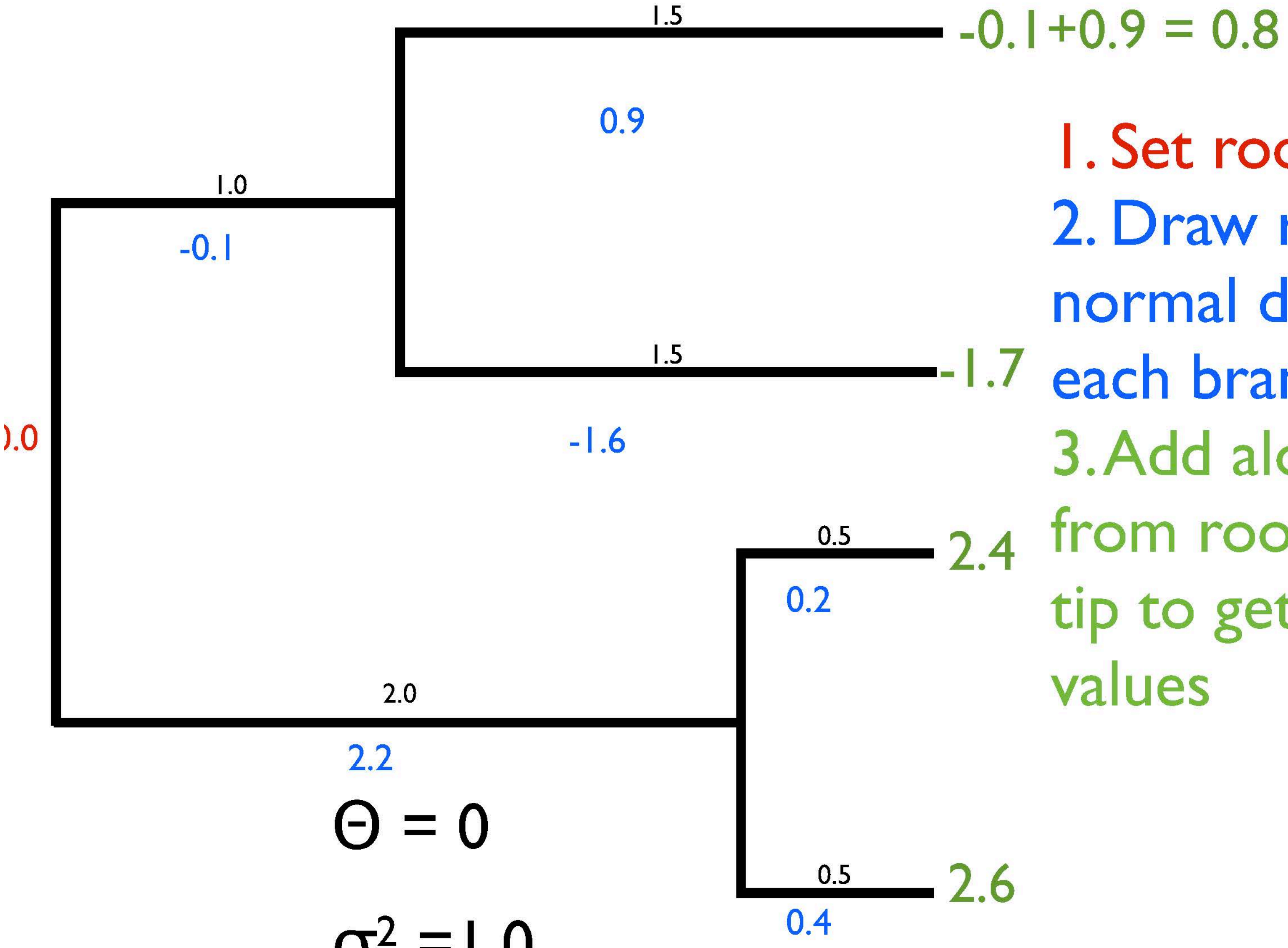
1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values



$\Theta = 0$

$\sigma^2 = 1.0$

1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values

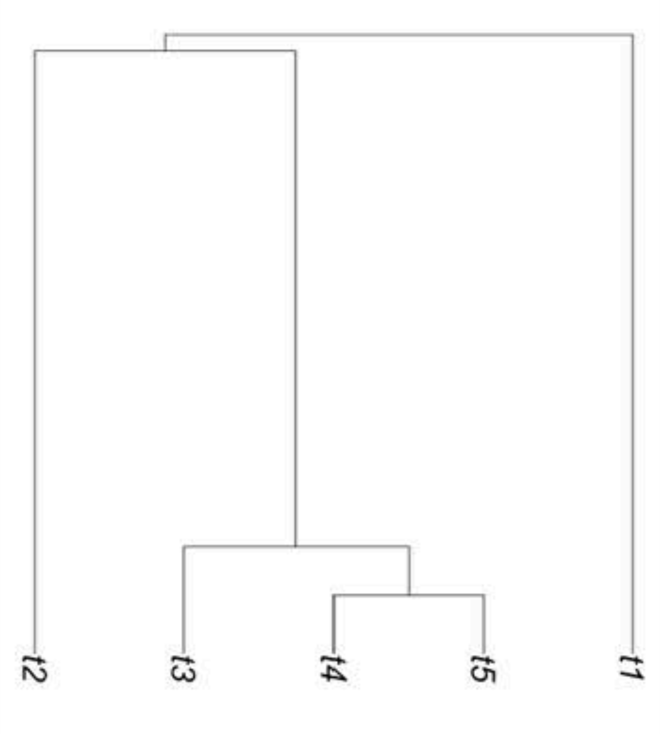


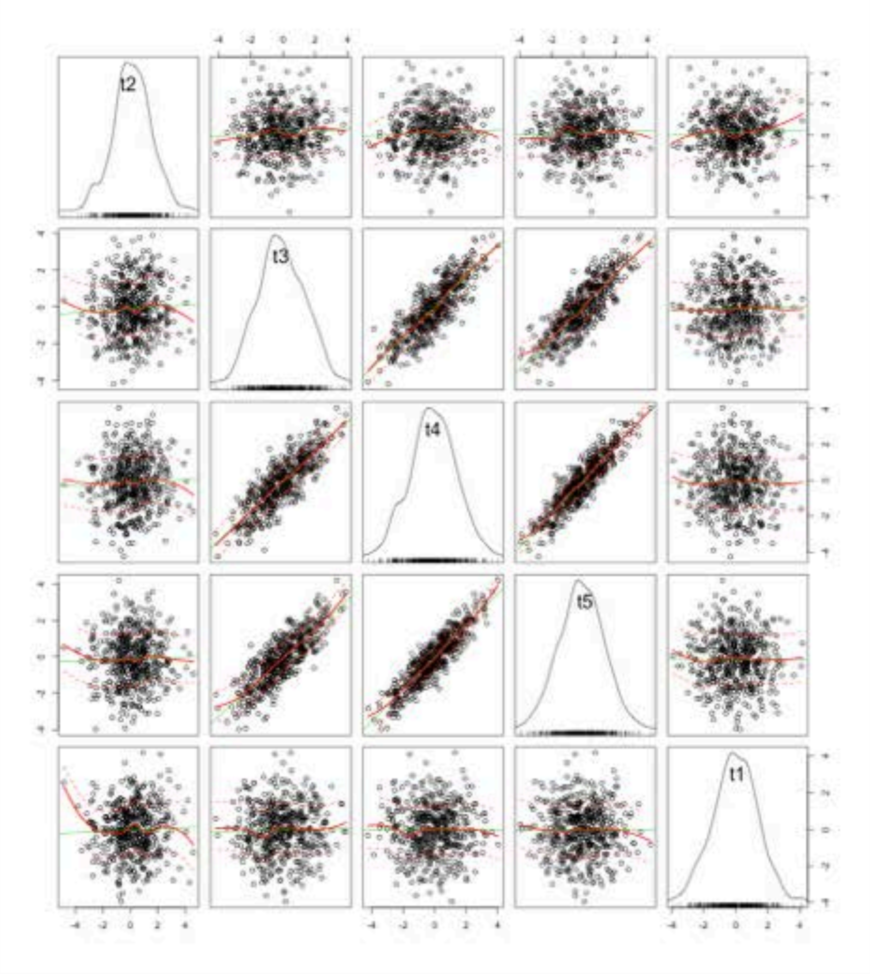
1. Set root state
2. Draw random normal deviate for each branch
3. Add along path from root to each tip to get tip values

$\Theta = 0$

$\sigma^2 = 1.0$

shared ancestry creates covariation across tips






YES!
AGORA BORA BOTAR A MÃO NA MASSA!

Ornstein-Uhlenbeck (OU)

- Captura a *idéia* de evolução estabilizadora
  - Valor de um caracter contínuo tem uma tendência a se mover para um valor médio
- Movimento Browniano com um ”salto”
- Tem 3 parâmetros
  - Valor inicial (θ)
  - Taxa ($\boldsymbol{\sigma}^2$)
  - Parâmetro de restrição (α)

Ornstein-Uhlenbeck

Leonard Ornstein

George Uhlenbeck

- Physics-based model of Brownian motion under friction
- also referred to as mean reverting
- can be modeled with attraction to a different mean than the starting state

species evolving towards adaptive peaks

that may impose limits on trait diversification

Ornstein-Uhlenbeck (OU)

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

Ornstein-Uhlenbeck (OU)

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

brownian motion

Ornstein-Uhlenbeck (OU)

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

change towards optimum

brownian motion

Ornstein-Uhlenbeck (OU)

$$dX_{(t)} = \alpha[\underset{\substack{\uparrow \\ \text{optimal value}}}{\Theta} - X_{(t)}]dt + \sigma dB_{(t)}$$

Ornstein-Uhlenbeck (OU)

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

pull towards “optimum”

Ornstein-Uhlenbeck (OU)

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

strength of selection is proportional to distance of trait from optimal value

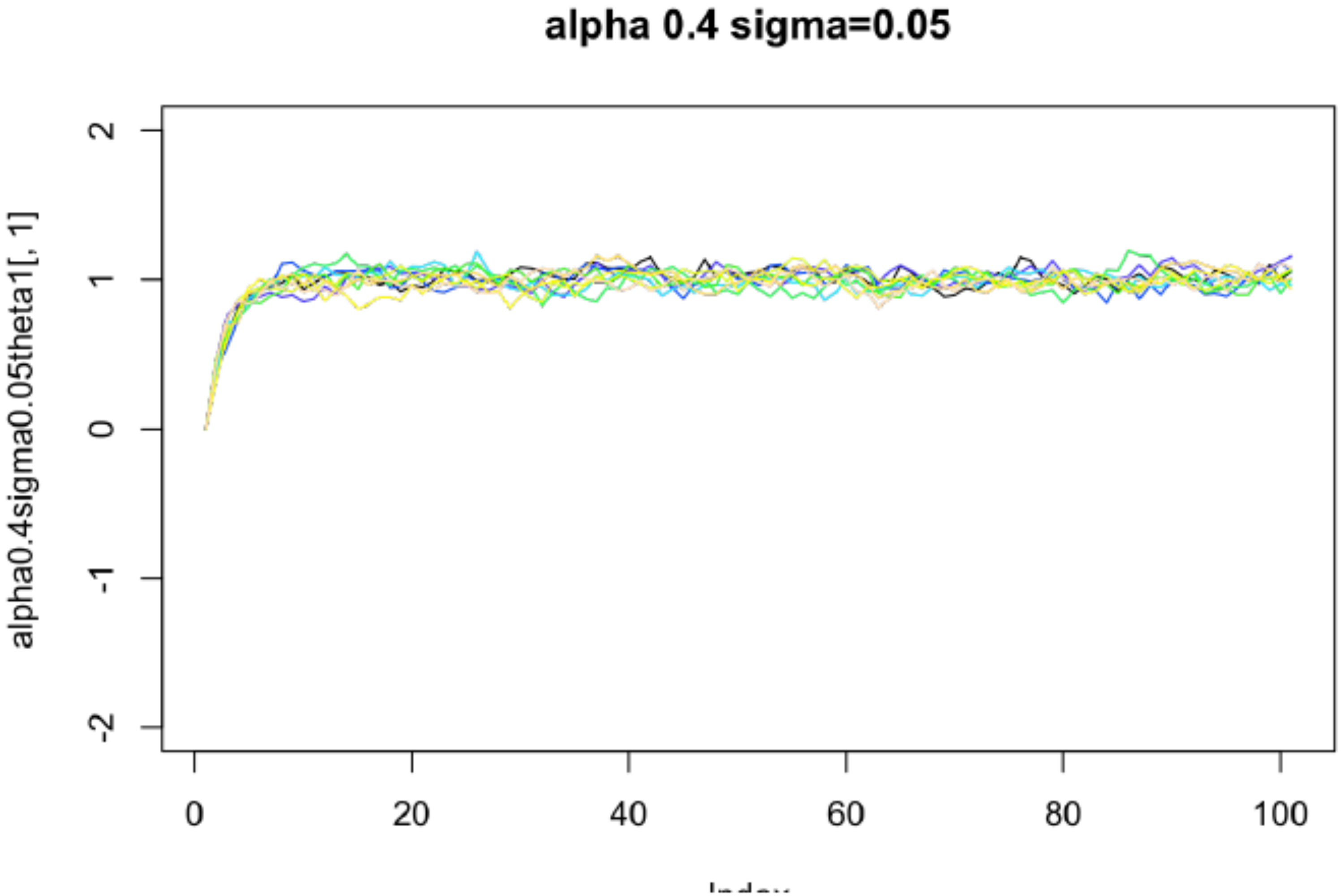
Quanto maior a distância entre valor do trait e o ótimo → Maior o ”puxão” em direção ao ótimo

Ornstein-Uhlenbeck (OU)

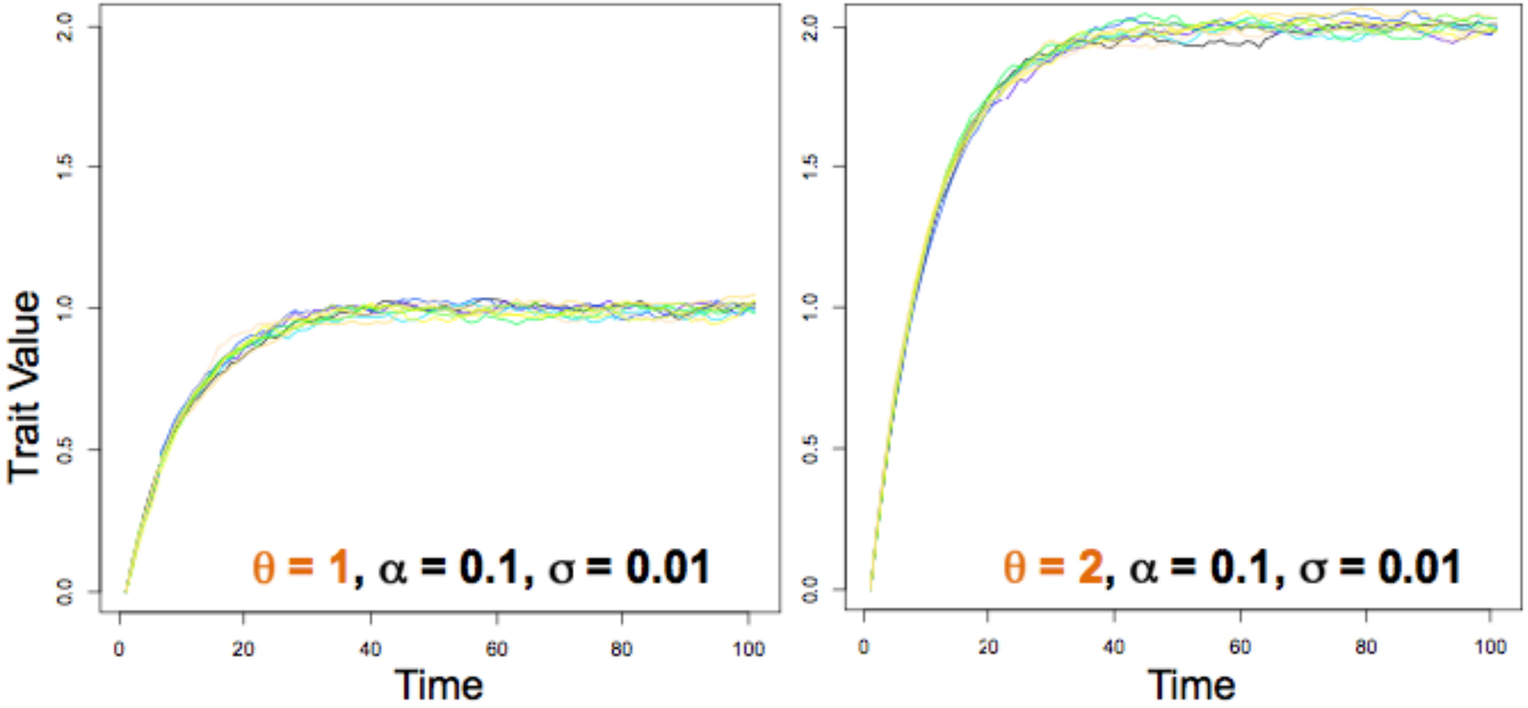
$$dX_{(t)} = \mathbf{0}[\Theta - X_{(t)}]dt + \sigma dB_{(t)}$$

when alpha is 0, OU
becomes BM

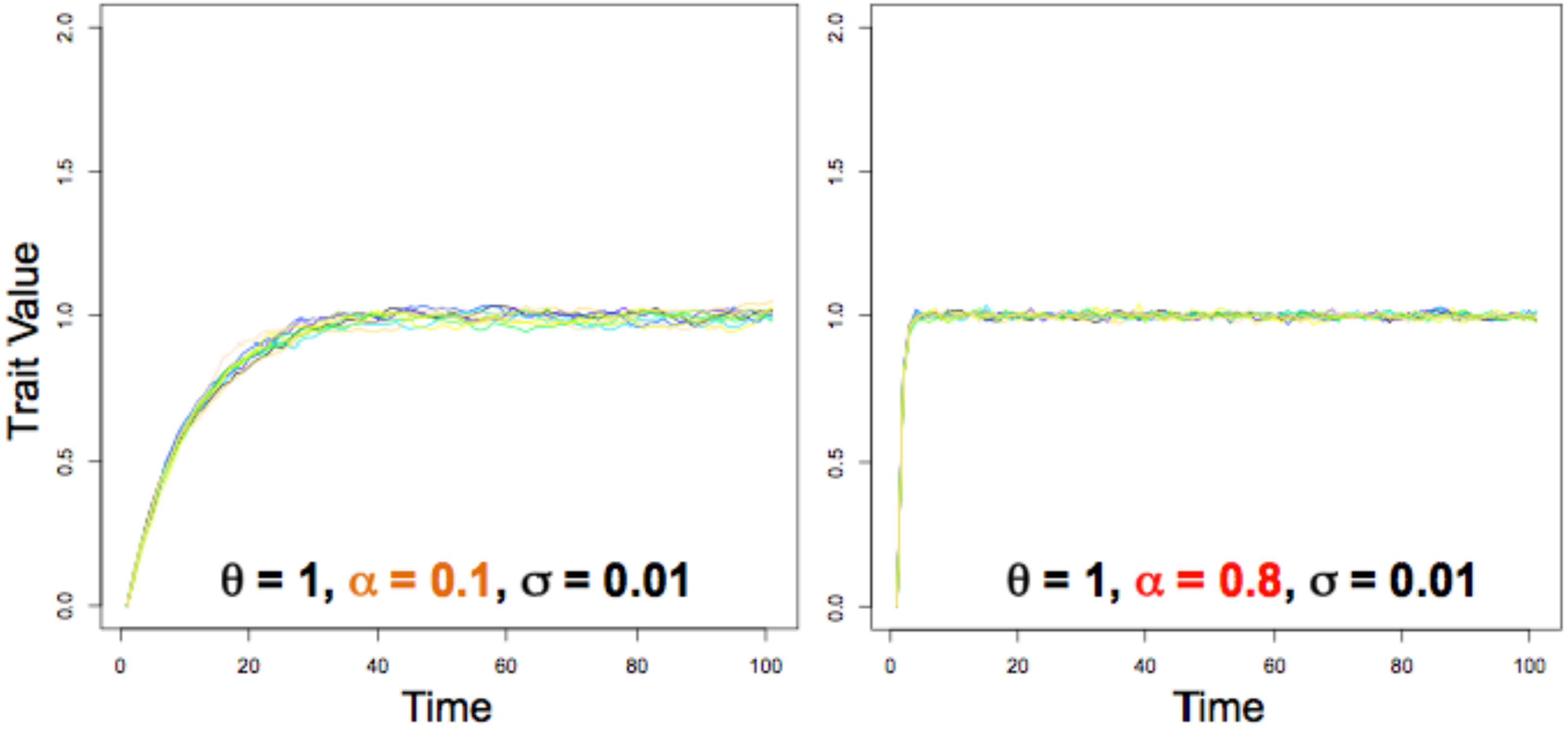
OU evolution



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

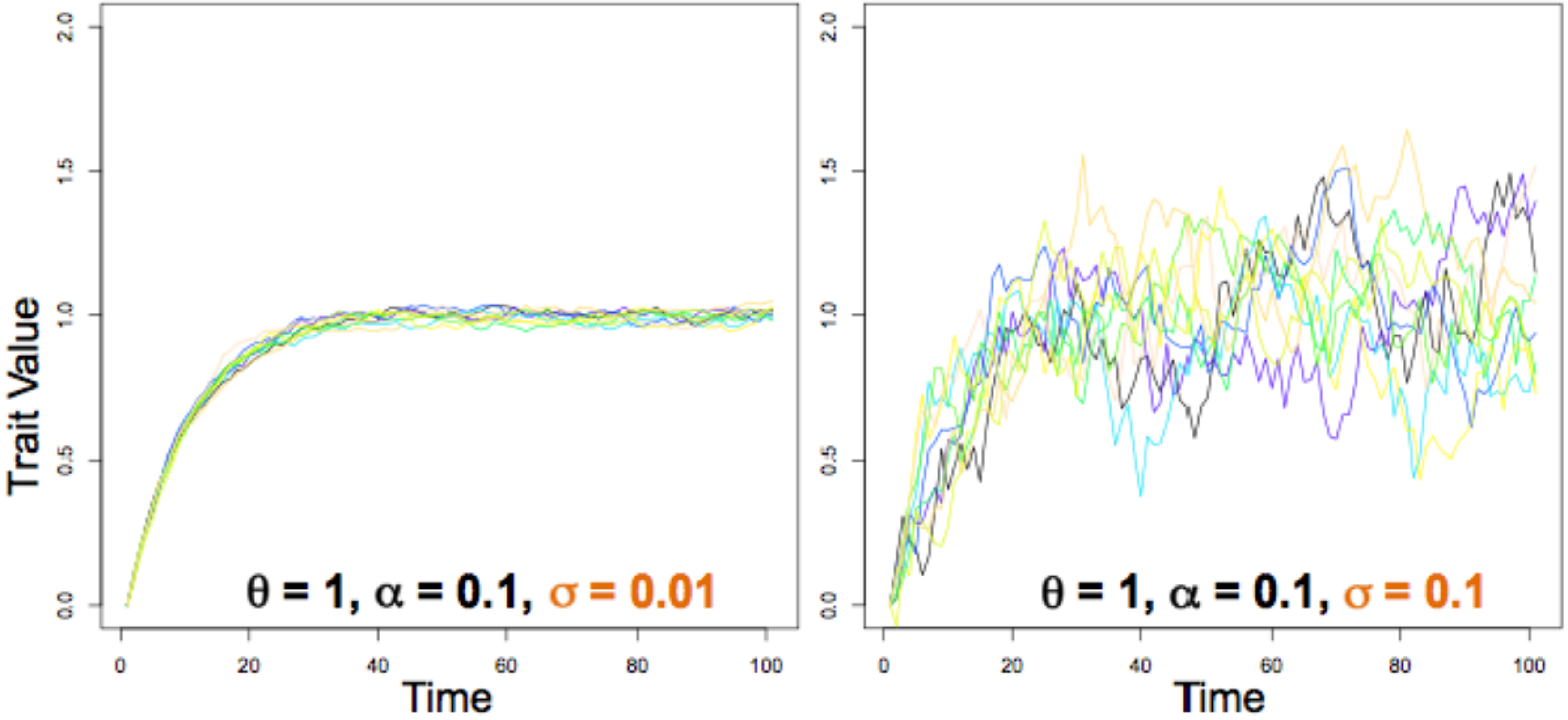


$$dX_{(t)} = \alpha[\theta - X_{(t)}]dt + \sigma dB_{(t)}$$

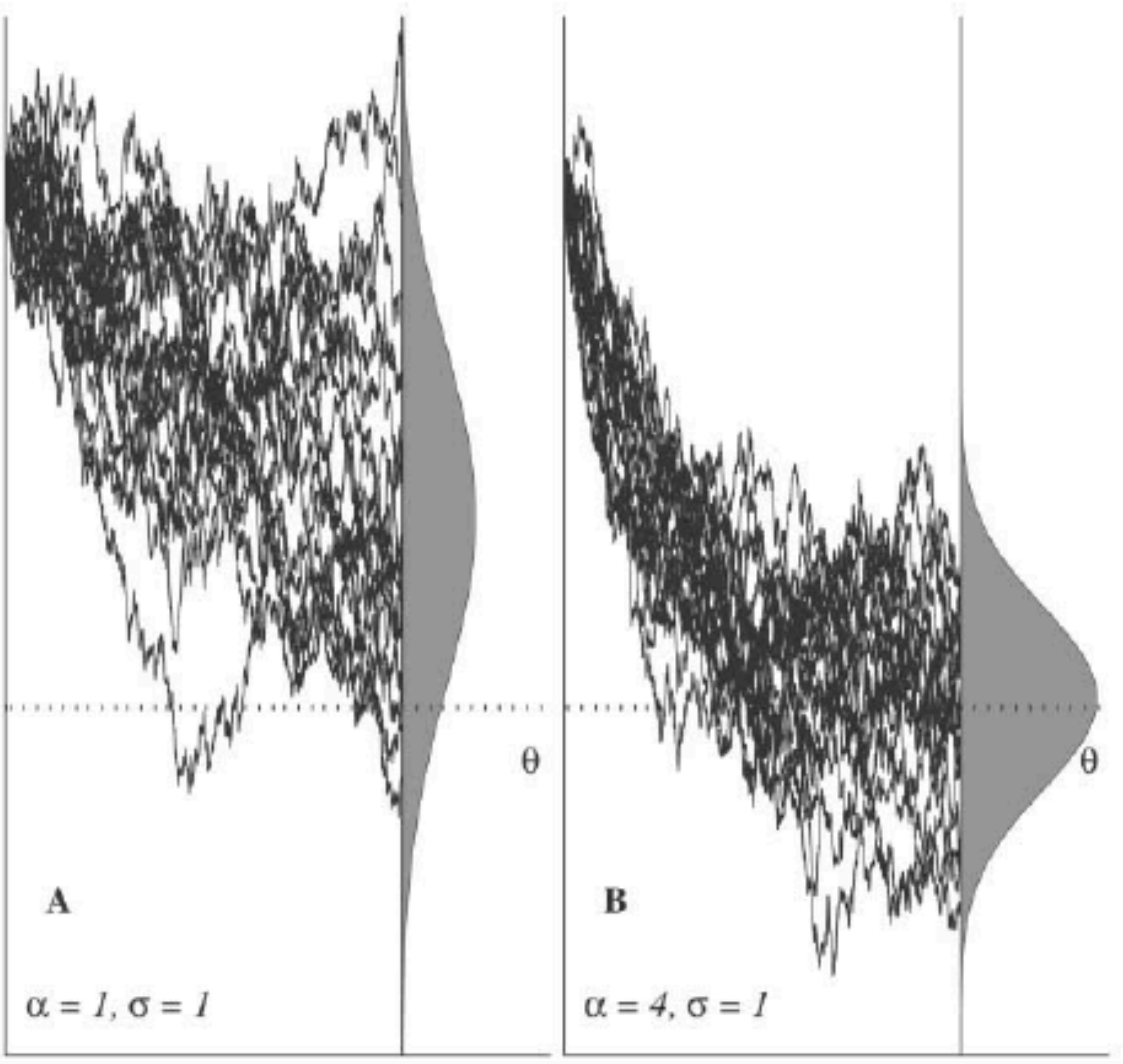


Stronger α the more quickly the optima is reached and the lower variance

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



Higher σ the greater the variance



Butler & King 2004

Early Burst (EB)

- Taxa de evolução desacelera com o tempo
- Maior taxa é na raiz da filogenia (valor inicial)
- Corresponde à idéia de **radiação adaptativa**
  - Espera-se que linhagens que entram em novas zonas adaptativas diversifiquem rápido
  - Taxas diminuem à medida que nichos são preenchidos
- Descrito por 3 parâmetros:
  - Valor inicial (θ)
  - Taxa inicial ($\boldsymbol{\sigma}^2_o$)
  - Taxa de mudança (r)

$$V_{ij} = \int_0^{s_{ij}} \sigma_o^2 e^{rt}\, dt = \sigma_o^2 \left(\frac{e^{r s_{ij}} - 1}{r} \right).$$

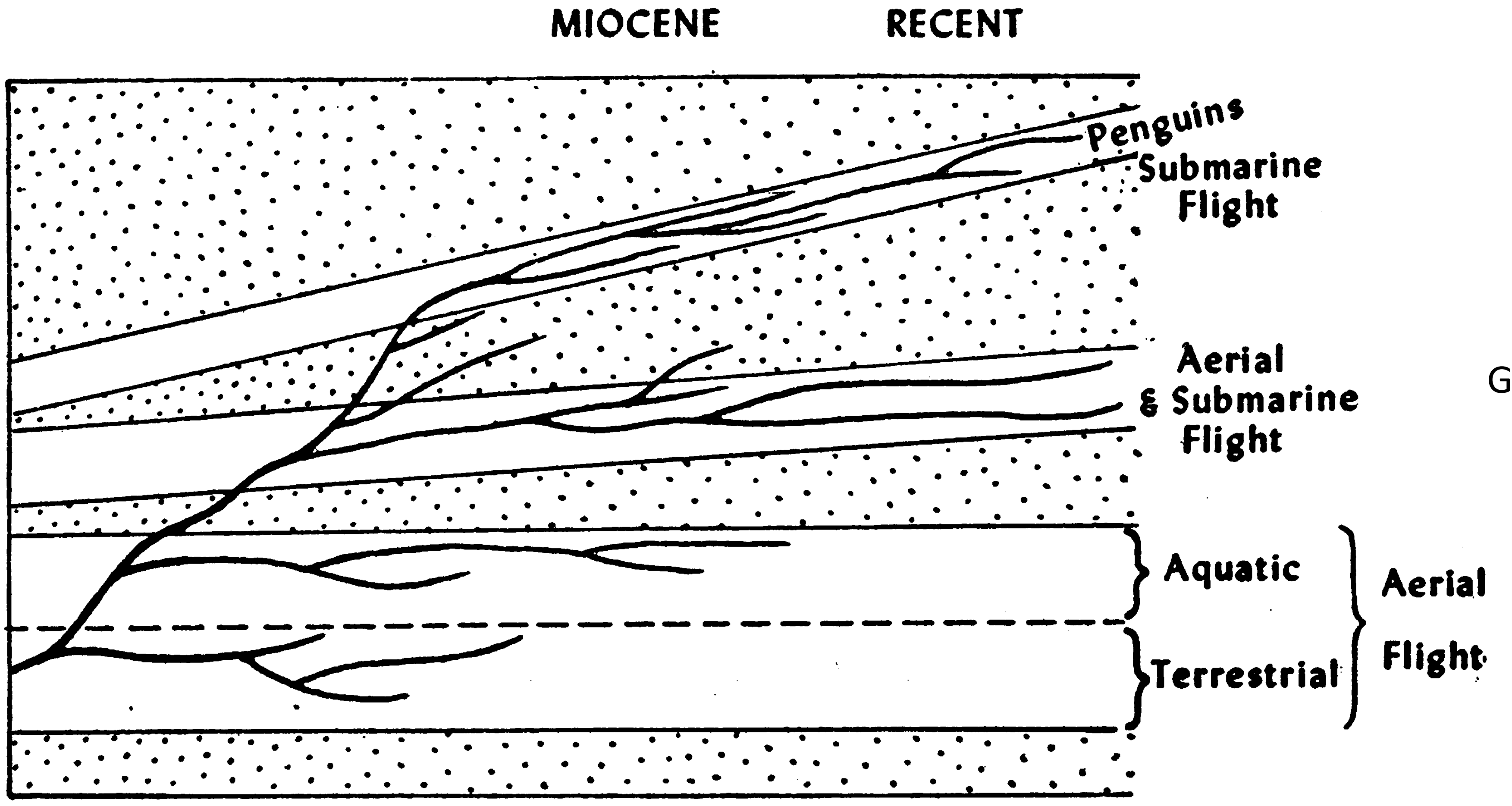


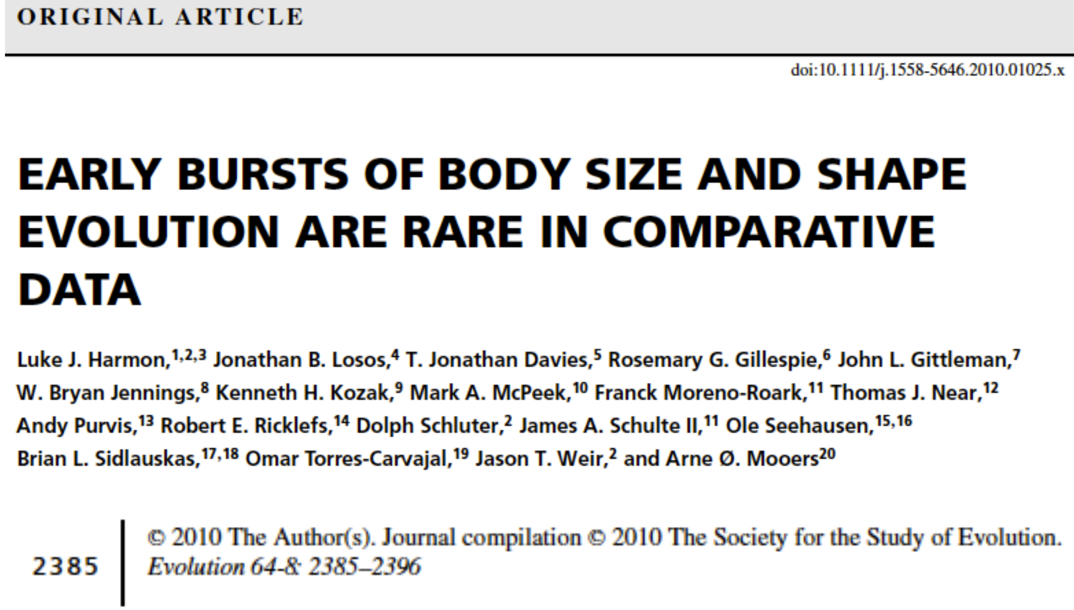
FIG. 31. Grid diagram of the theory of penguin evolution proposed in this paper.

George Gaylord Simpson

Evolution, 57(4), 2003, pp. 717–745

TESTING FOR PHYLOGENETIC SIGNAL IN COMPARATIVE DATA: BEHAVIORAL TRAITS ARE MORE LABILE

SIMON P. BLOMBERG,[1] THEODORE GARLAND, JR.,[1,2] AND ANTHONY R. IVES[3,4]



ORIGINAL ARTICLE

doi:10.1111/j.1558-5646.2010.01025.x

EARLY BURSTS OF BODY SIZE AND SHAPE EVOLUTION ARE RARE IN COMPARATIVE DATA

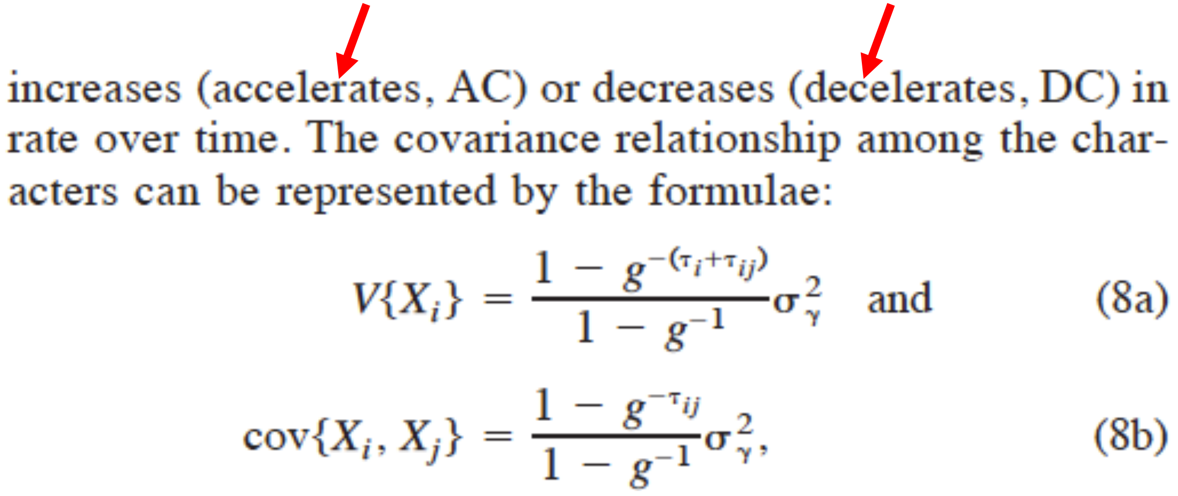
Luke J. Harmon,[1,2,3] Jonathan B. Losos,[4] T. Jonathan Davies,[5] Rosemary G. Gillespie,[6] John L. Gittleman,[7] W. Bryan Jennings,[8] Kenneth H. Kozak,[9] Mark A. McPeek,[10] Franck Moreno-Roark,[11] Thomas J. Near,[12] Andy Purvis,[13] Robert E. Ricklefs,[14] Dolph Schluter,[2] James A. Schulte II,[11] Ole Seehausen,[15,16] Brian L. Sidlauskas,[17,18] Omar Torres-Carvajal,[19] Jason T. Weir,[2] and Arne Ø. Mooers[20]

2385

© 2010 The Author(s). Journal compilation © 2010 The Society for the Study of Evolution.
Evolution 64-8: 2385–2396

Early burst

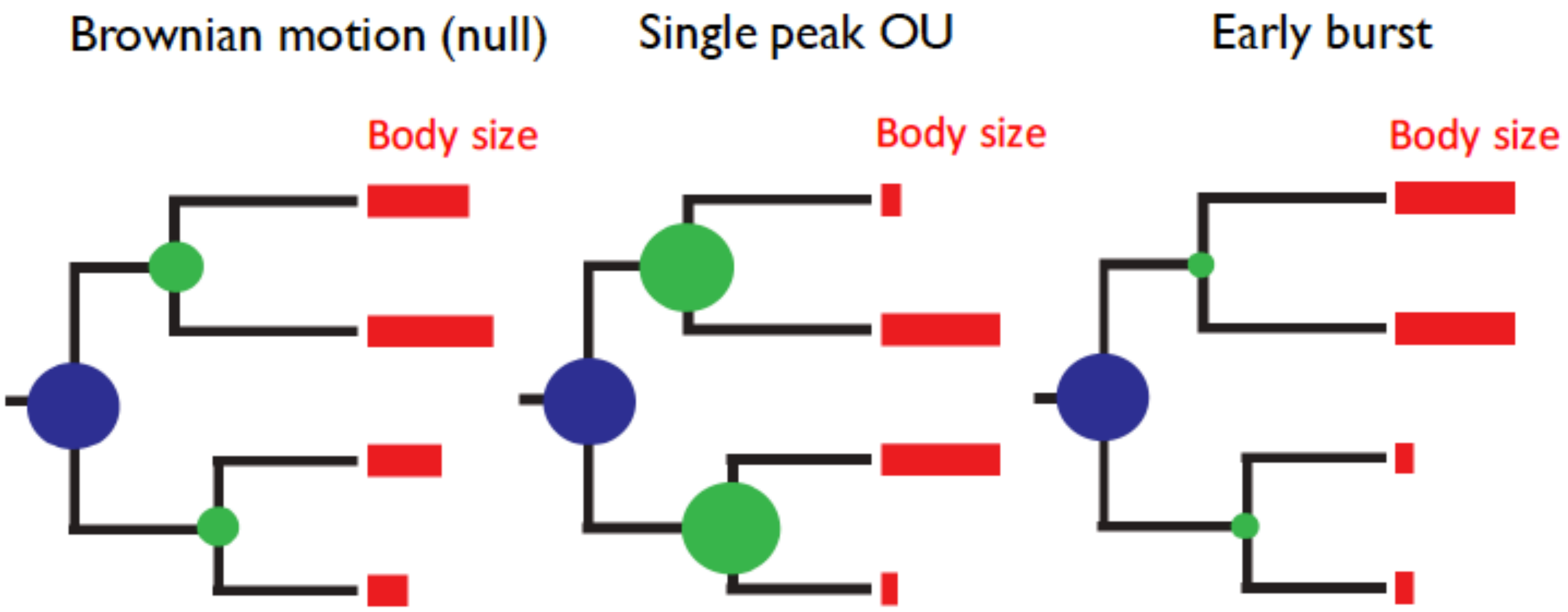
Late burst



increases (accelerates, AC) or decreases (decelerates, DC) in rate over time. The covariance relationship among the characters can be represented by the formulae:

$$V\{X_i\} = \frac{1 - g^{-(\tau_i+\tau_{ij})}}{1 - g^{-1}}\sigma_\gamma^2 \quad \text{and} \tag{8a}$$

$$\text{cov}\{X_i, X_j\} = \frac{1 - g^{-\tau_{ij}}}{1 - g^{-1}}\sigma_\gamma^2, \tag{8b}$$



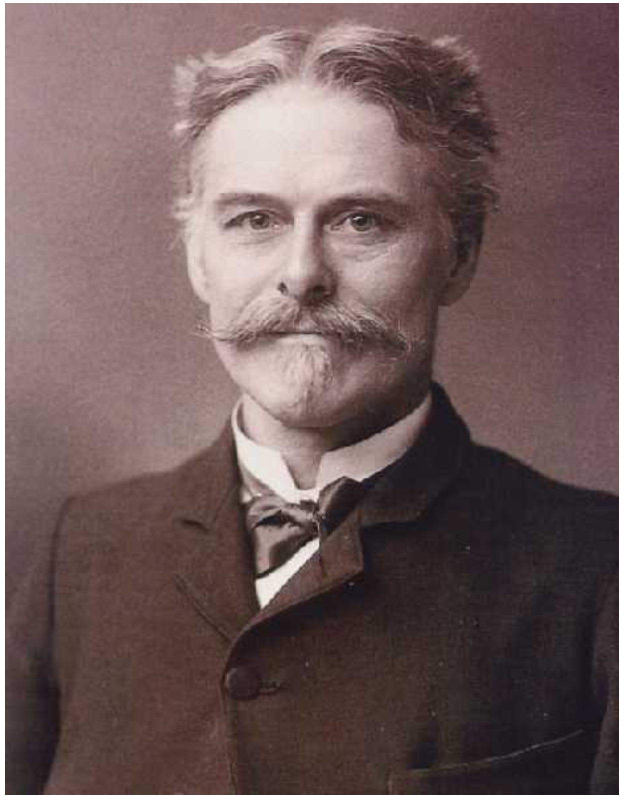
● Amount of variation in the whole clade

● Amount of variation within subclades

Harmon et al. (2010) - Evolution

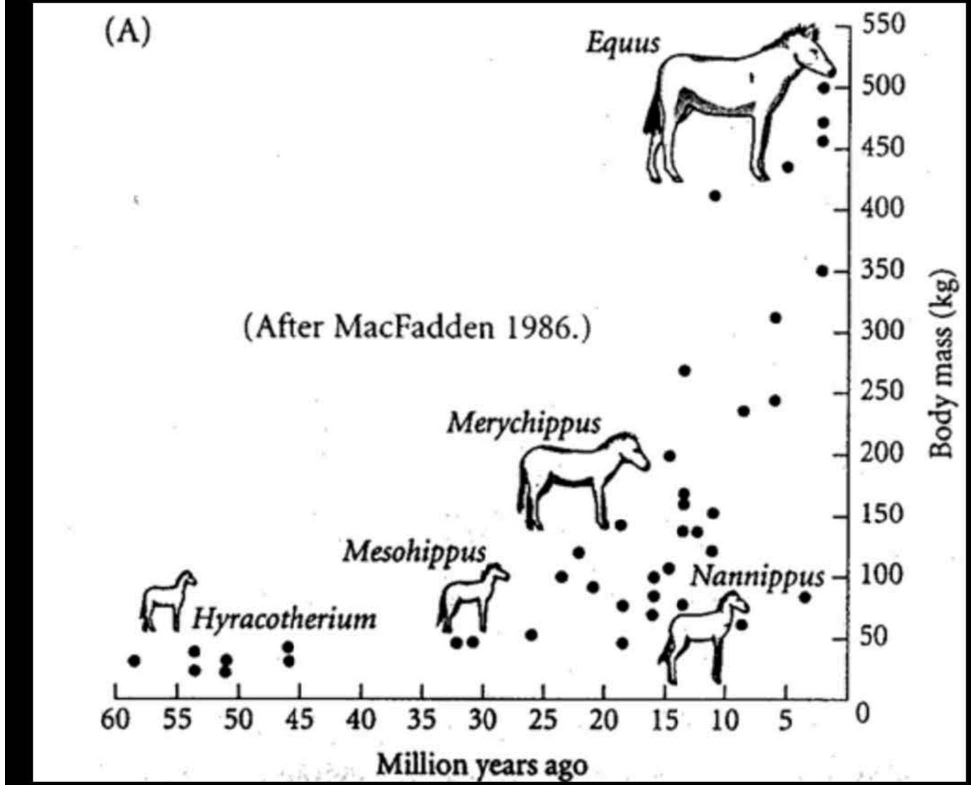
Movimento Browniano com tendência (Trend)

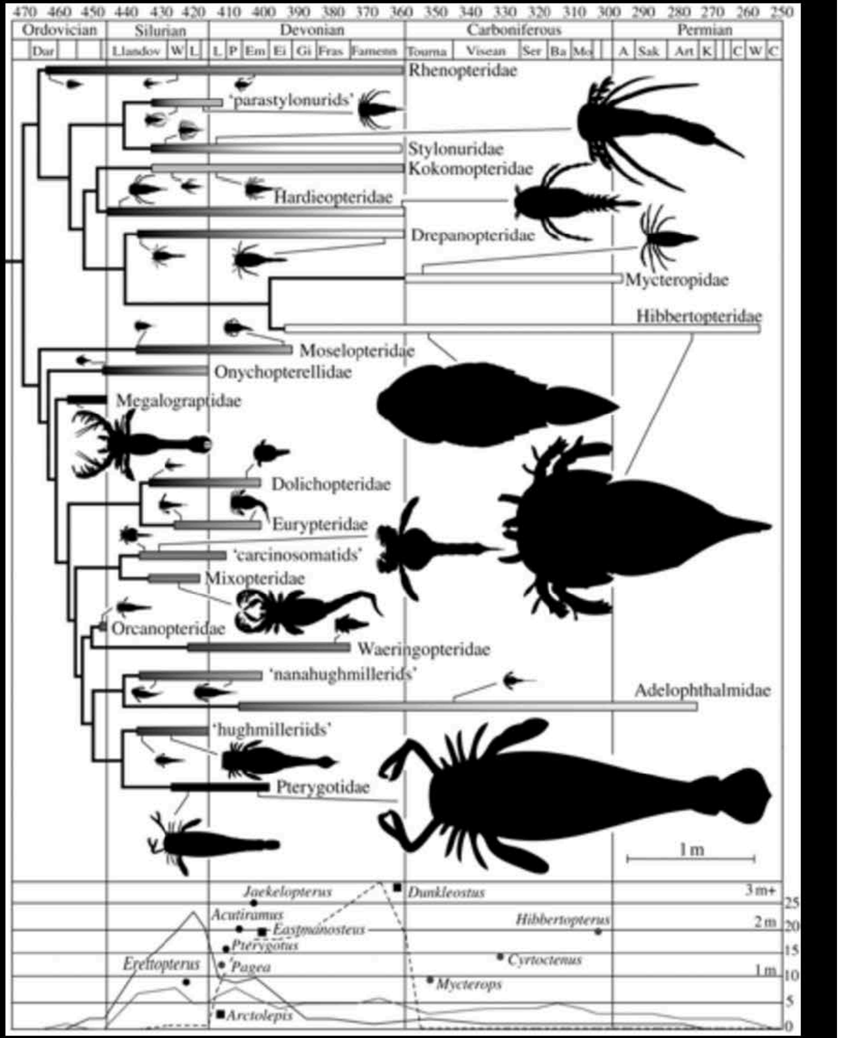
- Incorpora a idéia de mudança tendenciada do fenótipo (atributos) ao longo do tempo
    - Regra de Cope => aumento do tamanho de corpo ao longo do tempo



Trend

Species within lineages tend to get bigger over time (Cope's rule

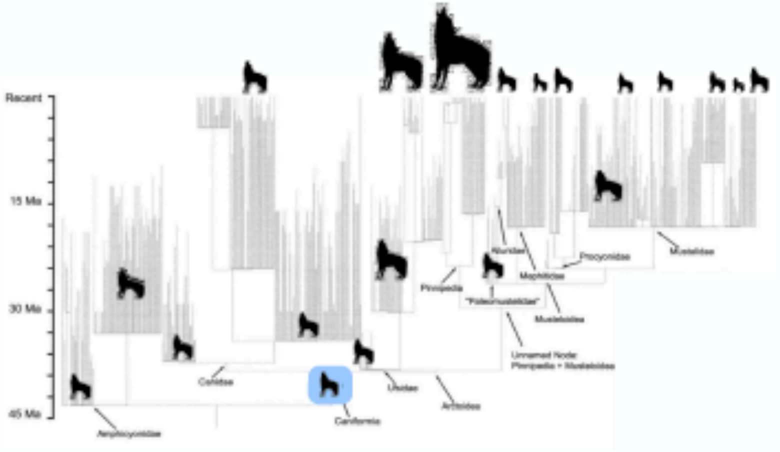


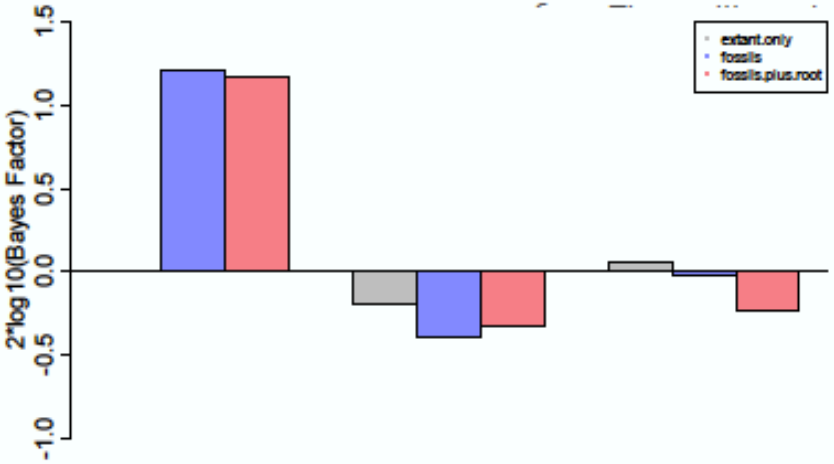


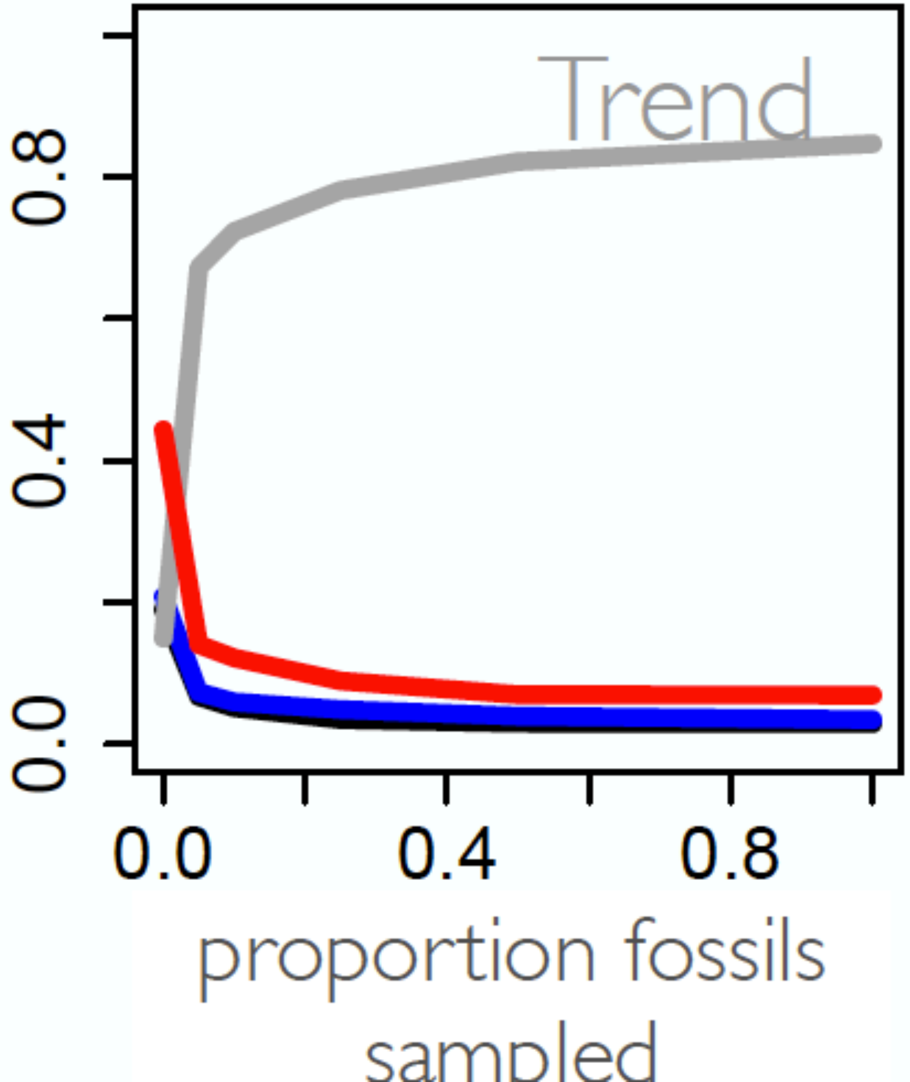
Futuyma 1998

Lamsdell and Braddy 2009

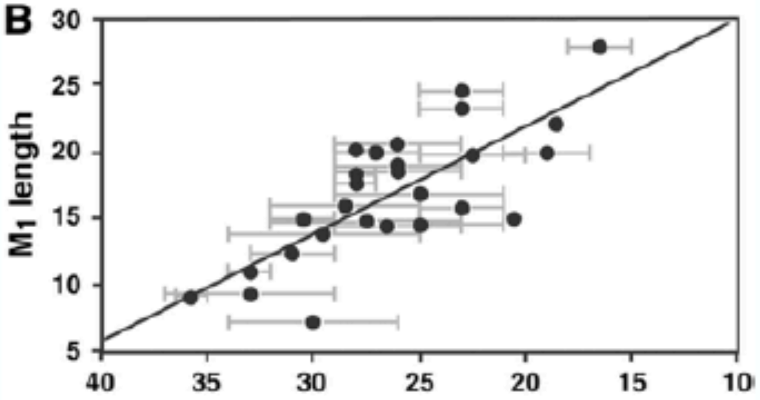
usually need fossils to detect trends







BM with trend



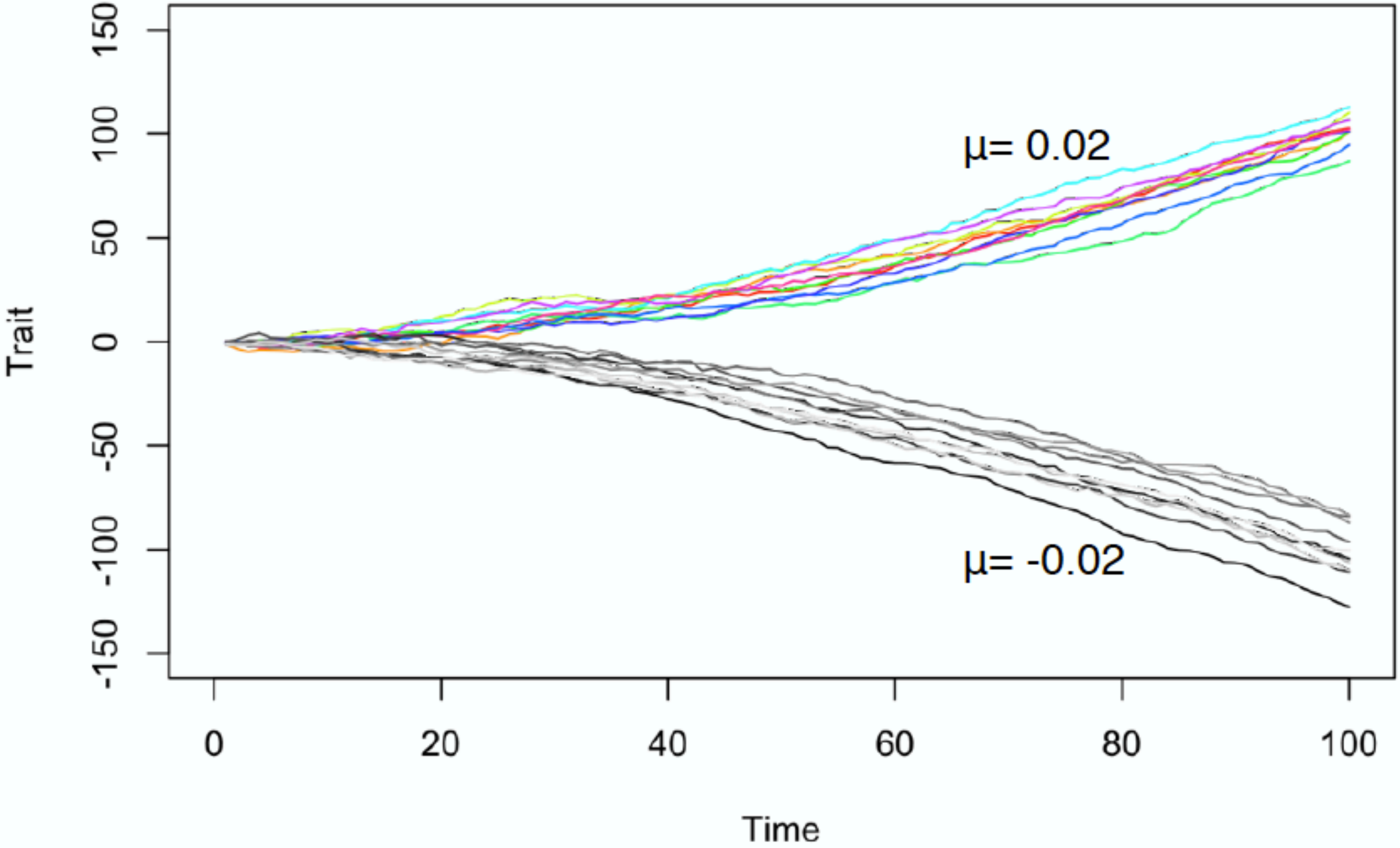
$$dX_{(t)} = \sigma dB_t$$

rate

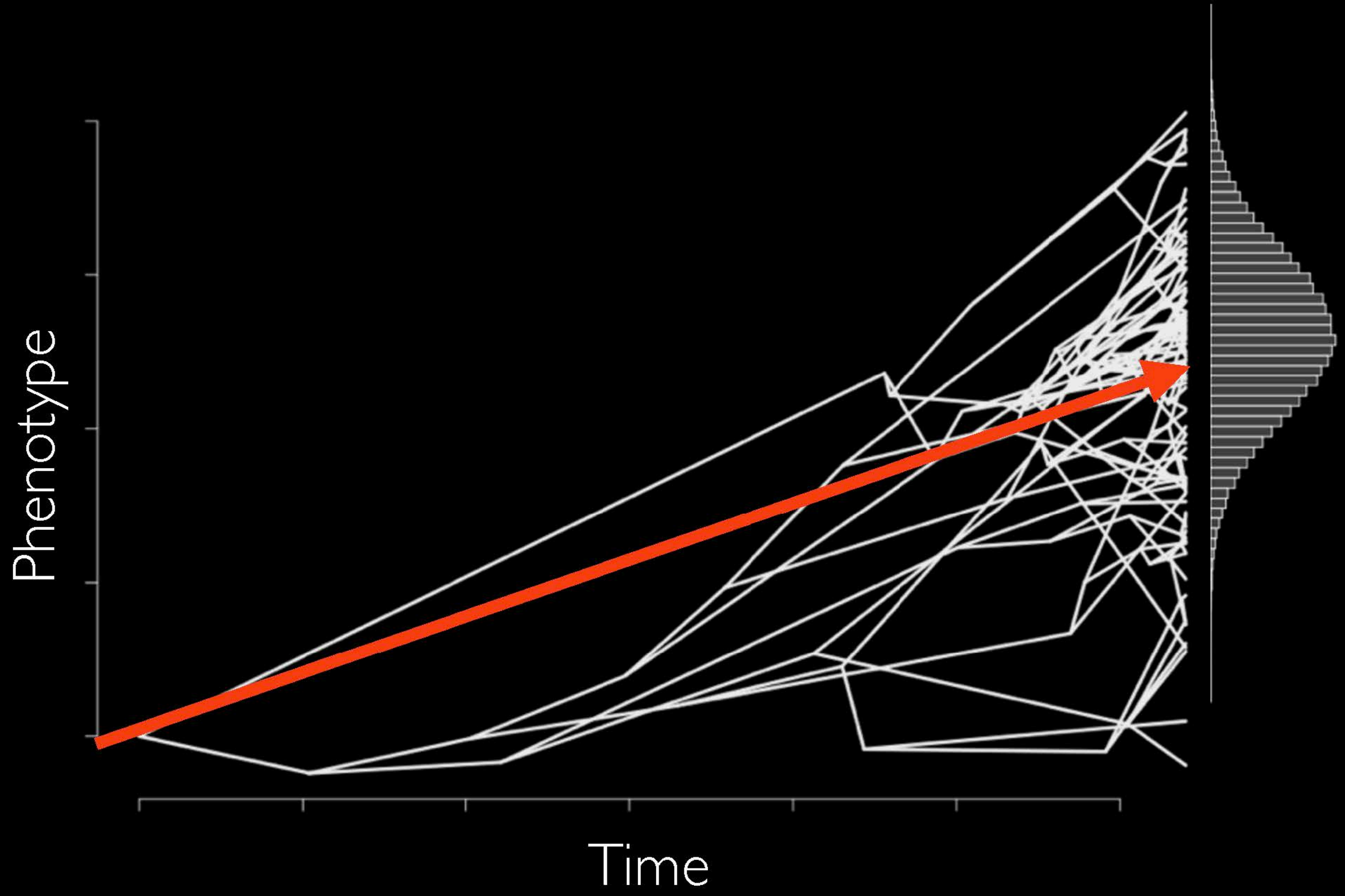
normal distribution where mean = t * μ

trait increases when μ > 0, decreases when μ < 0

BM with trend



trend



under Trend, the variance increases through time as with BM but the expected trait value does not stay the same

trait values for species

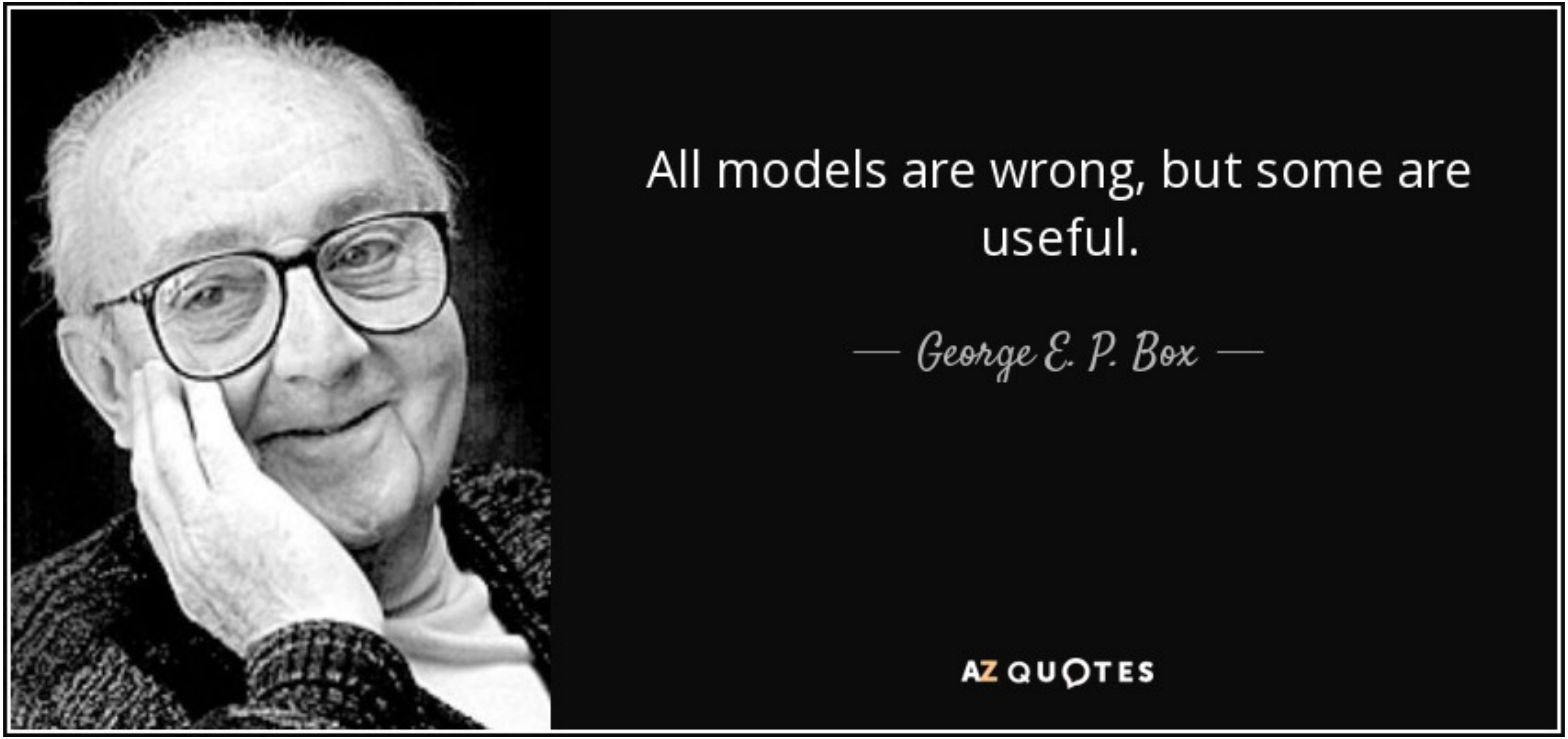
variance-covariance matrix * rate

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

expected value is a function of:

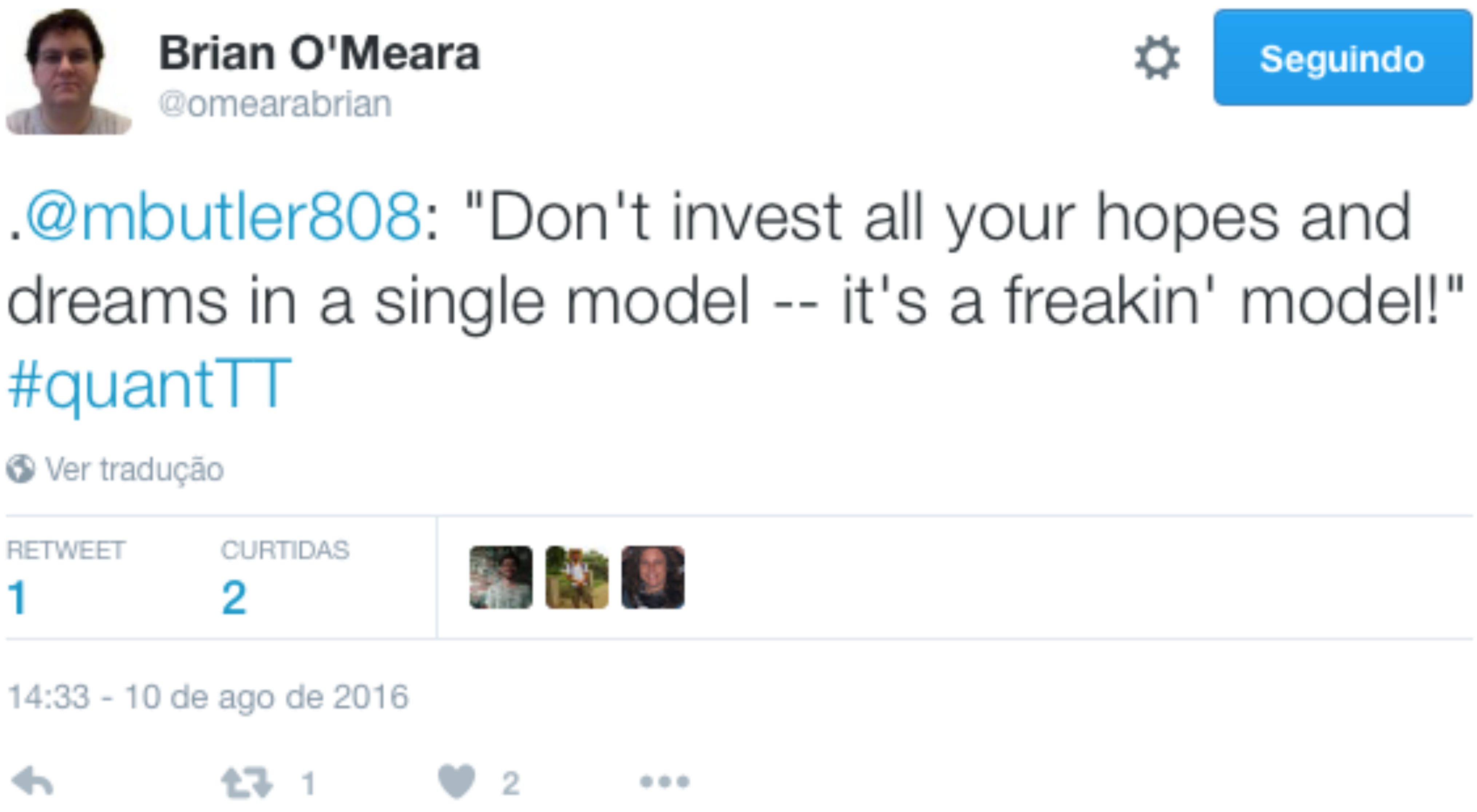
1. the starting state,
2. the strength of the trend,
3. the direction of the trend

Como comparar o ajuste de modelos aos dados?



All models are wrong, but some are useful.

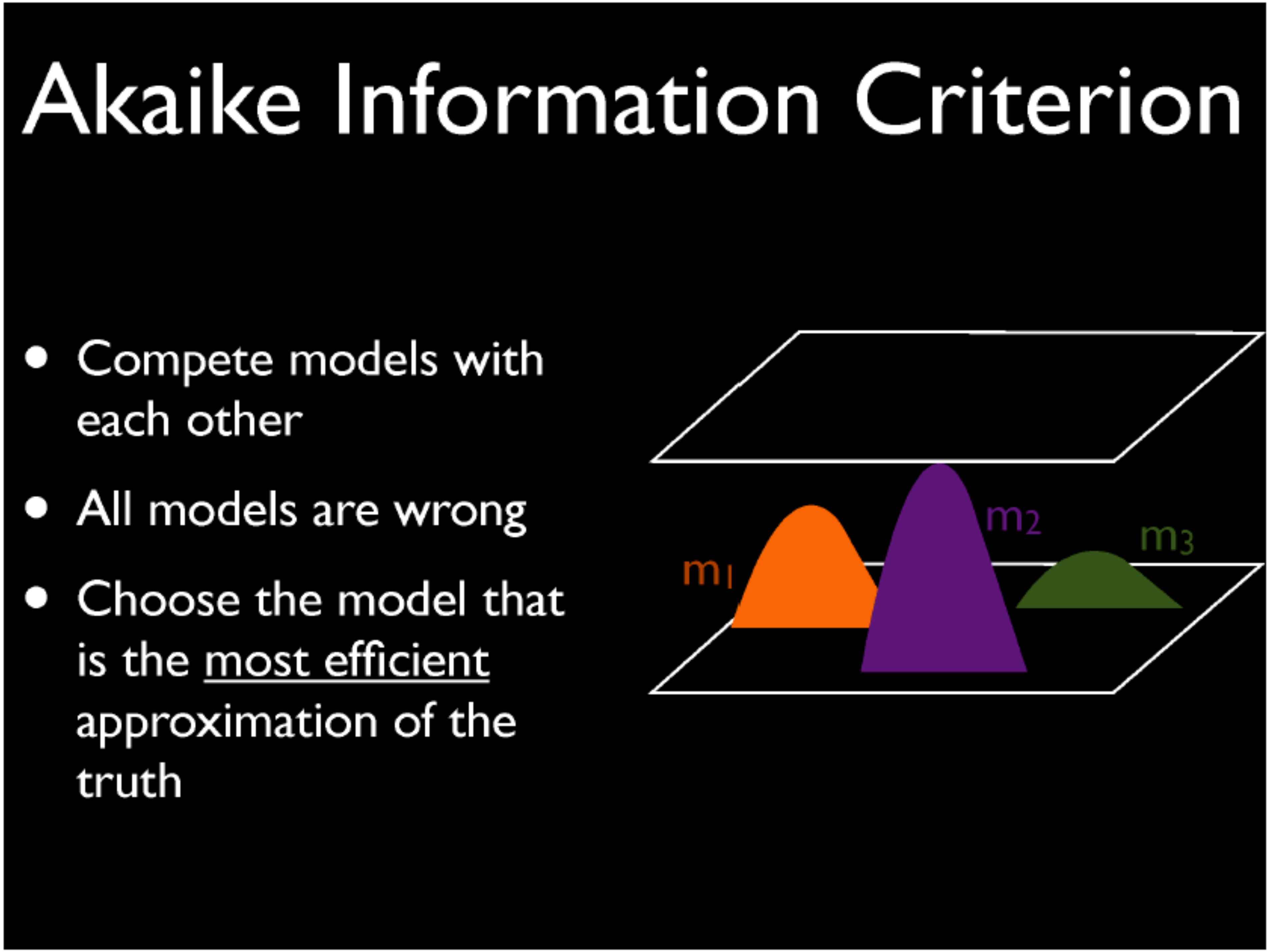
— George E. P. Box —



Brian O'Meara
@omearabrian

.@mbutler808: "Don't invest all your hopes and dreams in a single model -- it's a freakin' model!" #quantTT

14:33 - 10 de ago de 2016



Akaike Information Criterion

- Compete models with each other
- All models are wrong
- Choose the model that is the <u>most efficient</u> approximation of the truth

AIC penalizes the likelihood by the number of extra parameters

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

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2k(k+1)}{n-k-1}$$

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{r=1}^{R} \exp\left(-\frac{1}{2}\Delta_r\right)}$$

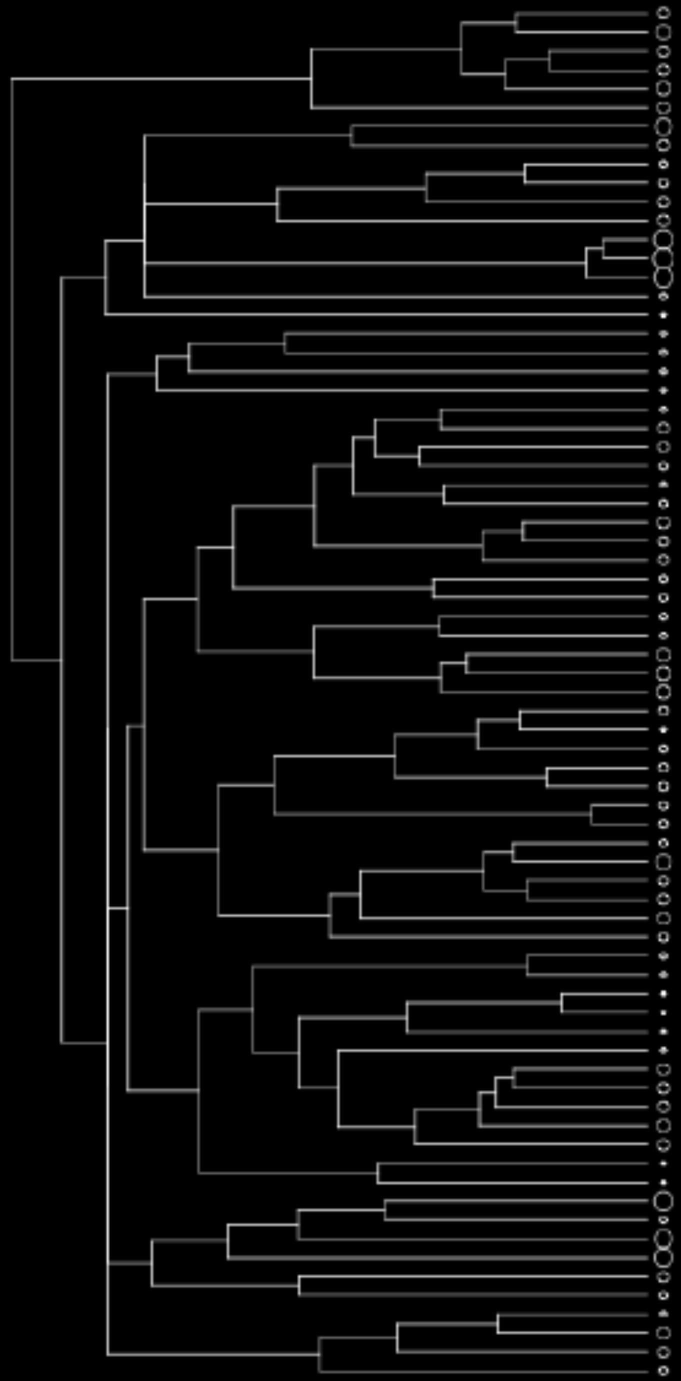
k	BM	EB	OU
root state	✔	✔	✔
σ^2	✔	✔	✔
scalar	×	✔	✔
total	2	3	3

Example: *Anolis* lizards

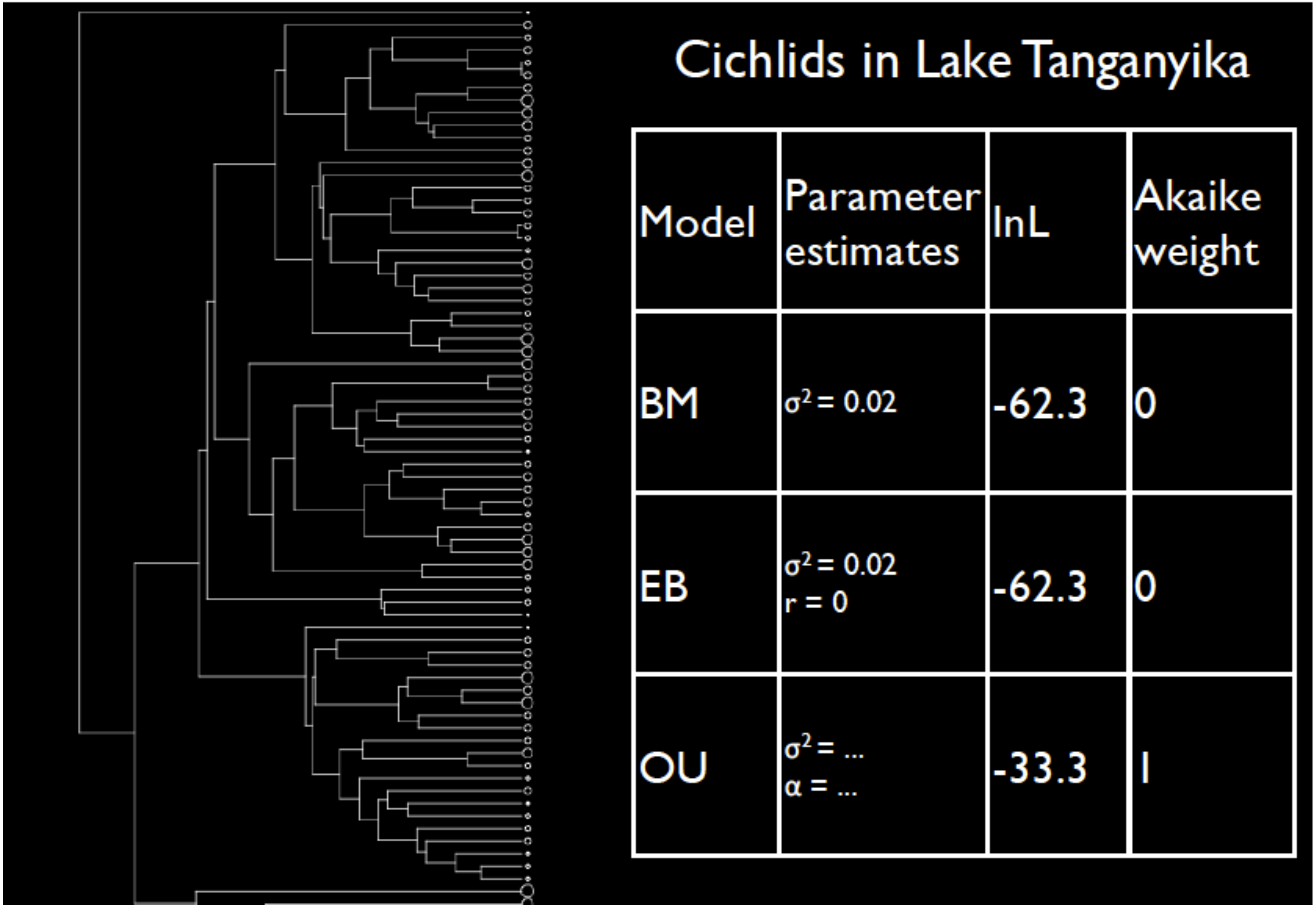
Lizards on Caribbean islands

Phylogenetic and body size data for 73 species (out of ~140 total)

Anolis baleatus



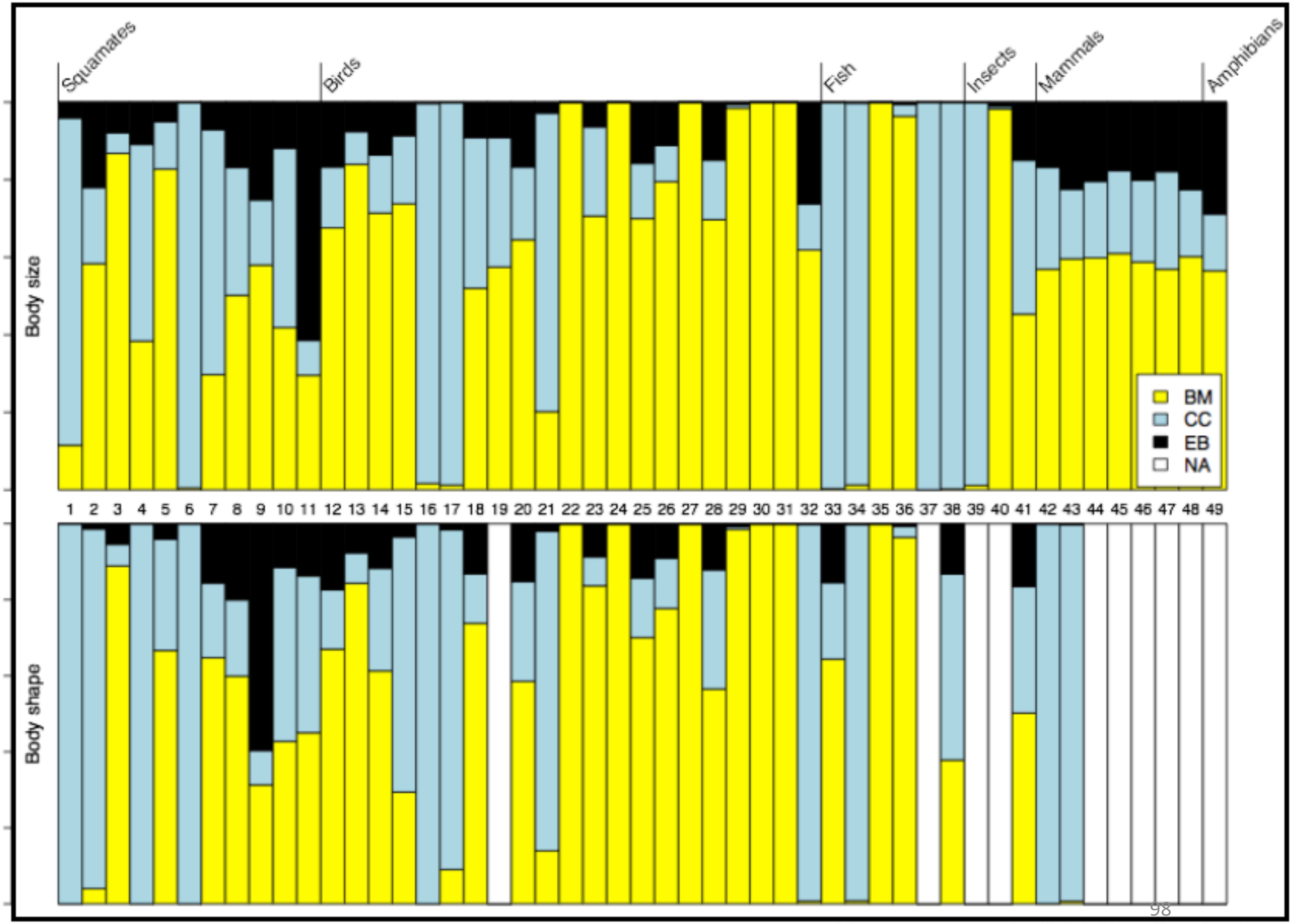
Model	Parameter estimates	lnL	Akaike weight
BM	$\sigma^2 = 0.004$	-18.2	0.58
EB	$\sigma^2 = 0.006$ $r = -0.01$	-18.1	0.2
OU	$\sigma^2 = 0.004$ $\alpha = 0$	-18.2	0.22



Cichlids in Lake Tanganyika

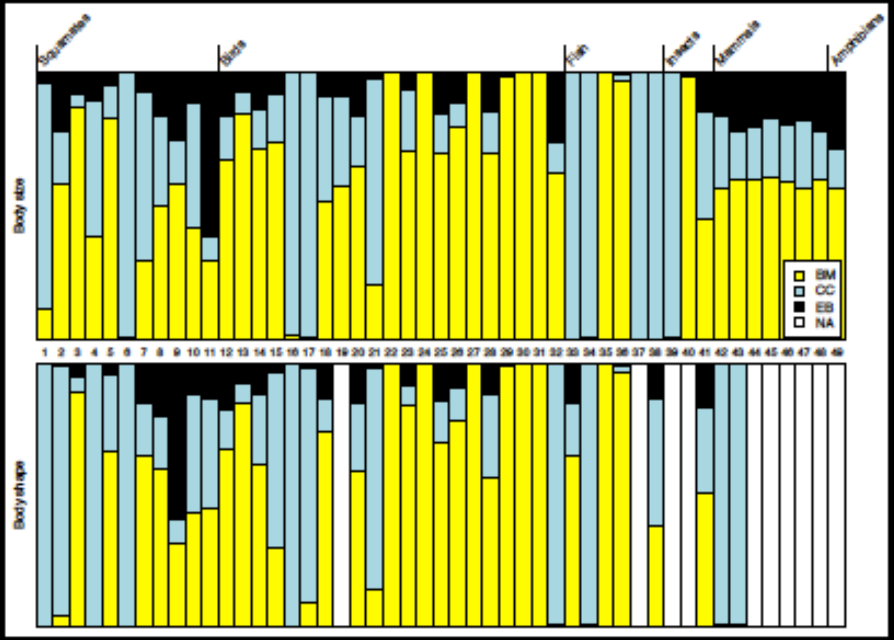
Model	Parameter estimates	lnL	Akaike weight
BM	$\sigma^2 = 0.02$	-62.3	0
EB	$\sigma^2 = 0.02$ $r = 0$	-62.3	0
OU	$\sigma^2 = \ldots$ $\alpha = \ldots$	-33.3	1

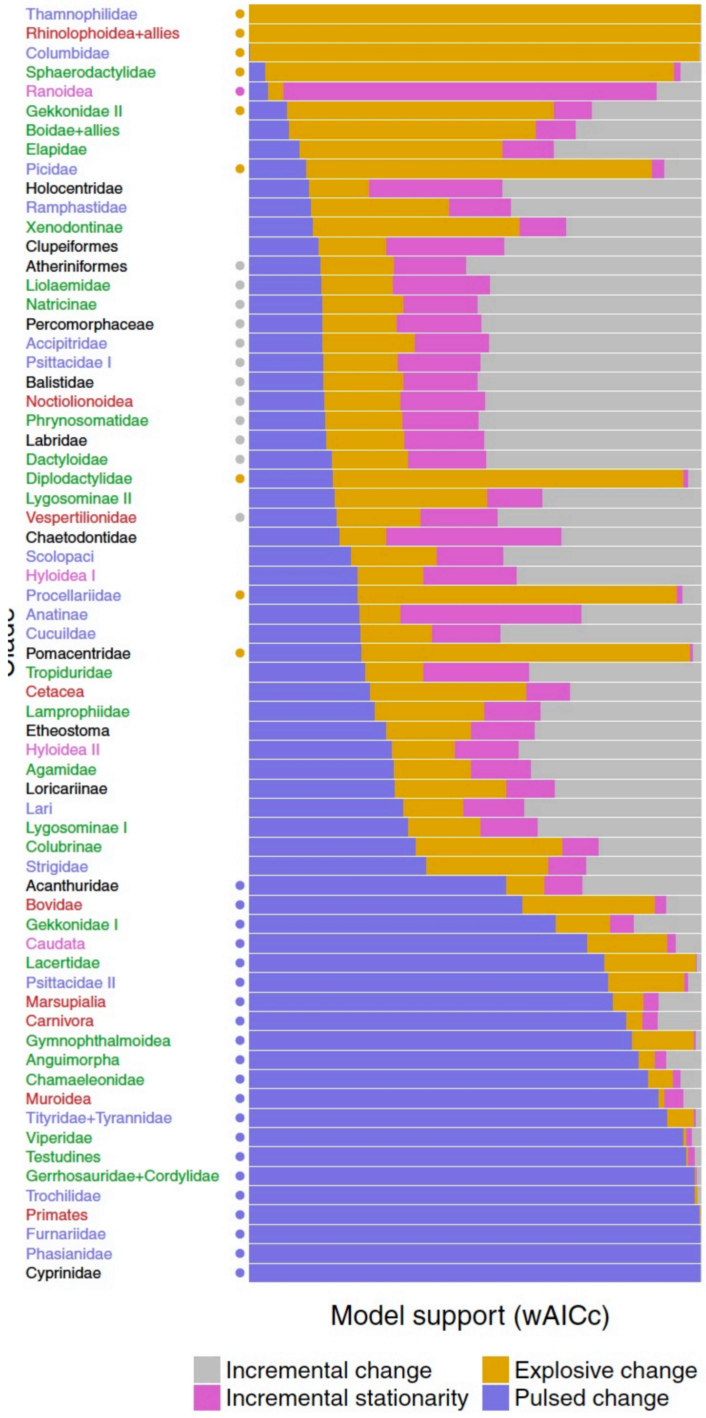
Pouca suporte para modelos EB para tamanho de corpo em geral, mesmo em exemplos clássicos de radiação adaptativa

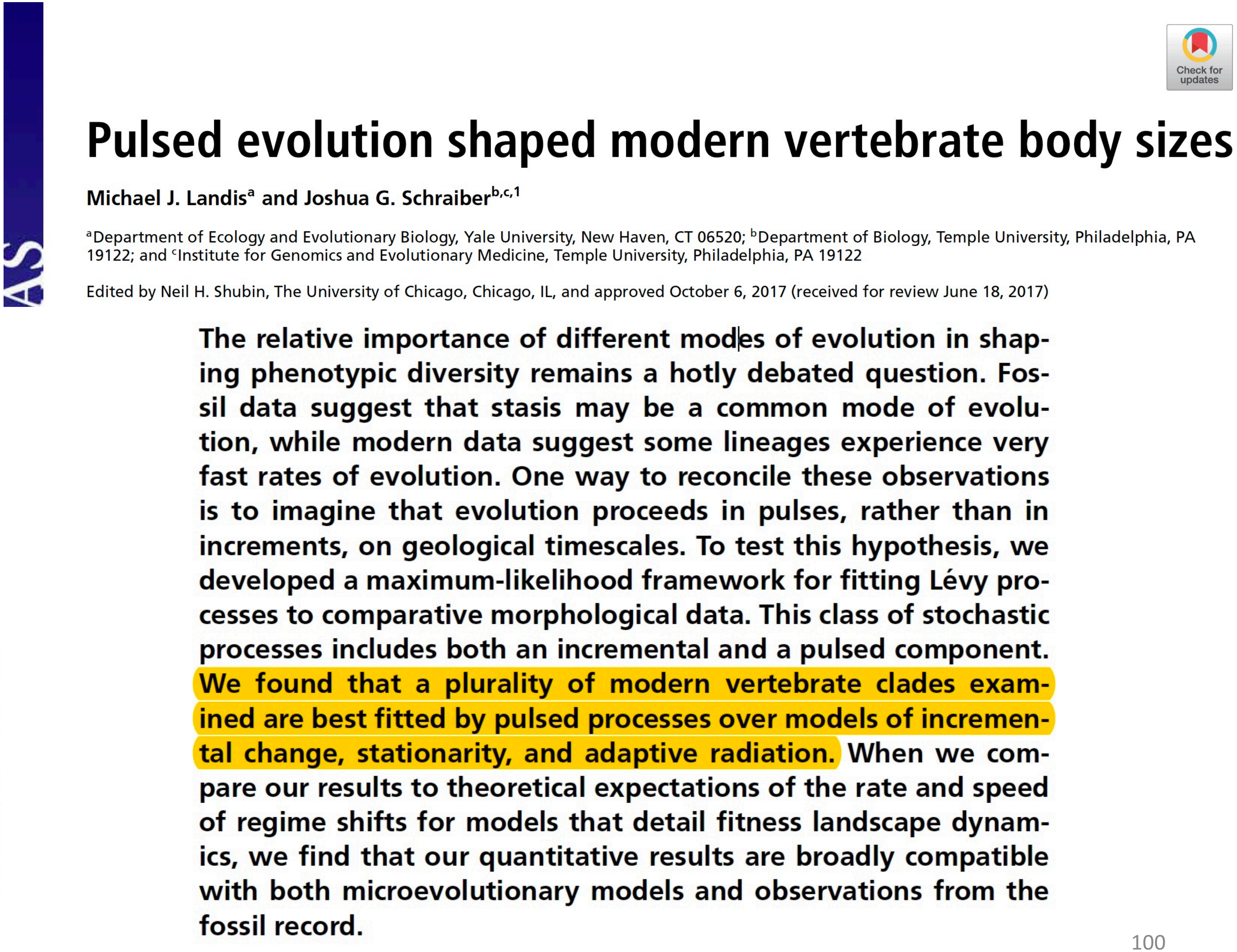


Harmon et al. (2010) - Evolution

- “Adaptive radiation” pattern very rare in this data set
- Constraints dominate over long time periods
- Brownian motion is sometimes a poor fit to real data







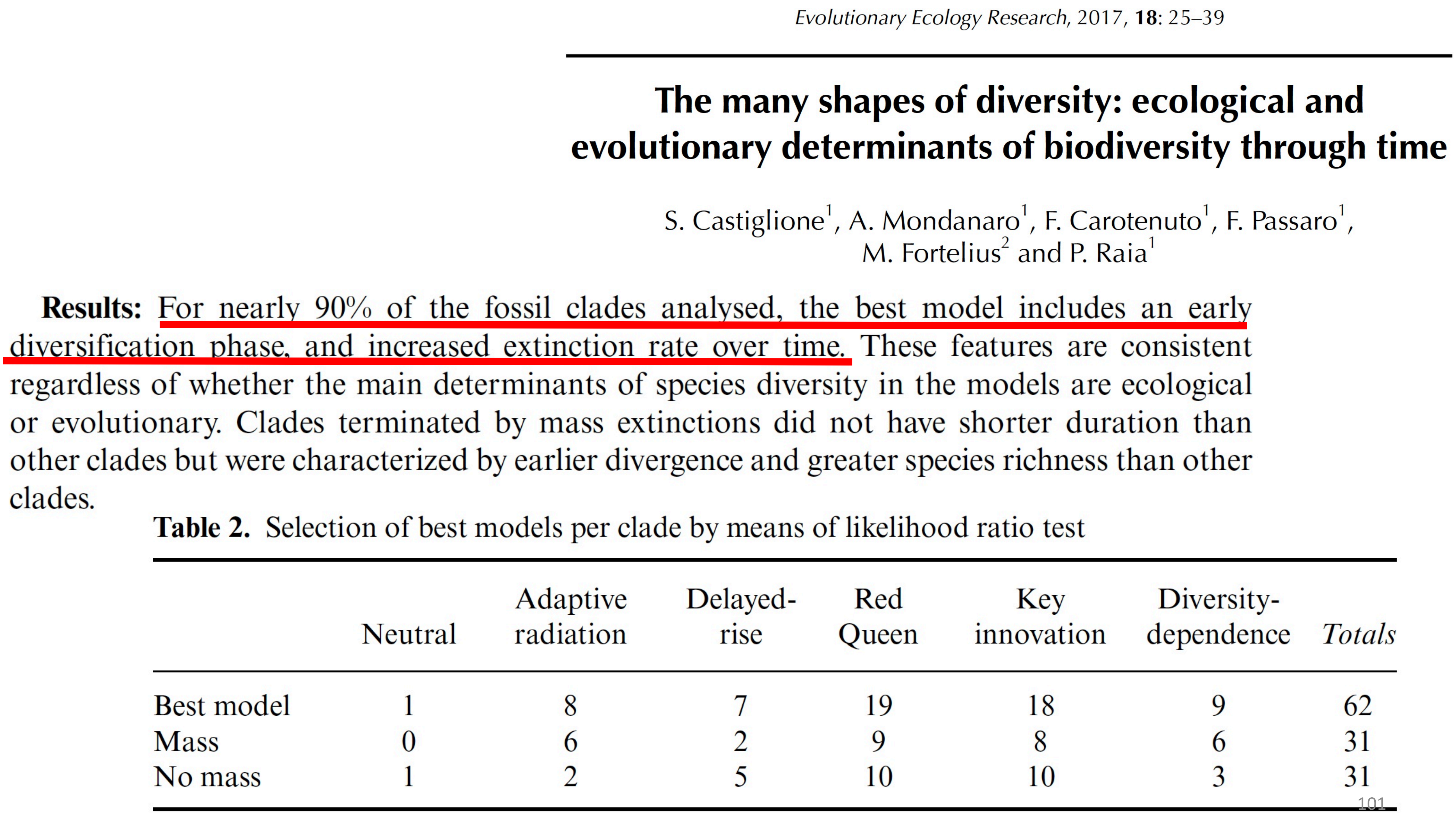
Pulsed evolution shaped modern vertebrate body sizes

Michael J. Landis[a] **and Joshua G. Schraiber**[b,c,1]

[a]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520; [b]Department of Biology, Temple University, Philadelphia, PA 19122; and [c]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA 19122

Edited by Neil H. Shubin, The University of Chicago, Chicago, IL, and approved October 6, 2017 (received for review June 18, 2017)

The relative importance of different modes of evolution in shaping phenotypic diversity remains a hotly debated question. Fossil data suggest that stasis may be a common mode of evolution, while modern data suggest some lineages experience very fast rates of evolution. One way to reconcile these observations is to imagine that evolution proceeds in pulses, rather than in increments, on geological timescales. To test this hypothesis, we developed a maximum-likelihood framework for fitting Lévy processes to comparative morphological data. This class of stochastic processes includes both an incremental and a pulsed component. We found that a plurality of modern vertebrate clades examined are best fitted by pulsed processes over models of incremental change, stationarity, and adaptive radiation. When we compare our results to theoretical expectations of the rate and speed of regime shifts for models that detail fitness landscape dynamics, we find that our quantitative results are broadly compatible with both microevolutionary models and observations from the fossil record.



Evolutionary Ecology Research, 2017, **18**: 25–39

The many shapes of diversity: ecological and evolutionary determinants of biodiversity through time

S. Castiglione[1], A. Mondanaro[1], F. Carotenuto[1], F. Passaro[1],
M. Fortelius[2] and P. Raia[1]

Results: For nearly 90% of the fossil clades analysed, the best model includes an early diversification phase, and increased extinction rate over time. These features are consistent regardless of whether the main determinants of species diversity in the models are ecological or evolutionary. Clades terminated by mass extinctions did not have shorter duration than other clades but were characterized by earlier divergence and greater species richness than other clades.

Table 2. Selection of best models per clade by means of likelihood ratio test

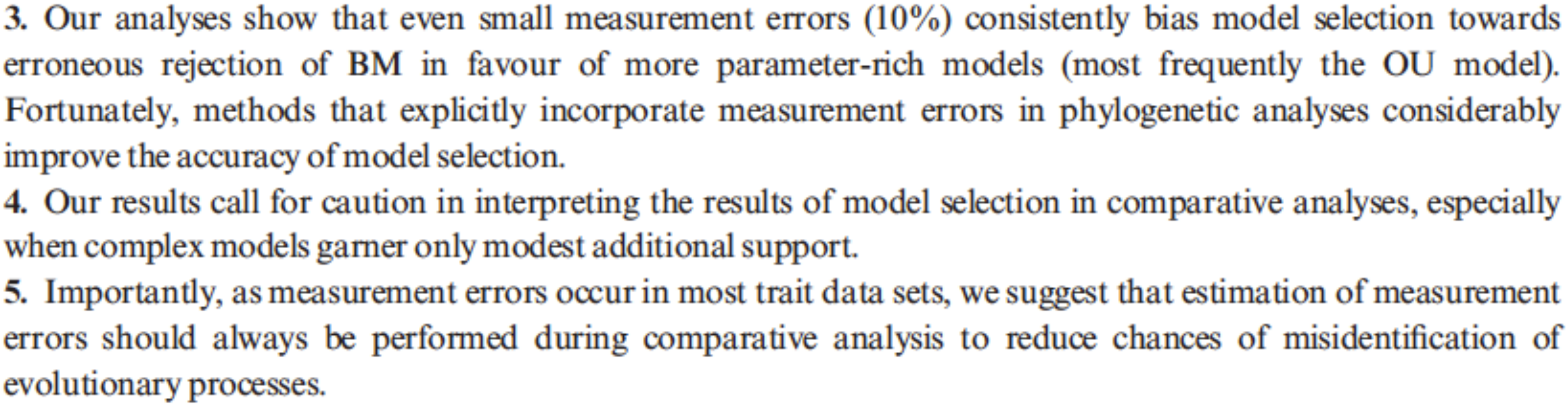
	Neutral	Adaptive radiation	Delayed-rise	Red Queen	Key innovation	Diversity-dependence	*Totals*
Best model	1	8	7	19	18	9	62
Mass	0	6	2	9	8	6	31
No mass	1	2	5	10	10	3	31

Methods in Ecology and Evolution

Methods in Ecology and Evolution 2015, **6**, 340–346 doi: 10.1111/2041-210X.12337

Measurement errors should always be incorporated in phylogenetic comparative analysis

Daniele Silvestro[1,2,3]†, Anna Kostikova[2,3]†, Glenn Litsios[2,3], Peter B. Pearman[4,5] and Nicolas Salamin[2,3]*



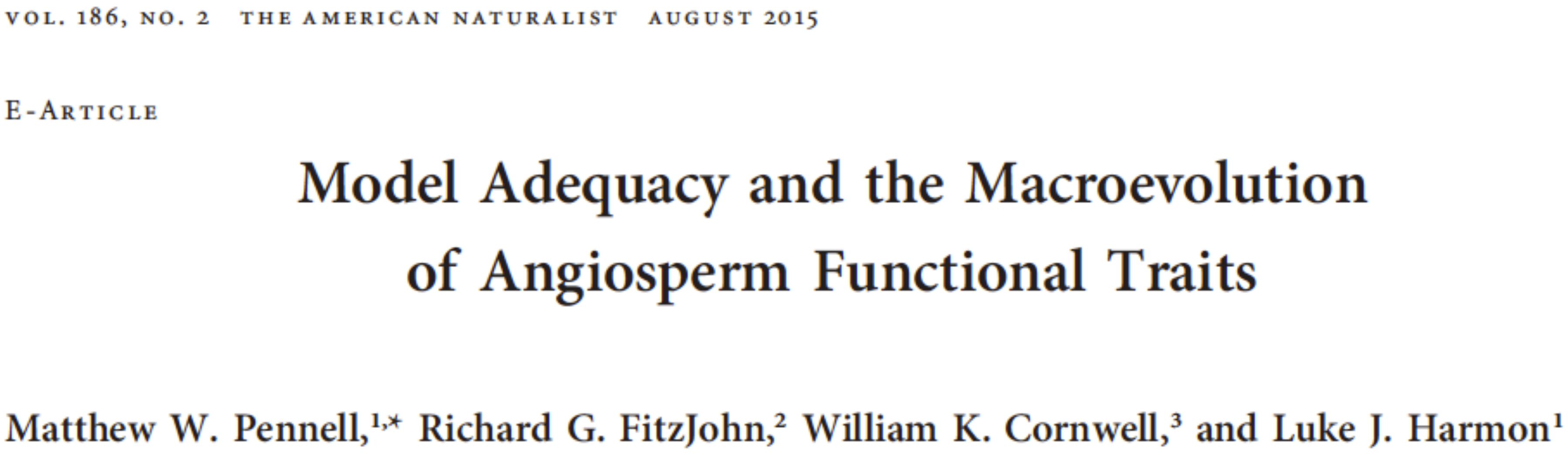
3. Our analyses show that even small measurement errors (10%) consistently bias model selection towards erroneous rejection of BM in favour of more parameter-rich models (most frequently the OU model). Fortunately, methods that explicitly incorporate measurement errors in phylogenetic analyses considerably improve the accuracy of model selection.

4. Our results call for caution in interpreting the results of model selection in comparative analyses, especially when complex models garner only modest additional support.

5. Importantly, as measurement errors occur in most trait data sets, we suggest that estimation of measurement errors should always be performed during comparative analysis to reduce chances of misidentification of evolutionary processes.

Diferença entre ajuste (AIC) e ”explicação” do modelo (R^2)

- Abordagem de teoria de informação (AIC-like) aponta o melhor modelo entre todos os disponíveis
  - Mas mesmo o melhor modelo (aquele com menor AIC) pode explicar pouco dos dados (ter R^2 baixo)
  - Peso de Akaike pode ser uma estatística melhor nesse caso
- O que fazer nesse caso?
  - Abordagens de ”model adequacy”
  - Arburtus (Pennell et al. 2015)



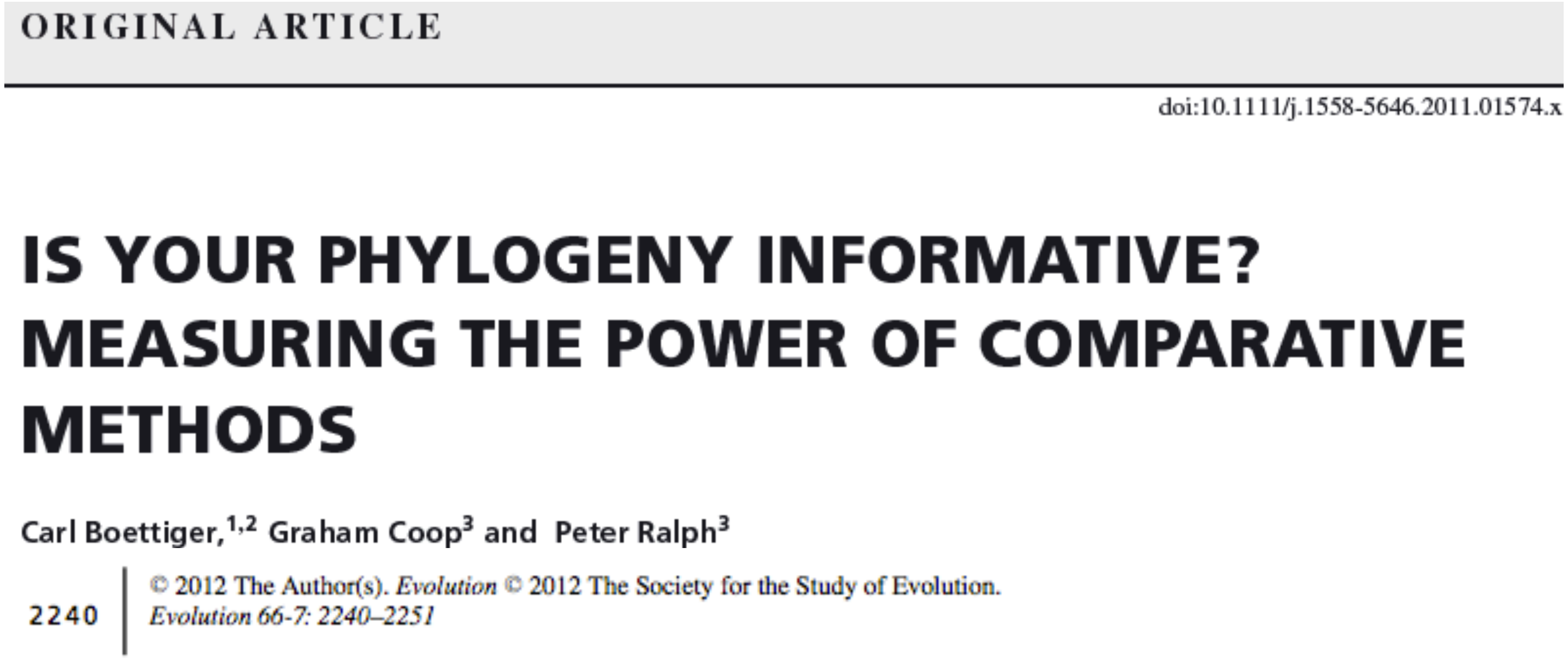
VOL. 186, NO. 2 THE AMERICAN NATURALIST AUGUST 2015

E-ARTICLE

Model Adequacy and the Macroevolution of Angiosperm Functional Traits

Matthew W. Pennell,[1,*] Richard G. FitzJohn,[2] William K. Cornwell,[3] and Luke J. Harmon[1]

Arbutus



ORIGINAL ARTICLE

doi:10.1111/j.1558-5646.2011.01574.x

IS YOUR PHYLOGENY INFORMATIVE? MEASURING THE POWER OF COMPARATIVE METHODS

Carl Boettiger,[1,2] Graham Coop[3] and Peter Ralph[3]

2240

© 2012 The Author(s). *Evolution* © 2012 The Society for the Study of Evolution.

Evolution 66-7: 2240–2251

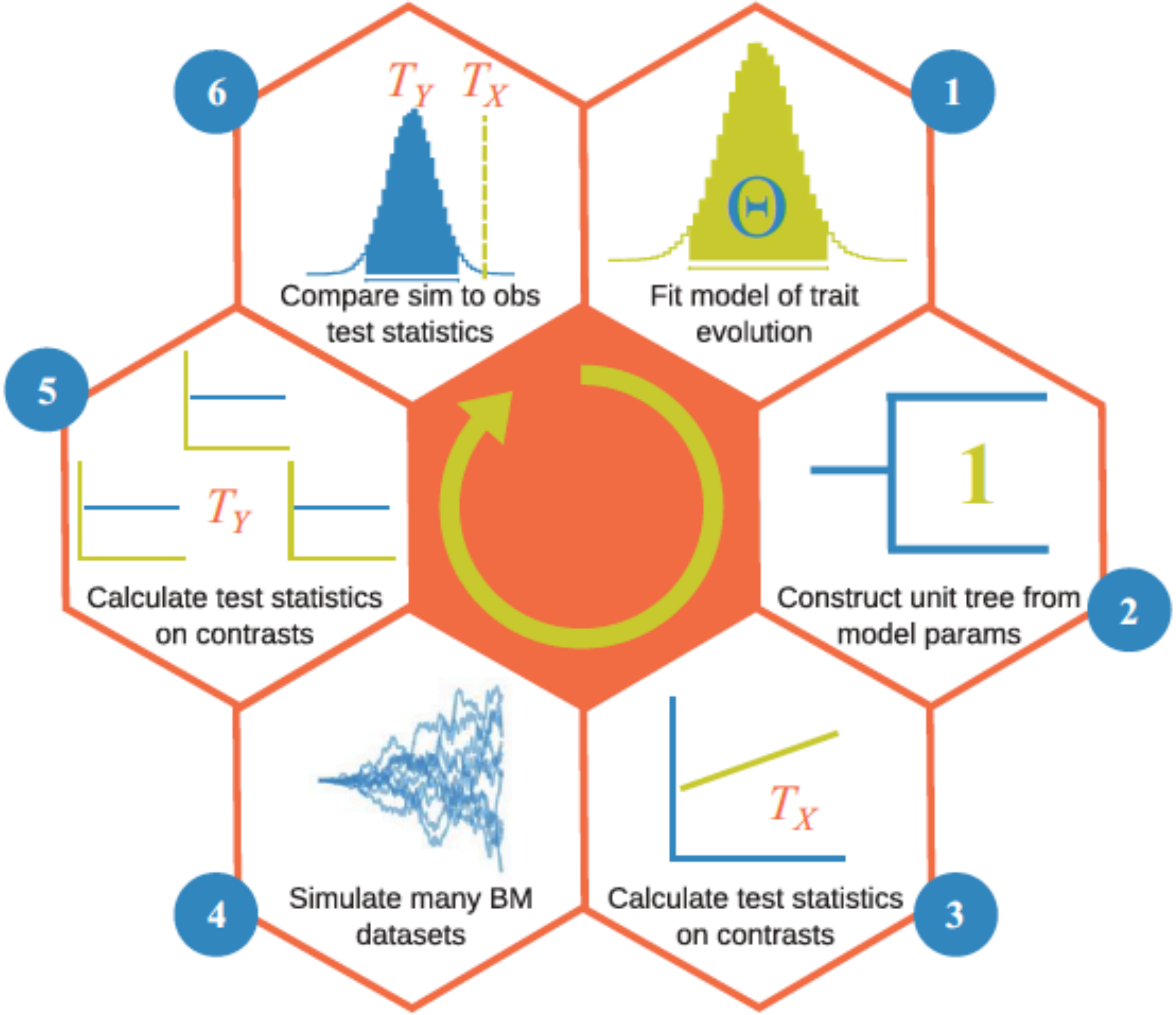
Phylogenetic
Monte
Carlo

Arbutus

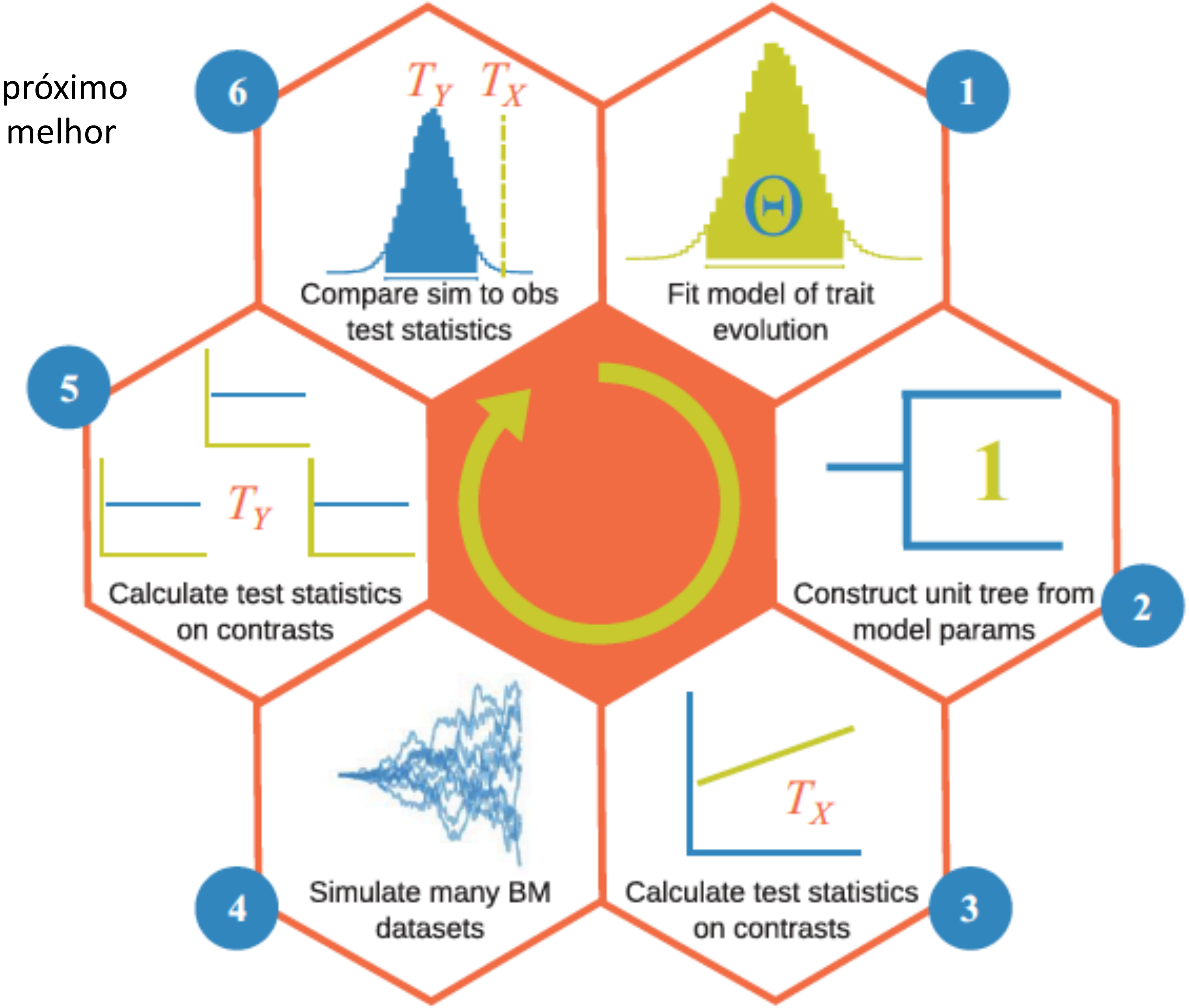
- Comparar o ajuste de modelos em relação ao universo de todos os modelos possíveis

- Focado na adequabilidade de modelos para evolução de atributos contínuos.

- Também pode ser usado para PGLS


6
T_Y T_X
Compare sim to obs test statistics
1
Θ
Fit model of trait evolution
5
T_Y
Calculate test statistics on contrasts
1
Construct unit tree from model params
2
4
Simulate many BM datasets
T_X
Calculate test statistics on contrasts
3

Quanto mais próximo do simulado melhor



Modificar a árvore de maneira a descrever a história do trait

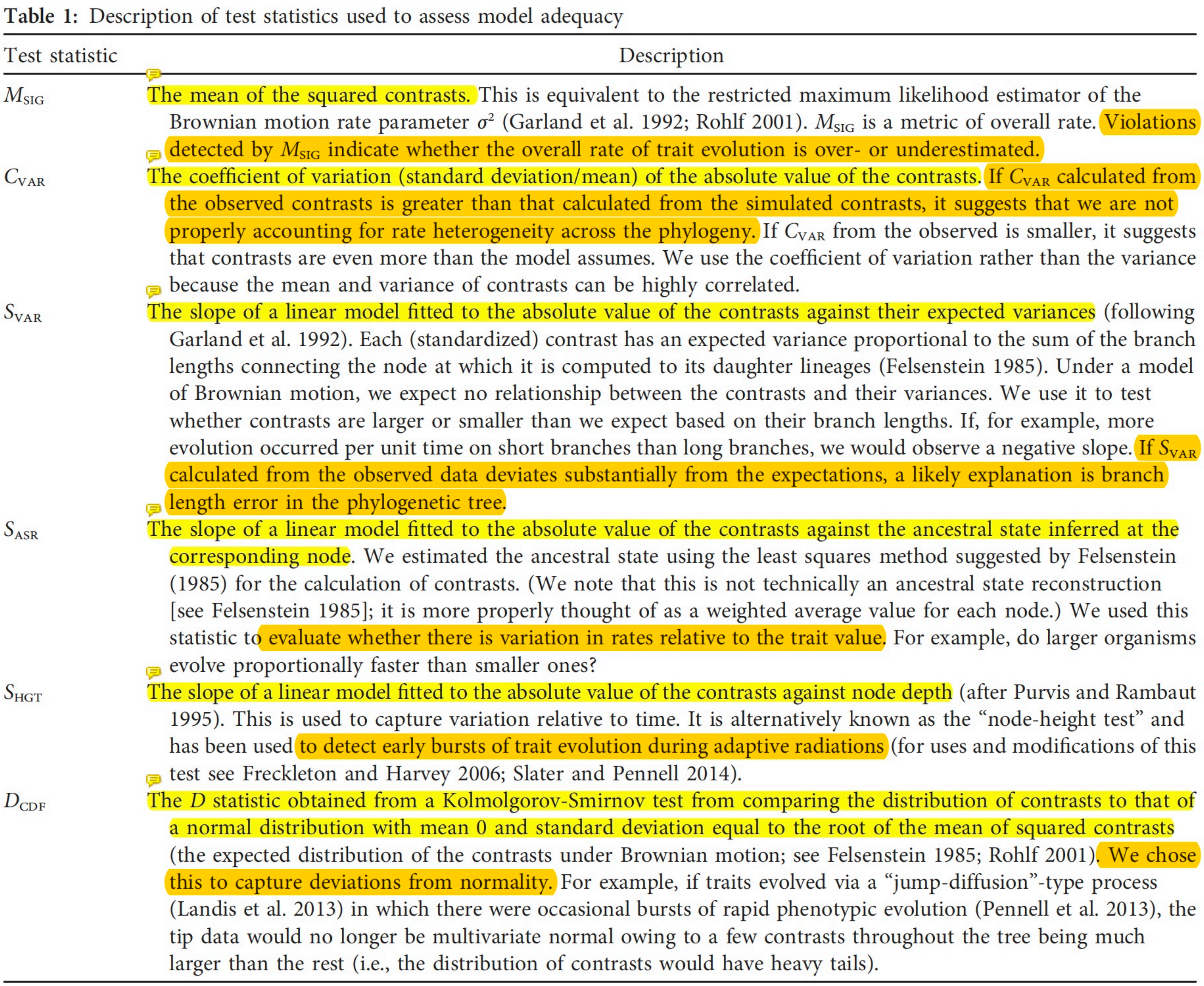


Table 1: Description of test statistics used to assess model adequacy

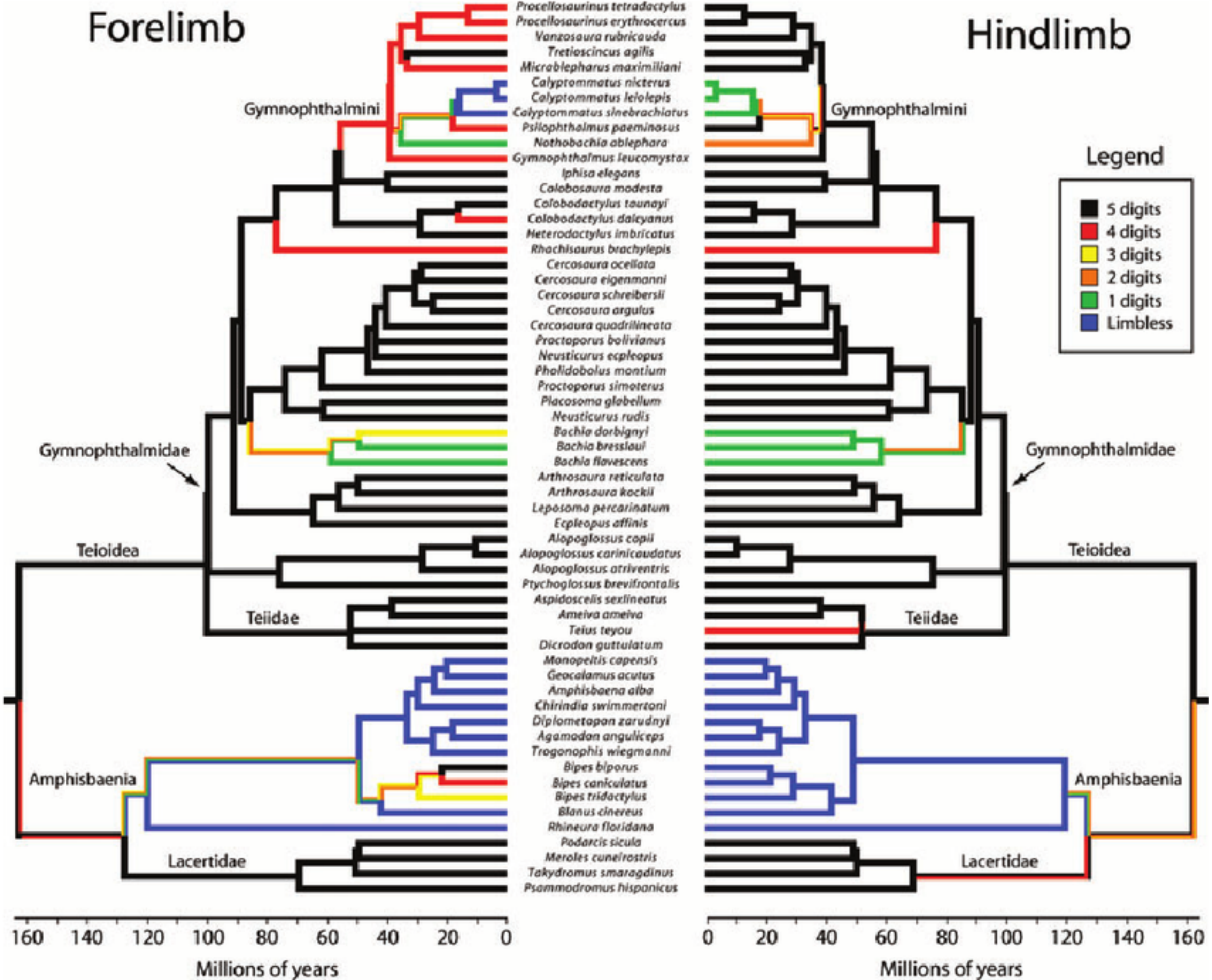
Test statistic	Description
M_{SIG}	The mean of the squared contrasts. This is equivalent to the restricted maximum likelihood estimator of the Brownian motion rate parameter σ^2 (Garland et al. 1992; Rohlf 2001). M_{SIG} is a metric of overall rate. Violations detected by M_{SIG} indicate whether the overall rate of trait evolution is over- or underestimated.
C_{VAR}	The coefficient of variation (standard deviation/mean) of the absolute value of the contrasts. If C_{VAR} calculated from the observed contrasts is greater than that calculated from the simulated contrasts, it suggests that we are not properly accounting for rate heterogeneity across the phylogeny. If C_{VAR} from the observed is smaller, it suggests that contrasts are even more than the model assumes. We use the coefficient of variation rather than the variance because the mean and variance of contrasts can be highly correlated.
S_{VAR}	The slope of a linear model fitted to the absolute value of the contrasts against their expected variances (following Garland et al. 1992). Each (standardized) contrast has an expected variance proportional to the sum of the branch lengths connecting the node at which it is computed to its daughter lineages (Felsenstein 1985). Under a model of Brownian motion, we expect no relationship between the contrasts and their variances. We use it to test whether contrasts are larger or smaller than we expect based on their branch lengths. If, for example, more evolution occurred per unit time on short branches than long branches, we would observe a negative slope. If S_{VAR} calculated from the observed data deviates substantially from the expectations, a likely explanation is branch length error in the phylogenetic tree.
S_{ASR}	The slope of a linear model fitted to the absolute value of the contrasts against the ancestral state inferred at the corresponding node. We estimated the ancestral state using the least squares method suggested by Felsenstein (1985) for the calculation of contrasts. (We note that this is not technically an ancestral state reconstruction [see Felsenstein 1985]; it is more properly thought of as a weighted average value for each node.) We used this statistic to evaluate whether there is variation in rates relative to the trait value. For example, do larger organisms evolve proportionally faster than smaller ones?
S_{HGT}	The slope of a linear model fitted to the absolute value of the contrasts against node depth (after Purvis and Rambaut 1995). This is used to capture variation relative to time. It is alternatively known as the "node-height test" and has been used to detect early bursts of trait evolution during adaptive radiations (for uses and modifications of this test see Freckleton and Harvey 2006; Slater and Pennell 2014).
D_{CDF}	The D statistic obtained from a Kolmolgorov-Smirnov test from comparing the distribution of contrasts to that of a normal distribution with mean 0 and standard deviation equal to the root of the mean of squared contrasts (the expected distribution of the contrasts under Brownian motion; see Felsenstein 1985; Rohlf 2001). We chose this to capture deviations from normality. For example, if traits evolved via a "jump-diffusion"-type process (Landis et al. 2013) in which there were occasional bursts of rapid phenotypic evolution (Pennell et al. 2013), the tip data would no longer be multivariate normal owing to a few contrasts throughout the tree being much larger than the rest (i.e., the distribution of contrasts would have heavy tails).


YES!
AGORA BORA BOTAR A MÃO NA MASSA!

Modelos evolutivos para caracteres categóricos


Barrow Island, W.A. © Ryan Ellis


Nick Hart


Forelimb
Hindlimb
Gymnophthalmini
Gymnophthalmini
Gymnophthalmidae
Gymnophthalmidae
Teioidea
Teioidea
Teiidae
Teiidae
Amphisbaenia
Amphisbaenia
Lacertidae
Lacertidae
Legend
5 digits
4 digits
3 digits
2 digits
1 digits
Limbless
Procellosaurinus tetradactylus
Procellosaurinus erythrocercus
Vanzosaura rubricauda
Tretioscincus agilis
Micrablepharus maximiliani
Calyptommatus nicterus
Calyptommatus leiolepis
Calyptommatus sinebrachiatus
Psilophthalmus paeminosus
Nothobachia ablephara
Gymnophthalmus leucomystax
Iphisa elegans
Colobosaura modesta
Colobodactylus taunayi
Colobodactylus dalcyanus
Heterodactylus imbricatus
Rhachisaurus brachylepis
Cercosaura ocellata
Cercosaura eigenmanni
Cercosaura schreibersii
Cercosaura argulus
Cercosaura quadrilineata
Proctoporus bolivianus
Neusticurus ecpleopus
Pholidobolus montium
Proctoporus simoterus
Placosoma glabellum
Neusticurus rudis
Bachia dorbignyi
Bachia bresslaui
Bachia flavescens
Arthrosaura reticulata
Arthrosaura kockii
Leposoma percarinatum
Ecpleopus affinis
Alopoglossus copii
Alopoglossus carinicaudatus
Alopoglossus atriventris
Ptychoglossus brevifrontalis
Aspidoscelis sexlineatus
Ameiva ameiva
Teius teyou
Dicrodon guttulatum
Monopeltis capensis
Geocalamus acutus
Amphisbaena alba
Chirindia swimmertoni
Diplometopon zarudnyi
Agamodon anguliceps
Trogonophis wiegmanni
Bipes biporus
Bipes caniculatus
Bipes tridactylus
Blanus cinereus
Rhineura floridana
Podarcis sicula
Meroles cuneirostris
Takydromus smaragdinus
Psammodromus hispanicus
160
140
120
100
80
60
40
20
0
Millions of years
0
20
40
60
80
100
120
140
160
Millions of years

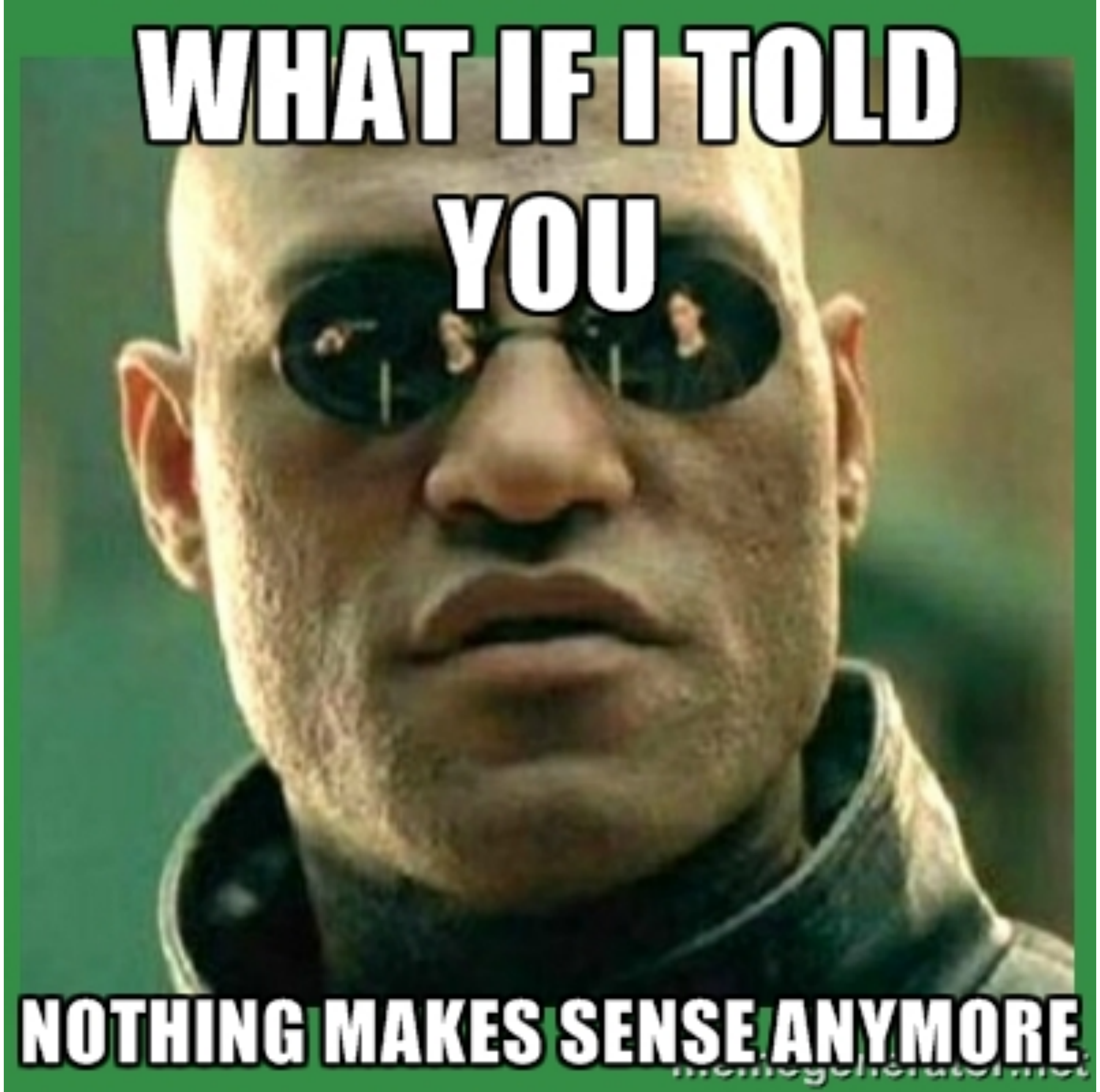
Quantas vezes répteis Squamata perderam seus membros?

Sem membros
Estado 0

Com membros
Estado 1

specimen of *Sirenoscincus yamagishii* Sakata & Hikida, 2003 from Ankarafantsika, Madagascar,

(Photographs E-G: Falk S. Eckhardt.)


WHAT IF I TOLD YOU
NOTHING MAKES SENSE ANYMORE
memegenerator.net

0

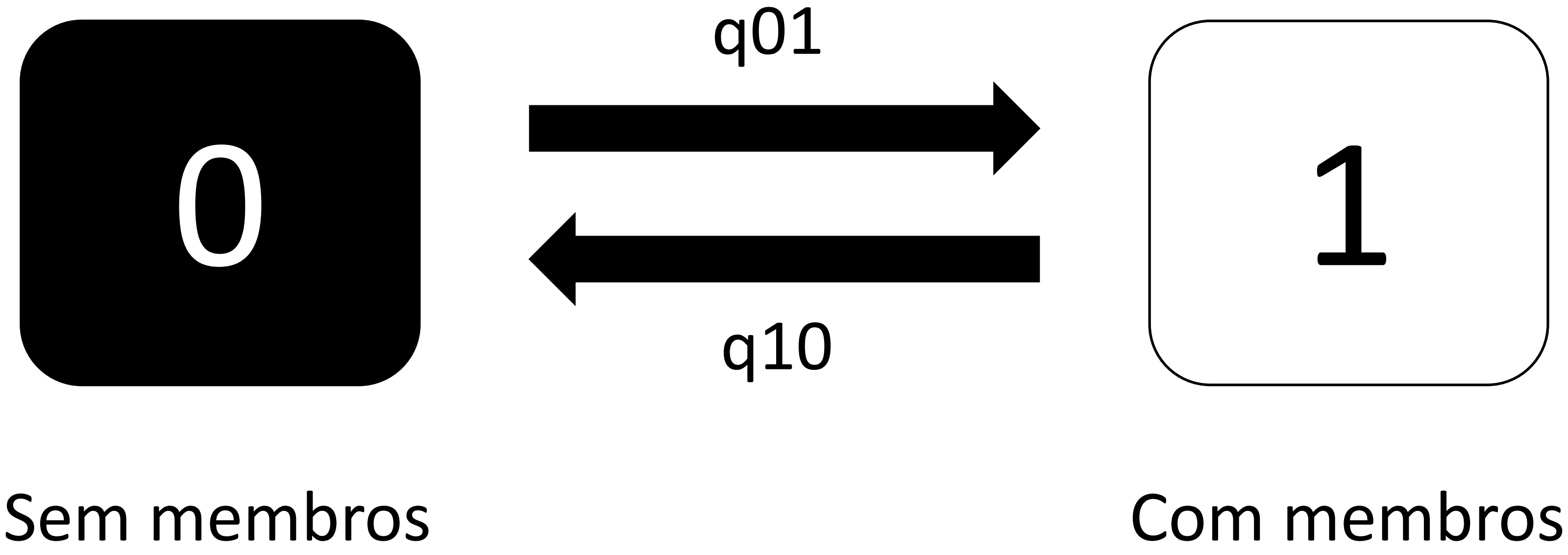
Sem membros

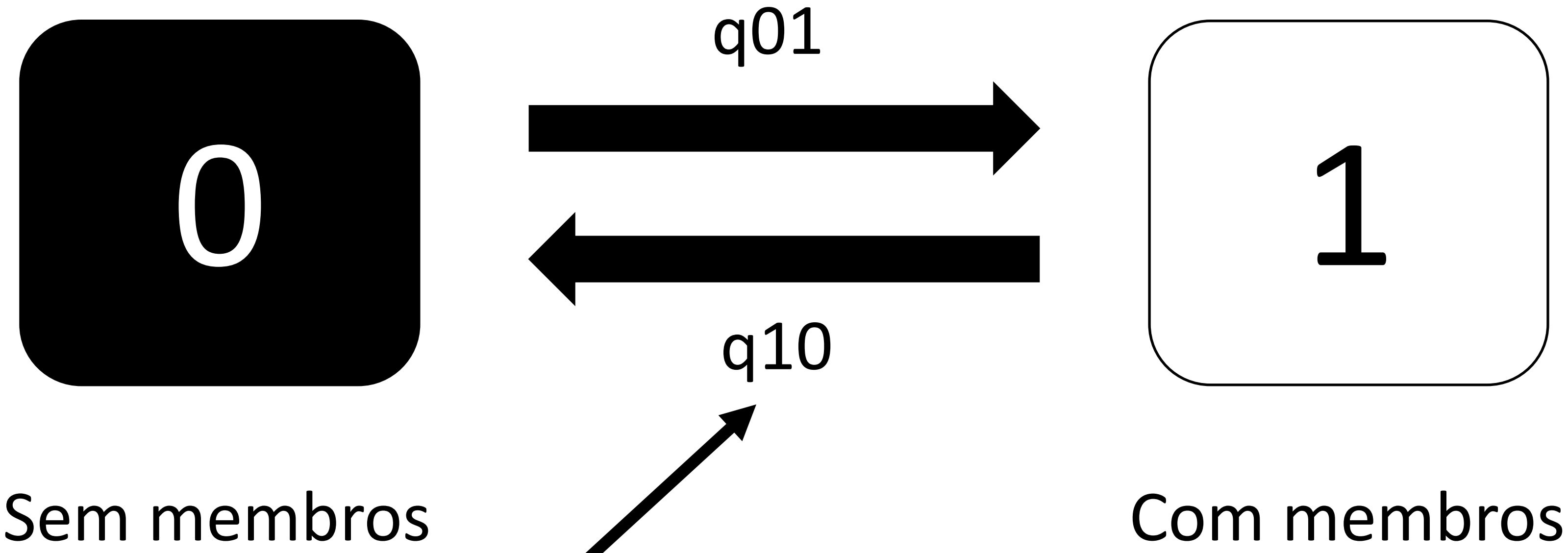
1

Com membros

Sem membros

Com membros


q01
0
1
q10
Sem membros
Com membros



q01

0

1

q10

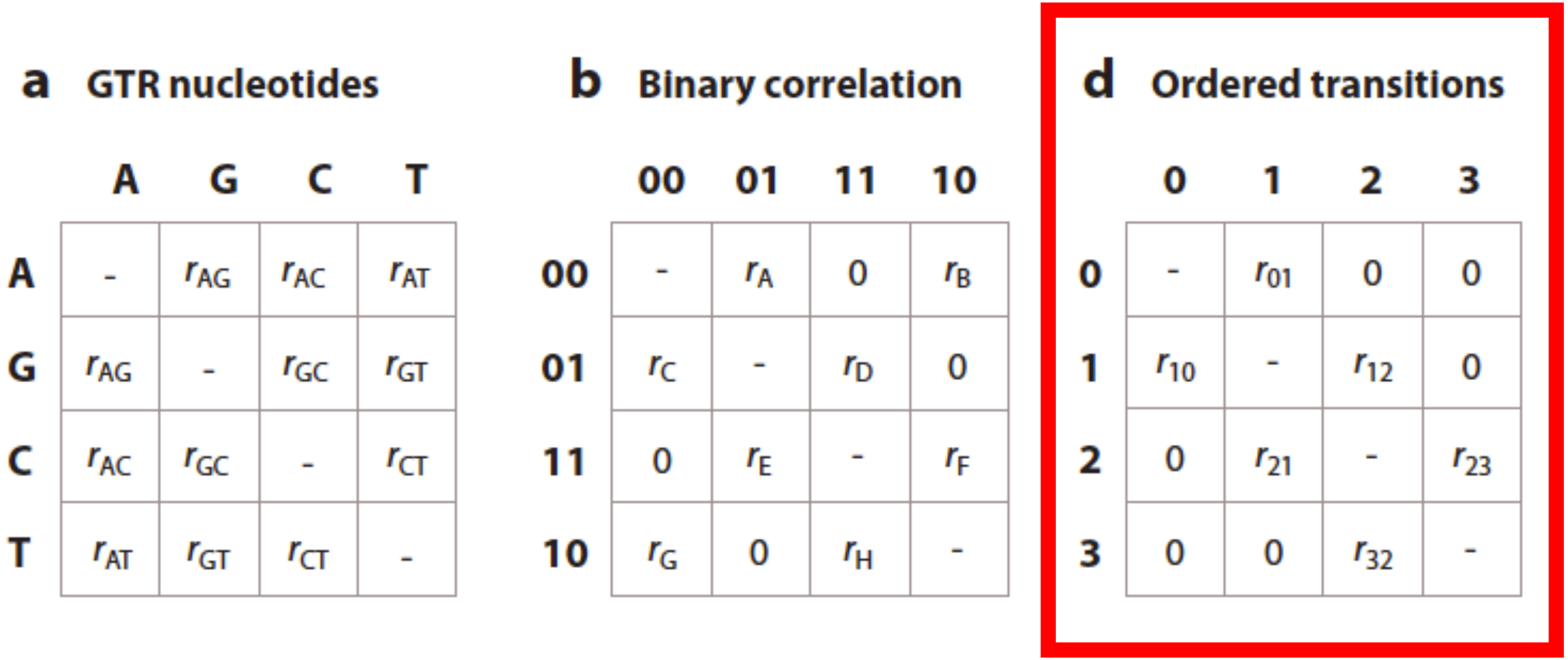
Sem membros

Com membros

Taxa instantânea
de mudança
de estado


THIS SEEMS SORT OF
FAMILIAR
DIYLOL.COM

Mas porque isso é importante?



a GTR nucleotides

	A	G	C	T
A	-	r_{AG}	r_{AC}	r_{AT}
G	r_{AG}	-	r_{GC}	r_{GT}
C	r_{AC}	r_{GC}	-	r_{CT}
T	r_{AT}	r_{GT}	r_{CT}	-

b Binary correlation

	00	01	11	10
00	-	r_A	0	r_B
01	r_C	-	r_D	0
11	0	r_E	-	r_F
10	r_G	0	r_H	-

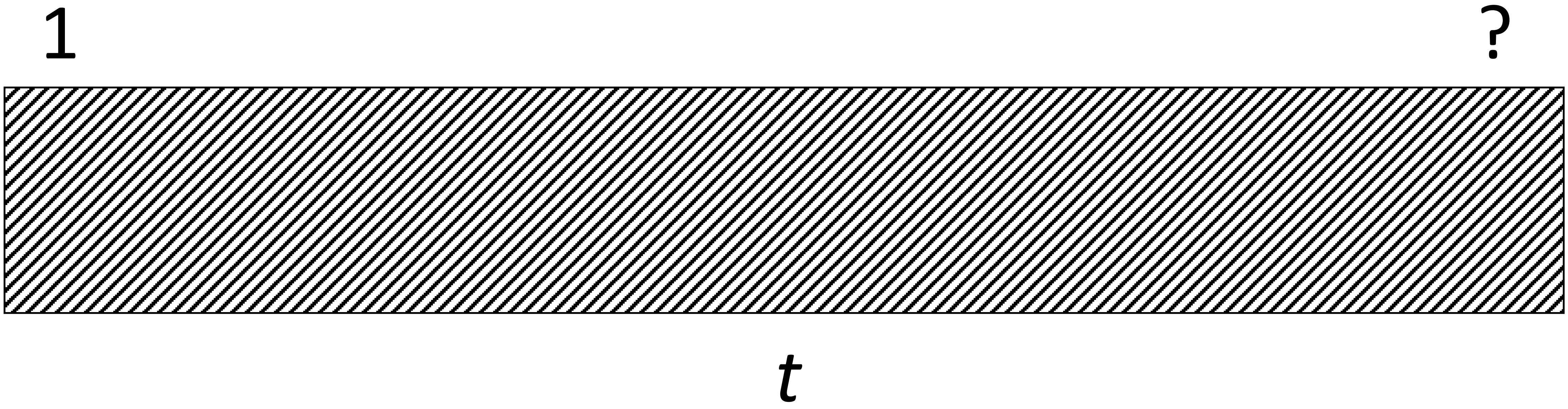
d Ordered transitions

	0	1	2	3
0	-	r_{01}	0	0
1	r_{10}	-	r_{12}	0
2	0	r_{21}	-	r_{23}
3	0	0	r_{32}	-

Além de serem necessários em métodos de inferência filogenética ”model-based”, esses modelos de substituição utilizam o mesmo arcabouço (processos de Markov) que métodos comparativos que iremos estudar, tais como correlação de caracteres binários e multiestado

O’Meara 2012

0	1	0



1

?

t

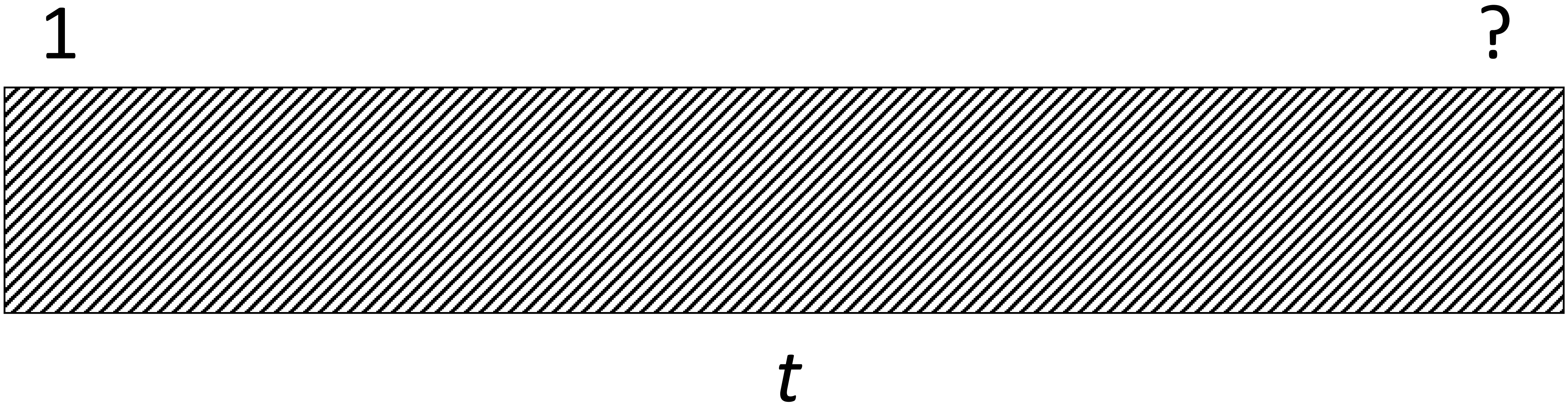
Matriz de transição instantânea

$$Q = \begin{bmatrix} -q01 & q01 \\ q10 & -q10 \end{bmatrix}$$

Matriz de transição instantânea

$$Q = \begin{bmatrix} -q01 & q01 \\ q10 & -q10 \end{bmatrix}$$

Soma da linhas = 0

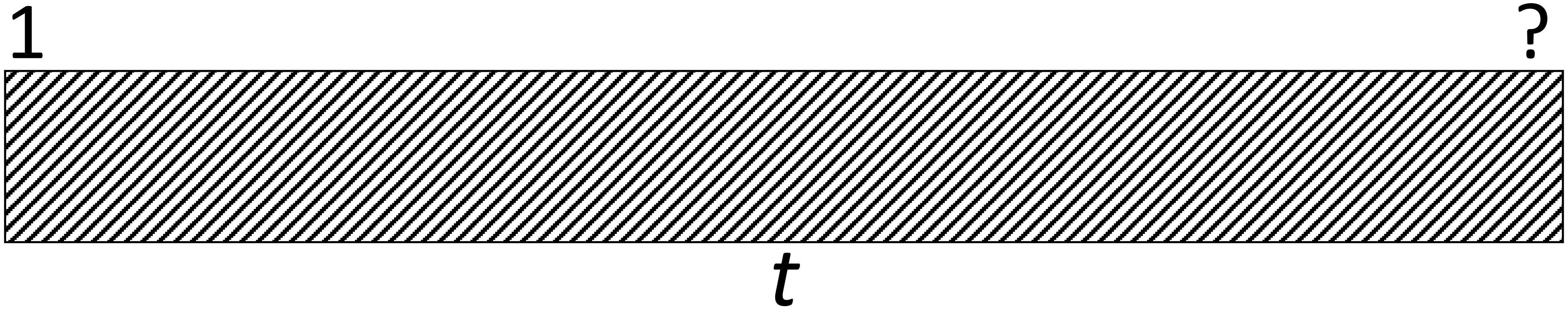


Qual a probabilidade de depois
do tempo *t* o estado 1 mudar para 0?

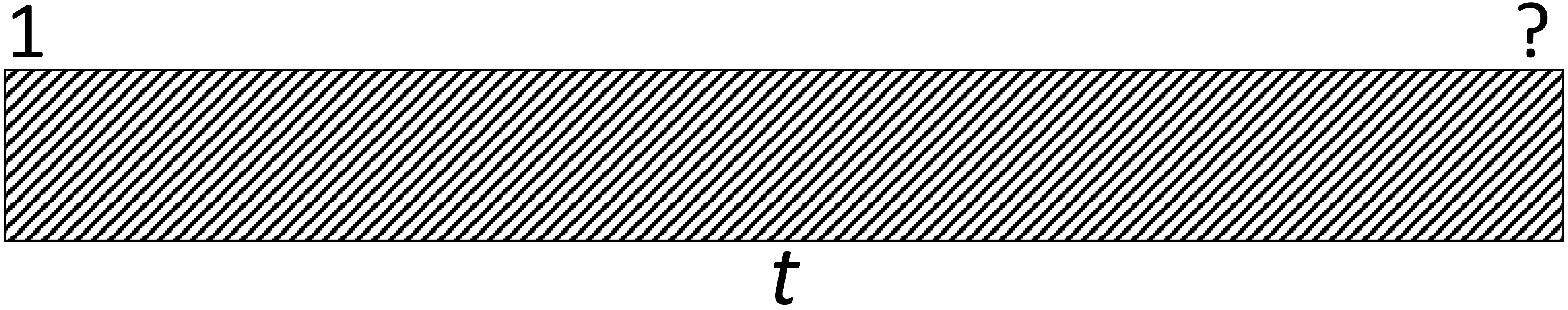
$P_{10}(t)$

Exponenciação de matriz

$$P(t)=e^{Qt}$$



$$Q = \begin{bmatrix} -q01 & q01 \\ q10 & -q10 \end{bmatrix}$$

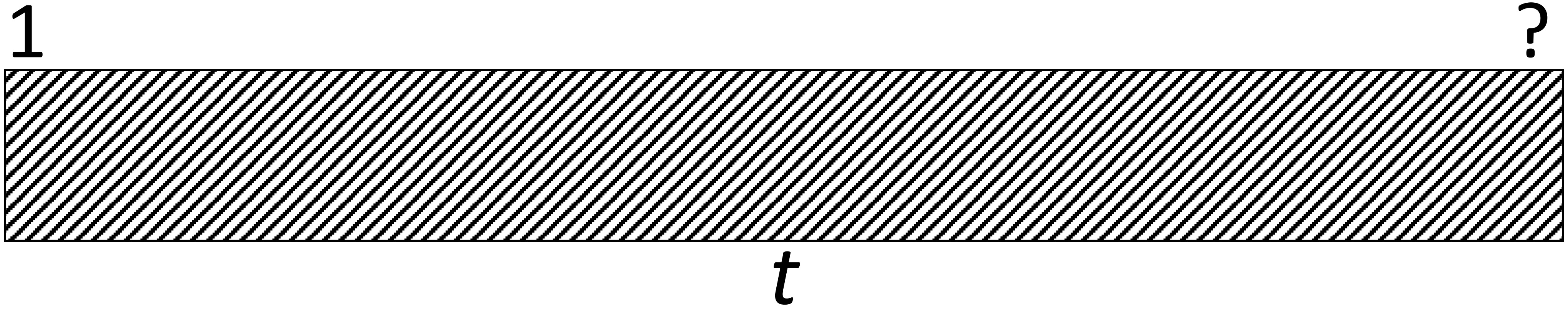


$$Q = \begin{bmatrix} -q01 & q01 \\ q10 & -q10 \end{bmatrix}$$

q01 = 0.1

q10 = 0.1

t = 5 mya

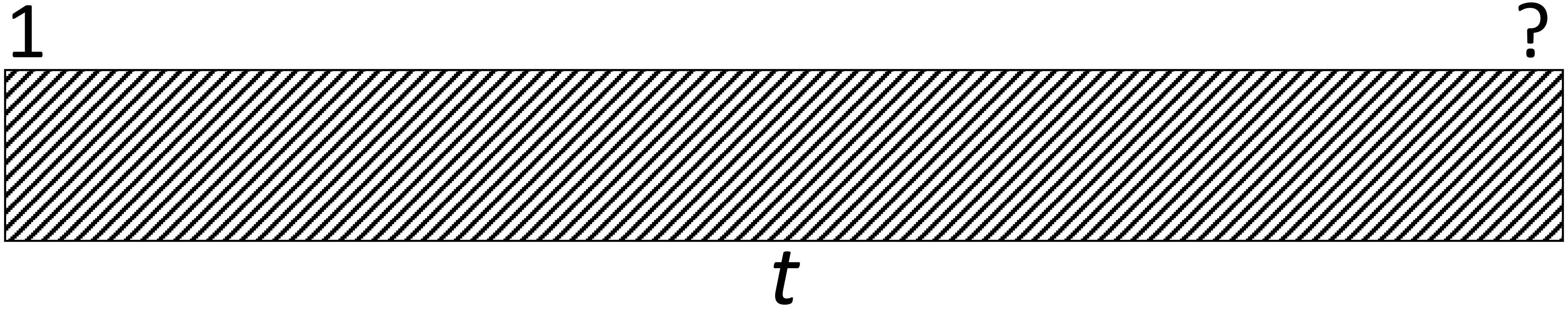


$q01 = 0.1$

$q10 = 0.1$

t = 5 mya

$$Qt = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}$$

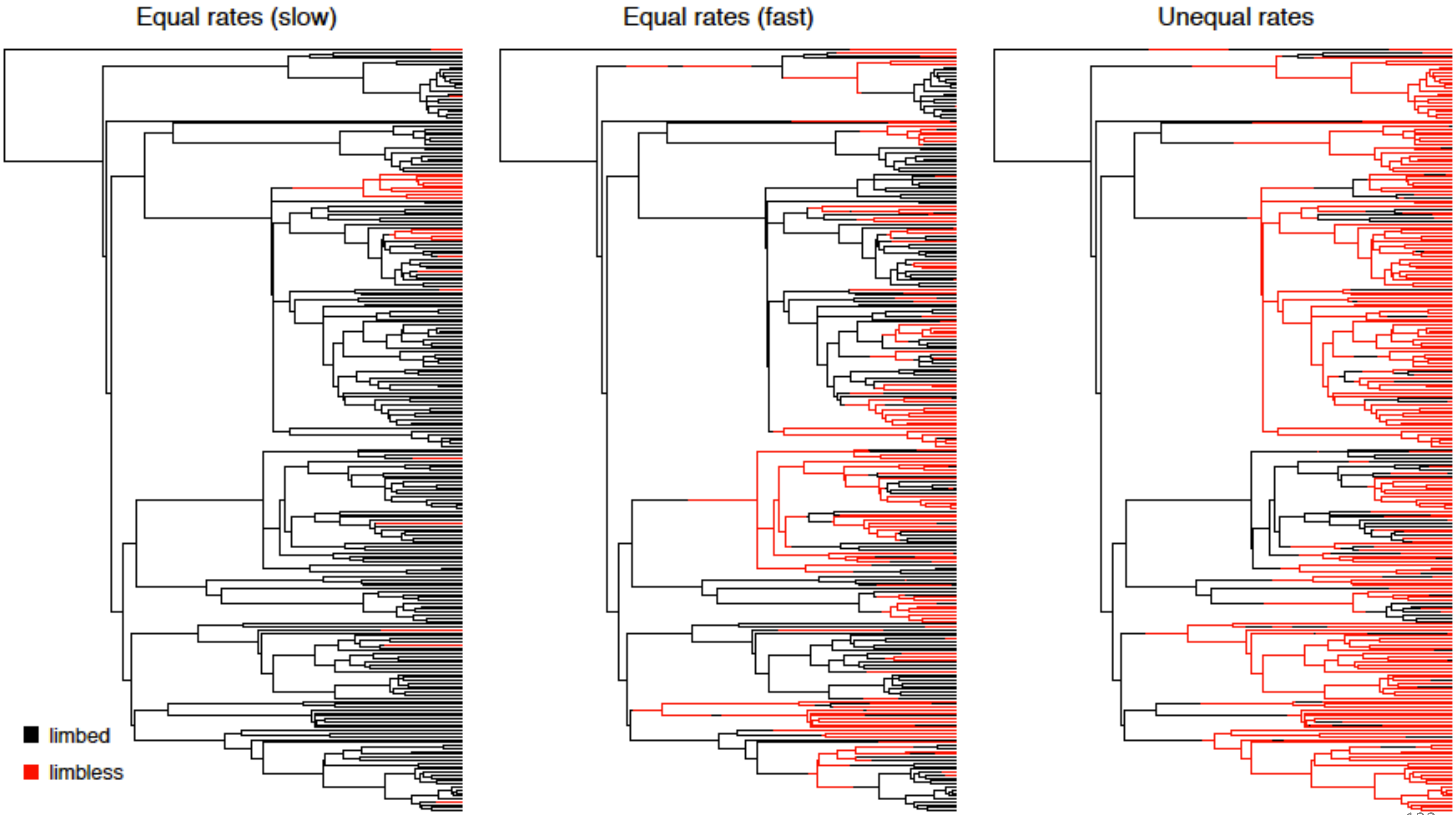


$$e^{Qt} = \begin{bmatrix} 0.68 & 0.32 \\ 0.32 & 0.68 \end{bmatrix}$$

q01 = 0.1

q10 = 0.1

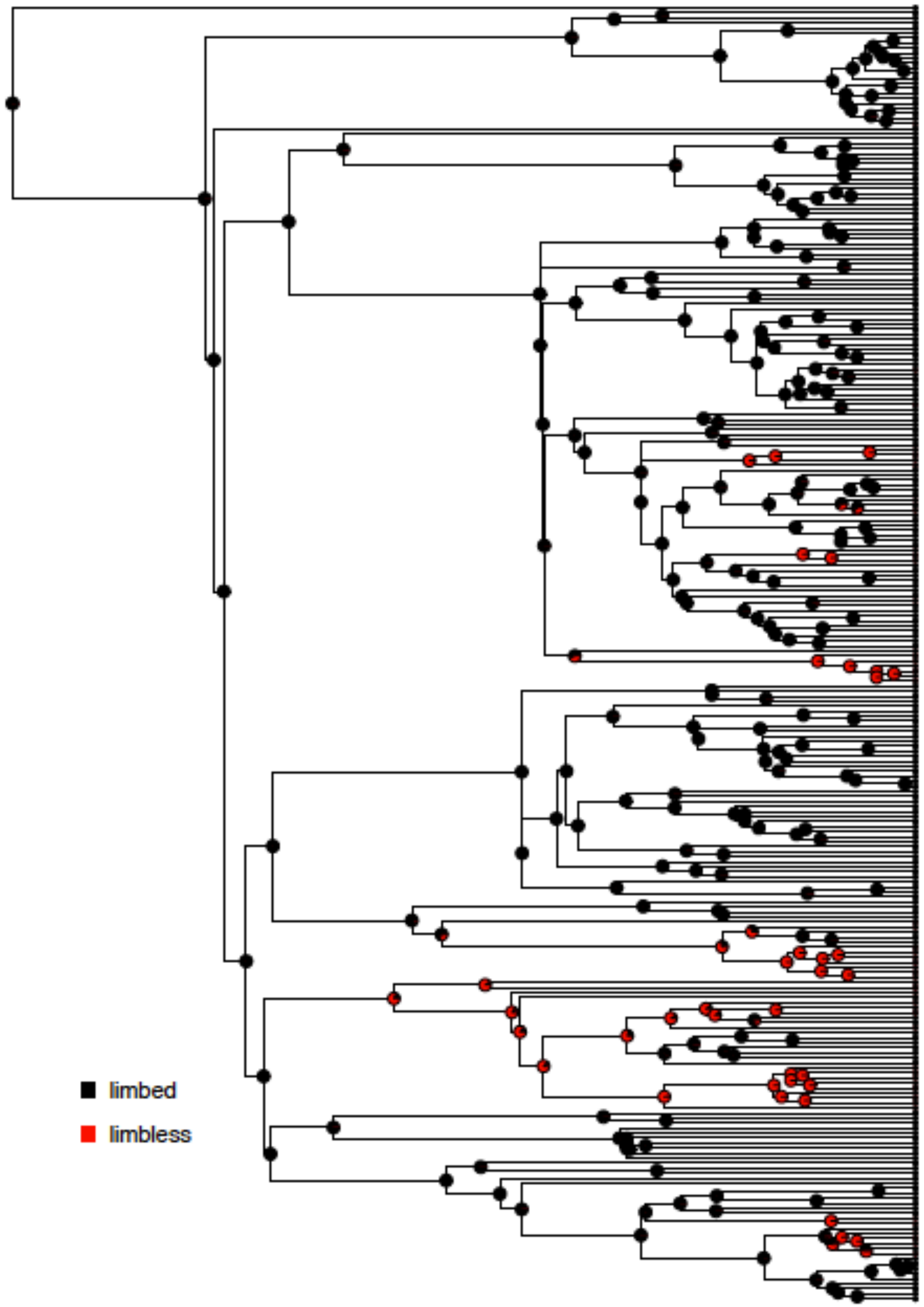
t = 5 mya



Equal rates (slow)

Equal rates (fast)

Unequal rates



Podemos também usar esses modelos para estimar estados ancestrais

Qual seria o resultado de uma análise com dados reais?

Equal rates

Sem membros

Com membros

Unequal rates

0

1

Sem membros

Com membros

Equal Rates Model

lnL = -80.5

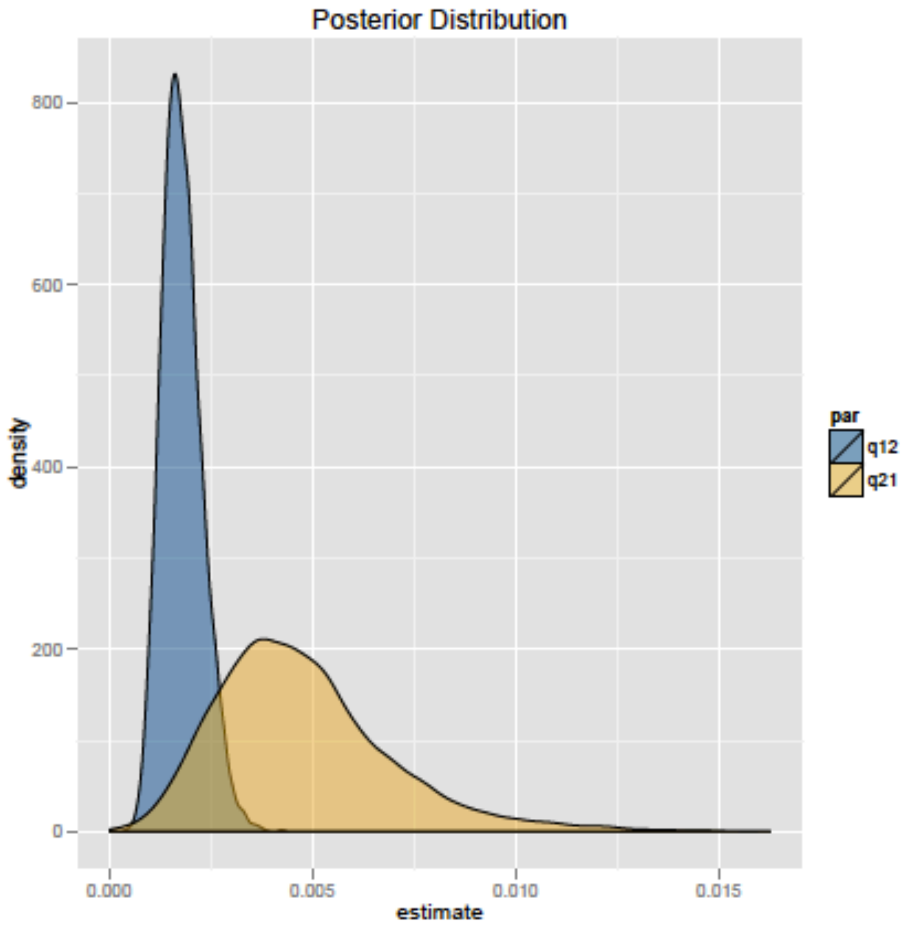
$$Q_{ER} = \begin{bmatrix} -0.0019 & 0.0019 \\ 0.0019 & -0.0019 \end{bmatrix}$$

Unequal Rates Model

lnL = -79.4

$$Q_{ARD} = \begin{bmatrix} -0.0016 & 0.0016 \\ 0.0038 & -0.0038 \end{bmatrix}$$

We can also do this with Bayesian analyses



geiger::fitDiscreteMCMC

geiger::fitContinuousMCMC

Caracteres multiestado

No limbs

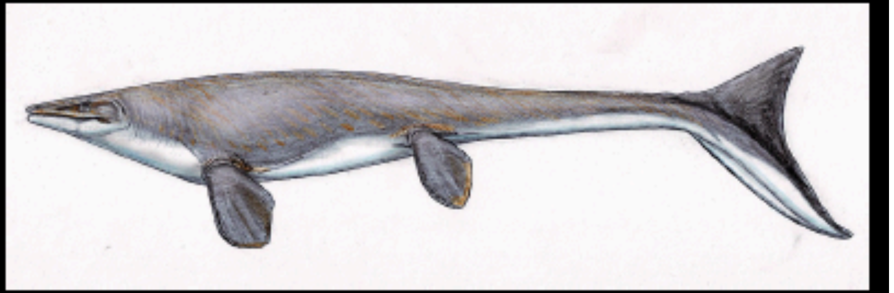
State 0

Front limbs only

State 1

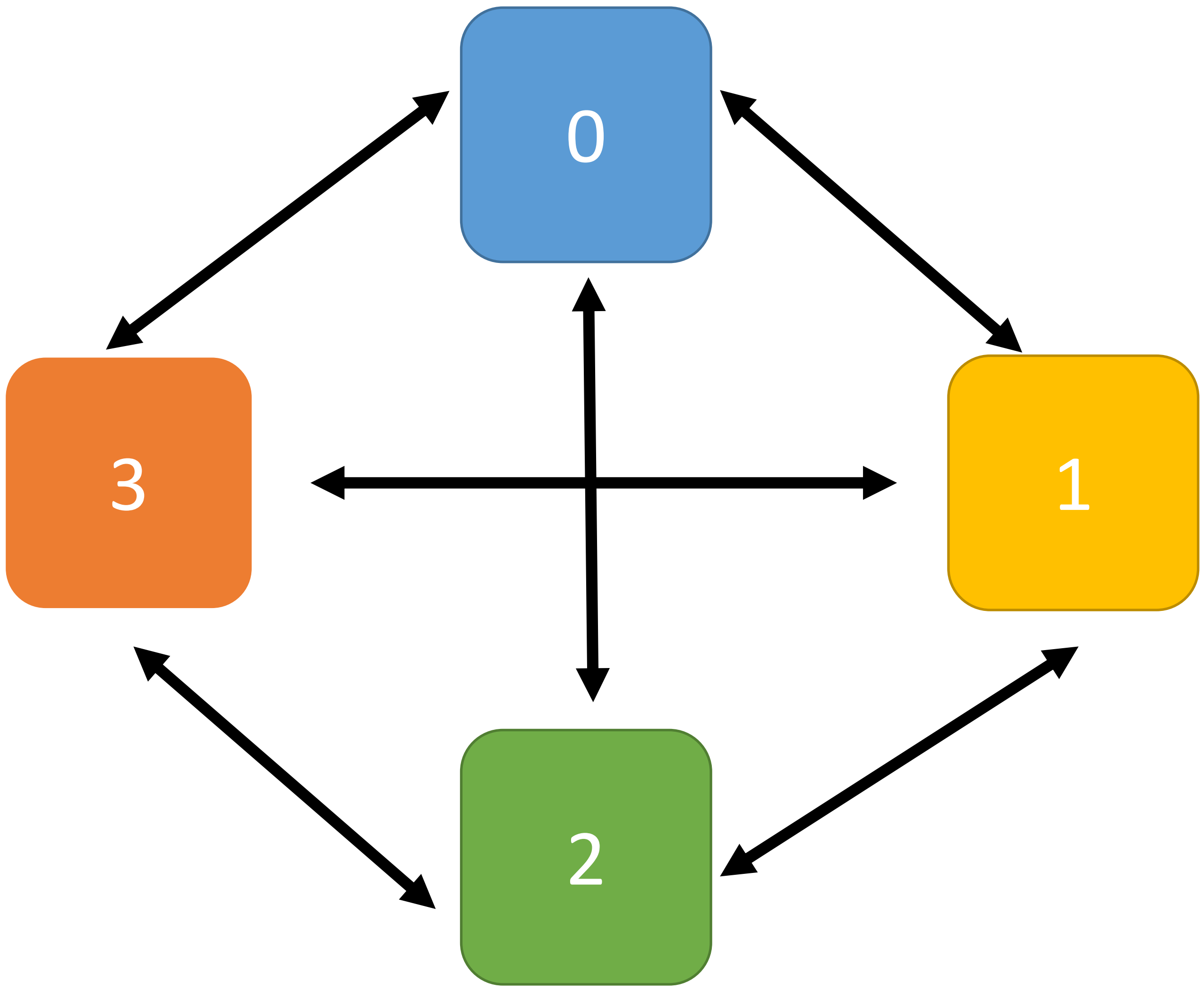
Limbs

State 2



“Fins”

State 3


0
3
1
2

Matriz de transição instantânea

$$Q = \begin{bmatrix} - & q01 & q02 & q03 \\ q10 & - & q12 & q13 \\ q20 & q21 & - & q23 \\ q30 & q31 & q32 & - \end{bmatrix}$$

Matriz de transição instantânea

$$Q = \begin{bmatrix} - & q01 & q02 & q03 \\ q10 & - & q12 & q13 \\ q20 & q21 & - & q23 \\ q30 & q31 & q32 & - \end{bmatrix}$$

Equal rates (ER)
1 parâmetro

Matriz de transição instantânea

$$Q = \begin{bmatrix} - & q01 & q02 & q03 \\ q10 & - & q12 & q13 \\ q20 & q21 & - & q23 \\ q30 & q31 & q32 & - \end{bmatrix}$$

Symmetric (SYM)
6 parâmetros

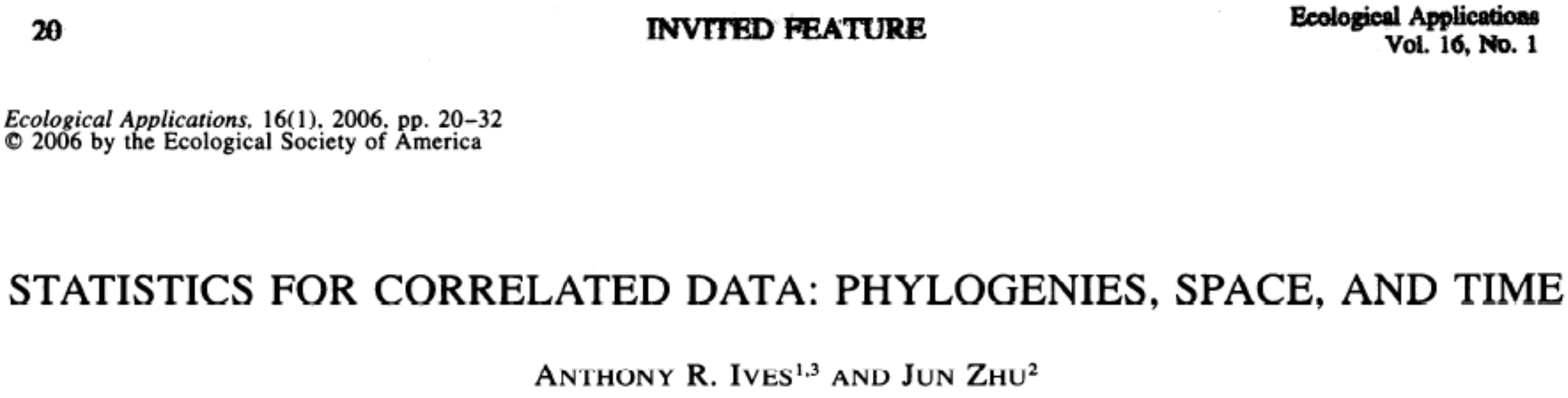
Matriz de transição instantânea

$$Q = \begin{bmatrix} - & q01 & q02 & q03 \\ q10 & - & q12 & q13 \\ q20 & q21 & - & q23 \\ q30 & q31 & q32 & - \end{bmatrix}$$

All Rates Differ (ARD)
12 parâmetros


YES!
AGORA BORA BOTAR A MÃO NA MASSA!

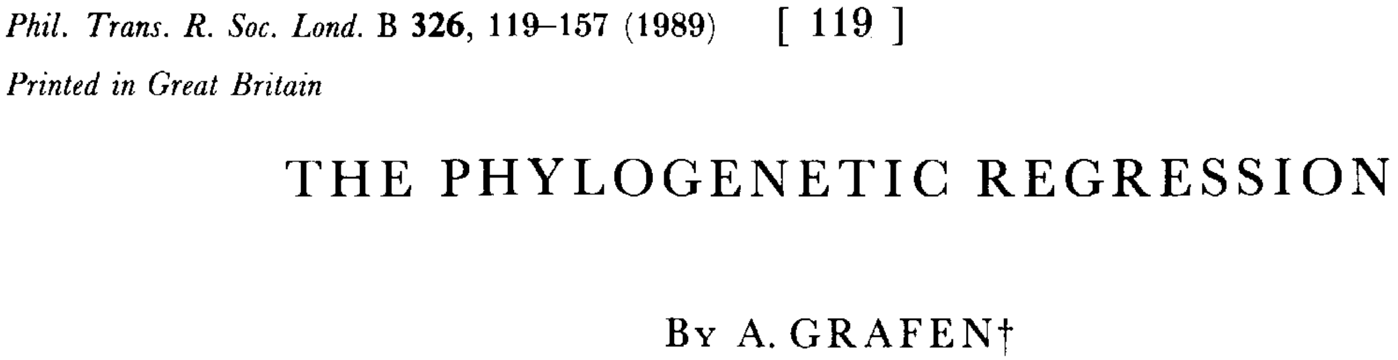
Mínimos Quadrados Generalizados (GLS)



Ecological Applications, 16(1), 2006, pp. 20–32

STATISTICS FOR CORRELATED DATA: PHYLOGENIES, SPACE, AND TIME

ANTHONY R. IVES[1,3] AND JUN ZHU[2]



Phil. Trans. R. Soc. Lond. B **326**, 119–157 (1989) [119]

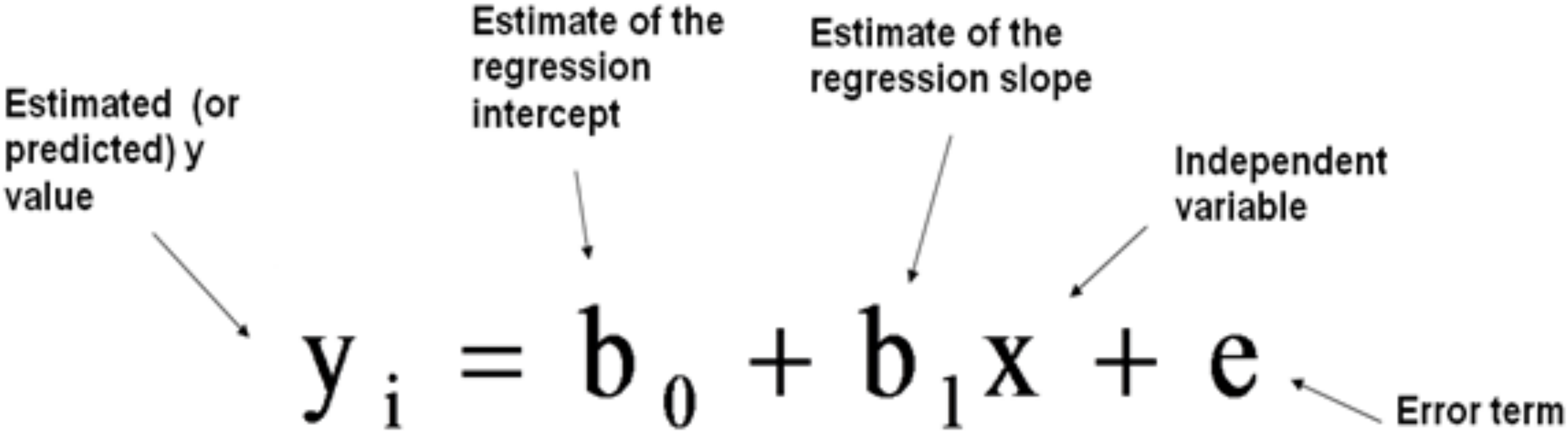
Printed in Great Britain

THE PHYLOGENETIC REGRESSION

By A. GRAFEN†

- Normalmente empregado para testar a associação entre dois atributos
- Muito flexível
  - Preditor e resposta contínuos
  - Preditores discretos binários ou ordinais (multiestado)
- Para resposta binária use uma regressão logística filogenética
  - Disponível no pacote phylolm::phyloglm
  - Ou uma abordagem Bayesiana no pacote MCMCglmm
- Pode incorporar diferentes modelos evolutivos
  - Manipulação da estrutura de correlação (matriz de variância-covariância)
  - Customizado ou já disponível

Antes, relembrando o modelo linear (OLS)



Estimated (or predicted) y value

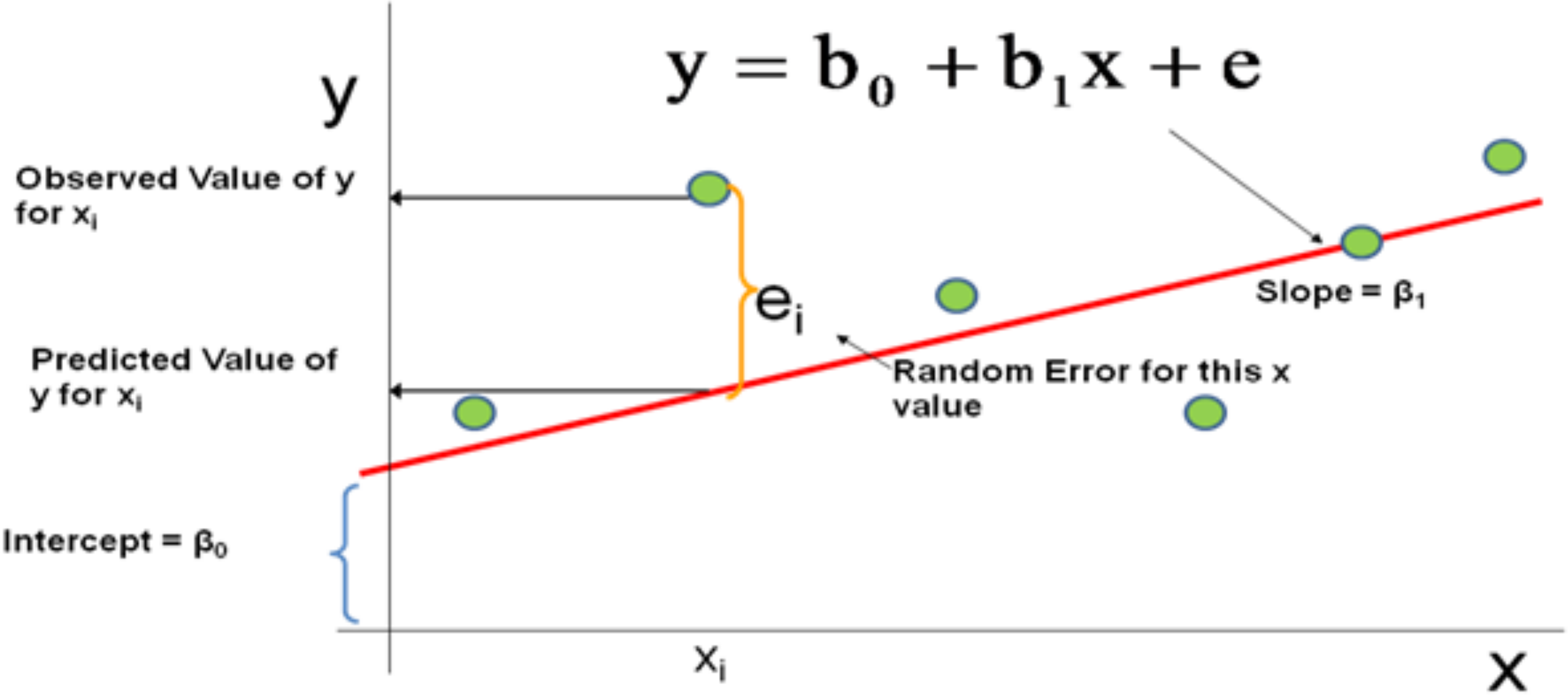
Estimate of the regression intercept

Estimate of the regression slope

Independent variable

Error term

$$y_i = b_0 + b_1 x + e$$



$$y = b_0 + b_1 x + e$$

y
Observed Value of y for x_i
Predicted Value of y for x_i
e_i
Random Error for this x value
Slope = β_1
Intercept = β_0
x_i
X

Pressupostos do OLS (Teorema Gauss-Markov)

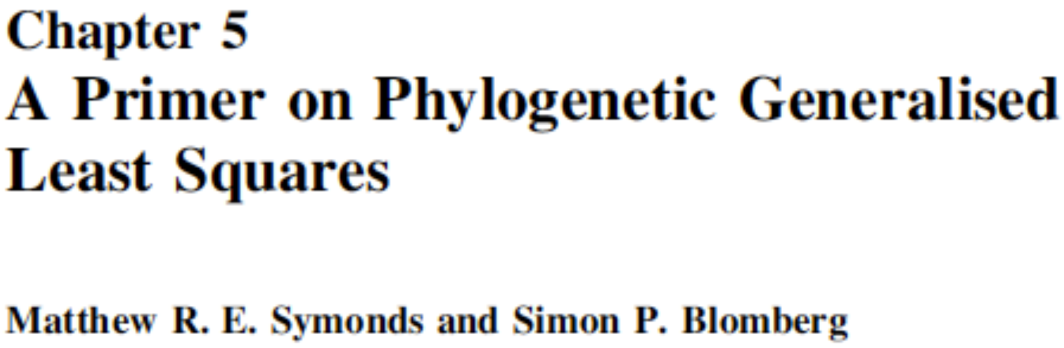
Matemática	Tradução pra língua humana
$y=\mathbf{X}\beta+\varepsilon$	Relação linear entre y e X
X é uma matriz n x k	Colunas de **X** são independentes
$E(\varepsilon\,\varepsilon'\|\mathbf{X})=\boldsymbol{\sigma}^2\mathbf{I}$	Homocedasticidade e ausência de autocorrelação
$\varepsilon\|\mathbf{X} \sim N(0,\boldsymbol{\sigma}^2\mathbf{I})$	Resíduos têm distribuição normal com média zero e variância homogênea

Pressuposto de independência e homoceidasticidade

Matemática	Tradução pra língua humana
$\text{var}[\varepsilon_i \vert X] = \boldsymbol{\sigma}^2$ para qualquer *i*	Variância no erro é a mesma ($\boldsymbol{\sigma}^2$) para qualquer elemento da matriz
$\text{cov}[\varepsilon_i, \varepsilon_j \vert X] = 0$ para qualquer *i*	Sabendo o valor da variável numa unidade amostral não nos diz nada do valor dela em outra u.a.

$$E(\epsilon\epsilon'|X) = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

$$E(\epsilon\epsilon'|X) = \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \sigma^2 I$$



Chapter 5
A Primer on Phylogenetic Generalised
Least Squares

Matthew R. E. Symonds and Simon P. Blomberg

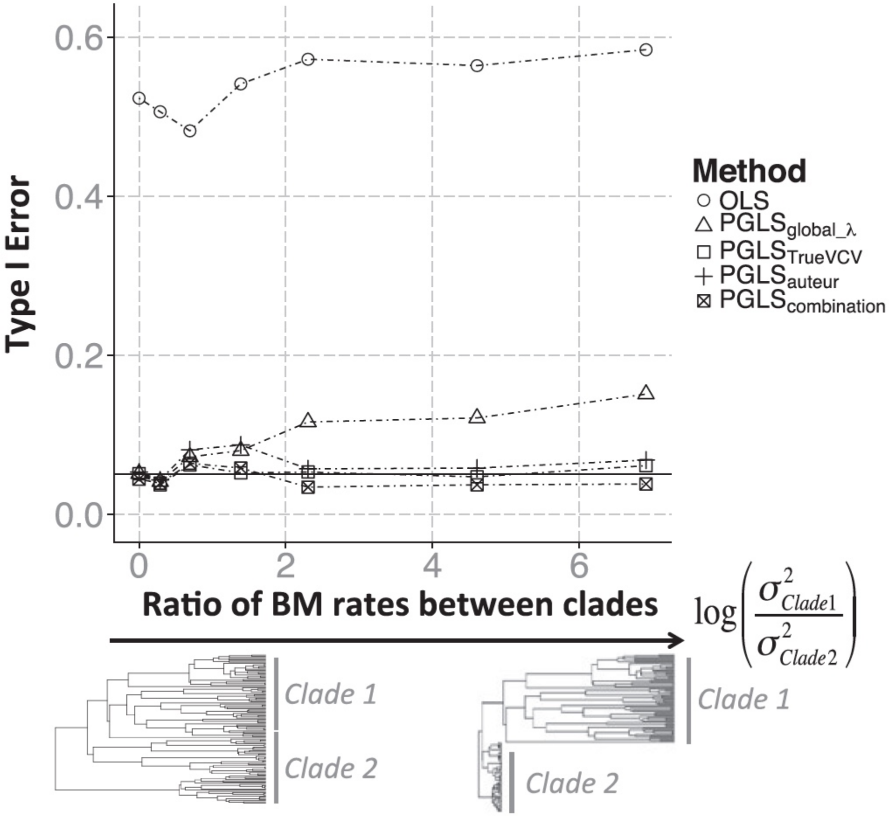
- Assume linearidade entre as variáveis resposta e preditora(s)

- Estimativa vai depender do modelo evolutivo incorporado

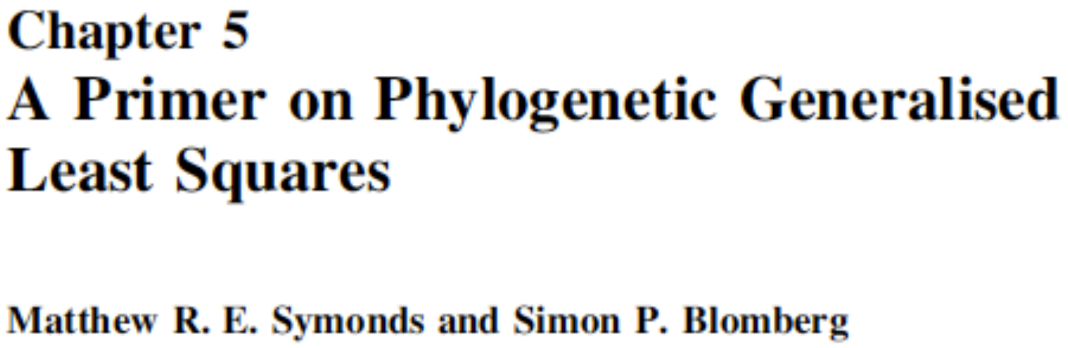
Ecology, 97(2), 2016, pp. 286–293

Improving phylogenetic regression under complex evolutionary models

Florent Mazel,[1,2,6] T. Jonathan Davies,[3,4] Damien Georges,[1,2] Sébastien Lavergne,[1,2]
Wilfried Thuiller,[1,2] and Pedro R. Peres-Neto[5]



regression. We found that PGLS has good power but unacceptable type I error rates. This finding is important since this method has been increasingly used in comparative analyses over the last decade. To address this issue, we propose a simple solution based on transforming the underlying variance–covariance matrix to adjust for model heterogeneity within PGLS. We suggest that heterogeneous rates of evolution might be particularly prevalent in large phylogenetic trees, while most current approaches assume a homogenous rate of evolution. Our analysis demonstrates that overlooking rate heterogeneity can result in inflated type I errors, thus misleading comparative analyses. We show that it is possible to correct for this bias even when the underlying model of evolution is not known a priori.



Chapter 5
A Primer on Phylogenetic Generalised Least Squares

Matthew R. E. Symonds and Simon P. Blomberg

- Assume linearidade entre as variáveis resposta e preditora(s)

- Estimativa vai depender do modelo evolutivo incorporado

- Erro comum num PGLS
  - Testar sinal filogenético nas variáveis resposta e/ou preditora

- O que é correto: testar sinal filogenético nos resíduos

Ajustando um PGLS

- O que é melhor: ajustar um modelo em que simultaneamente se estima o ”sinal” filogenético dos resíduos (usando ML), p.ex. usando o lambda do Pagel e se estima os parâmetros do modelo

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology & Evolution 2010, 1, 319–329

doi: 10.1111/j.2041-210X.2010.00044.x

Phylogenetic signal and linear regression on species data

Liam J. Revell*

Syst. Biol. 61(3):382–391, 2012
© The Author(s) 2012. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
DOI:10.1093/sysbio/syr118
Advance Access publication on January 3, 2012

Independent Contrasts and PGLS Regression Estimators Are Equivalent

SIMON P. BLOMBERG[1,*], JAMES G. LEFEVRE[2], JESSIE A. WELLS[1], AND MARY WATERHOUSE[3]

- Os estimadores dos Contrastes Filogenéticos Independentes (PIC) são equivalentes aos do PGLS quando se assume Movimento Browniano

Diagnosticando um PGLS