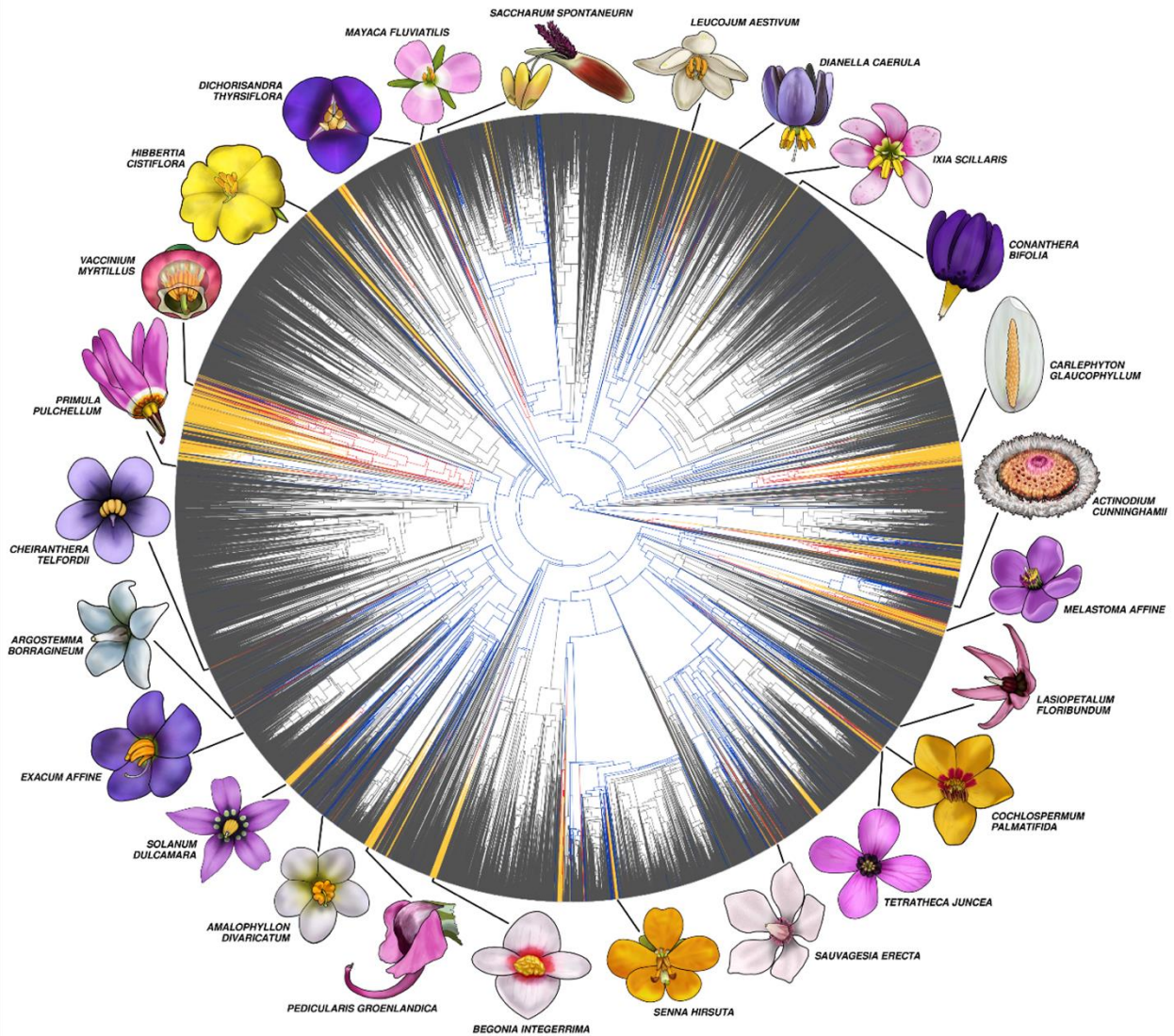


Todos los materiales del taller

https://roszenil.github.io/pcm_naturalistas/



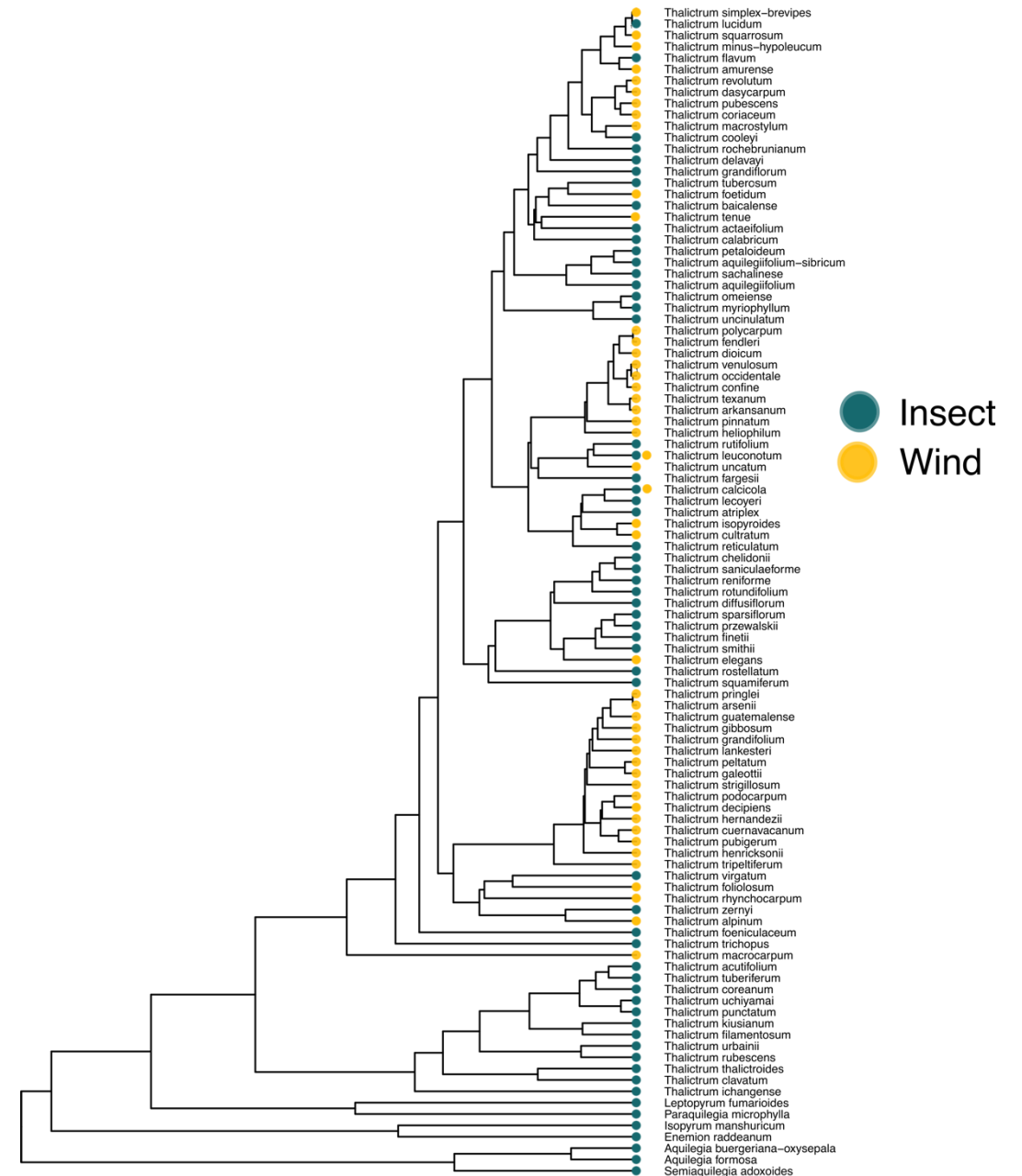
Introducción a las cadenas de Markov



Rosana Zenil-Ferguson
Assistant Professor. University of Kentucky

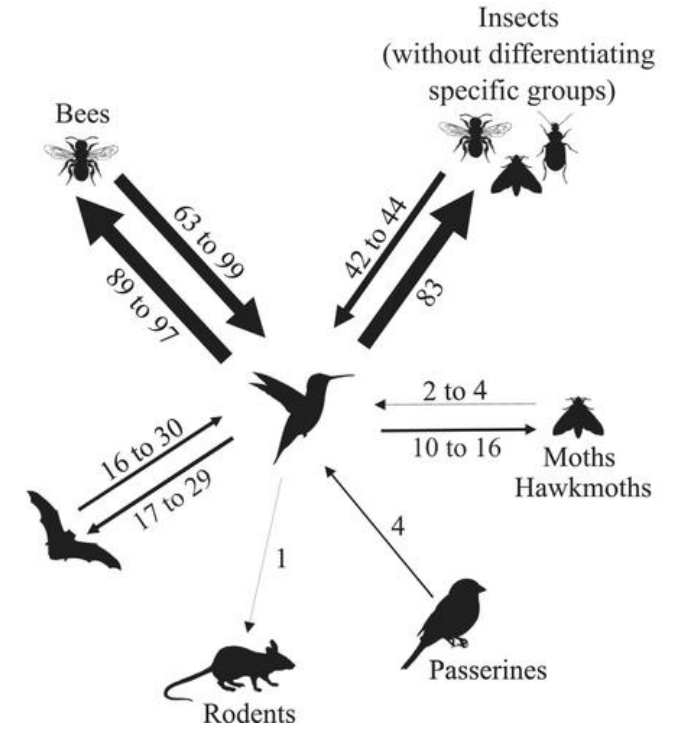
Árboles filogenéticos y Caracteres discretos

- Asumimos que el árbol filogenético está fijo
- Estamos interesados en saber la velocidad y el número de transiciones entre los dos estados.

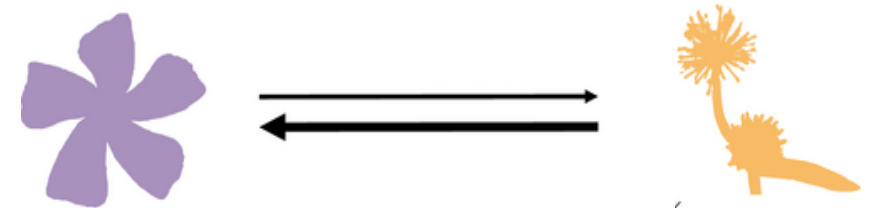


Polinización

- Plantas dependen de la polinización para asegurar su reproducción
- El polen se puede transferir por insectos, viento, vertebrados o incluso por el agua.
- Múltiples tipos de polinización, frecuentes cambios en las familias o géneros.
- La forma de la flor (anteras) ha coevolucionado con el polinizador



Barreto et al. 2024

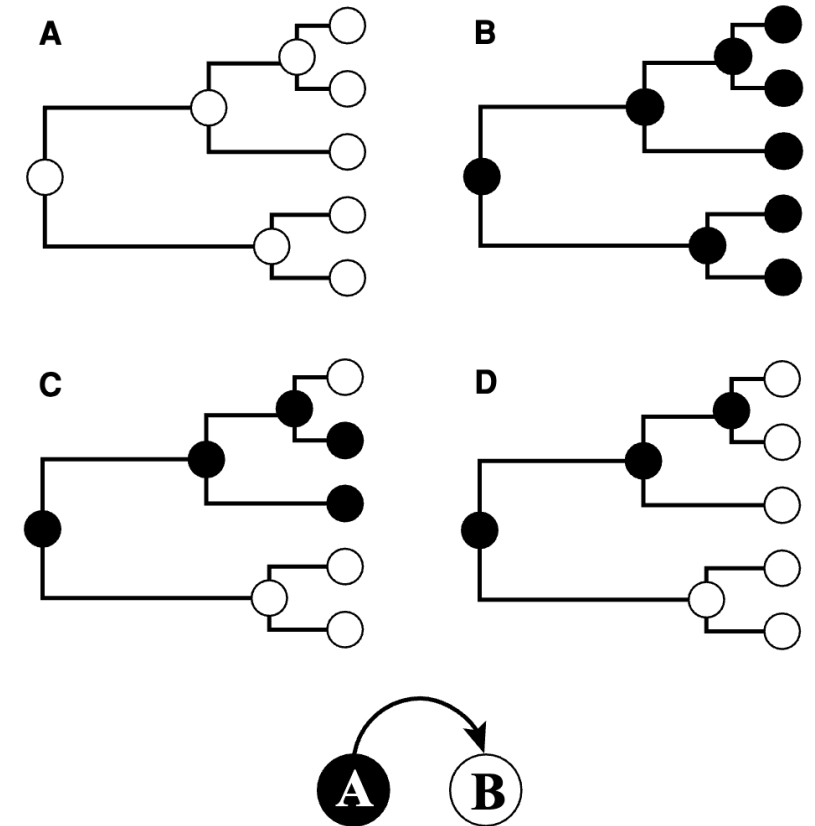


Cómo estimamos el número y la velocidad de las transiciones?

Qué es lo que estamos estudiando?

- La ley de la irreversibilidad de Dollo:
Un organism casi nunca regresa a un estado previo, incluso cuando dicho organism se encuentra en condiciones existenciales idénticas a las cuáles antes experimentó. - Louis Dollo (1893)

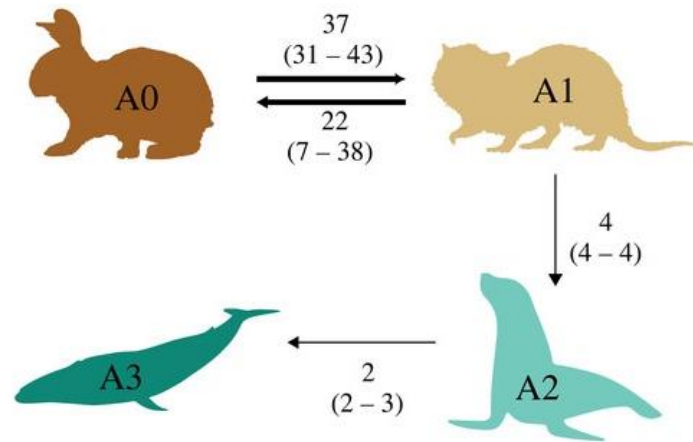
Goldberg and Igić. 2008



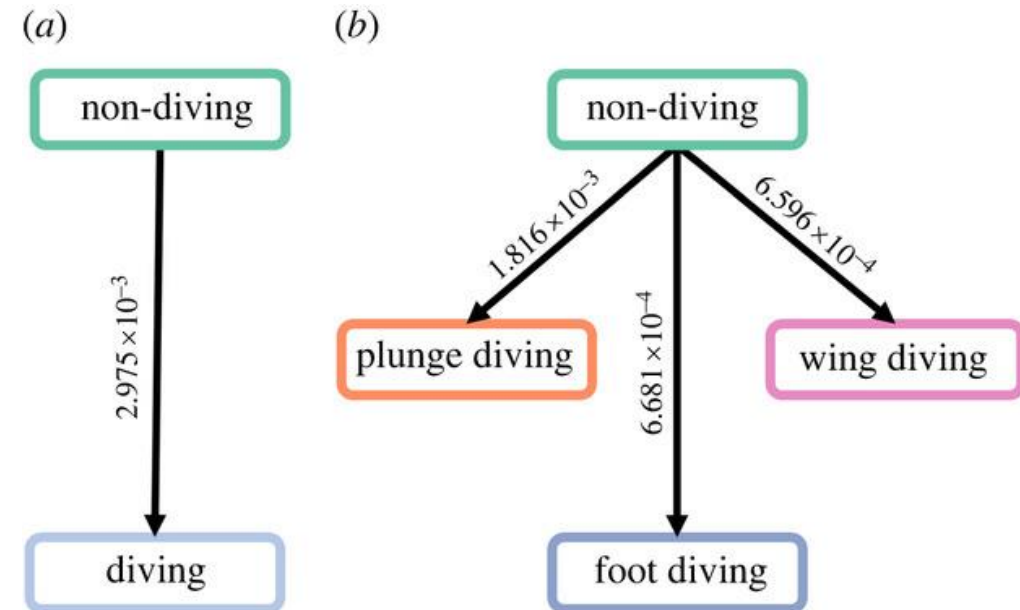
Qué es lo que estamos estudiando?

- La ley de la irreversibilidad de Dollo:

Un organism casi nunca regresa a un estado previo, incluso cuando dicho organism se encuentra en condiciones existenciales idénticas a las cuáles antes experimentó. - Louis Dollo (1893)

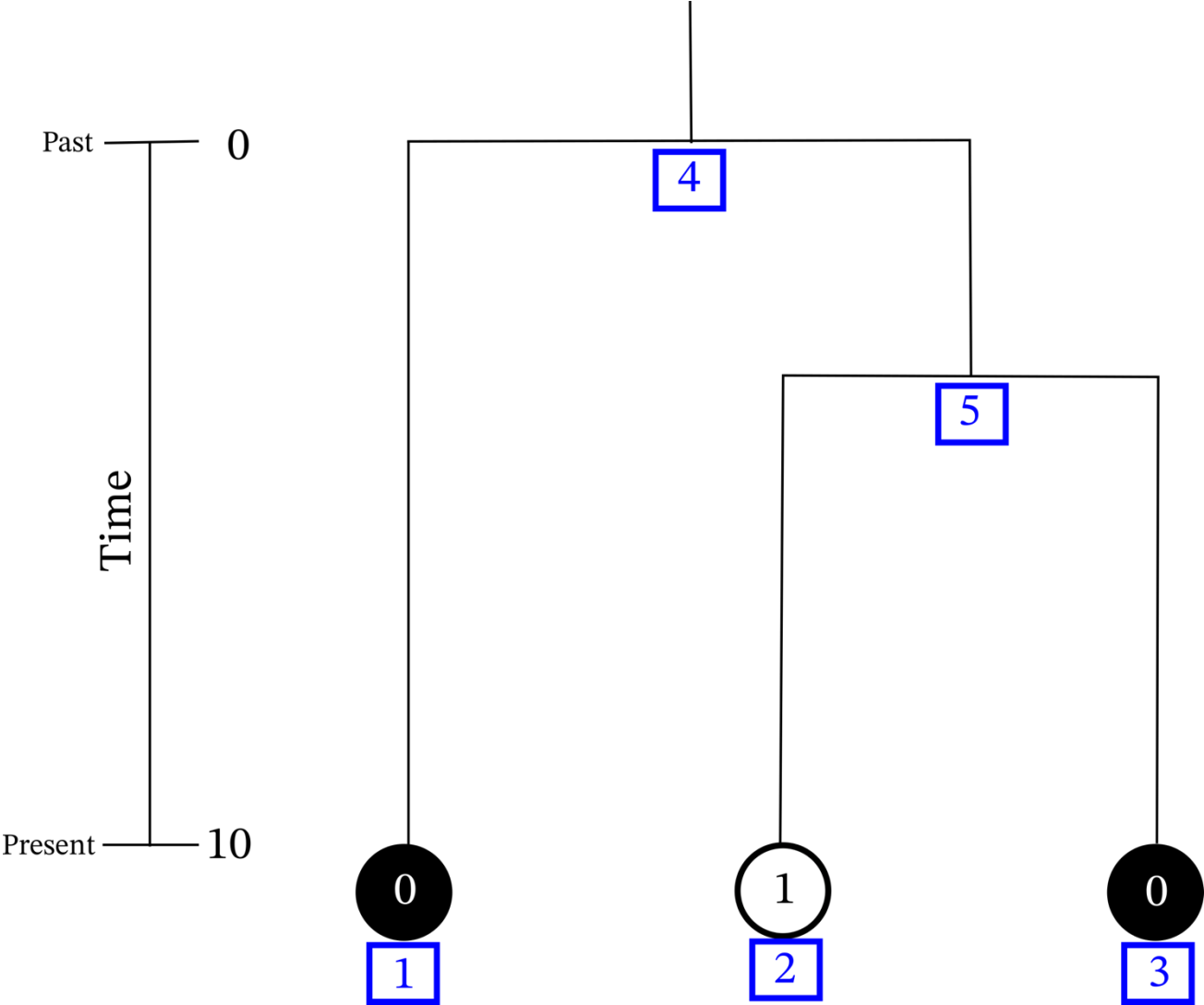


Adaptación secundaria a habitats acuáticos (Farina and Silvestro, 2023)



Buceo de aves acuáticas
Tyler and Younger, 2022

Enumerando los nodos de una filogenia (computadoras)



Definición de una variable aleatoria

Cadenas de Markov en tiempo continuo

Continuous-Time Markov Chains (CTMC)

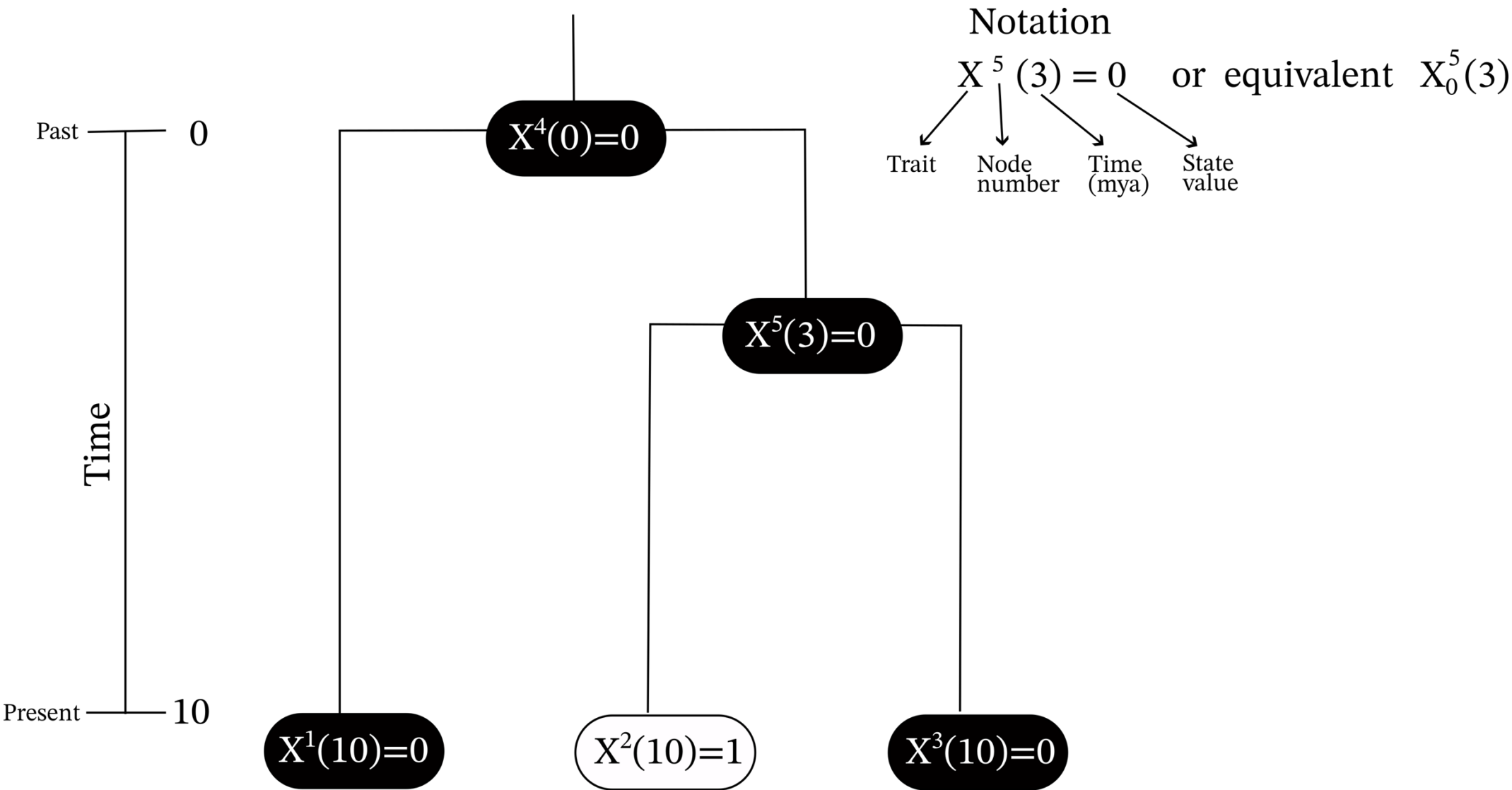
$$\{X(t), t \geq 0\}$$

Stochastic models that follow change in time with an associated **probability**

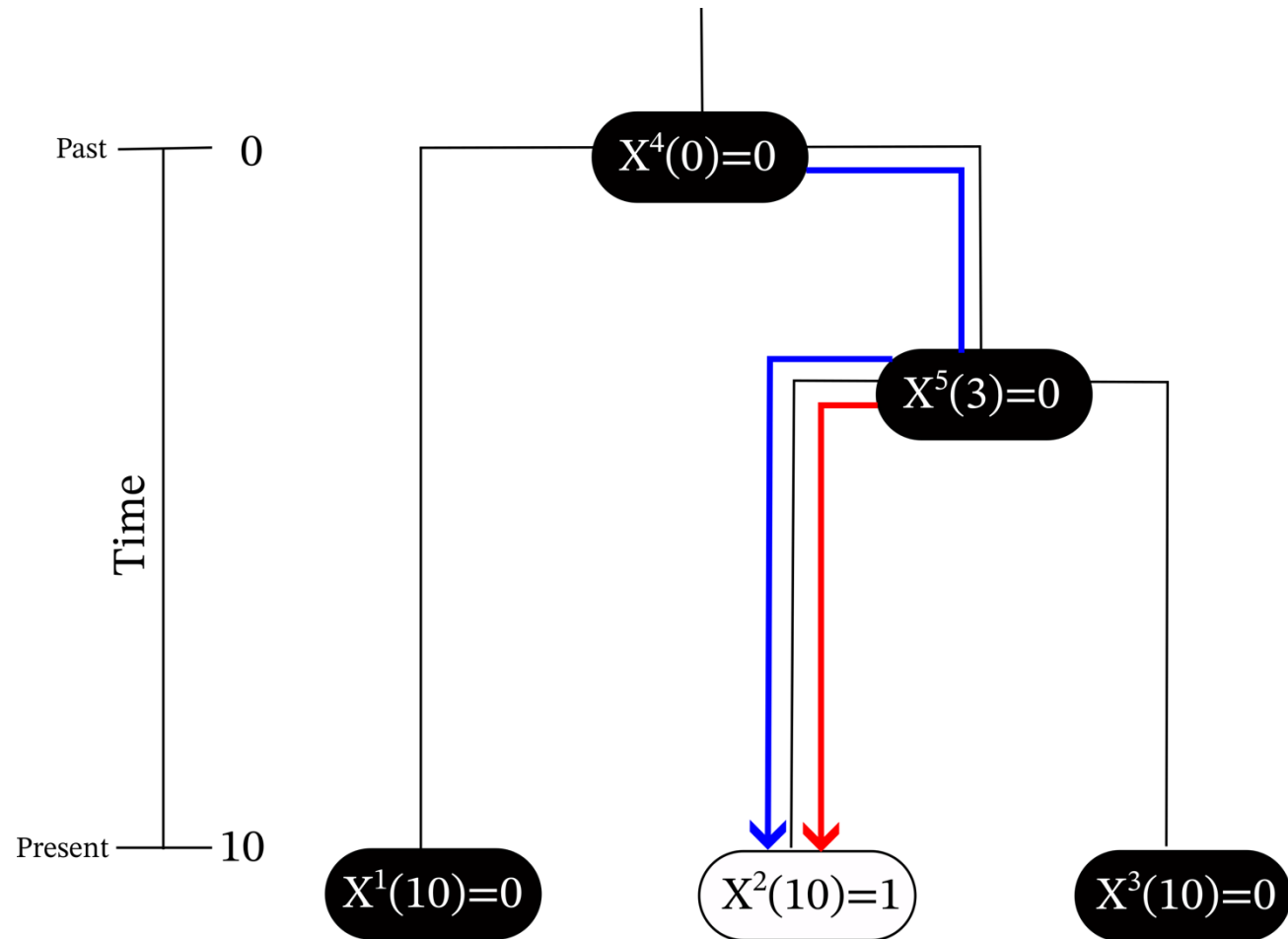
Modelos estocásticos que siguen el cambio en el **tiempo** y tienen asociados una **probabilidad**

$X(t)$ = el valor del fenotipo o caracter al tiempo t

$X(t)$ = Insecto (0), Viento (1)



Lo Markov en las cadenas de Markov



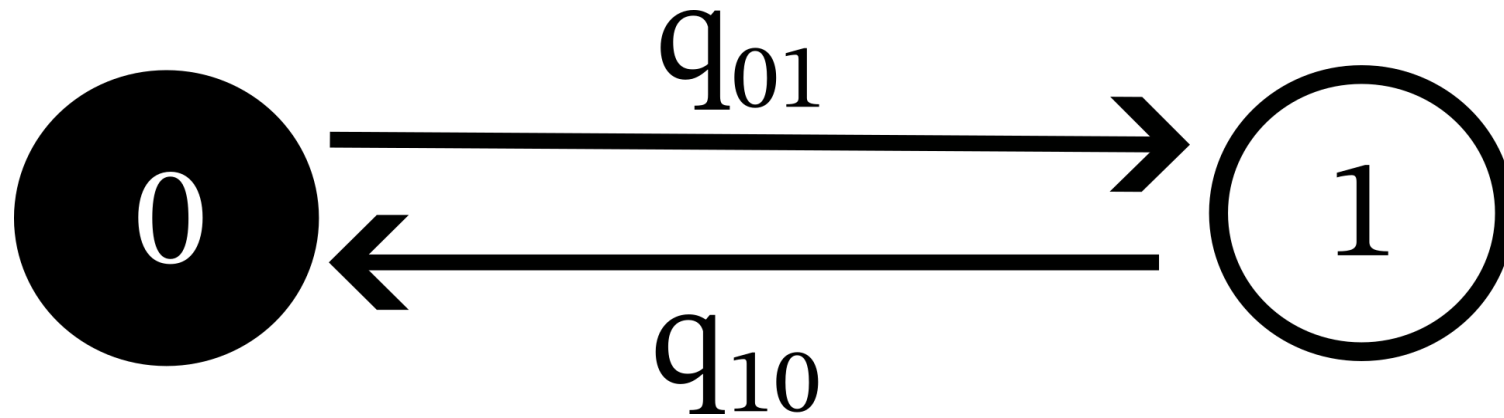
Probabilidades condicionales

$$P(X^2(10)=1|X^5(3)=0)=p_{01}(10-3)=p_{01}(7)$$

La matriz P- Las probabilidades

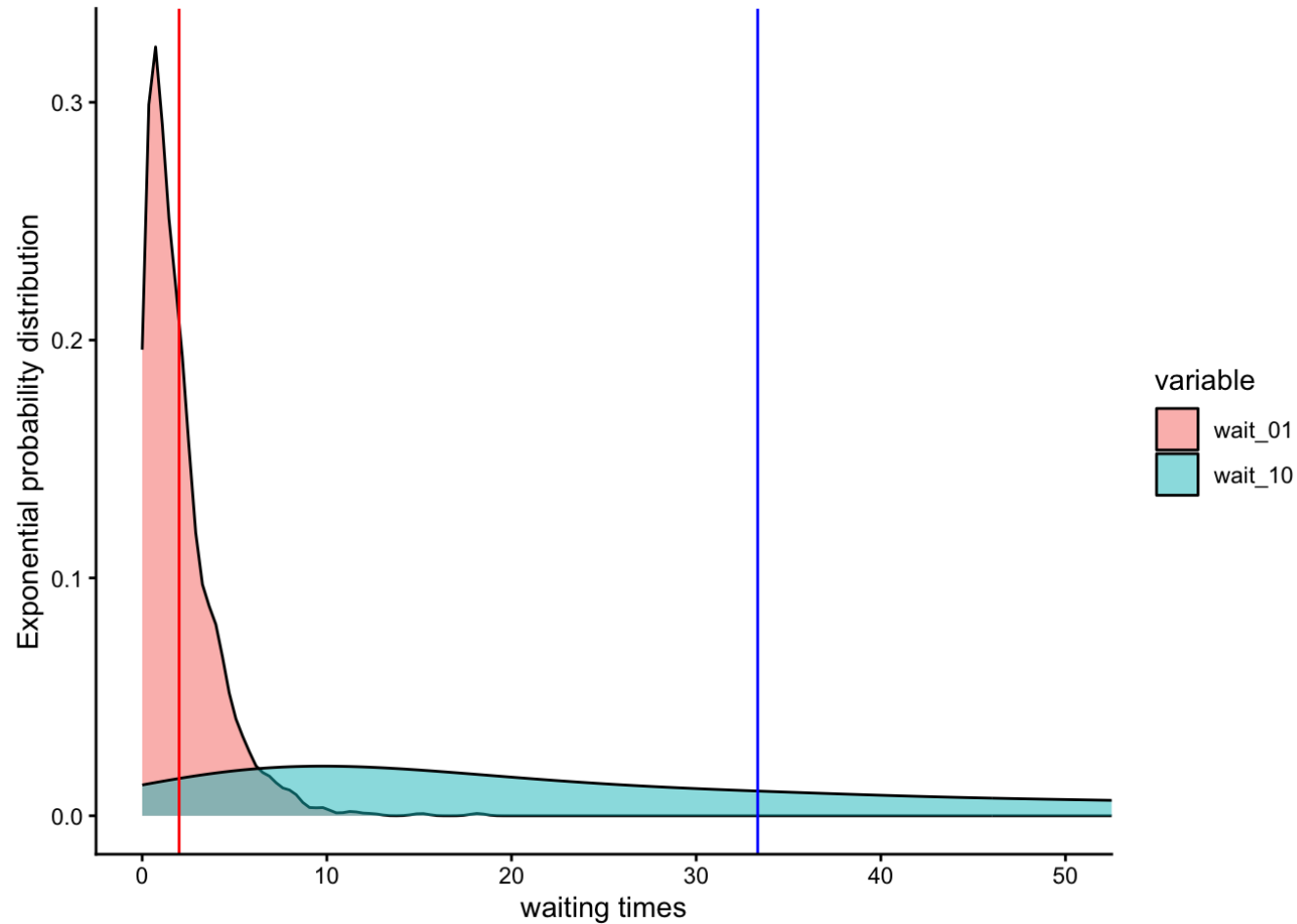
$$P(t) = \begin{pmatrix} p_{00}(t) & p_{01}(t) \\ p_{10}(t) & p_{11}(t) \end{pmatrix}$$

Tasas de Transición- Definición

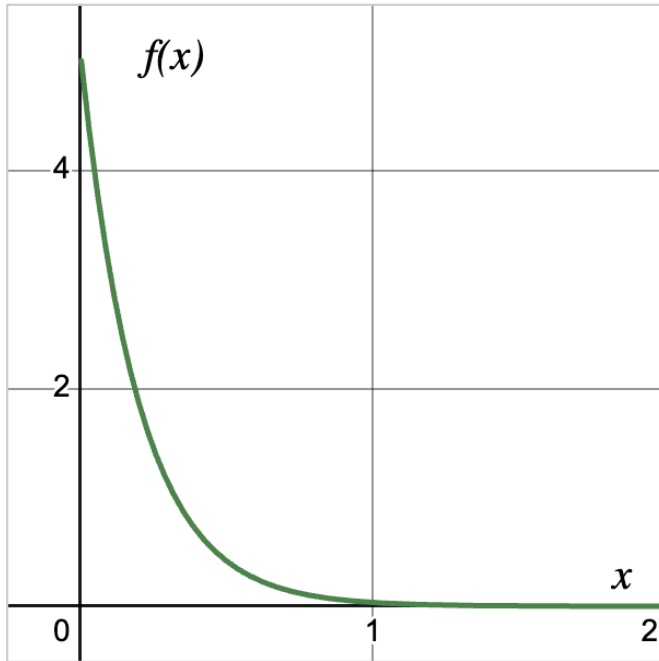


$$\lim_{h \rightarrow 0} P(X(t+h) = j | X(t) = i) = \lim_{h \rightarrow 0} p_{ij}(h) = q_{ij}$$

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



La exponencial es una distribución de tiempos de espera

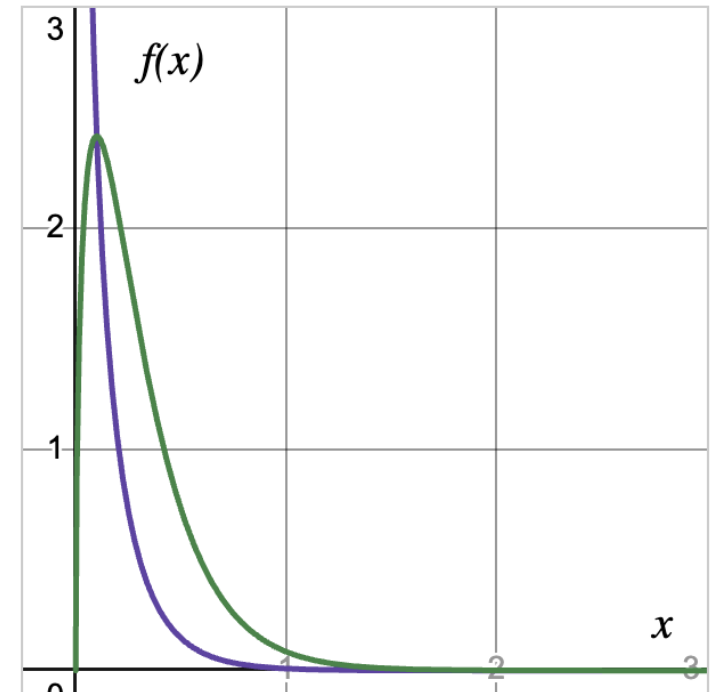


exponential distribution with parameter $\lambda > 0$, write $X \sim \text{exponential}(\lambda)$,

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

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

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

La Q y la P cómo se relacionan?

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

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

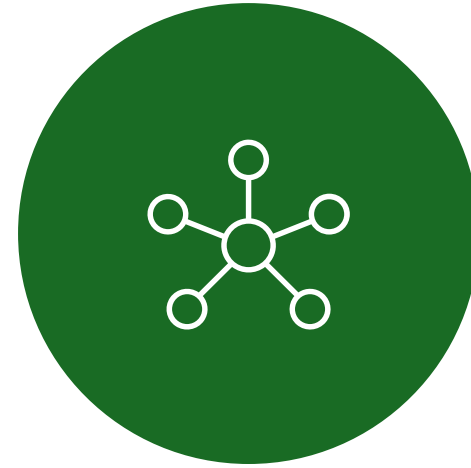
Introducción a RevBayes

Modelos Probabilísticos Bayesianos

Rev to the Bayes



BASICS OF BAYESIAN
STATISTICS



DRAWING PROBABILISTIC
MODELS

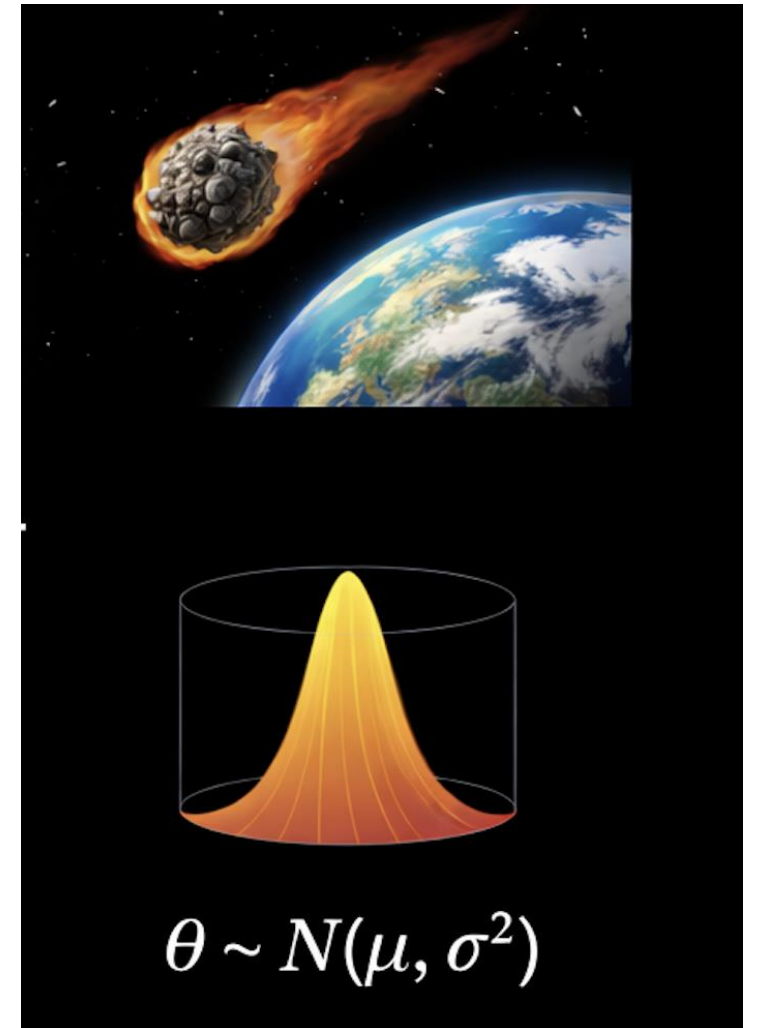
Dos conceptos esenciales en estadística Bayesiana

1. Las probabilidades iniciales reflejan **conocimiento previo o creencias sobre nuestra incertidumbre**.

E.g.Cuál es la probabilidad de que el Cotopaxi tenga una erupción hoy?

2. Cada **parametro** que queremos inferir es **desconocido e incierto**.

Los parámetros son variables aleatorias.



Cumberland Falls State Park, KY



Thalictrum
*Múltiples cambios de
tipo de polinización*

Verónica di Stilio
University of Washington



Thalictrum pubescences



Thalictrum thalictroides

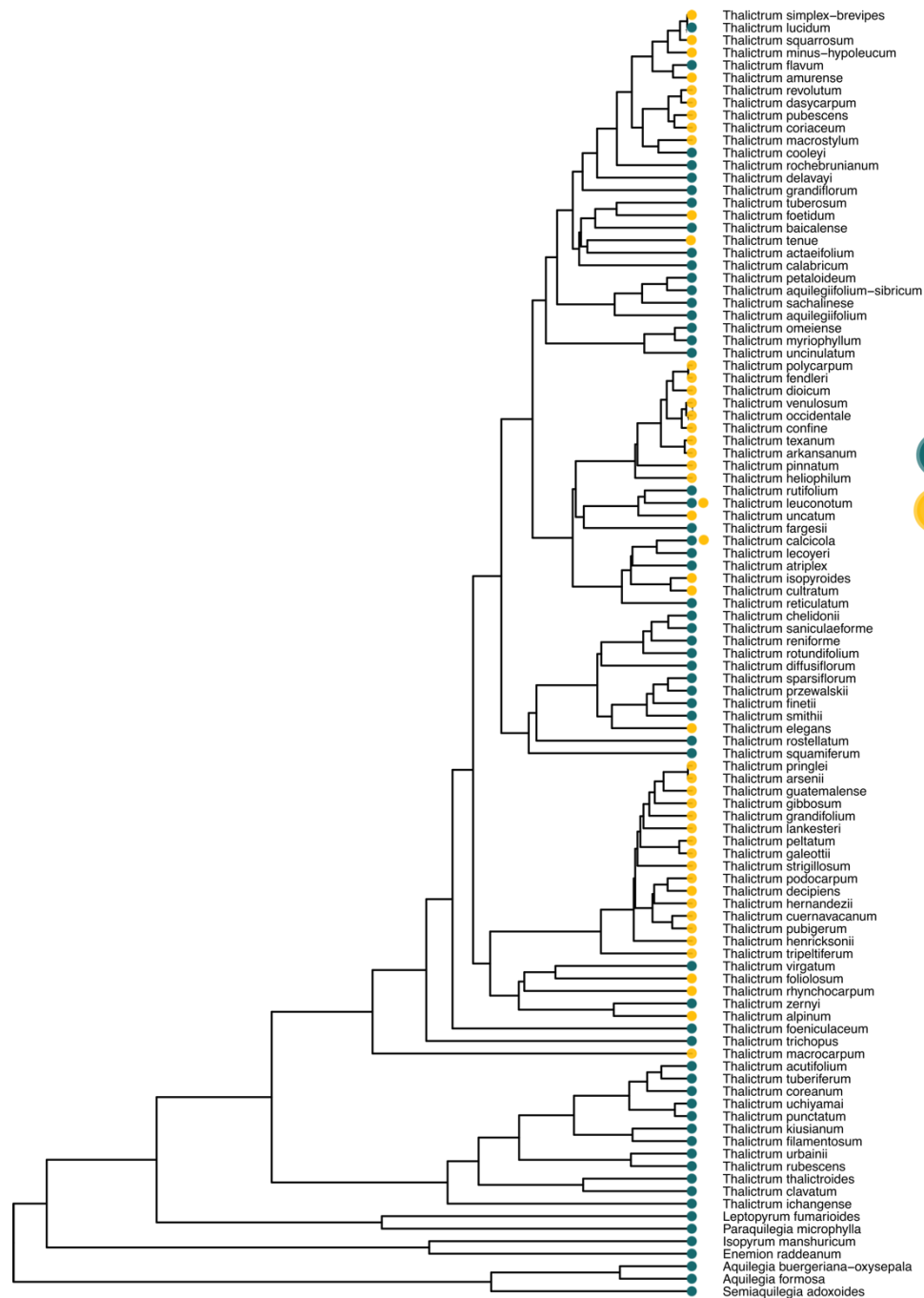
Tipos de polinización

Insecto= 0



Viento =1





● Insect
● Wind

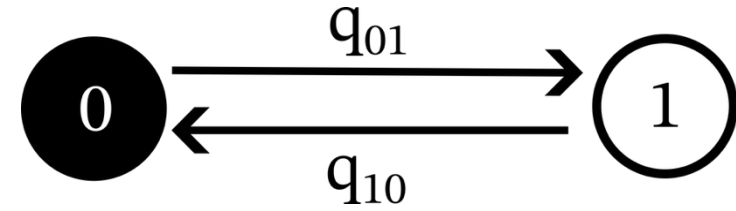
- Siempre grafiquen sus datos!
- Checar anidaciones en clados!
- Las transiciones es lo que les da poder para estimar los modelos que les queremos presentar.

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



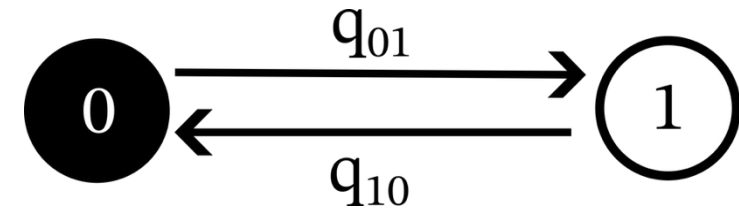
Parámetros desconocidos e inciertos del Mk2

1. Tasas de transición

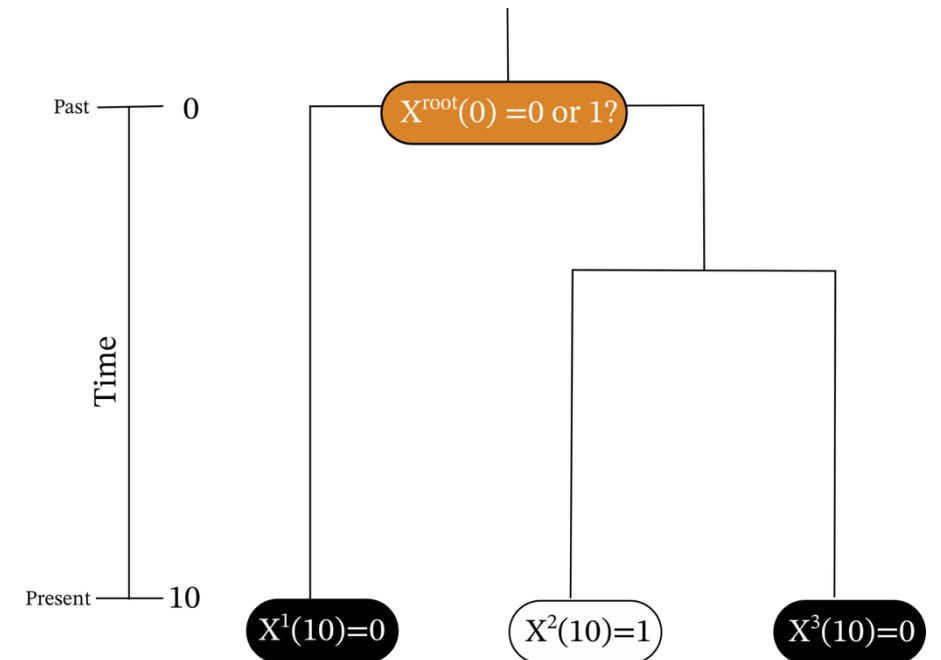


Parámetros desconocidos e inciertos del Mk2

1. Tasas de transición



2. Las frecuencias de los estados en la raíz.



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

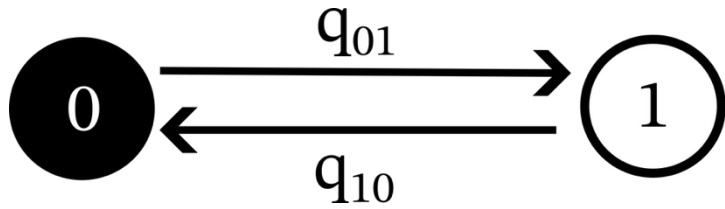


e) Plate

for (i in 1:N) { ... } plates (multiple connected nodes)

= for fixed operations (not part of graphical model)

Distribución Gamma para las tasas



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

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

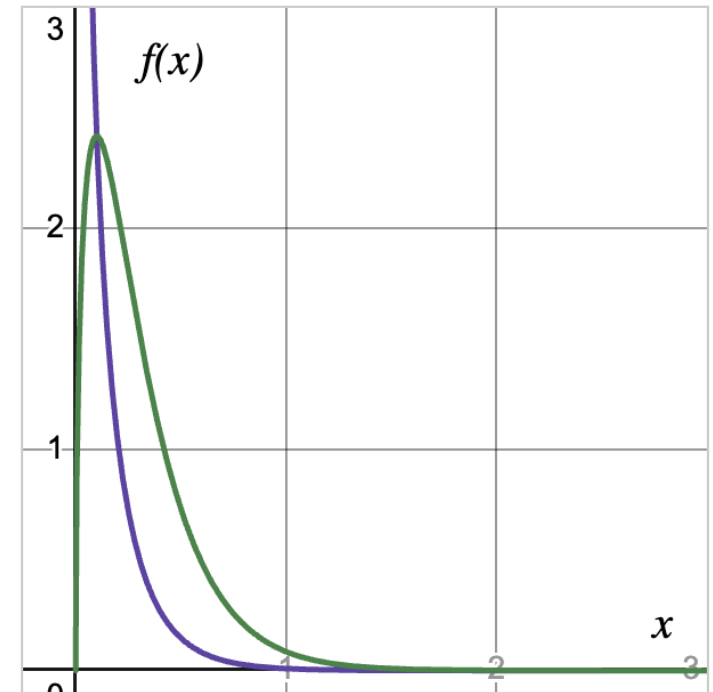


Gráfico probabilístico

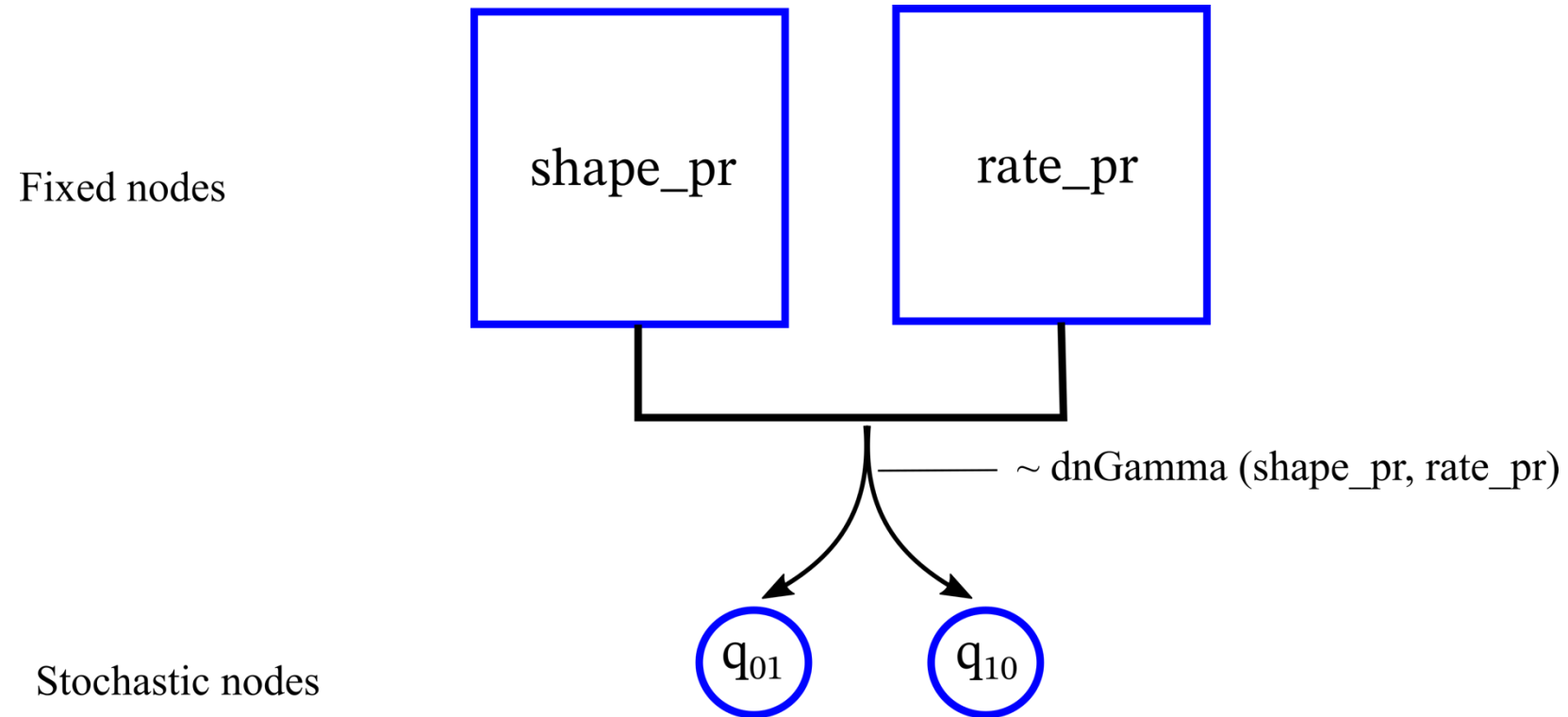
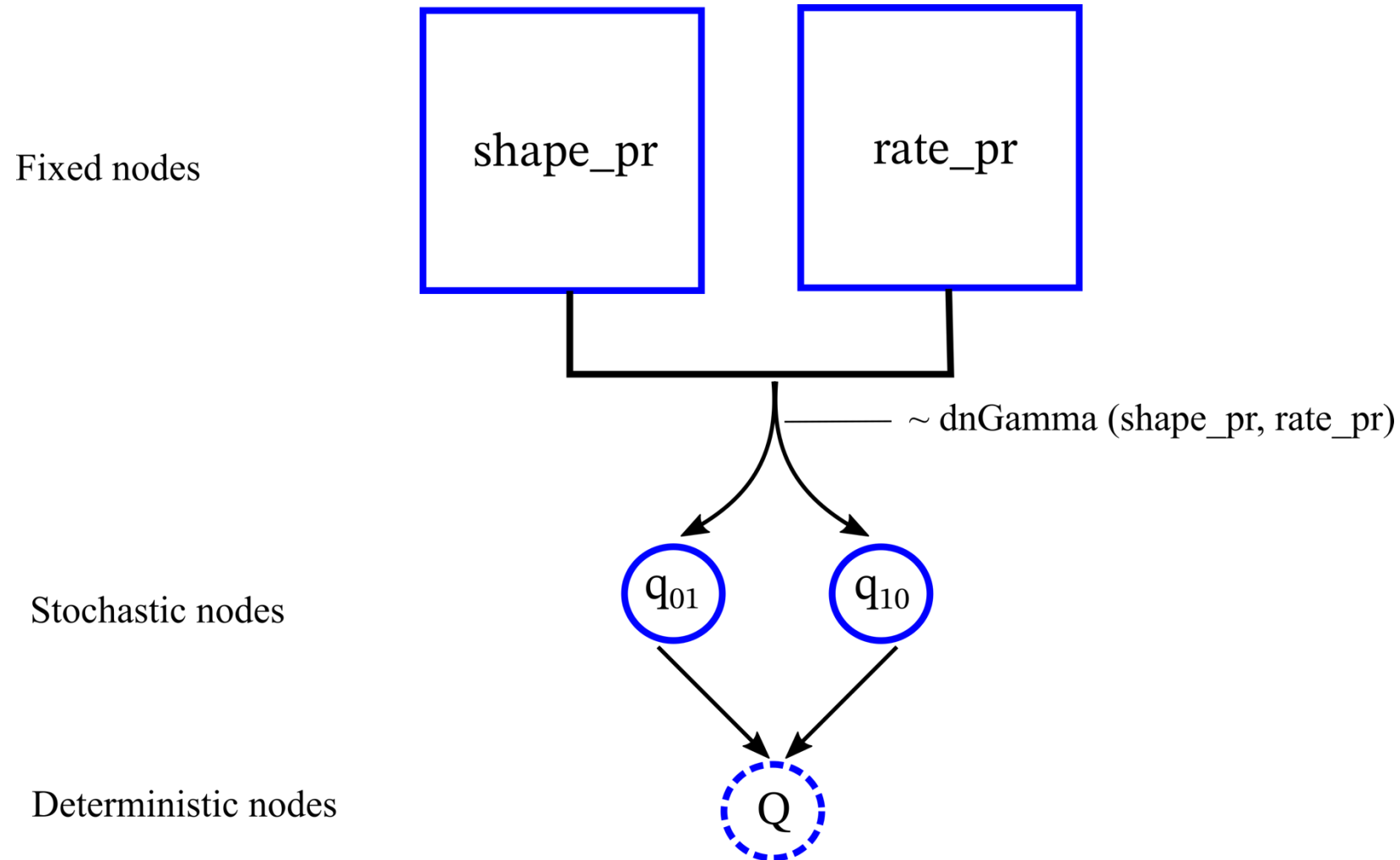
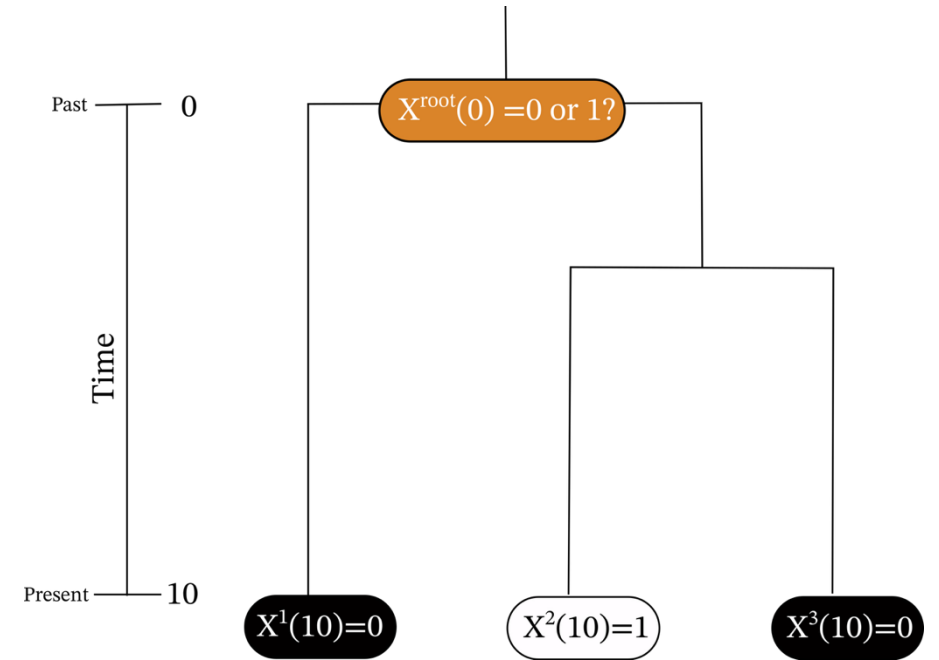
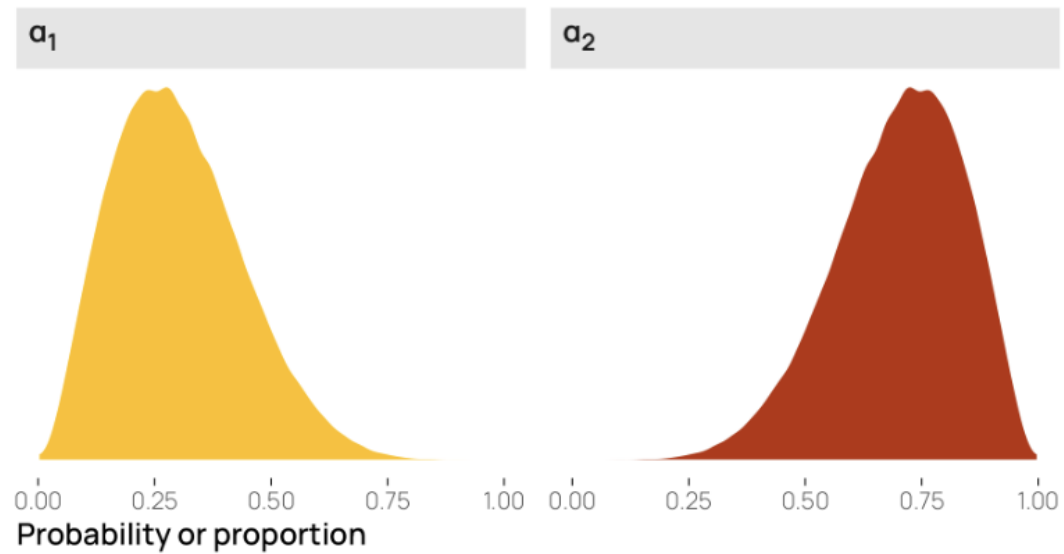


Gráfico probabilístico

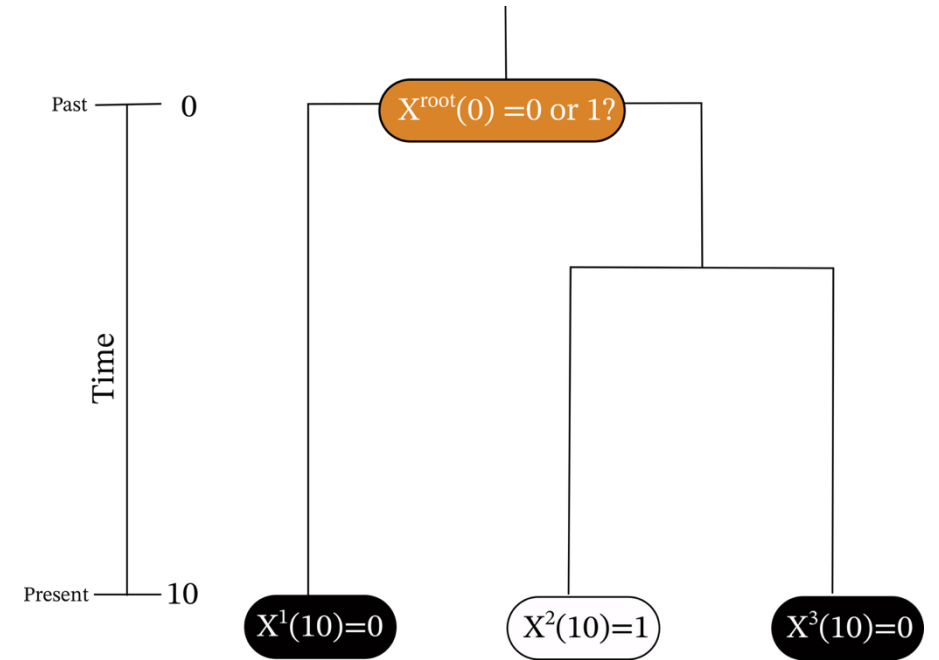
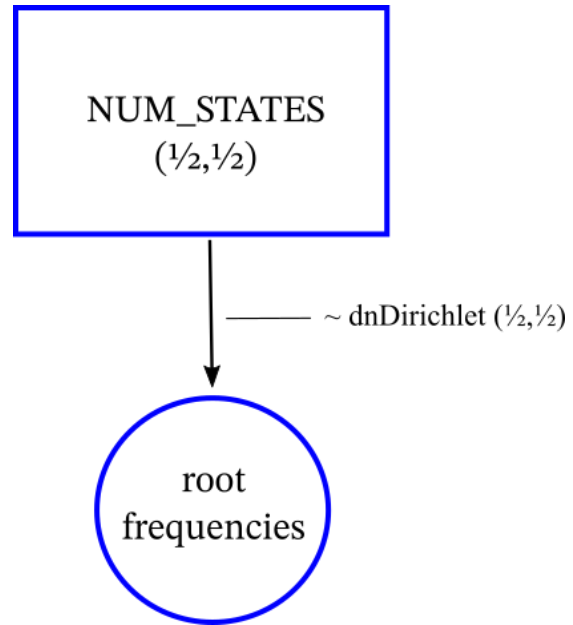


Frecuencias de la raíz

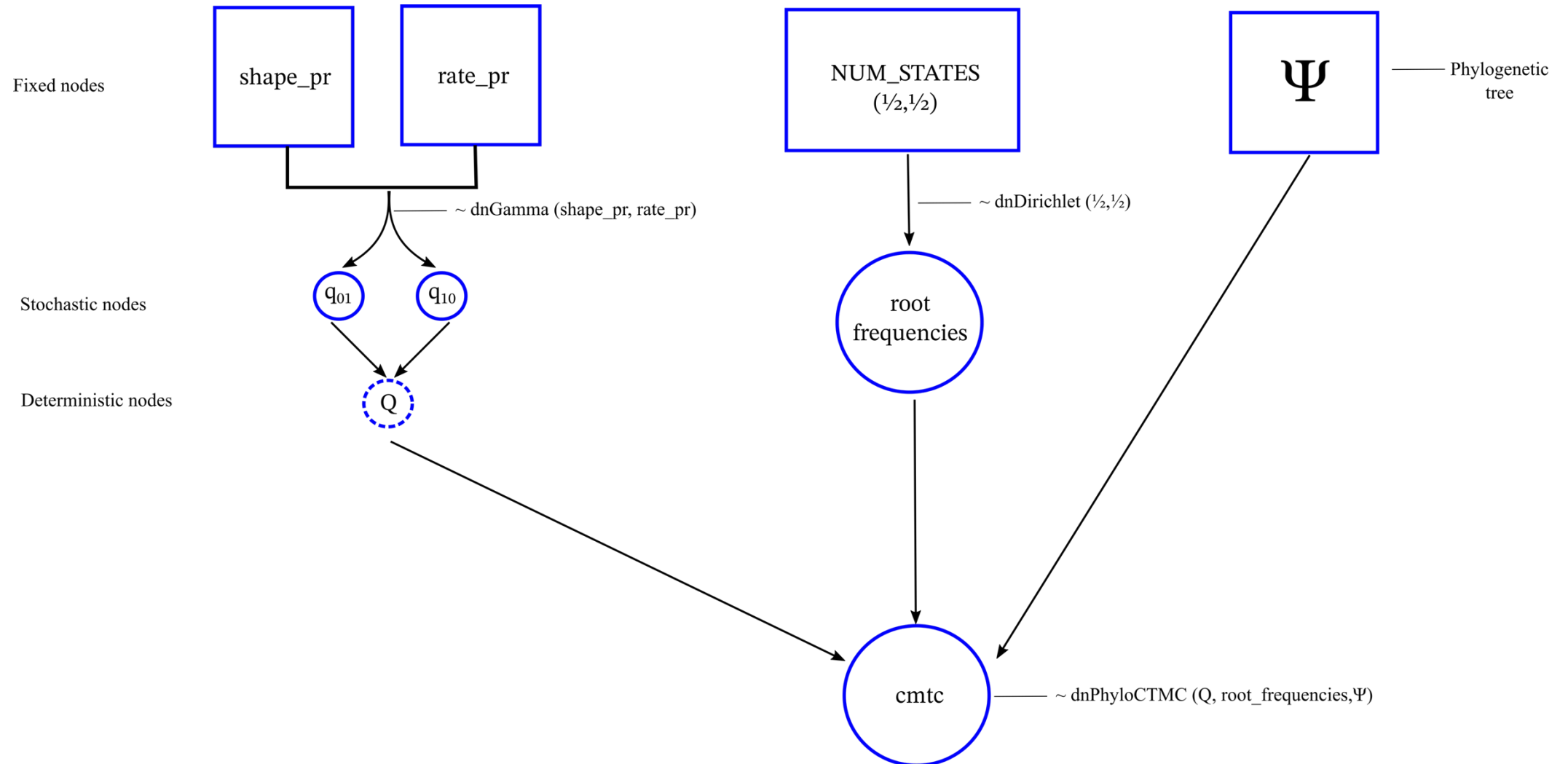
Dirichlet(3, 7)



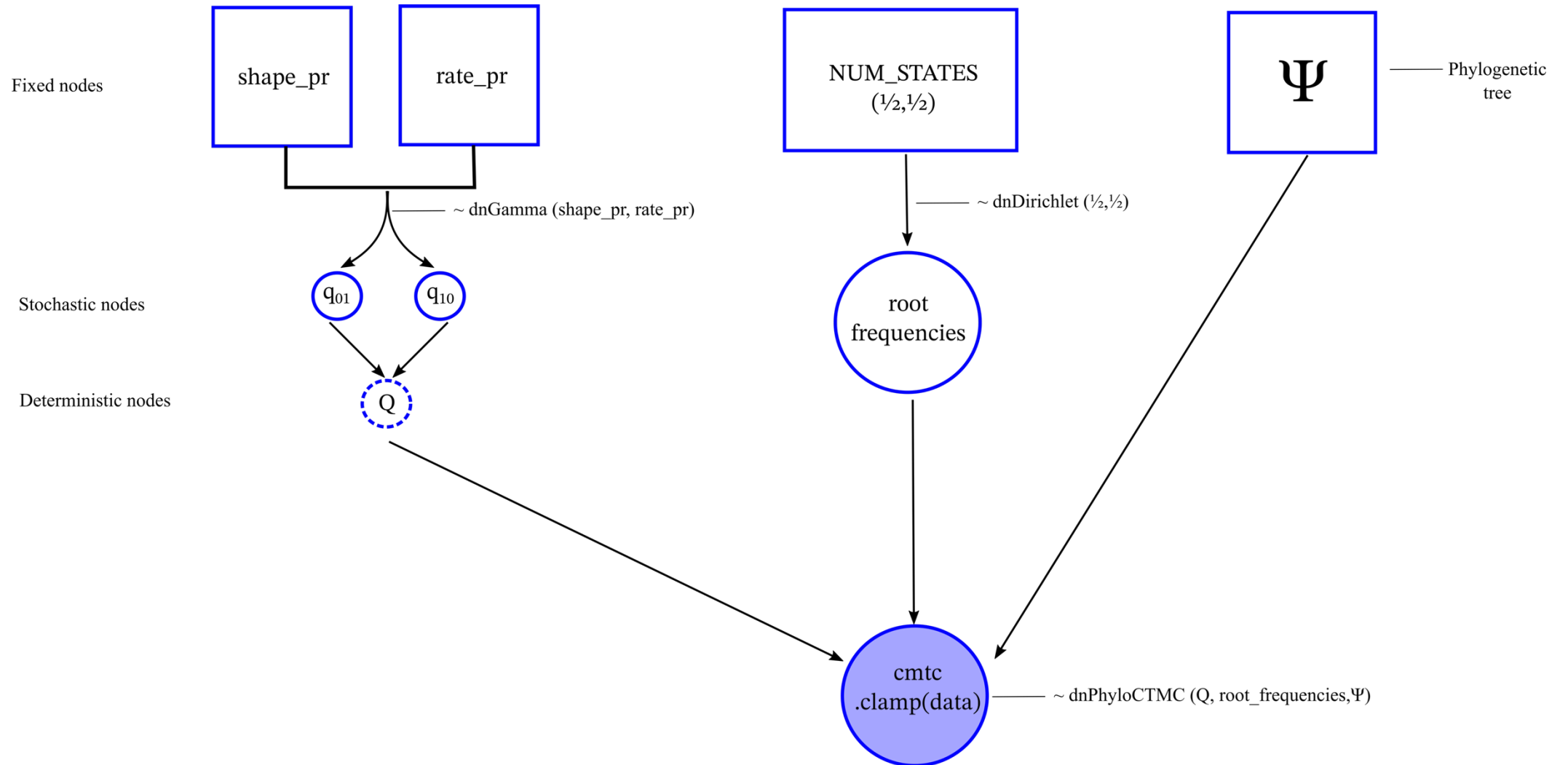
Frecuencias de la raíz



Mk2- Modelo gráfico probabilístico



Un paso más- Pinzar (clamping)



Movimientos- Propuestas

`moves.append()`

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.



Watch Dr. Paul Lewis
Bayesian inference and MCMC
Lecture

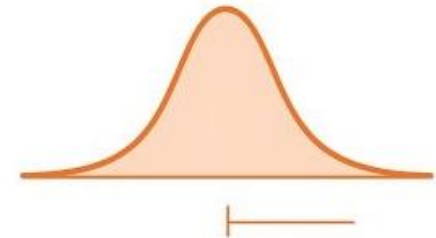
Monitores-Inferencia

monitors.append() –guardan la inferencia

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

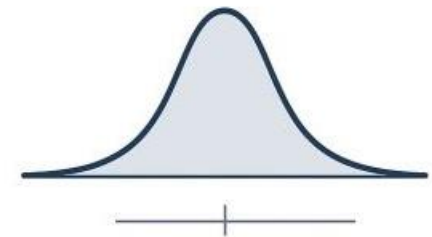
En la inferencia de verosimilitud (frecuentista) a menudo nos interesan las estimaciones de **máxima verosimilitud** y, con suerte, los intervalos de verosimilitud-confianza.

Credible Interval
(Bayesian)



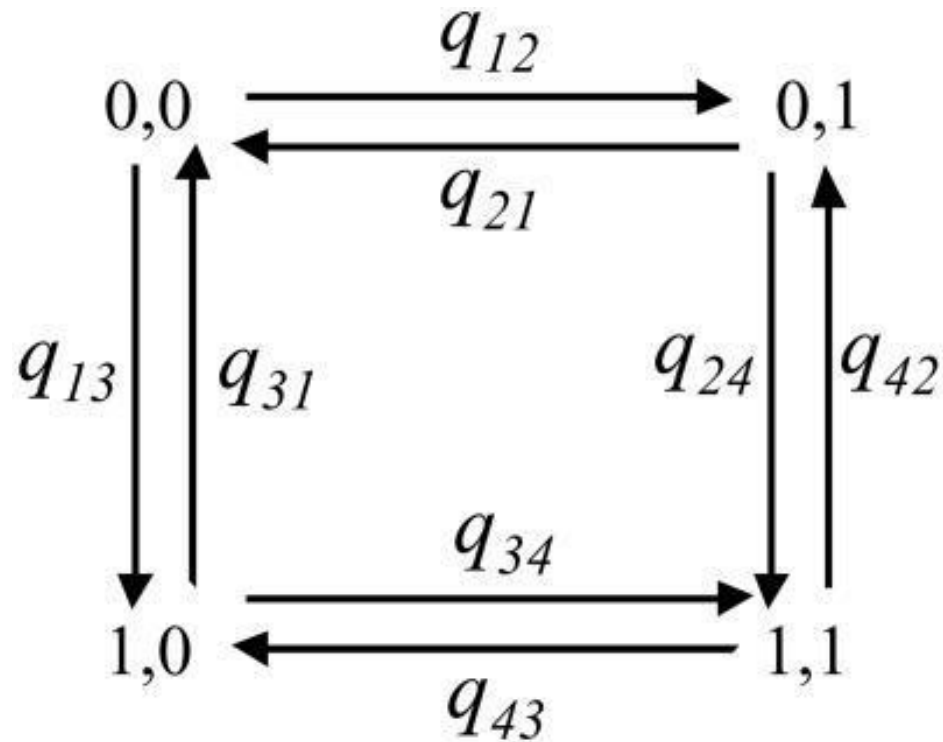
Given our data and prior, there's a 95% probability the true value lies here

Confidence Interval
(Frequentist)



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

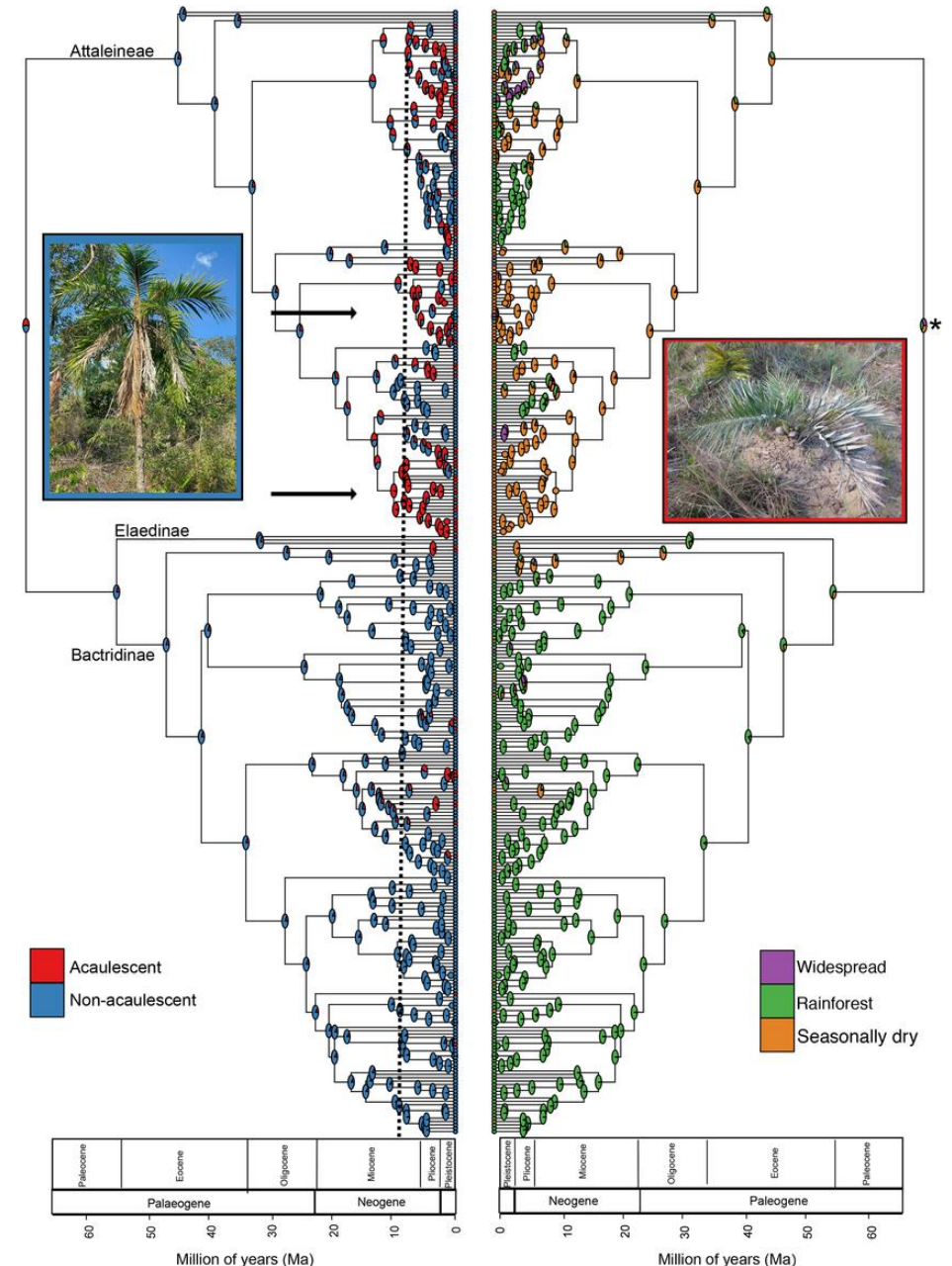
Las matrices Q son el centro del universo
(para caracteres discretos)



El test de correlación de Pagel

Pagel 1994

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

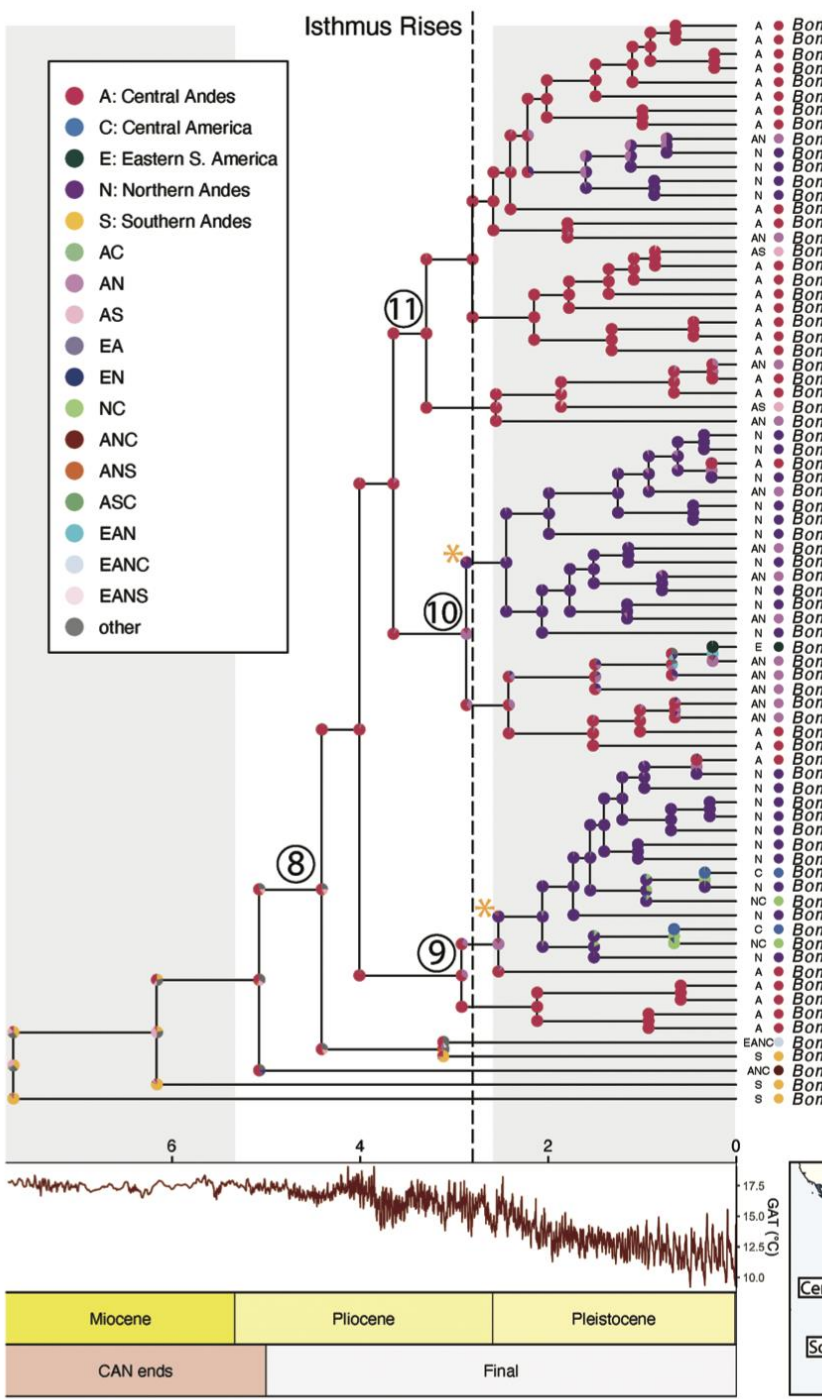


$Q =$

	\emptyset	A	B	C	AB	AC	BC	ABC
\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	—

Dispersal-Extinction Cladogenesis model
(DEC)

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



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

ChromEvol (Mayrose et al. 2010)

BiChrom (Zenil-Ferguson et al. 2017)

