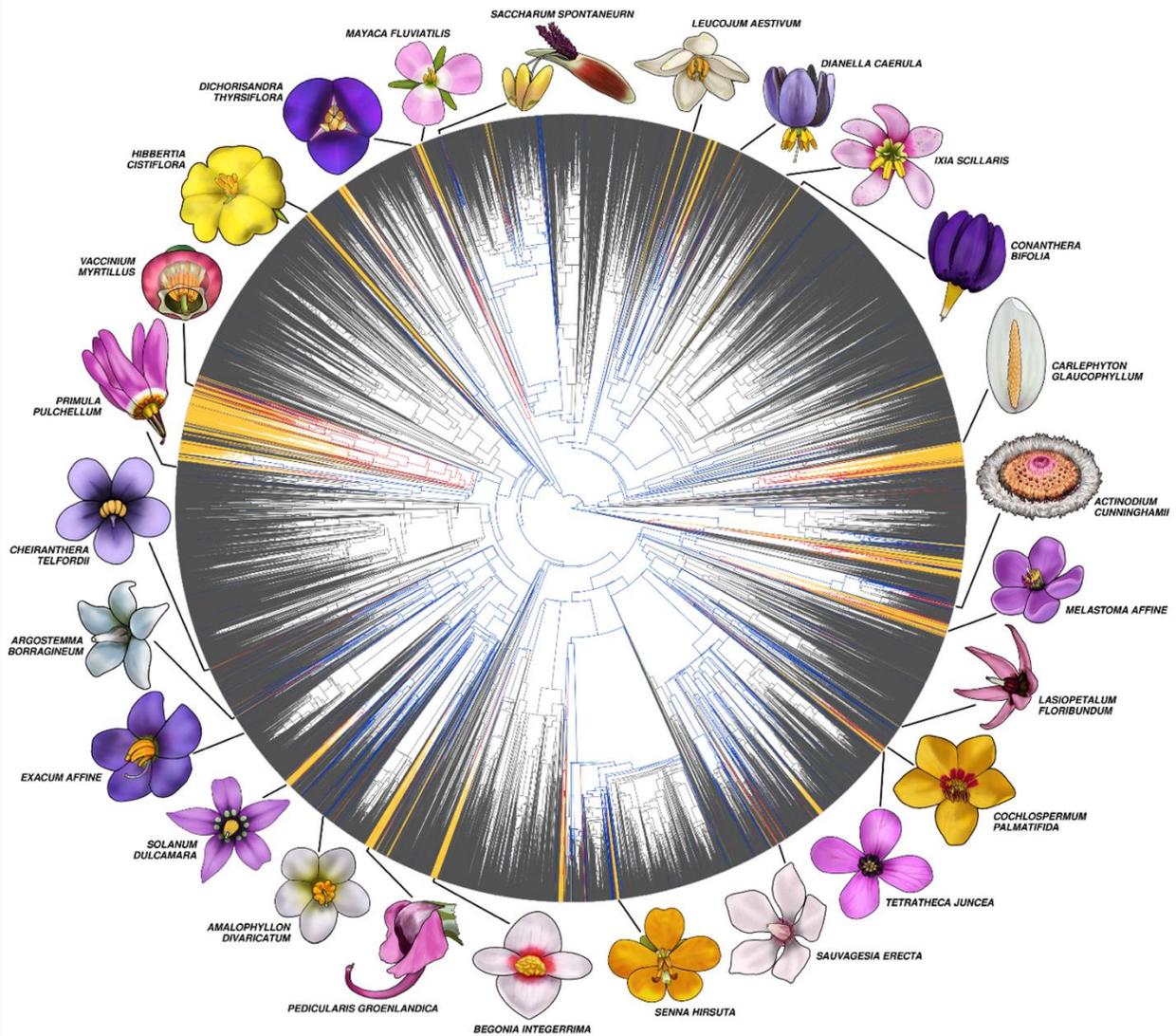


Todos los materiales del taller

<https://roszenil.github.io/mytutorials/contenido.html>



SCAN ME



Día 2

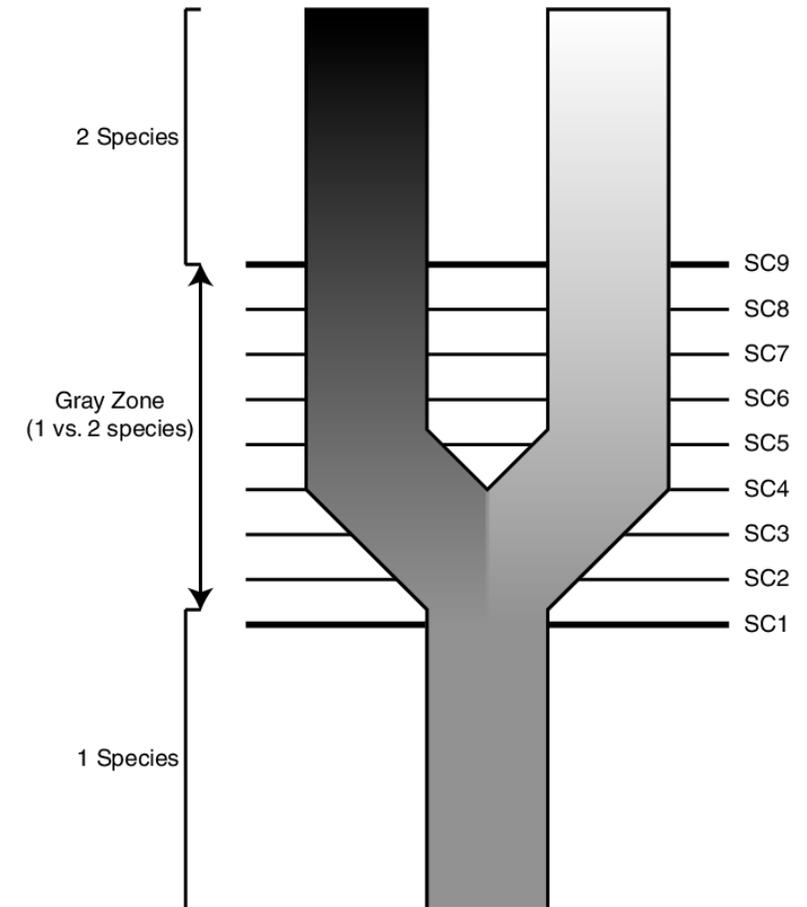
Introducción a Macroevolución

Inferencia Bayesiana con modelos de cadenas de Markov en tiempo continuo

Concepto moderno de especies

“All contemporary species concepts are variations on the general theme that species are segments of population-level lineages” (de Queiroz 2007)

Species concept	Property(ies)	Advocates/references
Biological	Interbreeding (natural reproduction resulting in viable and fertile offspring)	Wright (1940); Mayr (1942); Dobzhansky (1950)
Isolation	*Intrinsic reproductive isolation (absence of interbreeding between heterospecific organisms based on intrinsic properties, as opposed to extrinsic [geographic] barriers)	Mayr (1942); Dobzhansky (1970)
Recognition	*Shared specific mate recognition or fertilization system (mechanisms by which conspecific organisms, or their gametes, recognize one another for mating and fertilization)	Paterson (1985); Masters et al. (1987); Lambert and Spencer (1995)
Ecological	*Same niche or adaptive zone (all components of the environment with which conspecific organisms interact)	Van Valen (1976); Andersson (1990)
Evolutionary (some interpretations)	Unique evolutionary role, tendencies, and historical fate	Simpson (1951); Wiley (1978); Mayden (1997)
Cohesion	*Diagnosability (qualitative, fixed difference) Phenotypic cohesion (genetic or demographic exchangeability)	Grismer (1999, 2001) Templeton (1989, 1998a)
Phylogenetic Hennigian	Heterogeneous (see next four entries) Ancestor becomes extinct when lineage splits	(see next four entries) Hennig (1966); Ridley (1989); Meier and Willmann (2000)
Monophyletic	*Monophyly (consisting of an ancestor and all of its descendants; commonly inferred from possession of shared derived character states)	Rosen (1979); Donoghue (1985); Mishler (1985)
Genealogical	*Exclusive coalescence of alleles (all alleles of a given gene are descended from a common ancestral allele not shared with those of other species)	Baum and Shaw (1995); see also Avise and Ball (1990)
Diagnosable	*Diagnosability (qualitative, fixed difference)	Nelson and Platnick (1981); Cracraft (1983); Nixon and Wheeler (1990)
Phenetic	*Form a phenetic cluster (quantitative difference)	Michener (1970); Sokal and Crovello (1970); Sneath and Sokal (1973)
Genotypic cluster (definition)	*Form a genotypic cluster (deficits of genetic intermediates; e.g., heterozygotes)	Mallet (1995)



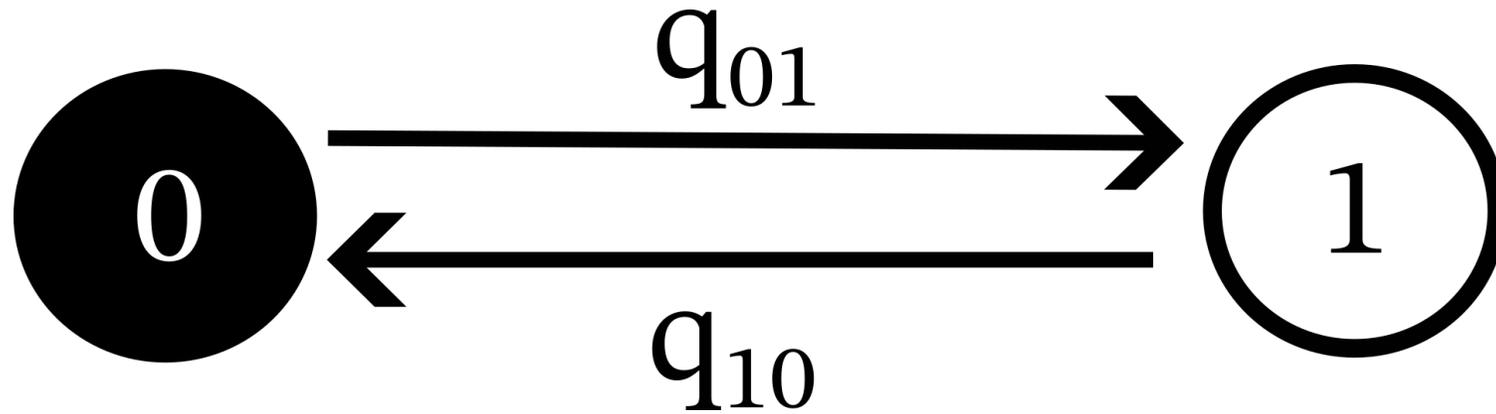
The work has been turning
out badly for me this
morning & I am sick
at heart & so by
God how I do
hate species & varieties.

The work has been turning out badly for me this morning & I am sick at heart & oh my God how I do hate species & varieties."

La matriz P- Las **probabilidades**

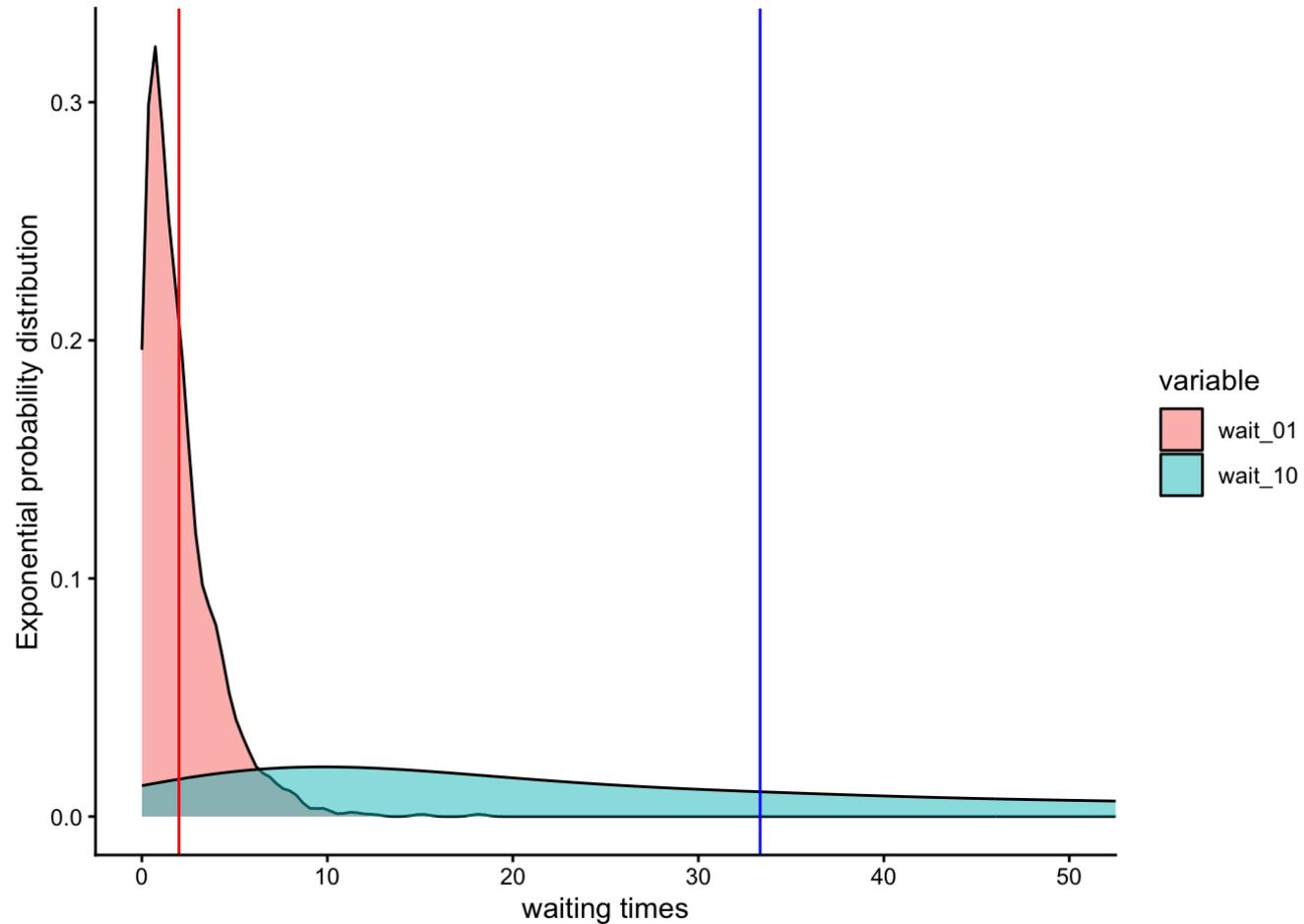
$$P(t) = \begin{pmatrix} p_{00}(t) & p_{01}(t) \\ p_{10}(t) & p_{11}(t) \end{pmatrix}$$

Tasas de Transición- Definición

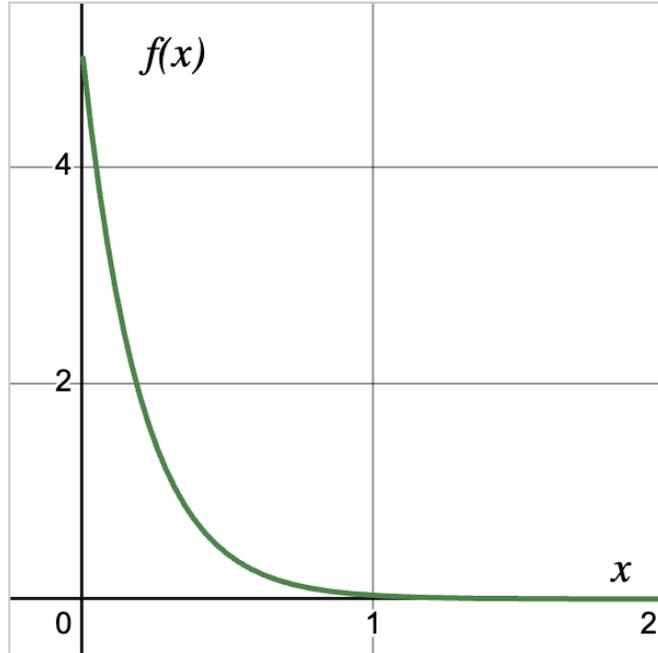


$$\lim_{h \rightarrow 0} P(X(t+h) = j | X(t) = i) = \lim_{h \rightarrow 0} p_{ij}(h) = q_{ij}$$

$1/(\text{tasa de transición}) = \text{Tiempo promedio de espera}$



La exponencial es una distribución de tiempos de espera

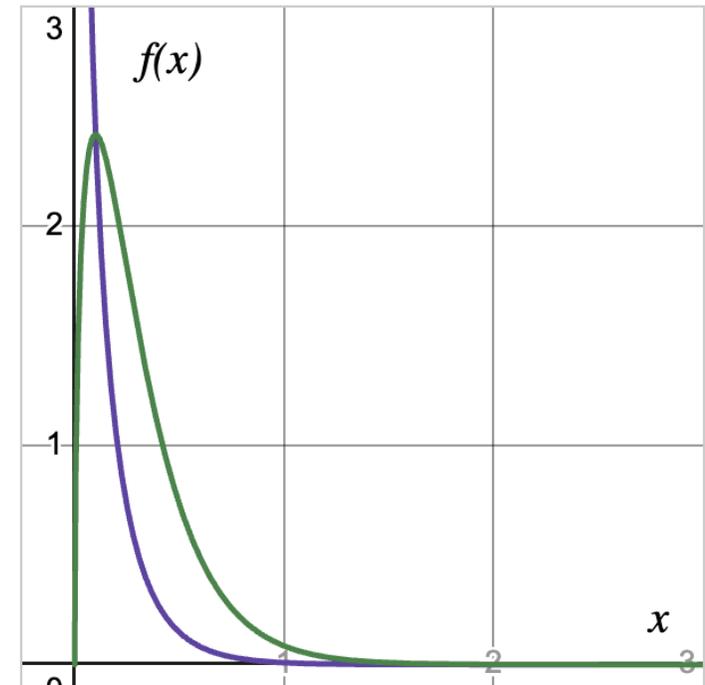


exponential distribution with parameter $\lambda > 0$, write $X \sim \text{exponential}(\lambda)$,

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{for } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

gamma distribution with parameters $\alpha, \lambda > 0$, write $X \sim \text{gamma}(\alpha, \lambda)$

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & \text{for } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$



La Q-matriz

$$Q = \begin{pmatrix} -q_{01} & q_{01} \\ q_{10} & -q_{10} \end{pmatrix}$$

¿Cómo se relacionan la P y la Q?

$$P(t) = \frac{1}{q_{01} + q_{10}} \begin{pmatrix} q_{10} + q_{01}e^{-(q_{01}+q_{10})t} & q_{01} - q_{01}e^{-(q_{01}+q_{10})t} \\ q_{10} - q_{10}e^{-(q_{01}+q_{10})t} & q_{01} + q_{10}e^{-(q_{01}+q_{10})t} \end{pmatrix}$$

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

SCIENTIFIC PAPER

The image shows a page of handwritten mathematical work on a grid background. On the left side, there are several columns of equations and terms, including $\lambda_1, \lambda_2, \dots, \lambda_n$ and $\mu_1, \mu_2, \dots, \mu_n$. Some terms are circled in red. The right side of the page is dominated by a large matrix of zeros, with some rows and columns highlighted or boxed. The matrix appears to be a sparse matrix with a diagonal structure. The overall appearance is that of a scientific or mathematical manuscript.

Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later*

Cleve Moler[†]
Charles Van Loan[‡]

Abstract. In principle, the exponential of a matrix could be computed in many ways. Methods involving approximation theory, differential equations, the matrix eigenvalues, and the matrix characteristic polynomial have been proposed. In practice, consideration of computational stability and efficiency indicates that some of the methods are preferable to others, but that none are completely satisfactory. Most of this paper was originally published in 1978. An update, with a separate bibliography, describes a few recent developments.

Key words. matrix, exponential, roundoff error, truncation error, condition

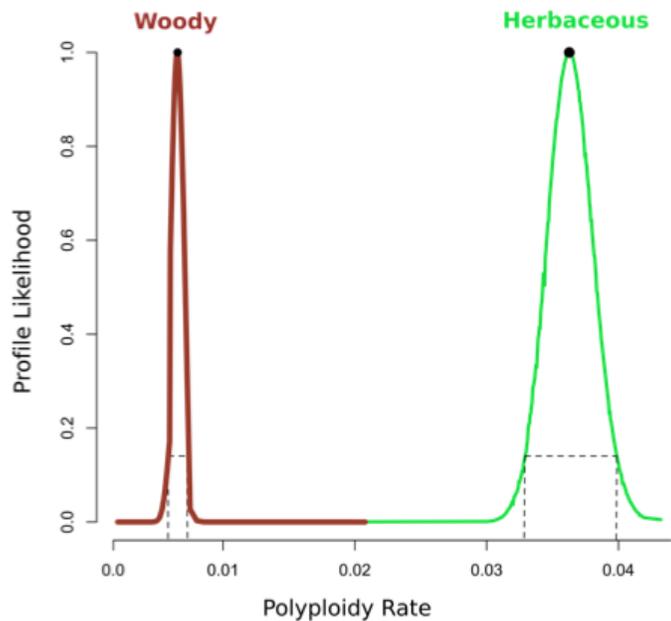
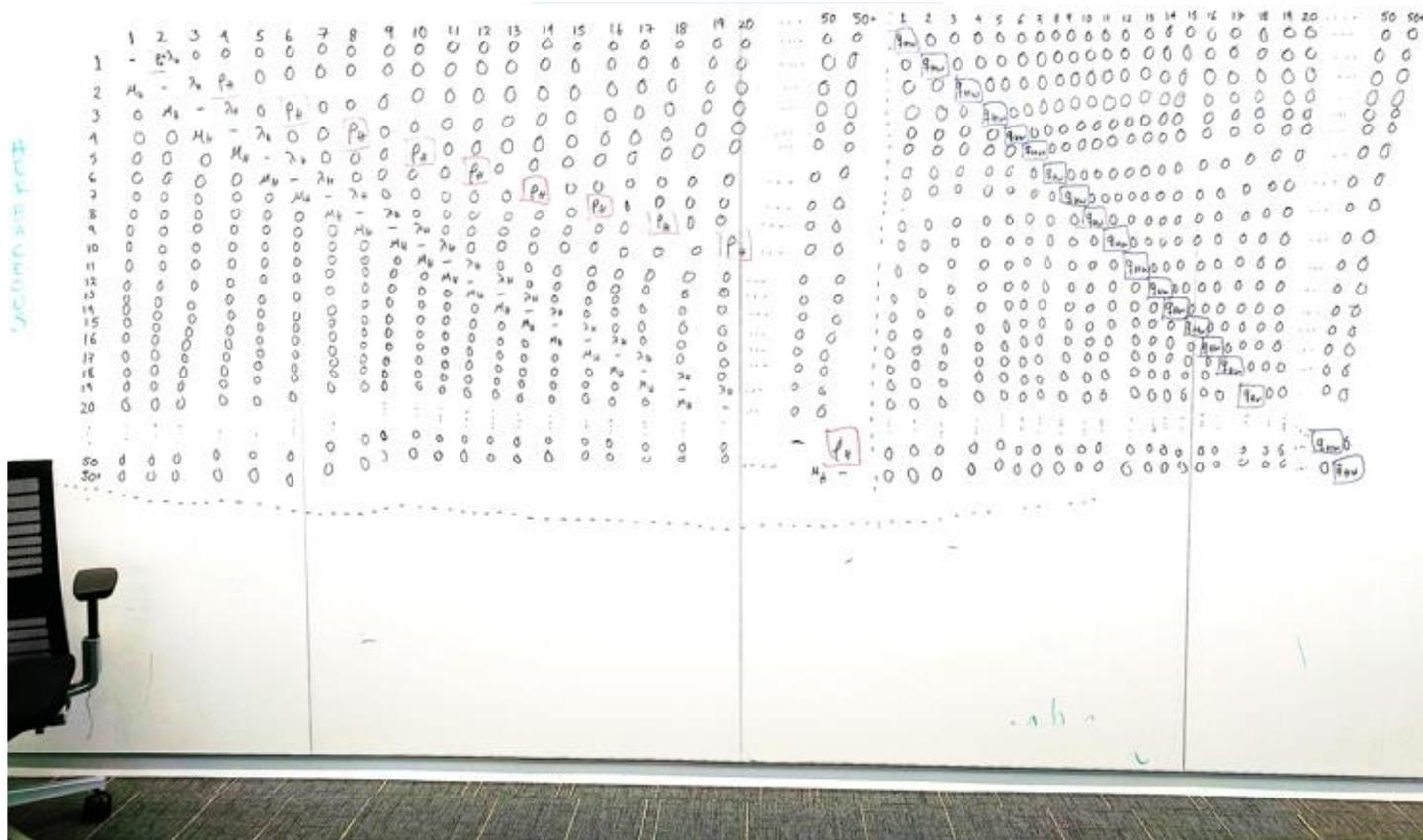
AMS subject classifications. 15A15, 65F15, 65F30, 65L99

PII. S0036144502418010

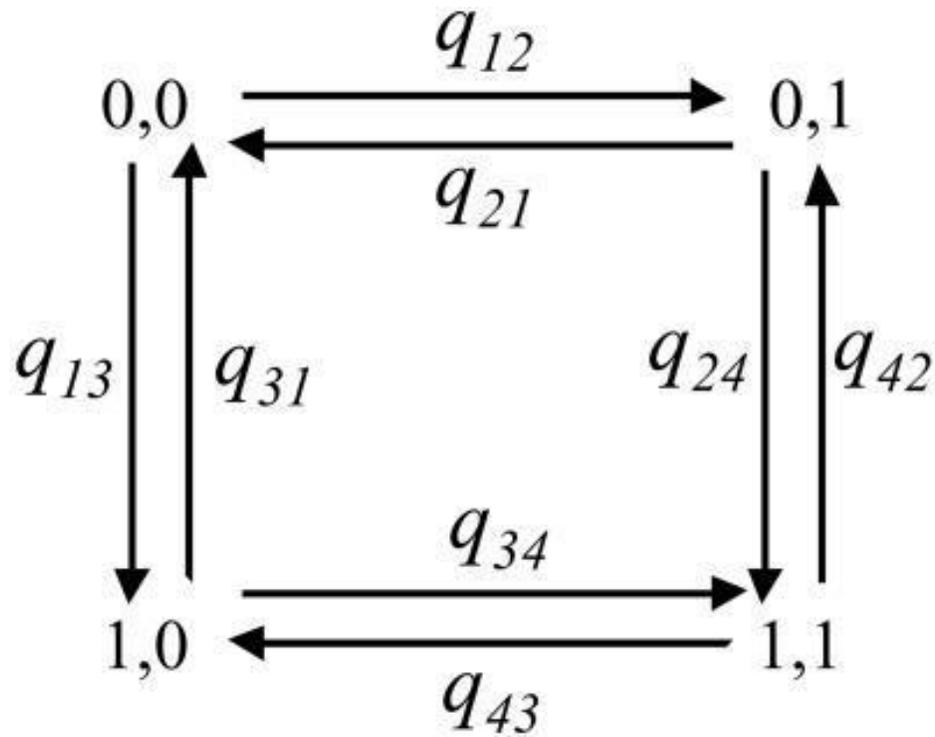
$$Q_{ij} = \begin{cases} \gamma_a & j = i + 1, \\ \delta_a & j = i - 1, \\ \rho_a & j = 2i, \\ \eta_a & j = 1.5i, \\ 0 & \text{otherwise,} \end{cases}$$

ChromEvol (Mayrose et al. 2010)

BiChrom (Zenil-Ferguson et al. 2017)



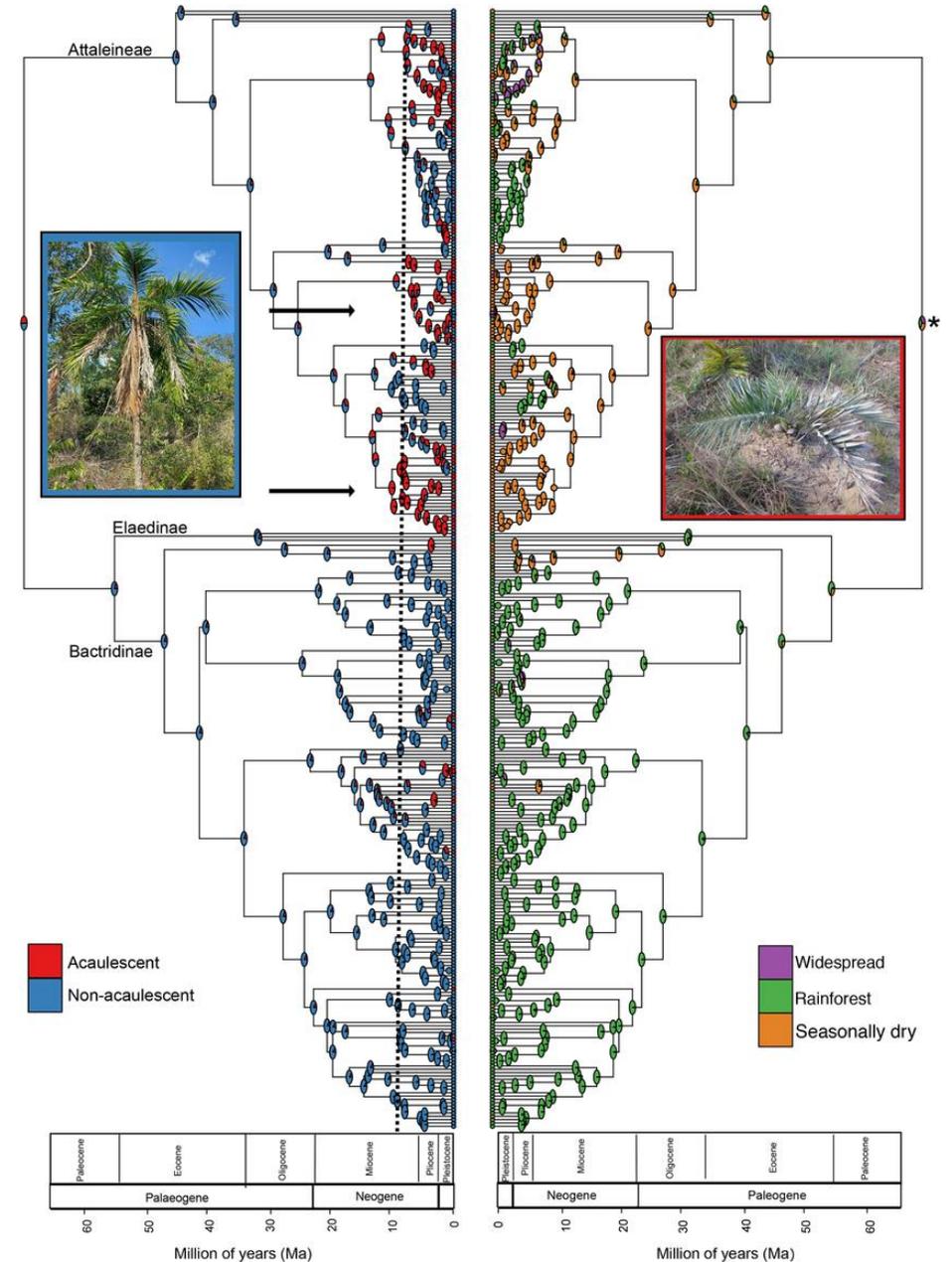
Las matrices Q son el centro del universo
(para caracteres discretos)



El test de correlación de Pagel

Pagel 1994

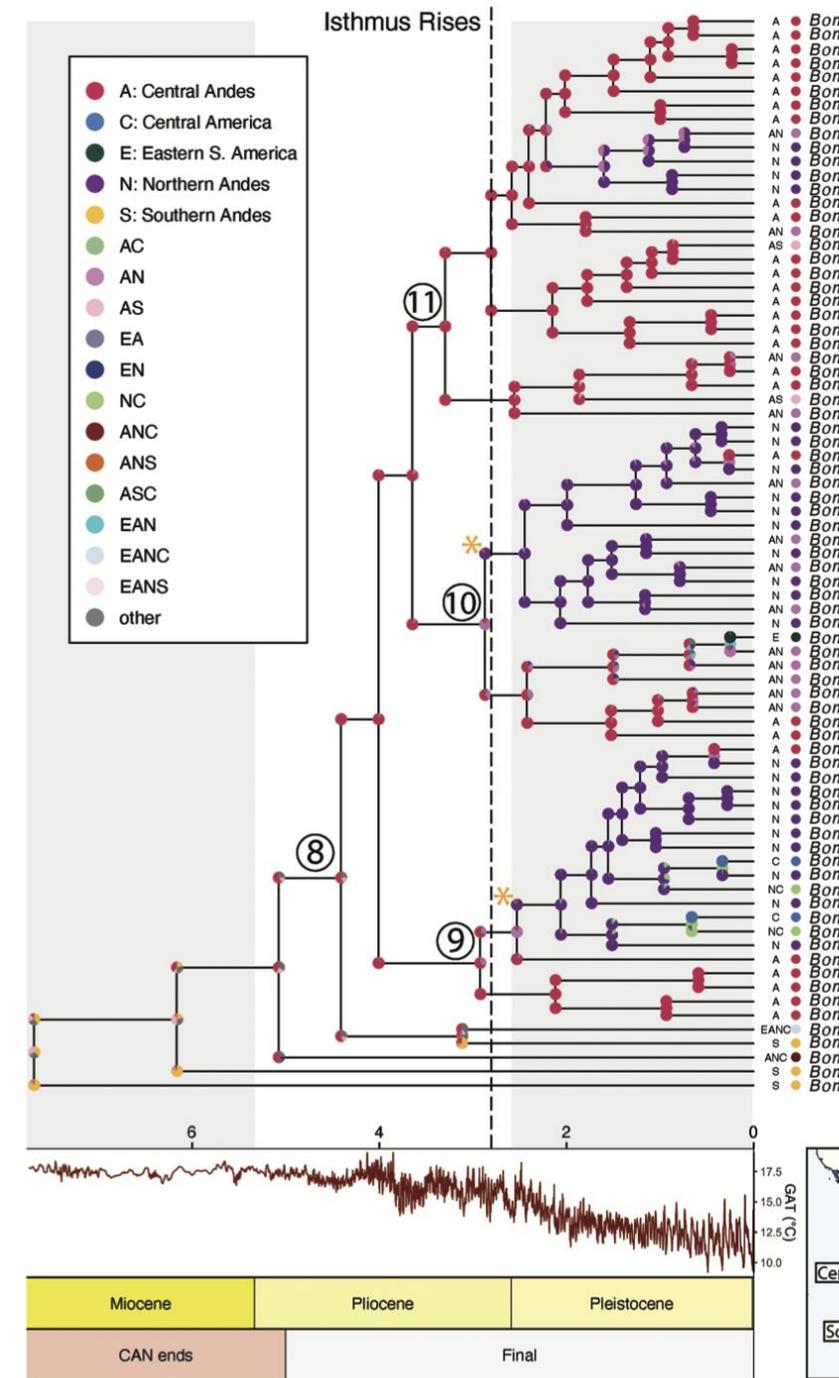
Acaulescencia en habitat seco
Cassia Silva et al. *Ecography*. 2022



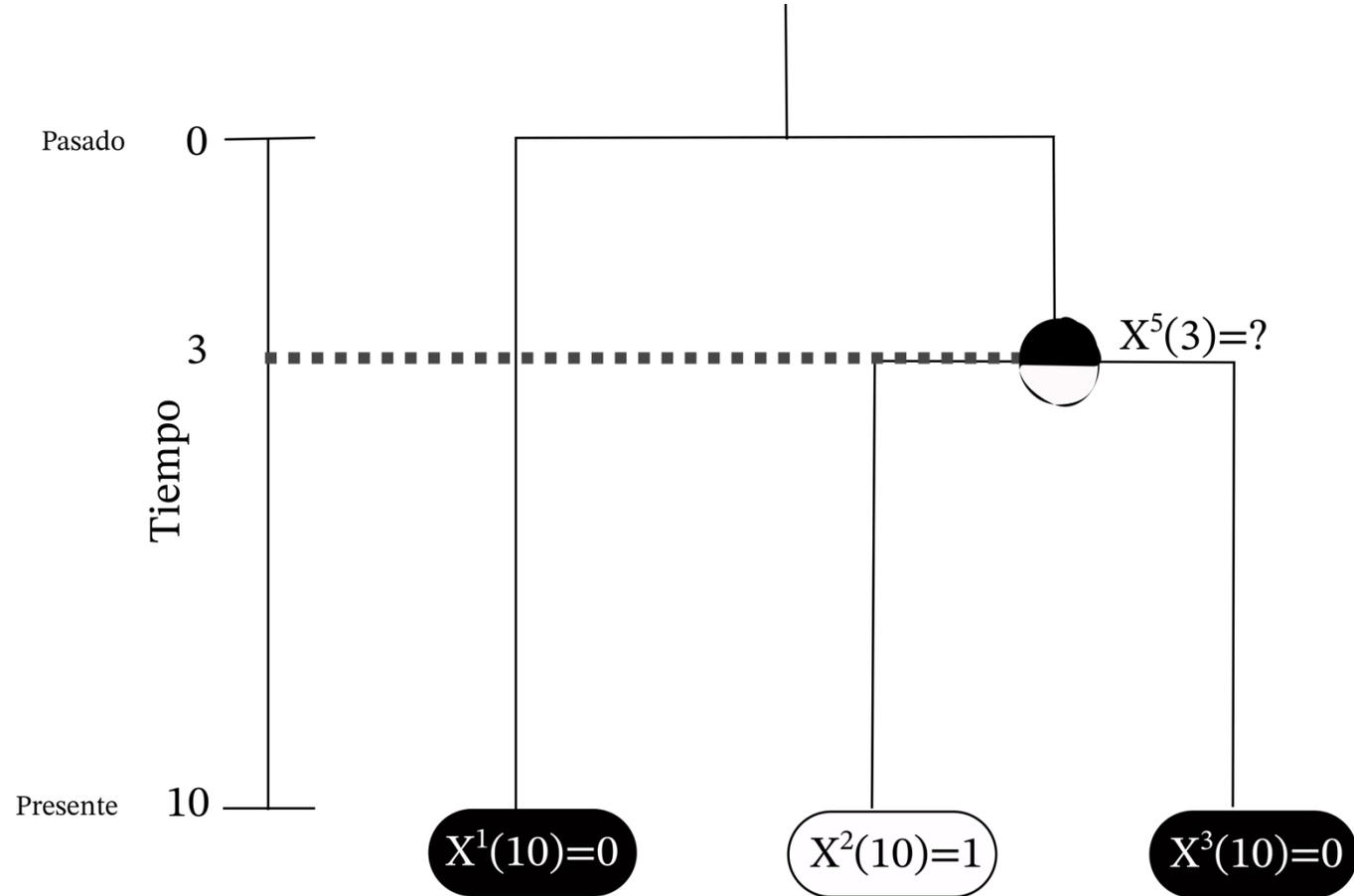
$$Q = \begin{array}{c|cccccccc} & \emptyset & A & B & C & AB & AC & BC & ABC \\ \hline \emptyset & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A & e_A & - & 0 & 0 & d_{AB} & d_{AC} & 0 & 0 \\ B & e_B & 0 & - & 0 & d_{BA} & 0 & d_{BC} & 0 \\ C & e_C & 0 & 0 & - & 0 & d_{CA} & d_{CB} & 0 \\ AB & 0 & e_B & e_A & 0 & - & 0 & 0 & d_{AC} + d_{BC} \\ AC & 0 & e_C & 0 & e_A & 0 & - & 0 & d_{AB} + d_{CB} \\ BC & 0 & 0 & e_C & e_B & 0 & 0 & - & d_{BA} + d_{CA} \\ ABC & 0 & 0 & 0 & 0 & e_C & e_B & e_A & - \end{array}$$

Dispersal-Extinction Cladogenesis model (DEC)

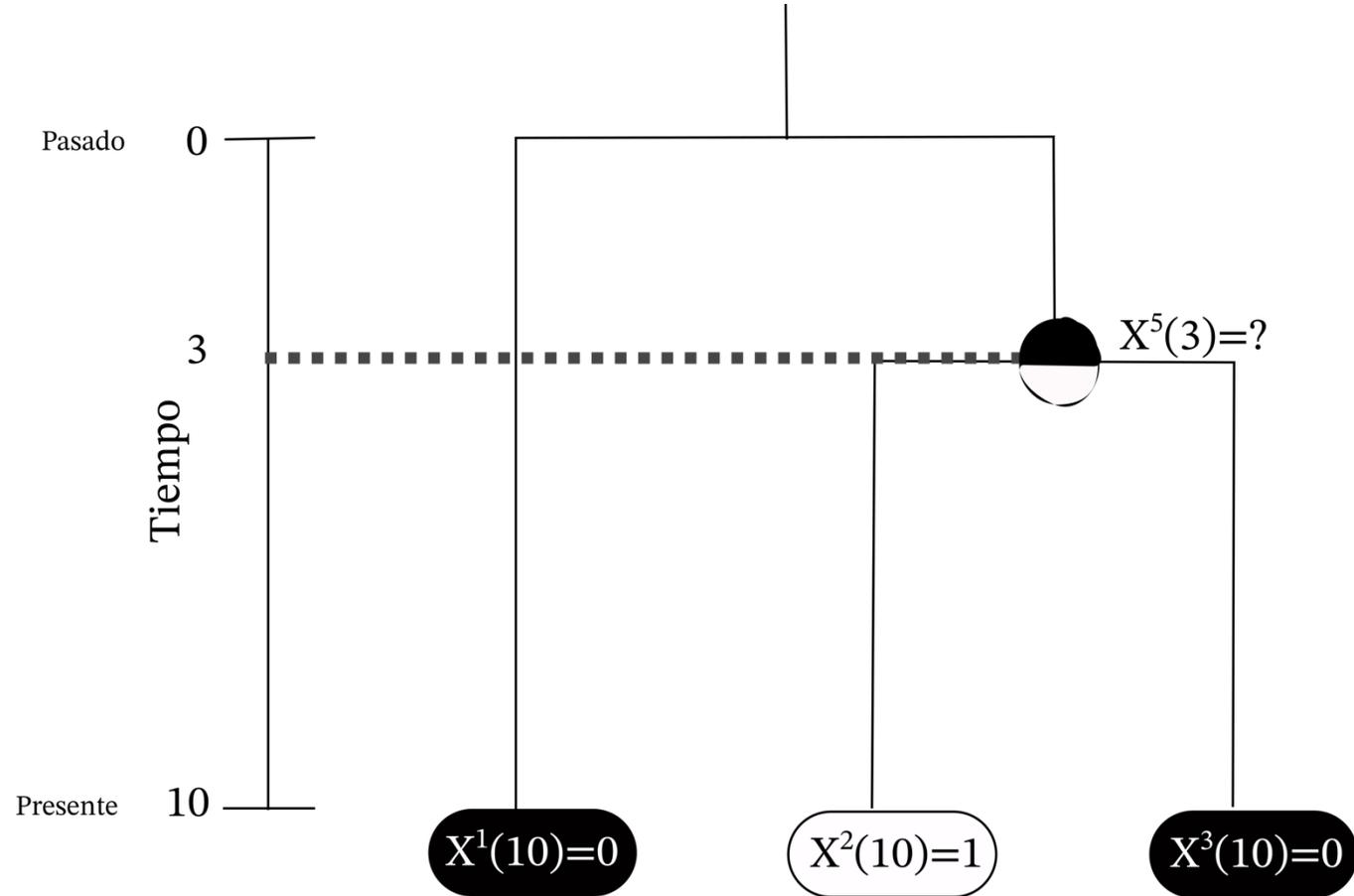
Ree et al. 2005; Ree and Smith (2008)



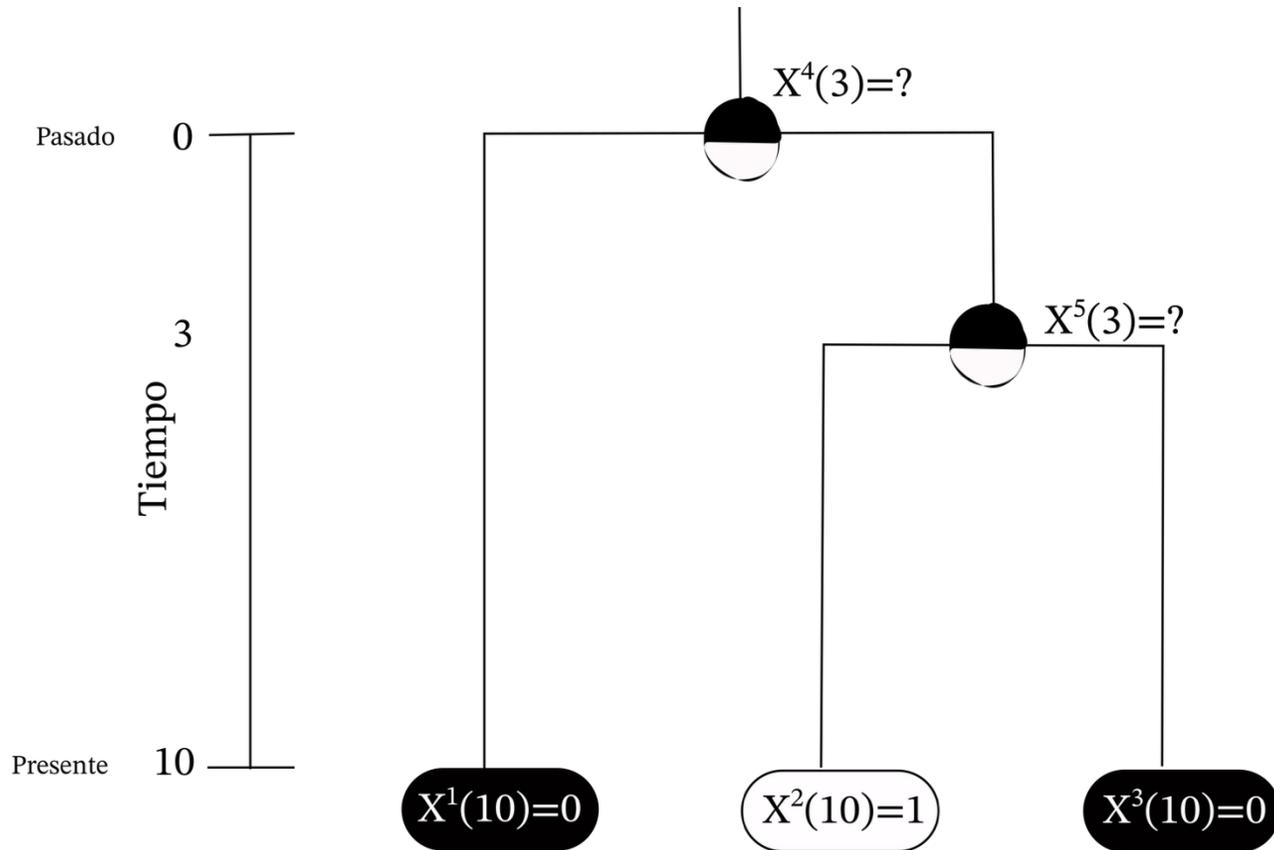
Gran problema: No sabemos nada del pasado



Gran problema: No sabemos nada del pasado



Gran problema: No sabemos nada del pasado



La función de verosimilitud para un modelo Mk2

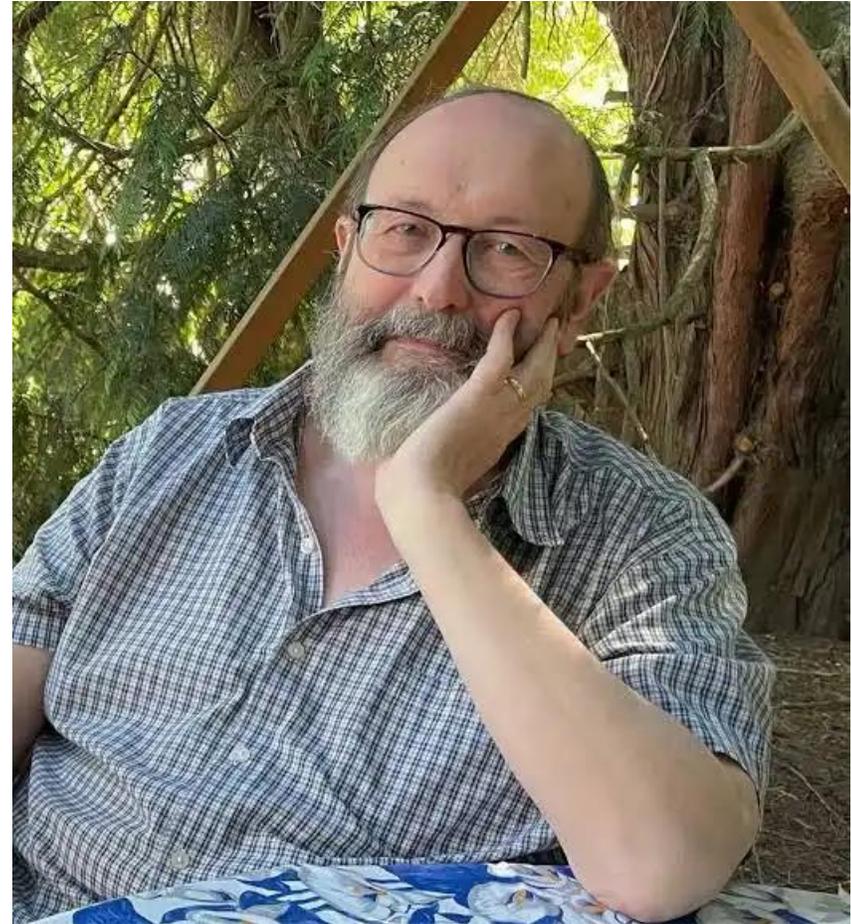
Evolutionary Trees From Gene Frequencies and Quantitative Characters: Finding Maximum Likelihood Estimates

Author(s): Joseph Felsenstein

Source: *Evolution*, Nov., 1981, Vol. 35, No. 6 (Nov., 1981), pp. 1229-1242

Published by: Society for the Study of Evolution

Stable URL: <https://www.jstor.org/stable/2408134>



Joseph Felsenstein

Algoritmo de poda (pruning algorithm-Felsenstein 1981)

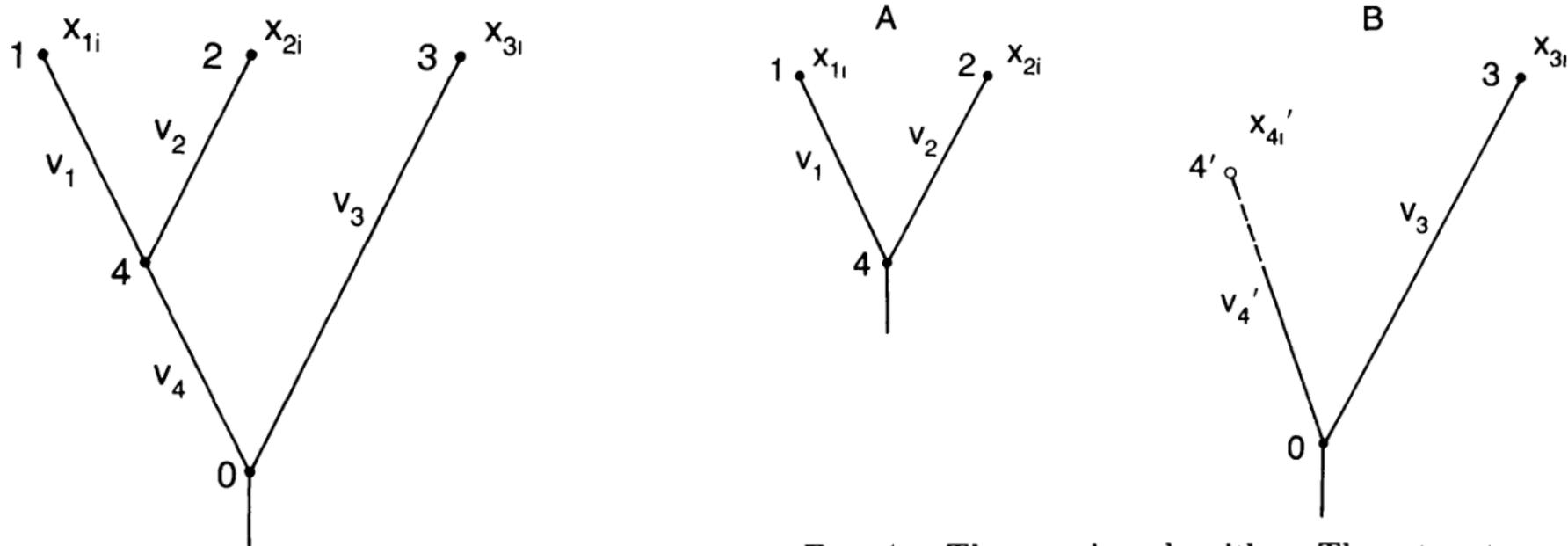


FIG. 5. The pruning algorithm. These two trees, taken together, have the same restricted likelihood as the tree of Figure 3. x_{4i}' and v_4' are defined in the text.

Cumberland Falls State Park, KY



Thalictrum pubescences



Thalictrum
Múltiples cambios de
tipo de polinización

Verónica di Stilio
University of Washington



Thalictrum thalictroides

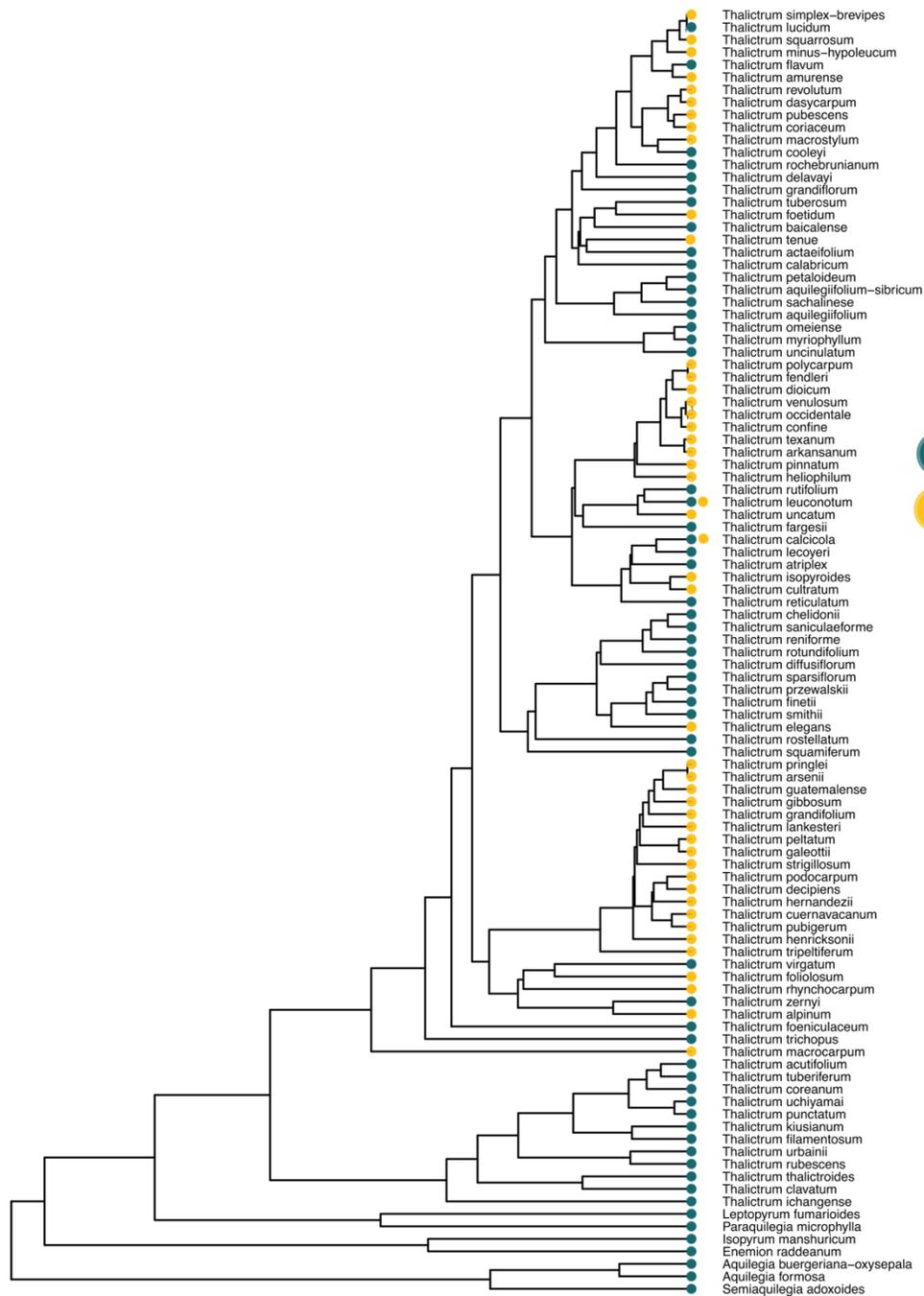
Tipos de polinización

Insecto= 0



Viento =1





● Insect
● Wind

- Siempre grafiquen sus datos!
- Checar anidaciones en clados!
- Las transiciones es lo que les da poder para estimar los modelos que les queremos presentar.

No pueden usar estos modelos si están en el "Felsenstein worst case scenario"



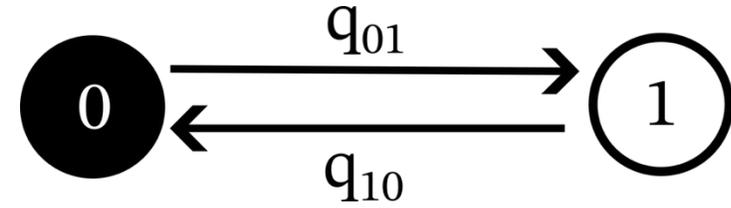
Maddison and Fitzjohn, 2014



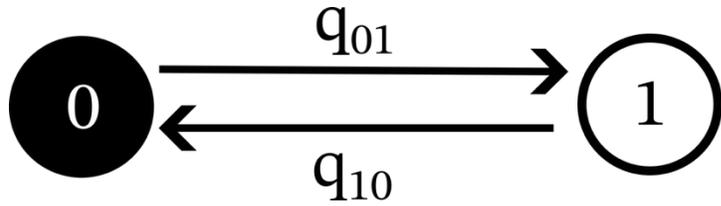
Uyeda, Zenil-Ferguson, and Pennell, 2018

Parámetros desconocidos e inciertos del Mk2

1. Tasas de transición

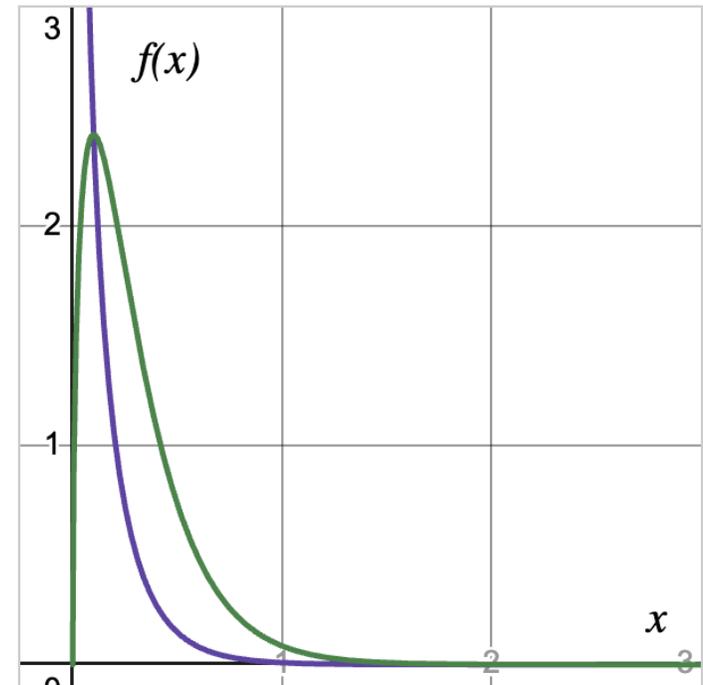


Distribución Gamma para las tasas



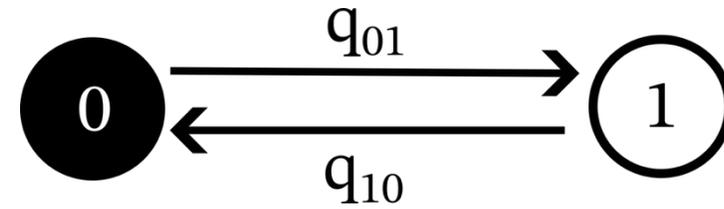
gamma distribution with parameters $\alpha, \lambda > 0$, write $X \sim \text{gamma}(\alpha, \lambda)$

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & \text{for } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

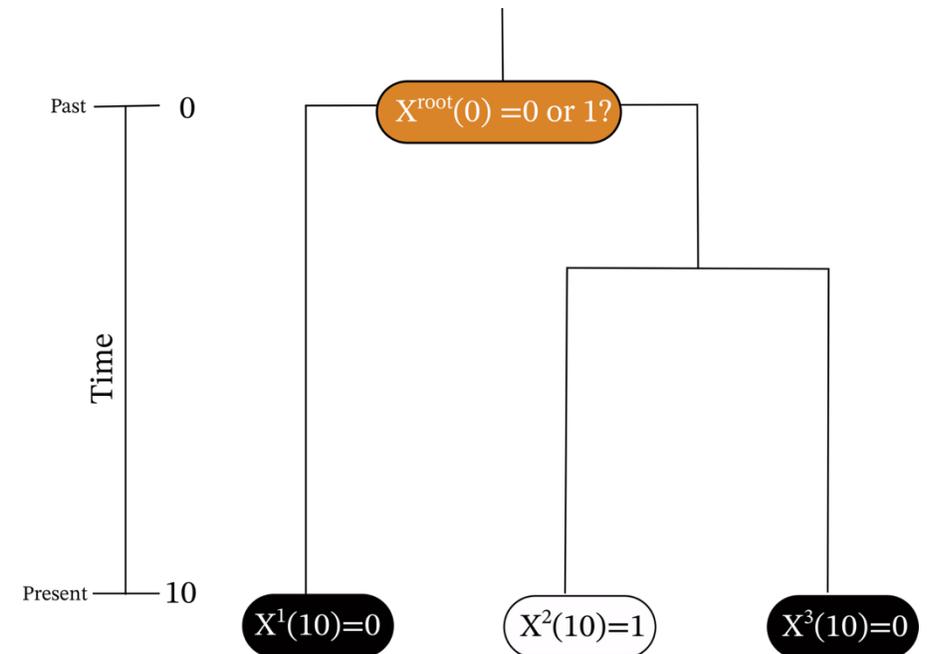


Parámetros desconocidos e inciertos del Mk2

1. Tasas de transición

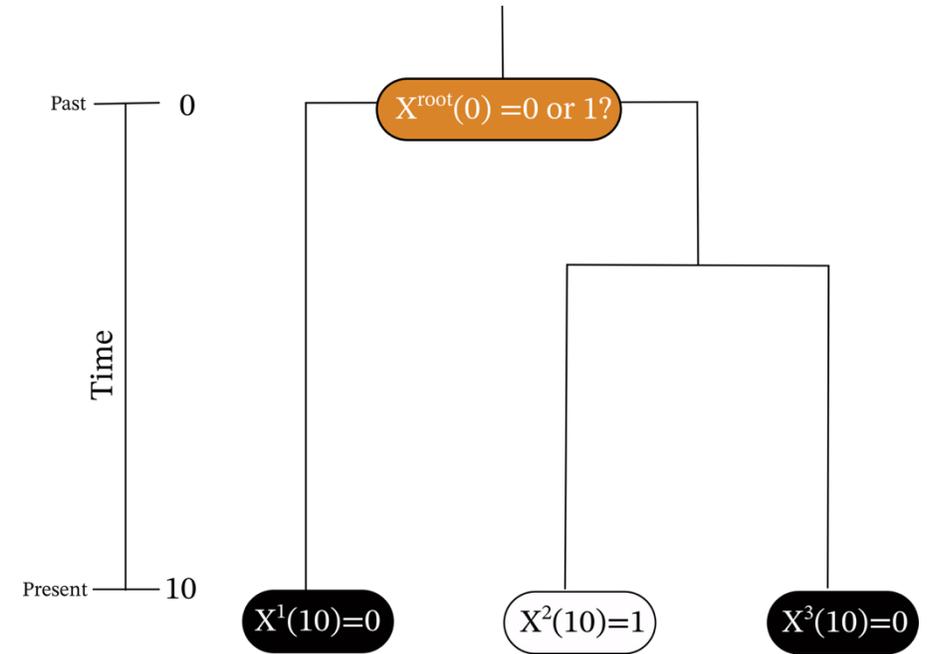
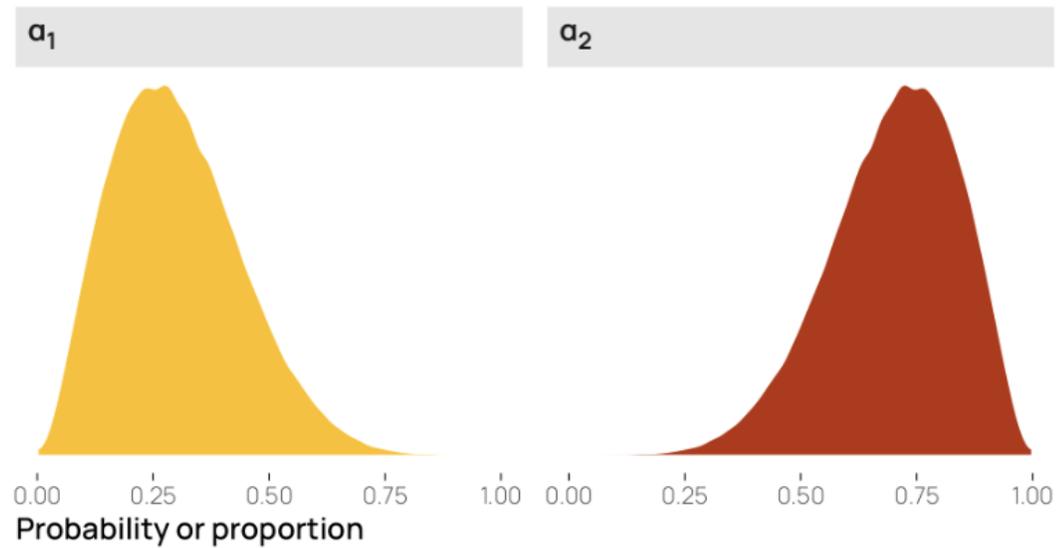


2. Las frecuencias de los estados en la raíz.



Frecuencias de la raíz

Dirichlet(3, 7)



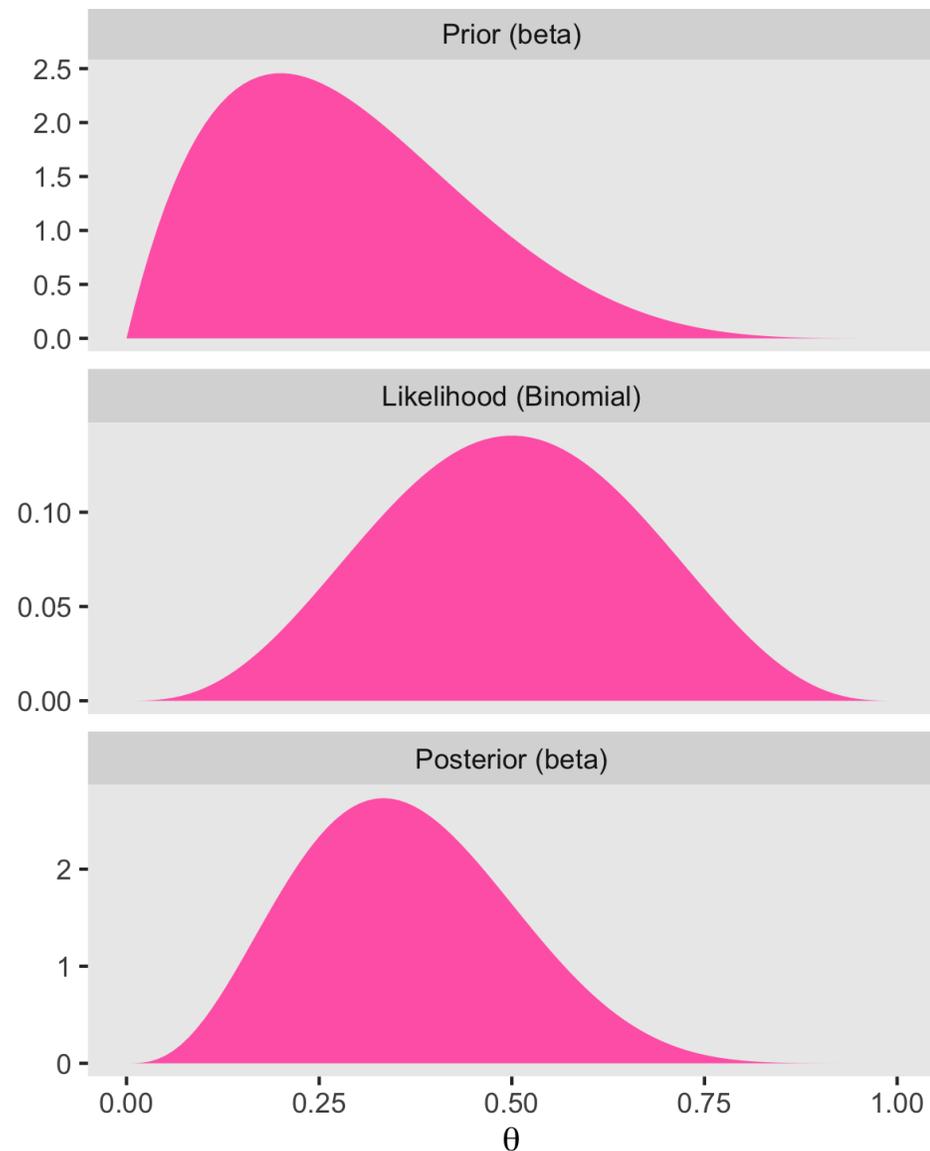
Inferencia Bayesiana

- Objetivo: Obtener la **distribución posterior** de los parámetros

Necesitamos

1. A priori para cada uno de los parámetros
2. Calcular la verosimilitud

Problema: La verosimilitud es difícil de calcular en el árbol filogenético



Metropolis-Hastings – Monte Carlo Markov Chain

1. Empieza con un valor para θ llamado θ_0
2. Haz $\theta_{viejo} = \theta_0$
3. Calcula la distribución posterior $P(\theta_{viejo} | Datos)$
4. Proponer una distribución aleatoria $g(\theta)$ para obtener un nuevo valor θ_{nuevo}
5. Calcula la distribución posterior $P(\theta_{nuevo} | Datos)$
6. Calcula los momios $momios = \frac{P(\theta_{nuevo} | Datos)}{P(\theta_{old} | Datos)}$
7. Calcula un valor aleatorio u entre 0 y 1
8. Si $u < momios$ entonces acepta. θ_{nuevo} , guárdalo, sino rechaza y no lo guardes.
9. Si lo aceptas haz $\theta_{viejo} = \theta_{nuevo}$ y vuelve al paso dos hasta acabar las iteraciones. Sino continua al paso dos con el mismo θ_{viejo} hasta acabar las iteraciones.

Movimientos (moves)- Propuestas

Propuestas para mover de valor los parámetros (MCMC)

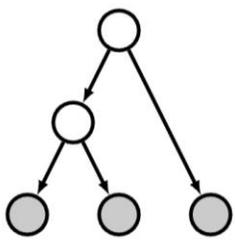
Se propone un valor de parámetro y se evalúa su probabilidad. Si es muy bueno, con alta probabilidad de ser aceptado. De lo contrario, se rechaza y se mantiene el valor anterior.



Ver la clase del Dr. Paul Lewis
Bayesian inference and MCMC

Importante

- El objetivo de la inferencia bayesiana es la **distribución posterior** no un estimador puntual
- Si la verosimilitud es difícil de calcular eso implica que la distribución posterior también. El MCMC es un algoritmo que nos permite **muestrear la posterior**
- Las muestras se pueden transformar para obtener muestras de las variables aleatorias transformadas



RevBayes

Bayesian phylogenetic inference using probabilistic graphical models and an interpreted language

<https://revbayes.github.io/tutorials/>

RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language

Sebastian Höhna , Michael J. Landis, Tracy A. Heath, Bastien Boussau, Nicolas Lartillot, Brian R. Moore, John P. Huelsenbeck, Fredrik Ronquist [Author Notes](#)

The screenshot shows the GitHub repository page for RevBayes. At the top, the repository name 'revbayes' is displayed as 'Public'. To the right, there are buttons for 'Watch 31', 'Fork 27', and 'Star 80'. Below this, the repository structure is shown with a search bar and 'Add file' and 'Code' buttons. A recent pull request by 'davidcerny' is highlighted, showing a merge of pull request #918 from the development branch. Below the pull request, a list of folders and their latest commit messages and dates is provided:

Folder	Commit Message	Time Ago
.github	Use macos-latest in release.yml	3 months ago
doc	Add initial release-workflow.md	4 years ago
doxygen	updated Doxyfile, removed obsolete tags	7 years ago
help	Merge branch 'development' into stochastic-only-vectors	3 months ago
projects	Update README.md for meson project	5 months ago
src	Update version numbers to v1.3.2	3 months ago
tests	Repurpose an existing test to further test stochasticOnly=...	3 months ago
validation	Update Validation_Poisson.Rev	last year

On the right side, the 'About' section provides details about the project: 'Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language', the website 'revbayes.com', and various icons for README, license (GPL-3.0), code of conduct, activity, custom properties, stars (80), watchers (31), and forks (27).

RevBayes

Bayesian Inference of Evolutionary Parameters

RevBayes core development activity
Dec 2019 to Feb 2026 (dev. branch)

The Code

- 390,000+ lines of code
- 3,400+ source files
- 16,000+ commits
- 19,500+ user downloads

The Developers

- 50+ project contributors
- 12+ cities in the US and Europe



Core Developers: Sebastian Höhna (lead), Tracy Heath, Michael Landis, Jeremy Brown, Benjamin Redelings, John Huelsenbeck (creator), Frederik Ronquist (creator)

Contributors & developers: J. Andréoletti, J. Barido-Sottani, R. Billenstein, R. Borges, B. Boussau, S. Brock, R. Castillo, D. Černý, W. Dismukes, W. Freyman, J. Gao, A. Hsiang, J. Justison, B. Kopperud, N. Lartillot, A. Magee, M. Manceau, M. May, B. Moore, B. Petrucci, K. Smith, M. Smith, N. Tiwari, C. Tribble, J. Walker, R. Warnock, A. Wright, R. Zenil-Ferguson, and many others...



IMPLEMENTACIÓN DE
LA ESTADÍSTICA BAYESIANA
EN MACROEVOLUCIÓN



MODELOS GRÁFICOS
PROBABILÍSTICOS

Notación de RevBayes

cada símbolo significa algo importante

	a) Constant node		\leftarrow constant variable
	b) Stochastic node		\sim stochastic variable
	c) Deterministic node		$:=$ deterministic variable
	d) Clamped node (observed)		node.clamp(data) observed value of stochastic variable
	e) Plate		for (i in 1:N) { ... } plates (multiple connected nodes)
			$=$ for fixed operations (not part of graphical model)

Gráfico probabilístico

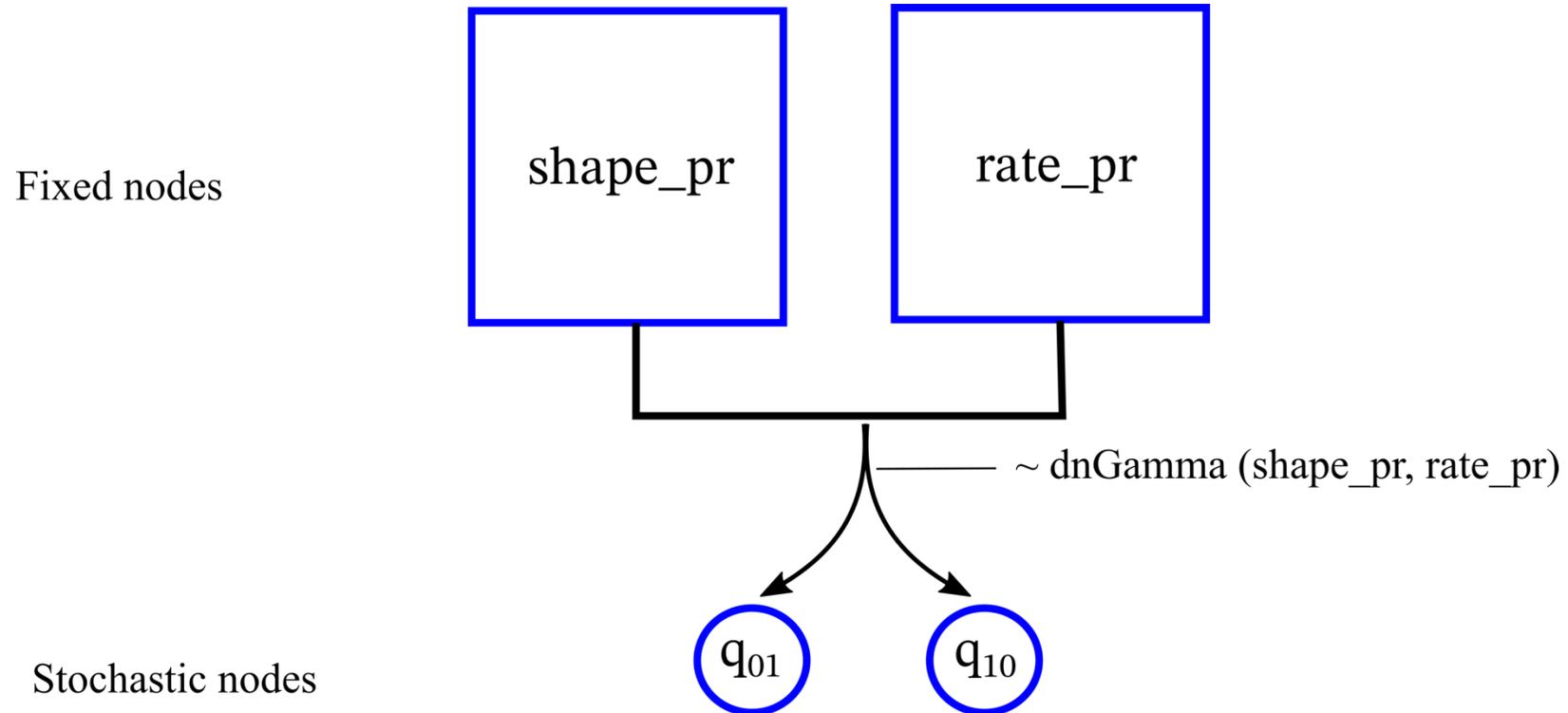
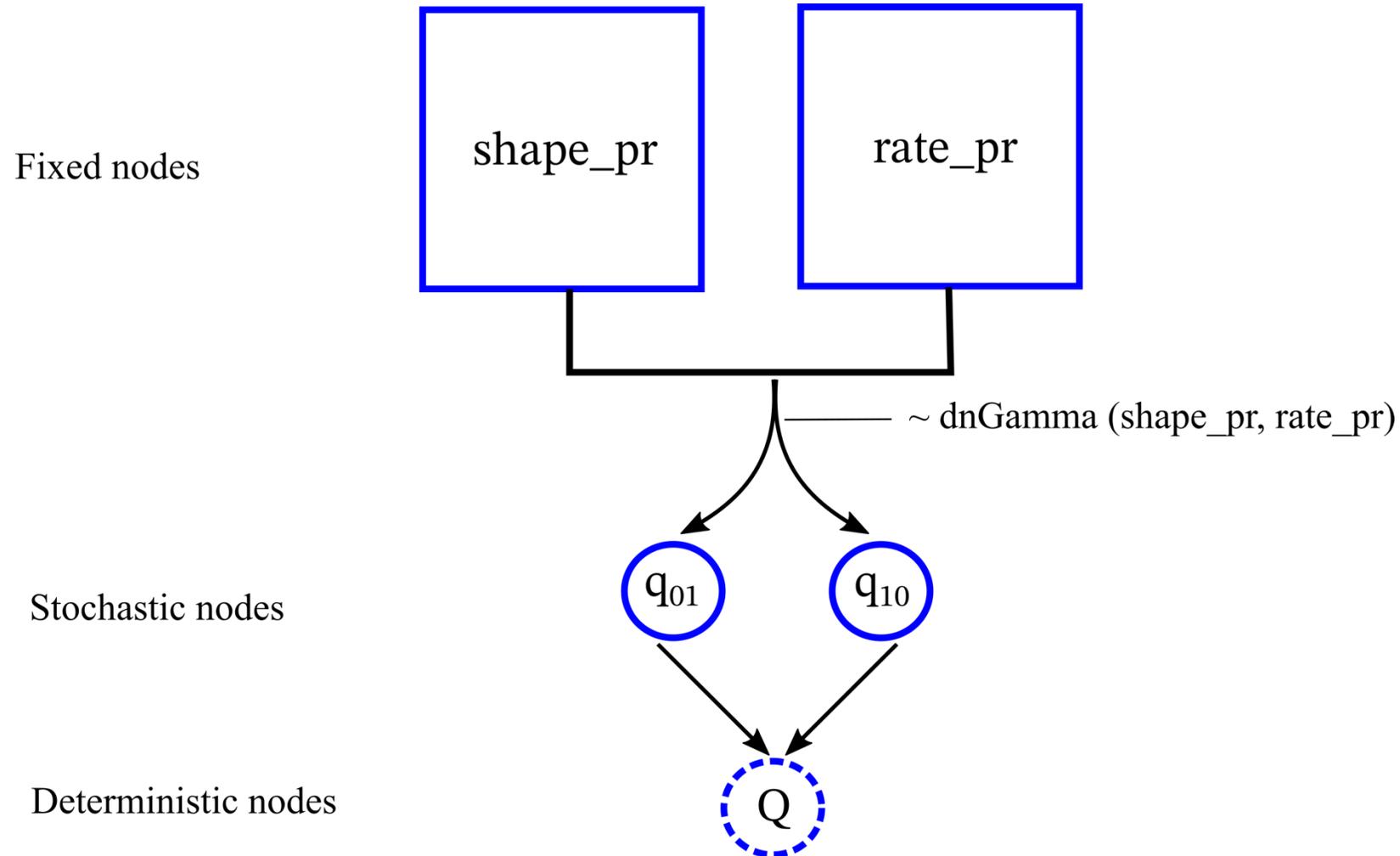
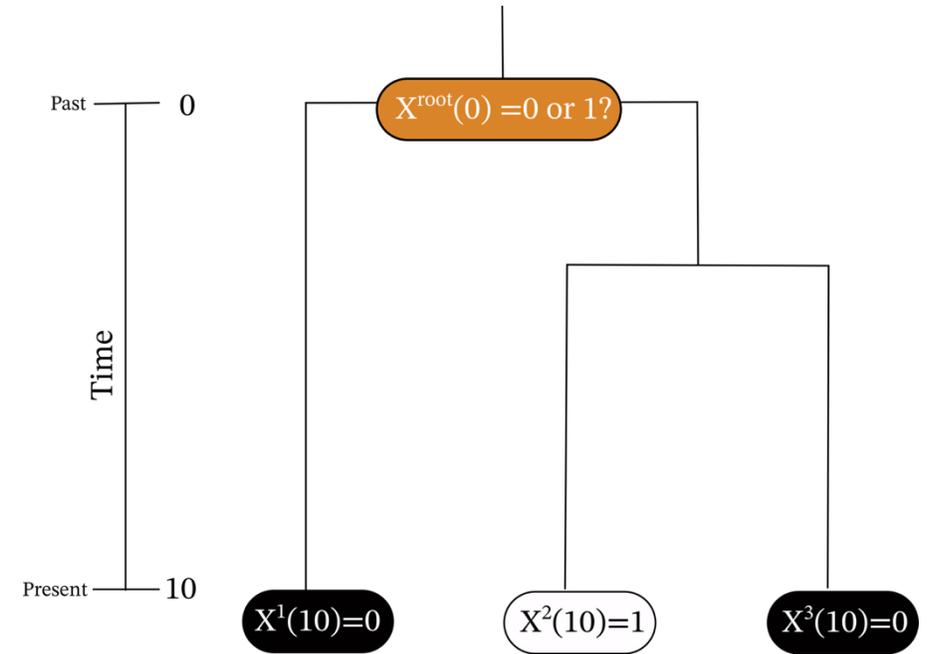
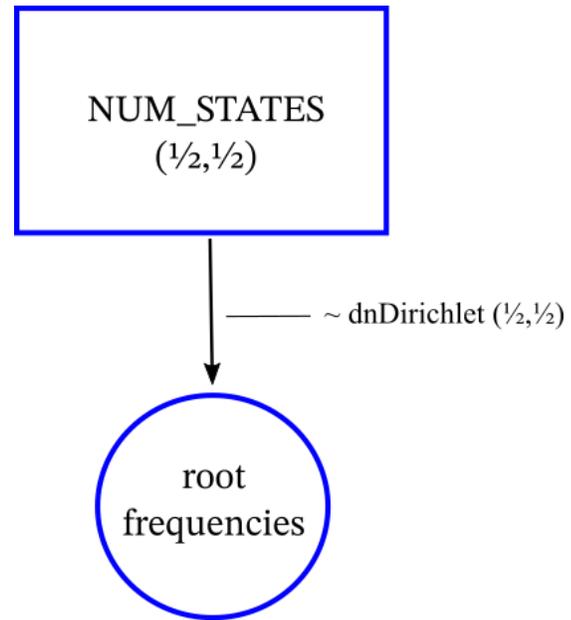


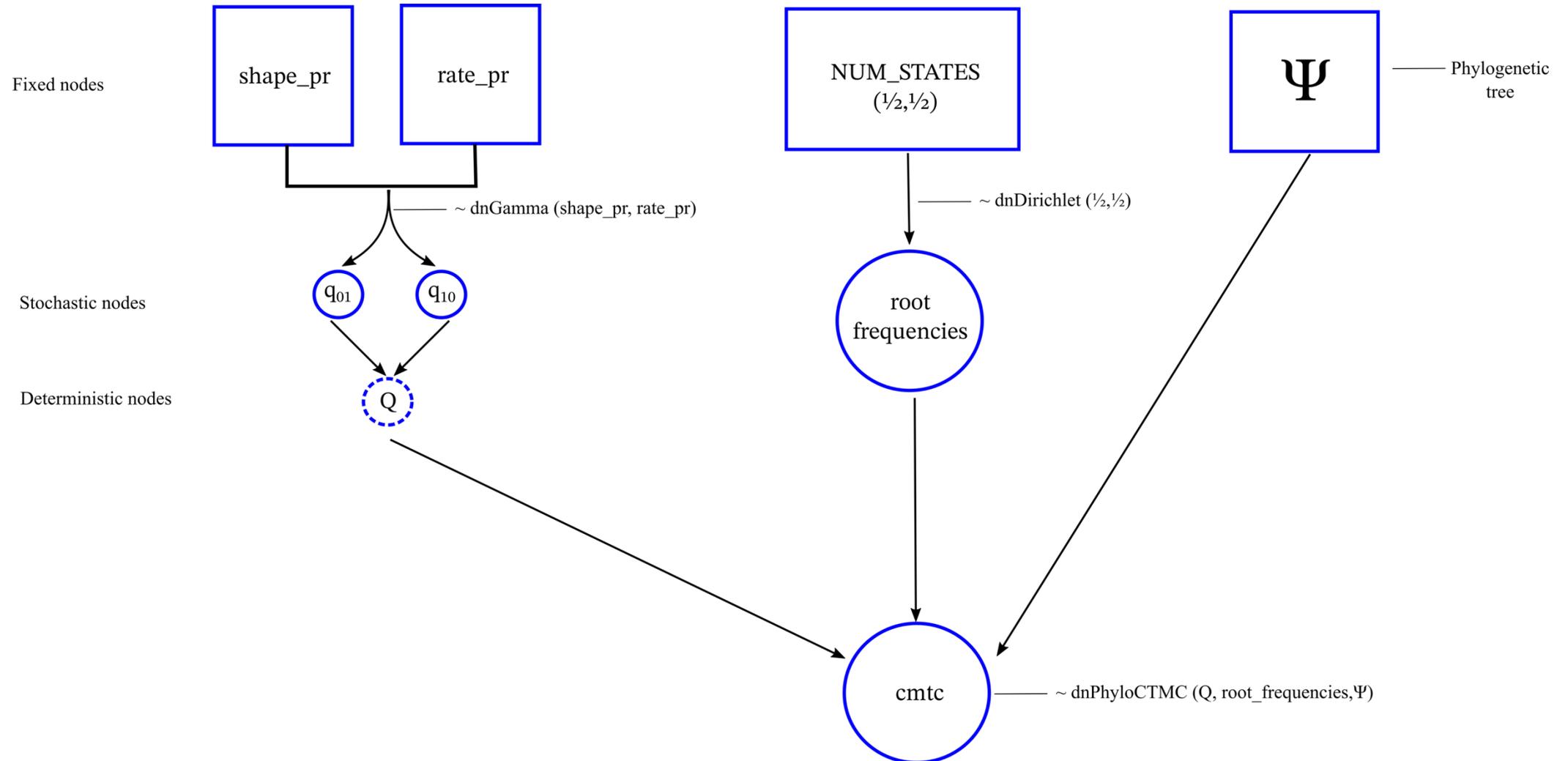
Gráfico probabilístico



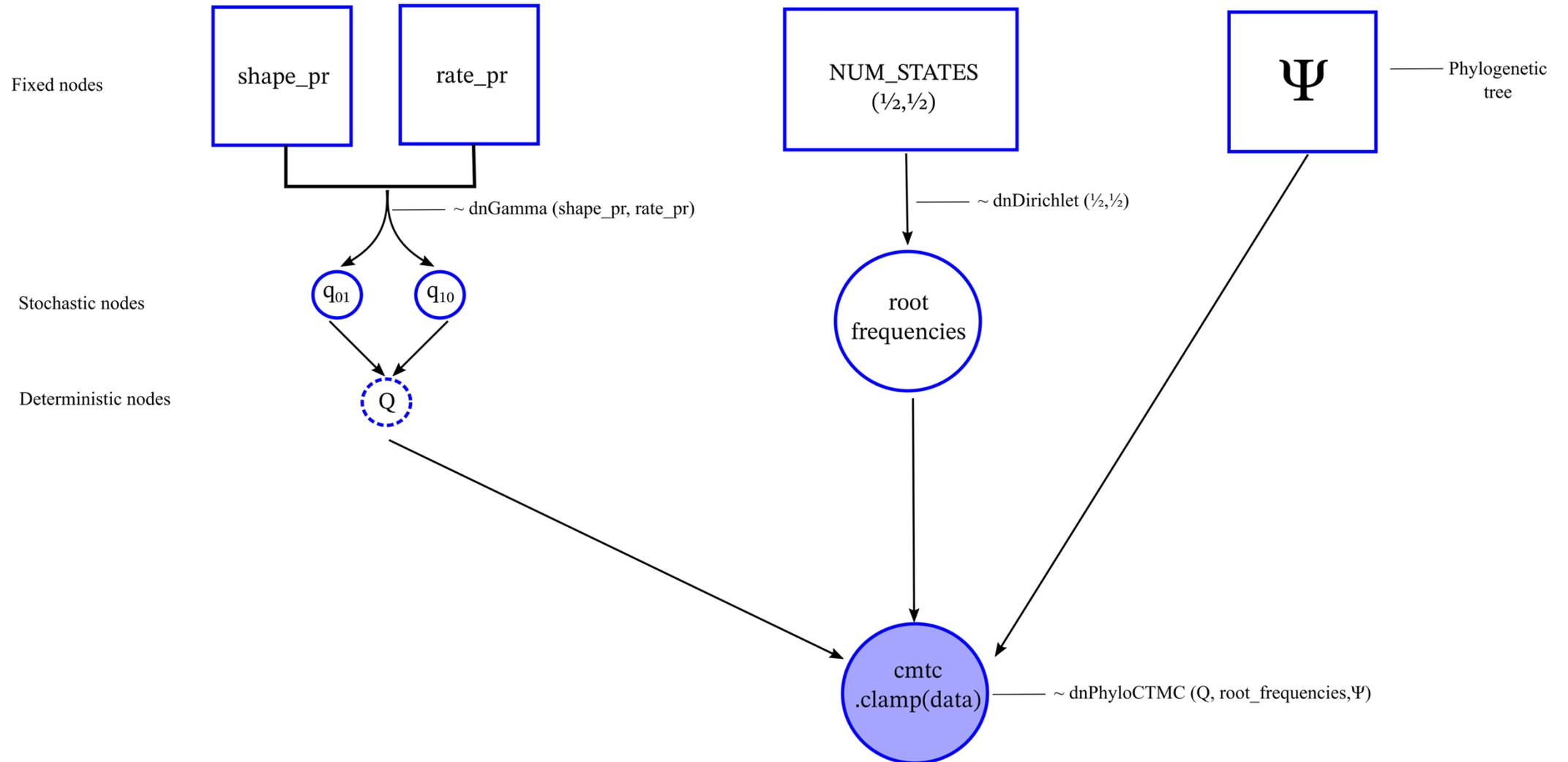
Frecuencias de la raíz



Mk2- Modelo gráfico probabilístico



Un paso más- Pinzar (clamping)



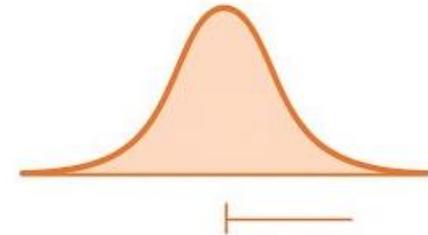
Monitores-Inferencia

monitors.append() –guardan la inferencia

En estadística bayesiana, el objeto de inferencia es la **distribución posterior**. Nos interesan los intervalos de máxima probabilidad (95%).

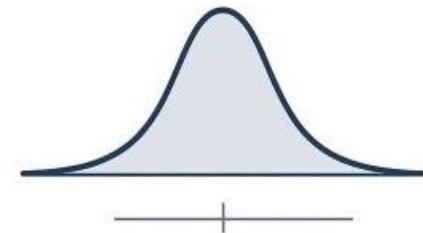
En la inferencia de verosimilitud (frecuentista) a menudo nos interesan las estimaciones de **máxima verosimilitud** y, con suerte, los intervalos de verosimilitud-confianza.

Credible Interval
(Bayesian)



Given our data and prior, there's a 95% probability the true value lies here

Confidence Interval
(Frequentist)



In 95 of 100 repeated experiments, the interval will contain the true value