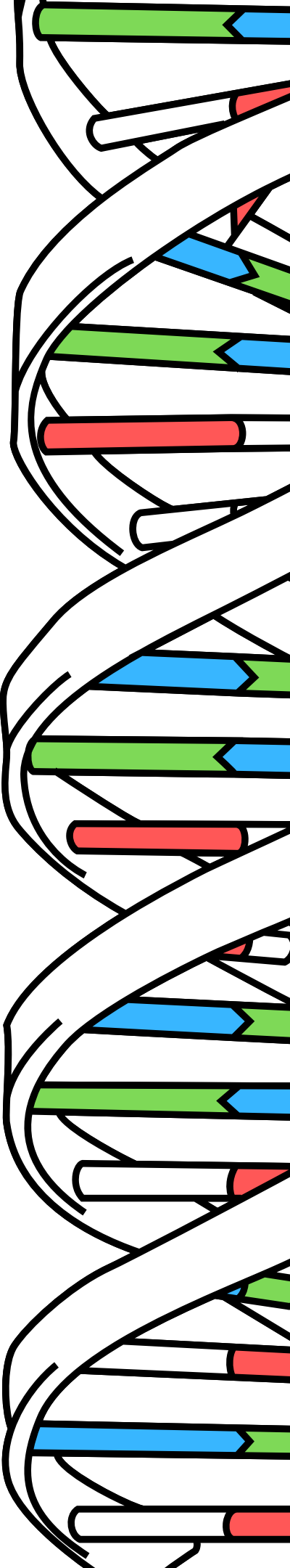


# Introducción a la filogenómica



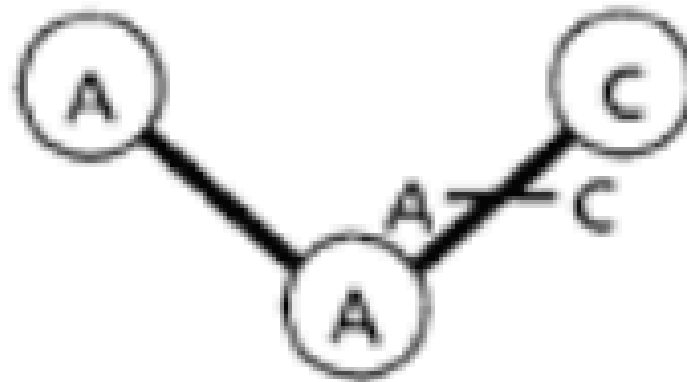
# Máxima Verosimilitud

Día	Temas
23 de septiembre	<ul style="list-style-type: none"><li>• Introducción a la Filogenómica. Conceptos básicos</li><li>• Introducción al sistema operativo Linux</li></ul>
24 de septiembre	<ul style="list-style-type: none"><li>• Métodos de construcción de librerías genómicas</li><li>• Control de calidad, ensamble y alineamiento de secuencias</li></ul>
25 de septiembre	<ul style="list-style-type: none"><li>• <b>Método de Máxima Verosimilitud</b></li><li>• Método de Inferencia Bayesiana</li></ul>
26 de septiembre	<ul style="list-style-type: none"><li>• Modelo coalescente multiespecies</li><li>• Evaluación de árboles de especies</li></ul>
27 de septiembre	<ul style="list-style-type: none"><li>• Estimación de tiempos de divergencia</li><li>• Evaluación de flujo genético</li></ul>

# Máxima Verosimilitud

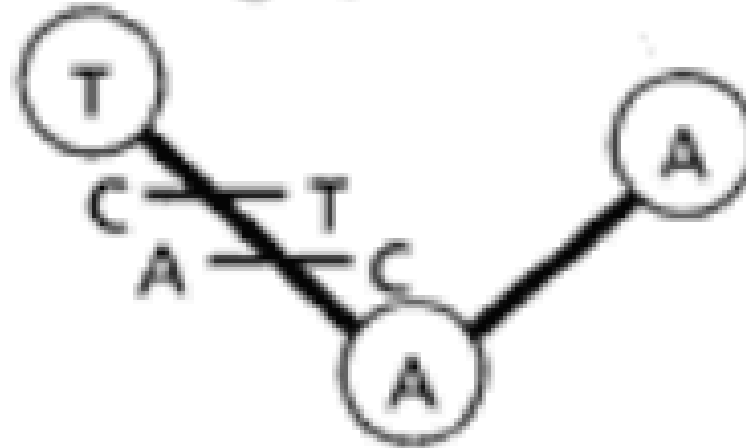
(a) Single substitution

1 change, 1 difference



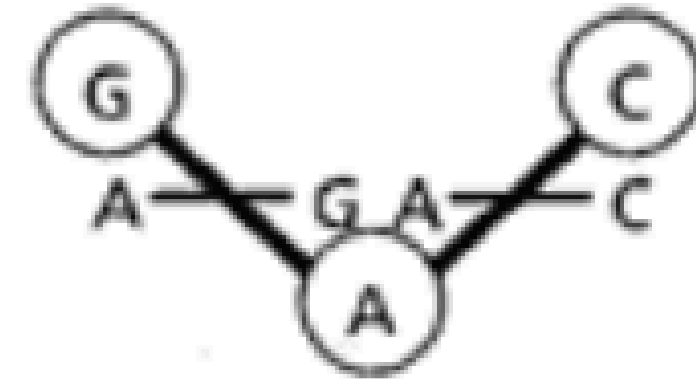
(b) Multiple substitution

2 changes, 1 difference



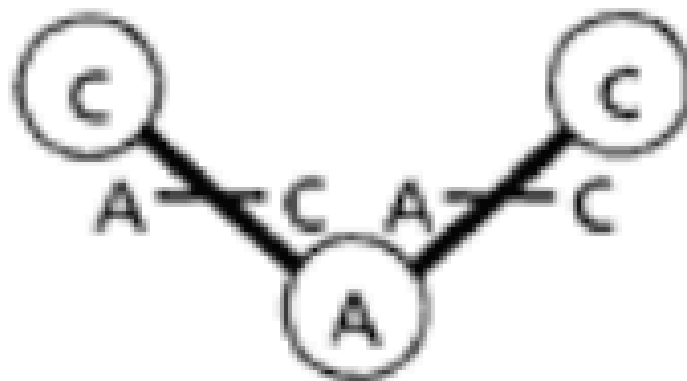
(c) Coincidental substitution

2 changes, 1 difference



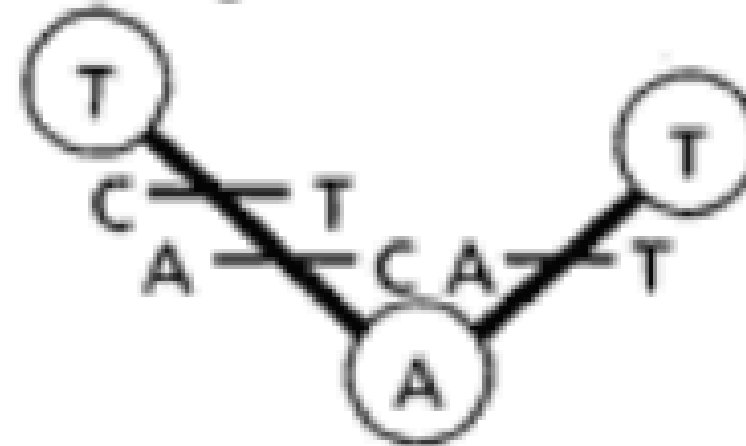
(d) Parallel substitution

2 changes, no difference



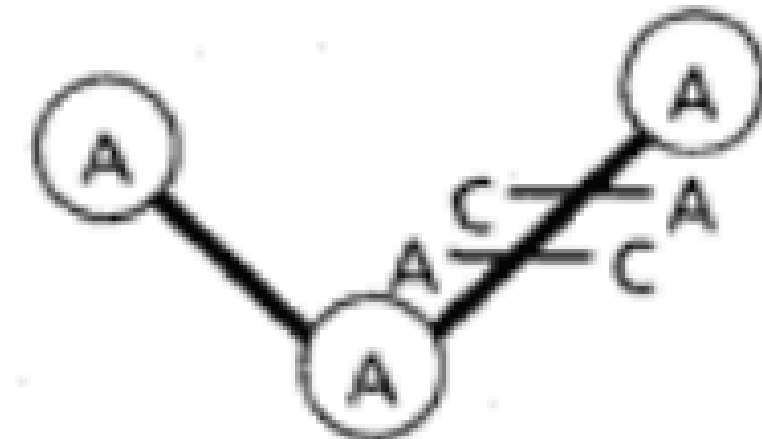
(e) Convergent substitution

3 changes, no difference

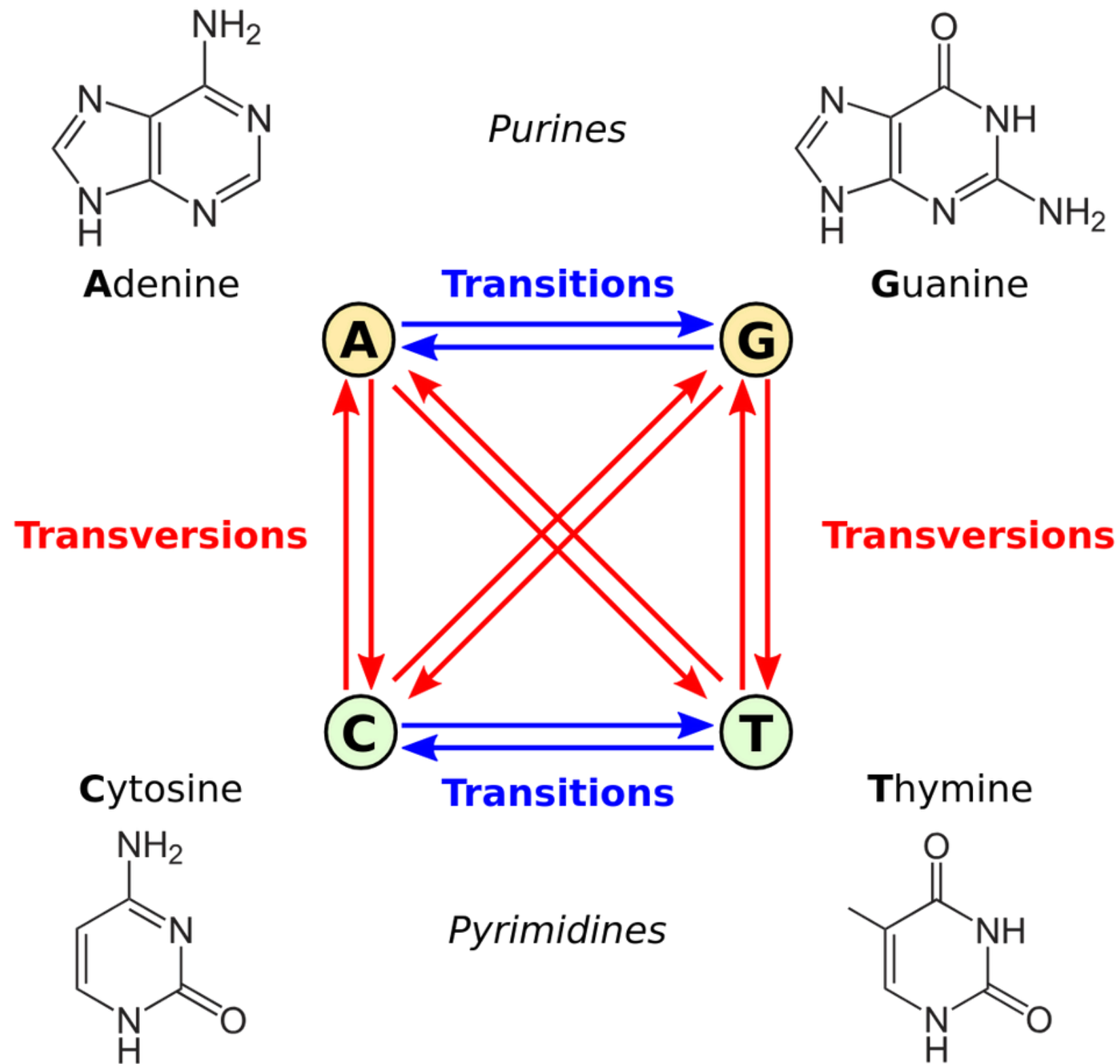


(f) Back substitution

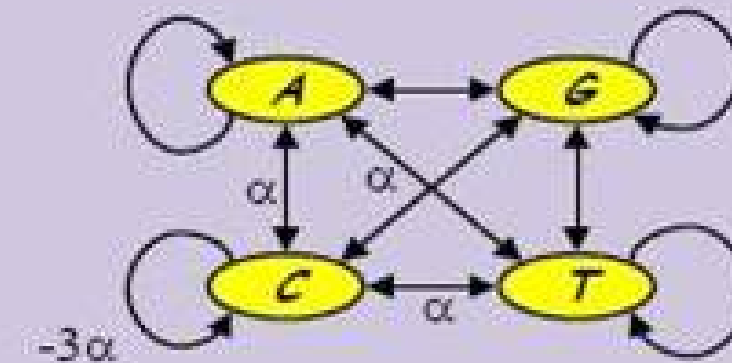
2 changes, no difference



# Máxima Verosimilitud



Jukes-Cantor assume equal rate of change:

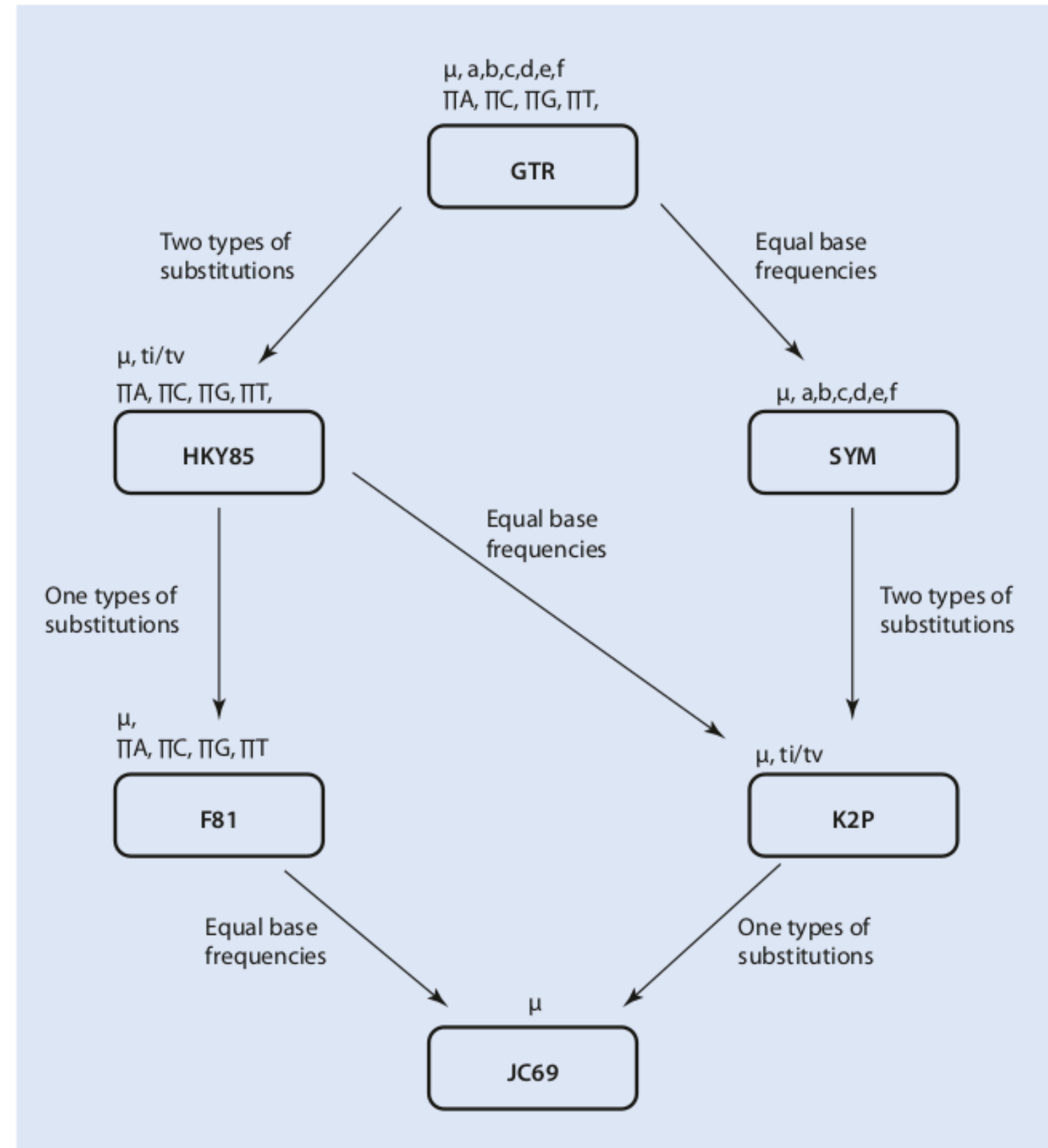


$$R = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{bmatrix} -3\alpha & \alpha & \alpha & \alpha \\ \alpha & -3\alpha & \alpha & \alpha \\ \alpha & \alpha & -3\alpha & \alpha \\ \alpha & \alpha & \alpha & -3\alpha \end{bmatrix} \end{matrix}$$

Kimura used a different rate matrix:

$$R = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{bmatrix} -2\beta - \alpha & \beta & \alpha & \beta \\ \beta & -2\beta - \alpha & \beta & \alpha \\ \alpha & \beta & -2\beta - \alpha & \beta \\ \beta & \alpha & \beta & -2\beta - \alpha \end{bmatrix} \end{matrix}$$

# Máxima Verosimilitud



## Máxima Verosimilitud

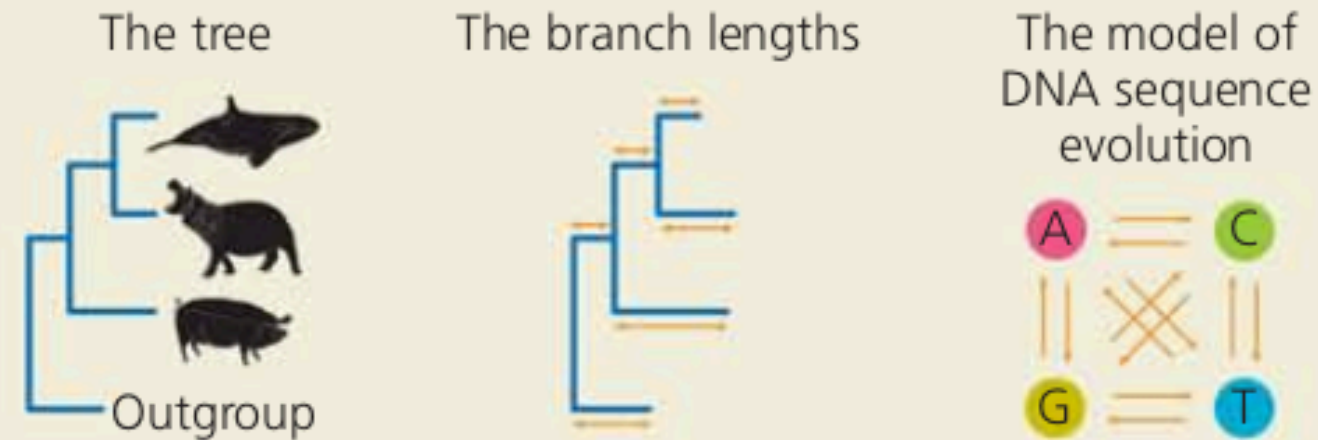
**Método estadístico que busca el árbol filogenético que mejor se ajusta a los datos observados, dado un modelo de evolución.**

Probabilidad de obtener los datos ( $D$ , la matriz) dada una determinada hipótesis ( $H$ , el árbol y el modelo de sustitución): cuál es la probabilidad de que, habiendo obtenido un determinado árbol, éste sea el resultado de nuestros datos.

$$P(D|H)$$

# Máxima Verosimilitud

## (a) The ingredients



## The data

AGTCCCCAXA GCTAAGGAGA ... TCCCTAAATC  
 AGTCCCCAAA GCAAAGGAGA ... TCTCTAAATC  
 AGATTCCAAA GCTAAGGAGA ... TCCCTAAATC  
 Outgroup AGTCCTCCAA ACTAAGGAGA ... CCCTTAAATC

## (b) The calculation

$$\begin{aligned}
 L \left( \begin{array}{c} \text{Tree} \\ \text{Outgroup} \end{array} \right) &= P \left( \begin{array}{c} \text{Species 1} \\ \text{Species 2} \\ \text{Species 3} \\ \text{Outgroup} \end{array} \middle| \begin{array}{c} \text{AGTCCCCAXA GCTAAGGAGA... TCCCTAAATC} \\ \text{AGTCCCCAAA GCAAAGGAGA... TCTCTAAATC} \\ \text{AGATTCCAAA GCTAAGGAGA... TCCCTAAATC} \\ \text{AGTCCTCCAA ACTAAGGAGA... CCCTTAAATC} \end{array} \right) \\
 &= P \left( \begin{array}{c} \text{Species 1} \\ \text{Species 2} \\ \text{Species 3} \\ \text{Outgroup} \end{array} \middle| \begin{array}{c} A \\ A \\ A \\ A \end{array} \right) \times P \left( \begin{array}{c} \text{Species 1} \\ \text{Species 2} \\ \text{Species 3} \\ \text{Outgroup} \end{array} \middle| \begin{array}{c} G \\ G \\ G \\ G \end{array} \right) \times \dots \\
 &= \sum_{\text{all possible state combinations}} P \left( \begin{array}{c} \text{Species 1} \\ \text{Species 2} \\ \text{Species 3} \\ \text{Outgroup} \end{array} \middle| \begin{array}{c} \text{State 1} \\ \text{State 2} \\ \text{State 3} \\ \text{State 4} \end{array} \right) = f \left( \begin{array}{c} \text{Tree} \\ \text{Outgroup} \end{array} \right)
 \end{aligned}$$

**Figure 4.26** How the likelihood of a tree is calculated  
 After Swofford et al. (1996); Graur and Li (2000); Huelsenbeck et al. (2001).

# Máxima Verosimilitud

Qué tan robusto es un árbol filogenético.

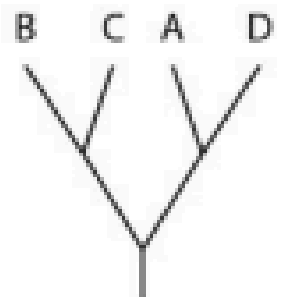
Medida de soporte para las relaciones entre las especies.

Las ramas que son respaldadas en un alto porcentaje de los pseudo-replicas son consideradas más confiables.

Conservativo 70%

Original  
sequence alignment

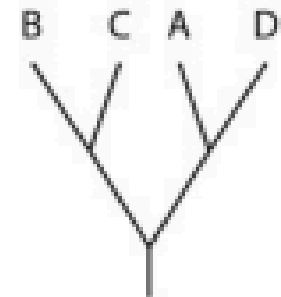
	Site number														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Species A	A	A	T	G	C	T	A	G	T	G	G	T	G	A	T
Species B	A	A	G	C	T	A	T	G	G	T	G	A	T	C	G
Species C	A	G	C	C	T	A	T	G	T	G	G	A	A	C	G
Species D	A	A	C	C	C	A	T	T	G	G	G	T	G	A	T



Original alignment  
tree

Bootstrap  
pseudo-replicate #1

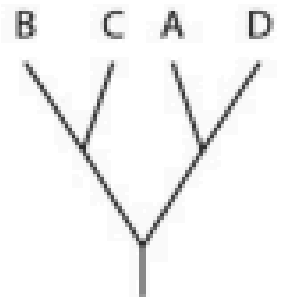
	5	3	3	1	12	9	2	4	11	13	10	14	8	11	13
Species A	C	T	T	A	T	T	A	G	G	G	G	A	G	G	G
Species B	T	G	G	A	A	G	A	C	G	T	T	C	G	G	T
Species C	T	C	C	A	A	T	G	C	G	A	G	C	G	G	A
Species D	C	C	C	A	T	T	A	C	G	G	G	A	T	G	G



Bootstrap tree #1

Bootstrap  
pseudo-replicate #2

	9	7	12	5	2	4	2	6	14	9	4	9	7	2	1
Species A	T	A	T	C	A	G	A	T	A	T	G	T	A	A	A
Species B	G	T	A	T	A	C	A	A	C	G	C	G	T	A	A
Species C	T	T	A	T	G	C	G	A	C	T	C	T	T	G	A
Species D	T	T	T	C	A	C	A	A	A	T	C	T	T	A	A



Bootstrap tree #2



# Inferencia Bayesiana

**Método de inferencia estadística que actualiza la probabilidad de una hipótesis a medida que se dispone de más evidencia o información.**

En el contexto de las filogenias, utilizamos la inferencia bayesiana para estimar la distribución de probabilidad de diferentes árboles filogenéticos dados los datos (secuencias de ADN o ARN) y un modelo de evolución.

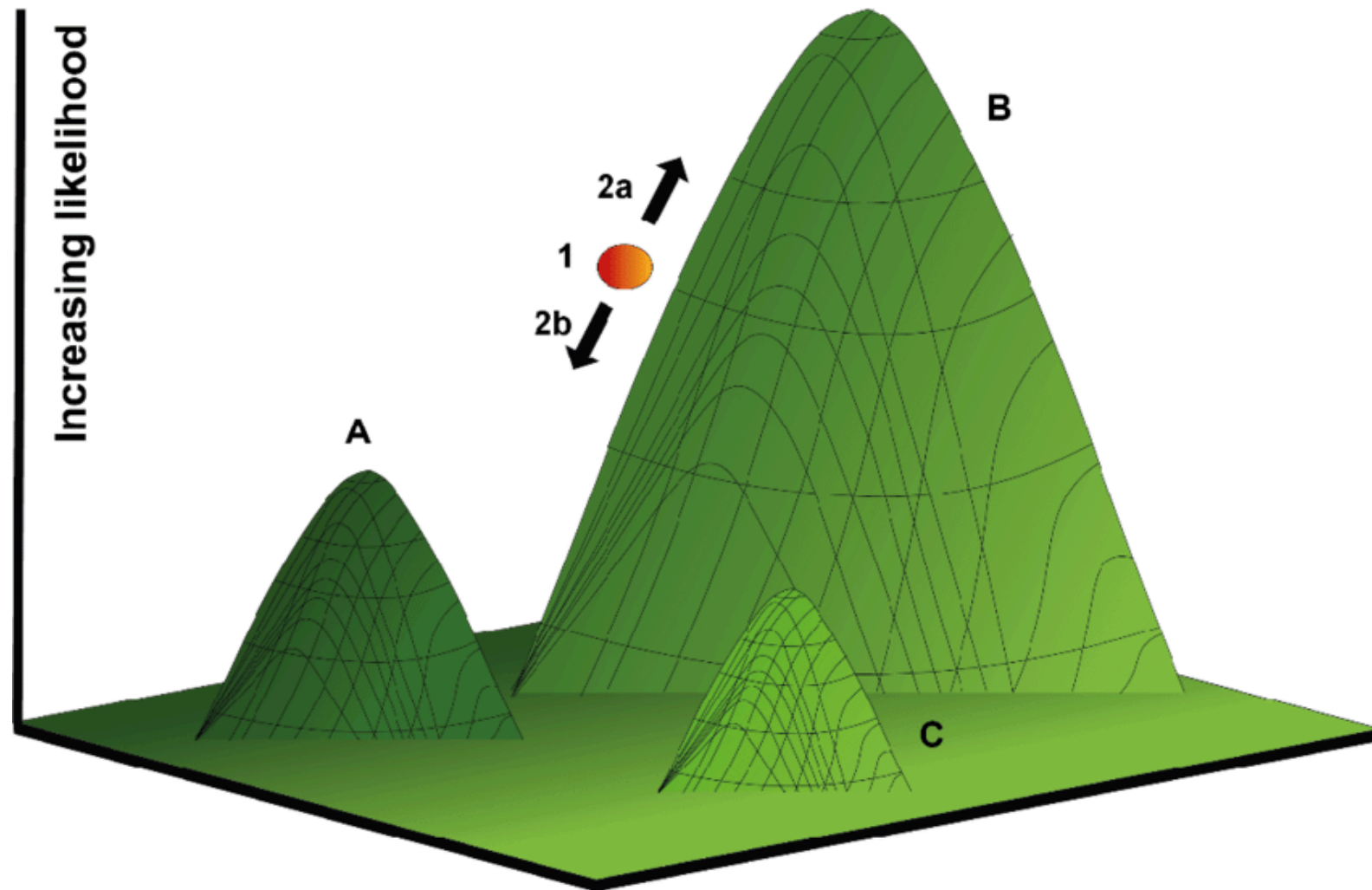
A diferencia de la máxima verosimilitud, que nos proporciona un único árbol "mejor", la inferencia bayesiana nos proporciona una distribución de probabilidad sobre todos los árboles posibles.

Esto nos permite tener en cuenta la incertidumbre en nuestras estimaciones y proporcionar intervalos de confianza para nuestras inferencias filogenéticas.

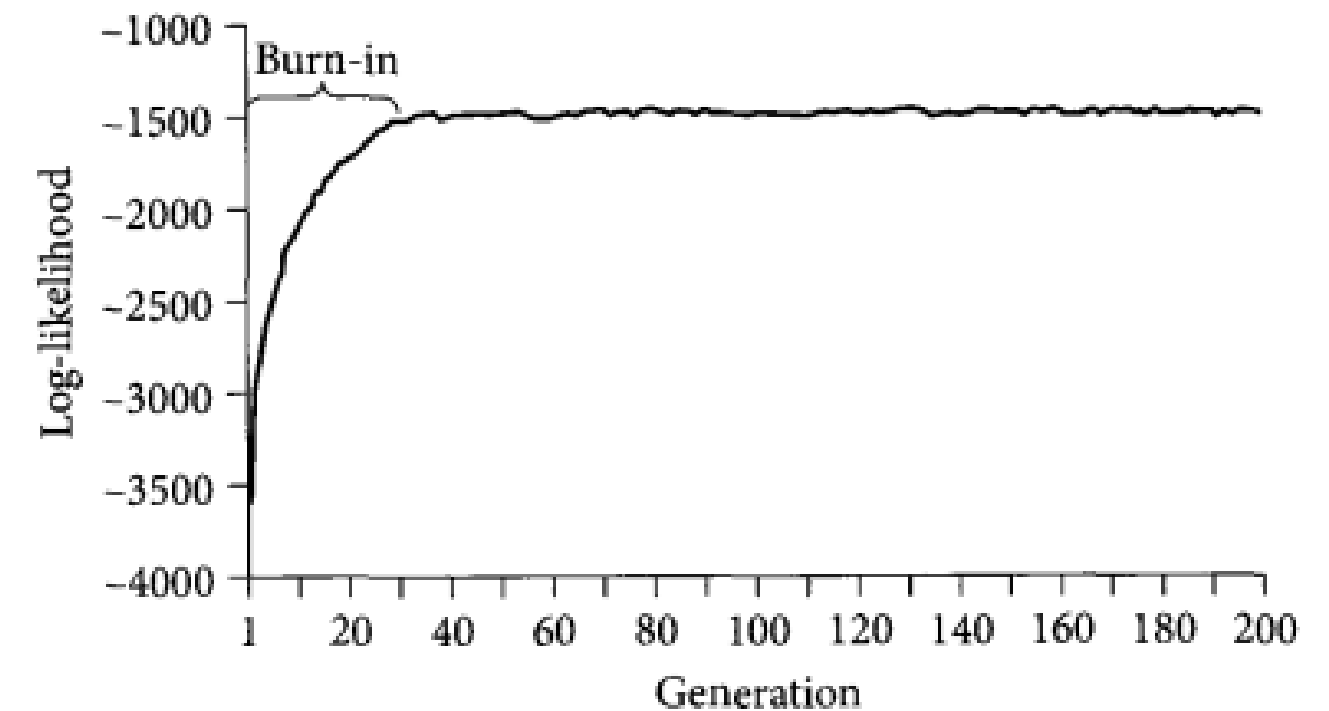
$$P(H|D) = \frac{\overset{\text{Likelihood}}{\boxed{P(D|H)}} \cdot \overset{\text{Probabilidad "a priori" del árbol}}{\boxed{P(H)}}}{\underset{\substack{\text{Probabilidad de los datos} \\ \text{Imposible de calcular. Se elimina al usar las MCMC.}}}{\boxed{P(D)}}}$$

No se sabe cómo calcularla, así que se asume que es igual a:  
 $1/N_{\text{árboles posibles}}$

# Inferencia Bayesiana



Verosimilitud de los diferentes conjuntos de parámetros que explican los datos.



El algoritmo mejora la verosimilitud de los parámetros encontrados.