

INFERENCIA FILOGENÉTICA

MÁXIMA VEROSIMILITUD

MÁXIMA VEROSIMILITUD

La hipótesis (árbol) con la máxima verosimilitud es aquella que tiene la mayor probabilidad de haber originado los datos observados



- 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

Modelo

- Cara en un lado y sello en otro
- Independencia en cada tiro
- Observador distingue caras de sellos

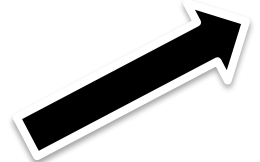
Datos





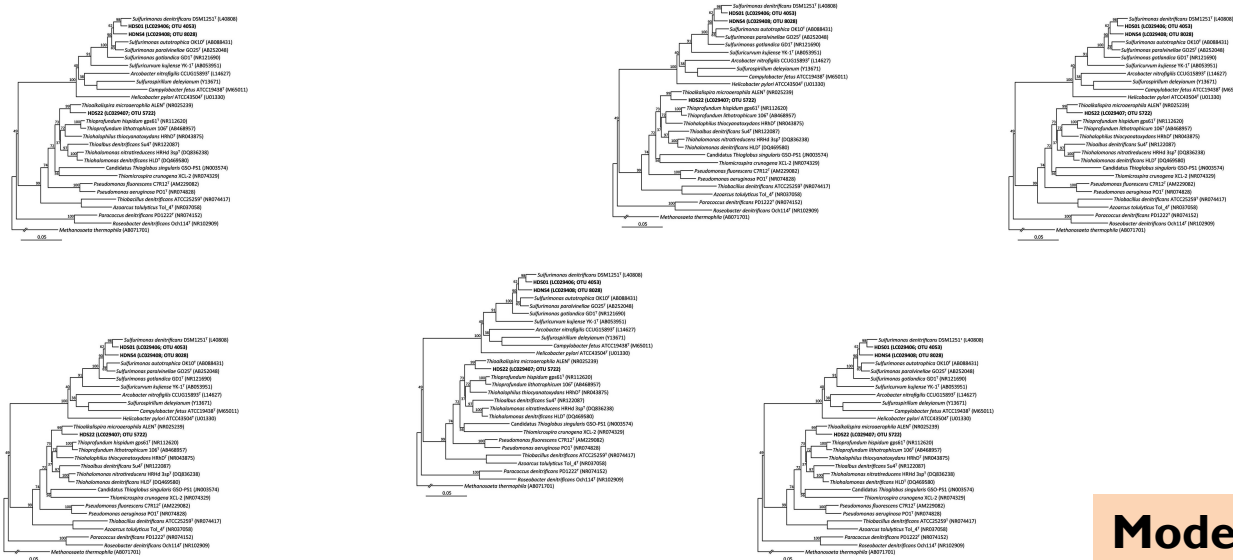
Prob. normal	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.001
Prob. sesgada	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.056

Verosimilitud



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

MÁXIMA VEROSIMILITUD EN FILOGENÉTICA



Hipótesis

Modelo

Mk, JC69, F81, HKY85, GTR...etc

Datos

Position	20	22	23	24	25	26	32	42	45	50	53	56	69	81	93	103	121	122	123	124	147	148	149	153	155	156	170	183	Total	
Consensus	A	A	A	A	G	C	A	C	T	A	C	C	C	C	A	A	A	A	T	T	A	T	T	T	G	C	G	C	G	
AXE	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	G	T	T	T	G	G	A	G	13
DW	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	G	T	T	G	G	G	A	G	12
FNZ	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	G	T	T	G	G	A	G	12	
NQ	A	A	A	A	A	A	A	A	A	G	A	A	T	A	A	A	A	A	A	A	A	G	T	T	G	G	G	A	G	12
ZT	A	A	A	A	A	A	A	A	A	G	A	A	T	A	A	A	A	A	A	A	A	G	T	T	G	G	G	A	G	12
MEM	A	A	A	A	A	A	A	A	A	G	A	A	T	A	A	A	A	A	A	A	A	G	T	T	G	G	G	A	G	13
GD	A	A	A	A	G	A	A	A	T	A	A	A	A	A	G	A	A	A	A	A	A	G	G	T	G	G	G	A	G	7
IM	A	A	A	A	G	A	A	A	T	A	A	A	A	A	G	A	A	A	A	A	A	G	G	T	G	T	G	A	G	7
NRS	A	A	A	A	G	A	A	A	T	A	A	A	A	A	G	A	A	A	A	A	A	G	G	T	G	T	G	A	G	7
ZPJ	A	A	A	A	G	A	A	A	T	A	A	A	A	A	G	A	A	A	A	A	A	G	G	T	G	T	G	A	G	7
AB	A	A	A	A	G	A	A	A	T	A	A	A	A	A	A	A	A	A	T	T	A	G	T	T	G	G	G	A	G	0
XC	A	A	A	A	G	A	A	A	T	A	A	A	A	A	A	A	A	A	T	T	A	G	T	T	G	G	G	A	G	0
MR	T	A	A	A	G	A	A	A	T	A	A	A	A	A	A	A	A	T	T	A	G	T	T	G	G	G	A	G	2	
UBG	T	A	A	A	G	A	A	A	T	A	A	A	A	A	A	A	A	T	T	A	G	T	T	G	G	G	A	G	2	
FYC	A	A	A	A	A	A	A	A	T	A	A	A	A	A	T	G	A	T	T	A	G	T	T	G	G	T	A	A	6	
JRR	A	A	A	A	G	A	A	A	T	A	A	A	T	A	T	G	A	T	T	A	G	T	T	G	G	T	A	A	6	
UM	A	A	A	A	G	A	A	A	T	A	A	A	T	A	T	G	A	T	T	A	G	T	T	G	G	T	A	A	6	
WZ	A	A	A	A	G	A	A	A	T	A	A	A	T	A	A	G	A	T	T	A	G	T	T	G	G	T	A	A	5	
DGO	A	A	A	A	G	A	A	A	T	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	6	
Summary:	17A	13A	13A	13A	13G	13C	17A	14C	13T	16A	18C	15C	13C	15C	16A	15A	9A	9T	9T	9A	18G	14T	18T	18G	15C	14G	16C	14G	135	
	2T	6-	6-	6-	6-	6-	2C	5A	6C	3G	1T	4T	6T	4G	3T	4C	10-	10-	10-	10-	1-	4C	1-	1T	4T	5T	3A	5-		

MÁXIMA VEROSIMILITUD EN FILOGENÉTICA

PASOS:

1. Se escoge un árbol cualquiera con longitud de ramas y un modelo de sustitución
2. Se calcula la verosimilitud de cada posición (carácter)
3. Se multiplican las verosimilitudes de todas las posiciones (caracteres)
4. Se usa un algoritmo para optimizar la longitud de ramas y otros parámetros (repetiendo pasos 1–3) hasta que se maximice la verosimilitud del árbol
5. Se repiten estos pasos en otros árboles hasta encontrar el árbol de máxima verosimilitud

MÁXIMA VEROSIMILITUD EN FILOGENÉTICA

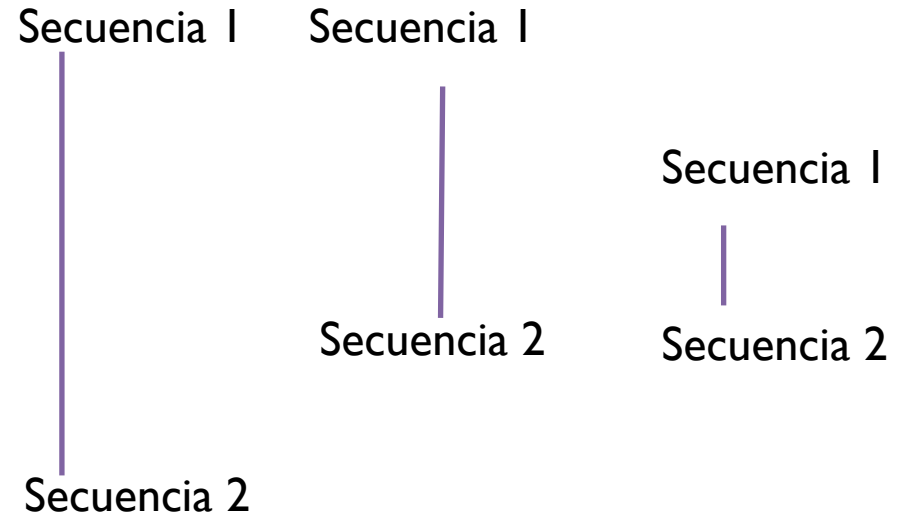
PASOS:

- I. Se escoge un árbol cualquiera con longitud de ramas y un modelo de sustitución

DATOS

Secuencia 1	A	G	G	T	C	T
Secuencia 2	A	G	A	T	A	T

HIPÓTESIS



Modelo

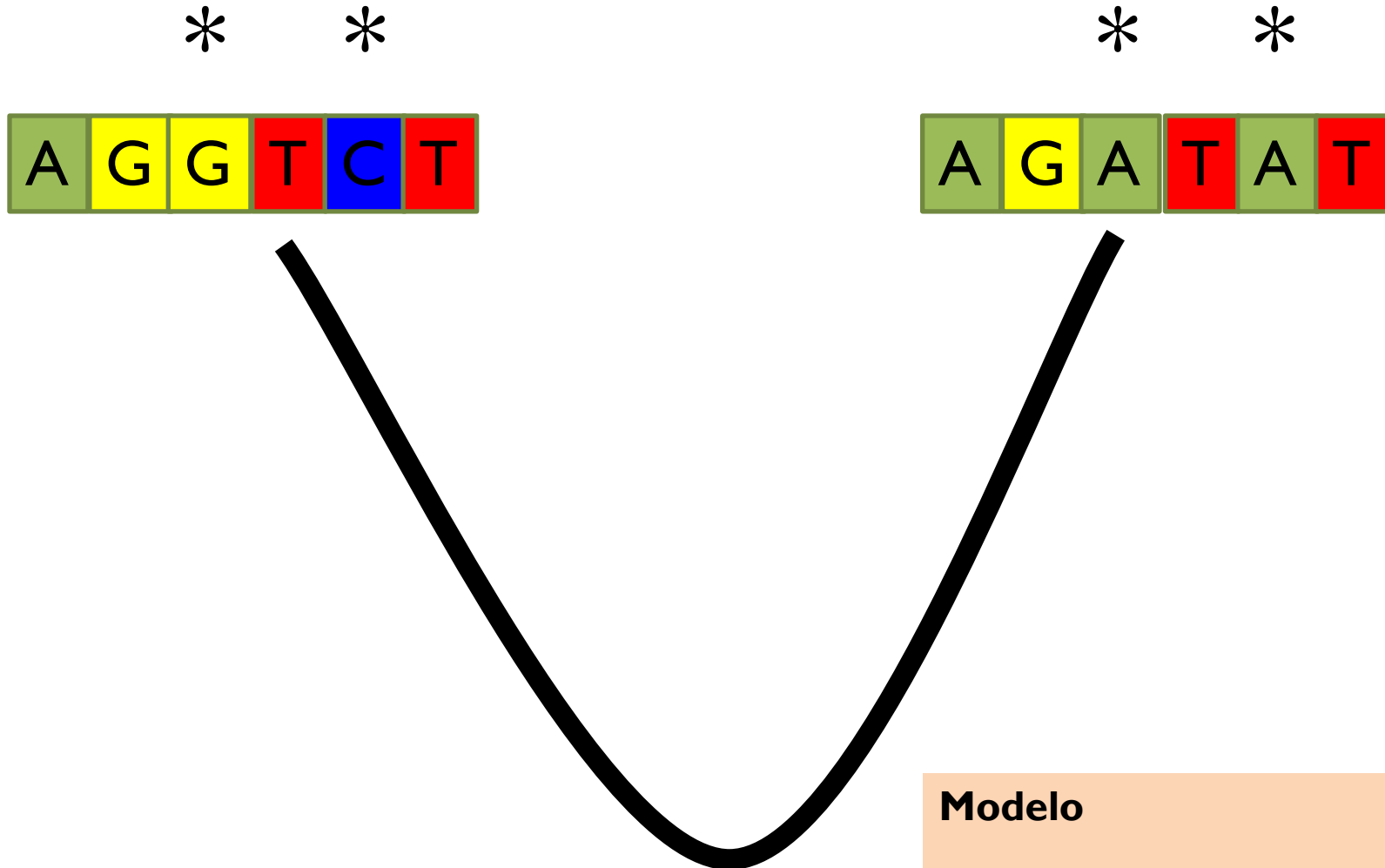
Jukes-Cantor (1969)

MÁXIMA VEROSIMILITUD EN FILOGENÉTICA

PASOS:

1. Se escoge un árbol cualquiera con longitud de ramas y un modelo de sustitución
2. Se calcula la verosimilitud de cada posición (carácter)

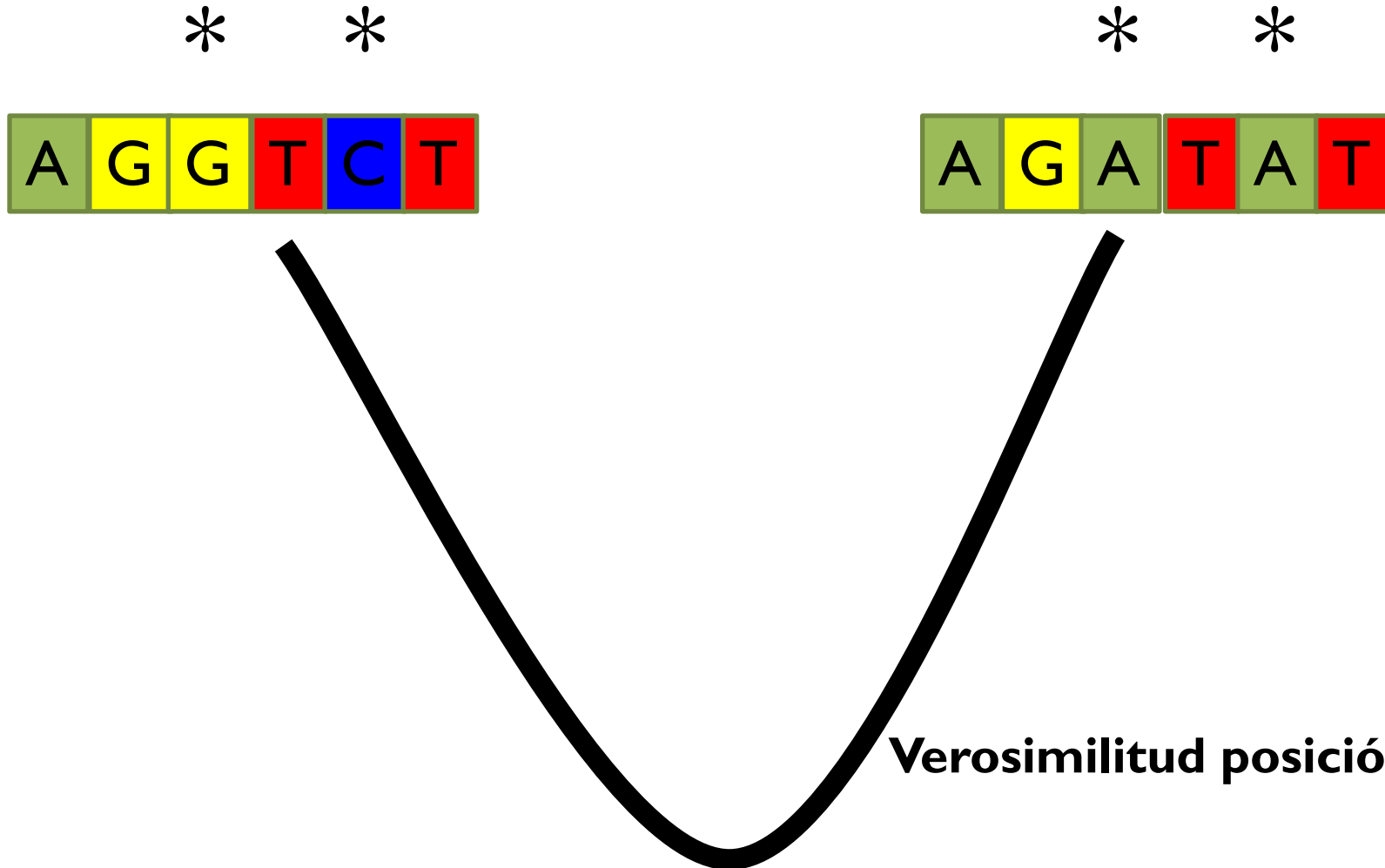
MÁXIMA VEROSIMILITUD EN FILOGENÉTICA



Modelo

Jukes-Cantor (1969)

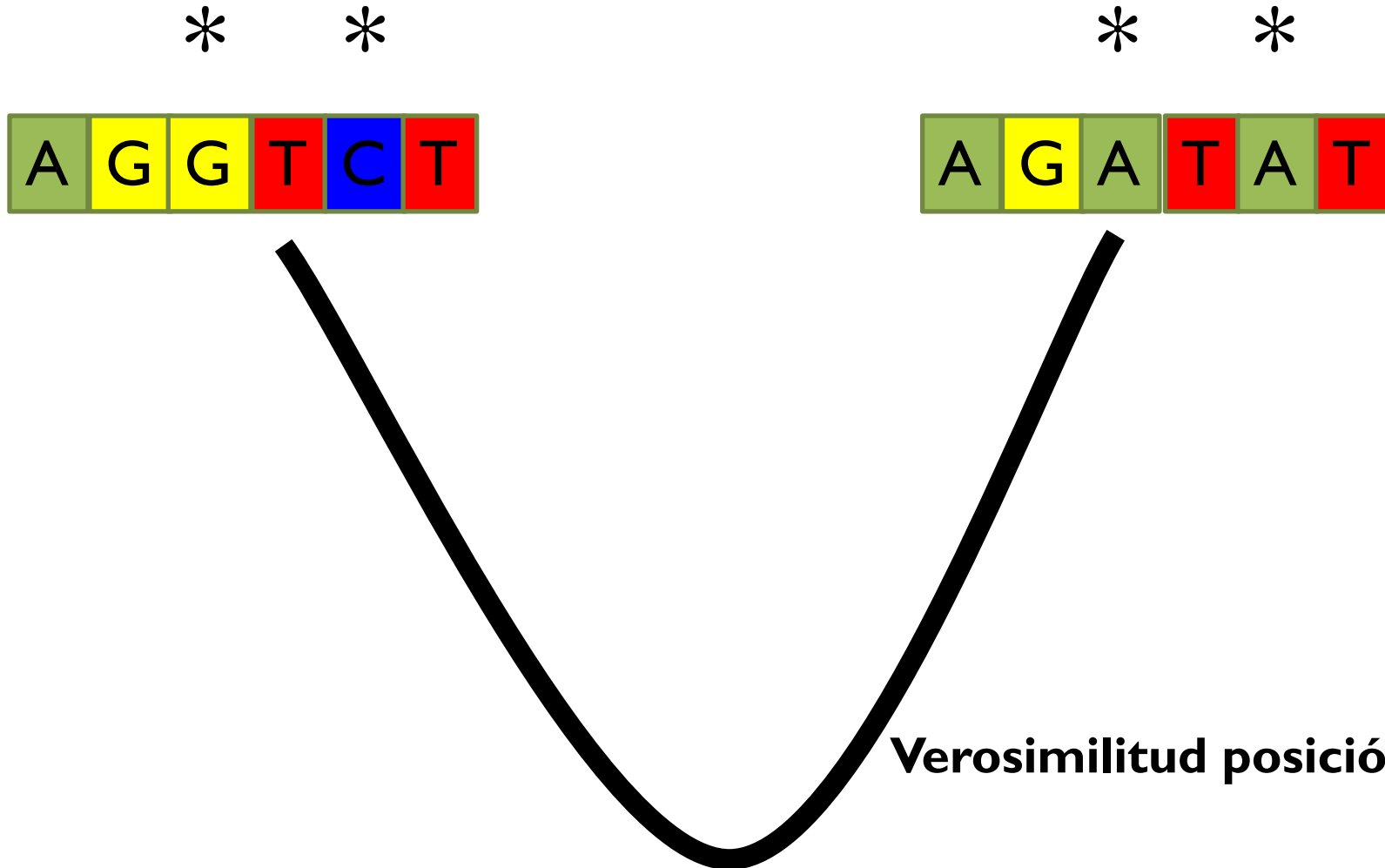
MÁXIMA VEROSIMILITUD EN FILOGENÉTICA



Verosimilitud posición # 1:

$$P(A \leftrightarrow A) = \frac{1}{4} \left(\frac{1}{4} + \frac{3}{4} e^{-4/3 \mu t} \right)$$

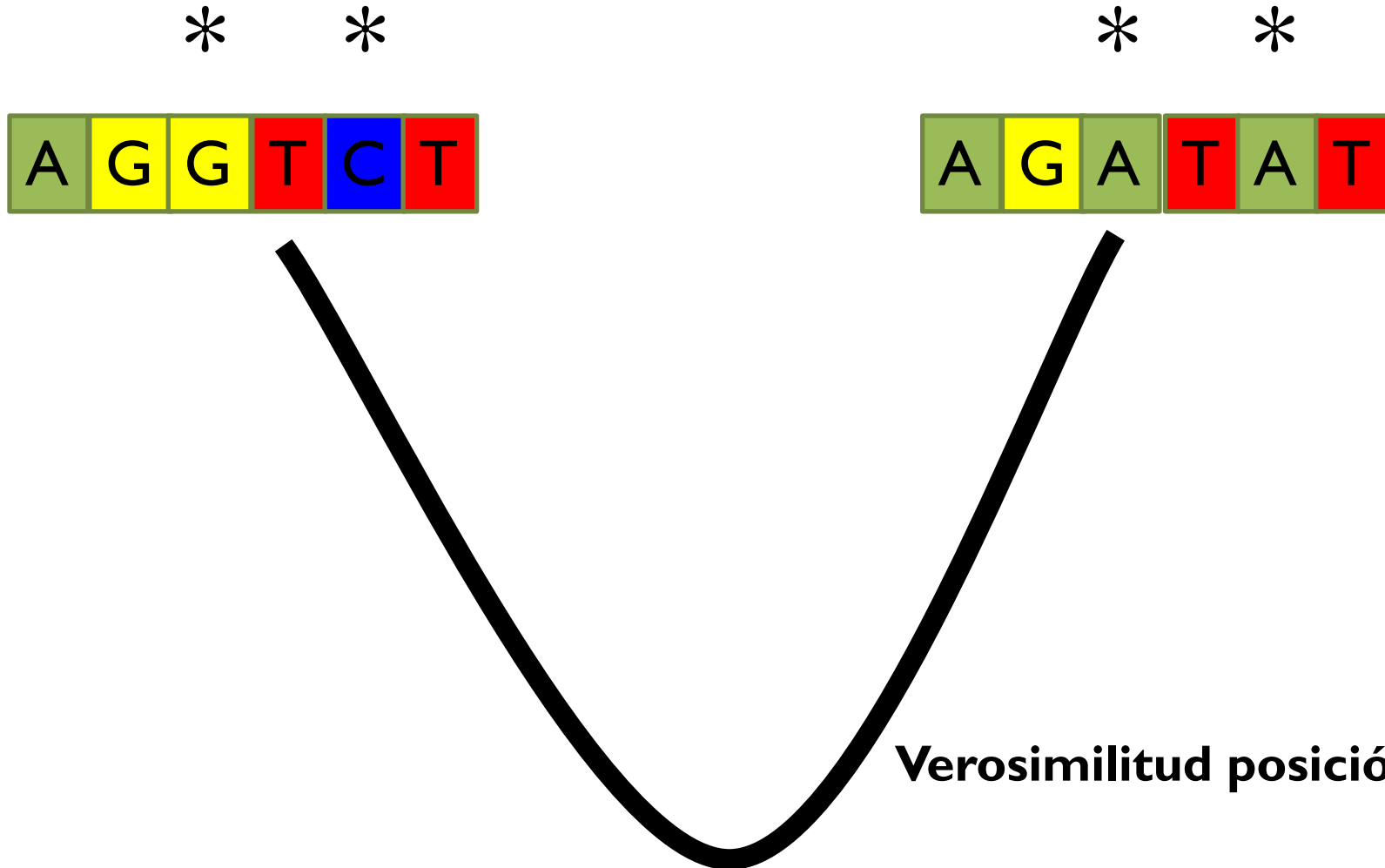
MÁXIMA VEROSIMILITUD EN FILOGENÉTICA



Verosimilitud posición # 2:

$$P(G \leftrightarrow G) = \frac{1}{4} \left(\frac{1}{4} + \frac{3}{4} e^{-4/3 \mu t} \right)$$

MÁXIMA VEROSIMILITUD EN FILOGENÉTICA



Verosimilitud posición # 3:

$$P(G \leftrightarrow A) = \frac{1}{4} \left(\frac{1}{4} - \frac{1}{4} e^{-4/3 \mu t} \right)$$

MÁXIMA VEROSIMILITUD EN FILOGENÉTICA

PASOS:

1. Se escoge un árbol cualquiera con longitud de ramas y un modelo de sustitución
2. Se calcula la verosimilitud de cada posición (carácter)
3. Se multiplican las verosimilitudes de todas las posiciones (caracteres)

Verosimilitud total (L) del árbol

$$L = \underbrace{\left[\frac{1}{4} \left(\frac{1}{4} + \frac{3}{4} e^{-4/3 \mu t} \right) \right]^4}_{\text{A} \leftrightarrow \text{A}, \text{G} \leftrightarrow \text{G}, \text{T} \leftrightarrow \text{T}, \text{T} \leftrightarrow \text{T}} * \underbrace{\left[\frac{1}{4} \left(\frac{1}{4} - \frac{1}{4} e^{-4/3 \mu t} \right) \right]^2}_{\text{G} \leftrightarrow \text{A}, \text{C} \leftrightarrow \text{A}}$$

A ↔ A

G ↔ G

T ↔ T

T ↔ T

G ↔ A

C ↔ A

*

*

*

*

A	G	G	T	C	T
---	---	---	---	---	---

A	G	A	T	A	T
---	---	---	---	---	---

MÁXIMA VEROSIMILITUD EN FILOGENÉTICA

PASOS:

1. Se escoge un árbol cualquiera con longitud de ramas y un modelo de sustitución
2. Se calcula la verosimilitud de cada posición (carácter)
3. Se multiplican las verosimilitudes de todas las posiciones (caracteres)
4. Se usa un algoritmo para optimizar la longitud de ramas y otros parámetros (repetiendo pasos 1–3) hasta que se maximice la verosimilitud del árbol

Verosimilitud total (ML) del árbol

¿¿¿ μt ???

$$L = \underbrace{\left[\frac{1}{4} \left(\frac{1}{4} + \frac{3}{4} e^{-4/3 \mu t} \right) \right]^4}_{\text{A} \leftrightarrow \text{A}, \text{G} \leftrightarrow \text{G}, \text{T} \leftrightarrow \text{T}, \text{T} \leftrightarrow \text{T}} * \underbrace{\left[\frac{1}{4} \left(\frac{1}{4} - \frac{1}{4} e^{-4/3 \mu t} \right) \right]^2}_{\text{G} \leftrightarrow \text{A}, \text{C} \leftrightarrow \text{A}}$$

A ↔ A

G ↔ G

T ↔ T

T ↔ T

G ↔ A

C ↔ A

*

*

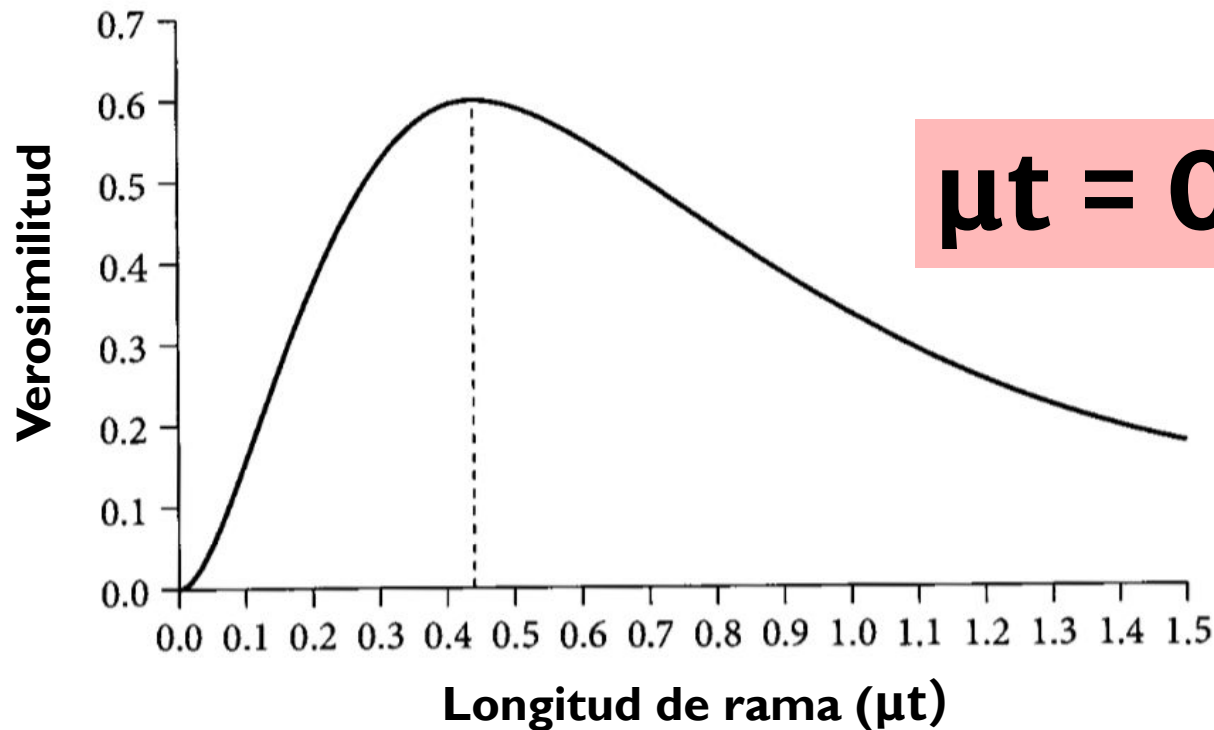
*

*

A G G T C T

A G A T A T

$$L = \left[\frac{1}{4} \left(\frac{1}{4} + \frac{3}{4} e^{-4/3 \mu t} \right) \right]^4 * \left[\frac{1}{4} \left(\frac{1}{4} - \frac{1}{4} e^{-4/3 \mu t} \right) \right]^2$$



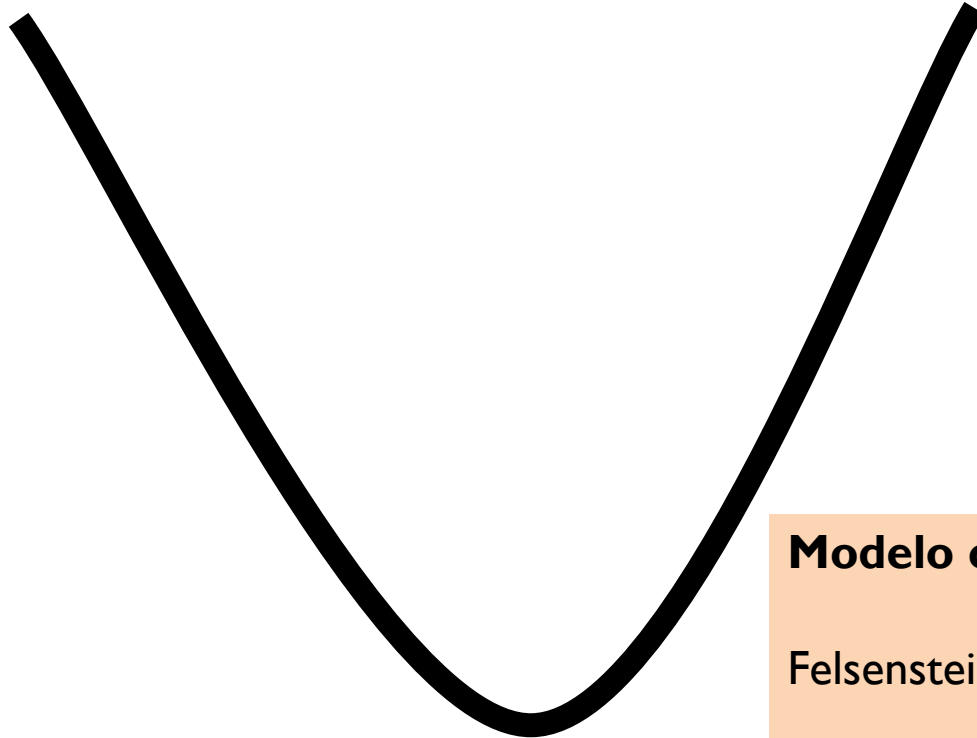
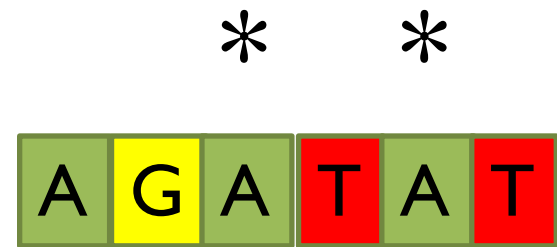
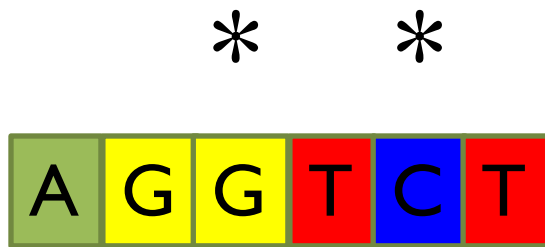
$$\mu t = 0.44$$

Reemplazando...

$$L = \left[\frac{1}{4} \left(\frac{1}{4} + \frac{3}{4} e^{-4/3(0.44)} \right) \right]^4 * \left[\frac{1}{4} \left(\frac{1}{4} - \frac{1}{4} e^{-4/3(0.44)} \right) \right]^2$$

$$L = 0.000000595$$

$$\ln(L) = -14.33$$

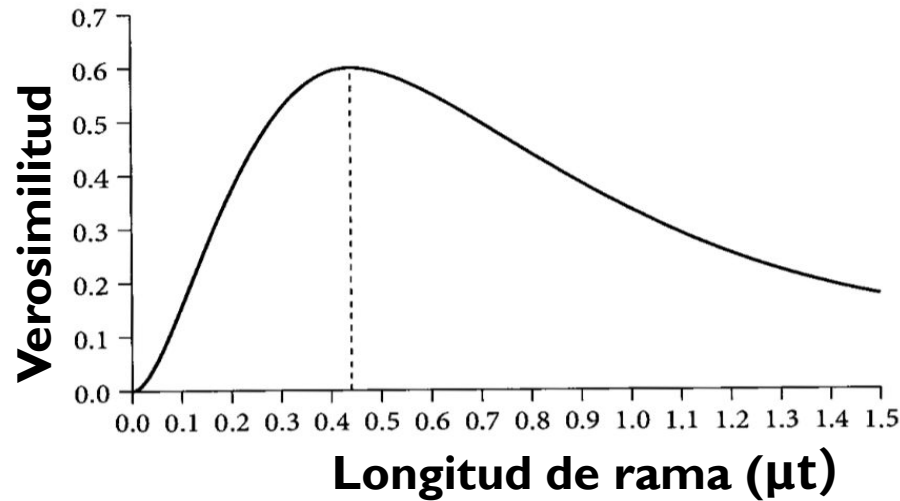


Modelo con más parámetros

Felsenstein (1981): F81

- Bases con frecuencias desiguales

Modelo F8I: además de estimar μ_t ...



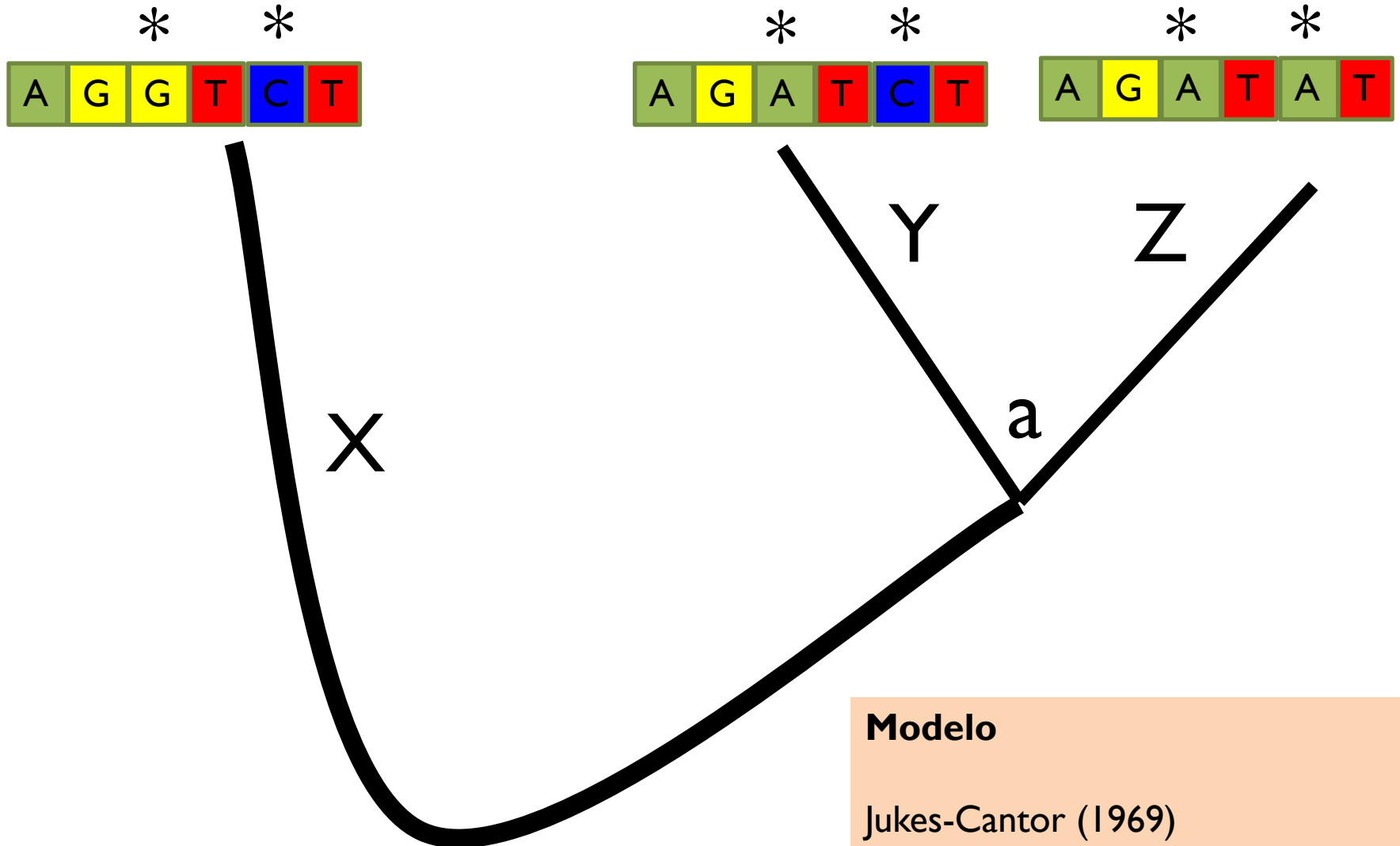
Estimar simultáneamente $\pi_A, \pi_G, \pi_C, \pi_T$

MÁXIMA VEROSIMILITUD EN FILOGENÉTICA

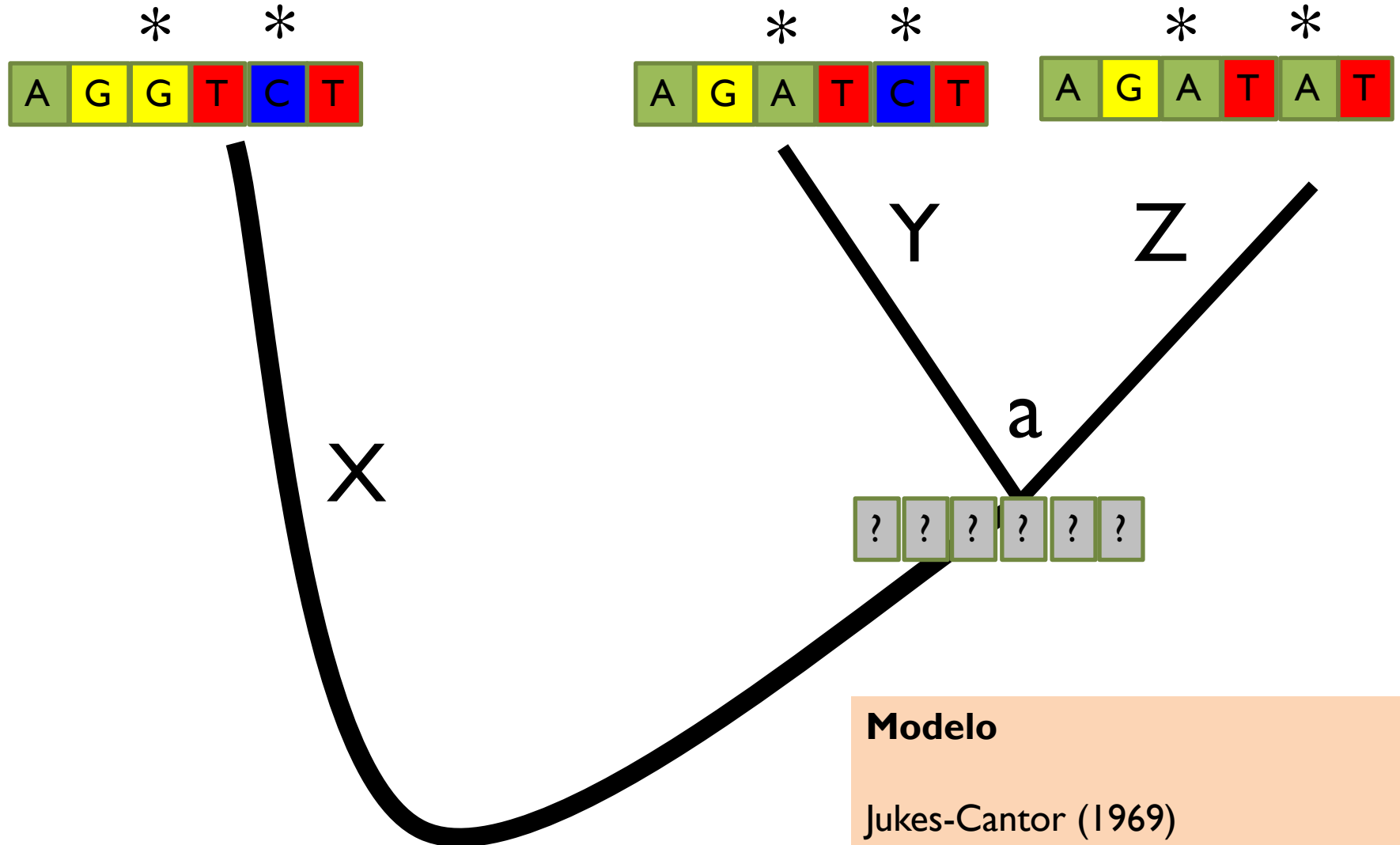
RESUMÉN:

1. Se escoge un árbol cualquiera con longitud de ramas y un modelo de sustitución
2. Se calcula la verosimilitud de cada posición (carácter)
3. Se multiplican las verosimilitudes de todas las posiciones (caracteres)
4. Se usa un algoritmo para optimizar la longitud de ramas y otros parámetros (repetiendo pasos 1–3) hasta que se maximice la verosimilitud del árbol
5. Se repiten estos pasos en otros árboles hasta encontrar el árbol de máxima verosimilitud

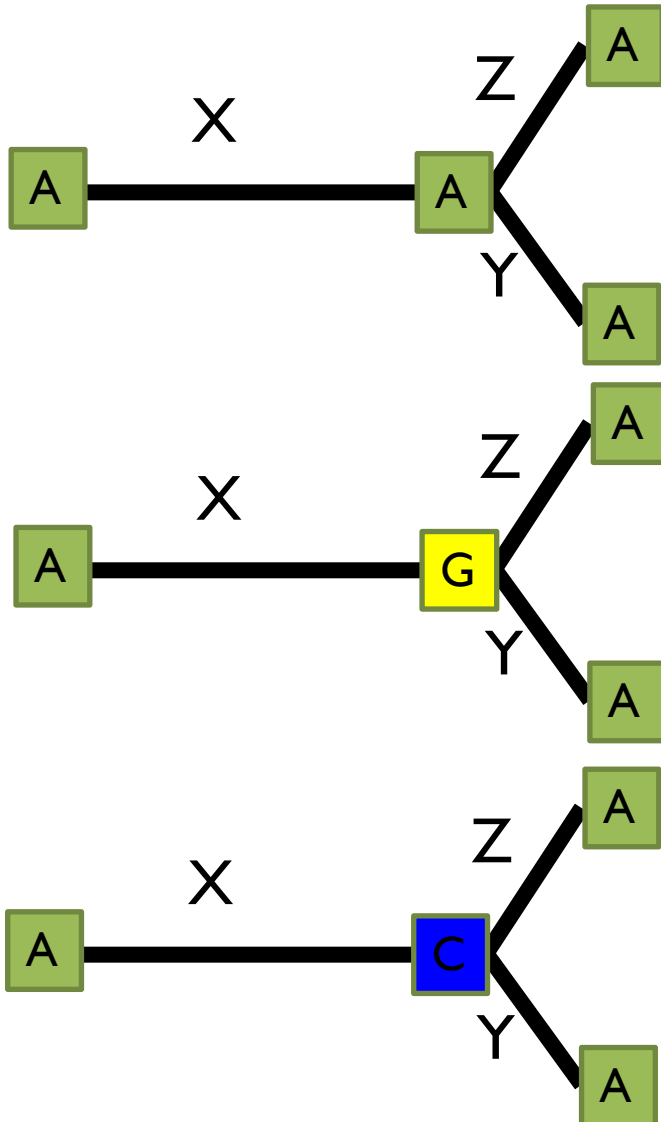
¿Y si adicionamos terminales al árbol?



¿Y si adicionamos terminales al árbol?



Verosimilitud posición # 1

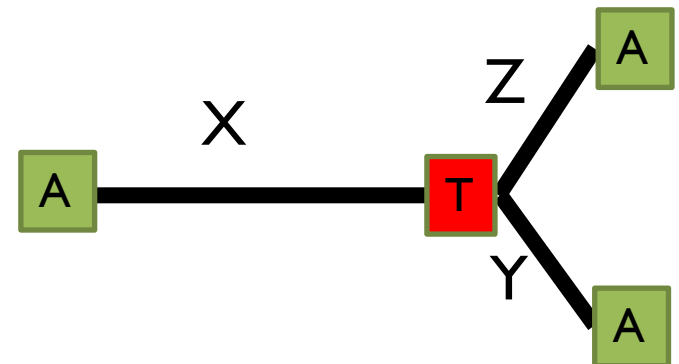


$$P(A \leftrightarrow A) = \left[\frac{1}{4} \left(\frac{1}{4} + \frac{3}{4} e^{-4/3 \mu t} \right) \right]^3$$

+

$$P(A \leftrightarrow G) = \left[\frac{1}{4} \left(\frac{1}{4} - \frac{1}{4} e^{-4/3 \mu t} \right) \right]^3$$

+



+

$$P(A \leftrightarrow T) = \left[\frac{1}{4} \left(\frac{1}{4} - \frac{1}{4} e^{-4/3 \mu t} \right) \right]^3$$

$$P(A \leftrightarrow C) = \left[\frac{1}{4} \left(\frac{1}{4} - \frac{1}{4} e^{-4/3 \mu t} \right) \right]^3$$