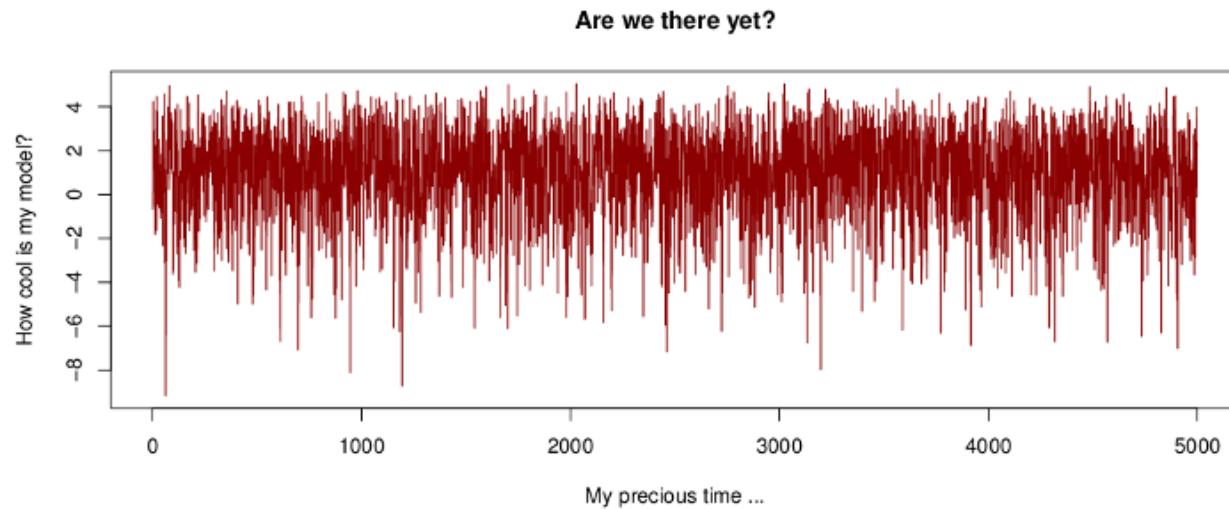


# WORKSHOP - MCMC passo a passo

Uma visão prática de métodos Monte Carlo para inferência Bayesiana  
e aplicações com modelos comparativos filogenéticos



Daniel S. Caetano  
Diogo Melo

Fevereiro 2016

Universidade de São Paulo  
Instituto de Biociências

University of Idaho

# Por quê?

Análises Bayesianas estão se tornando cada vez mais comuns.

Em análises filogenéticas e em métodos comparativos, Bayes está para todos os lados!



**Mr Bayes**  
Bayesian Inference  
of Phylogeny



**RevBayes**



# Por quê?

Análises Bayesianas estão se tornando cada vez mais comuns.

Em análises filogenéticas e em métodos comparativos Bayes está para todos os lados!



Mr Bayes  
Bayesian Inference  
of Phylogeny



RevBayes

# Objetivos

Objetivo principal é criar uma intuição/noção de como a análise Bayesiana funciona.

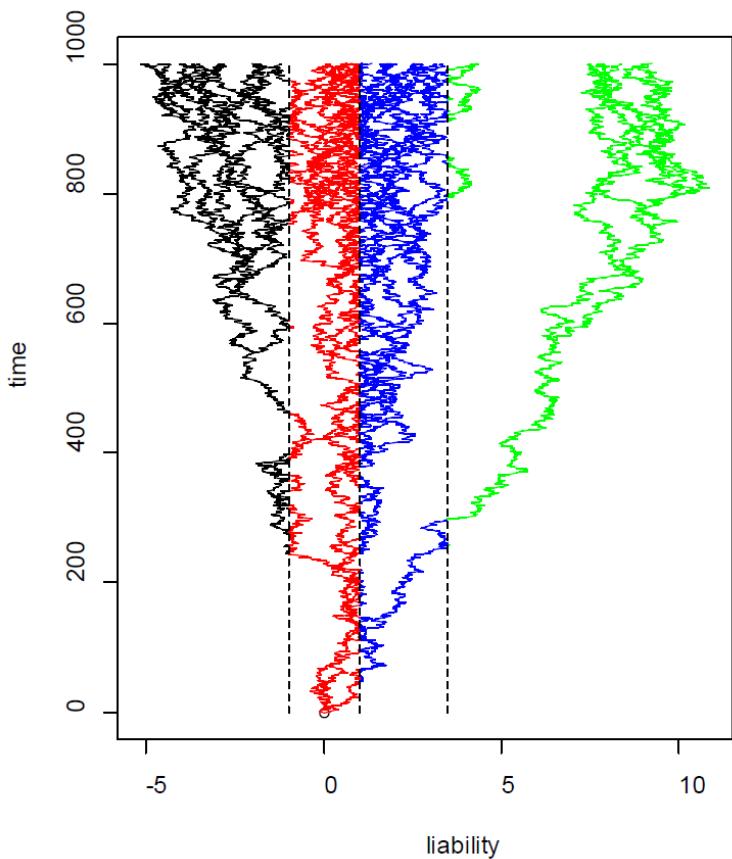
Para tal, nós vamos...

- I) Trabalhar de uma forma simples.
- II) Usar experimentos ao invés de provas matemáticas.
- III) Explorar e replicar os passos do Markov chain Monte Carlo.
- IV) Aplicar em modelos simples e expandir para modelos usados na vida real.

# Objetivos

Modelos de estudo serão métodos comparativos filogenéticos  
**PCMs**

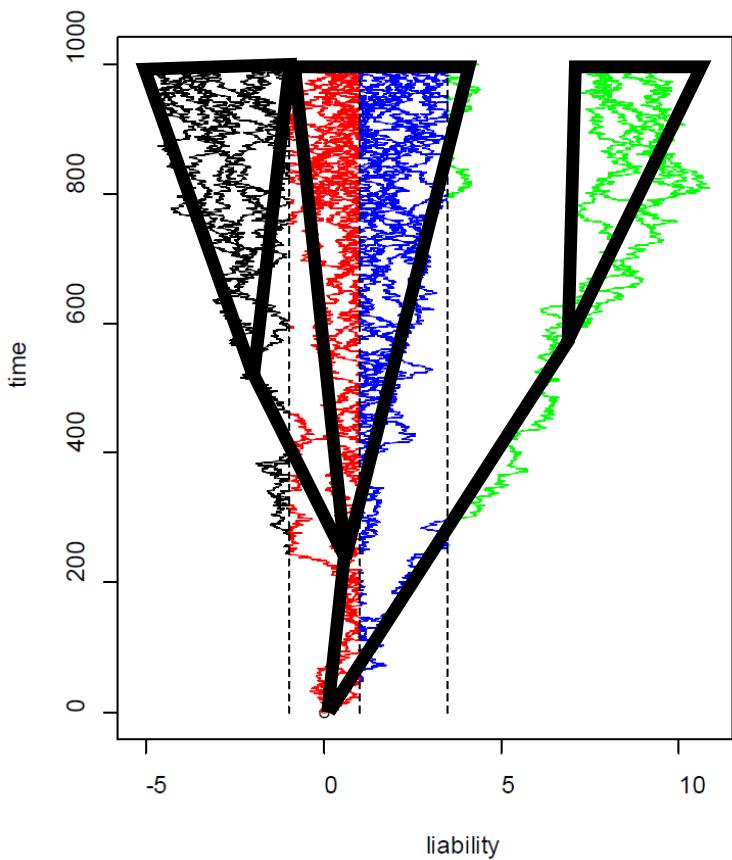
- Brownian motion
- ACDC (accelerate / decelerate model)
- OU (Ornstein Uhlenbeck)
- entre outros...



# Objetivos

Modelos de estudo serão métodos comparativos filogenéticos  
**PCMs**

- Brownian motion
- ACDC (accelerate / decelerate model)
- OU (Ornstein Uhlenbeck)
- entre outros...



# Apresentação

Antes de começar ...

Quem sou eu?

# Apresentação

- Graduação em Ciências Biológicas na USP.
- Mestrado em Ecologia (Dr. Glauco Machado)



The ecological tale of Gonyleptidae (Arachnida, Opiliones) evolution: phylogeny of a Neotropical lineage of armoured harvestmen using ecological, behavioural and chemical characters

Daniel S. Caetano<sup>a,b</sup> and Glauco Machado<sup>a,\*</sup>

<sup>a</sup>Programa de Pós-graduação em Ecologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, 05508-900, Brazil; <sup>b</sup>Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, 05508-900, Brazil

Accepted 30 September 2012

Comportamento animal

+

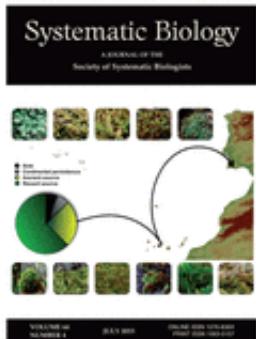
Ecologia



Árvore filogenética

# Apresentação

-- Cursando Doutorado na University of Idaho (Dr. Luke J. Harmon)



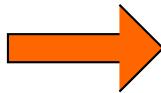
## Comparative Analysis of Principal Components Can be Misleading

JOSEF C. UYEDA\*, DANIEL S. CAETANO, AND MATTHEW W. PENNELL

*Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844, USA.  
Correspondence to be sent to: Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, University of Idaho,  
Moscow, ID 83844, USA; E-mail: josefuyeda@gmail.com*

*Received 21 July 2014; reviews returned 28 October 2014; accepted 30 March 2015  
Associate Editor: Tanja Stadler*

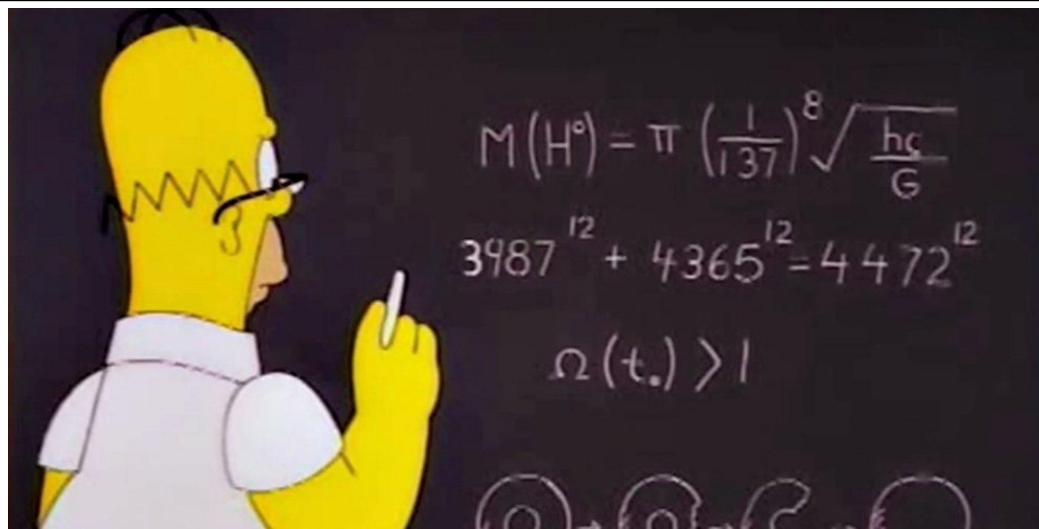
Criação/Avaliação de  
Métodos Comparativos  
Filogenéticos



Dados secundários  
(Bases de dados)

# Apresentação

Parte da minha tese envolve novos métodos usando MCMC ...



$$M(H^\circ) = \pi \left(\frac{1}{137}\right)^8 \sqrt{\frac{hc}{G}}$$

$$3987^{12} + 4365^{12} = 4472^{12}$$

$$\Omega(t_*) > 1$$

Aqui estou!

# O que é um modelo?

"Um *modelo científico* é uma **idealização simplificada** de um sistema que possui maior complexidade, mas que ainda assim supostamente reproduz na sua essência o comportamento do sistema complexo que é o alvo de estudo e entendimento."

-- Wikipedia/pt

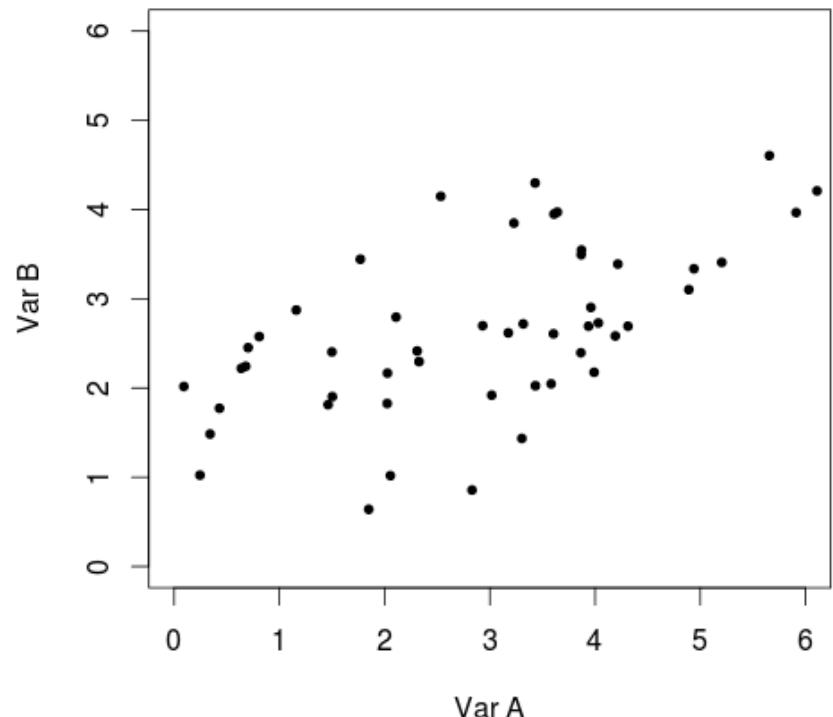
"A *scientific model* seeks to represent empirical objects, phenomena, and physical processes in a logical and objective way. All models are *simulacra*, that is, **simplified reflections of reality** that, despite being approximations, can be extremely useful."

-- Wikipedia/en

Box, George E.P. & Draper, N.R. (1987)

# O que é um modelo?

## Exemplos



Objetivo:

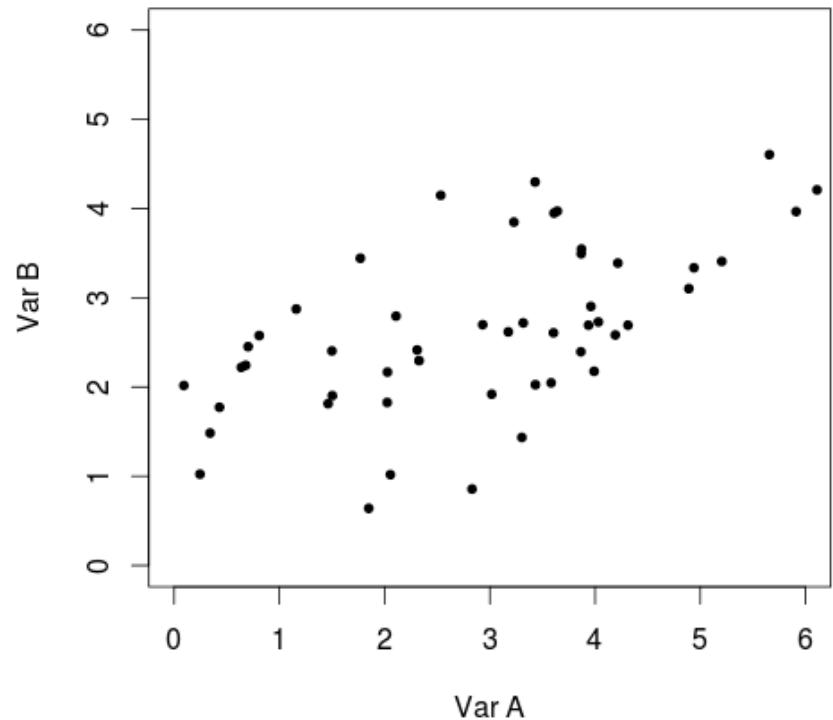
Entender o *mecanismo* que explica a relação entre A e B.

Pergunta:

Qual modelo melhor explica o *padrão* entre A e B?

# O que é um modelo?

## Exemplos

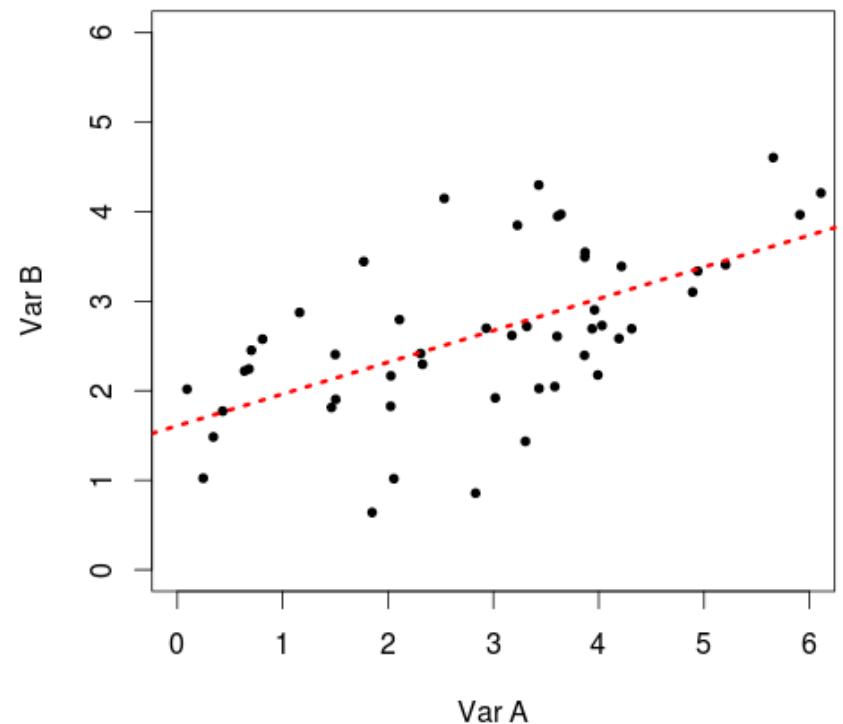


Método:

- a)** Escolher modelo(s) dada as hipóteses a serem testadas.
- b)** Estimar os parâmetros do(s) modelo(s).

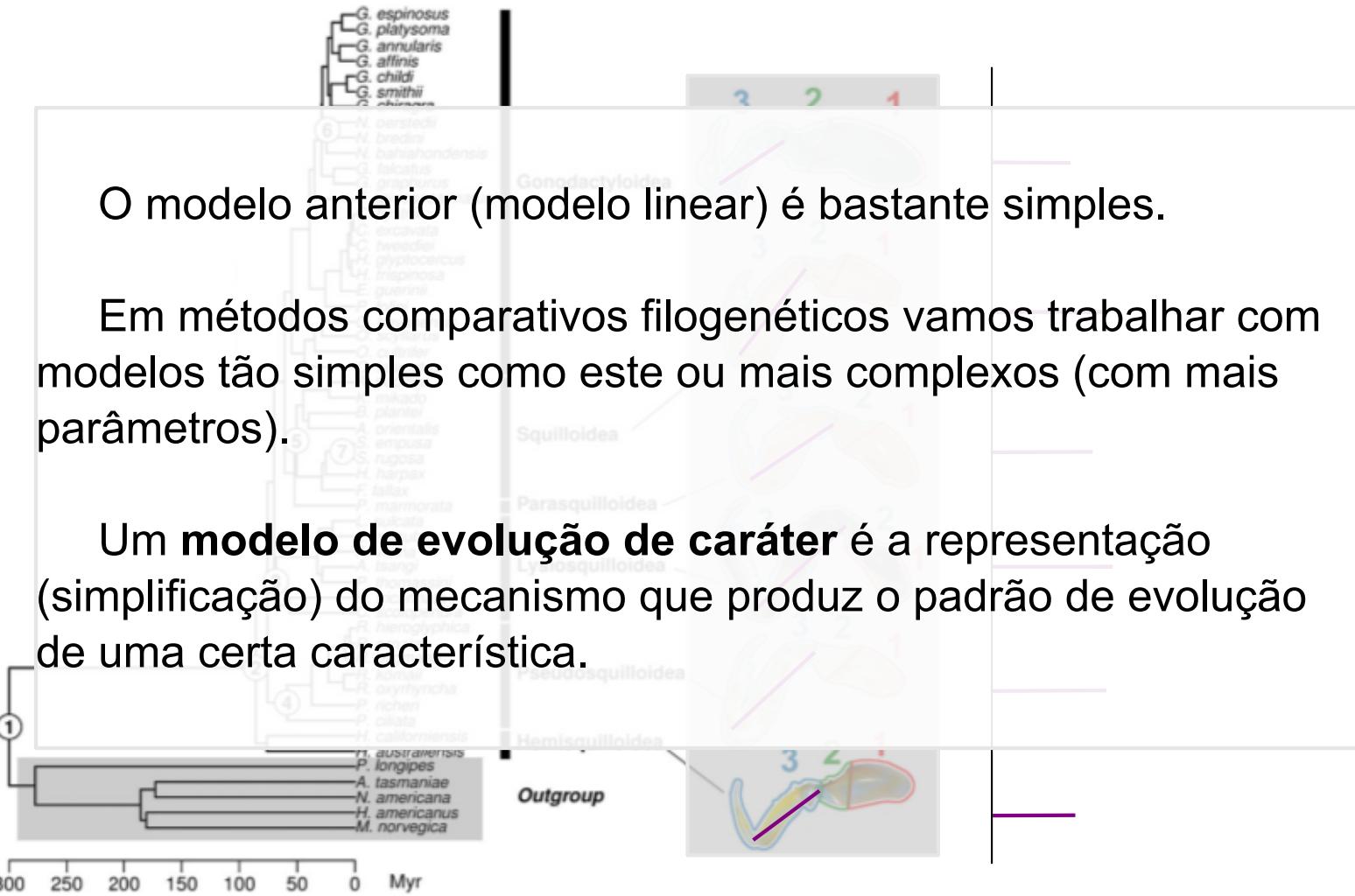
# O que é um modelo?

## Exemplos

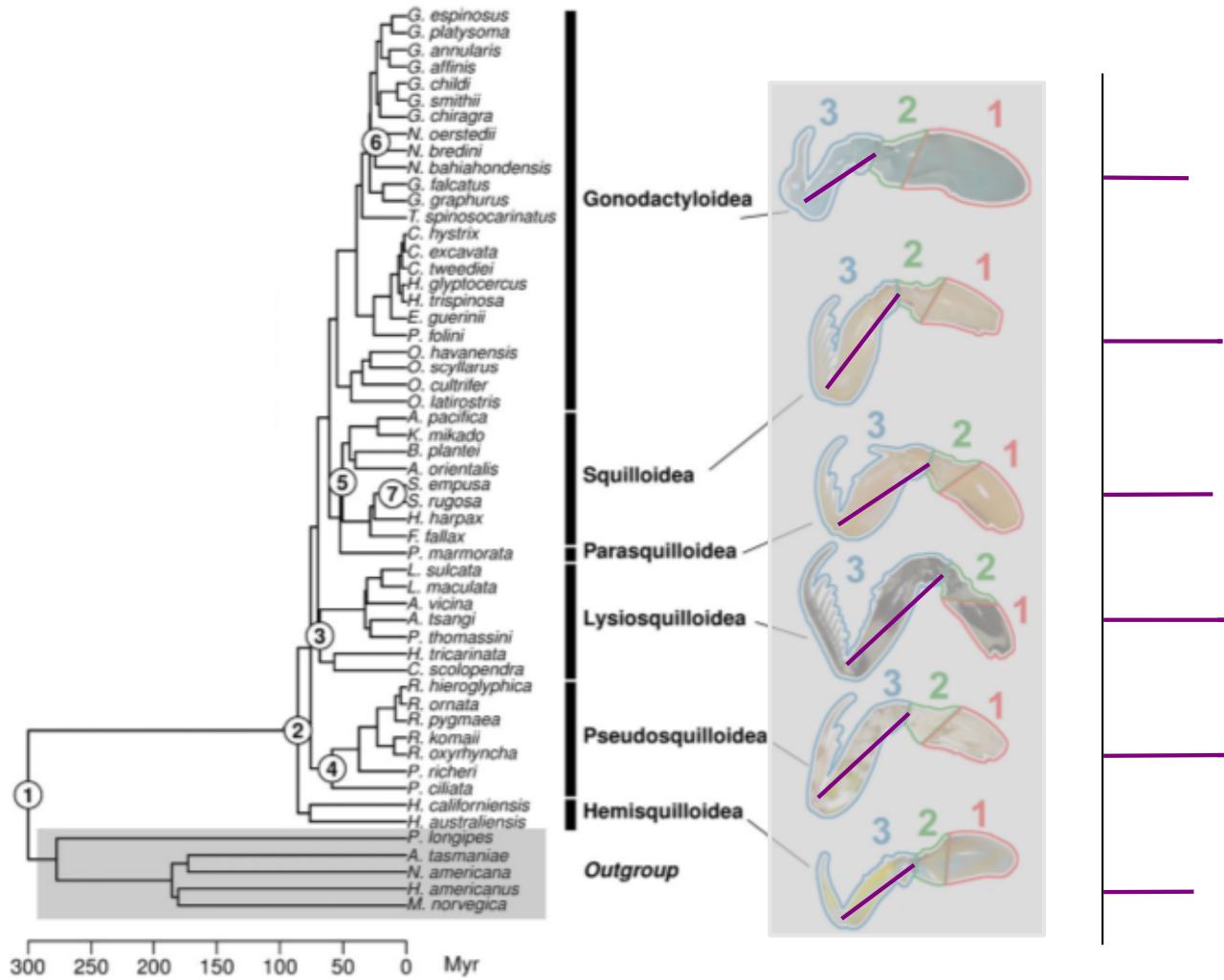


Método:  
modelo escolhido  
a)  $ax + b = y$   
estimativa dos parâmetros  
b)  $0.3x + 1.6 = y$

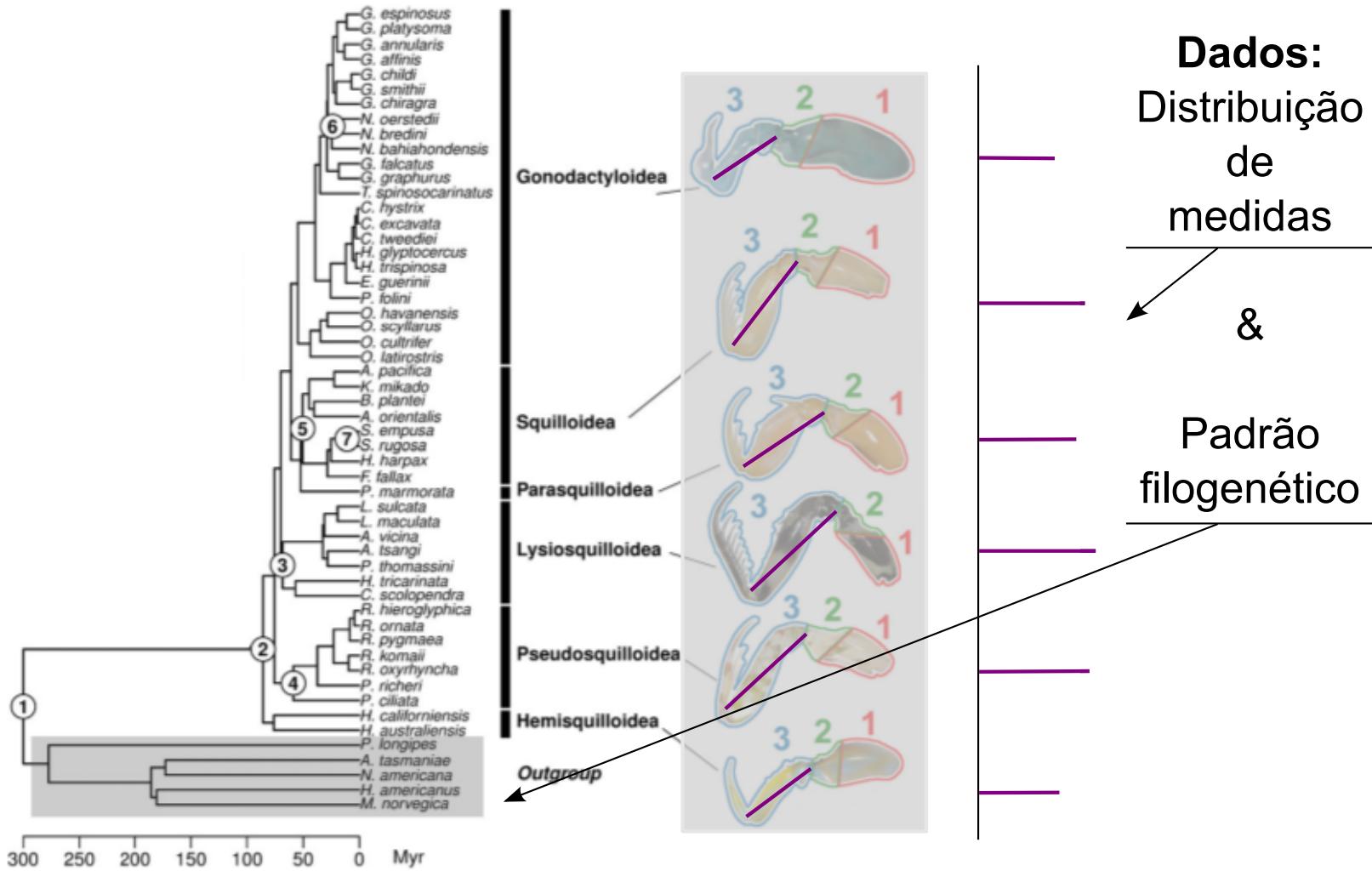
# O que é um modelo?



# O que é um modelo?



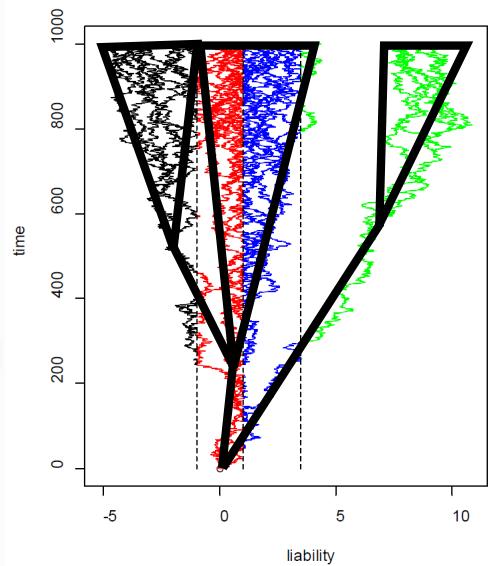
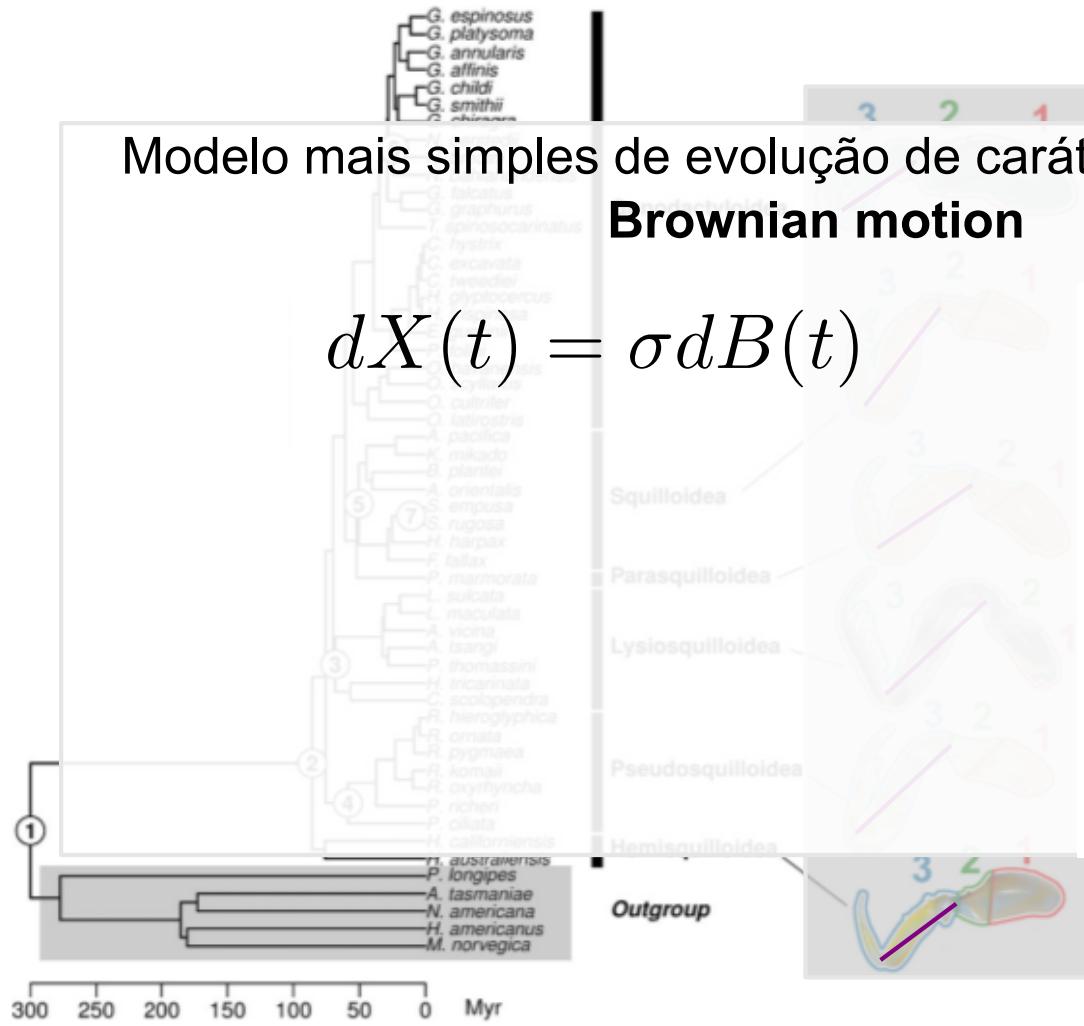
# O que é um modelo?



# O que é um modelo?

Modelo mais simples de evolução de caracteres contínuos é o **Brownian motion**

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



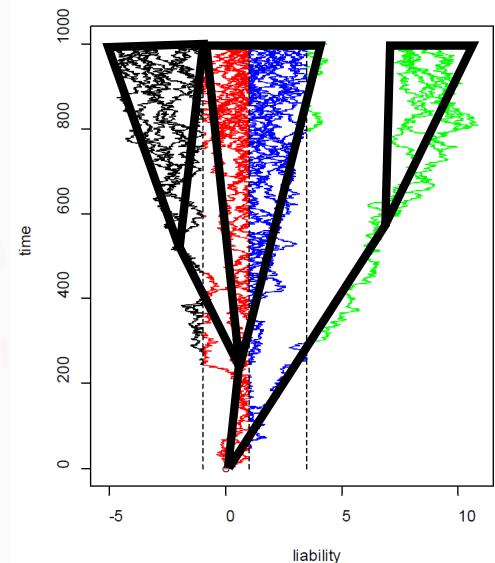
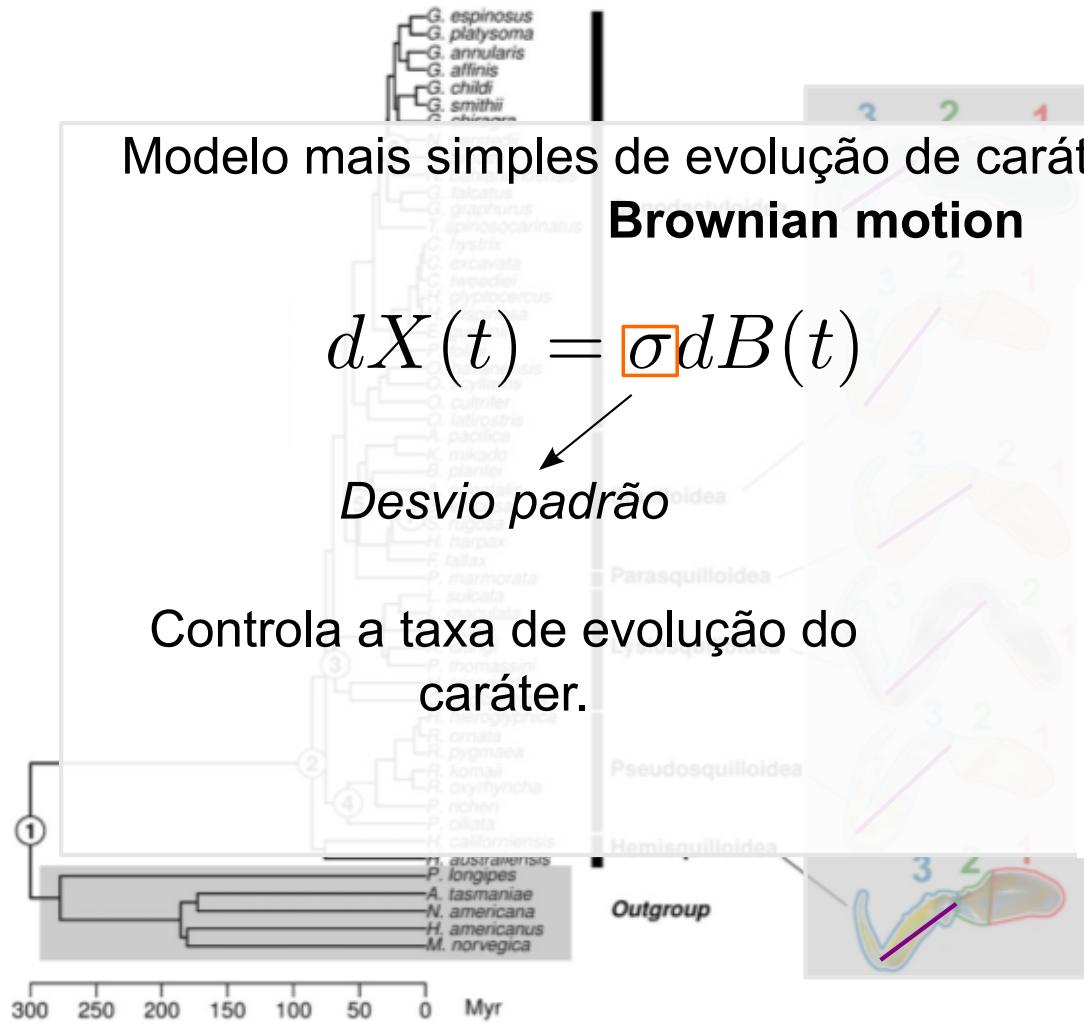
# O que é um modelo?

Modelo mais simples de evolução de caracteres contínuos é o **Brownian motion**

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

*Desvio padrão*

Controla a taxa de evolução do caráter.

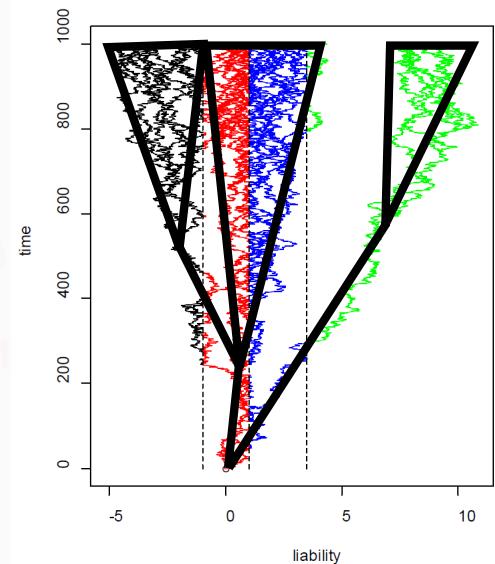
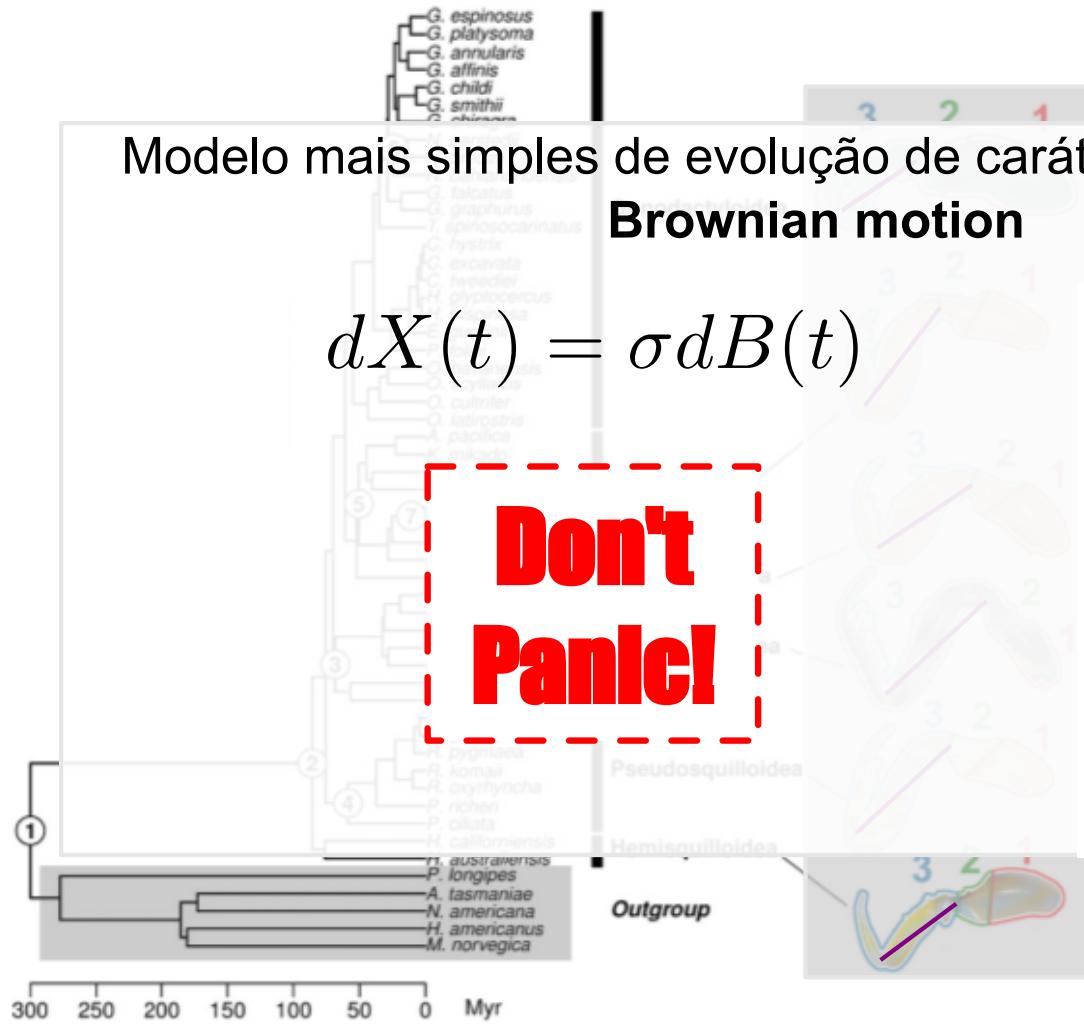


# O que é um modelo?

Modelo mais simples de evolução de caracteres contínuos é o  
**Brownian motion**

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

Don't  
Panic!



# Parâmetros dos modelos

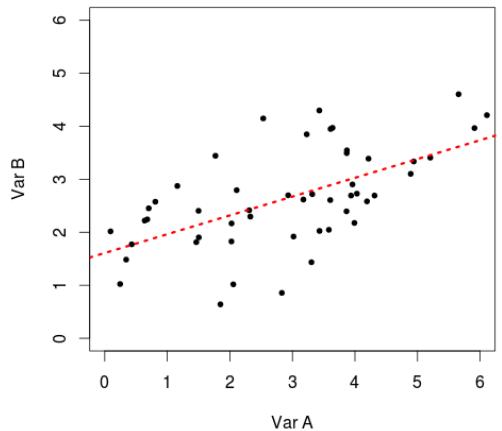
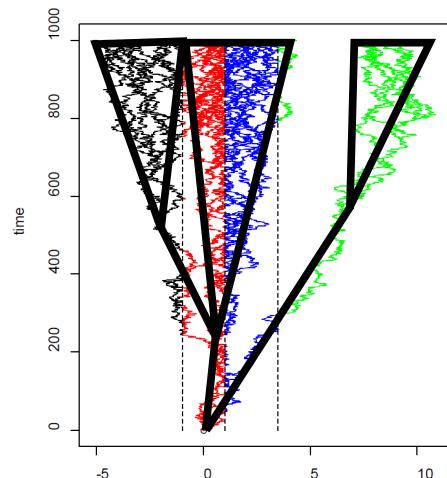
Número de parâmetros

1

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

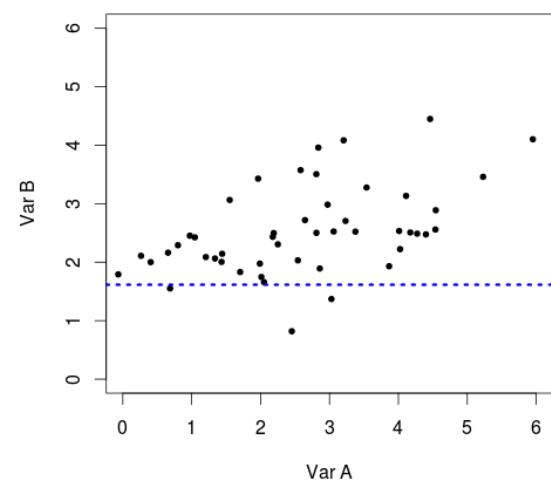
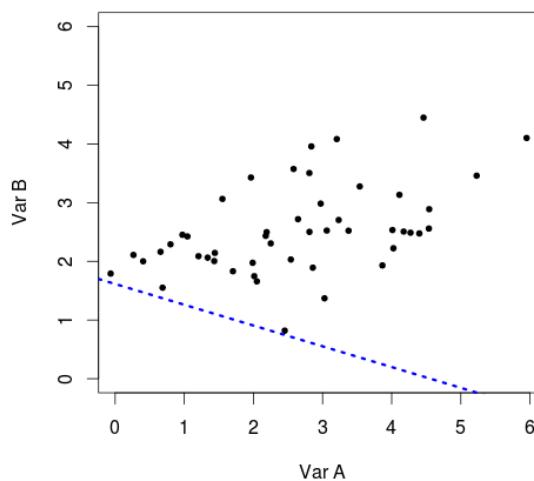
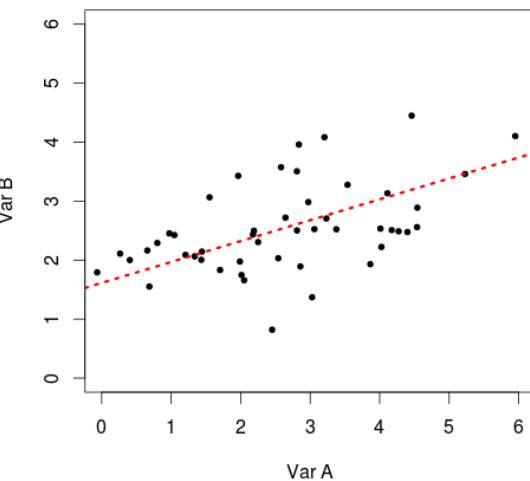
2

$$ax + b = y$$



# Parâmetros dos modelos

Valores dos parâmetros afetam o ajuste dos modelos



$$0.3x + 1.6 = y$$

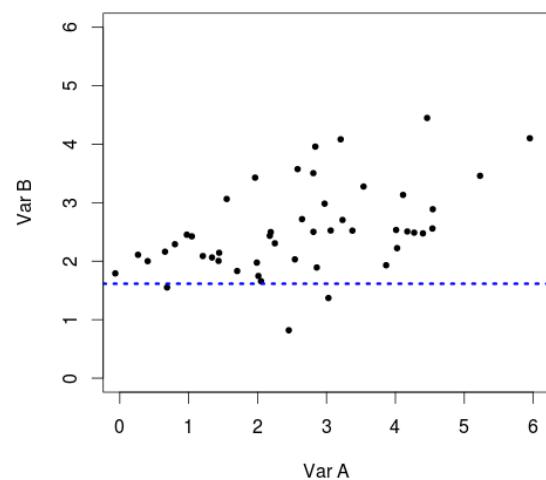
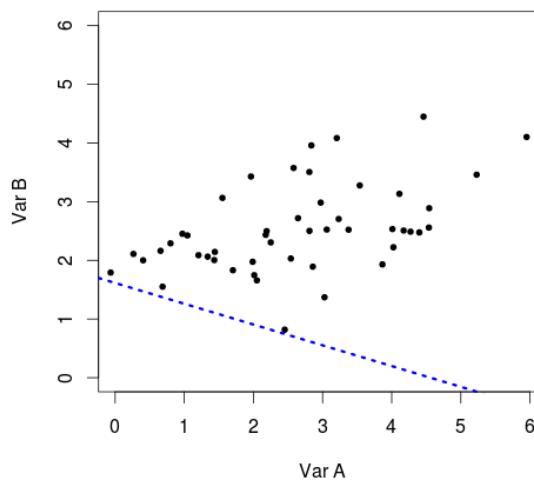
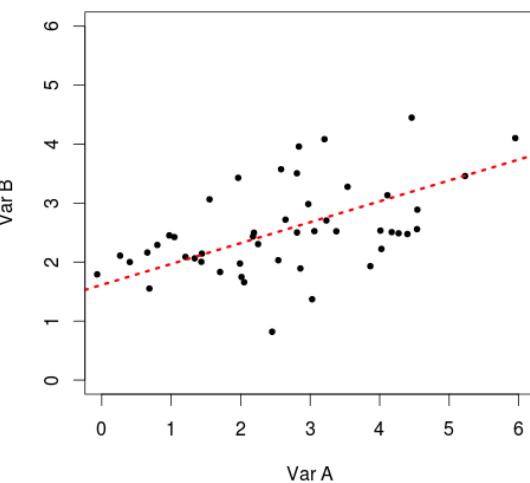
(melhor ajuste)

$$-0.3x + 1.6 = y$$

$$0x + 1.6 = y$$

# Parâmetros dos modelos

O principal objetivo desse curso é explorar um método para estimar os melhores valores para os parâmetros dos modelos incorporando incerteza (distribuição posterior).



$$0.3x + 1.6 = y$$

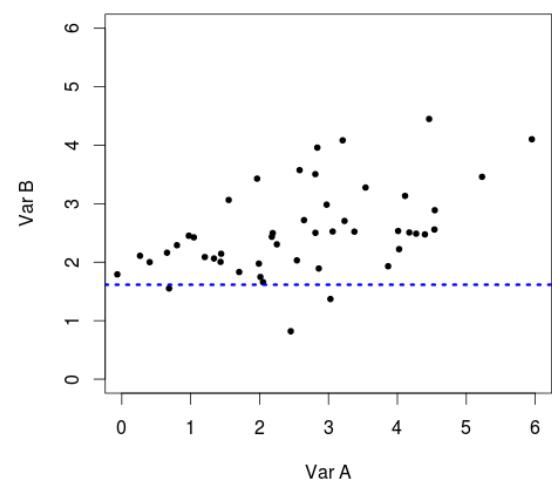
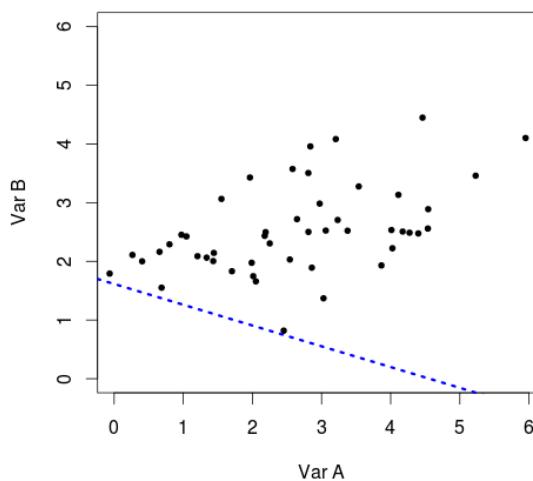
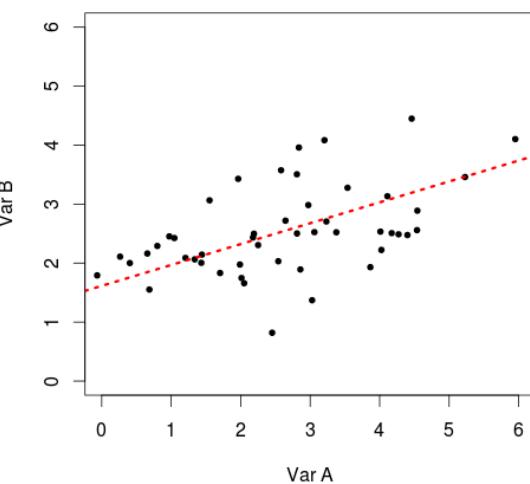
(melhor ajuste)

$$-0.3x + 1.6 = y$$

$$0x + 1.6 = y$$

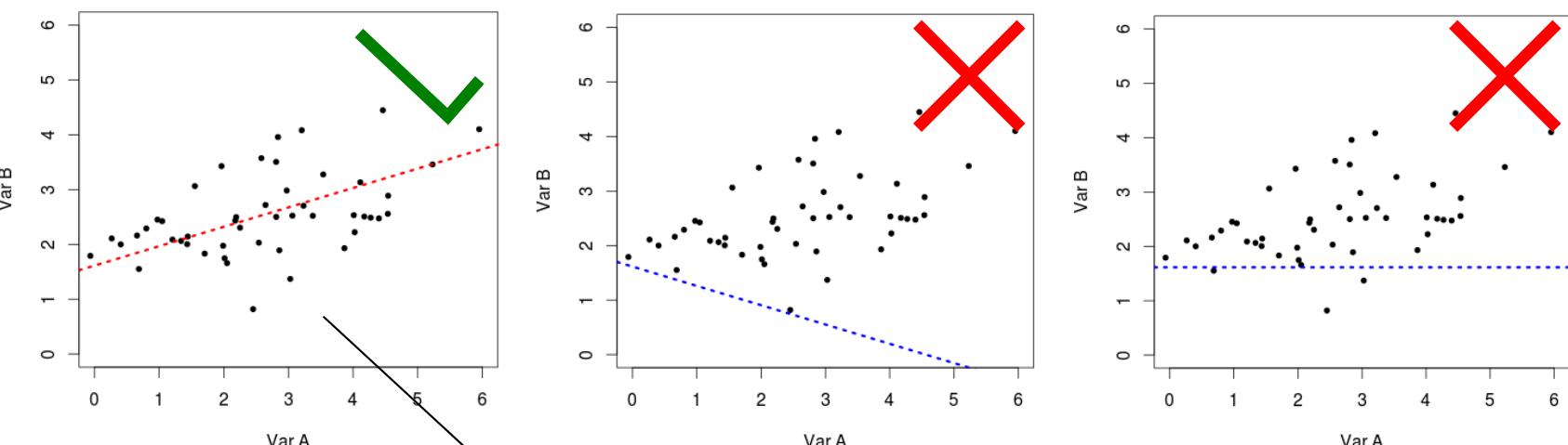
# Ideia: Estimando parâmetros

*Parameter estimate:* encontrar, usando algum método (direto ou heurístico), o(s) valore(s) dos parâmetro(s) que produzem o melhor ajuste do modelo aos dados.



# Ideia: Estimando parâmetros

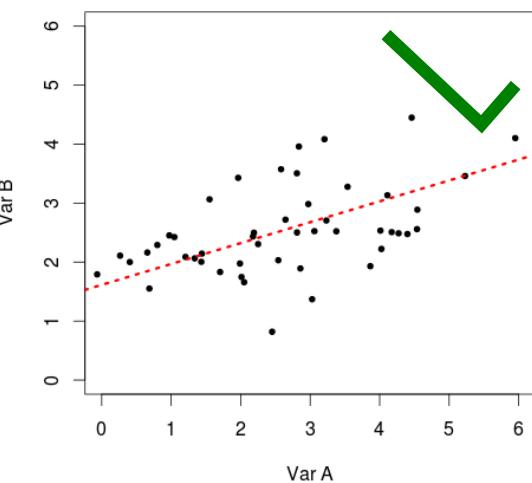
*Parameter estimate:* encontrar, usando algum método (direto ou heurístico), o(s) valore(s) dos parâmetro(s) que produzem o melhor ajuste do modelo aos dados.



→ Resultado esperado da estimativa.

# Ideia: Estimando parâmetros

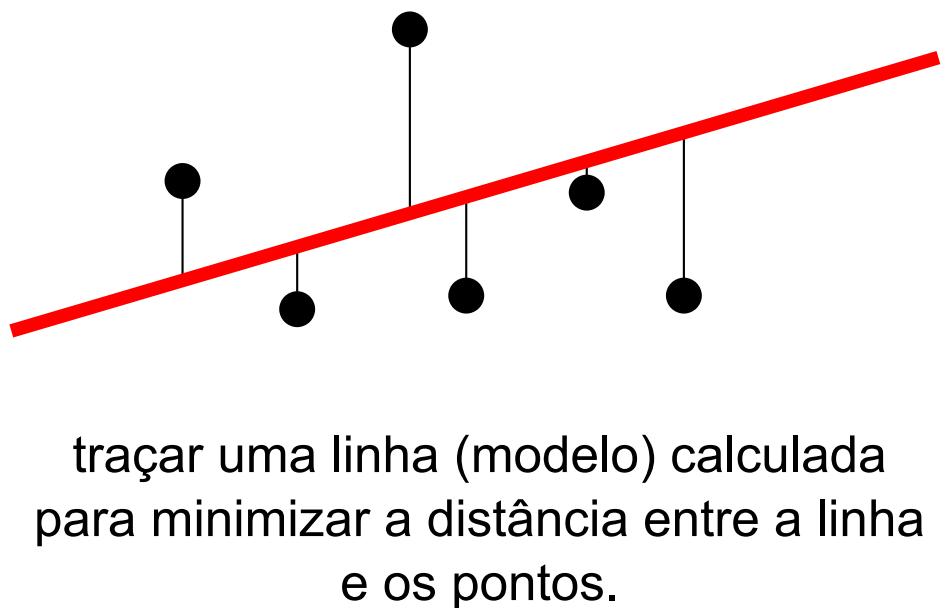
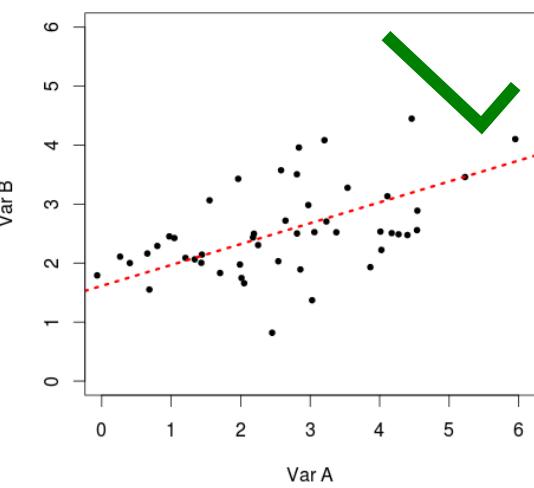
*Parameter estimate:* encontrar, usando algum método (direto ou heurístico), o(s) valore(s) dos parâmetro(s) que produzem o melhor ajuste do modelo aos dados.



- a) Método direto: *least squares*
- b) Método heurístico: no olho... sério.

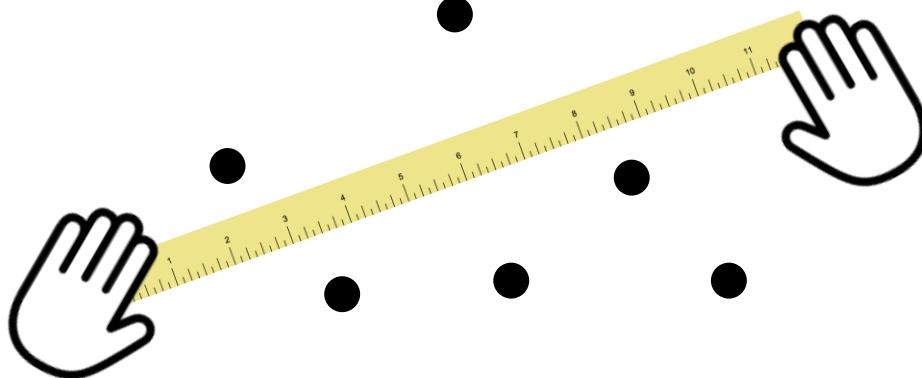
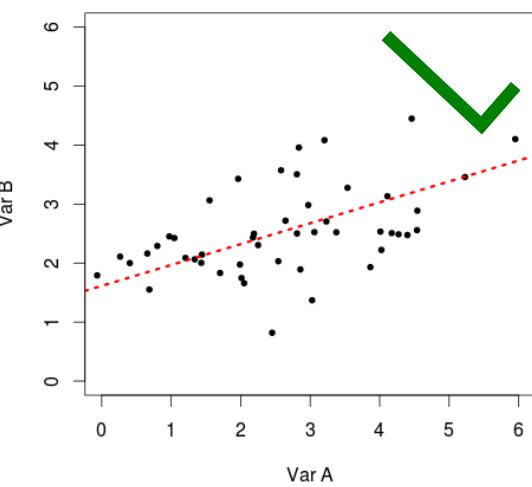
# Ideia: Estimando parâmetros

a) Método direto: *least squares*



# Ideia: Estimando parâmetros

b) Método heurístico: no olho... sério.



encontrar uma linha que minimiza a distância entre os pontos, ou seja, que parece "passar ali no meio de tudo".

# Ideia: Estimando parâmetros

Método heurístico: ajustando a antena da TV.

# Ideia: Estimando parâmetros

Método heurístico: ajustando a antena da TV.



# Ideia: Estimando parâmetros

Método heurístico: ajustando a antena da TV.



→ Muda o valor do parâmetro ...

... verifica o ajuste nos dados.



# Ideia: Estimando parâmetros

Em métodos comparativos a maioria dos modelos são ajustados por meio da busca pelos melhores valores para os parâmetros.

Somente poucos modelos, como os baseados em contraste filogenético, podem ser estimados utilizando um método direto.

A analogia de mudar a posição da antena e verificar a imagem da TV é uma boa descrição para a busca de **Markov chain Monte Carlo** e para a **estimativa de máxima verossimilhança**.