

Aula 3

Estimativa de estado ancestral de caracteres contínuos e categóricos

Estimativa e teste de "sinal" filogenético de caracteres contínuos e categóricos

Estimativa de estado ancestral para atributos
contínuos

O que vai ser coberto nesse módulo

- Teoria geral e motivação para estimar estados ancestrais
- Principais métodos
 - Squared-change parsimony
 - Maximum Likelihood
 - Generalised Least Squares (GLS)
 - Bayesiano
- Limitações
 - Efeito dos modelos de evolução nas estimativas
 - Muita incerteza associada com as estimativas
 - Ou seja, basicamente se você não precisa deles, não use! 😊

O que vai ser coberto nesse módulo

- Teoria geral e motivação para estimar estados ancestrais

- Principais métodos

- Squared-change parsimony
- **Maximum Likelihood**
- **Generalised Least Squares (GLS)**
- **Bayesiano**



Disponível na
função `ape::ace`



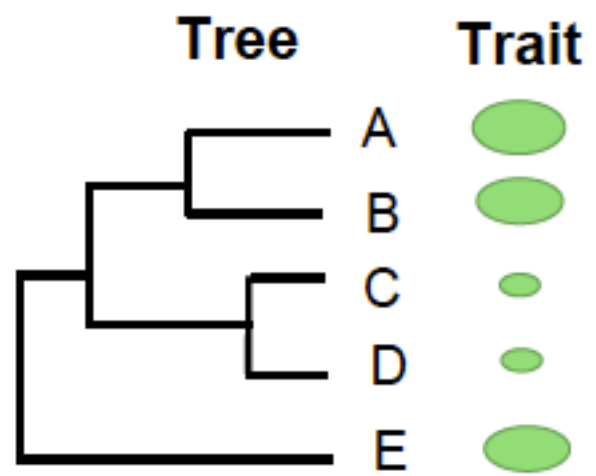
Métodos estatísticos
("model-based")

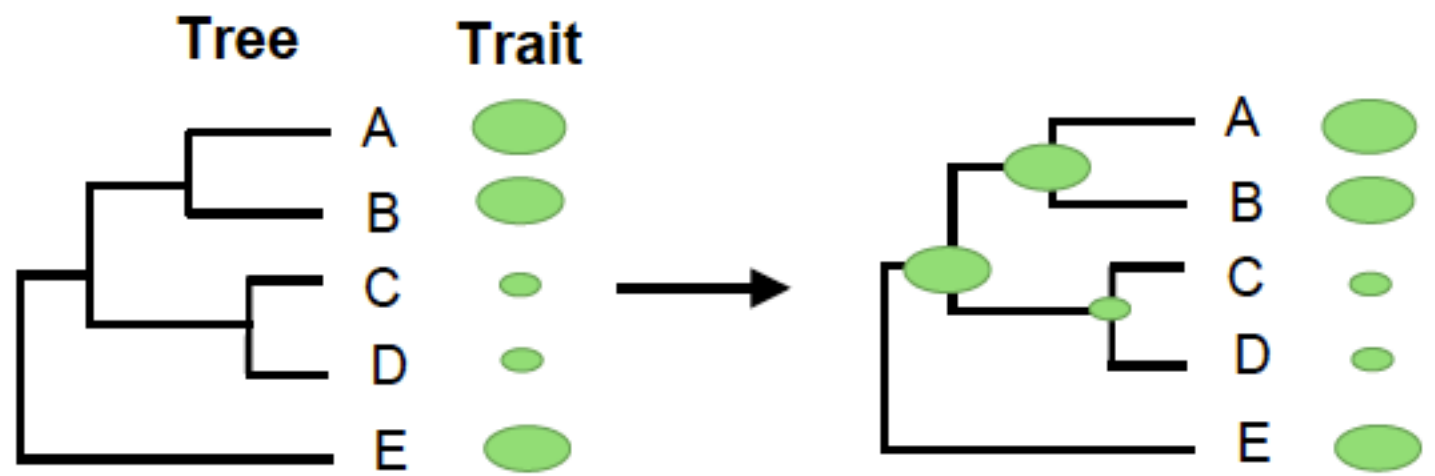
- Limitações

- Efeito dos modelos de evolução nas estimativas
- Muita incerteza associada com as estimativas
- Ou seja, basicamente se você não precisa deles, não use! 😊

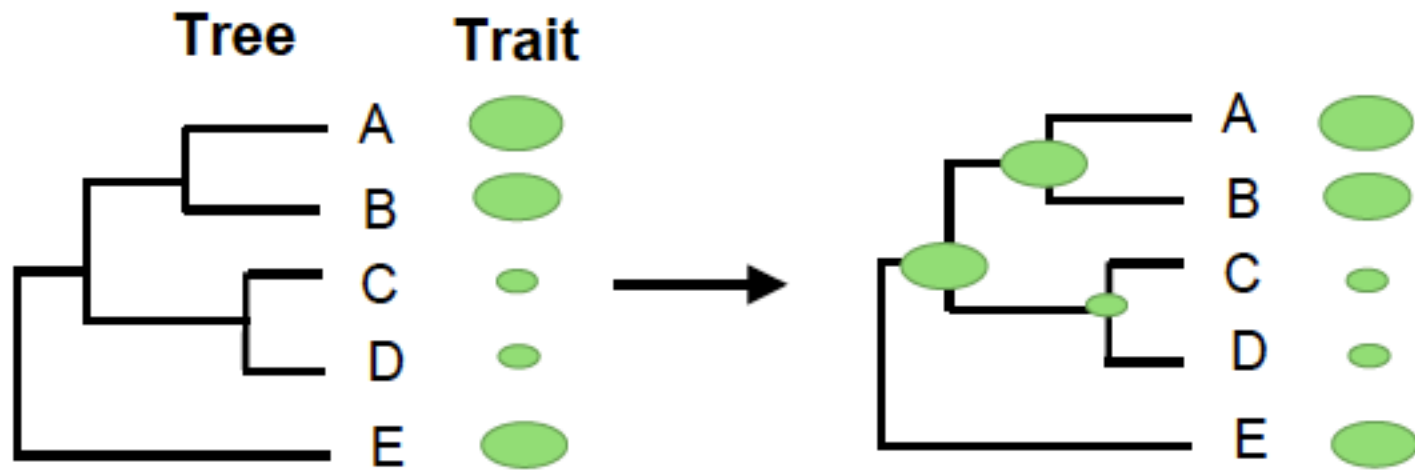
Porque estimar o estado ancestral de um caracter

- "Reconstuir" a história ou evolutiva sequência de mudanças num caracter, normalmente nos nós de filogenias
 - P.ex. entender como a colonização de uma ilha afetou a evolução de tamanho de corpo
- Correlacionar com alguma característica ambiental passada
- Idealmente precisamos obter medidas de incerteza associadas aos estados
 - Nem todos os métodos permitem calcular 95% IC ao redor das estimativas





Infer how traits have changed through time



Infer how traits have changed
through time

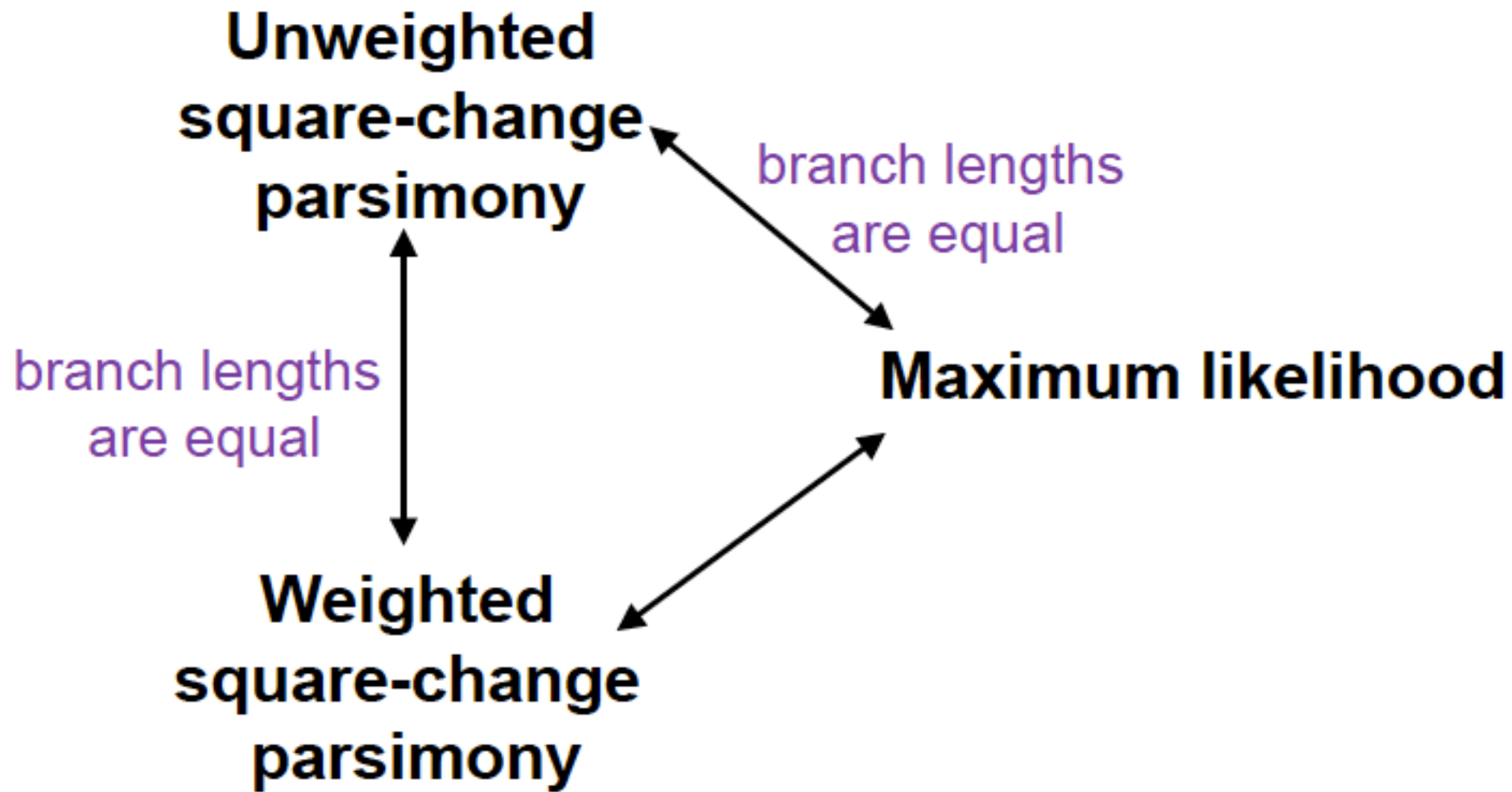
↓
Visualize/identify complex
macroevolutionary/macroecological patterns

Squared-change parsimony

- Basicamente funciona dentro do contexto de inferência filogenética de MP
- Minimiza mudanças ao longo da árvore
- Dois tipos: ponderada (weighted) e não ponderada (unweighted)
 - Não ponderada => ignora comprimentos de ramo (ok se tiver um cladograma antigo sem comprimento de ramo)
 - Ponderada => considera comprimentos de ramo
- Não permite calcular incerteza ao redor da estimativa
- Para caracteres discretos: assume que taxas de transição de estados é muito pequena. Logo, pode levar a resultados errados quando não é o melhor modelo

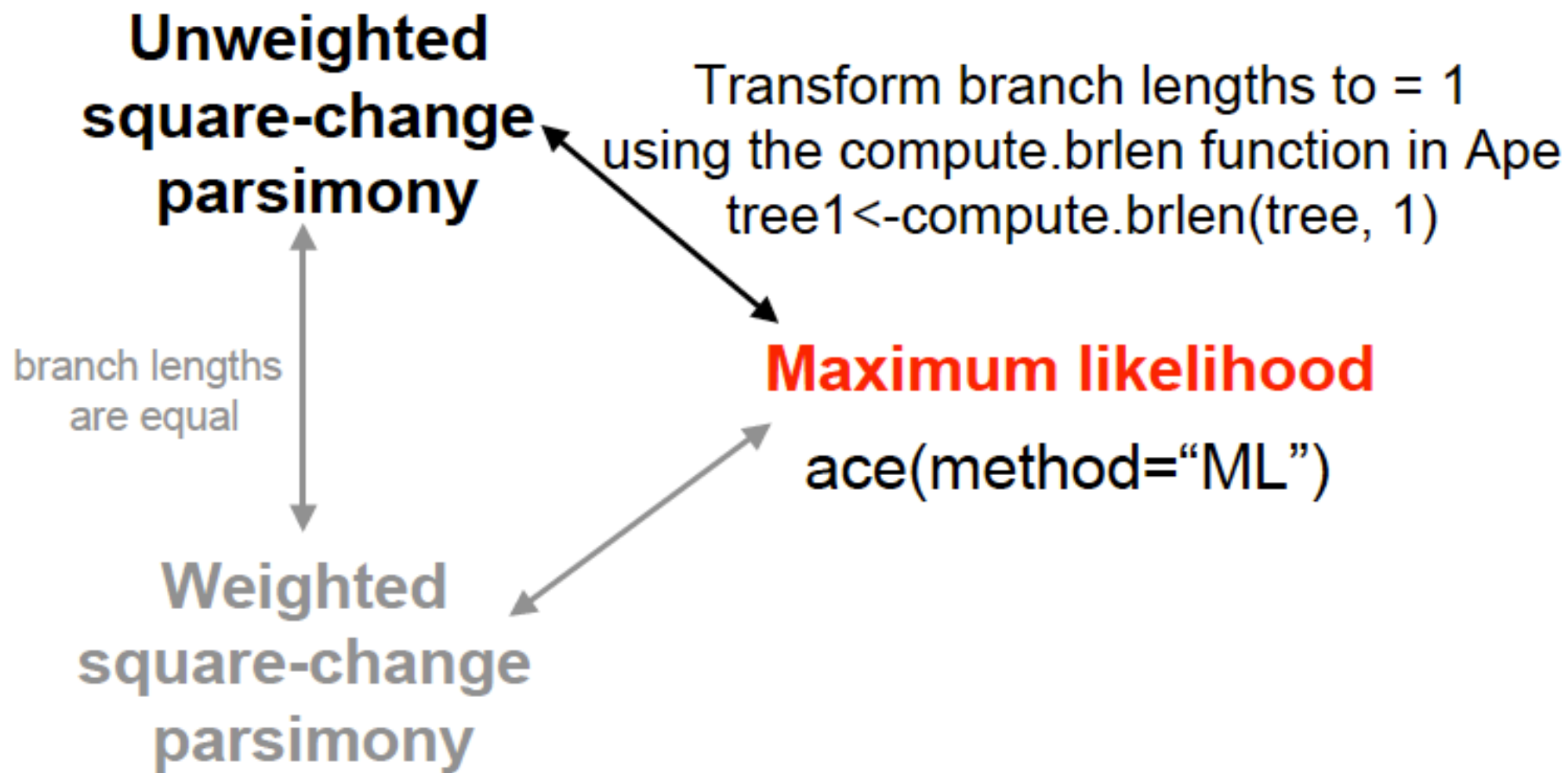
Maximum likelihood

- Conjunto de estados ancestrais mais verosimilar (likely) considerando toda a filogenia
- Incorpora comprimento de ramo
- Necessita de um modelo de evolução para estimar taxa de evolução
 - BM no ape
 - OU no bayou
 - Outra solução: "esticar" a árvore de acordo com um escalar
- Equivalente ao squared-change parsimony quando não árvore não tem comprimento de ramo



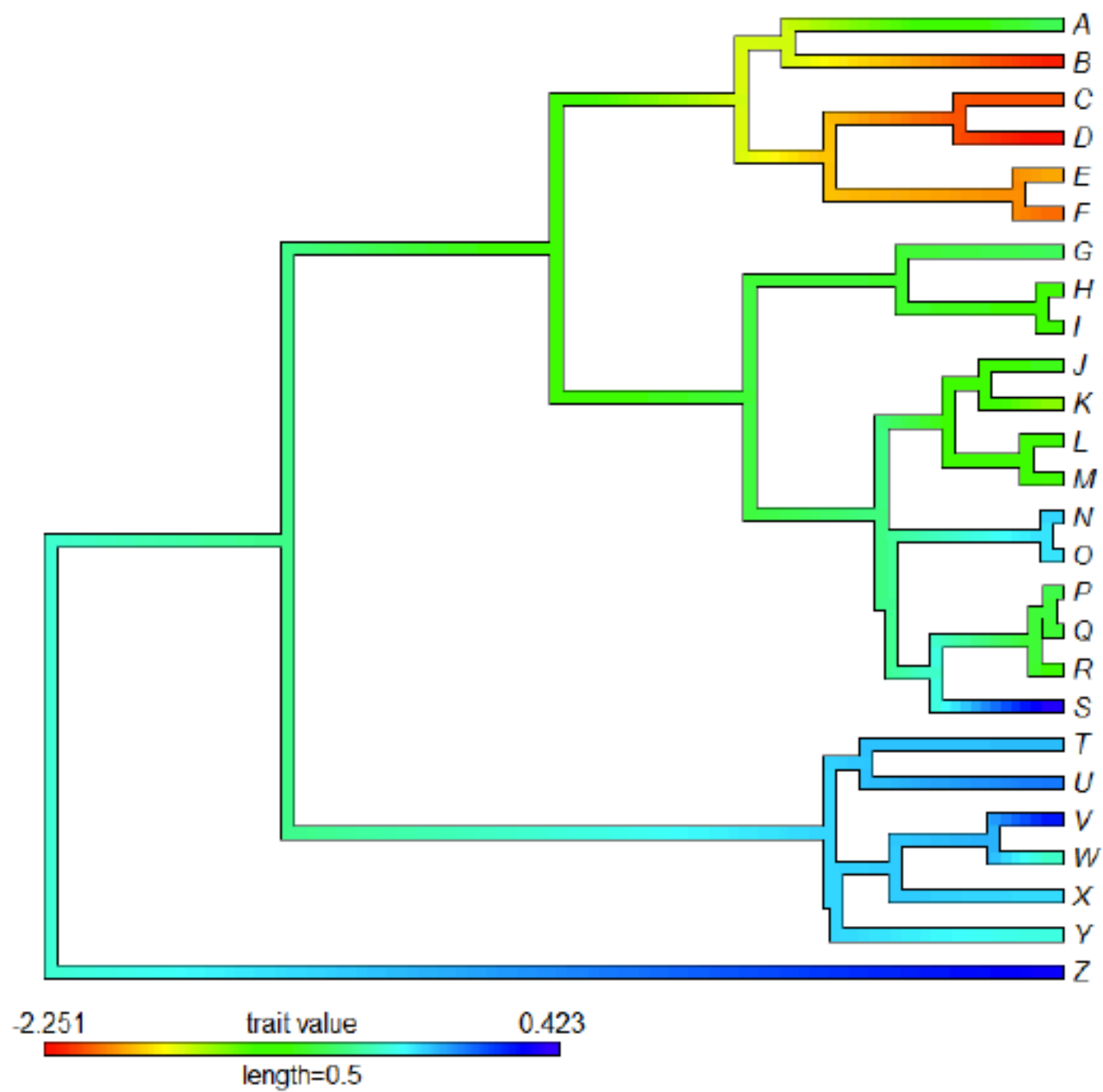


Implementation



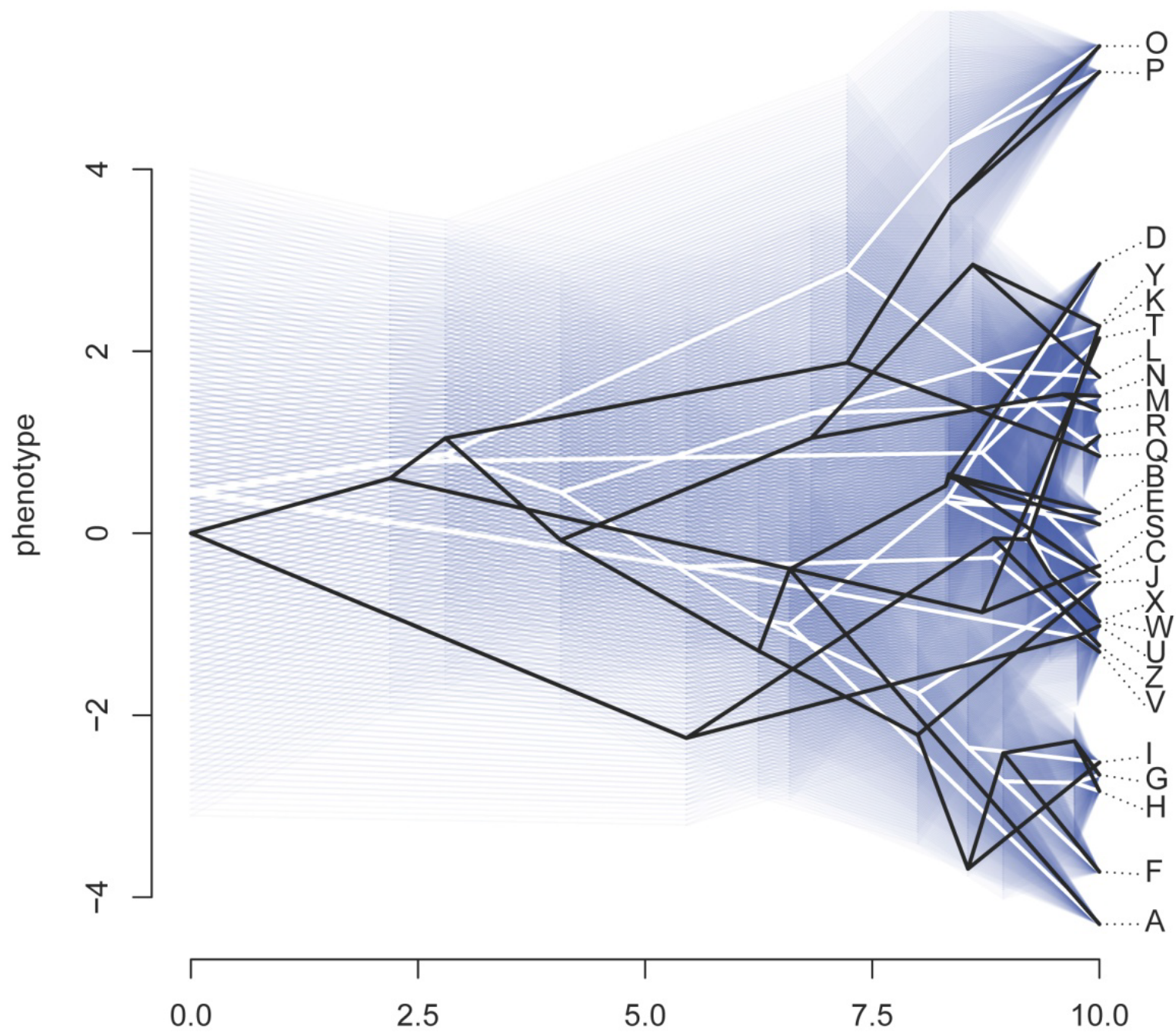
Caracteres contínuos: métodos estatísticos

- Valores dos estados ancestrais é tal que maximiza as probabilidades dado a árvore e os dados
 - Tentativa e erro a partir de um valor inicial => ineficiente para árvores grandes
 - "Pruning algorithm" que calcula os PICs
 - Contraste da raiz da filogenia é o MLE para a raiz
- Para estimar os estados ancestrais por ML precisamos de um modelo de evolução
 - Movimento Browniano é o mais comum
 - Diferença entre espécies é linearmente proporcional ao tempo de divergência
 - Permite estimar incerteza, calculando variâncias de cada estimativa para obter o 95% IC



Incertezas

- Variâncias ao redor das estimativas é sempre muito grande
- Aumenta à medida que se vai dos tips pra raiz => acúmulo de incertezas
- Raiz da árvore sempre terá a maior incerteza



Especificação do modelo

- Se especificarmos um modelo de evolução para estimar os estados ancestrais que não é o mais adequado para descrever a evolução daquele atributo, as estimativas serão erradas
 - Diferença entre BM e Trend
- Detecção de tendência é sempre difícil sem dados sobre fósseis
 - Métodos bayesianos permitem incorporar fósseis como *priors*

Caracteres discretos

Caracteres discretos

- Utiliza um modelo M_k (Markov k -state)

PRETTY SURE

**I'VE HEARD THAT ONE
BEFORE..**

Caracteres discretos

- Utiliza ML e um modelo Mk (Markov k -state) para ajustar aos dados
- Necessita de uma matriz Q de transição especificando as taxas instantâneas de transição entre estados de caracter
- A partir da matriz Q calcula-se a probabilidade de se observar um dado estado de caracter num dado tempo t , seguindo a fórmula

$$\mathbf{p}_t = \exp(\mathbf{Q}t)\mathbf{p}_0$$

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

Estimativa *joint* e *marginal*

- Joint => encontra o conjunto de caracteres que *conjuntamente* maximizam a verossimilhança *em todos os nós*
 - Restringe-se a *uma história* do caracter que é a mais verossimel
- Marginal => encontra o estado num dado nó que maximiza a verossimilhança integrando sobre todos os estados em todos os nós, proporcionalmente à sua probabilidade

Mapeamento estocástico de caracter

- Amostra estados de caracter nos nós e história dos caracteres (mudanças ao longo dos ramos) a partir das distribuição posterior Bayesiana conjunta
- Primenramente se amostra uma matriz Q a partir da distribuição de probabilidade posterior, então se amostra um conjunto de estados ancestrais nos nós a partir da sua distribuição de probabilidade conjunta condicional dado a matriz Q . Por fim, simula as mudanças ao longo dos ramos dado a matriz Q e a amostra dos estados nos nós.

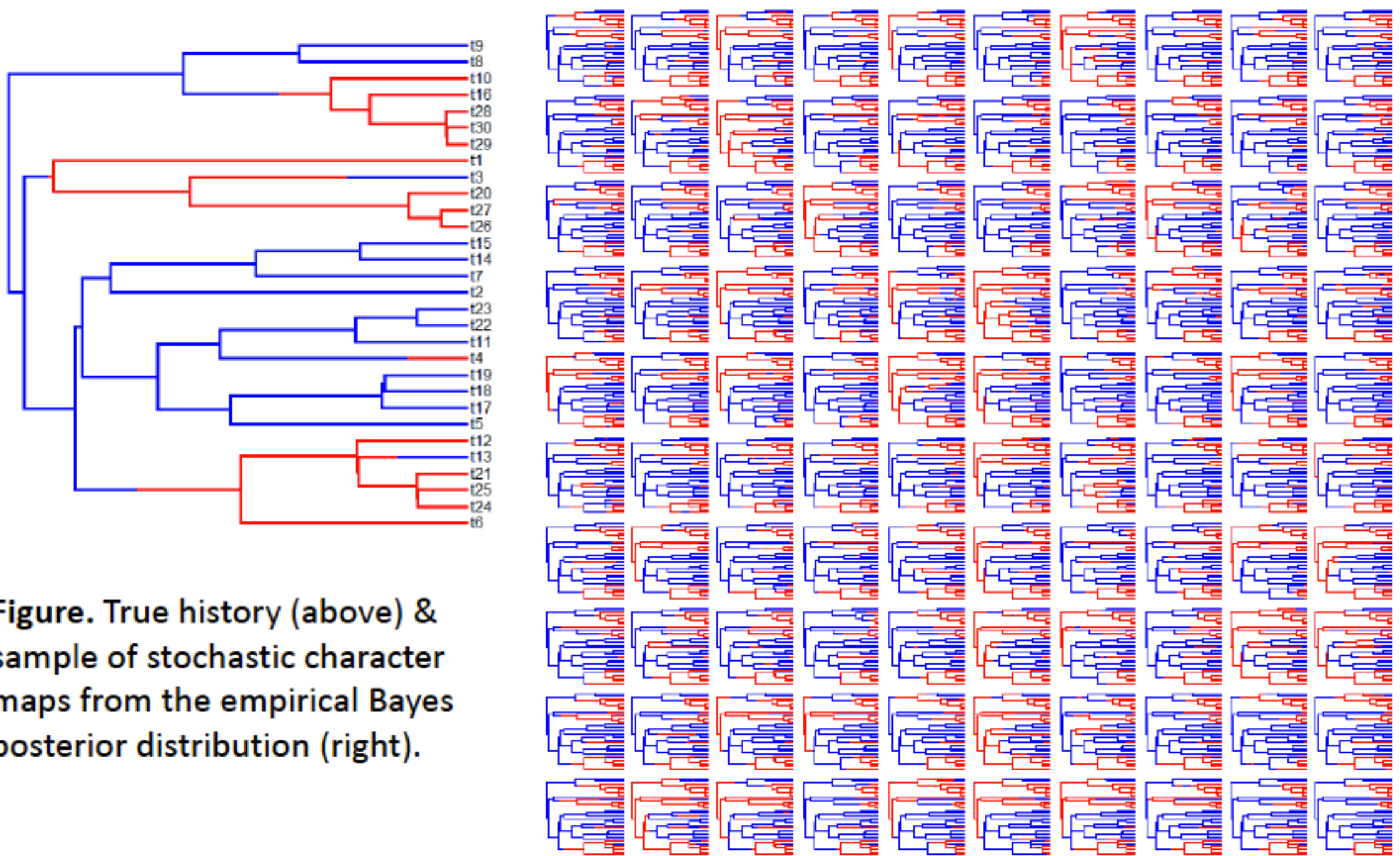


Figure. True history (above) & sample of stochastic character maps from the empirical Bayes posterior distribution (right).

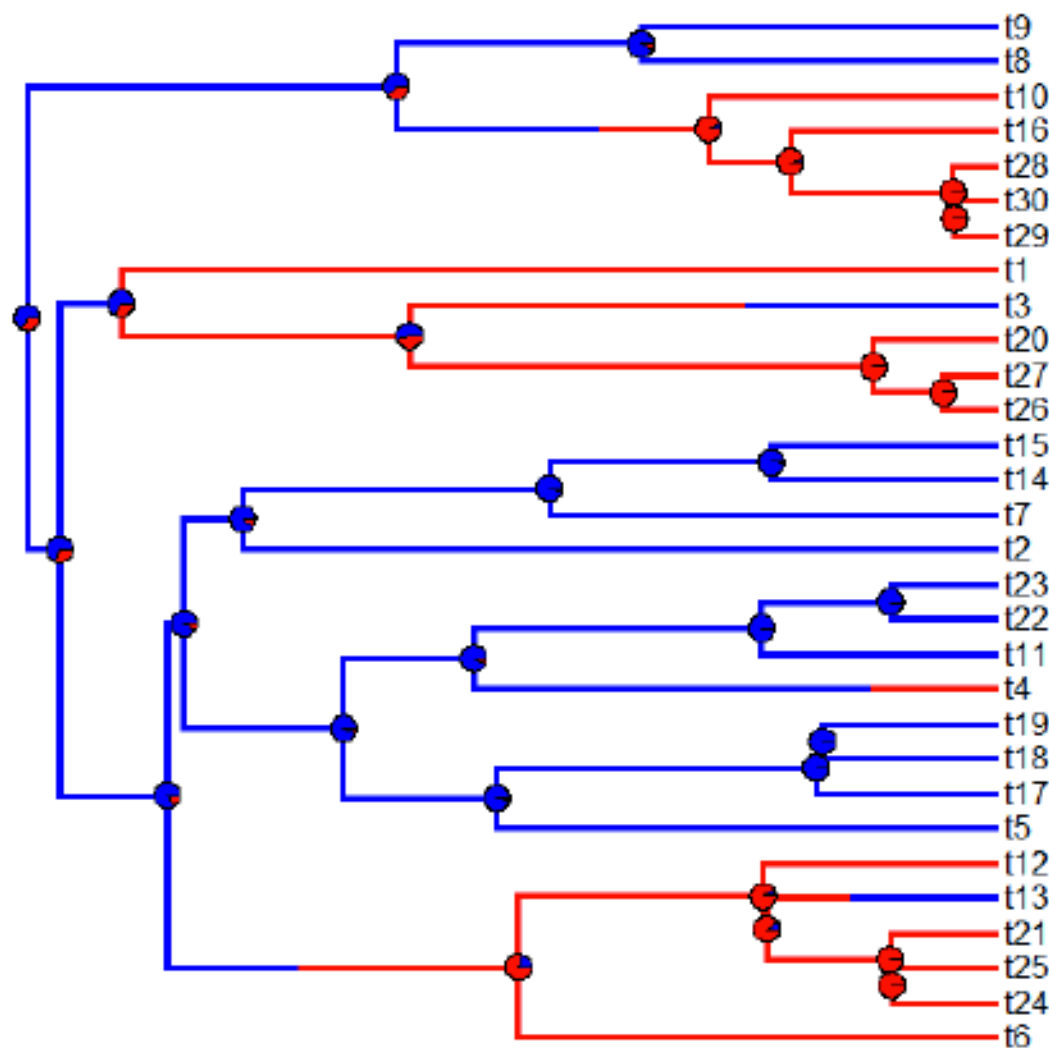


Figure. True history with posterior probabilities from stochastic mapping.

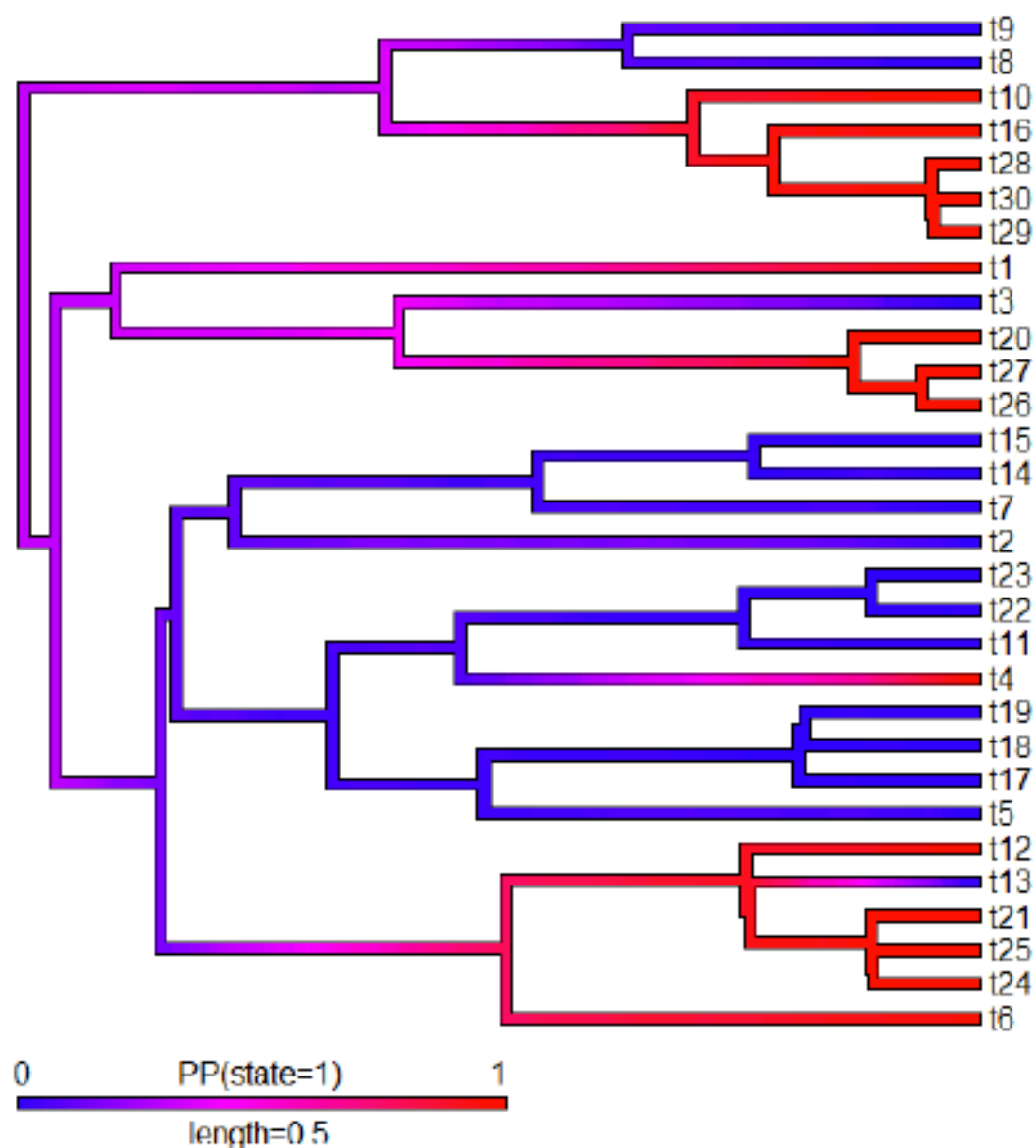


Figure. Posterior density map from stochastic mapping.

”Sinal” Filogenético de atributos contínuos e discretos

Definição

“A tendência de espécies relacionadas serem mais similares entre si do que com aquelas tomadas aleatoriamente na filogenia” (Blomberg & Garland 2002)

“Sinal”
filogenético

```
graph LR; A["Sinal filogenético"] --> B["INFERÊNCIA FILOGENÉTICA"]; A --> C["Literatura de método comparativo, ecofilogenética, e macroevolução"]; style B fill:#4a86e8,color:#fff; style C fill:#4a86e8,color:#fff;
```

INFERÊNCIA FILOGENÉTICA
“Habilidade de um caracter
de reconstruir a história
evolutiva de espécies”

Literatura de método
comparativo, ecofilogenética,
e macroevolução (definição
slide anterior)

Phylogenetic Signal, Evolutionary Process, and Rate

LIAM J. REVELL,¹ LUKE J. HARMON,² AND DAVID C. COLLAR¹

- Vários modelos evolutivos podem gerar um padrão de “sinal” filogenético
 - BM
 - OU com alpha pequeno
 - OU com múltiplos picos
 - Early burst

”sinal” filogenético é um padrão e não um processo

”sinal” filogenético não é sinônimo de
restrição (*constraint*)

Modelos não restritos (BM) podem gerar alto sinal filogenético,
enquanto modelos com restrição (tipo OU) podem gerar baixo sinal

Métodos para medir sinal filogenético

Atributos contínuos

- K de Blomberg
- Lambda de Pagel
- I de Moran
- C de Abouheif
- R^2 do PVR
- Correlação de Mantel
- Teste de viés para decomposição da entropia quadrática nos nós de uma filogenia

Atributos categóricos

- D de Fritz (binários)
- Teste de viés para decomposição da entropia quadrática nos nós de uma filogenia (binários e multiestado)
- Correlação de Mantel

How to measure and test phylogenetic signal

Tamara Münkemüller^{1*}, Sébastien Lavergne¹, Bruno Bzeznik², Stéphane Dray³,
Thibaut Jombart⁴, Katja Schiffers¹ and Wilfried Thuiller¹



Genetics and Molecular Biology Online Ahead of Print
Copyright © 2012, Sociedade Brasileira de Genética. Printed in Brazil
www.sbg.org.br

A comparison of metrics for estimating phylogenetic signal under alternative evolutionary models

José Alexandre F. Diniz-Filho¹, Thiago Santos², Thiago Fernando Rangel¹ and Luis Mauricio Bini¹

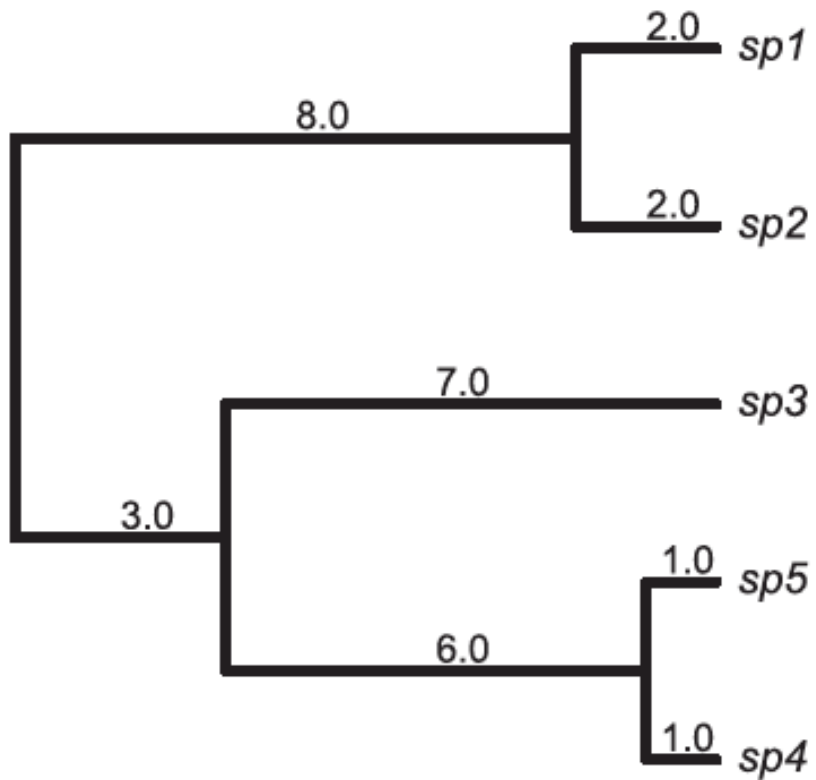
K de Blomberg

- K de Blomberg expressa a força do sinal filogenético como a razão entre a média quadrada do erro das espécies (MSE_0) medido da média corrigida pela filogenia e a média quadrada do erro baseada na matriz de variância-covariância derivada da filogenia (assumindo o Movimento Browniano)

$$\frac{MSE_0}{MSE} = \left(\frac{1}{n-1} \right) \left(tr \mathbf{V} - \frac{n}{\Sigma \Sigma \mathbf{V}^{-1}} \right), \quad K = \text{observed} \frac{MSE_0}{MSE} / \text{expected} \frac{MSE_0}{MSE}.$$

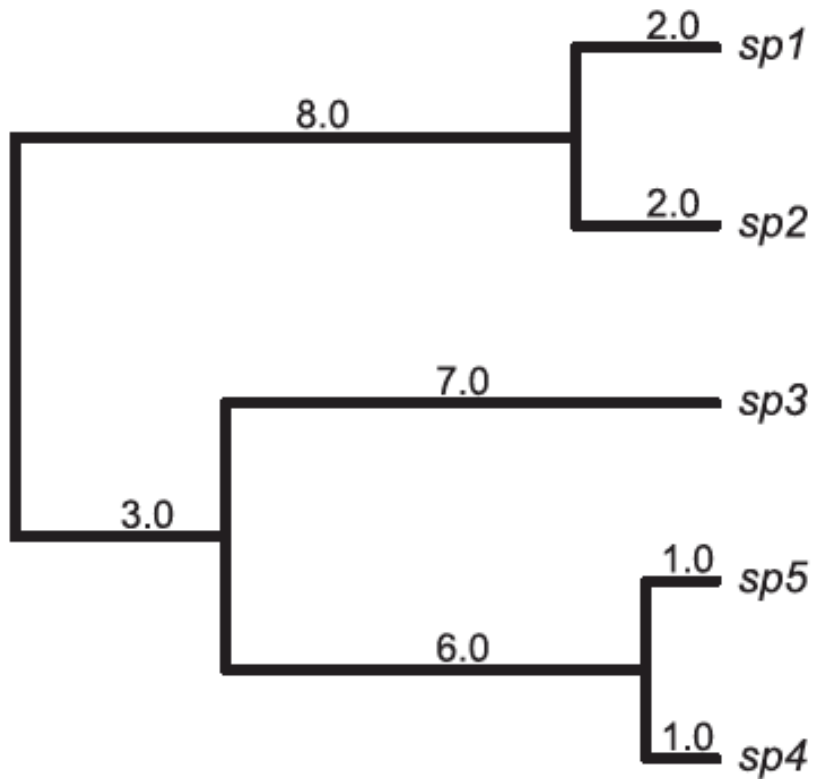
- Se a similaridade nos valores dos atributos forem preditos pela filogenia, o MSE será pequeno e o número MSE_0/MSE será grande
- Vantagem da medida: devido à essa padronização pelo esperado sob o BM, os valores podem ser comparados entre filogenias de diferentes tamanhos e formatos

lambda de Pagel



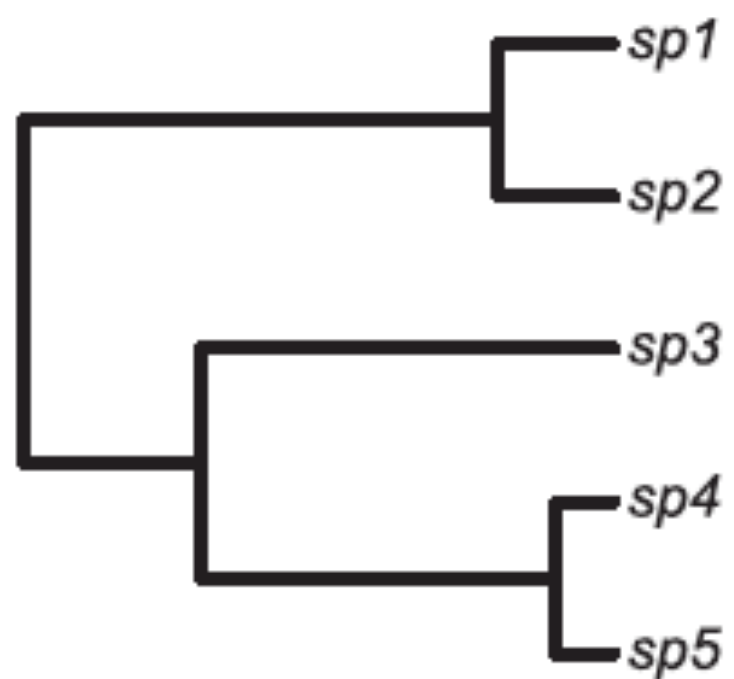
	<i>sp1</i>	<i>sp2</i>	<i>sp3</i>	<i>sp4</i>	<i>sp5</i>
<i>sp1</i>	10	8	0	0	0
<i>sp2</i>	8	10	0	0	0
<i>sp3</i>	0	0	10	3	3
<i>sp4</i>	0	0	3	10	9
<i>sp5</i>	0	0	3	9	10

Lambda de Pagel



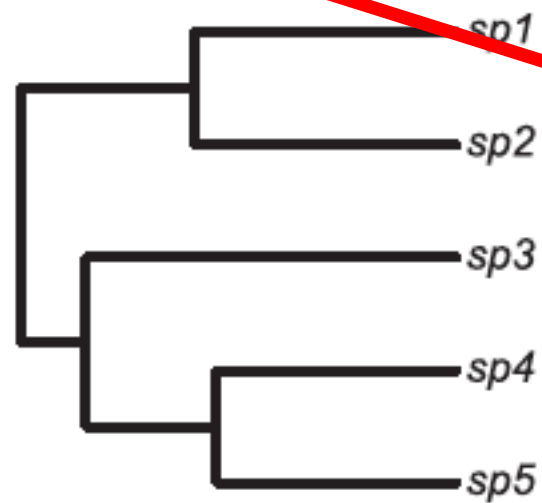
	<i>sp1</i>	<i>sp2</i>	<i>sp3</i>	<i>sp4</i>	<i>sp5</i>
<i>sp1</i>	10	$\lambda \cdot 3$	0	0	0
<i>sp2</i>	$\lambda \cdot 3$	10	0	0	0
<i>sp3</i>	0	0	10	$\lambda \cdot 3$	$\lambda \cdot 3$
<i>sp4</i>	0	0	$\lambda \cdot 3$	10	$\lambda \cdot 9$
<i>sp5</i>	0	0	$\lambda \cdot 3$	$\lambda \cdot 9$	10

$$\lambda = 1$$



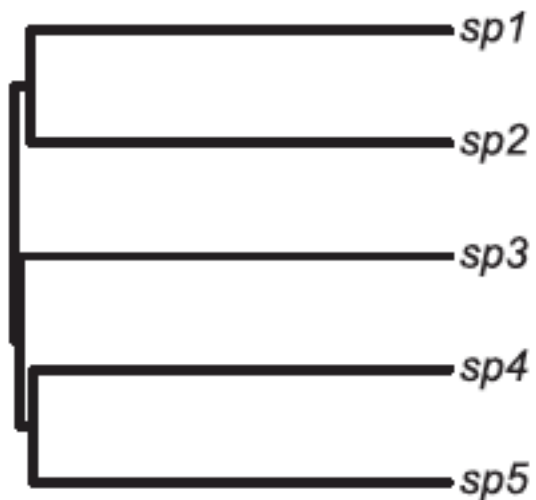
	<i>sp1</i>	<i>sp2</i>	<i>sp3</i>	<i>sp4</i>	<i>sp5</i>
<i>sp1</i>	10	8	0	0	0
<i>sp2</i>	8	10	0	0	0
<i>sp3</i>	0	0	10	3	3
<i>sp4</i>	0	0	3	10	9
<i>sp5</i>	0	0	3	9	10

$\lambda = 0.5 * 8$



	sp1	sp2	sp3	sp4	sp5
sp1	10	4	0	0	0
sp2	4	10	0	0	0
sp3	0	0	10	1.5	1.5
sp4	0	0	1.5	10	4.5
sp5	0	0	1.5	4.5	10

$\lambda = 0.05$



	sp1	sp2	sp3	sp4	sp5
sp1	10	0.4	0	0	0
sp2	0.4	10	0	0	0
sp3	0	0	10	0.15	0.15
sp4	0	0	0.15	10	0.45
sp5	0	0	0.15	0.45	10

I de Moran

Número de spp

Matriz **W** de pesos (inverso das distâncias filogenéticas entre as spp)

Característica nas spp *i* e *j* (centrada pela média), gerando uma matriz **Z** (Z_{ij})

$$I = \frac{\sum_{i,j} w_{ij} z_i z_j}{\sum_{i=1}^n w_{ij} z_i^2}$$

Soma de pesos na matriz **W**

ADAPTATION: STATISTICS AND A NULL MODEL FOR ESTIMATING PHYLOGENETIC EFFECTS

JOHN L. GITTLEMAN¹ AND MARK KOT²

¹*Department of Zoology and Graduate Programs in Ecology and Ethology,
University of Tennessee, Knoxville, Tennessee 37996-0810*

²*Department of Applied Mathematics, University of Washington, Seattle, Washington 98195*

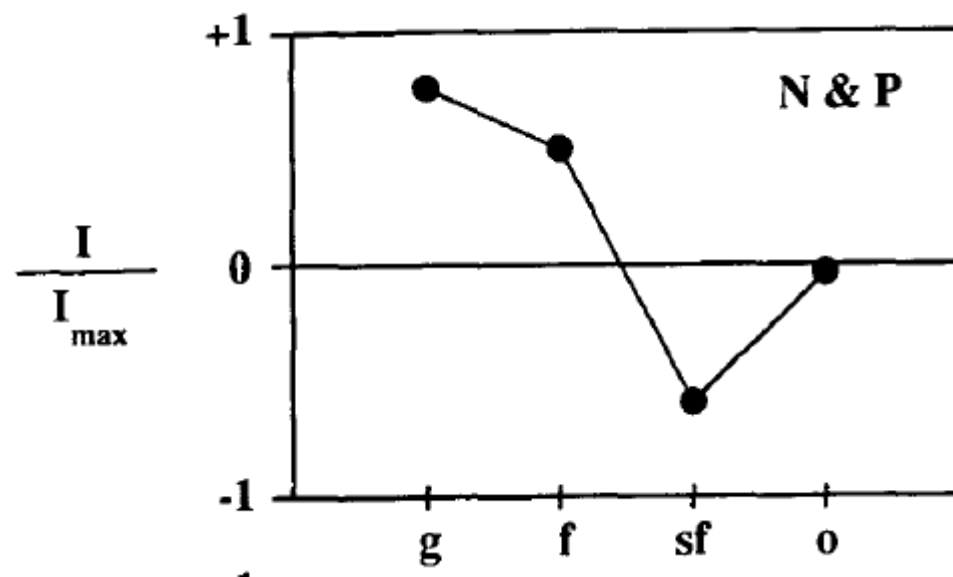
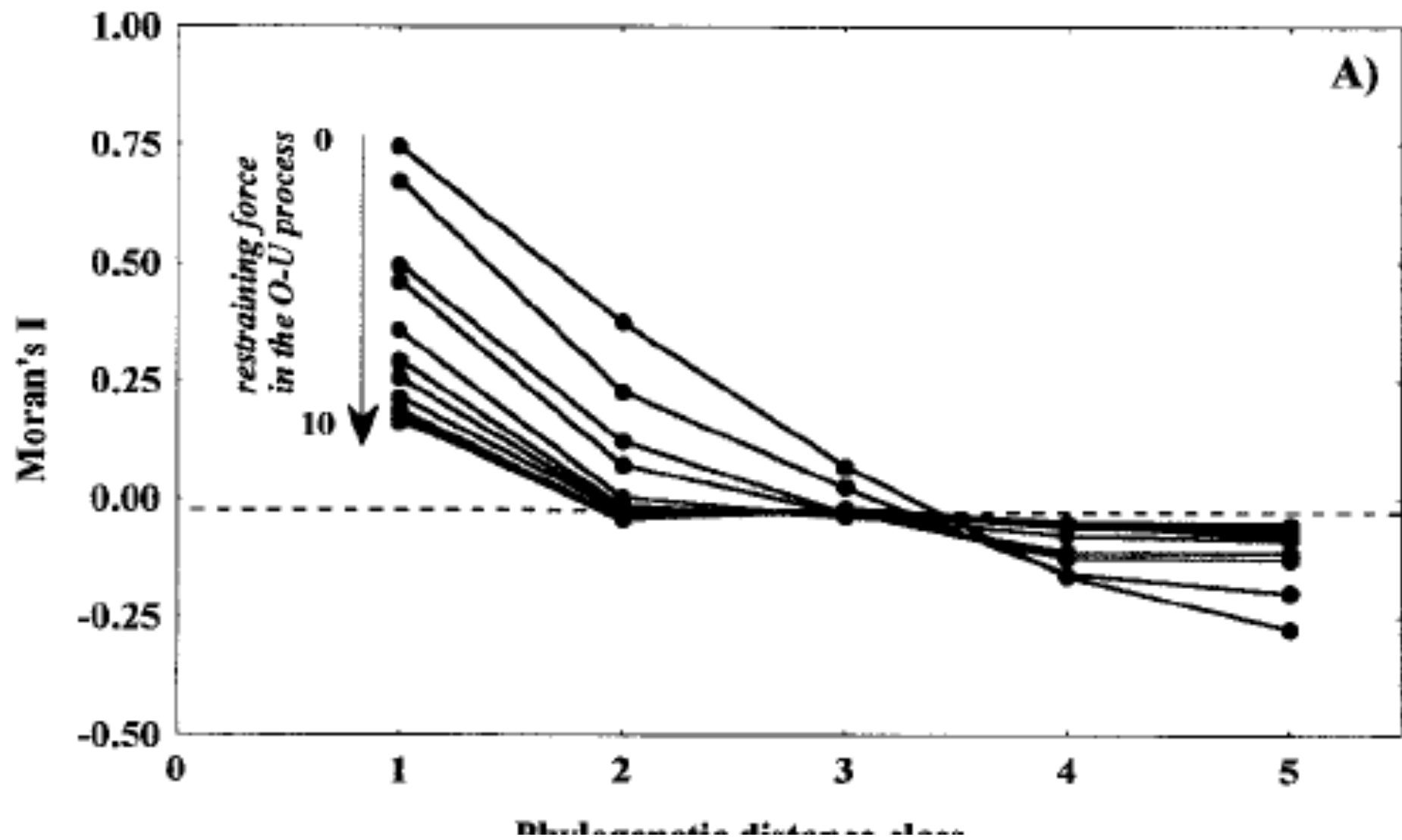


TABLE 4. Phylogenetic correlograms for carnivore body weight using a molecular phylogeny and two different taxonomic classifications.

Divergence date (mybp)	Normalized Moran's <i>I</i>	<i>z</i>
Wayne et al. (1989) molecular phylogeny (39 spp.)		
0-15	0.766	6.316
15-30	0.541	3.600
30-45	-0.633	-4.778
45-60	-0.316	-0.834
Taxonomic rank	Normalized Moran's <i>I</i>	<i>z</i>
Nowak and Paradiso (1983) classification (39 spp.)		
Genus	0.778	6.240
Family	0.712	5.220
Superfamily	-0.454	-1.882
Order	-0.307	-1.884
Ewer (1973) classification (39 spp.)		
Genus	0.642	4.766
Family	0.688	5.124
Superfamily	-0.397	-1.423
Order	-0.307	-1.884

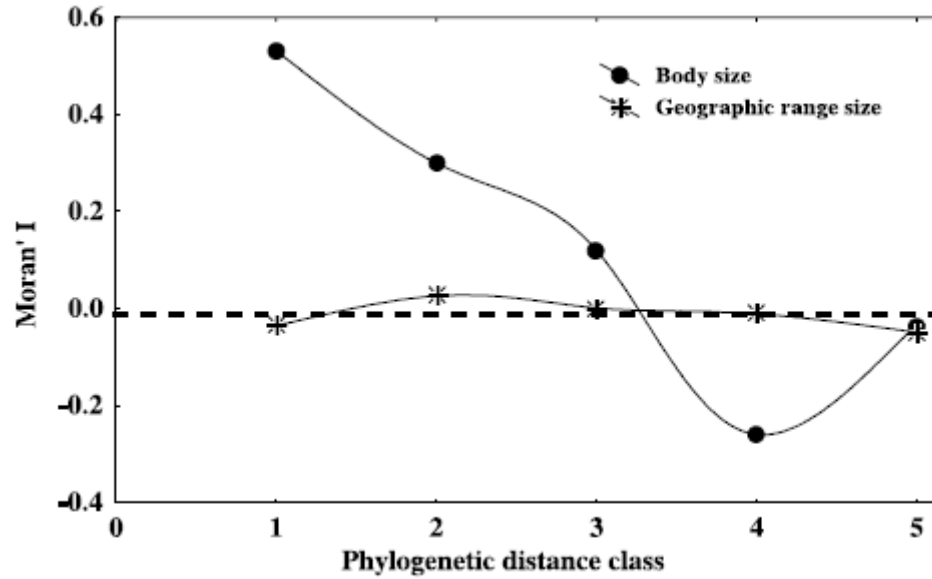
/ de Moran



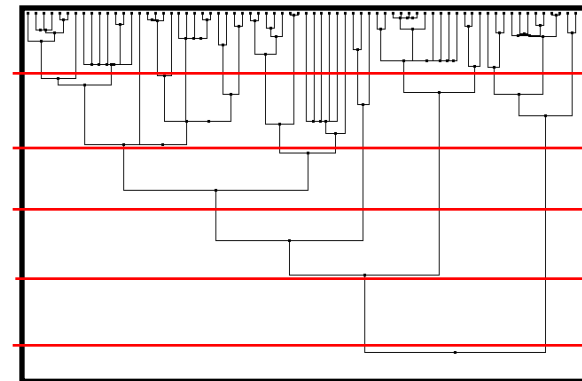
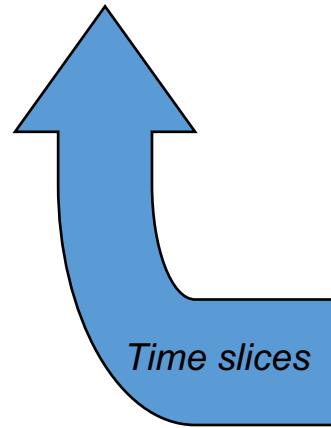
CORRELOGRAMAS

Moran's I

$$I = \frac{\sum_{i,j} w_{ij} z_i z_j}{\sum_{i=1}^n W z_i^2}$$



Allows evaluation of more complex structures in the matrix **W** of phylogenetic relationships...

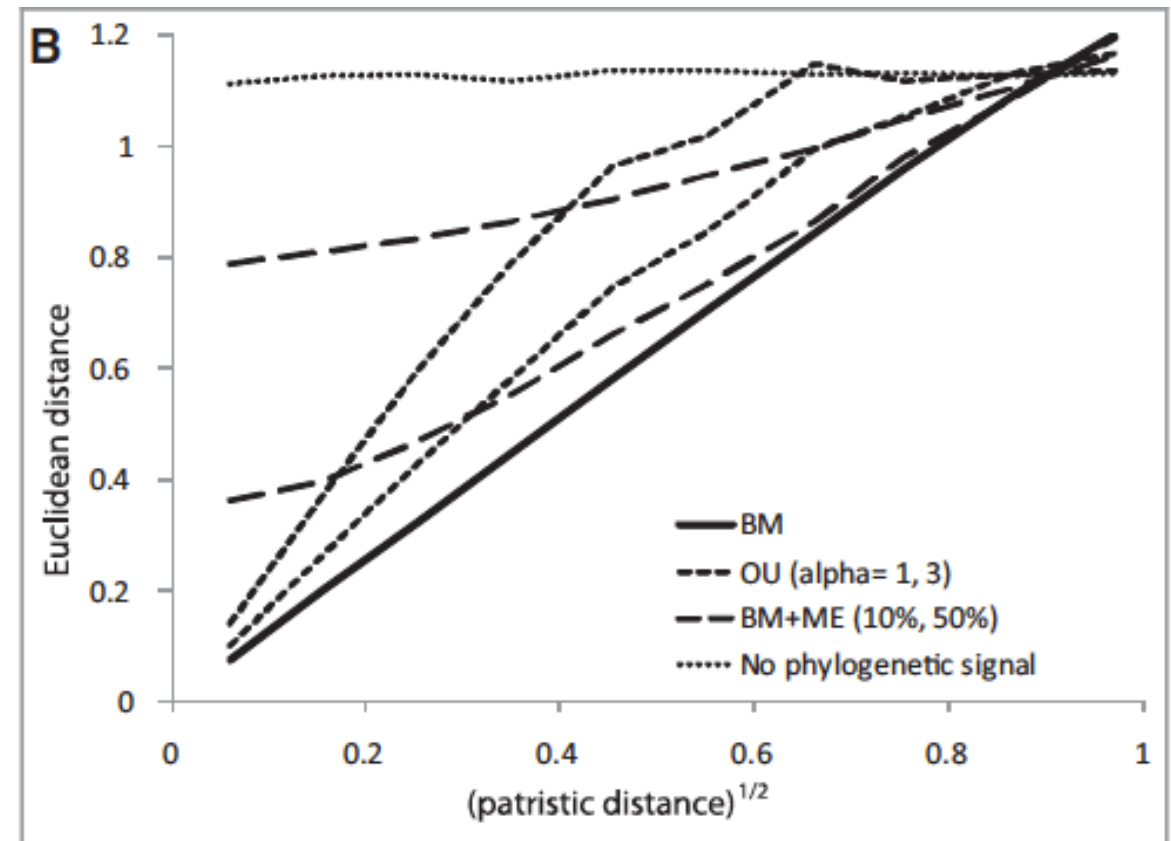
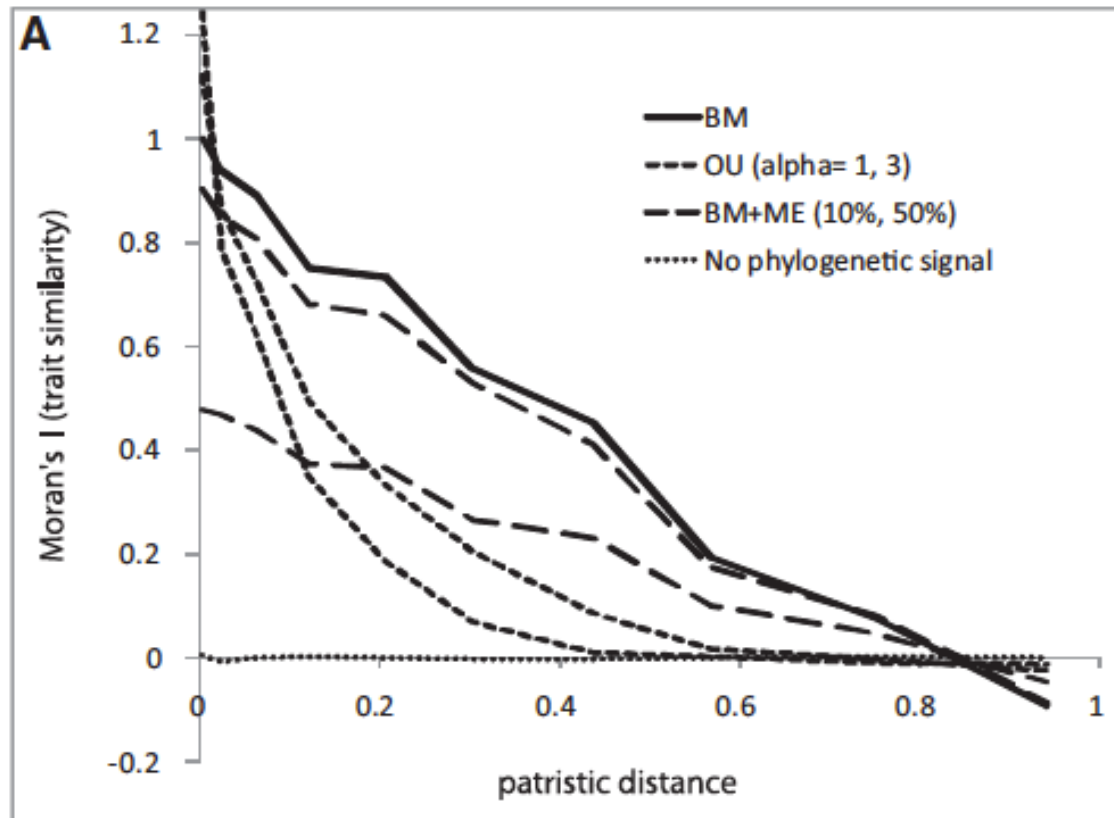


Teste de Mantel tem mais poder para detectar sinal filogenético quando há erro de mensuração (variação intraespecífica) do que K de Blomberg

Distogramas com distância Euclidiana podem ser usados para representar sinal filogenético e desvios do BM

ASSESSING PHYLOGENETIC SIGNAL WITH MEASUREMENT ERROR: A COMPARISON OF MANTEL TESTS, BLOMBERG ET AL.'S K , AND PHYLOGENETIC DISTOGRAMS

Olivier J. Hardy^{1,2} and Sandrine Pavoine^{3,4}



(a)

	Body Mass (kg)	d^2	
	Y_i	$(Y_{i+1} - Y_i)^2$	
A	C	53.6	5.76
	B	51.2	4.00
	D	49.2	2.89
	E	47.5	1391.3
	F	10.2	0.81
	G	11.1	4.41
		13.2	0.36
	12.6	...	
	$\sum y^2$	$\sum d^2$	
	(= 3006.3)	(=1409.5)	

$$\eta = \sum d^2 / \sum y^2 = 0.469$$

$$\text{Calculated } C = 1 - \eta/2 = 0.766$$

$$\text{Critical } C = 0.587$$

$$P < 0.01$$

(b) **Contrasts**
Body Mass

Proposed Ordering of
Nodes and their Contrasts

A
B
C
D
E
F
G

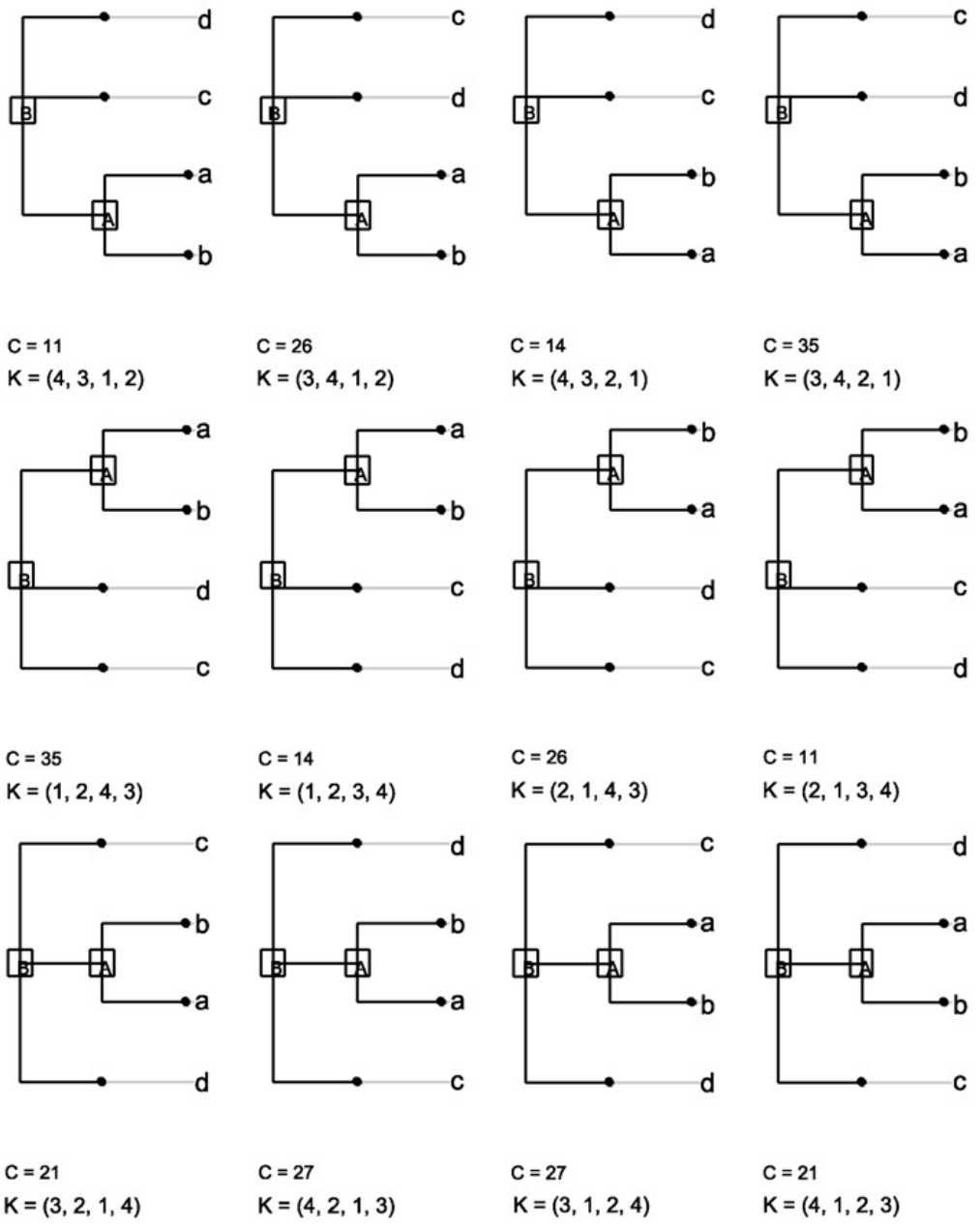
A method for testing the assumption of phylogenetic independence in comparative data

Ehab Abouheif*

*Department of Ecology and Evolution, State University of New York at Stony Brook,
Stony Brook, NY 11794-5245, USA*

Baseado só na topologia,
sem considerar os
comprimentos de ramo

B



the set of consistent representations of the tree topology

D



Testing for phylogenetic signal in phenotypic traits: New matrices of phylogenetic proximities

Sandrine Pavoine^{a,*}, Sébastien Ollier^b, Dominique Pontier^a, Daniel Chessel^a

TFSP's statistic

$$C_{mean} = 1 - \frac{mean(C)}{2 \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$C_{mean} = 0.2024$$

Moran's statistic

$$I = \frac{(\mathbf{x} - \bar{x})^t \mathbf{A} (\mathbf{x} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

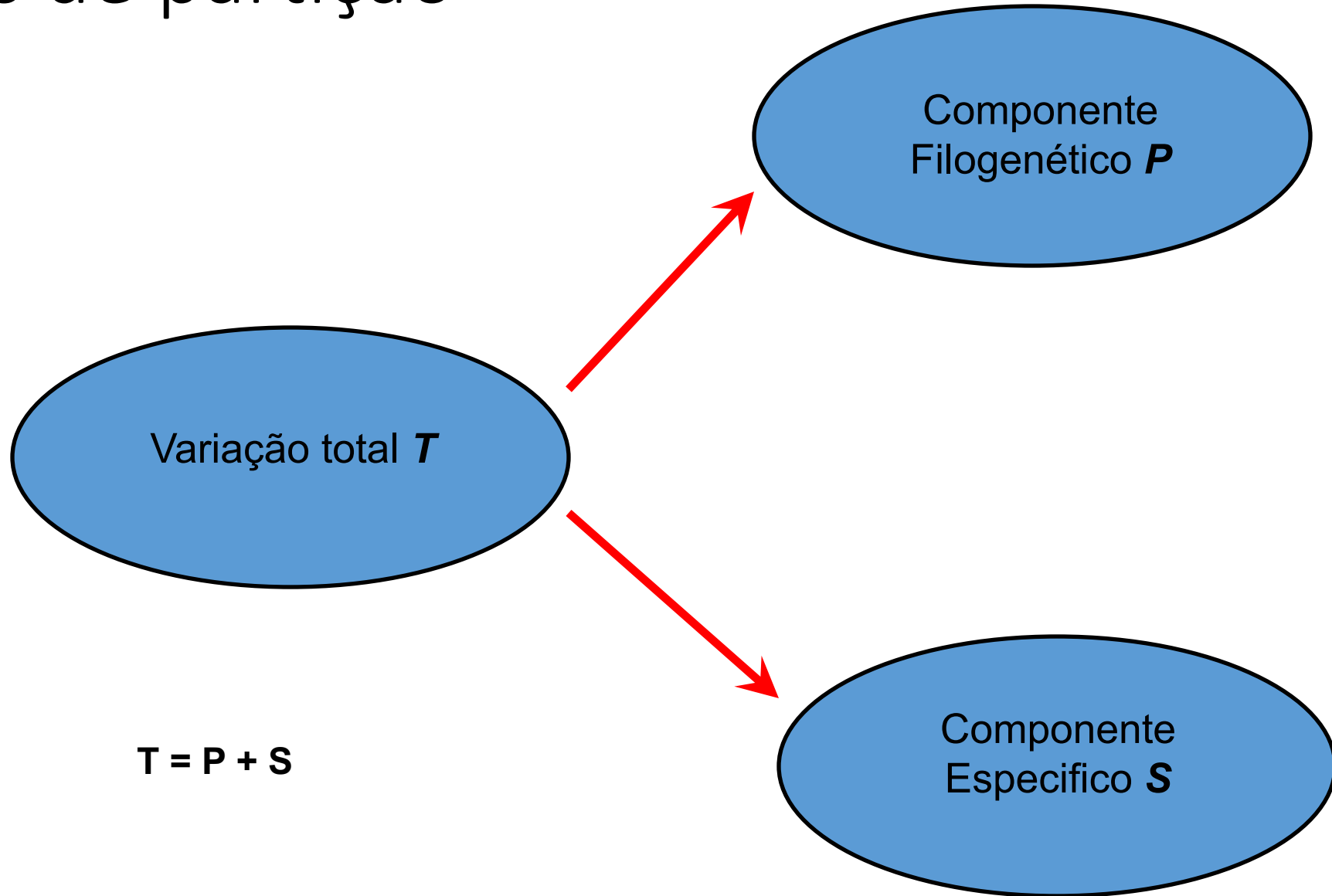
$$I = \frac{\mathbf{z}^t \mathbf{A} \mathbf{z}}{\sum_{i=1}^n \sum_{j=1}^n a_{ij}}$$

$$I = 0.2024$$

phylogenetic signal statistics

I de Moran e C de Abouheif são equivalentes

Métodos de partição



EVOLUTION

INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

PUBLISHED BY

THE SOCIETY FOR THE STUDY OF EVOLUTION

Vol. 52

October 1998

No. 5

Evolution, 52(5), 1998, pp. 1247-1262

AN EIGENVECTOR METHOD FOR ESTIMATING PHYLOGENETIC INERTIA

JOSÉ ALEXANDRE FELIZOLA DINIZ-FILHO,^{1,2} CARLOS EDUARDO RAMOS DE SANT'ANA,^{3,4}
AND LUIS MAURICIO BINI^{1,5}

¹*Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Goiás. Cx.P. 131,
74.001-970, Goiânia, GO, Brasil*

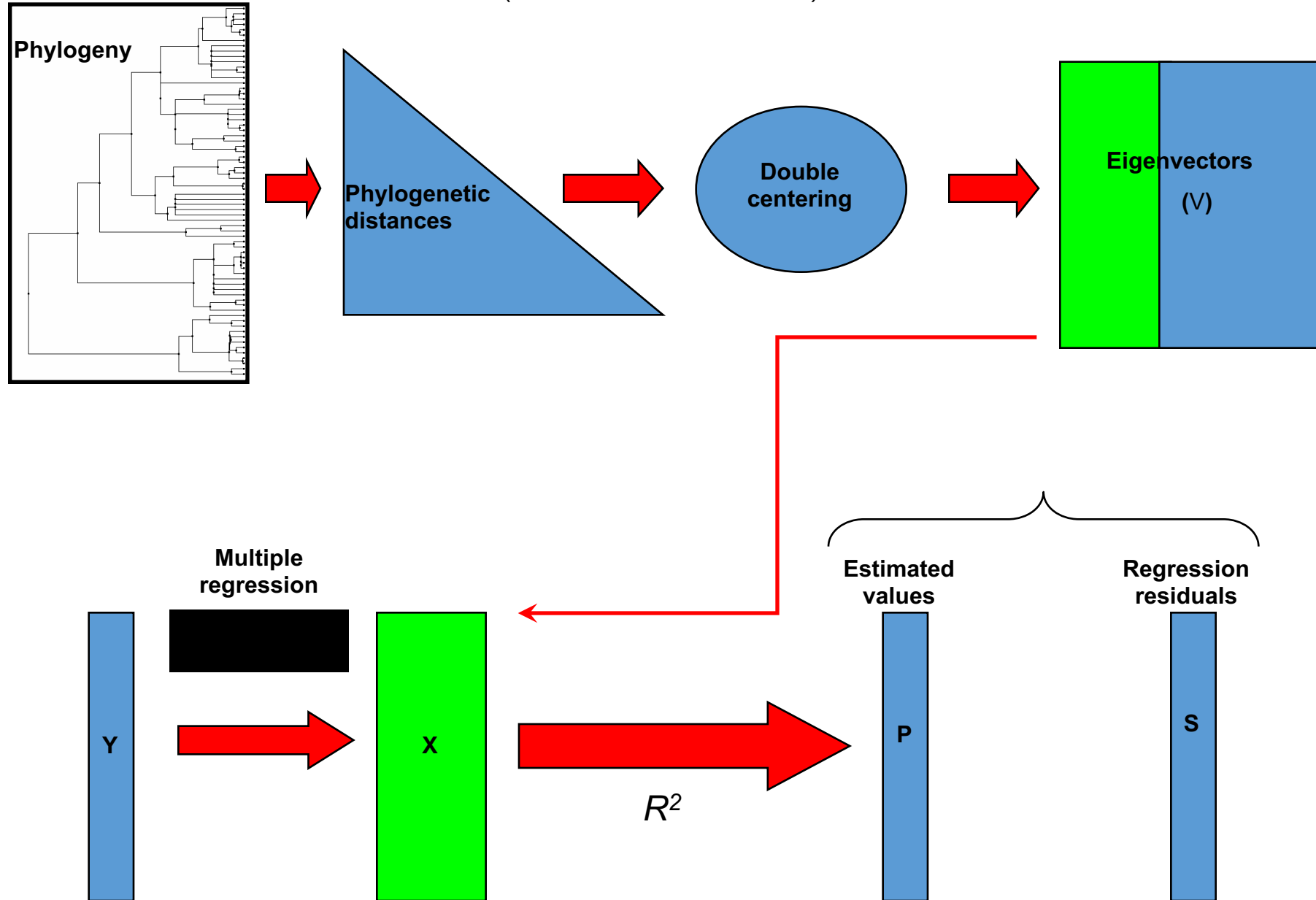
²*E-mail: diniz@icbl.ufg.br*

³*Escola Técnica Federal de Goiás, Coordenação de Química e Biologia. Rua 75, n. 46, 74.055-110, Goiânia, GO, Brasil*

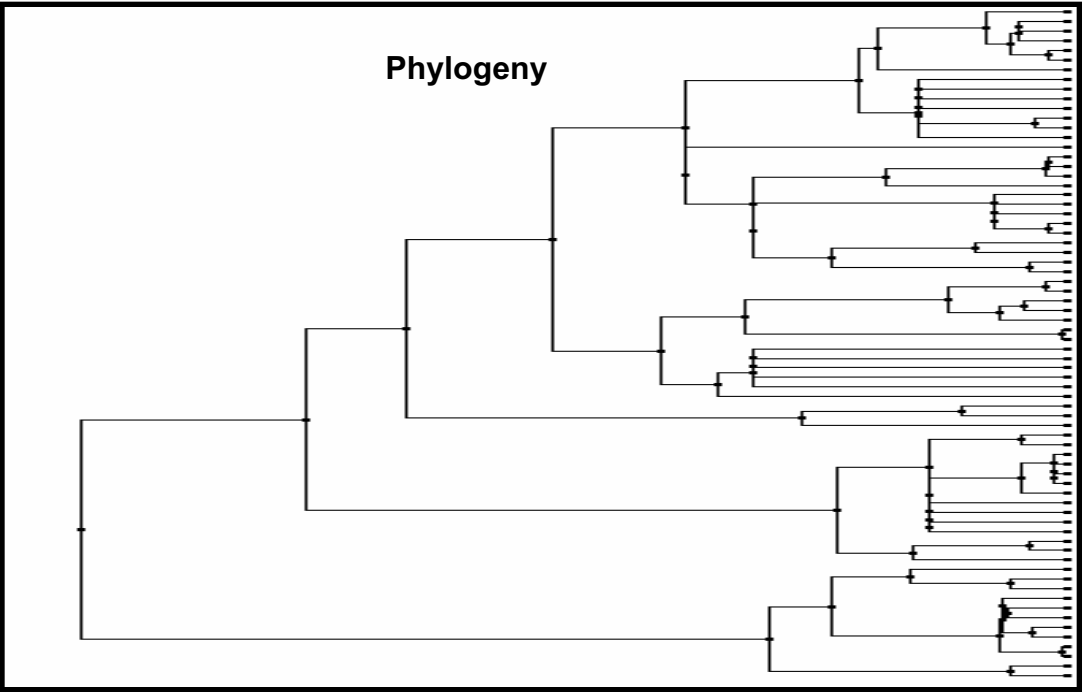
⁴*E-mail: kadu@etfgo.br*

⁵*E-mail: bini@icbl.ufg.br*

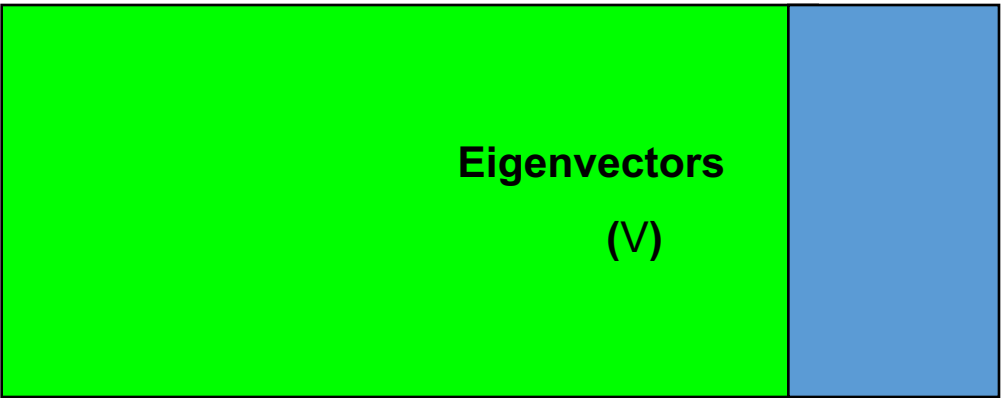
Diniz-Filho's et al. (1998) Phylogenetic eigenVector Regression (PVR)
(*Evolution* 52: 1247-1262.)



Diniz-Filho`s et al. (1998) phylogenetic eigenvector regression (PVR)



Eigenvalues



Phylogenetic eigenvectors represent linearly different cuts of phylogeny, allowing evaluation of phylogenetic effects at different `scales`

Crítica do Rohlf

Vol. 55

November 2001

No. 11

Evolution, 55(11), 2001, pp. 2143–2160COMPARATIVE METHODS FOR THE ANALYSIS OF CONTINUOUS VARIABLES:
GEOMETRIC INTERPRETATIONS

F. JAMES ROHLF

which to investigate correlations with other variables. However, Diniz-Filho et al. (1998) only retained the first few eigenvectors—those for which the corresponding λ_i exceed the values expected from the broken stick model. The fit is then no longer perfect and the R^2 values simply measure the proportion of the variance attributable to average differences among the nodes corresponding to the retained eigenvectors (Diniz-Filho et al. 1998, table 3). Leaving out any eigenvector with $\lambda_i > 0$ corresponds to ignoring some part of the structure of the tree near the tips and thus not taking the entire phylogeny into account. By ignoring eigenvectors with small eigenvalues, this approach confounds variation due to divergence among species that diverge near the tips of the tree with variation within the branches connecting each species to its more recent common ancestor. Thus, this only estimates

EXPLORING PATTERNS OF INTERSPECIFIC VARIATION IN QUANTITATIVE TRAITS USING SEQUENTIAL PHYLOGENETIC EIGENVECTOR REGRESSIONS

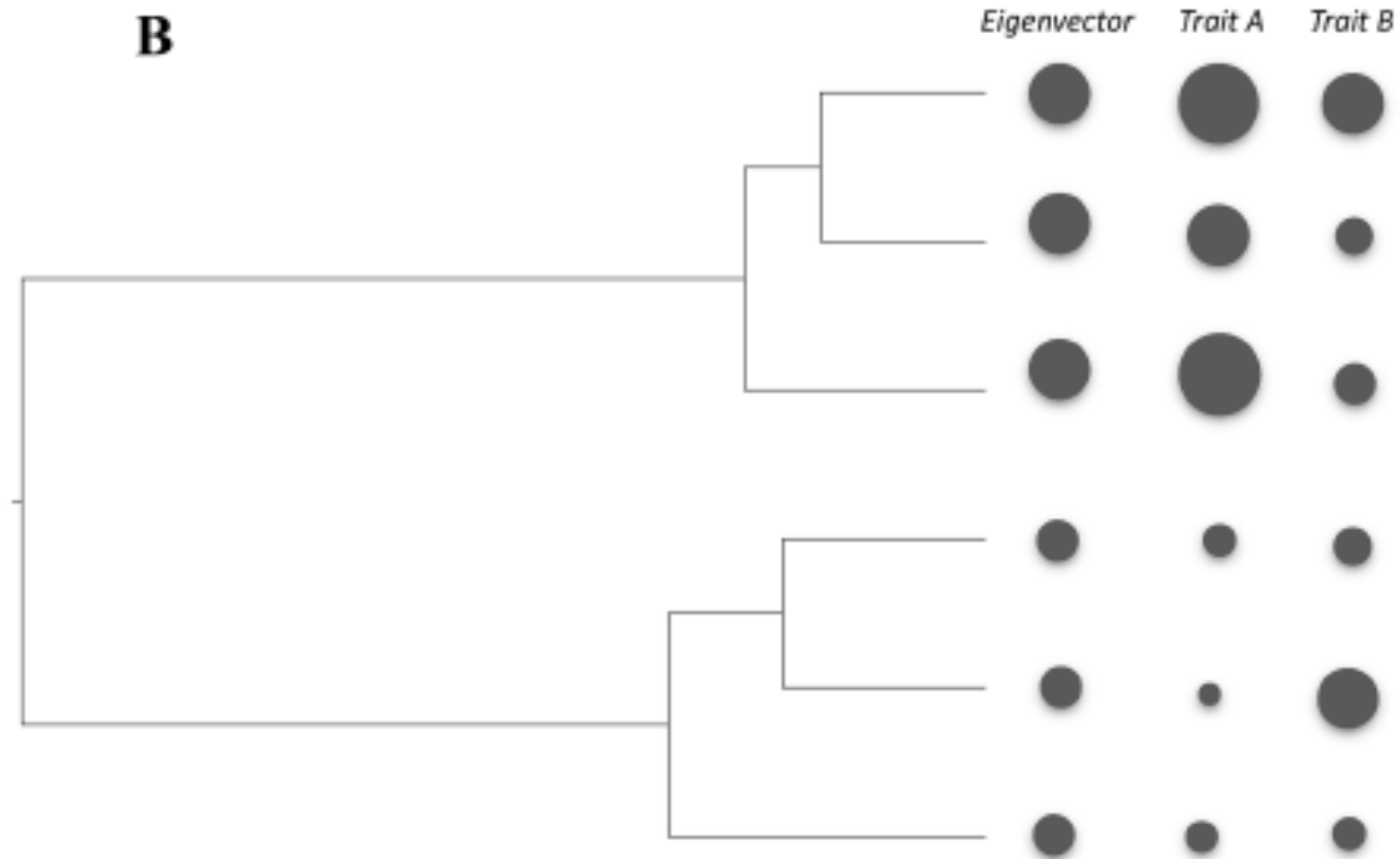
José Alexandre Felizola Diniz Filho,^{1,2} Thiago F. Rangel,¹ Thiago Santos,³ and Luis Mauricio Bini¹

¹*Departamento de Ecologia, ICB, Universidade Federal de Goiás. CxP 131, 74001-970 Goiânia, GO, Brasil*

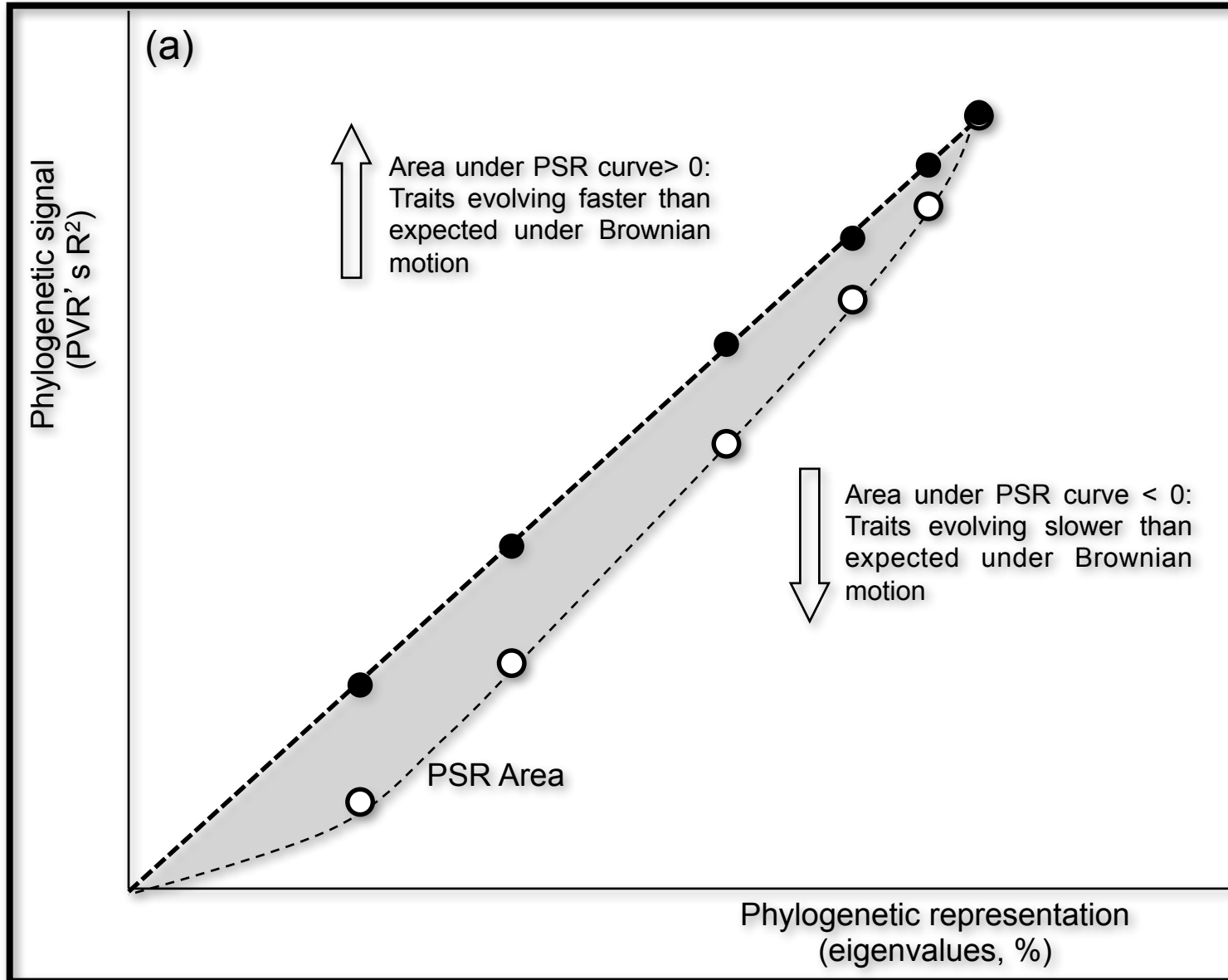
²*E-mail: diniz@icb.ufg.br*

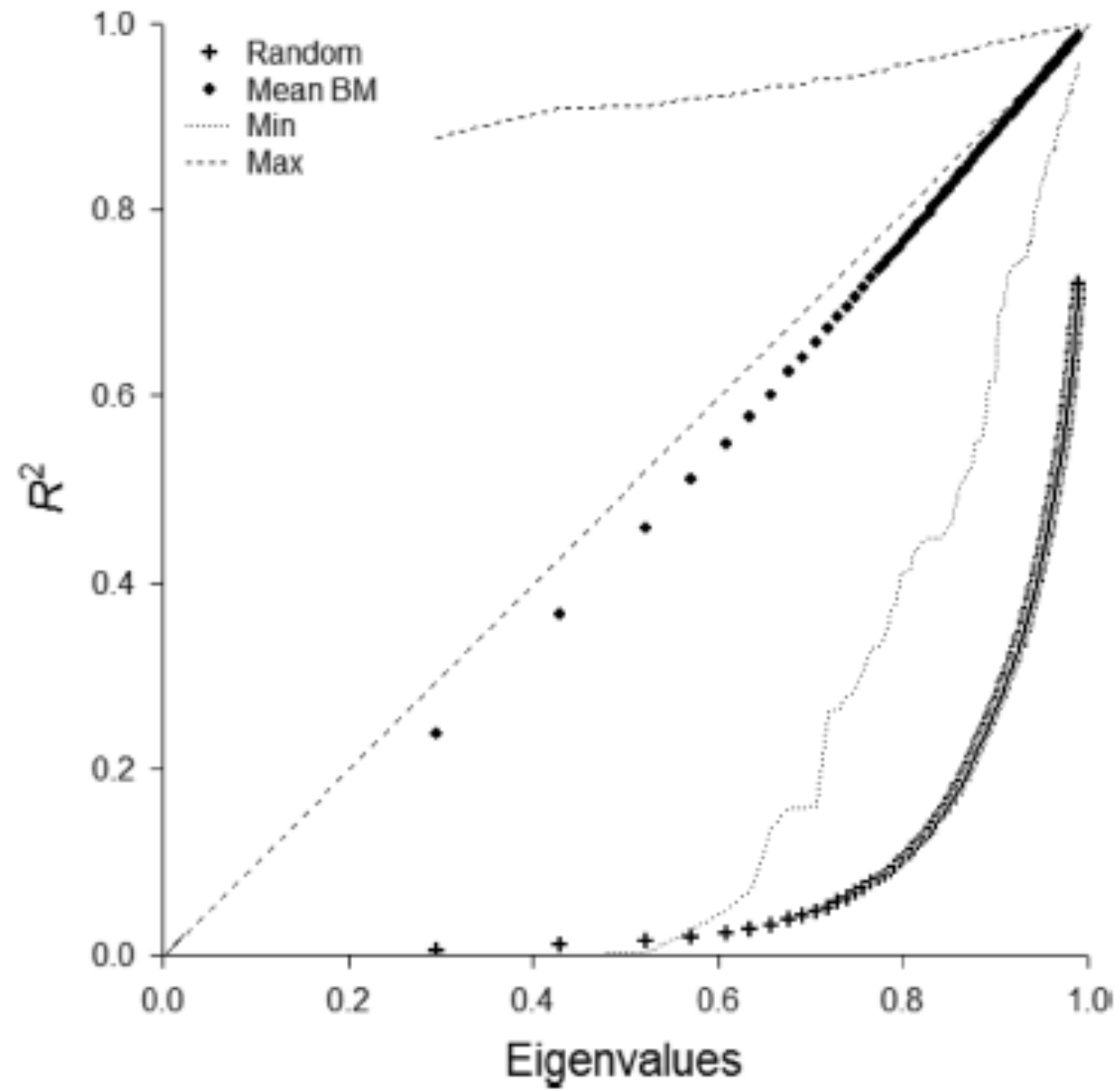
³*Programa de Pós-Graduação em Ecologia and Evolução, Universidade Federal de Goiás, Goiânia, GO, Brasil*

PVR e Modelos de Evolução:
Curva PSR (Phylogenetic Signal-Representation curve)



Curva PSR







On the selection of phylogenetic eigenvectors for ecological analyses

Jose Alexandre F. Diniz-Filho, Luis Mauricio Bini, Thiago Fernando Rangel, Ignacio Morales-Castilla, Miguel Á. Olalla-Tárraga, Miguel Á. Rodríguez and Bradford A. Hawkins

The criterion that minimizes Moran's I in the residuals yields a model with only 14 eigenvectors, and with small autocorrelation. The R^2 is not as high as in the other non-sequential models (0.70), but still much higher than expected for a sequential model with similar k (i.e. under sequential approach approximately 20 eigenvectors would be needed to achieve a similar R^2 ; Table 1). Mean S -components were also stable and similar to geographic patterns from other PVR

Teste de Mantel



Nathan Mantel (1919 - 2002)

Cancer Research

VOLUME 27

FEBRUARY 1967

NUMBER 2

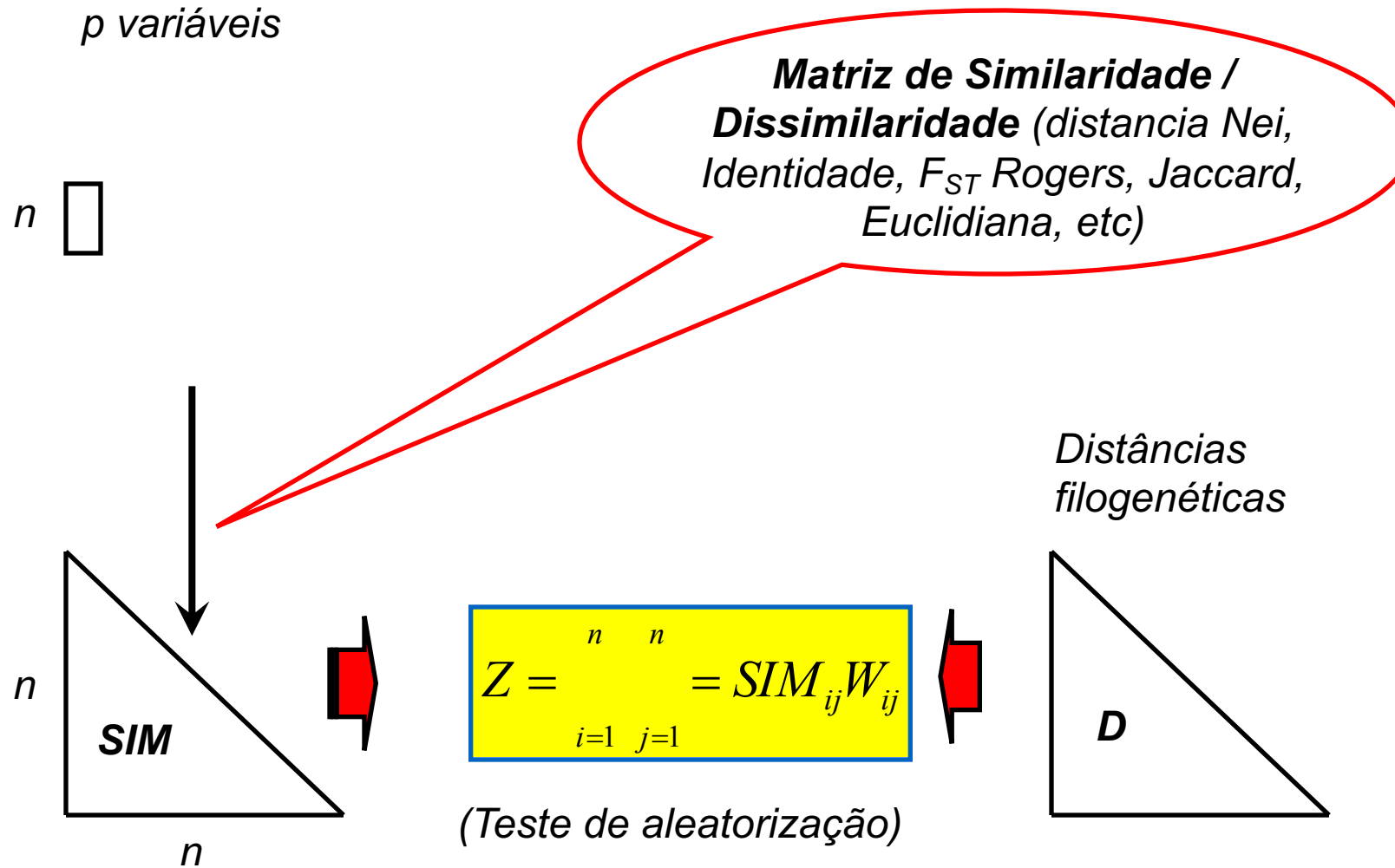
[CANCER RESEARCH 27 Part 1, 209-220, February 1967]

The Detection of Disease Clustering and a Generalized Regression Approach

NATHAN MANTEL

Pode ser usado pra vários traits ao mesmo tempo

AUTOCORRELAÇÃO MULTIDIMENSIONAL: O Teste de MANTEL



POOR STATISTICAL PERFORMANCE OF THE MANTEL TEST IN PHYLOGENETIC COMPARATIVE ANALYSES

Luke J. Harmon^{1,2} and Richard E. Glor³

¹*Department of Biological Sciences, University of Idaho, Moscow, Idaho, 83844*

²*E-mail: lukeh@uidaho.edu*

³*Department of Biology, RC Box 270211, University of Rochester, Rochester, New York, 14627*

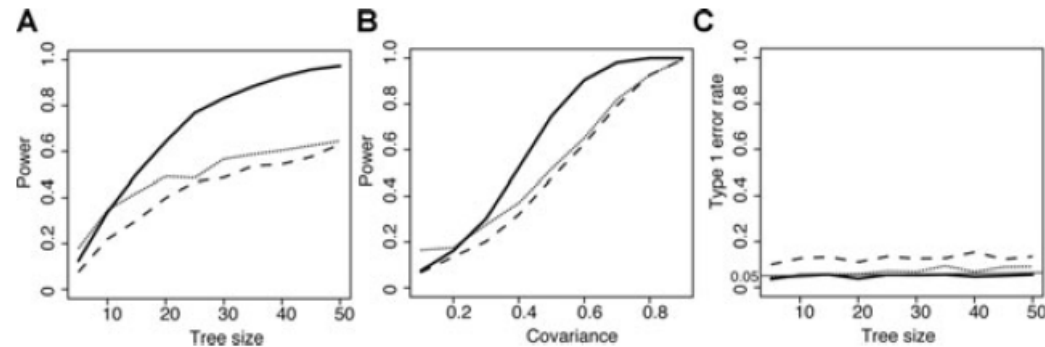


Figure 2. Statistical power (A & B) and type-I error (C) for phylogenetic tests of character correlation using the standard partial Mantel test (dashed line), PP Mantel test (dotted line), and independent contrasts (solid line). In each case, 1000 simulated datasets were analyzed for a range of tree sizes with an expected covariance of 0.5 (in A and C) or for trees with 25 taxa and a range of covariance values (B).

Decomposition of trait diversity among the nodes of a phylogenetic tree

SANDRINE PAVOINE,^{1,2,5} MICHEL BAGUETTE,³ AND MICHAEL B. BONSALE^{1,4}

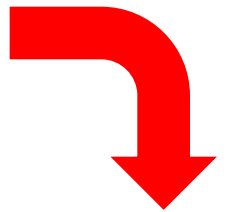
Calcula a Entropia Quadrática de Rao para o conjunto de espécies



Decompõe a Entropia ao longo dos nós da filogenia (só considera topologia)



Testa se há viés da diversidade na raiz/tips



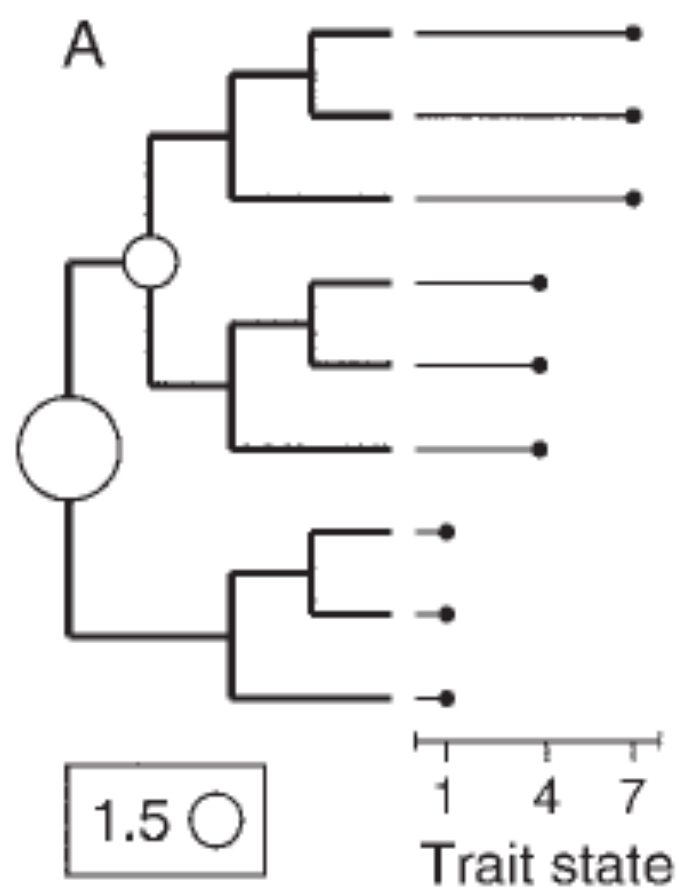
Aleatorização para determinar significância

$$H_{\Delta}(P) = \sum_{k=1}^S \sum_{l=1}^S p_k p_l \frac{(\delta_{kl})^2}{2}$$

$$w_i = \lambda_{n_i} \left[\sum_{g=1}^{\text{card}(\text{dd}_{n_i})} \sum_{g'=1}^{\text{card}(\text{dd}_{n_i})} \frac{\lambda_{\text{dd}_{n_i}[g]} \lambda_{\text{dd}_{n_i}[g']}}{\lambda_{n_i} \lambda_{n_i}} \frac{D_{\Delta}^2(P_{\text{dd}_{n_i}[g]}, P_{\text{dd}_{n_i}[g']})}{2} \right]$$

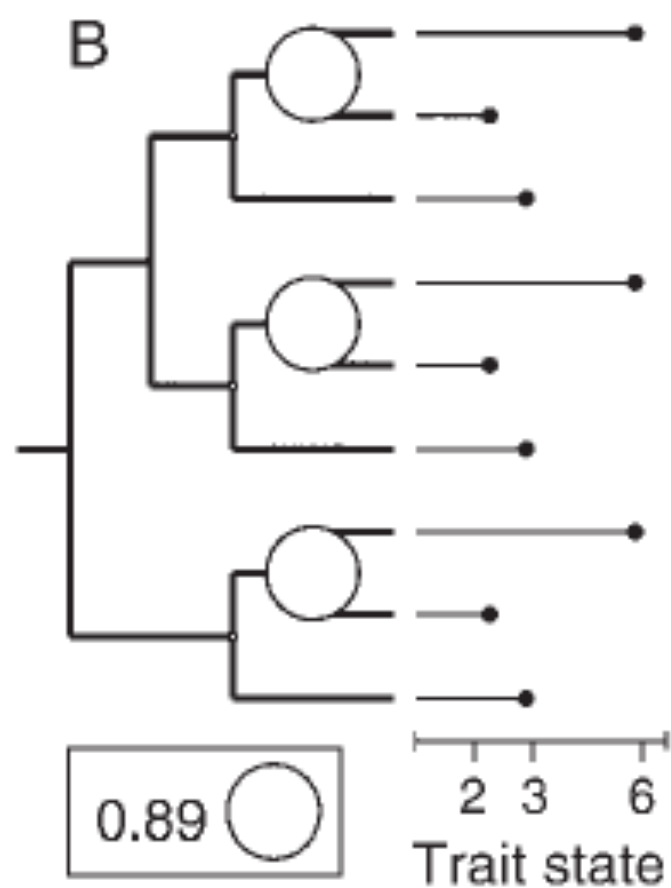
$$S_3 = \frac{1}{N} \sum_{i=1}^N i v_i$$

between zero and unity. This third test provides a description of phylogenetic signal where species have more different trait values if they are distantly related on the phylogeny (trait diversity is skewed to root) and a description of the absence of phylogenetic signal where closely related species have highly different trait values (e.g., due to convergent evolution and trait diversity is then skewed to tips; Fig. 2A, B). This tips/root skewness test is related to the phylogenetic signal randomization test of Blomberg et al. (2003) based on branch-length transformations and to the divergence order test (DOT) developed by Ackerly et al. (2006). In the statistic of Ackerly et al. i is replaced with the age of the node and v_i with unstandardized contrasts for a single trait across the nodes (according to Felsenstein 1985). In contrast to our statistic S_3 , statistics of Ackerly et al. and Blomberg et al. can only be applied to a single quantitative trait at a time, although the values taken by the statistic of



QE = 6

SN test $P = 0.027$
 FN test $P = 0.004$
 Ro test $P < 0.001$
 (skewed to root)

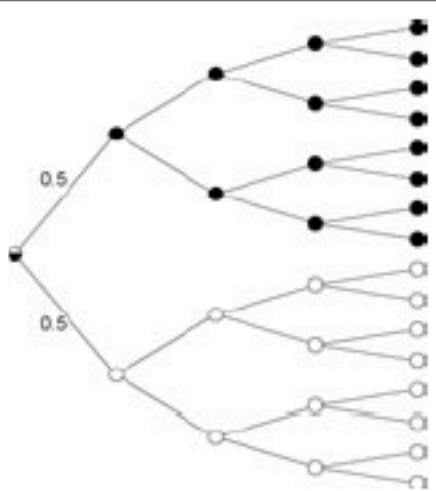
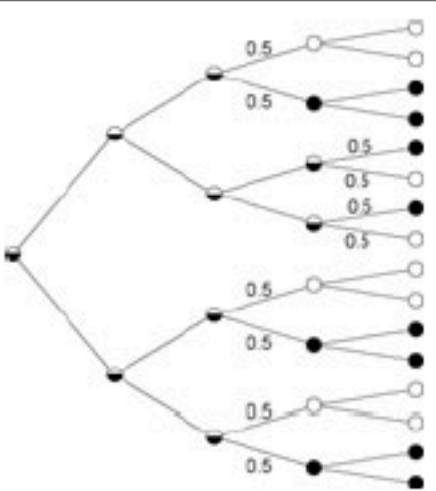
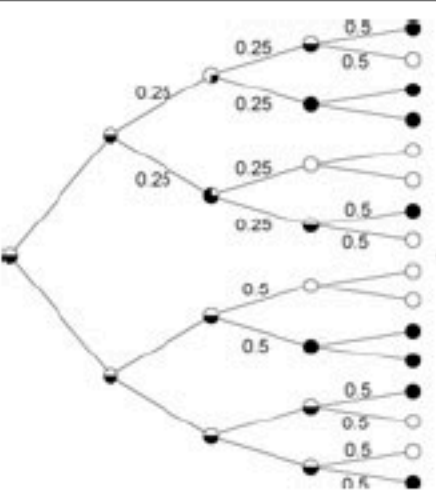
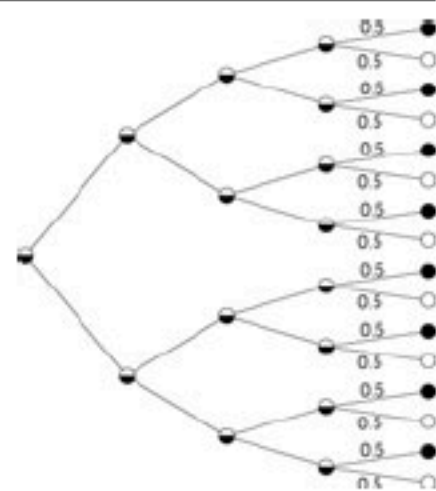


QE = 1.86

NS
 NS
 $P < 0.001$
 (skewed to tips)

Selectivity in Mammalian Extinction Risk and Threat Types: a New Measure of Phylogenetic Signal Strength in Binary Traits

Table 1. The calculation of D for different types of phylogenetic pattern in binary traits.

	Phylogenetic pattern ^a			
	<i>extremely clumped</i>	<i>clumped (Brownian)</i>	<i>random</i>	<i>overdispersed</i>
Phylogeny ^b				
$\sum d_{\text{obs}}^c$	1	5	6.5	8
D^d	-2.4	0.0	1.0	1.9

We computed our measure of phylogenetic signal (or character dispersion on a phylogeny), D , by scaling the observed sum of sister-clade differences with the mean values of the two expected distributions as follows:

$$D = [\Sigma d_{\text{obs}} - \text{mean}(\Sigma d_{\text{b}})] / [\text{mean}(\Sigma d_{\text{r}}) - \text{mean}(\Sigma d_{\text{b}})].$$

The D statistic is equal to 1 if the observed binary trait has a phylogenetically random distribution across the tips of the phylogeny and to 0 if the observed trait is as clumped as if it had evolved by Brownian motion under our threshold model (Table 1). Values of D can fall outside this that are not at either end of the possible range of values. Increasing phylogenetic clumping in the binary trait is indicated by values of D decreasing from 1. Our method also tests D for significant departure from 0 (Brownian expectation) and 1 (random expectation) by seeing where Σd_{obs} lies within the two expected distributions (Σd_{b} and Σd_{r}).