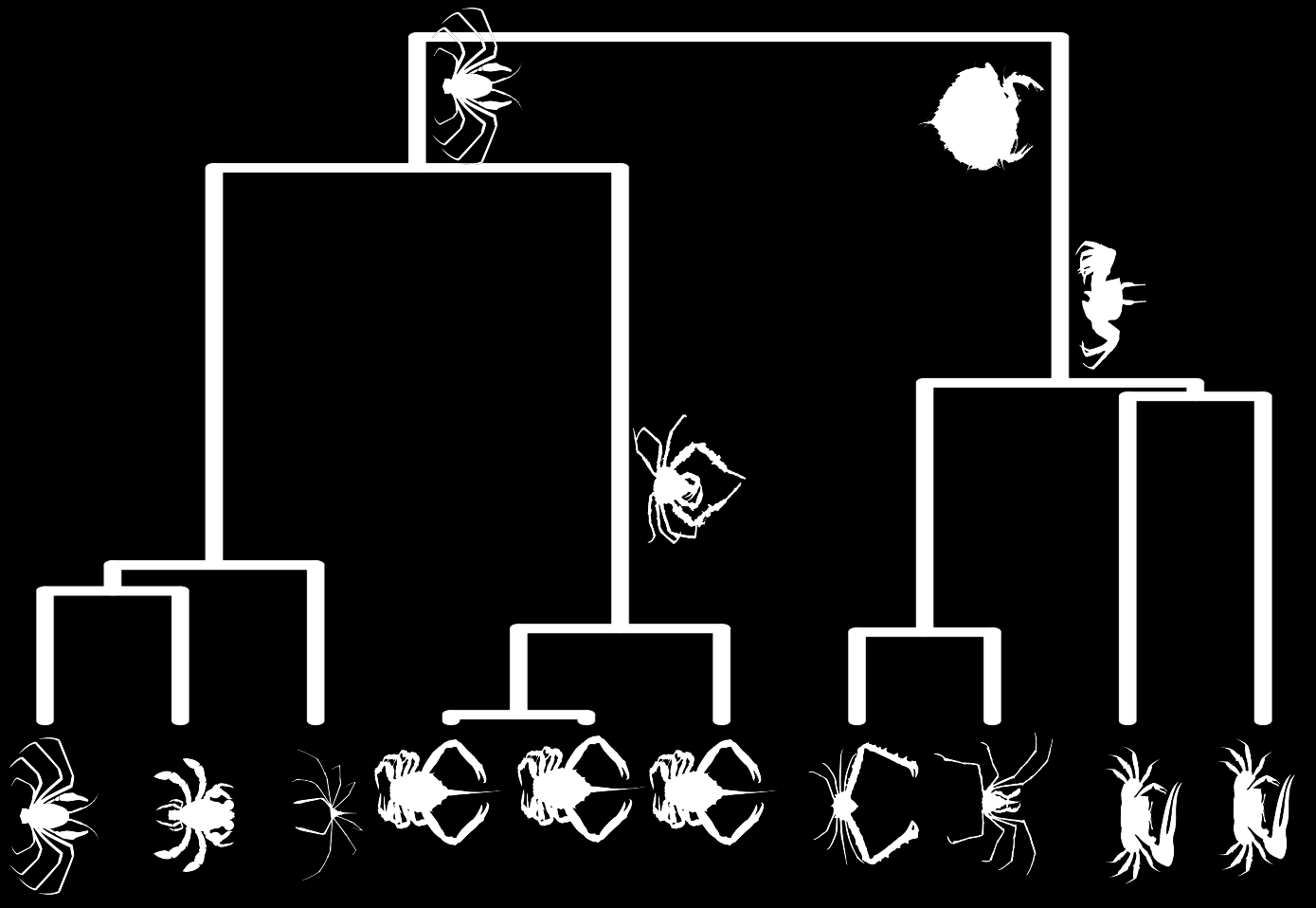
Disparidad a lo largo del tiempo

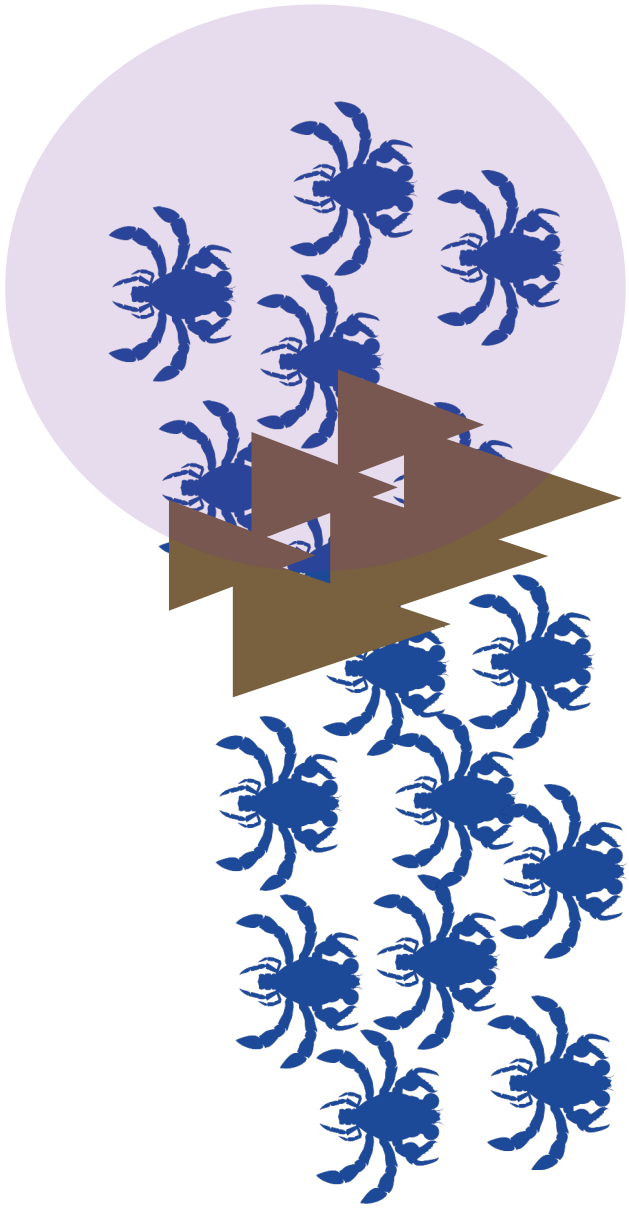
Patrones filogenéticos

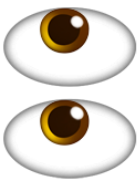
Dra. Karen López y Dr. Cristian Cervantes

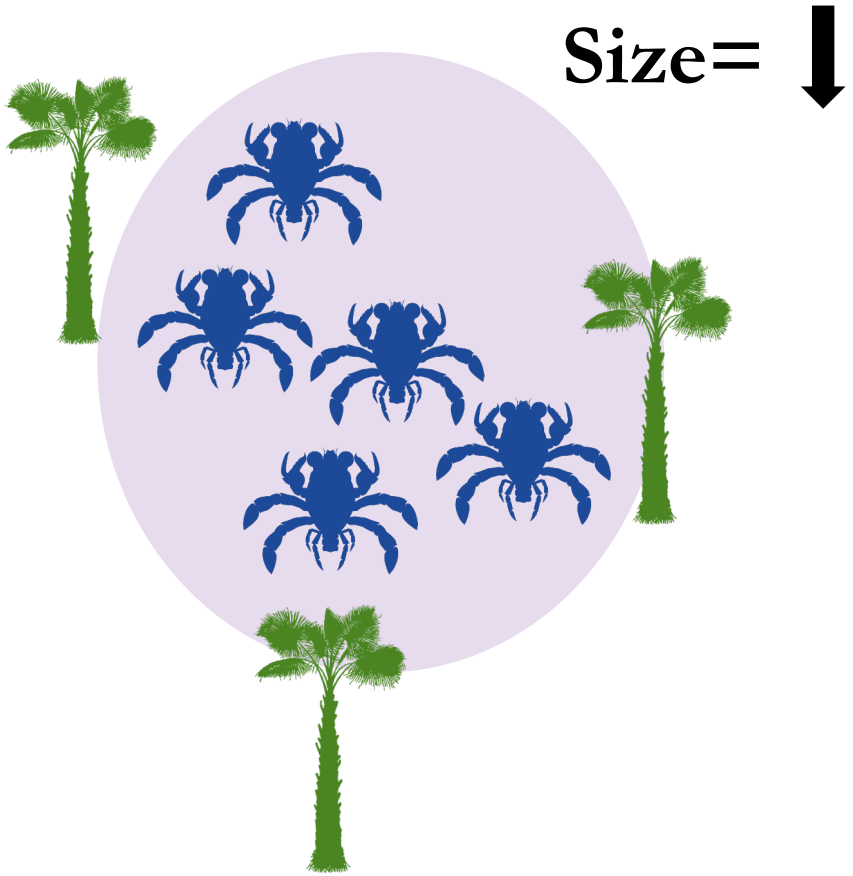
Painting by D.W. Miller. Courtesy of Smithsonian Institution

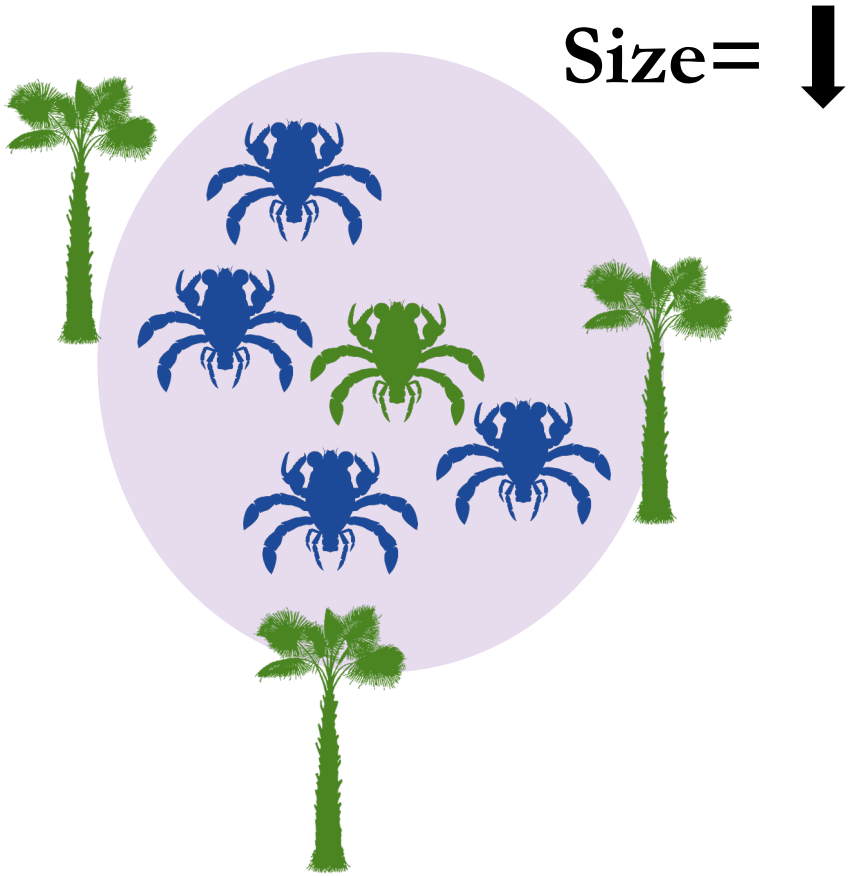


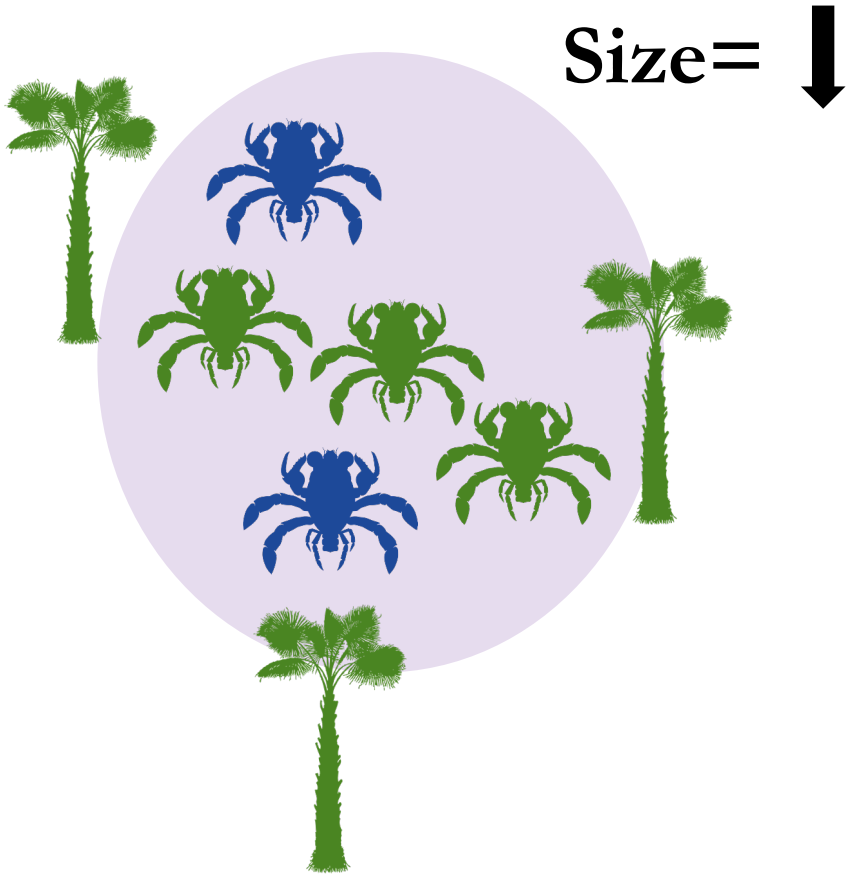
Time

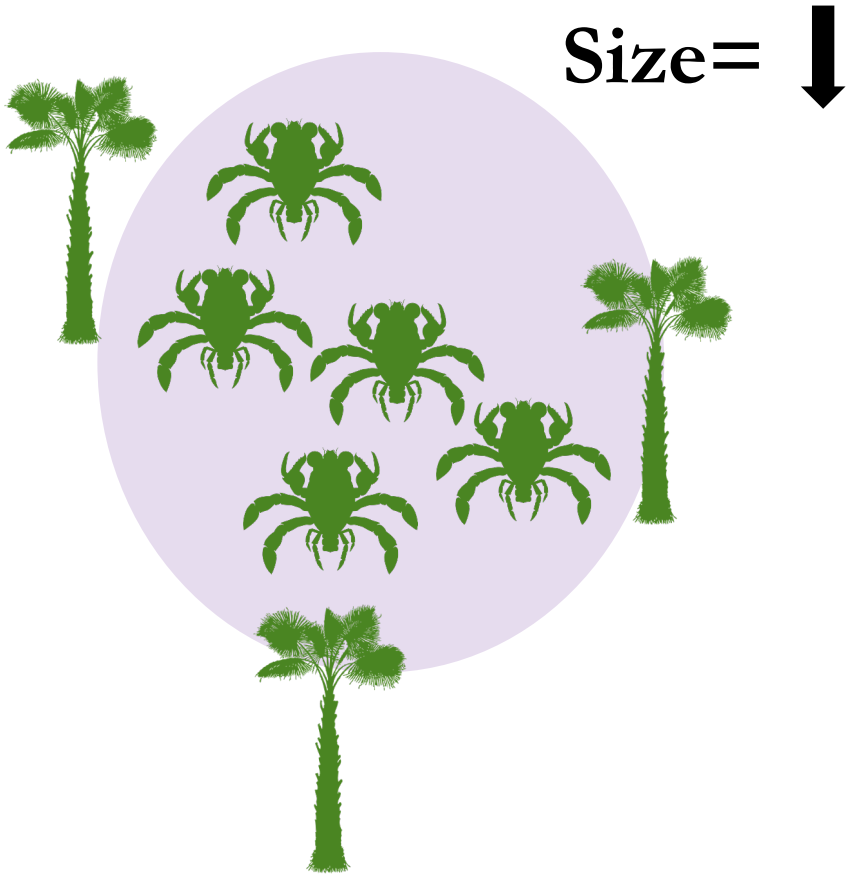


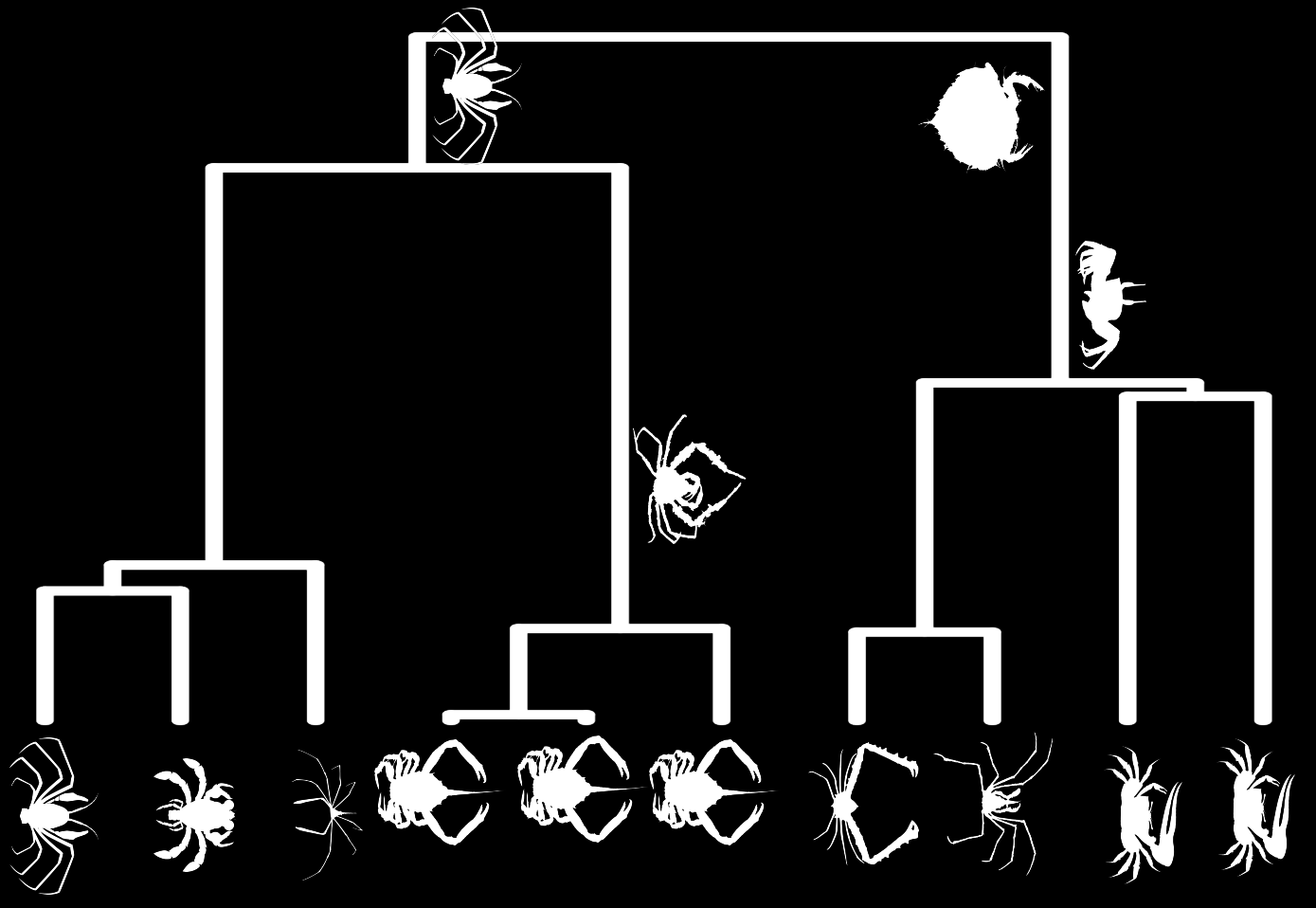



Size=

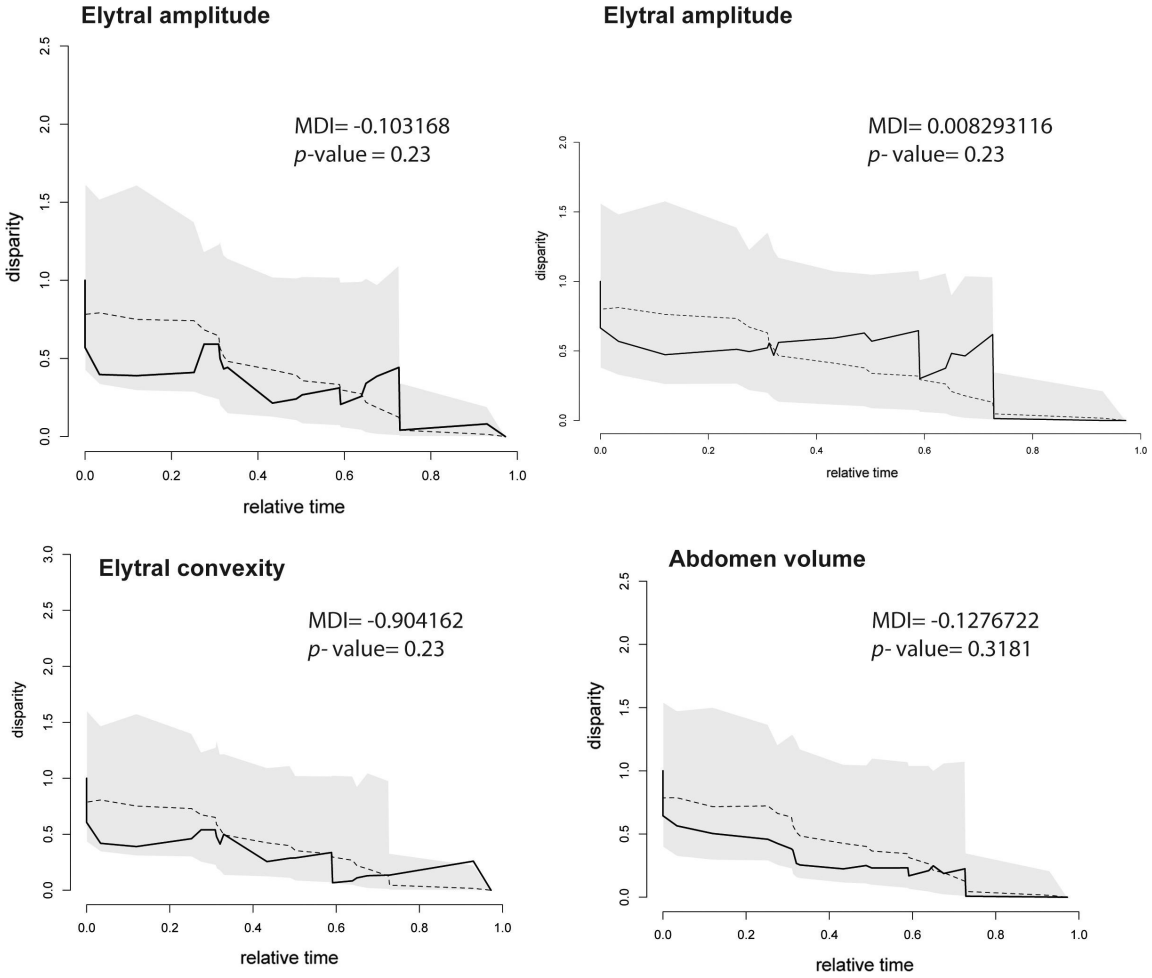

Size=


Size=


Size=



Disparity through time
Diasparidad a lo largo del tiempo
Harmon et al., 2008



Coevolución de carácteres

A Motivating Example - Peto's Paradox

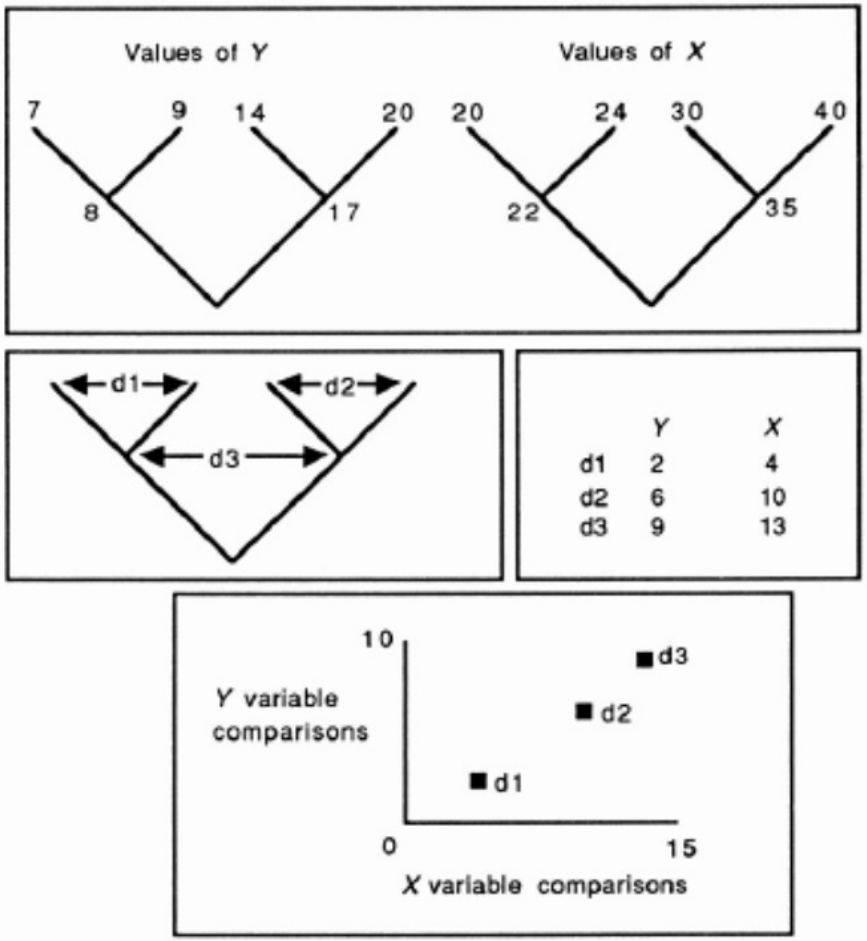
- At the most basic level, cancer is caused by certain cell mutations.
- Therefore, we would intuitively expect that the higher an organism's body mass, the more cells it has and the higher the probability that it develops cancer.
- Similarly, we would expect that the longer an organism's lifespan, the more time for a mutation to occur and the higher the probability that it develops cancer.
- This has been observed within humans.
- But at an interspecies level, evidence hasn't been found to support this seemingly intuitive hypothesis - Peto's paradox.
- We could thus employ phylogenetic comparative methods to examine, for example, the evolutionary relationship between the traits of body mass and cancer mortality.

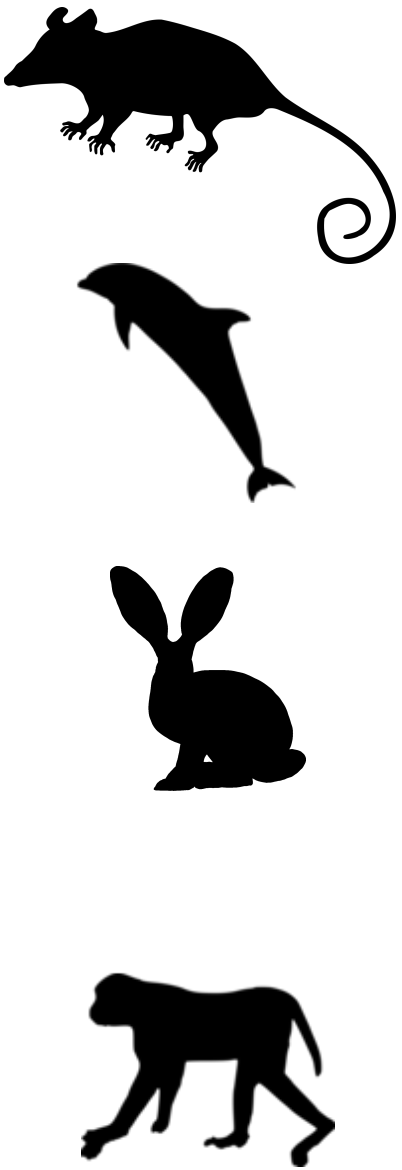
Comparative Methods: Phylogenetically Independent Contrasts (PICS)

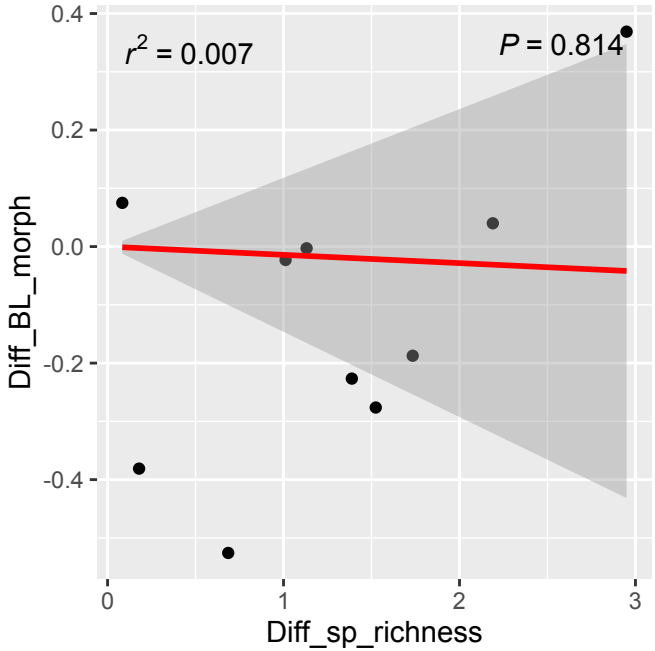
- The logic behind PICs is to transform the original tip data into values that are statistically independent and identically distributed.
- We can do this by taking a series of contrasts between character values at the tips.
- Under a Brownian motion model, a contrast is due to displacement that has occurred since the two species split - ie the non-shared part of their evolutionary histories.
- Hence, the contrasts are statistically independent.
- We can then standardize the contrasts by dividing each by the square root of its variance.
- We then perform ordinary least squares regression on the contrasts.

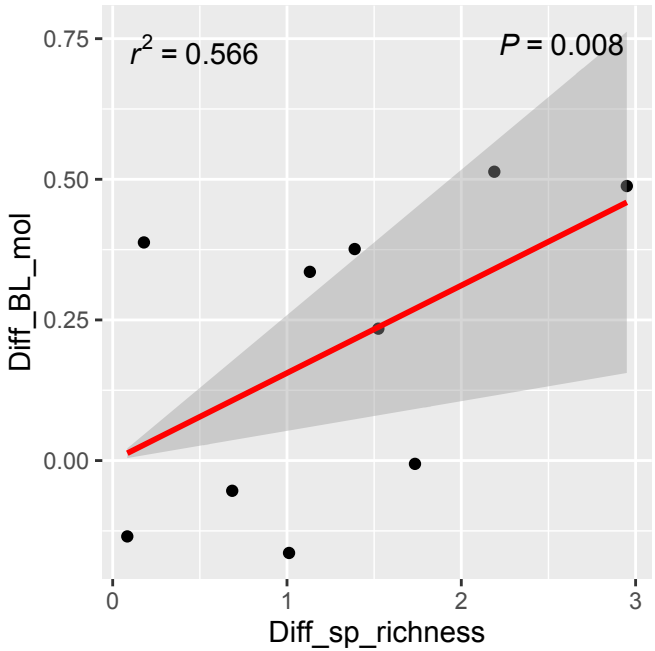
Phylogenetically independent contrasts

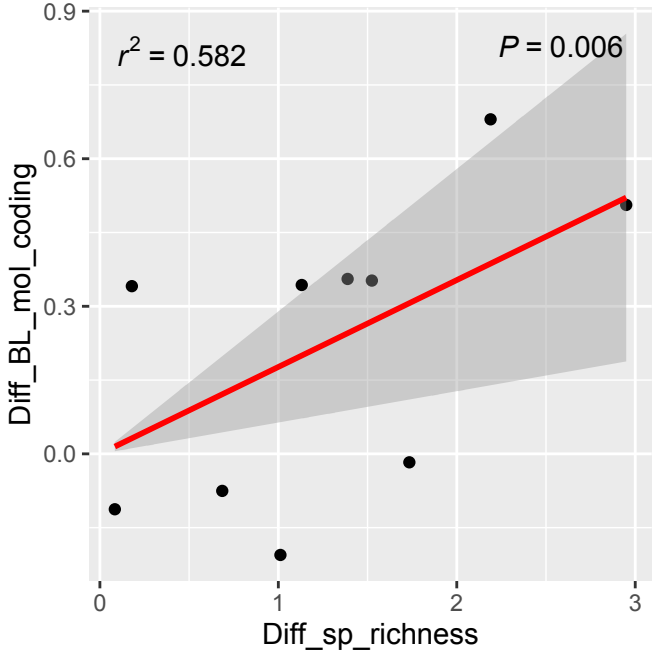
Under a Brownian motion model of evolution, d1, d2, and d3 provide independent comparisons. Path length differences are ignored in this illustration.

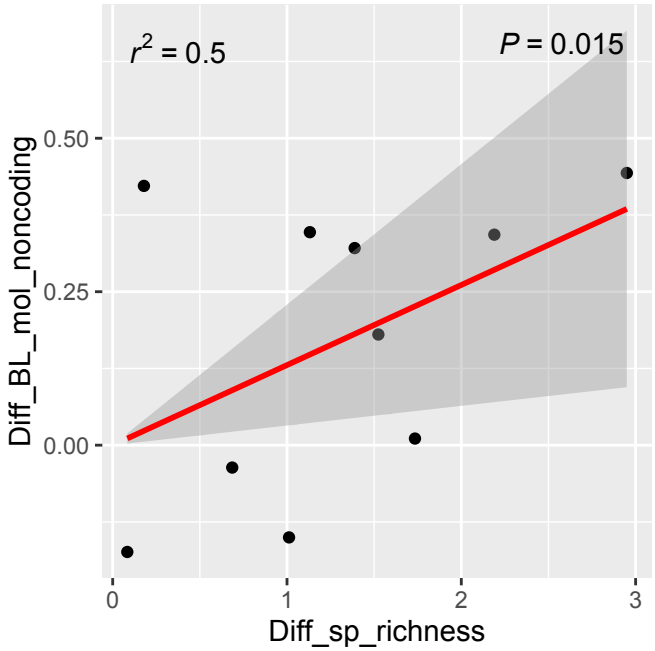





r^2 = 0.007
P = 0.814
Diff_BL_morph
Diff_sp_richness
0.4
0.2
0.0
-0.2
-0.4
0
1
2
3

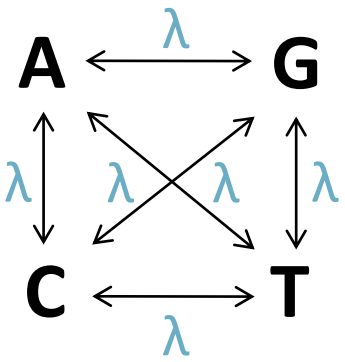

r^2 = 0.566
P = 0.008
Diff_BL_mol
Diff_sp_richness
0.75
0.50
0.25
0.00
0
1
2
3


r^2 = 0.582
P = 0.006
Diff_BL_mol_coding
Diff_sp_richness
0.9
0.6
0.3
0.0
0
1
2
3


r^2 = 0.5
P = 0.015
Diff_BL_mol_noncoding
Diff_sp_richness
0.50
0.25
0.00
0
1
2
3

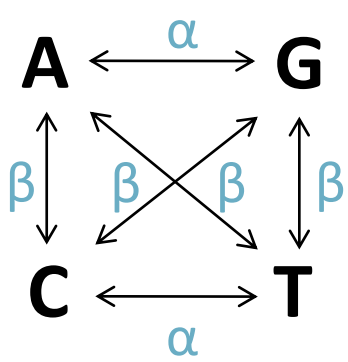
Nucleotide substitution models

JC



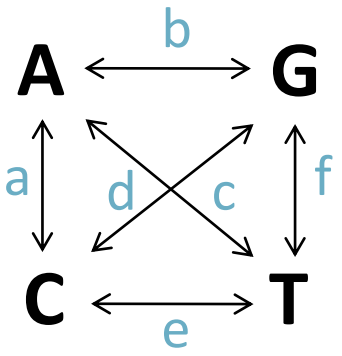
$\pi_A=\pi_C=\pi_G=\pi_T$

HKY



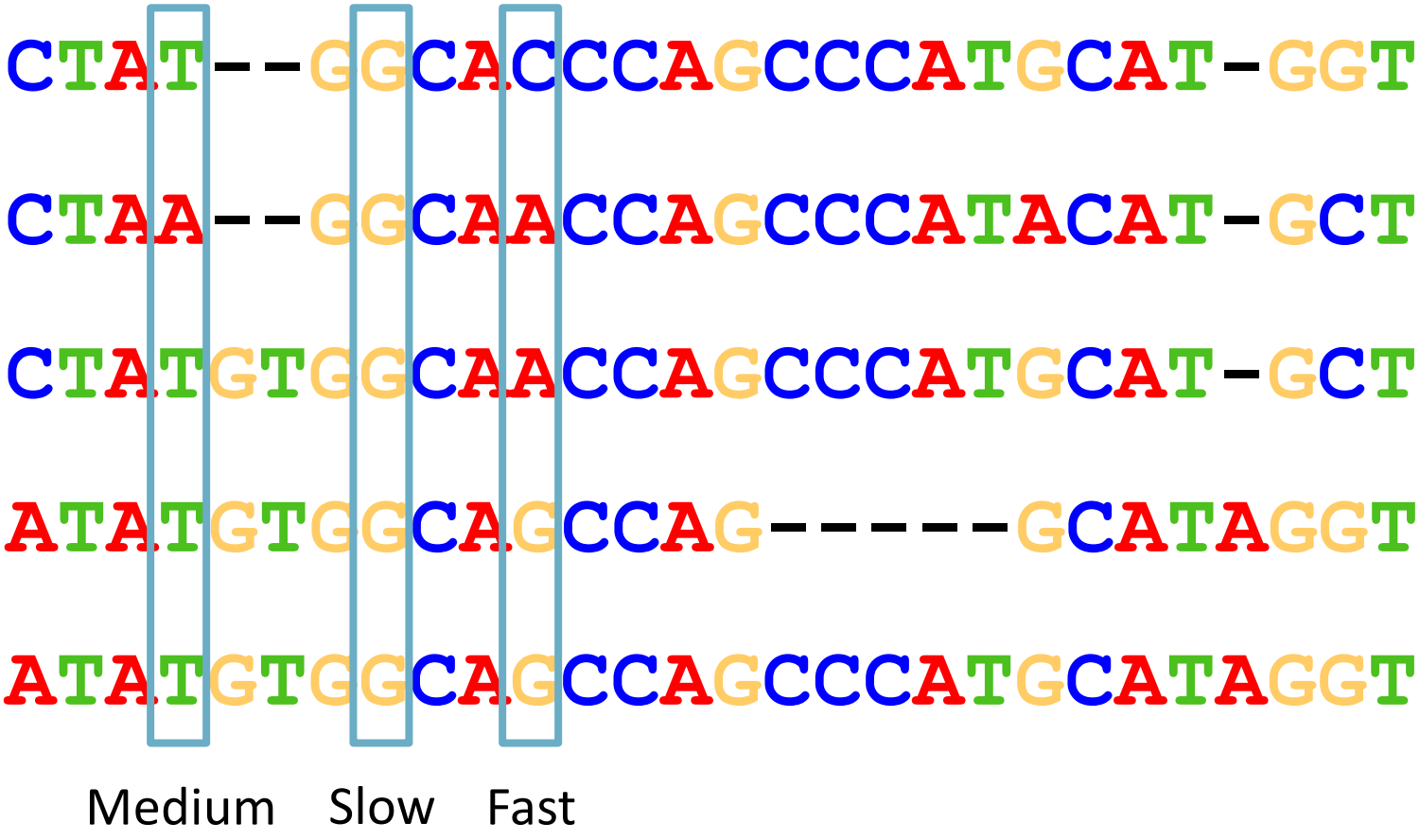
$\pi_A, \pi_C, \pi_G, \pi_T$

GTR



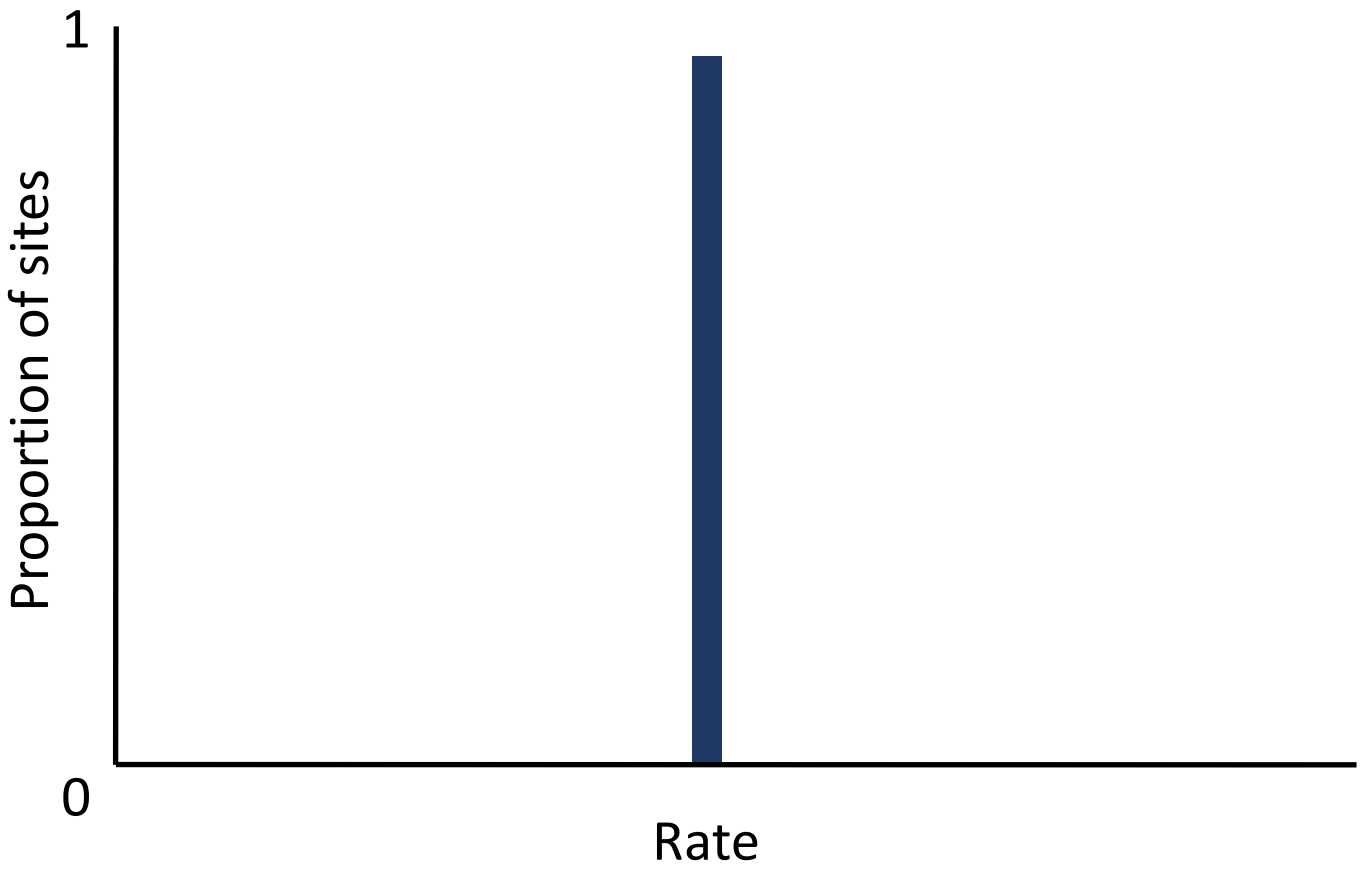
$\pi_A, \pi_C, \pi_G, \pi_T$

Rate variation across sites



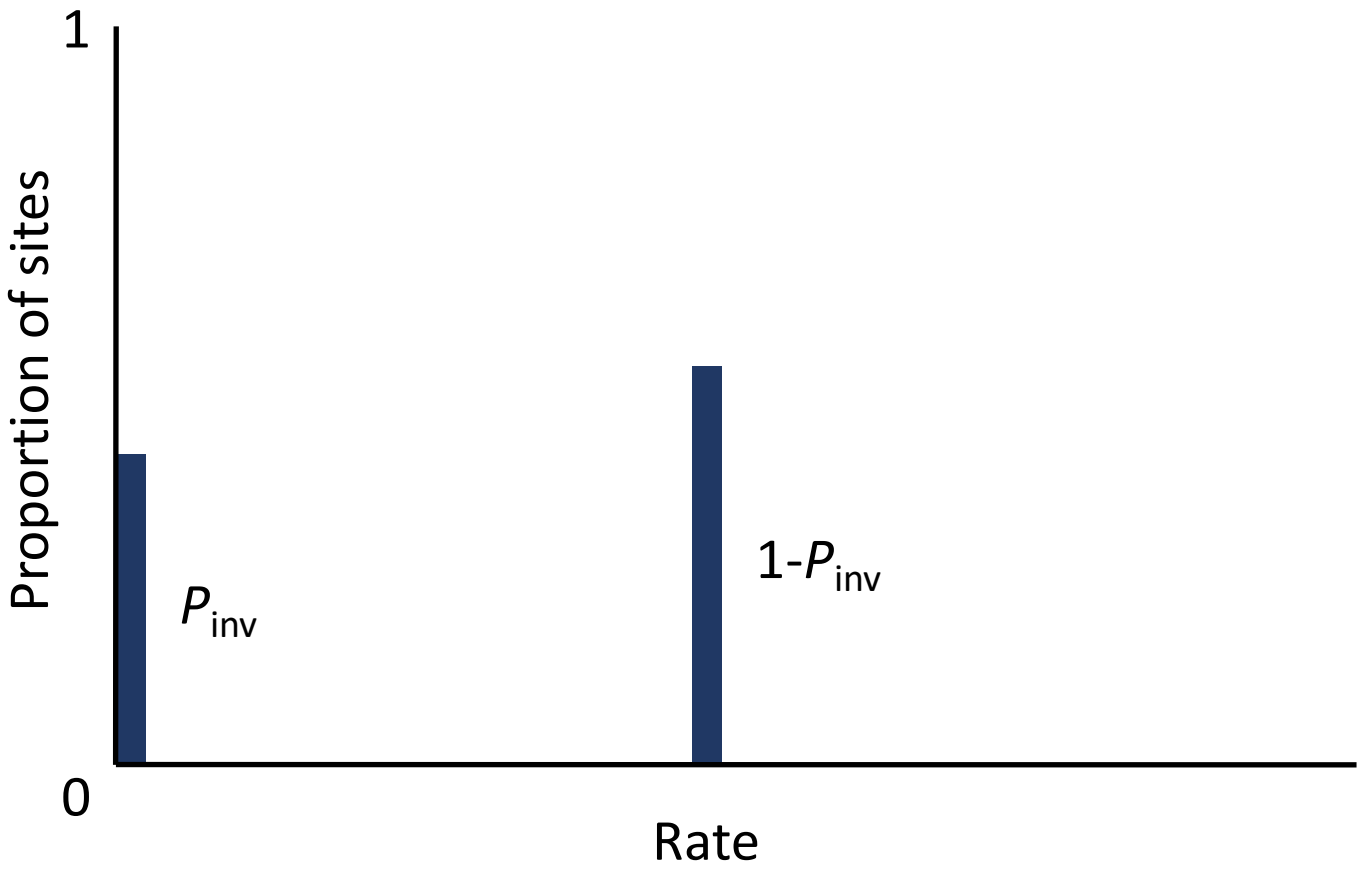
Rate variation across sites

- Equal rates among sites



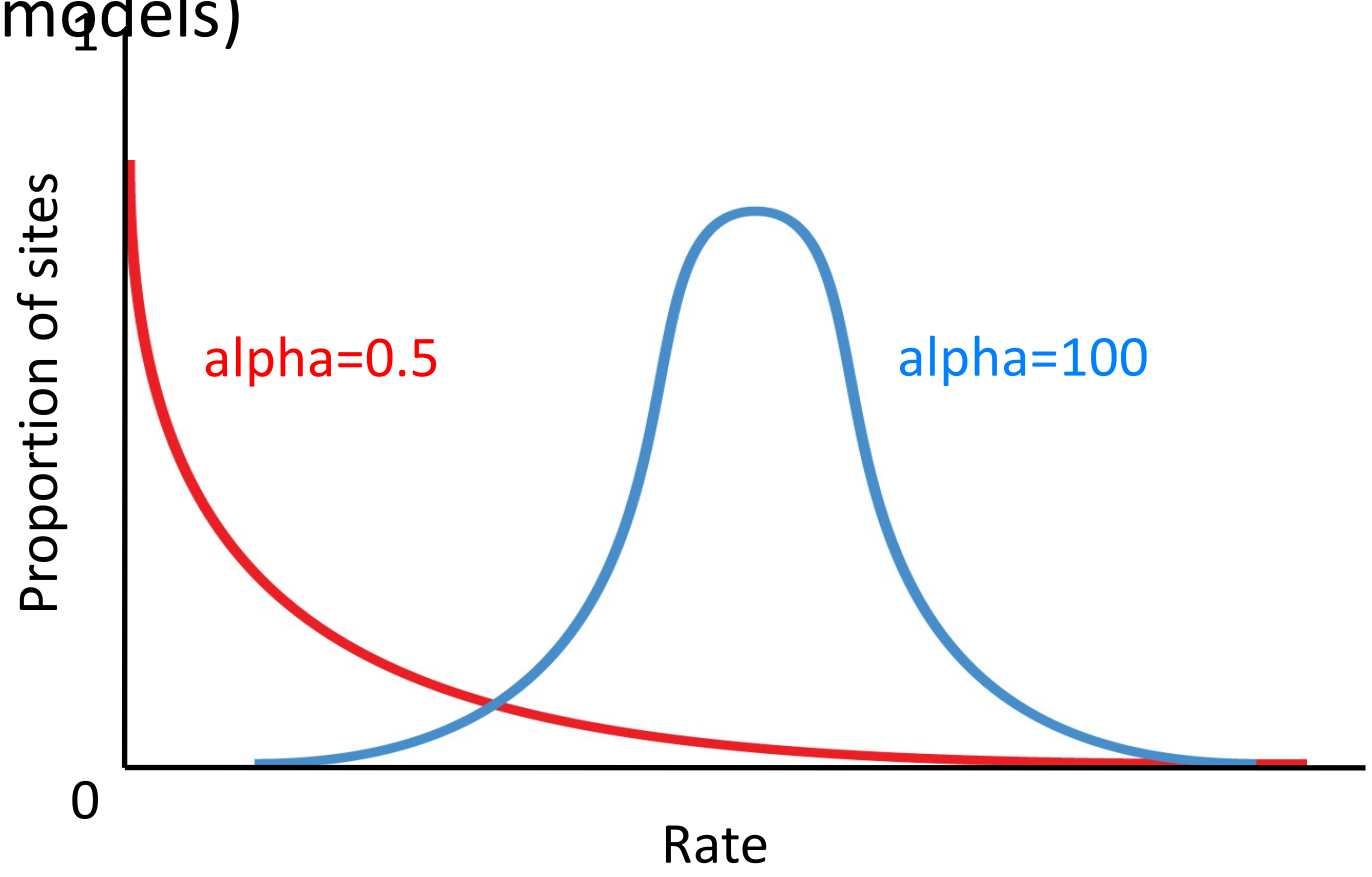
Rate variation across sites

- Proportion of invariable sites (**+I** models)



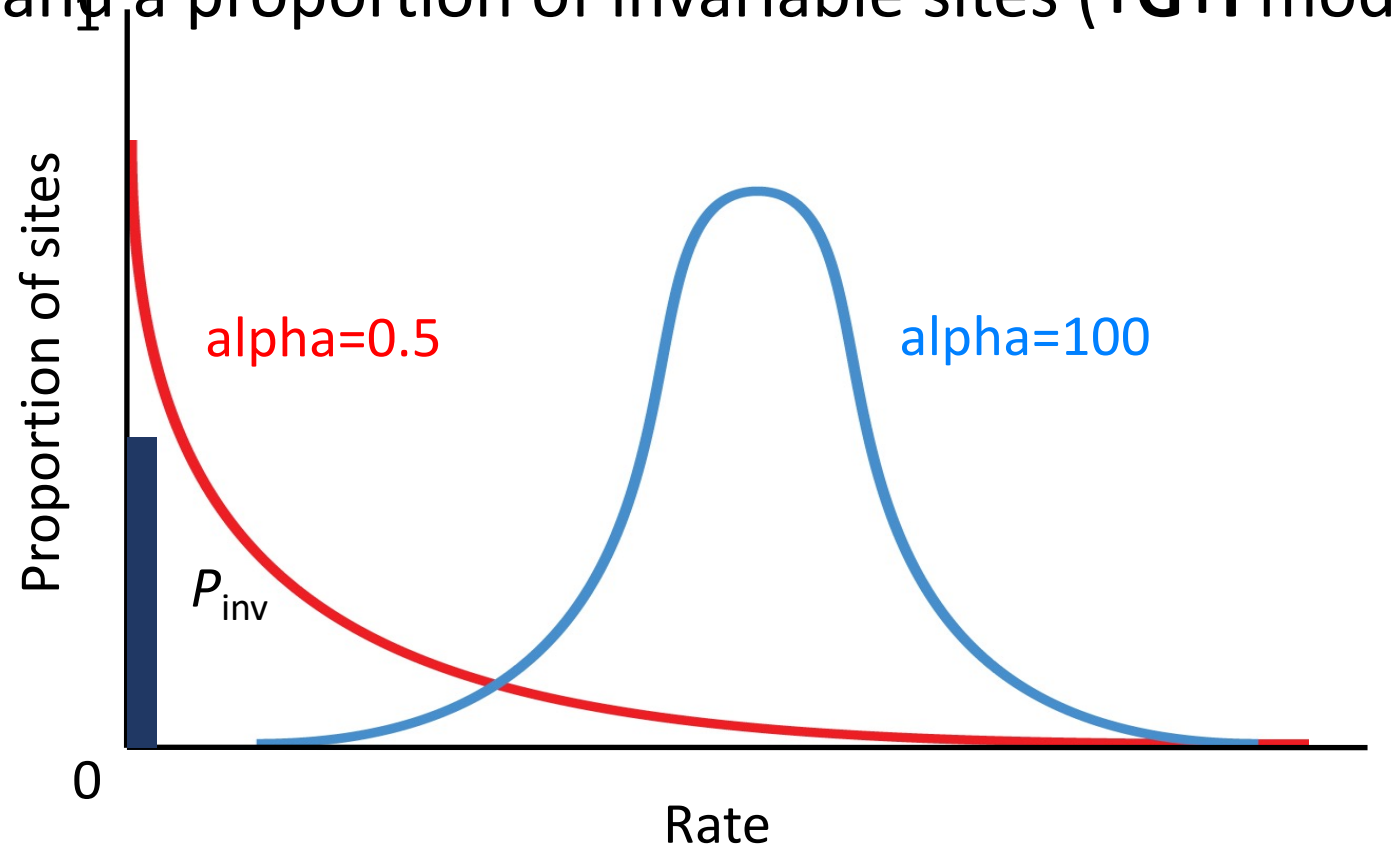
Rate variation across sites

- Gamma-distributed rate variation across sites (**+G** models)



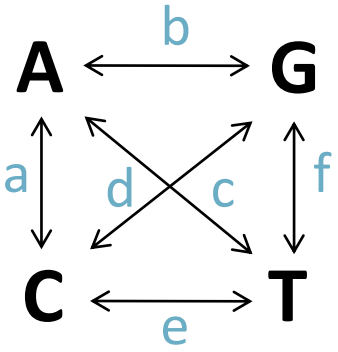
Rate variation across sites

- Gamma-distributed rate variation across sites and a proportion of invariable sites (**+G+I** models)



Nucleotide substitution models

rate matrix



base frequencies

$\pi_A + \pi_C + \pi_G + \pi_T = 1$

site rates

+ I + G

most complex time-reversible model:	**GTR+I+G**
	a, b, c, d, e, f
	π_A, π_C, π_G, π_T
	I, G

JOURNAL ARTICLE

A Phylogenetic Mixture Model for Detecting Pattern-Heterogeneity in Gene Sequence or Character-State Data

Mark Pagel , Andrew Meade

Systematic Biology, Volume 53, Issue 4, August 2004, Pages 571–581,
https://doi.org/10.1080/10635150490468675
Published: 01 August 2004 **Article history** ▾

Abstract

We describe a general likelihood-based 'mixture model' for inferring phylogenetic trees from gene-sequence or other character-state data. The model accommodates cases in which different sites in the alignment evolve in qualitatively distinct ways, but does not require prior knowledge of these patterns or partitioning of the data. We call this qualitative variability in the pattern of evolution across sites "pattern-heterogeneity" to distinguish it from both a homogenous process of evolution and from one characterized principally by differences in rates of evolution. We present studies to show that the model correctly retrieves the signals of pattern-heterogeneity from simulated gene-sequence data, and we apply the method to protein-coding genes and to a ribosomal 12S data set. The mixture model outperforms conventional partitioning in both these data sets. We implement the mixture model such that it can simultaneously detect rate- and pattern-heterogeneity. The model simplifies to a homogeneous model or a rate-variability model as special cases, and therefore always performs at least as well as these two approaches, and often considerably improves upon them. We make the model available within a Bayesian Markov-chain Monte Carlo framework for phylogenetic inference, as an easy-to-use computer program.

$$Q = \begin{array}{c} \\ 00 \\ 10 \\ 01 \\ 11 \end{array} \begin{array}{c} \begin{array}{cccc} 00 & 10 & 01 & 11 \end{array} \\ \begin{pmatrix} - & \alpha_A & \alpha_B & 0 \\ \beta_A & - & 0 & \alpha_B \\ \beta_B & 0 & - & \alpha_A \\ 0 & \beta_A & \beta_B & - \end{pmatrix} \end{array}$$

$$Q = \begin{array}{c} \\ 00 \\ 10 \\ 01 \\ 11 \end{array} \begin{array}{c} \begin{array}{cccc} 00 & 10 & 01 & 11 \end{array} \\ \begin{pmatrix} - & \mu_{1,2} & \mu_{1,3} & 0 \\ \mu_{2,1} & - & 0 & \mu_{2,4} \\ \mu_{3,1} & 0 & - & \mu_{3,4} \\ 0 & \mu_{4,2} & \mu_{4,3} & - \end{pmatrix} \end{array}$$

$\mu_{1,2} \neq \mu_{3,4}$: a gain in character A does not depend on the state of character B

- $\mu_{2,1} \neq \mu_{4,3}$: a loss in character A does not depend on the state of character B
- $\mu_{1,3} \neq \mu_{2,4}$: a gain in character B does not depend on the state of character A
- $\mu_{3,1} \neq \mu_{4,2}$: a loss in character B does not depend on the state of character A

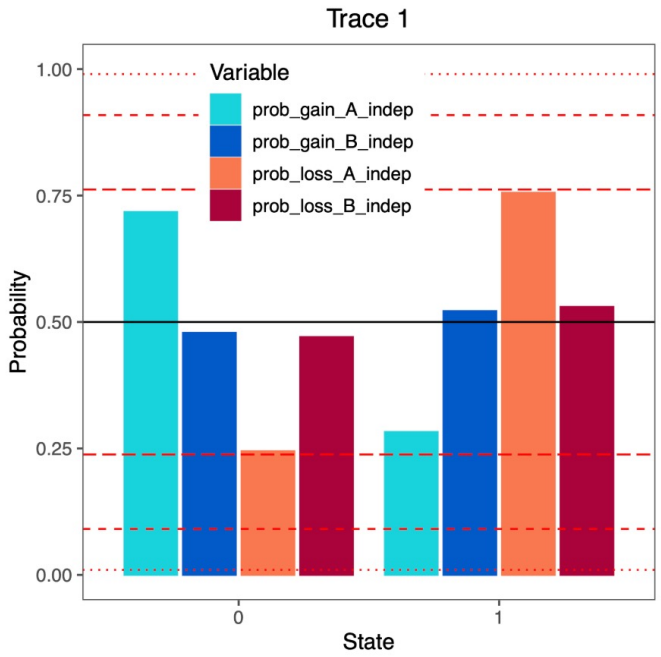


Figure 2. Probability that a rate of gain or loss of a character was independent. If we see a high posterior probability for 1, then that means that the rate of gain or loss for the character is independent of the other character. We also show the prior (black solid line), weak support (BF < 3.2, long-dashed red line), substantial support ($3.2 <$ BF < 10, dashed red line), and strong support ($10 <$ BF < 100, dotted red line). Even though the support varies, for this specific analysis we don't see any significant support of either correlated or independent rates.