

INFERENCIA FILOGENÉTICA

INFERENCIA BAYESIANA

Teorema de Bayes

Diagram illustrating Bayes' Theorem with labels and arrows:

- Probabilidad posterior (Posterior Probability) points to $\Pr(\text{Hipótesis}|\text{Datos})$
- Verosimilitud (Likelihood) points to $\Pr(\text{Datos}|\text{hipótesis})$
- Probabilidad *a priori* de la hipótesis (Prior Probability of the hypothesis) points to $\Pr(\text{Hipótesis})$
- Probabilidad *a priori* de los datos (Prior Probability of the data) points to $\Pr(\text{Datos})$

$$\Pr(\text{Hipótesis}|\text{Datos}) = \frac{\Pr(\text{Datos}|\text{hipótesis}) \times \Pr(\text{Hipótesis})}{\Pr(\text{Datos})}$$



Probabilidad *a priori* de
la hipótesis = 0.5

- Mitad monedas normales (50% chance cara o sello)
- Mitad monedas sesgadas (75% chance sello, 25% chance cara)

Hipótesis 1: La moneda es normal

Hipótesis 2: La moneda es sesgada

Datos





Prob. normal	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.001
Prob. sesgada	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.056

Verosimilitud

- Normal: 0.5^{10}
- Sesgada: 0.75^{10}

Teorema de Bayes

Probabilidad posterior
de que la moneda está
sesgada

Verosimilitu
d

Probabilidad *a priori* de
que la moneda está
sesgada

$$\Pr(\text{Hipótesis}|\text{Datos}) = \frac{0.056 \times 0.5}{\Pr(\text{Datos})}$$

Probabilidad *a priori* de
los datos

Teorema de Bayes

Probabilidad posterior
de que la moneda está
sesgada



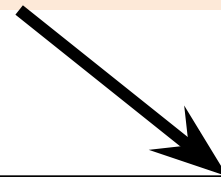
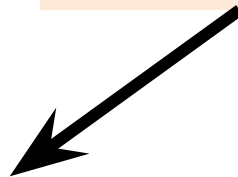
$$\Pr(\text{Hipótesis}|\text{Datos}) = \frac{0.02181}{\Pr(\text{Datos})}$$



Prob. de obtener los
datos bajo todas la
hipótesis



Probabilidad *a priori* de
los datos



$$0.5 \times \left(\begin{array}{|l|} \hline \text{Prob. de moneda} \\ \text{normal: } 0.5^{10} \\ \hline \end{array} + \begin{array}{|l|} \hline \text{Prob. de moneda} \\ \text{sesgada: } 0.75^{10} \\ \hline \end{array} \right) = 0.0286$$

Teorema de Bayes

Probabilidad posterior
de que la moneda está
sesgada



$$\Pr(\text{Hipótesis}|\text{Datos}) = \frac{0.02181}{0.0286} = 0.98$$

Probabilidad *a priori*

Probabilidad posterior

Sesgada

0.5



0.98

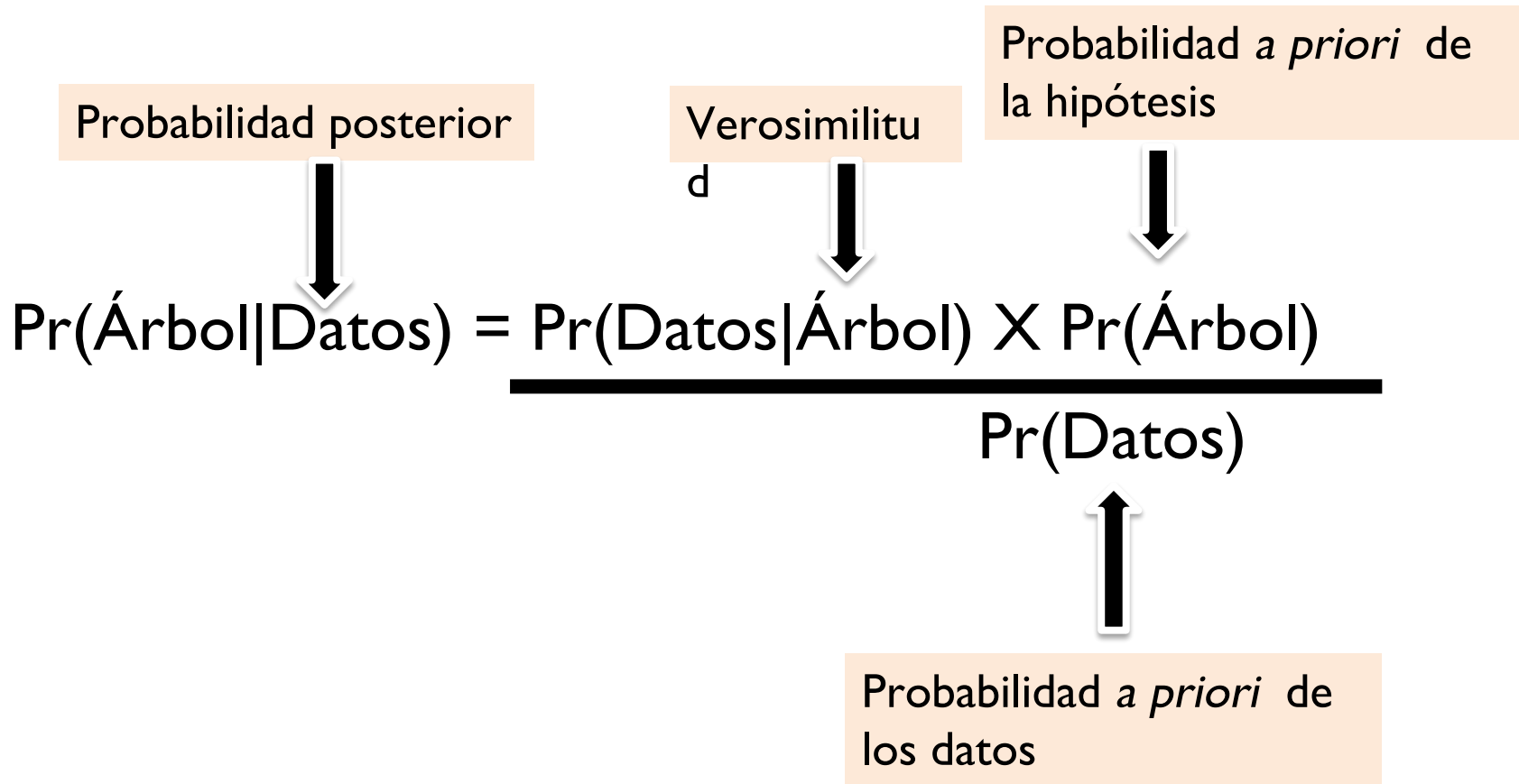
Normal

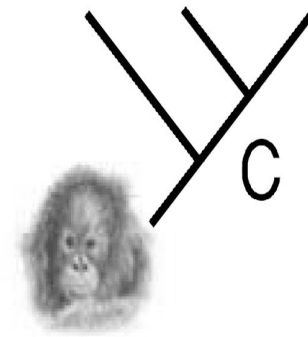
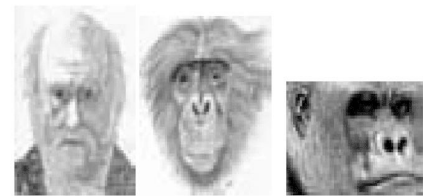
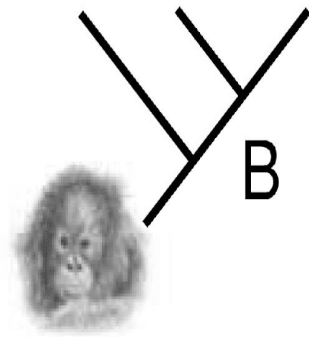
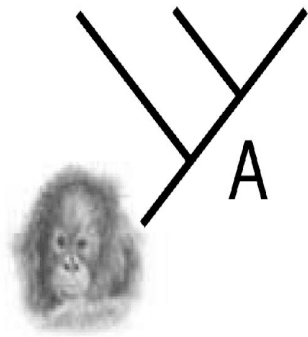
0.5

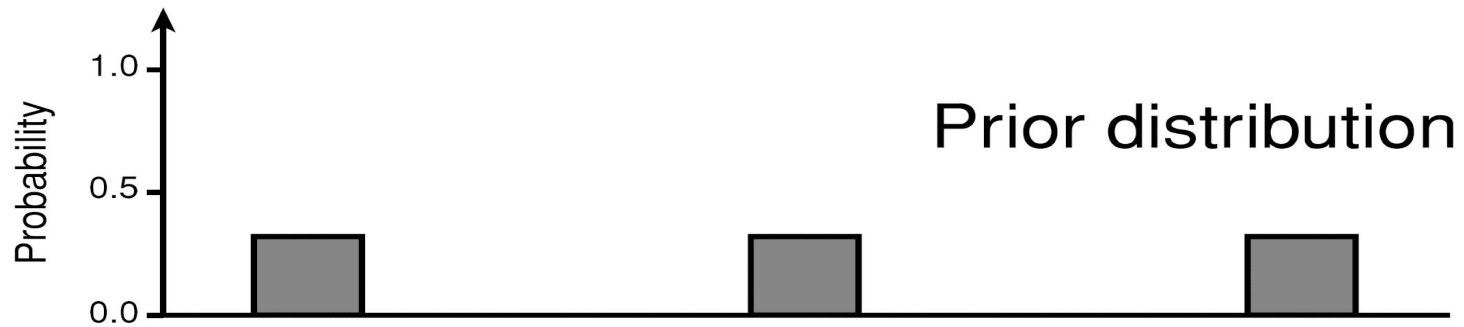
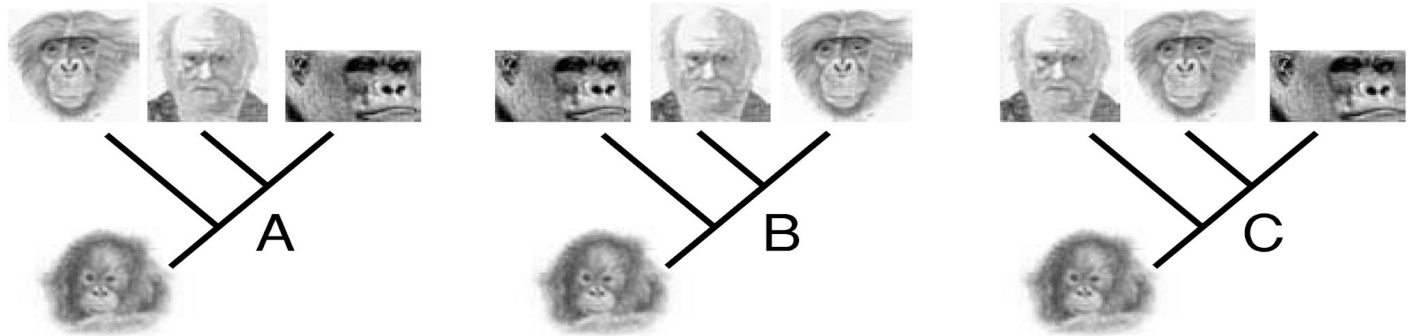


0.02

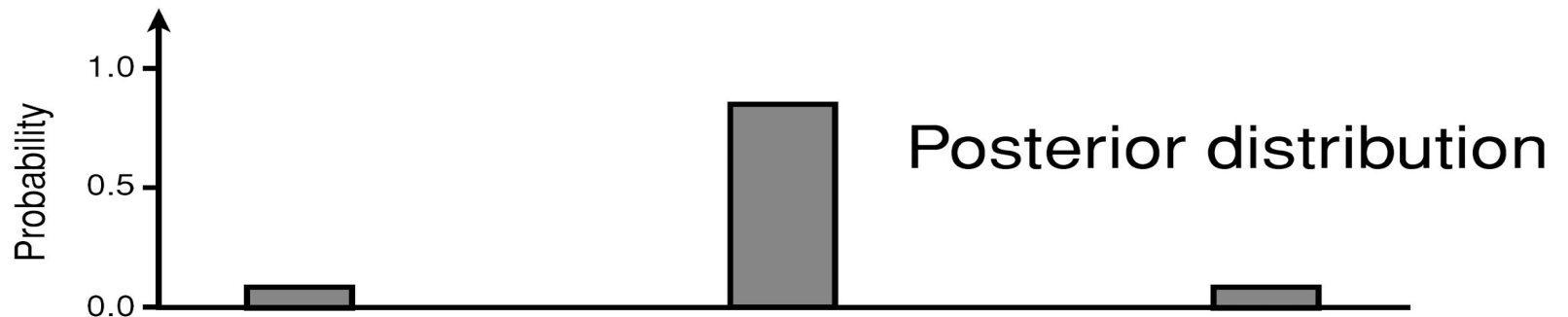
INFERENCIA BAYESIANA EN FILOGENÉTICA







↓ Datos y modelo de sustitución ↓



Probabilidad posterior



Verosimilitu

d



Probabilidad *a priori* de
la hipótesis



$$\Pr(\text{Árbol}|\text{Datos}) = \frac{\Pr(\text{Datos}|\text{Árbol}) \times \Pr(\text{Árbol})}{\Pr(\text{Datos})}$$

$\Pr(\text{Datos})$



Probabilidad *a priori* de
los datos

INFERENCIA BAYESIANA EN FILOGENÉTICA

¿Como obtener la probabilidad de los datos bajo todas la hipótesis posibles?

$\text{Pr}(\text{Datos})$

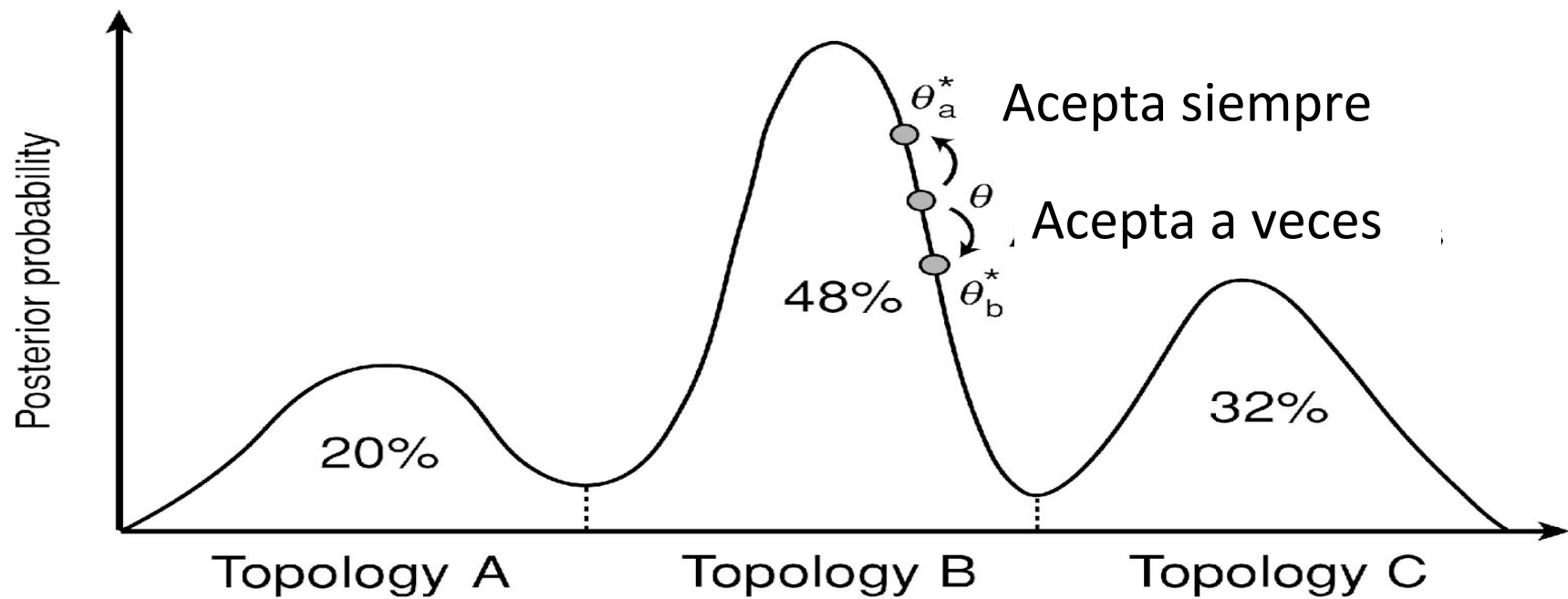


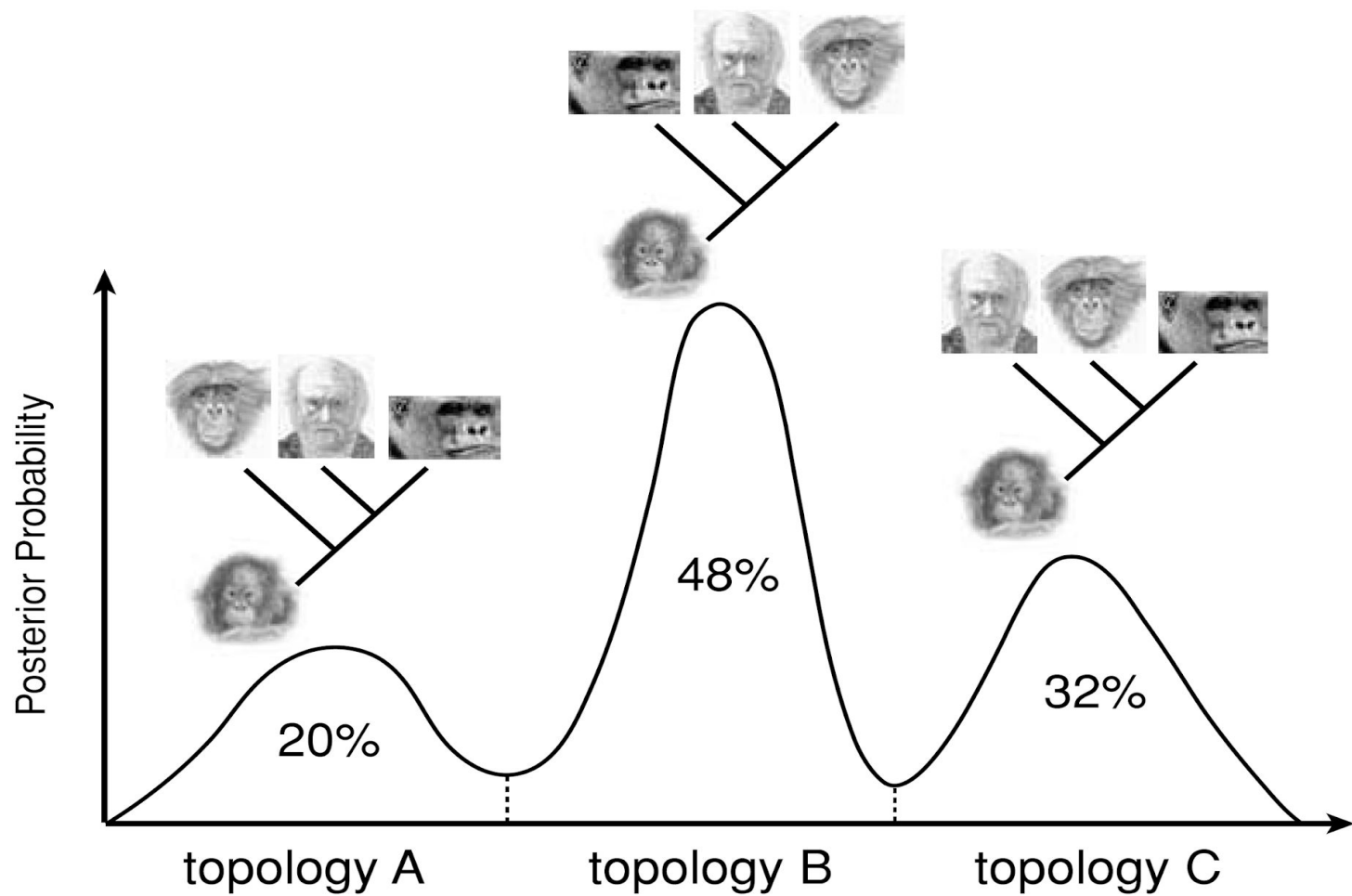
Probabilidad *a priori* de
los datos

Cadena de Markov Monte Carlo (MCMC)

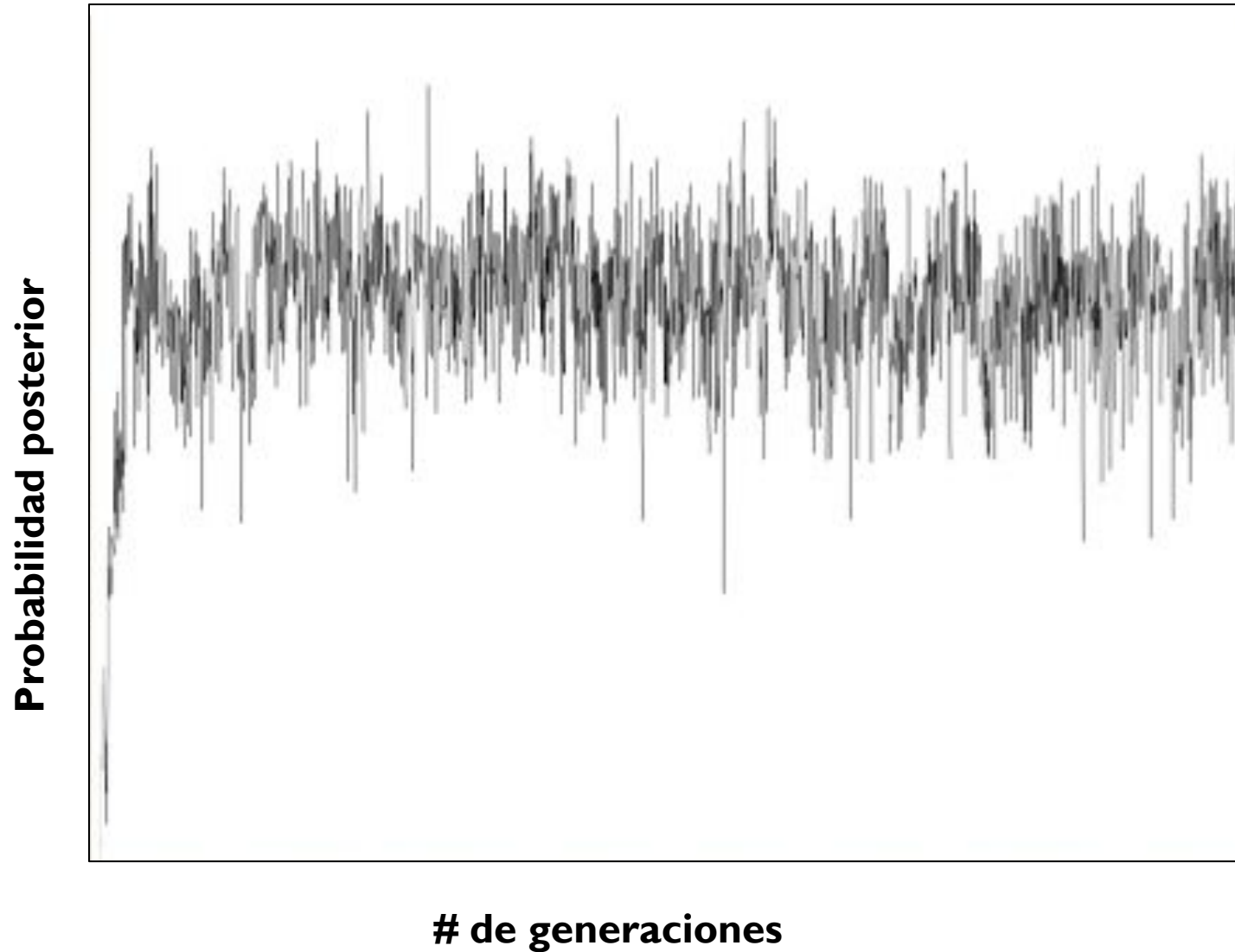
Permite estimar la distribución de probabilidades posteriores sin importar el punto de inicio en un paisaje de parámetros (topologías, ramas, parámetros de modelos) multidimensional

1. Comenzamos en un punto arbitrario de parámetros (θ)
2. Se hace un movimiento aleatorio hacia θ'
3. Se calcula la relación (r) entre θ' y θ
 - Si $r > 1$, aceptamos el nuevo estado θ'
 - Si $r < 1$, aceptamos el nuevo estado θ' con probabilidad r . Si se rechaza, nos quedamos con θ
4. Volvemos al paso 2 y repetimos millones de veces (generaciones)
5. Guardar árbol y parámetros cada n generaciones





Cadena de Markov Monte Carlo (MCMC)



Complicaciones del MCMC

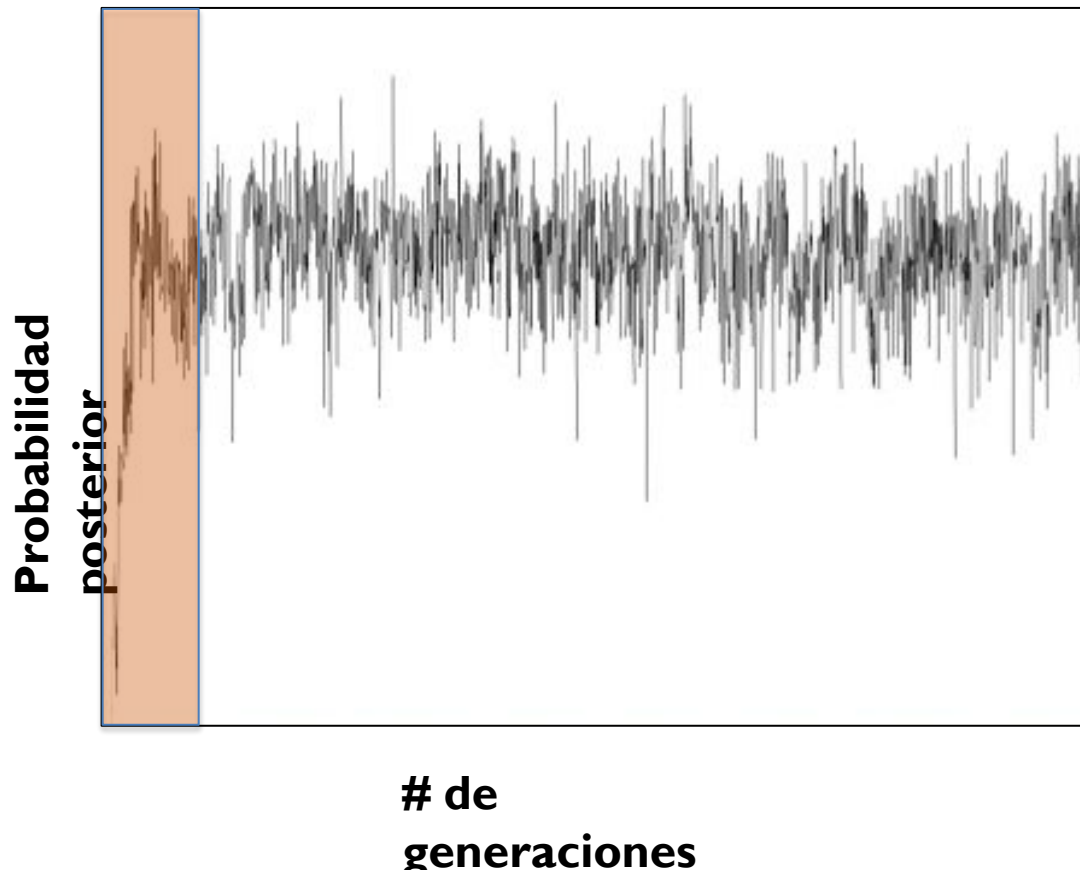
I. Selección *a priori* de un modelo de sustitución de caracteres

SOLUCIÓN: Salto entre modelos (reverse-jump MCMC)

Complicaciones del MCMC

2. La cadena del MCMC necesita alcanzar estacionalidad

SOLUCIÓN: Burn-in



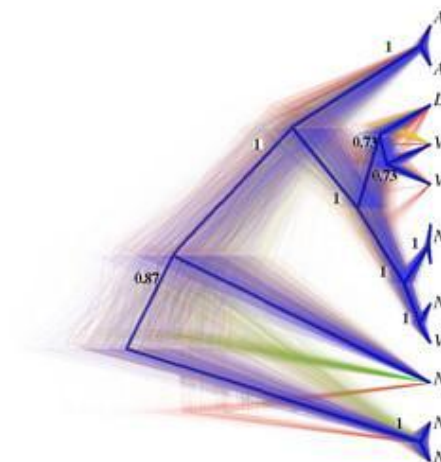
Complicaciones del MCMC

3. Es necesario garantizar que durante el periodo de estacionalidad la cadena haya explorado todo el espacio de parámetros (“mixing”).

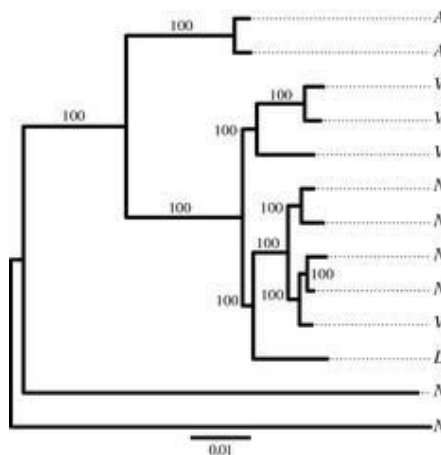
- Estrategia 1: Varias corridas independientes
- Estrategia 2: Modificar la forma en que nuevos puntos de parámetros son propuestos: Cadenas calientes y cadena fría (Metropolis-Coupled)

¿Cómo entender los resultados de MCMC?

En la zona estacionaria hay muchos árboles con longitudes de ramas y topología similares



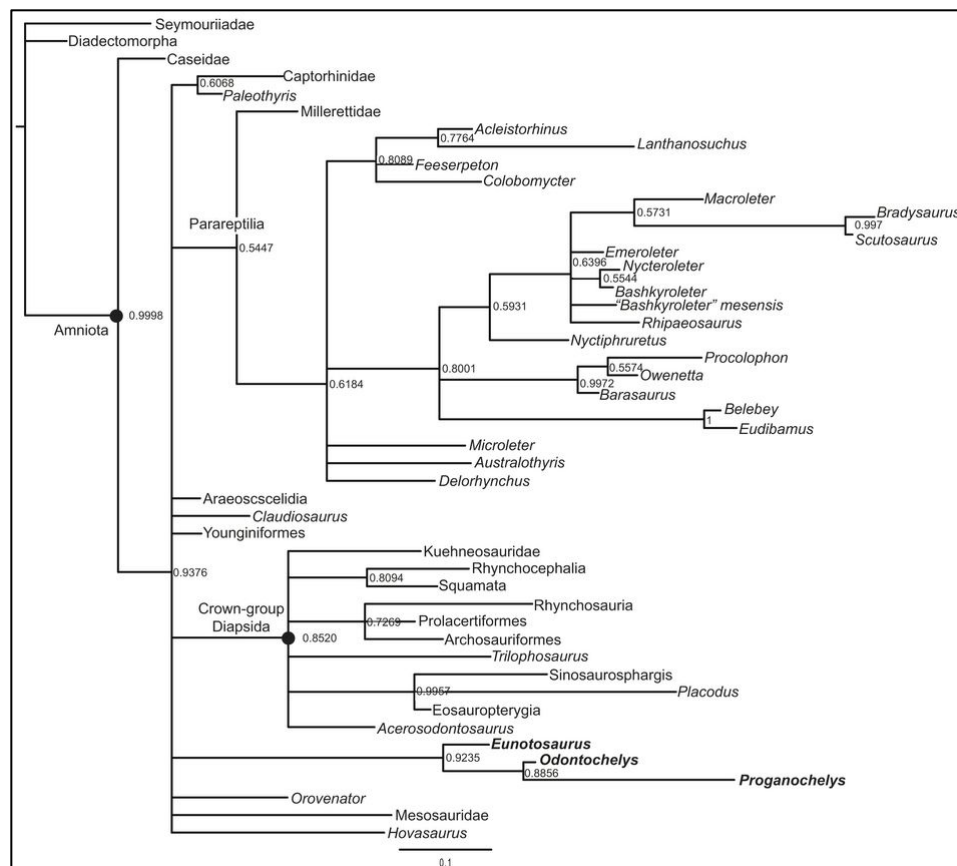
Opción 1: observar todos los árboles



Opción 2: Árbol de máxima credibilidad

¿Cómo entender los resultados de MCMC?

En la zona estacionaria hay muchos árboles con longitudes de ramas y topología similares



Opción 3: Árbol de 50% consenso de mayoría

¿Cómo entender los resultados de MCMC?

3. Probabilidad Posterior de los clados como medida de soporte

