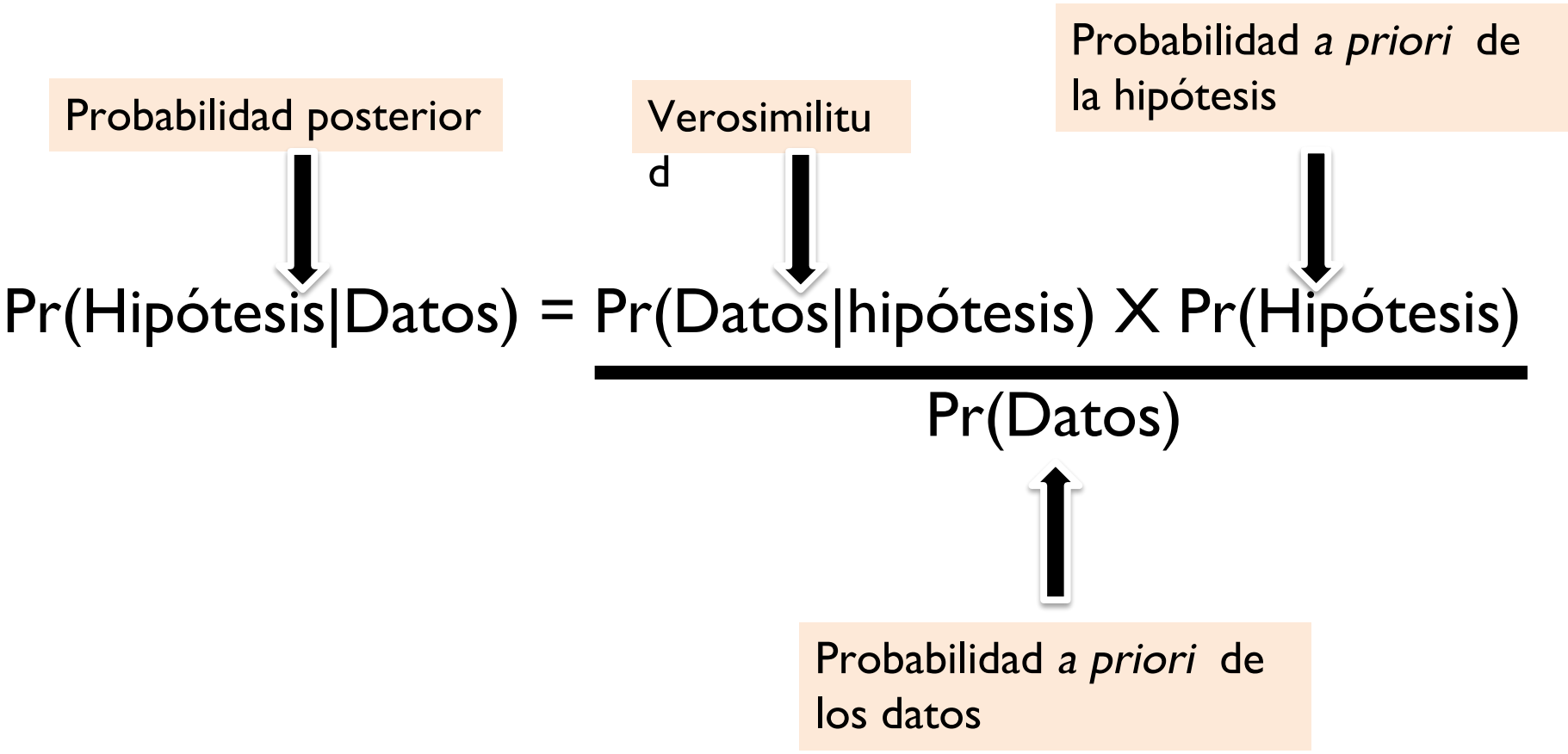
INFERENCIA FILOGENÉTICA

INFERENCIA BAYESIANA

Teorema de Bayes



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

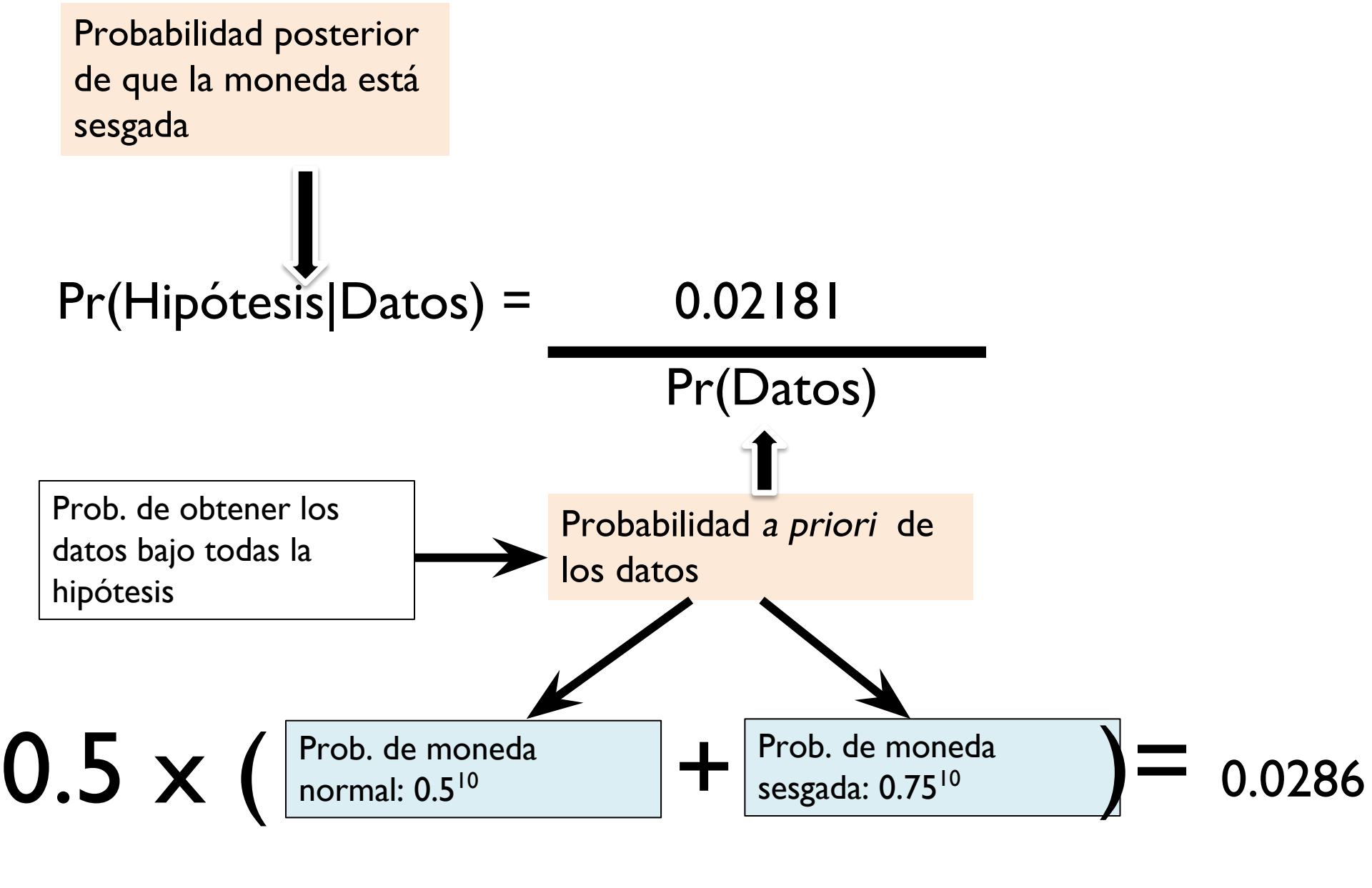
Verosimilitud

Probabilidad *a priori* de que la moneda está sesgada

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

Probabilidad *a priori* de los datos

Teorema de Bayes



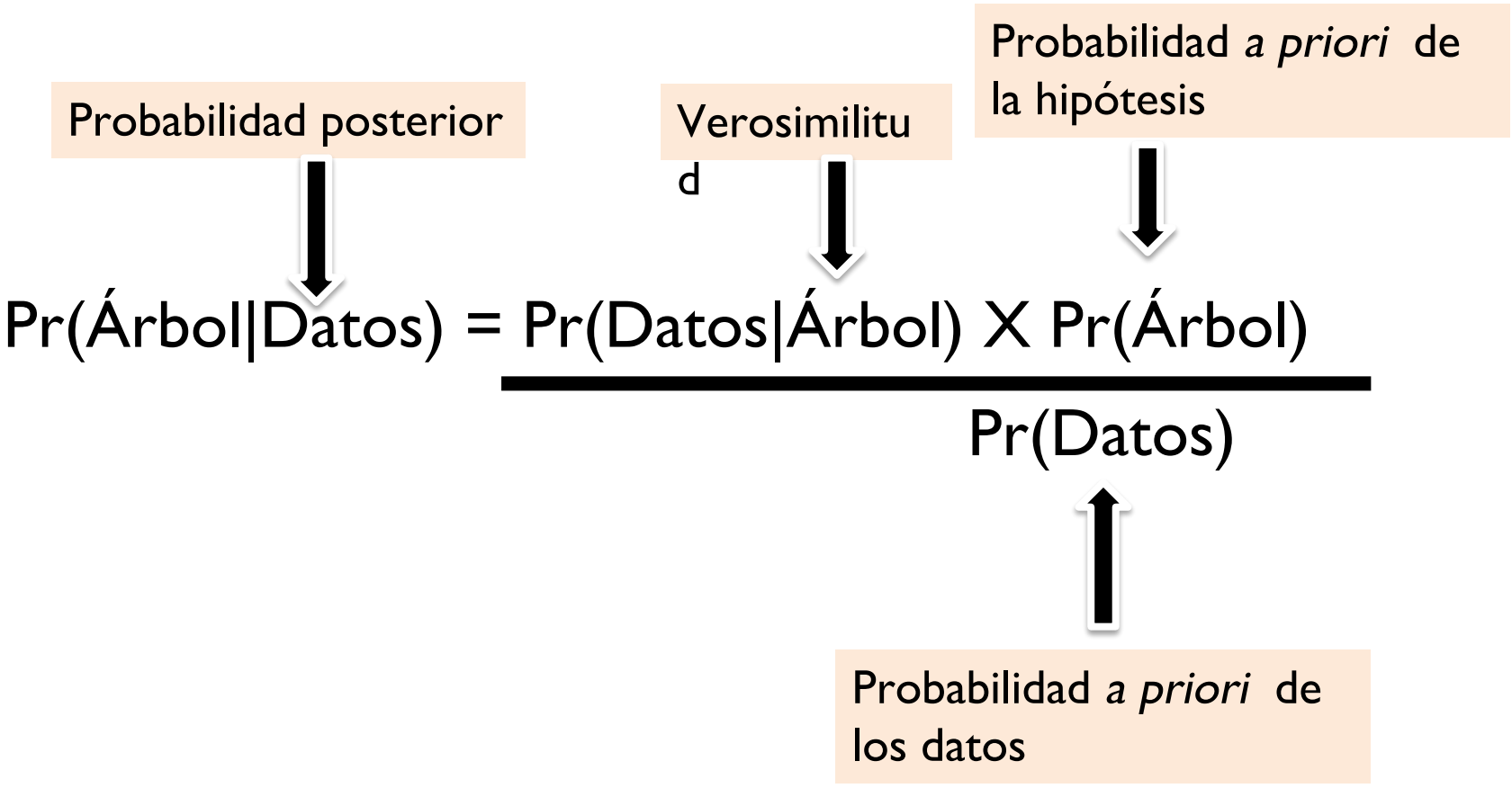
Teorema de Bayes

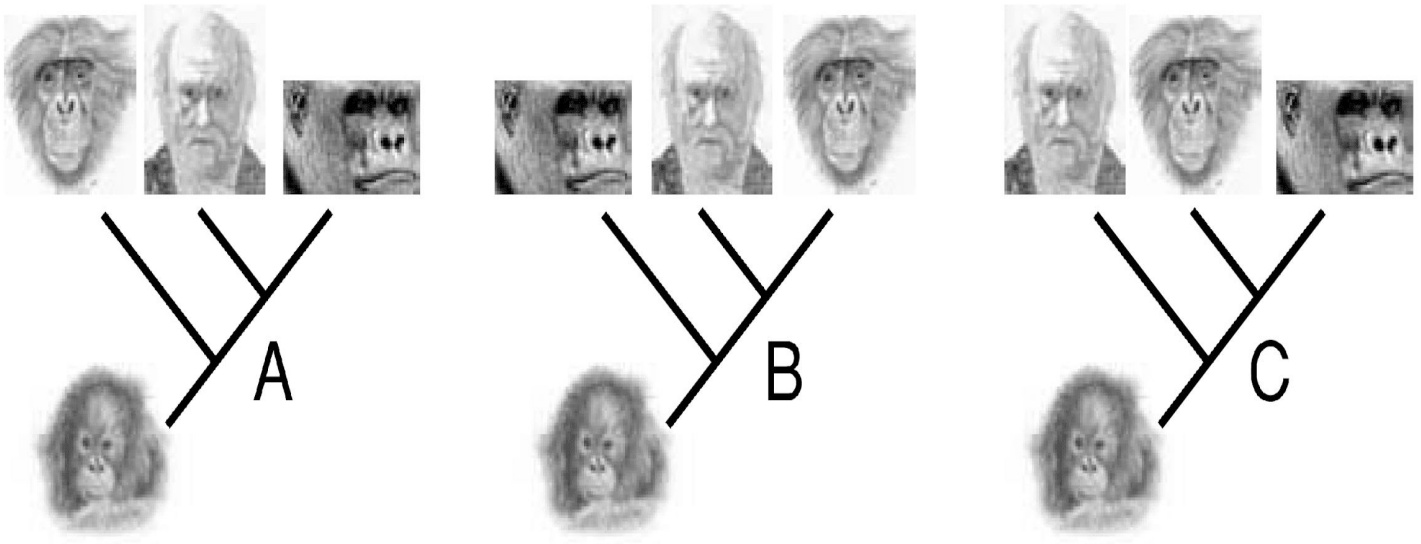
Probabilidad posterior de que la moneda está sesgada

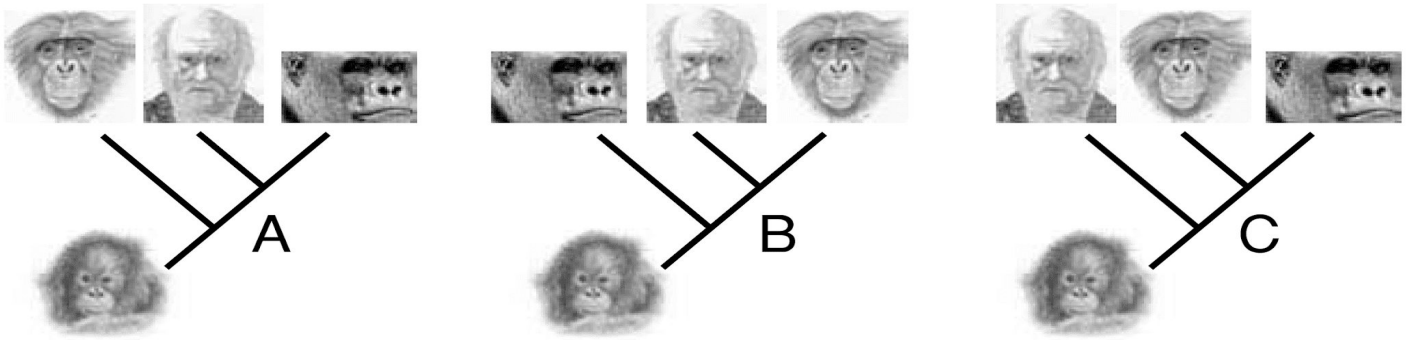
$$\Pr(\text{Hipótesis}|\text{Datos}) = \frac{0.02181}{0.0286} = 0.98$$

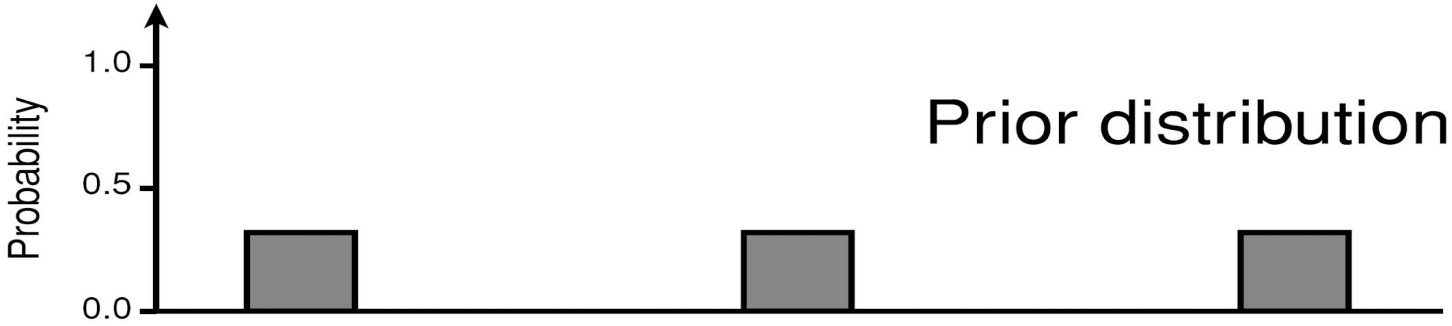
	Probabilidad *a priori*		Probabilidad posterior
Sesgada	**0.5**	→	**0.98**
Normal	**0.5**	→	**0.02**

INFERENCIA BAYESIANA EN FILOGENÉTICA

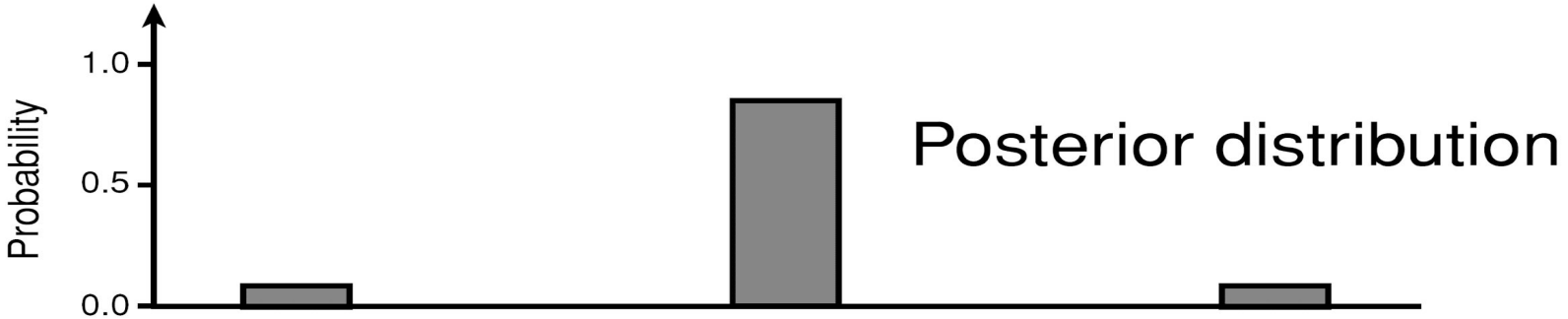


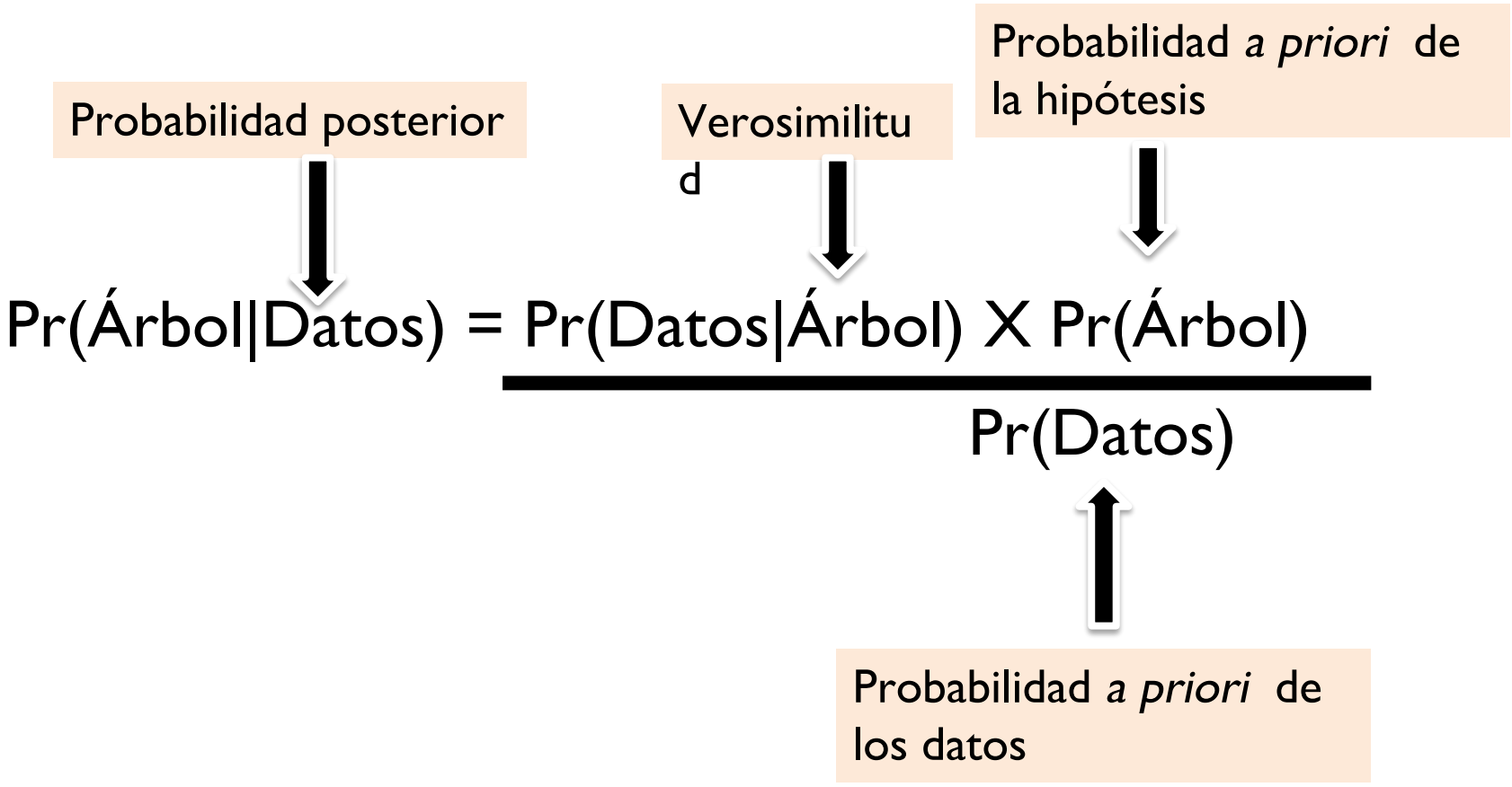

A
B
C





Datos y modelo de sustitución




Probabilidad posterior
Verosimilitud
Probabilidad *a priori* de la hipótesis
Pr(Árbol|Datos) = Pr(Datos|Árbol) X Pr(Árbol)
Pr(Datos)
Probabilidad *a priori* de los datos

INFERENCIA BAYESIANA EN FILOGENÉTICA

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

Pr(Datos)

Probabilidad *a priori* de los datos

Cadena de Markov Monte Carlo (MCMC)

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

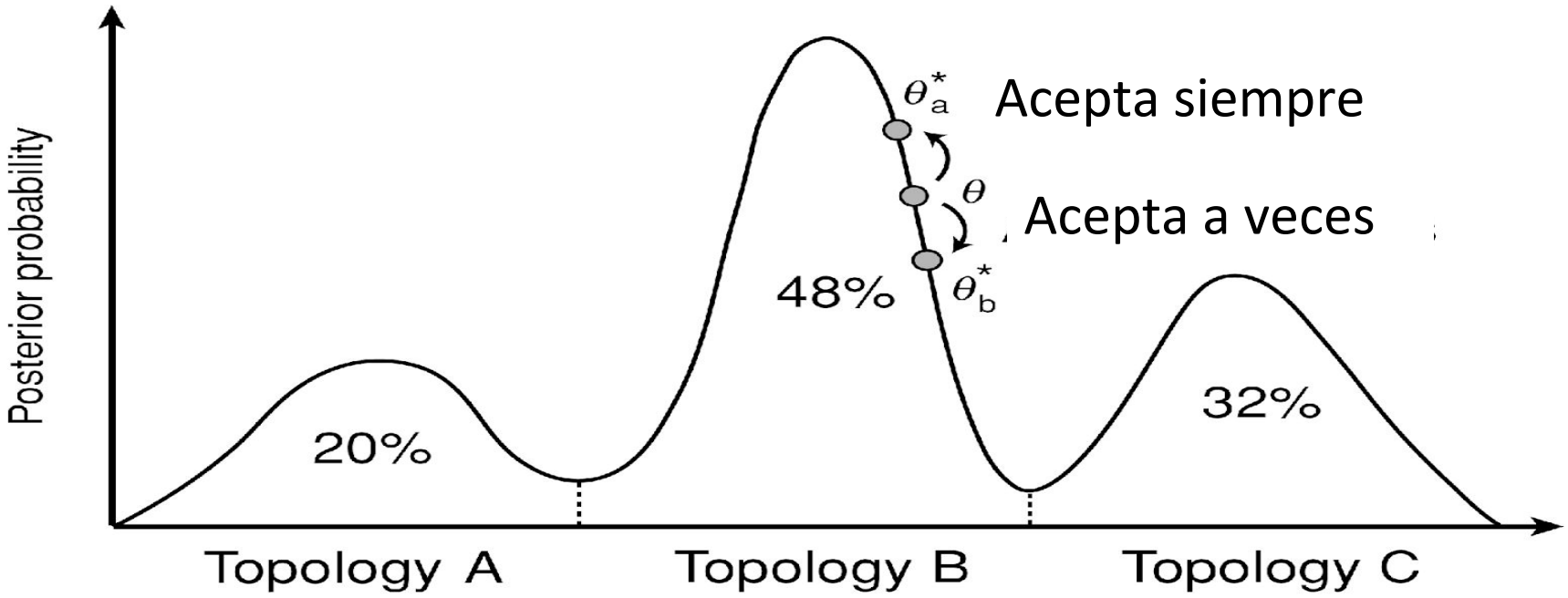
1. Comenzamos en un punto arbitrario de parámetros (θ)

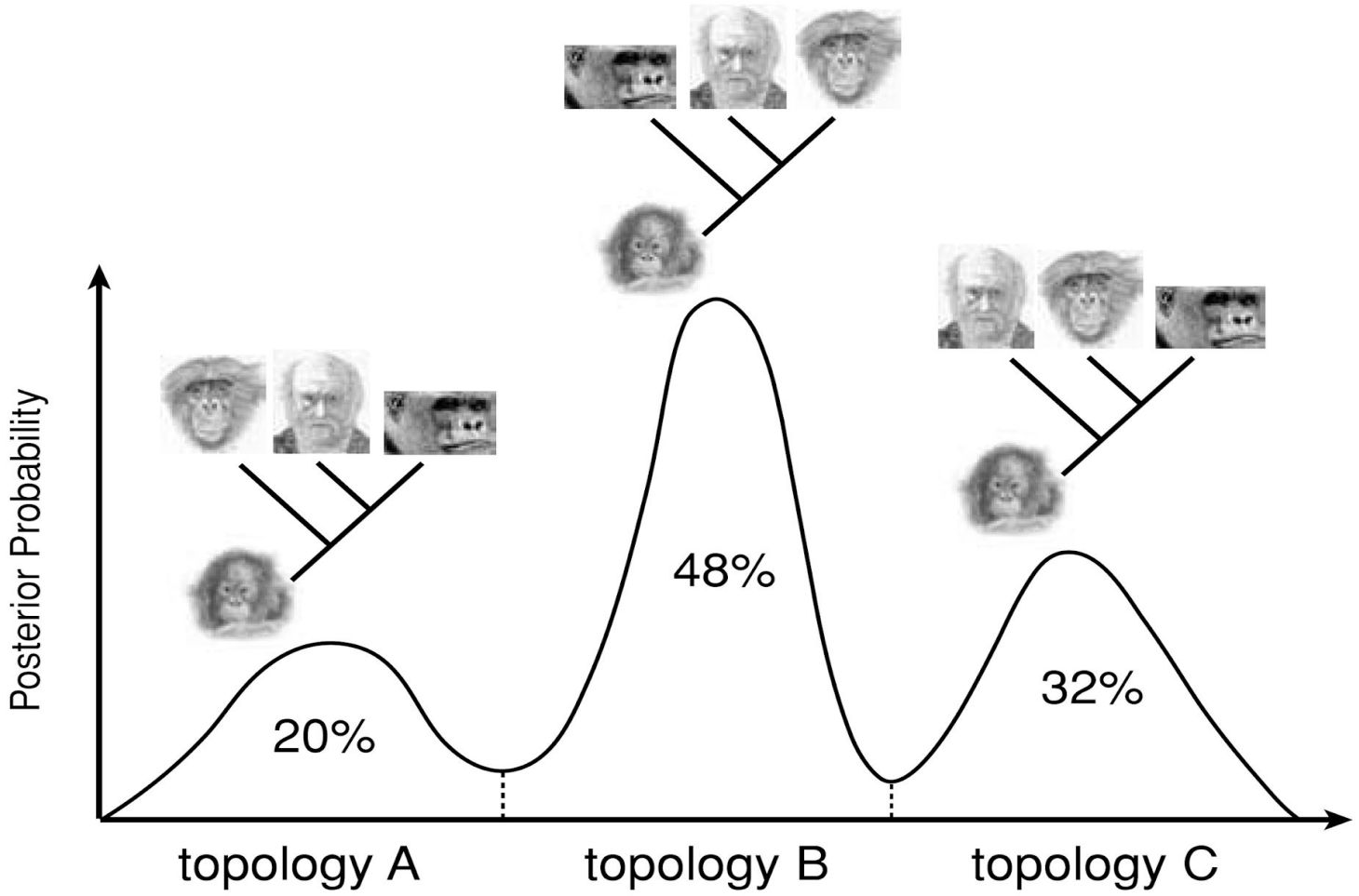
2. Se hace un movimiento aleatorio hacia θ'

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

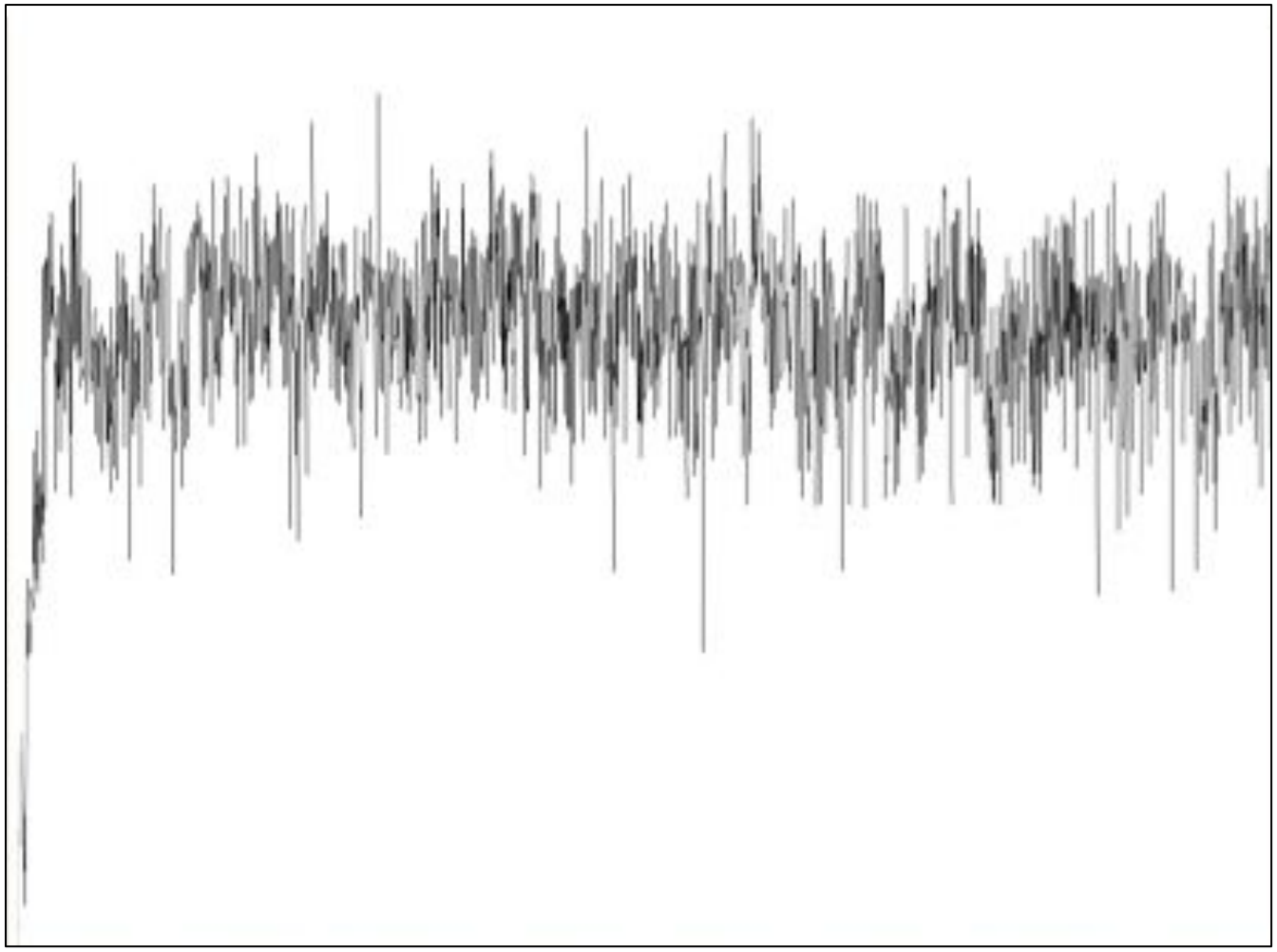
4. Volvemos al paso 2 y repetimos millones de veces (generaciones)

5. Guardar árbol y parámetros cada n generaciones


Posterior probability
θ^*_a
Acepta siempre
θ
Acepta a veces
48%
θ^*_b
20%
32%
Topology A
Topology B
Topology C


Posterior Probability
20%
48%
32%
topology A
topology B
topology C

Cadena de Markov Monte Carlo (MCMC)



Complicaciones del MCMC

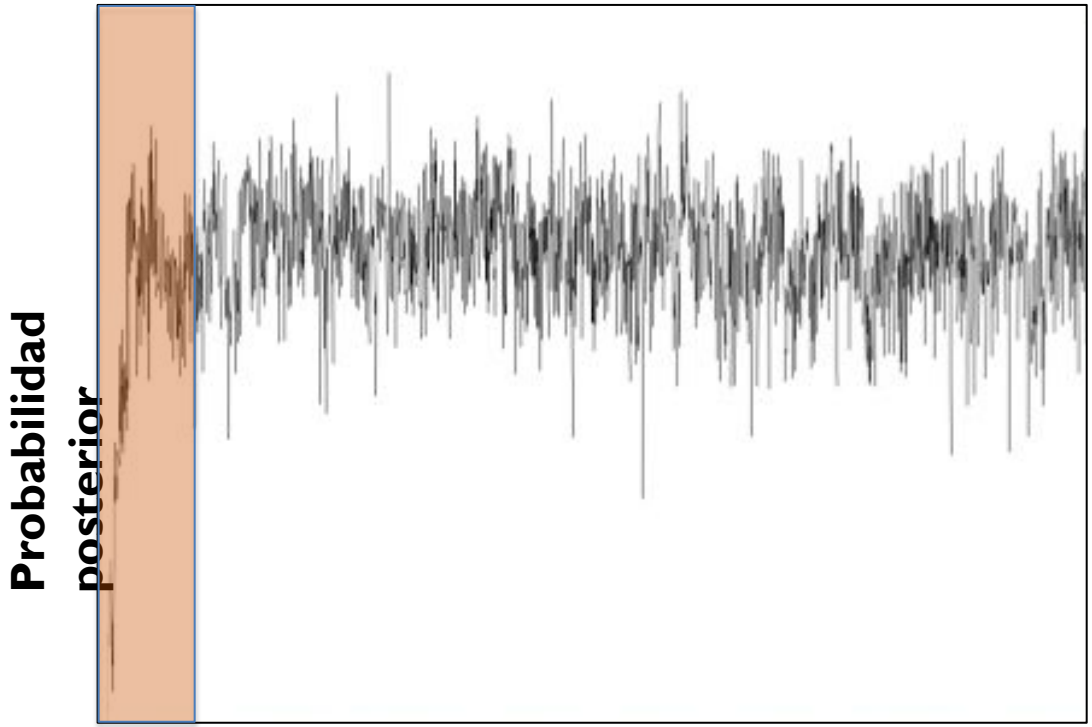
1. Seleccion *a priori* de un modelo de sustitución de caracteres

SOLUCIÓN: Salto entre modelos (reverse-jump MCMC)

Complicaciones del MCMC

2. La cadena del MCMC necesita alcanzar estacionalidad

SOLUCIÓN: Burn-in



Complicaciones del MCMC

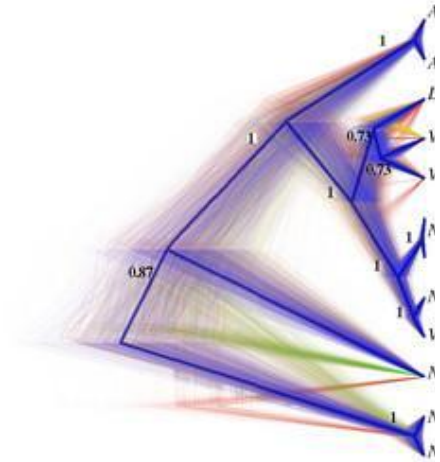
3. Es necesario garantizar que durante el periodo de estacionalidad la cadena haya explorado todo el espacio de parámetros ("mixing").

- Estrategia 1: Varias corridas independientes

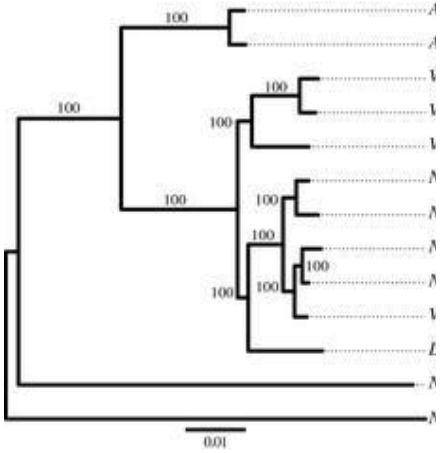
- Estrategia 2: Modificar la forma en que nuevos puntos de parámetros son propuestos: Cadenas calientes y cadena fría (Metropolis-Coupled)

¿Cómo entender los resultados de MCMC?

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



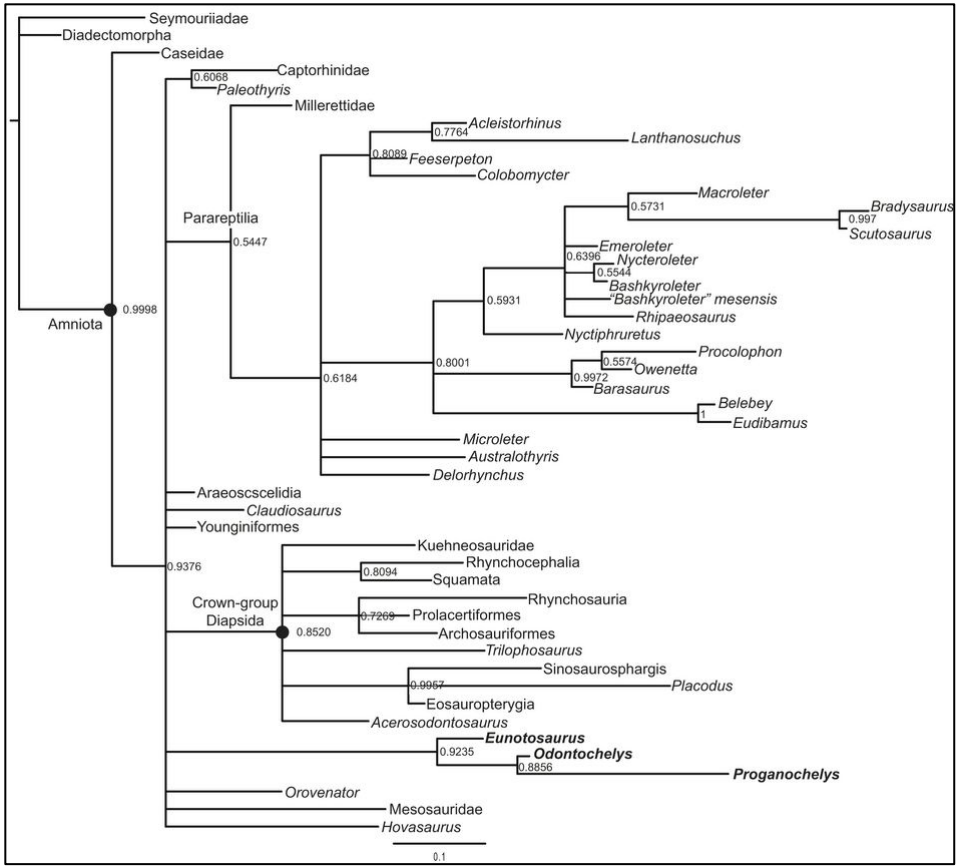
Opción 1: observar todos los árboles



Opción 2: Árbol de máxima credibilidad

¿Cómo entender los resultados de MCMC?

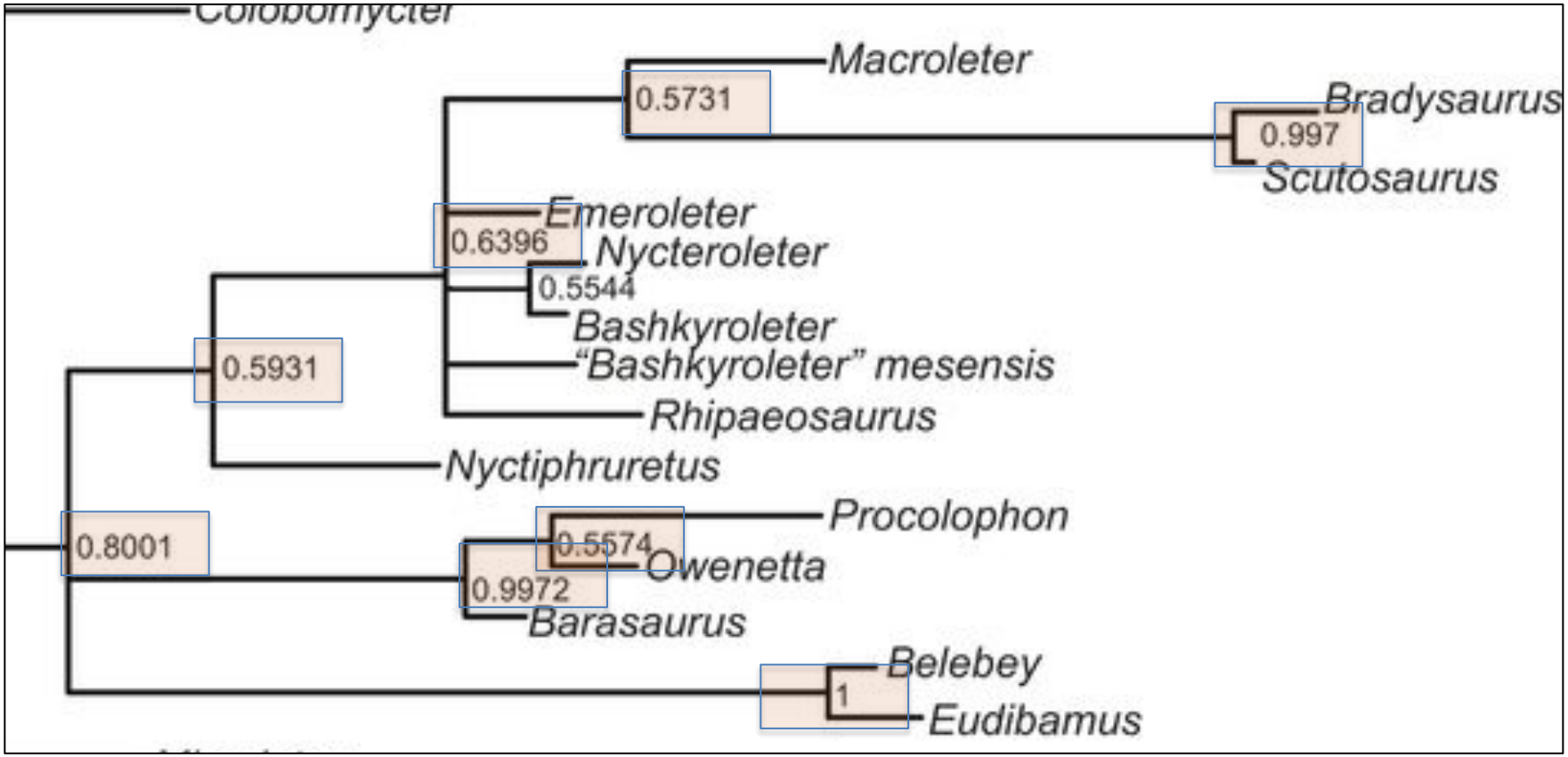
En la zona estacionaria hay muchos árboles con longitudes de ramas y topología similares



Opción 3: Árbol de 50% consenso de mayoría

¿Cómo entender los resultados de MCMC?

3. Probabilidad Posterior de los clados como medida de soporte



CONFIANZA EN HIPÓTESIS FILOGENÉTICAS

MEDIDAS DE SOPORTE DE LOS CLADOS

1. Bootstrap No Paramétrico

Original data set

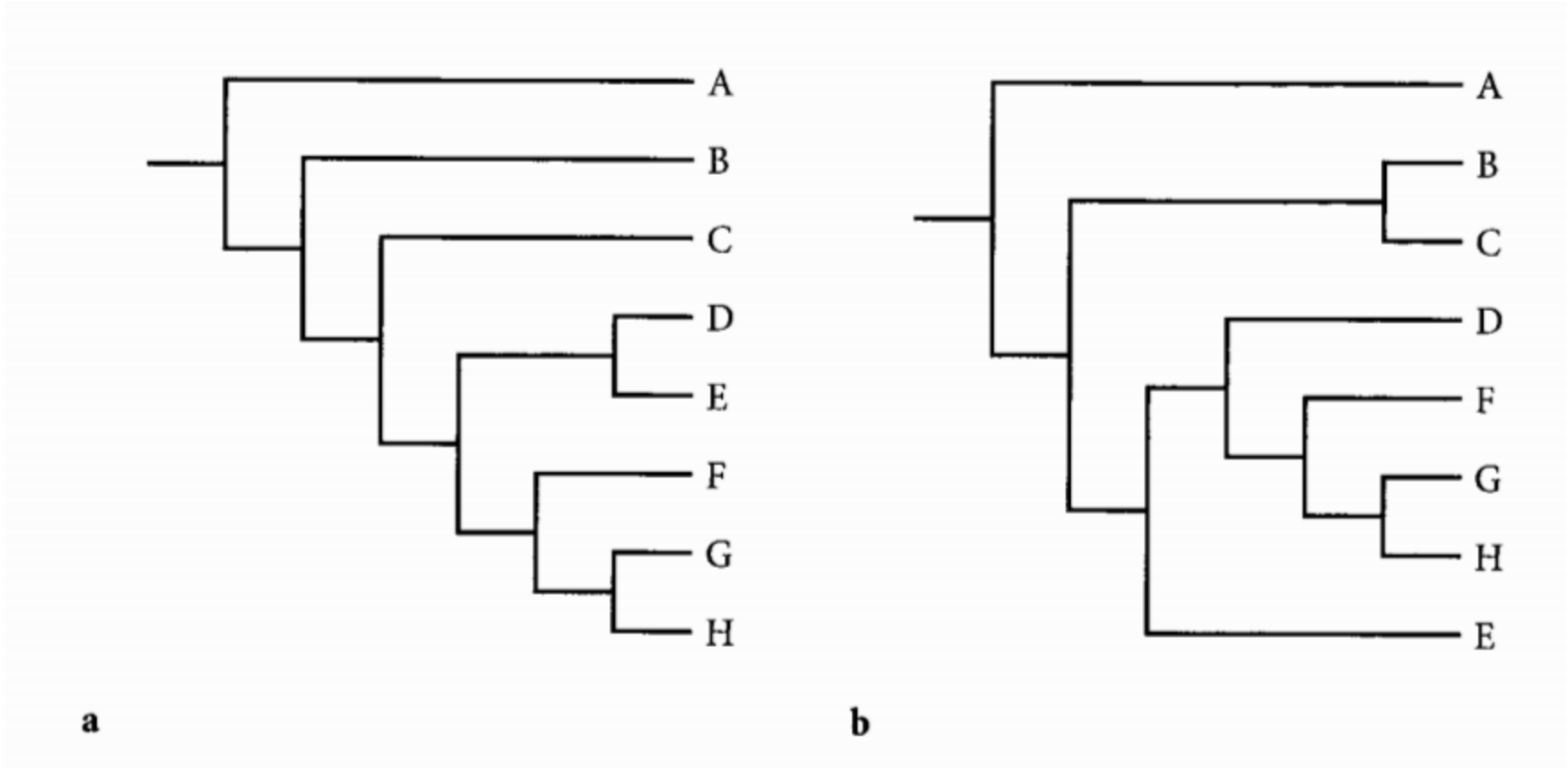
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
A	T	T	T	C	C	T	T	T	C	A	G	G	T	A	T	T	A	T	G	A	G	A	T	A	C	G	T	A	C	T	G	A	A	A	A	A	G	T	C	C
B	T	T	T	C	C	T	T	T	T	A	G	G	T	T	T	G	A	T	G	A	G	A	T	A	C	A	T	T	A	C	G	A	A	A	G	A	G	T	C	A
C	T	T	T	G	C	T	T	C	T	C	G	G	T	A	C	T	A	C	A	A	T	A	T	A	T	A	T	A	C	C	A	G	A	A	A	A	G	T	C	A
D	T	T	T	G	C	T	T	C	C	G	A	C	T	A	C	A	A	A	G	G	C	A	T	A	C	G	T	A	G	C	T	G	A	A	A	A	G	G	C	G
E	C	T	T	G	C	C	T	A	C	T	G	T	T	G	C	A	A	T	A	A	T	A	T	A	C	G	A	A	G	C	T	A	A	A	A	A	G	T	C	G
F	T	T	C	G	T	C	C	C	C	G	G	C	T	A	C	A	A	T	G	G	T	A	T	A	T	G	T	A	C	T	C	G	A	A	A	A	G	A	T	G
G	G	T	T	G	T	T	T	C	C	G	G	C	T	A	C	A	G	T	G	A	T	A	T	A	C	G	T	A	C	C	C	G	A	G	A	A	C	T	T	G
H	T	T	T	A	T	T	T	C	C	G	G	C	T	A	C	A	G	T	G	A	T	A	T	A	C	G	T	G	C	C	C	G	A	G	A	A	G	T	T	G

Bootstrap data set

	02	39	35	22	36	31	40	05	16	23	15	35	35	40	03	06	24	33	06	07	14	20	35	01	36	09	13	22	11	25	26	33	03	09	16	20	08	18	17	32
A	T	C	A	A	A	G	C	C	T	T	T	A	A	C	T	T	A	A	T	T	A	A	A	T	A	C	T	A	G	C	G	A	T	C	T	A	T	T	A	A
B	T	C	G	A	A	G	A	C	G	T	T	G	G	A	T	T	A	A	T	T	T	A	G	T	A	T	T	A	G	C	A	A	T	T	G	A	T	T	A	A
C	T	C	A	A	A	A	A	C	T	T	C	A	A	A	T	T	A	A	T	T	A	A	A	T	A	T	T	A	G	T	A	A	T	T	T	A	C	C	A	G
D	T	C	A	A	A	T	G	C	A	T	C	A	A	G	T	T	A	A	T	T	A	G	A	T	A	C	T	A	A	C	G	A	T	C	A	G	C	A	A	G
E	T	C	A	A	A	T	G	C	A	T	C	A	A	G	T	C	A	A	C	T	G	A	A	C	A	C	T	A	G	C	G	A	T	C	A	A	A	T	A	A
F	T	T	A	A	A	C	G	T	A	T	C	A	A	G	C	C	A	A	C	C	A	G	A	T	A	C	T	A	G	T	G	A	C	C	A	G	C	T	A	G
G	T	T	A	A	A	C	G	T	A	T	C	A	A	G	T	T	A	A	T	T	A	A	A	G	A	C	T	A	G	C	G	A	T	C	A	A	C	T	G	G
H	T	T	A	A	A	C	G	T	A	T	C	A	A	G	T	T	A	A	T	T	A	A	A	T	A	C	T	A	G	C	G	A	T	C	A	A	C	T	G	G

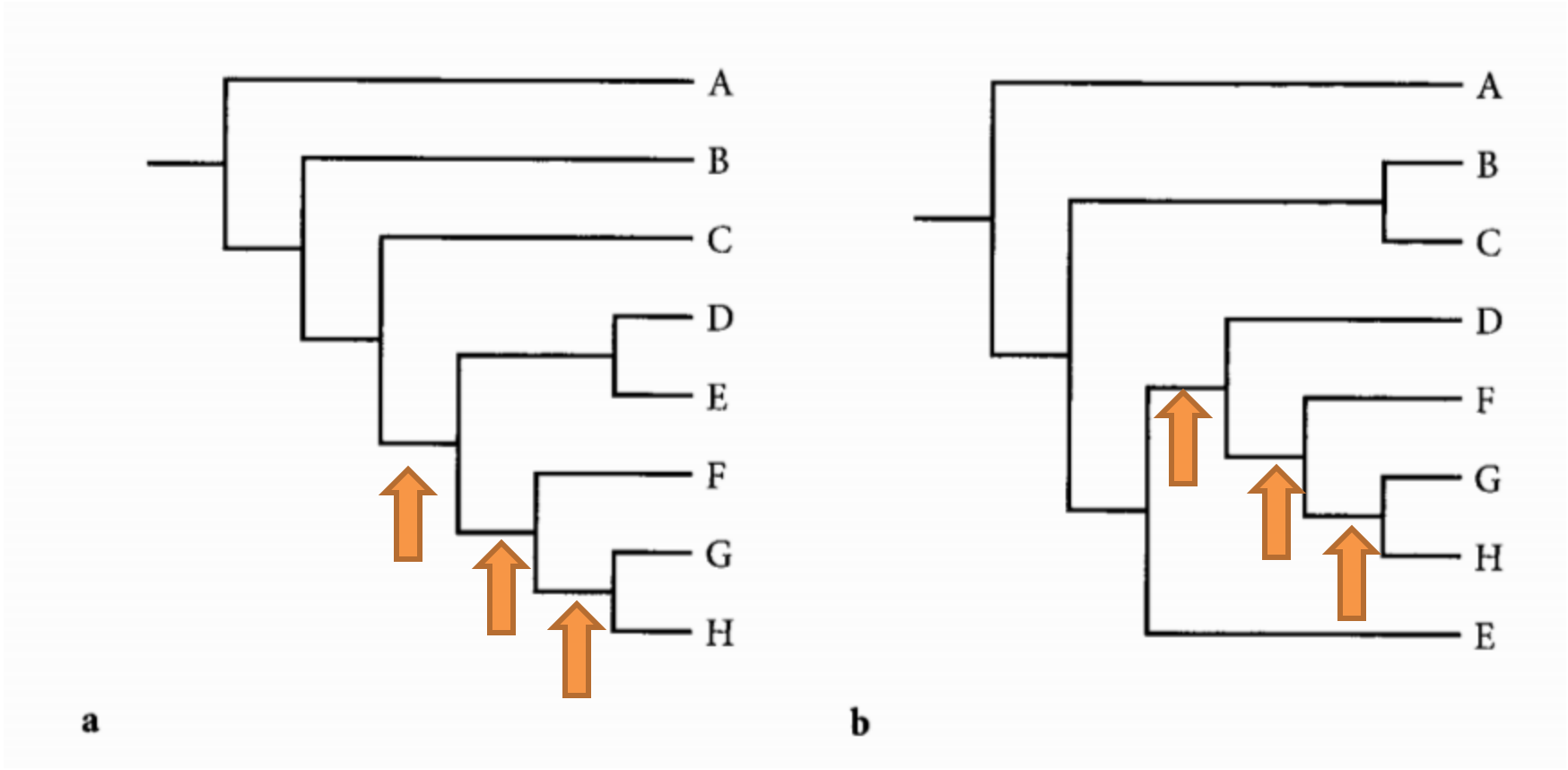
MEDIDAS DE SOPORTE DE LOS CLADOS

1. Bootstrap No Paramétrico



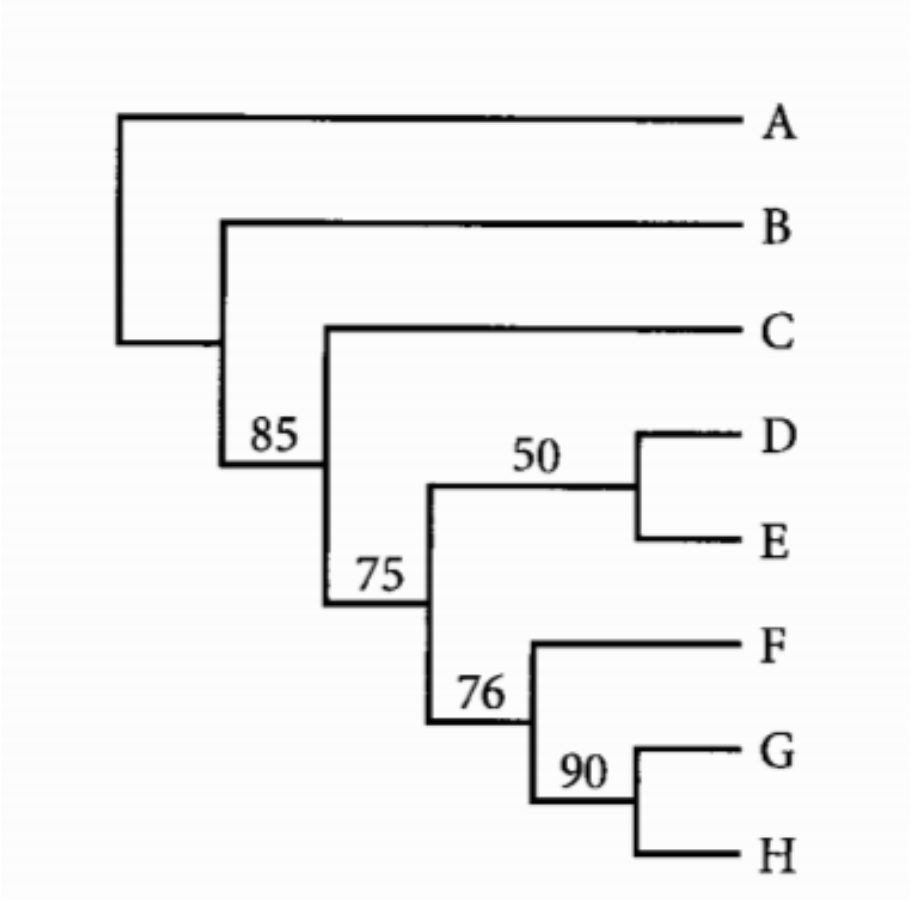
MEDIDAS DE SOPORTE DE LOS CLADOS

1. Bootstrap No Paramétrico



MEDIDAS DE SOPORTE DE LOS CLADOS

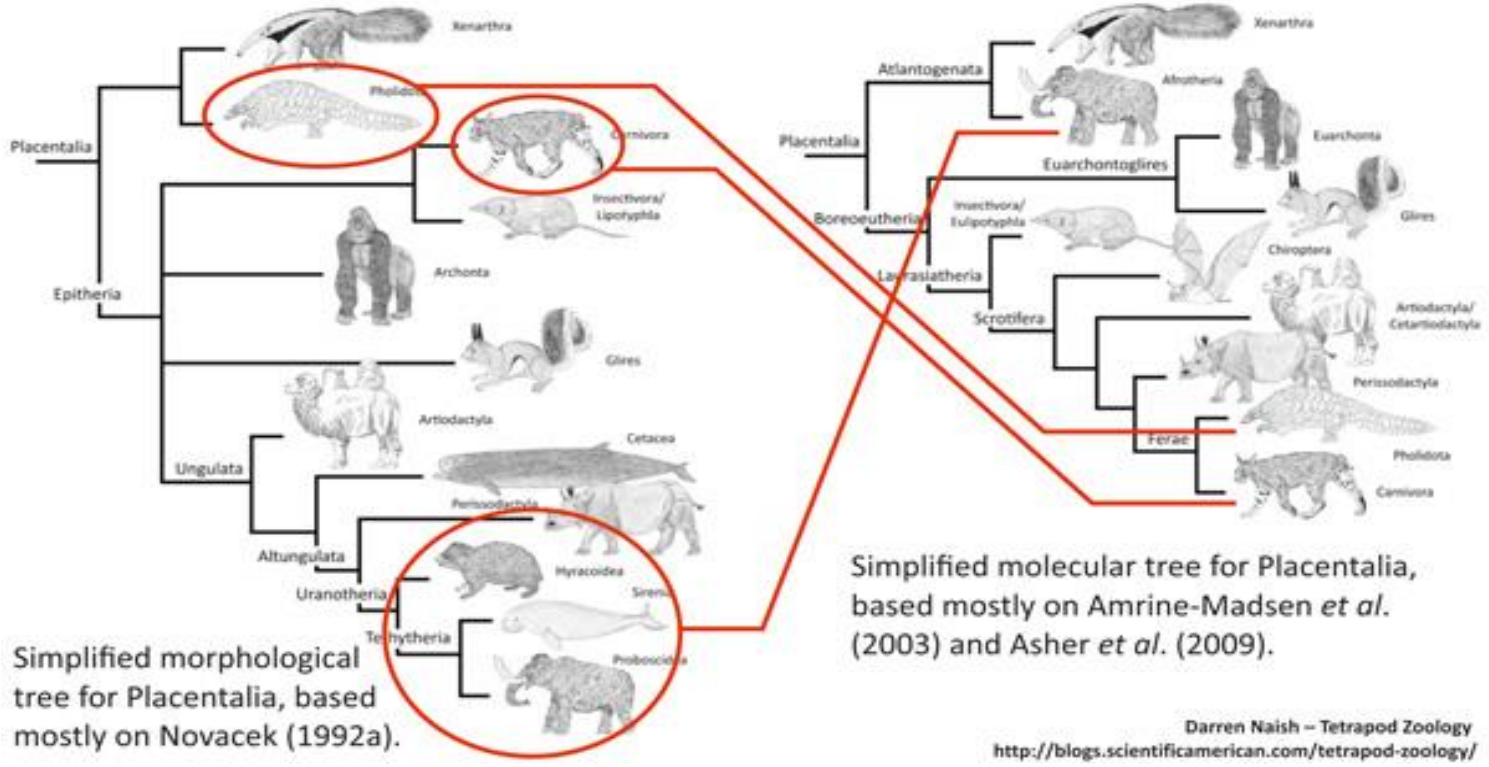
1. Bootstrap No Paramétrico

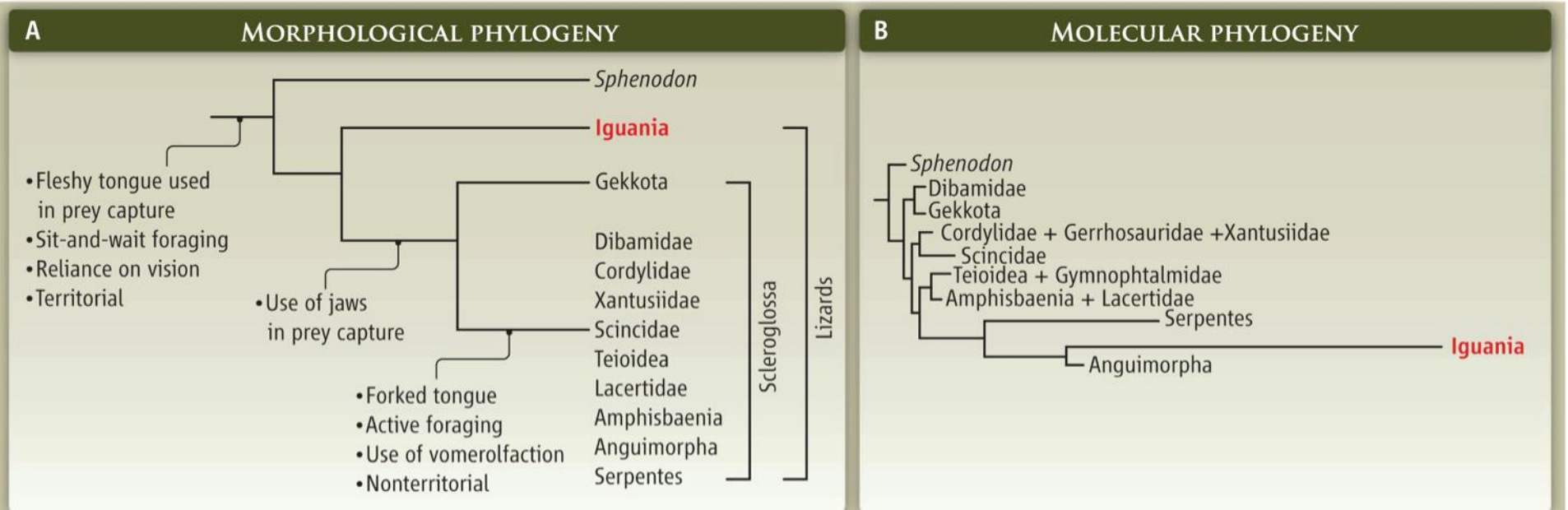


CONFLICTO EN HIPÓTESIS FILOGENÉTICAS

Particiones a veces generan hipótesis conflictivas

- Morfología vs. molecular
- Genomas diferentes
- Genes codifcadores y no codificadores
- Posiciones en codón
- Intrón vs. exón
- Proteína intra vs. extracelular


Xenarthra
Pholidota
Carnivora
Insectivora/
Lipotyphla
Placentalia
Archonta
Epitheria
Glires
Artiodactyla
Cetacea
Ungulata
Perissodactyla
Hyracoidea
Altungulata
Sirenia
Uranotheria
Tethytheria
Proboscidea
Simplified morphological tree for Placentalia, based mostly on Novacek (1992a).
Xenarthra
Atlantogenata
Afrotheria
Euarchonta
Placentalia
Euarchontoglires
Boreoeutheria
Insectivora/
Eulipotyphla
Glires
Laurasiatheria
Chiroptera
Scrotifera
Artiodactyla/
Cetartiodactyla
Perissodactyla
Ferae
Pholidota
Carnivora
Simplified molecular tree for Placentalia, based mostly on Amrine-Madsen et al. (2003) and Asher et al. (2009).
Darren Naish – Tetrapod Zoology
http://blogs.scientificamerican.com/tetrapod-zoology/


A
MORPHOLOGICAL PHYLOGENY
Sphenodon
Iguania
Gekkota
Dibamidae
Cordylidae
Xantusiidae
Scincidae
Teioidea
Lacertidae
Amphisbaenia
Anguimorpha
Serpentes
Scleroglossa
Lizards
•Fleshy tongue used in prey capture
•Sit-and-wait foraging
•Reliance on vision
•Territorial
•Use of jaws in prey capture
•Forked tongue
•Active foraging
•Use of vomerolfaction
•Nonterritorial
B
MOLECULAR PHYLOGENY
Sphenodon
Dibamidae
Gekkota
Cordylidae + Gerrhosauridae +Xantusiidae
Scincidae
Teioidea + Gymnophtalmidae
Amphisbaenia + Lacertidae
Serpentes
Iguania
Anguimorpha

CONFLICTO EN HIPÓTESIS FILOGENÉTICAS

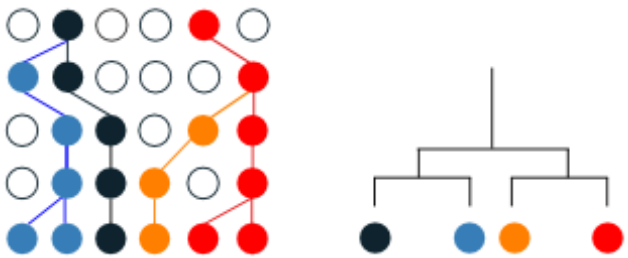
Razones de conflicto

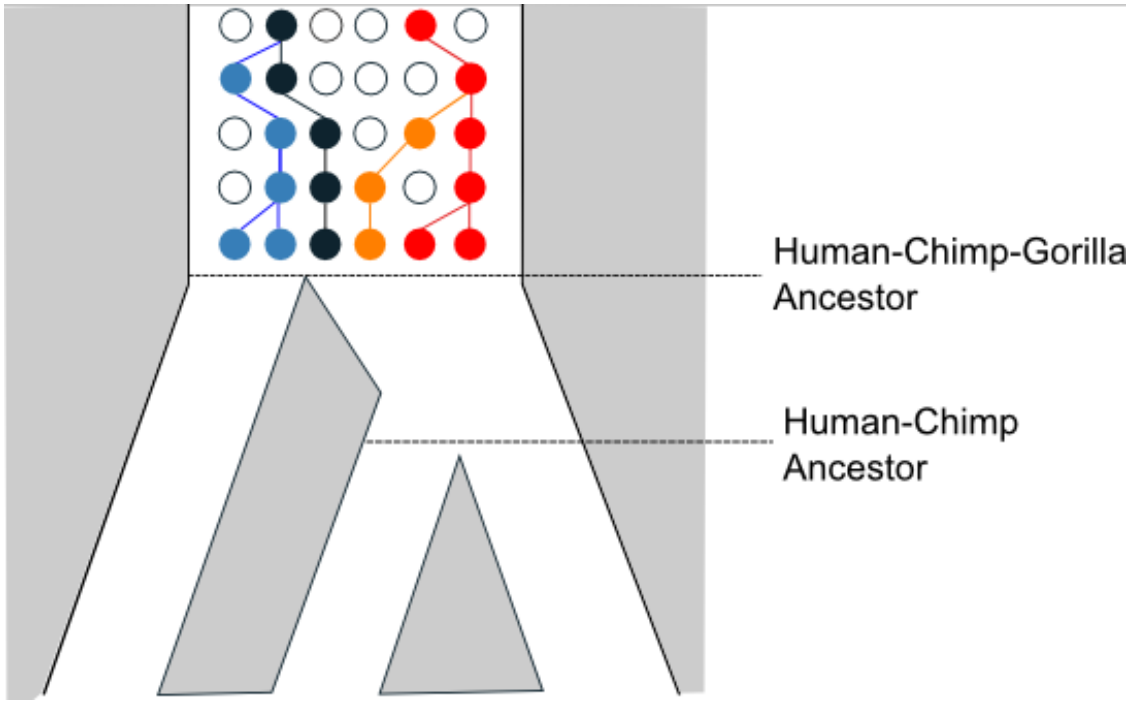
- Metodológicas
  - Contaminación
  - Mala identificación
  - Errores de laboratorio/computacional
  - Genes parálogos

- Biológicas:

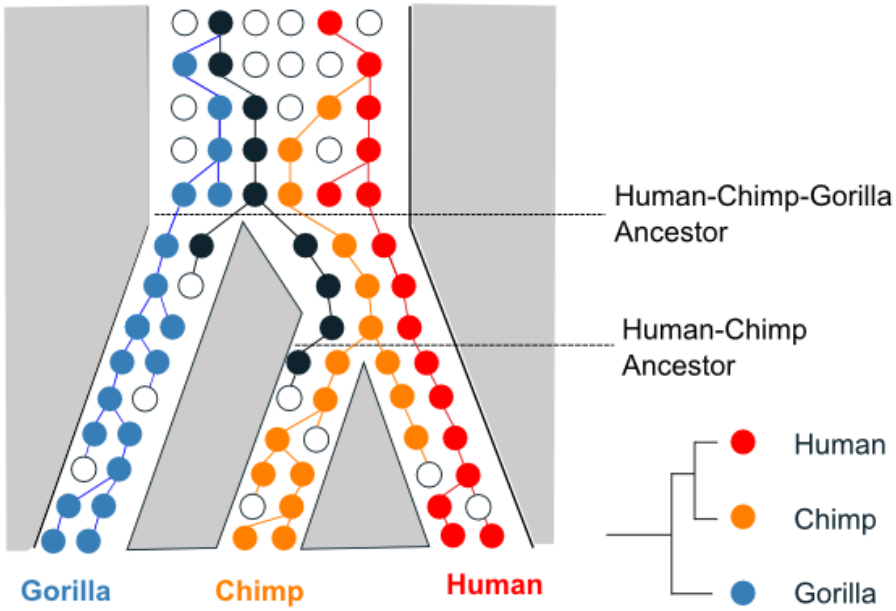
  - Separación incompleta de linajes
  - Introgresión
  - Transferencia Horizontal de genes

SEPARACIÓN INCOMPLETA DE LINAJES



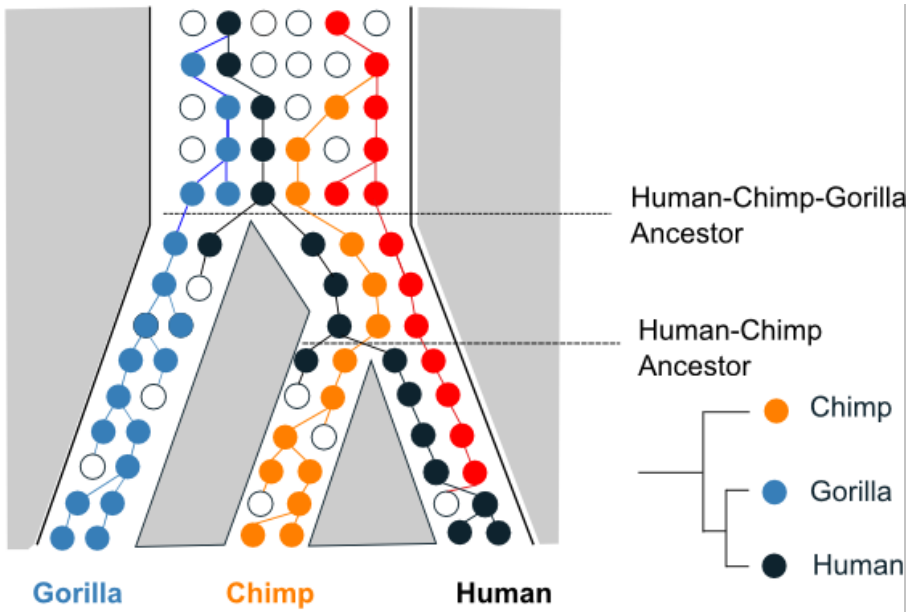


SEPARACIÓN INCOMPLETA DE LINAJES

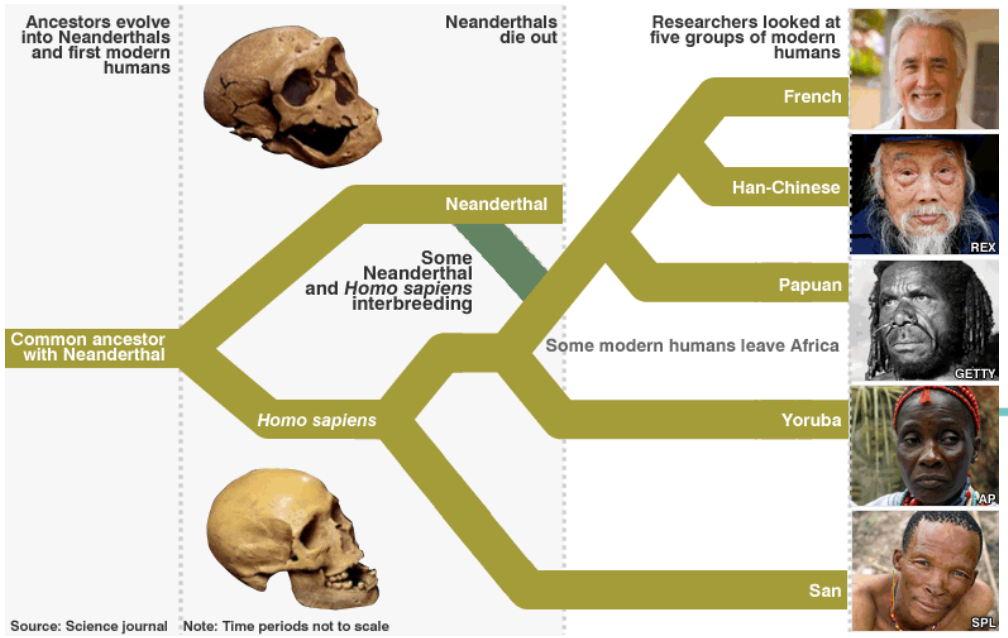


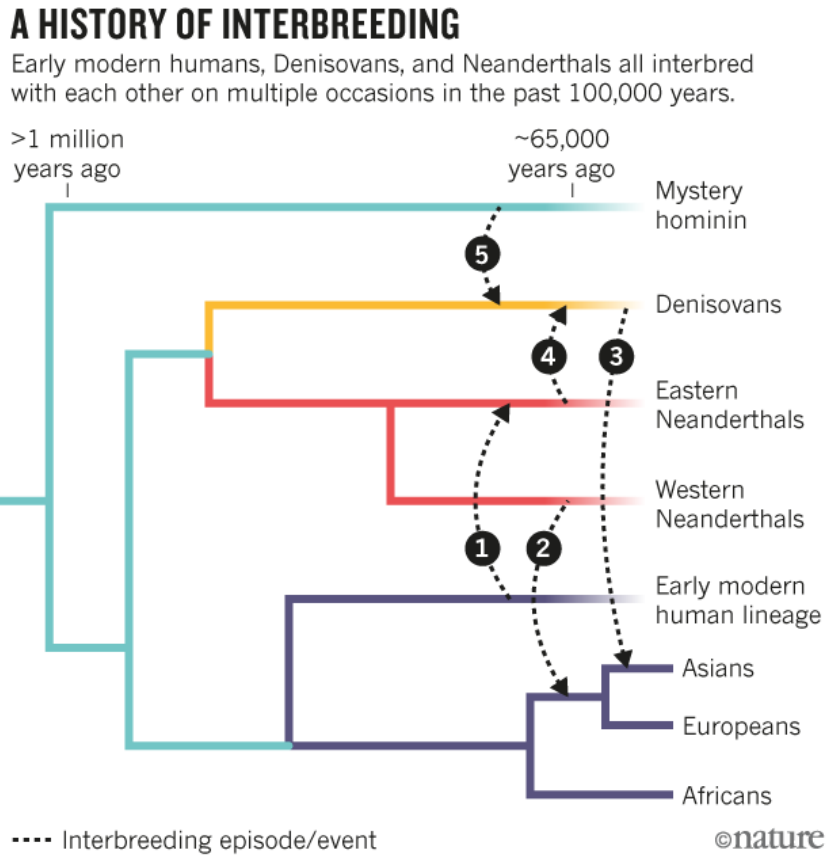
Separación completa

Separación incompleta

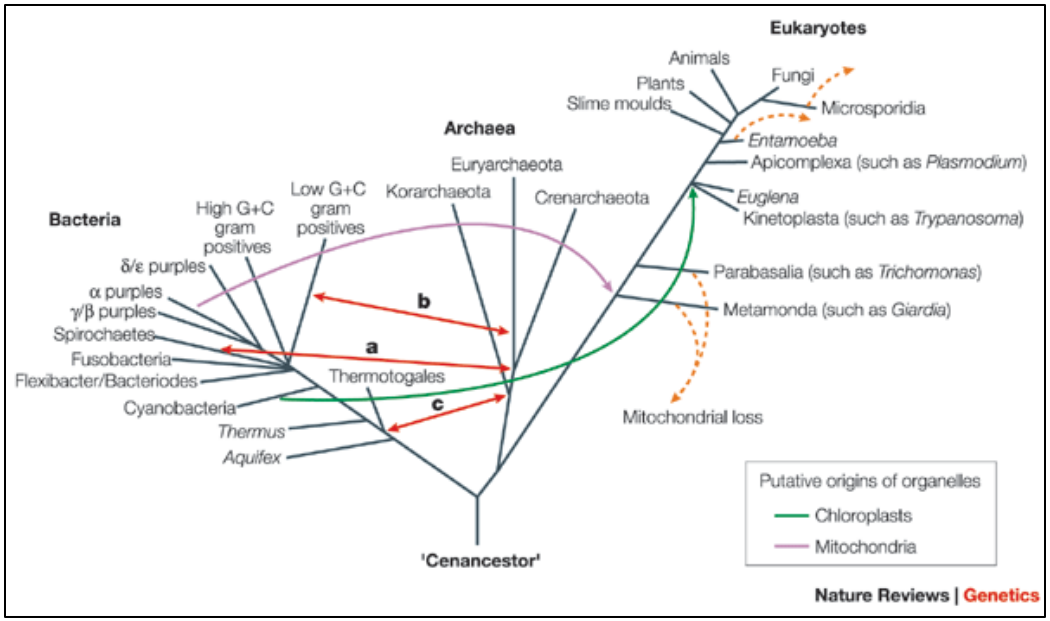


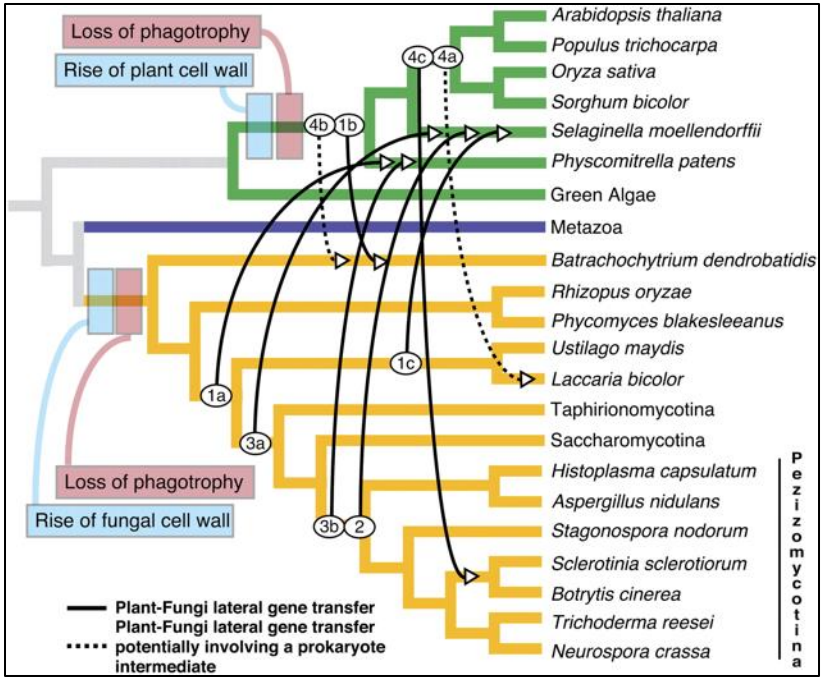
INTROGRESIÓN





TRANSFERENCIA HORIZONTAL DE GENES

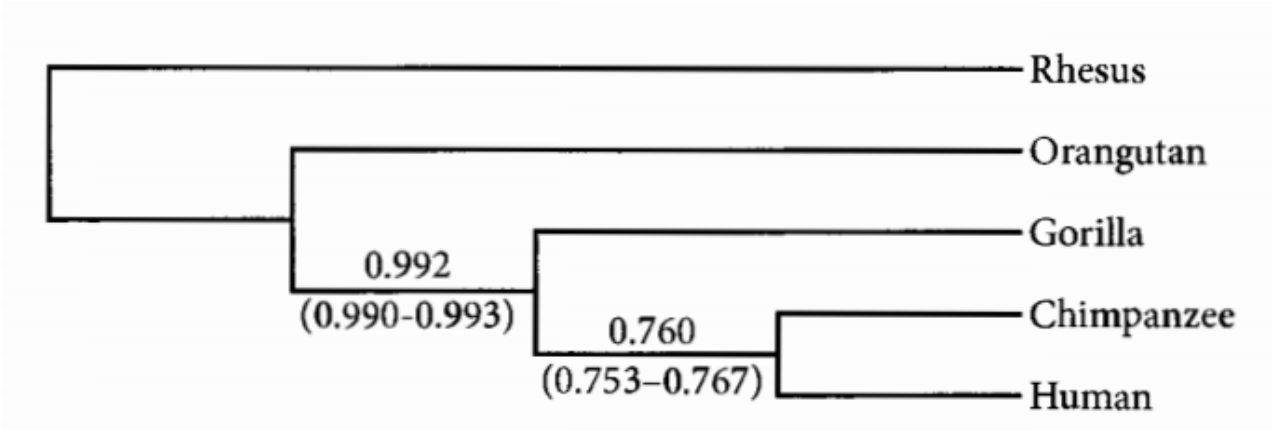




PUESTA A PRUEBA DE HIPÓTESIS FILOGENÉTICAS

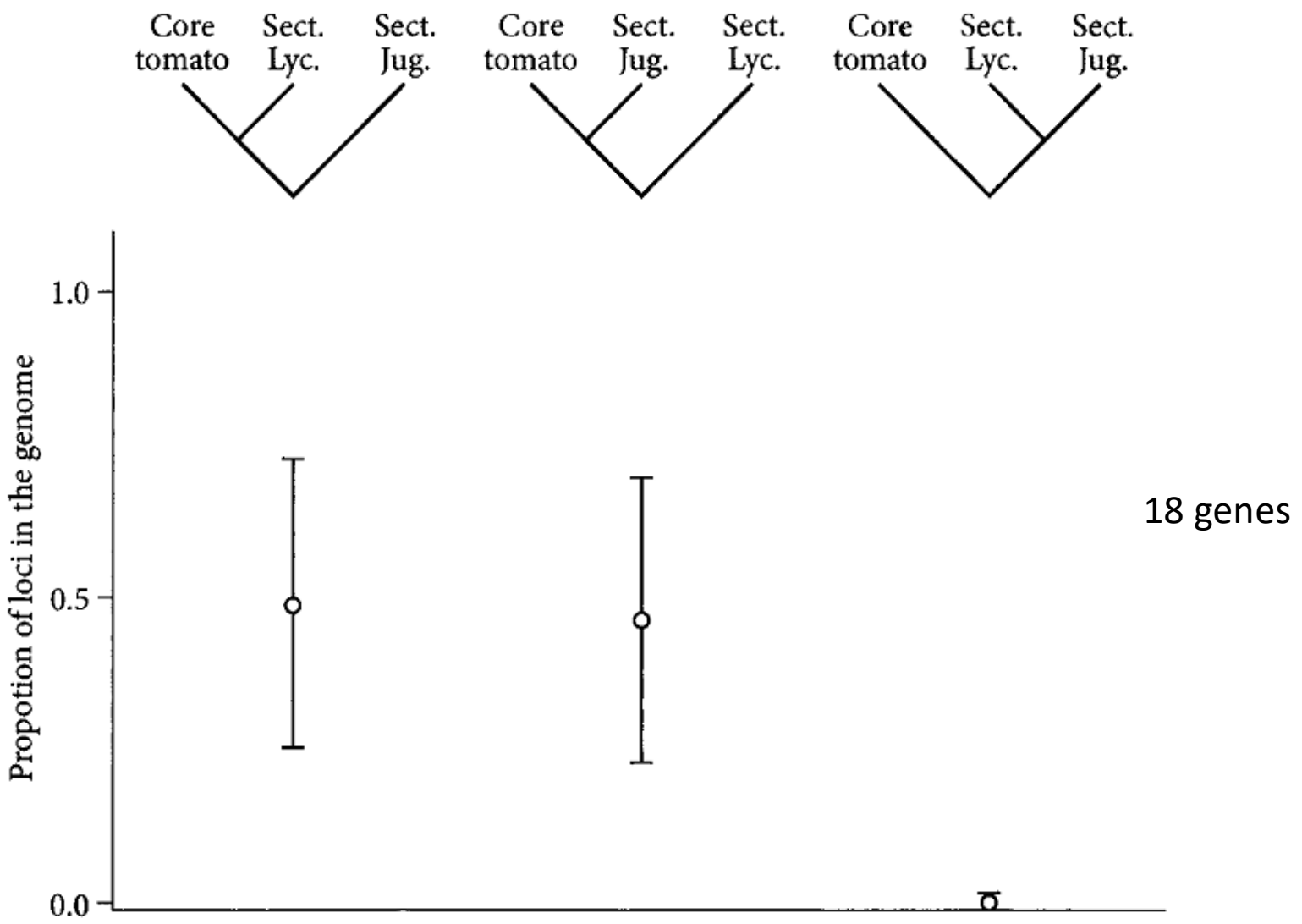
Comparación de conjuntos de datos

- **Métodos para árboles de especies**
  - Análisis Bayesiano de Concordancia (no limitado a sorteo incompleto de linajes)



30.000 genes

Analisis Bayesiano de Concordancia



18 genes