INFERENCIA FILOGENÉTICA USANDO MODELOS EVOLUTIVOS

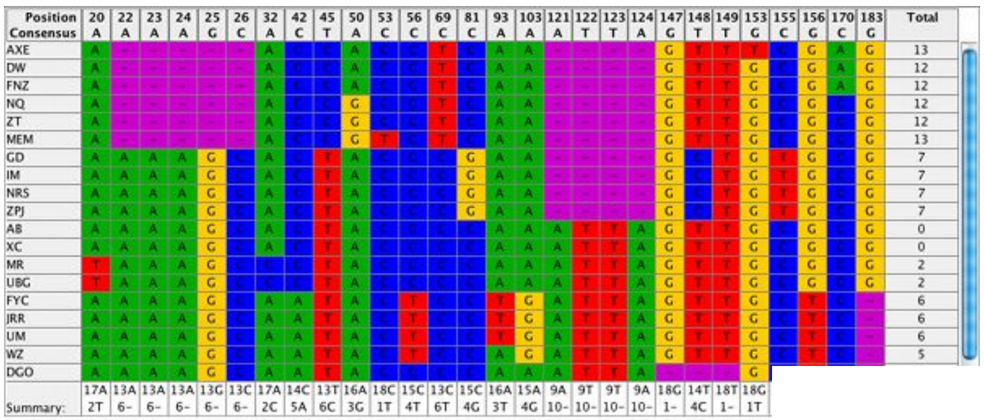
¿QUÉ ES UN MODELO?

Descripción de un sistema o proceso usando un lenguaje matemático para poder hacer predicciones

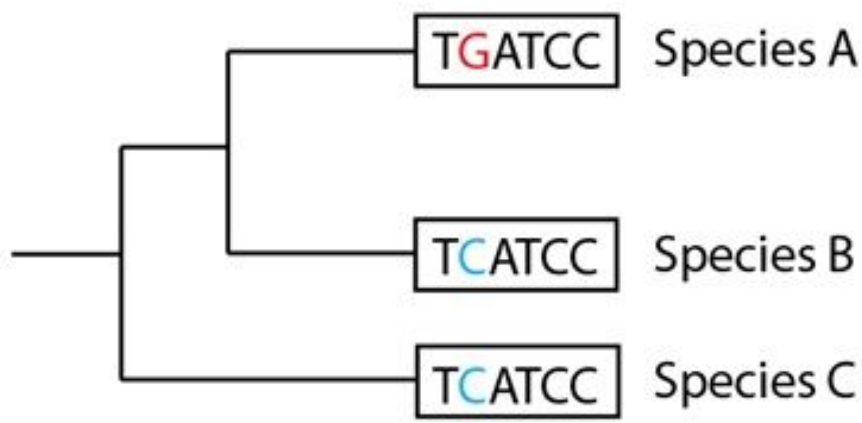
- Parámetros (variables)
- Función matemática

¿QUÉ ES UN MODELO DE SUSTITUCIÓN DE CARACTERES?

Modelos matemáticos que predicen como los caracteres evolucionan entre sus estados y tasas relativas de cambio en las ramas de un árbol



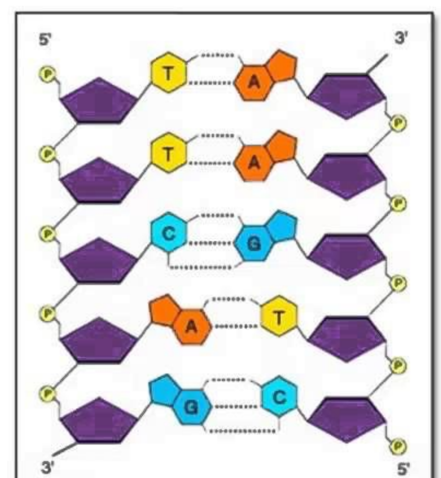
Position	20	22	23	24	25	26	32	42	45	50	53	56	69	81	93	103	121	122	123	124	147	148	149	153	155	156	170	183	Total
Consensus	A	A	A	A	G	C	A	C	T	A	C	C	C	C	A	A	A	T	T	A	G	T	T	G	C	G	C	G	
AXE	A	-	-	-	-	-	A	C	C	A	C	C	T	C	A	A	-	-	-	-	G	T	T	T	C	G	A	G	13
DW	A	-	-	-	-	-	A	C	C	A	C	C	T	C	A	A	-	-	-	-	G	T	T	G	C	G	A	G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

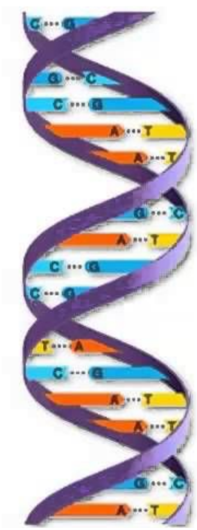


Homología en secuencias de ADN

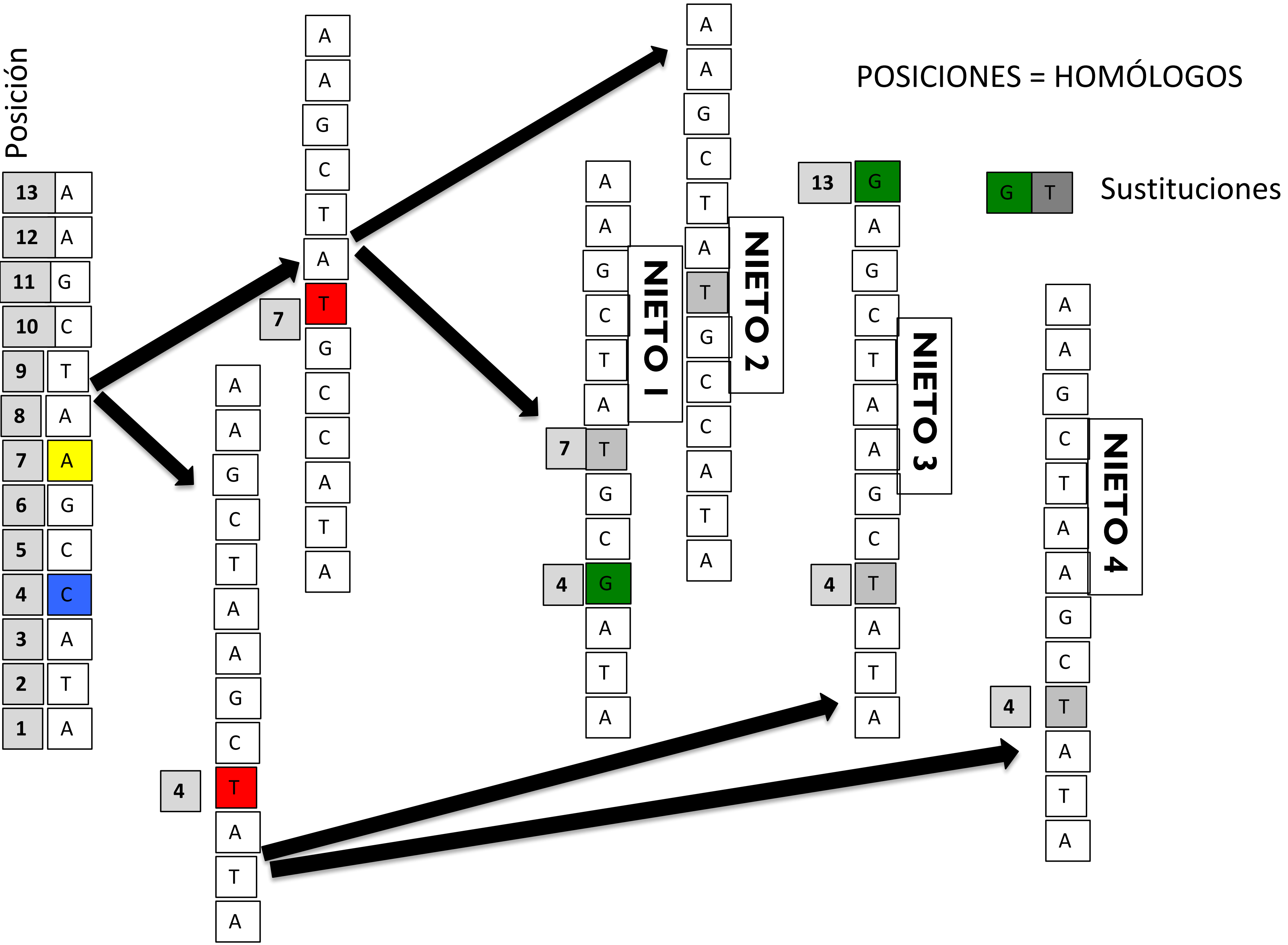
Alineamiento

Scarites	C	T	T	A	G	A	T	C	G	T	A	C	C	A	A	-	-	-	A	A	T	A	T	T	A	C
Carenum	C	T	T	A	G	A	T	C	G	T	A	C	C	A	C	A	-	T	A	C	-	T	T	T	A	C
Pasimachus	A	T	T	A	G	A	T	C	G	T	A	C	C	A	C	T	A	T	A	A	G	T	T	T	A	C
Pheropsophus	C	T	T	A	G	A	T	C	G	T	T	C	C	A	C	-	-	-	A	C	A	T	A	T	A	C
Brachinus armiger	A	T	T	A	G	A	T	C	G	T	A	C	C	A	C	-	-	-	A	T	A	T	A	T	T	C
Brachinus hirsutus	A	T	T	A	G	A	T	C	G	T	A	C	C	A	C	-	-	-	A	T	A	T	A	T	A	C
Aptinus	C	T	T	A	G	A	T	C	G	T	A	C	C	A	C	-	-	-	A	C	A	A	T	T	A	C
Pseudomorpha	C	T	T	A	G	A	T	C	G	T	A	C	C	-	-	-	-	-	A	C	A	A	A	T	A	C





Evolución de secuencias de ADN



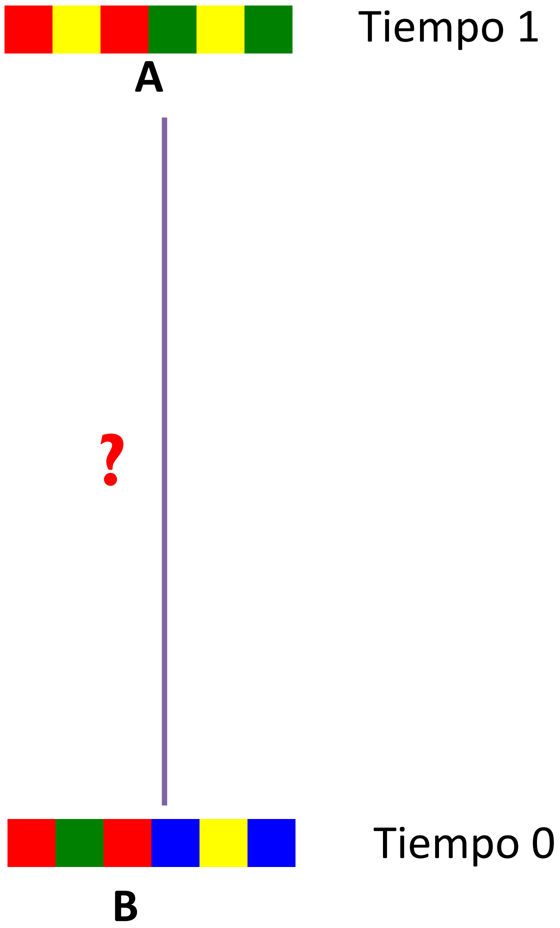
SUPUESTOS DE LOS PARÁMETROS DE LOS MODELOS

- **Propiedad de Markov:** La probabilidad de sustitución de un caracter depende únicamente del estado presente (independiente de estados anteriores)
- **Homegeneidad:** Las tasas de sustitución en un sitio no cambian con el tiempo
- **Estacionalidad:** Las frecuencias relativas de los estados de caracter están en equilibrio
- Las tasas de cambio son reversibles en el tiempo

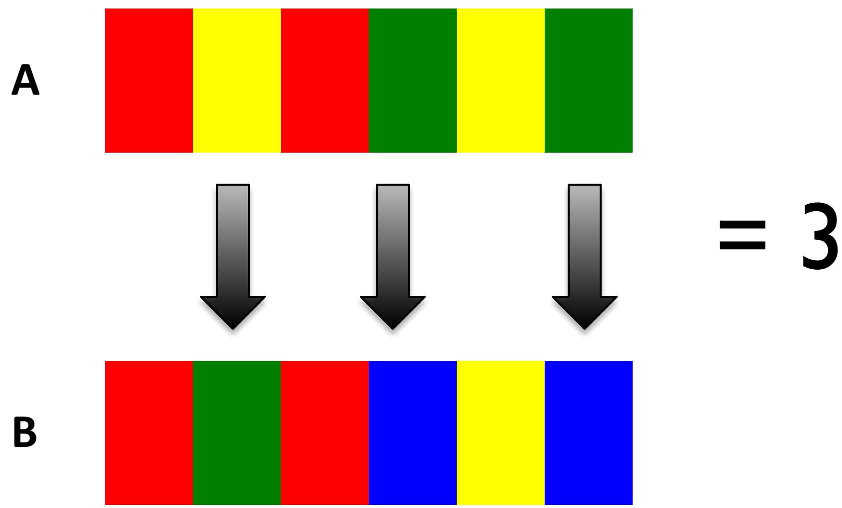
SUPUESTOS DE EVOLUCIÓN DE LOS CARACTERES

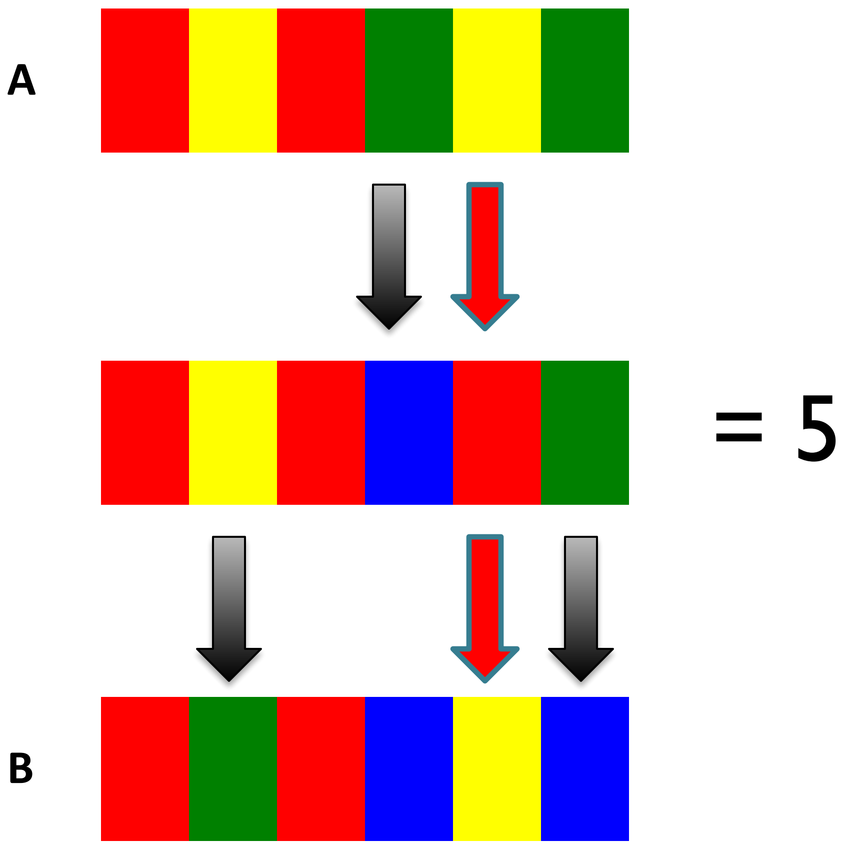
- Caracteres son neutrales
- Caracteres evolucionan independientemente
- Caracteres cambian a un número finito de estados
- Caracteres evolucionan a lo largo de las ramas
- Las ramas tienen una duración específica

Distancia evolutiva



Promedio de sustituciones que han ocurrido en cada caracter


A
B
= 3


A
B
= 5

¿Cómo estimar el promedio de cambios que hubo entre las dos secuencias observadas a través del tiempo?

Calculando la probabilidad de observar cada cambio en función de:

- Tasa de sustitución: μ
- Tiempo: t

MATRIZ DE PROBABILIDAD DE SUSTITUCIÓN

		To:			
		A	C	G	T
From:	A	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	C	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	G	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	T	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$

¿Cuál es la probabilidad de comenzar con A y terminar en A?

$$\frac{1}{4} + \frac{3}{4} e^{-4/3\mu t}$$

Si μt es pequeño, entonces $e^{-4/3\mu t}$ es cercano a uno

Si μt es grande, entonces $e^{-4/3\mu t}$ es cercano a cero

Toma en cuenta todas las historias posibles de transformación

MATRIZ DE PROBABILIDAD DE SUSTITUCIÓN

Jukes-Cantor (1969) = JC69

		To:			
		A	C	G	T
From:	A	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	C	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	G	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	T	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$

- Cuatro bases son igualmente frecuentes
- Los tipos de sustituciones ocurren a la misma tasa
- La tasa de sustitución en igual a través de todos los sitios de la secuencia

Pero sabemos que:

- Las cuatro bases generalmente no se encuentran con la misma frecuencia
- Algunos tipos de sustitución ocurren a diferentes tasa que otros (Ts vs. Tv)
- Algunas posiciones de la secuencia evolucionan a tasa más rápidas que otras

MODELOS MÁS REALISTAS DE EVOLUCIÓN MOLECULAR

Felsenstein (1981) = F81

		To:			
		A	C	G	T
From:	A	$\pi_A + (1 - \pi_A)e^{-mt}$	$\pi_C(1 - e^{-mt})$	$\pi_G(1 - e^{-mt})$	$\pi_T(1 - e^{-mt})$
	C	$\pi_A(1 - e^{-mt})$	$\pi_C + (1 - \pi_C)e^{-mt}$	$\pi_G(1 - e^{-mt})$	$\pi_T(1 - e^{-mt})$
	G	$\pi_A(1 - e^{-mt})$	$\pi_C(1 - e^{-mt})$	$\pi_G + (1 - \pi_G)e^{-mt}$	$\pi_T(1 - e^{-mt})$
	T	$\pi_A(1 - e^{-mt})$	$\pi_C(1 - e^{-mt})$	$\pi_G(1 - e^{-mt})$	$\pi_T + (1 - \pi_T)e^{-mt}$

Permite que las cuatro bases estén a differentes frecuencias

Parámetro Π

$(\Pi_A \neq \Pi_T \neq \Pi_G \neq \Pi_C)$

MODELOS MÁS REALISTAS DE EVOLUCIÓN MOLECULAR

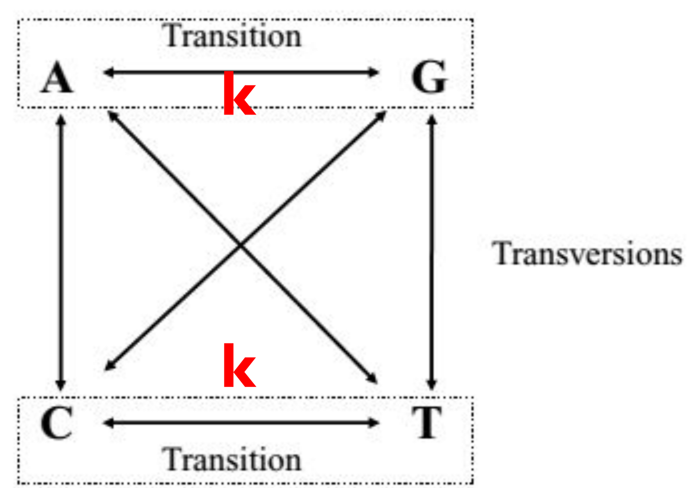
HKY85

From:	To: A (freq = π_A)	To: C (freq = π_C)	To: G (freq = π_G)	To: T (freq = π_T)
A (freq = π_A)	$-m(\pi_C + \kappa\pi_G + \pi_T)$	$\pi_C m$	$\pi_G \kappa m$	$\pi_T m$
C (freq = π_C)	$\pi_A m$	$-m(\pi_A + \pi_G + \kappa\pi_T)$	$\pi_G m$	$\pi_T \kappa m$
G (freq = π_G)	$\pi_A \kappa m$	$\pi_C m$	$-m(\kappa\pi_A + \pi_C + \pi_T)$	$\pi_T m$
T (freq = π_T)	$\pi_A m$	$\pi_C \kappa m$	$\pi_G m$	$-m(\pi_A + \kappa\pi_C + \pi_G)$

Permite que las tasas relativas de los tipos de sustitución sean diferentes

Transiciones más probables que transversiones

Parámetro **"k"** de tasa relativa de cambio



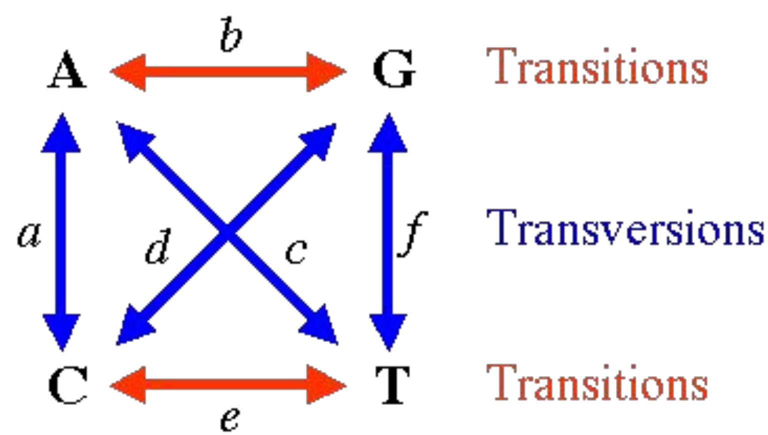
MODELOS MÁS REALISTAS DE EVOLUCIÓN MOLECULAR

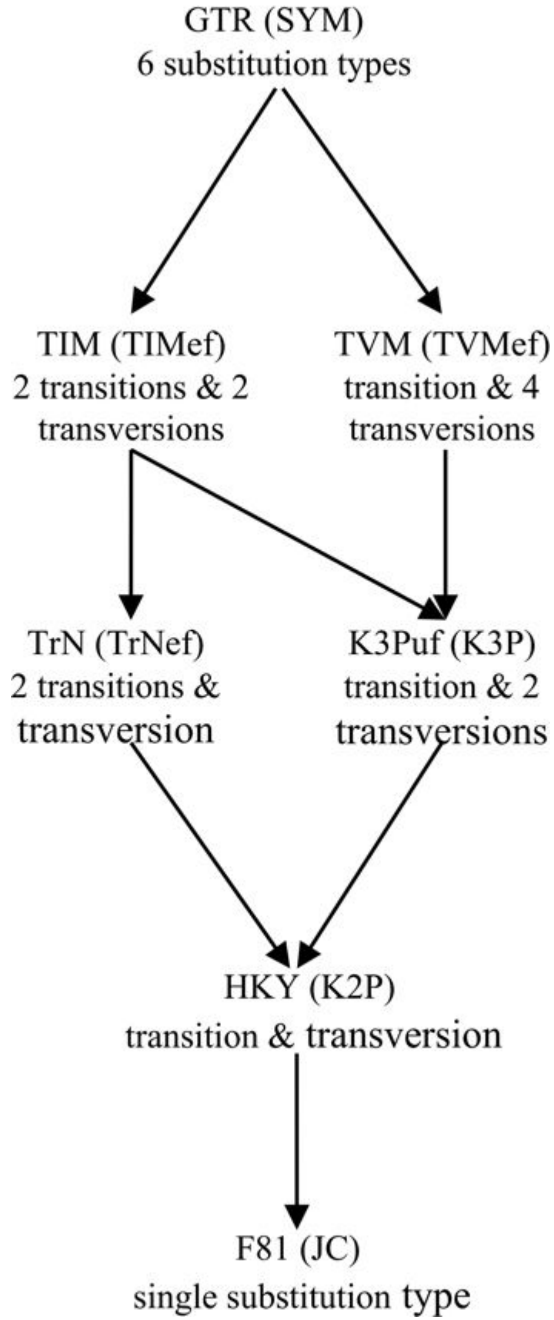
General time-reversible model (GTR)

$$
Q=\begin{array}{c}
\begin{array}{cccc} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \end{array} \\
\begin{pmatrix}
-\mu(a\pi_C+b\pi_G+c\pi_T) & a\mu\pi_C & b\mu\pi_G & c\mu\pi_T \\
a\mu\pi_A & -\mu(a\pi_A+d\pi_G+e\pi_T) & d\mu\pi_G & e\mu\pi_T \\
b\mu\pi_A & d\mu\pi_C & -\mu(b\pi_A+d\pi_C+f\pi_T) & f\mu\pi_T \\
c\mu\pi_A & e\mu\pi_C & f\mu\pi_G & -\mu(c\pi_A+e\pi_C+f\pi_G)
\end{pmatrix}
\end{array}
$$

El más complejo: todas las tasas relativas de cambio son diferentes

Adiciona un parámetro diferente para cada tipo de cambio



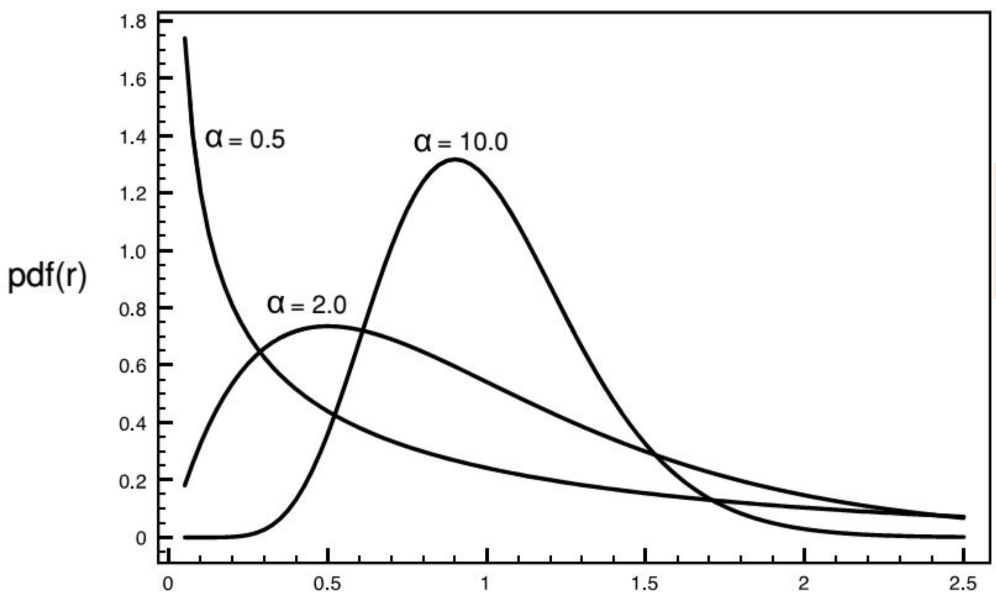

GTR (SYM)
6 substitution types
TIM (TIMef)
2 transitions & 2 transversions
TVM (TVMef)
transition & 4 transversions
TrN (TrNef)
2 transitions & transversion
K3Puf (K3P)
transition & 2 transversions
HKY (K2P)
transition & transversion
F81 (JC)
single substitution type

¿Que pasa si las posiciones evolucionan a tasas diferentes dentro de la secuencia de ADN?

P.e. Codones (lento en bases que modifican aminoácidos)

ESTRATÉGIAS:

- Particionar los datos y agruparlos de acurdo a sus tasas de evolución.
- Asumir que la tasa de sustitución es diferente a través de las posiciones y muestrear valores con respecto a una distribución de tasas.



Distribución gamma de tasas de sustitución

¿EXISTEN MODELOS DE EVOLUCIÓN MORFOLÓGICA?

Modelo "Mk" (Markov y número de estados observados)

- Generalización de JC69 (k = 4 estados)
- Usado también para AFLPs, INDELS, proteínas

Syst. Biol. 50(6):913–925, 2001

A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data

PAUL O. LEWIS

Department of Ecology and Evolutionary Biology, The University of Connecticut, Storrs, Connecticut 06269-3043, USA; E-mail: paul.lewis@uconn.edu

Abstract.—Evolutionary biologists have adopted simple likelihood models for purposes of estimating ancestral states and evaluating character independence on specified phylogenies; however, for purposes of estimating phylogenies by using discrete morphological data, maximum parsimony remains the only option. This paper explores the possibility of using standard, well-behaved Markov models for estimating morphological phylogenies (including branch lengths) under the likelihood criterion. An important modification of standard Markov models involves making the likelihood conditional on characters being variable, because constant characters are absent in morphological data sets. Without this modification, branch lengths are often overestimated, resulting in potentially serious biases in tree topology selection. Several new avenues of research are opened by an explicitly model-based approach to phylogenetic analysis of discrete morphological data, including combined-data likelihood analyses (morphology + sequence data), likelihood ratio tests, and Bayesian analyses. [Discrete morphological character; Markov model; maximum likelihood; phylogeny.]

Lewis (2001)

Syst. Biol. 65(4):602–611, 2016

DOI:10.1093/sysbio/syv122
Advance Access publication December 28, 2015

Modeling Character Change Heterogeneity in Phylogenetic Analyses of Morphology through the Use of Priors

APRIL M. WRIGHT[1,*], GRAEME T. LLOYD[2], AND DAVID M. HILLIS[1]

[1]Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA; [2]Department of Biological Sciences, Macquarie University, NSW 2109, Australia
**Correspondence to be sent to: Department of Integrative Biology, University of Texas at Austin, 2401 Speedway Austin, TX 78712, USA; E-mail: wright.aprilm@gmail.com.*

Received 30 April 2015; reviews returned 14 December 2015; accepted 15 December 2015
Associate Editor: Peter Foster

Abstract.—The Mk model was developed for estimating phylogenetic trees from discrete morphological data, whether for living or fossil taxa. Like any model, the Mk model makes a number of assumptions. One assumption is that transitions between character states are symmetric (i.e., the probability of changing from 0 to 1 is the same as 1 to 0). However, some characters in a data matrix may not satisfy this assumption. Here, we test methods for relaxing this assumption in a Bayesian context. Using empirical data sets, we perform model fitting to illustrate cases in which modeling asymmetric transition rates among characters is preferable to the standard Mk model. We use simulated data sets to demonstrate that choosing the best-fit model of transition-state symmetry can improve model fit and phylogenetic estimation. [Bayesian estimation, morphology, paleontology, phylogeny, priors.]

Wright et al. (2016)

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

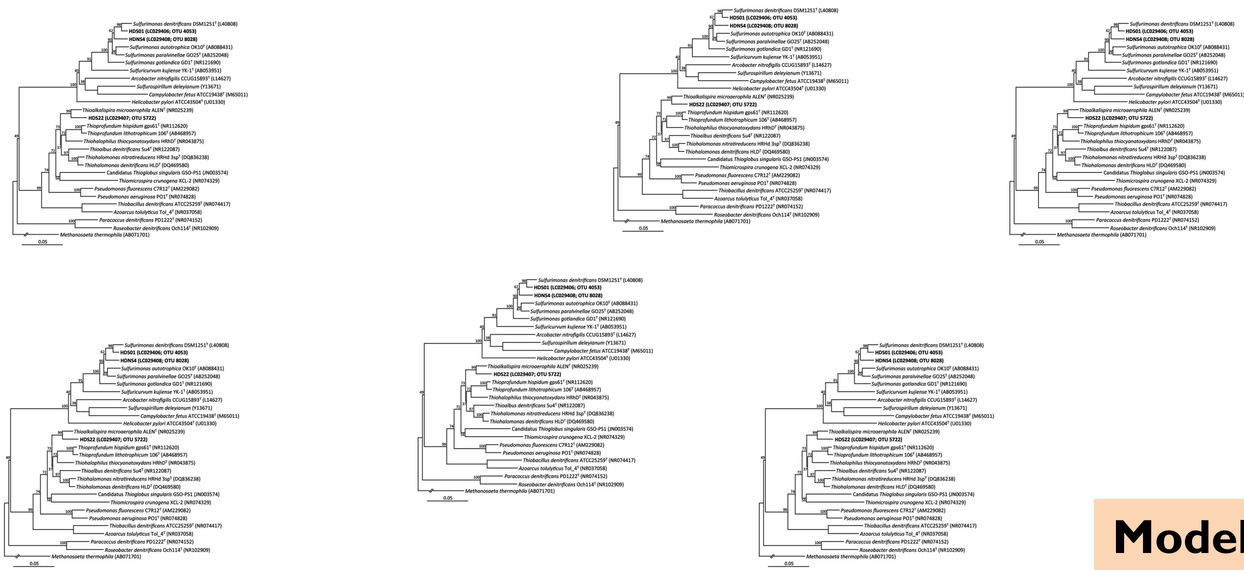
Datos

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

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

Datos

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 (caracter)

3. Se multiplican las verosimilitudes de todas las posiciones (caracteres)

4. Se usa un algoritmo para optimizar la longitud de ramas y otros parámetros (repitiendo 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

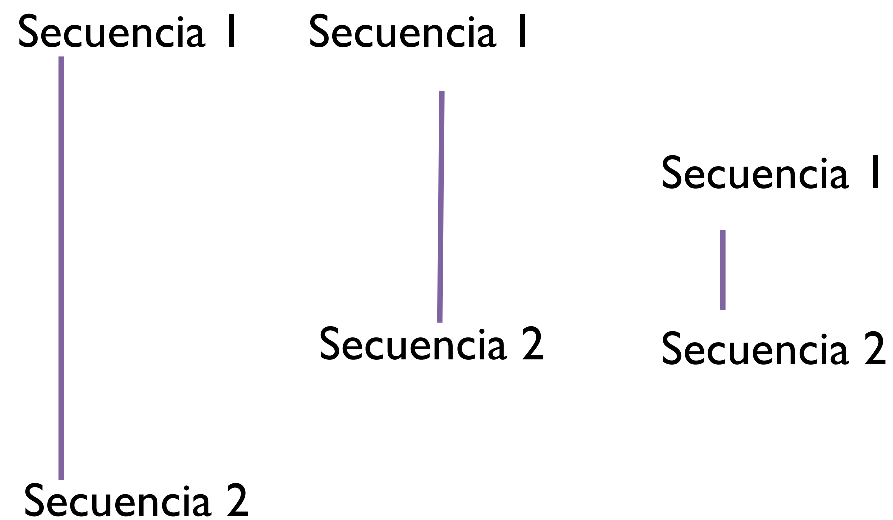
PASOS:

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

DATO S

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

HIPÓTESIS



Modelo

Jukes-Cantor (1969)

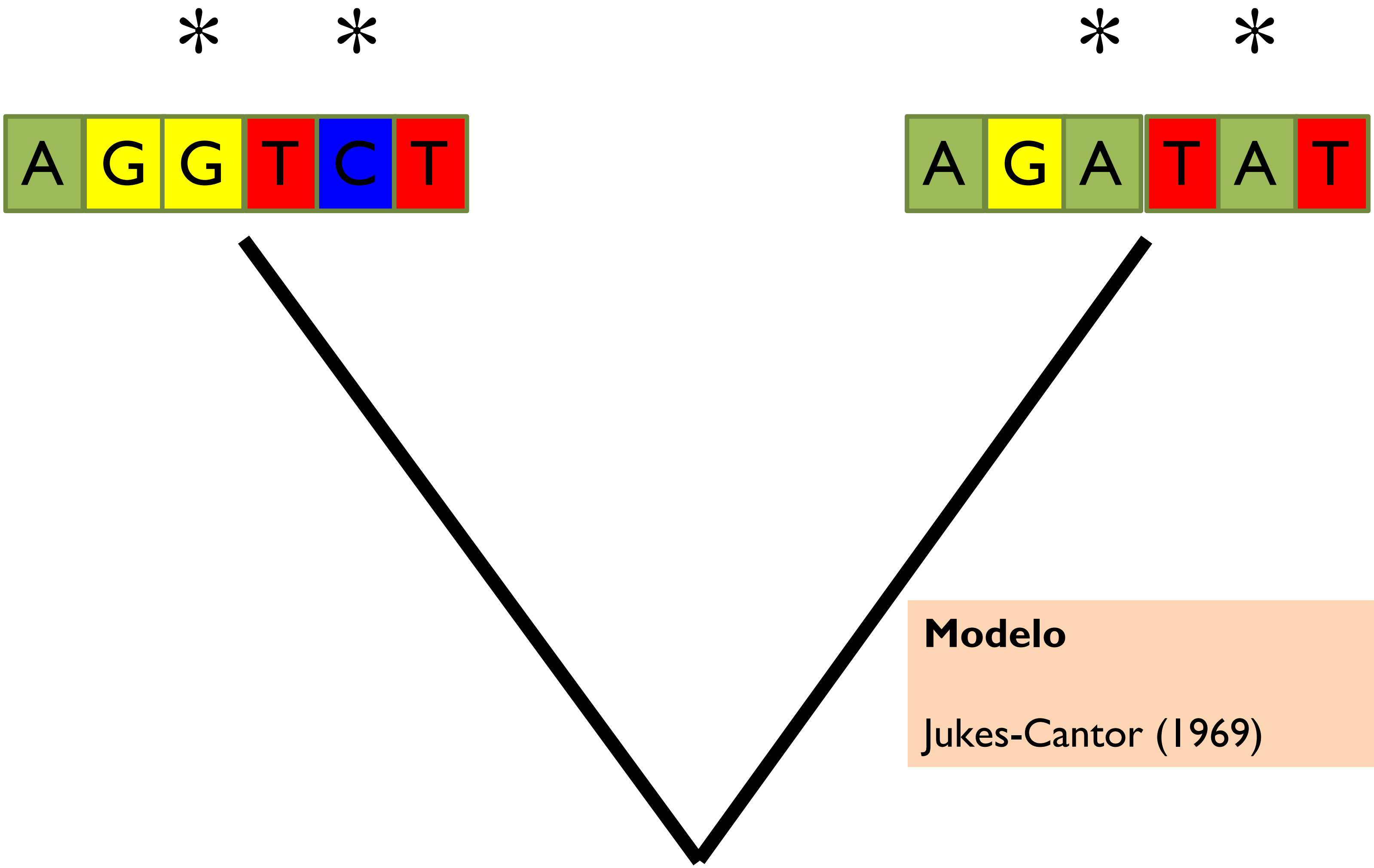
MÁXIMA VEROSIMILITUD EN FILOGENÉTICA

PASOS:

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

2. Se calcula la verosimilitud de cada posición (caracter)

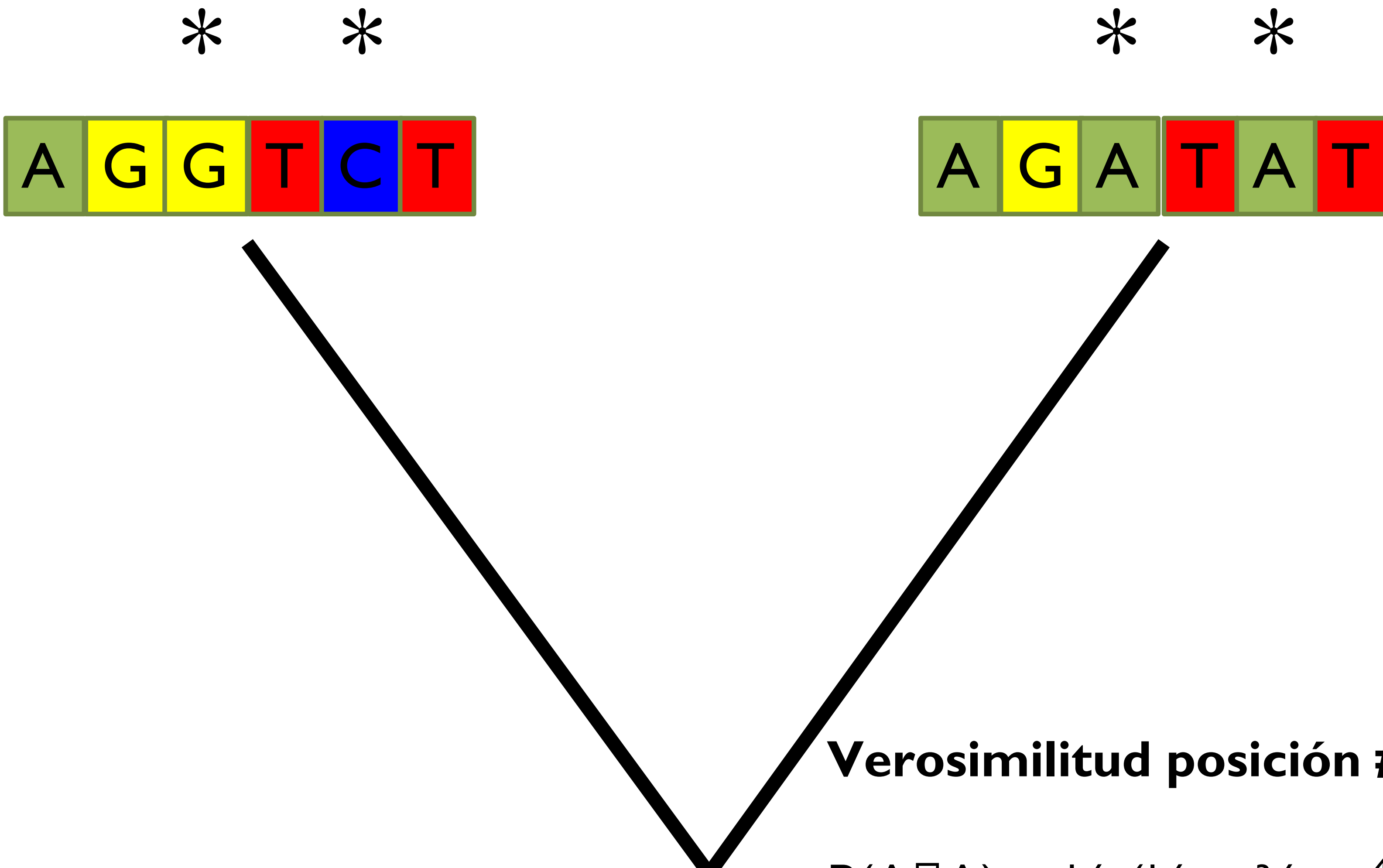
MÁXIMA VEROSIMILITUD EN FILOGENÉTICA



Modelo

Jukes-Cantor (1969)

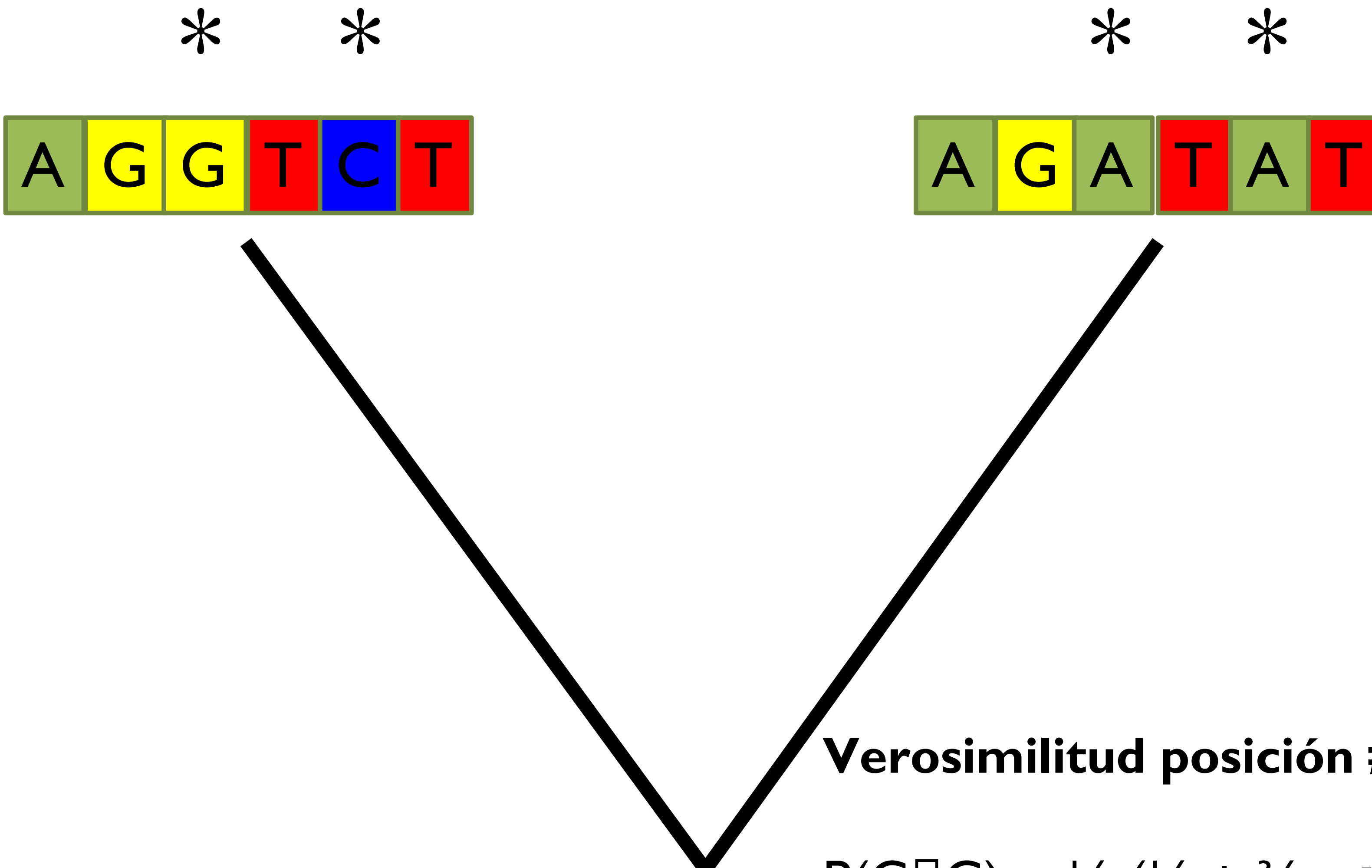
MÁXIMA VEROSIMILITUD EN FILOGENÉTICA



Verosimilitud posición # 1:

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

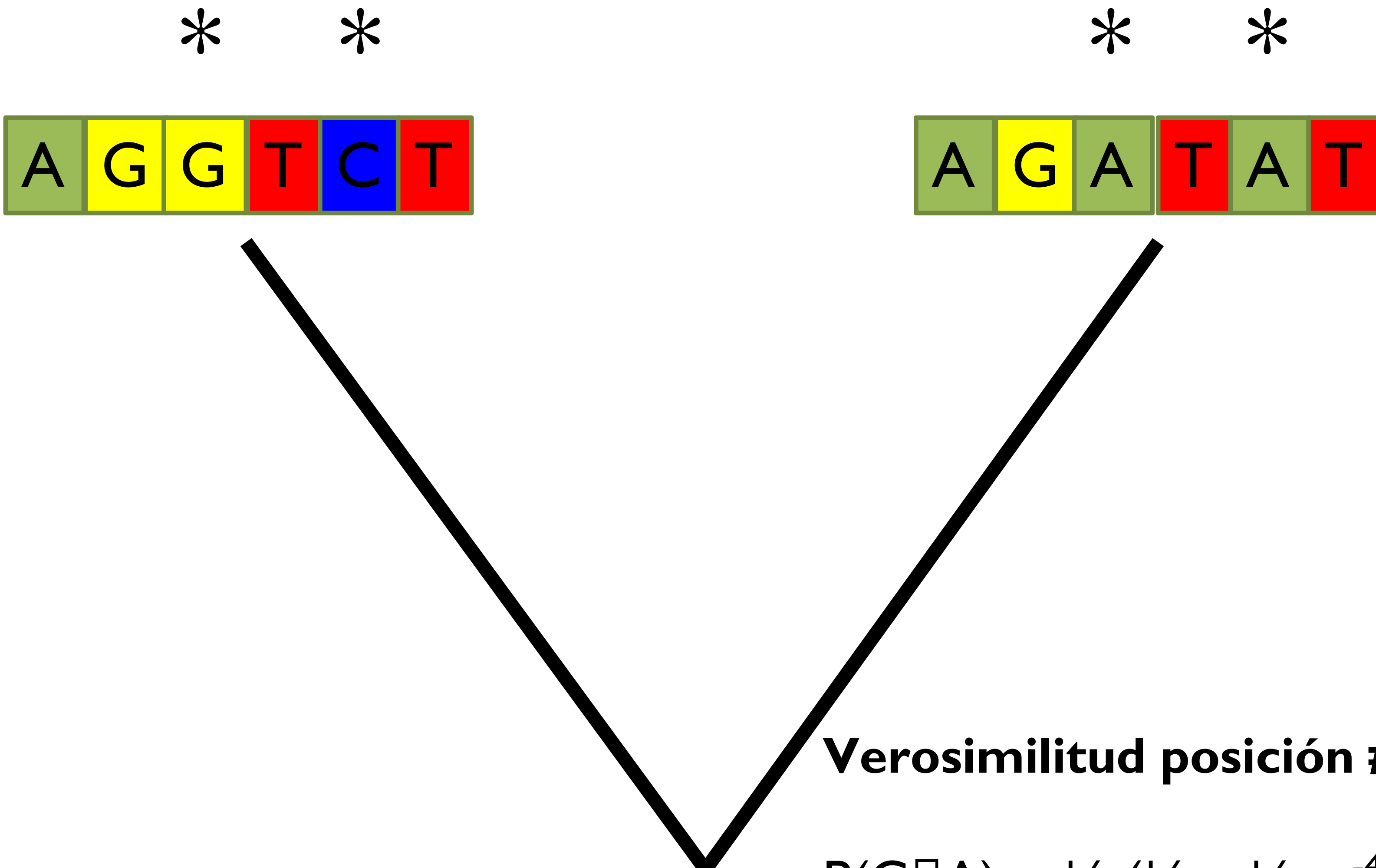
MÁXIMA VEROSIMILITUD EN FILOGENÉTICA



Verosimilitud posición # 2:

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

MÁXIMA VEROSIMILITUD EN FILOGENÉTICA



Verosimilitud posición # 3:

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

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 (caracter)

3. Se multiplican las verosimilitudes de todas las posiciones (caracteres)

Verosimilud total (L) del árbol

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

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 (caracter)

3. Se multiplican las verosimilitudes de todas las posiciones (caracteres)

4. Se usa un algoritmo para optimizar la longitud de ramas y otros parámetros (repitiendo pasos 1–3) hasta que se maximice la verosimilitud del árbol

Verosimilud total (ML) del árbol

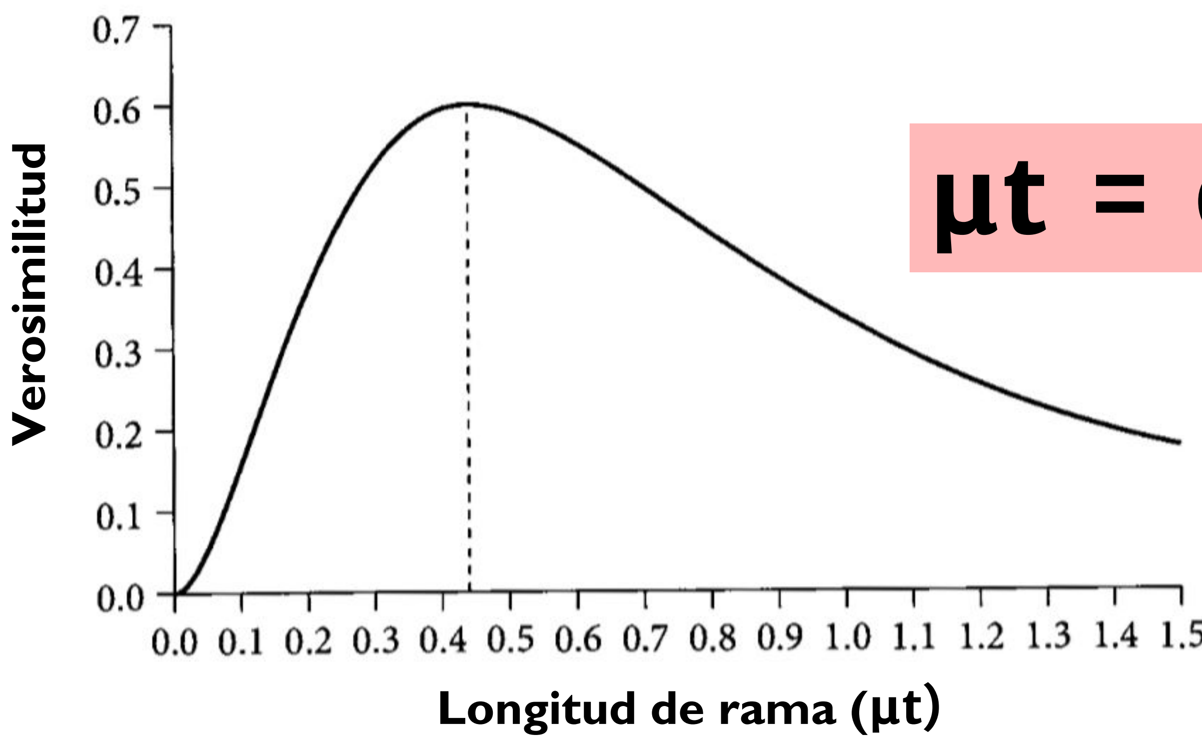
¿¿¿ μt ???

$$L = \underbrace{[\tfrac{1}{4}\,(\tfrac{1}{4} + \tfrac{3}{4}\, e^{-4/3\boxed{\mu t}})]^4}_{\substack{A \to A \\ G \to G \\ T \to T \\ T \to T}} * \underbrace{[\tfrac{1}{4}\,(\tfrac{1}{4} - \tfrac{1}{4}\, e^{-4/3\boxed{\mu t}})]^2}_{\substack{G \to A \\ C \to A}}$$

	*		*		
A	G	G	T	C	T

	*		*		
A	G	A	T	A	T

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



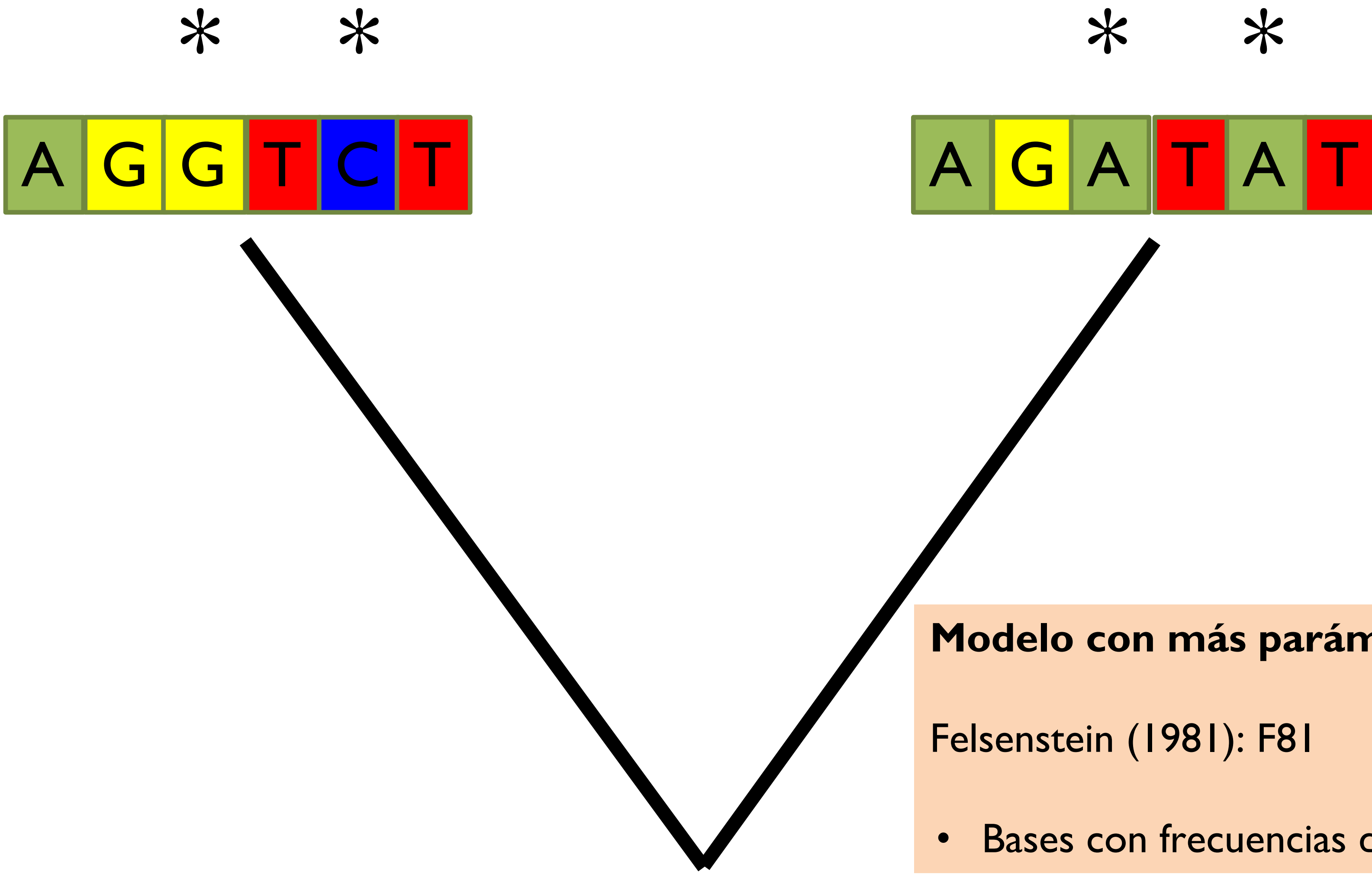
$\mu t = 0.44$

Reemplazando…

$$L = [\tfrac{1}{4}\,(\tfrac{1}{4} + \tfrac{3}{4}\, e^{-4/3\boxed{(0.44)}})]^4 * [\tfrac{1}{4}\,(\tfrac{1}{4} - \tfrac{1}{4}\, e^{-4/3\boxed{(0.44}})]^2$$

$$L = 0.000000595$$

$$\boxed{Ln(L) = -14.33}$$

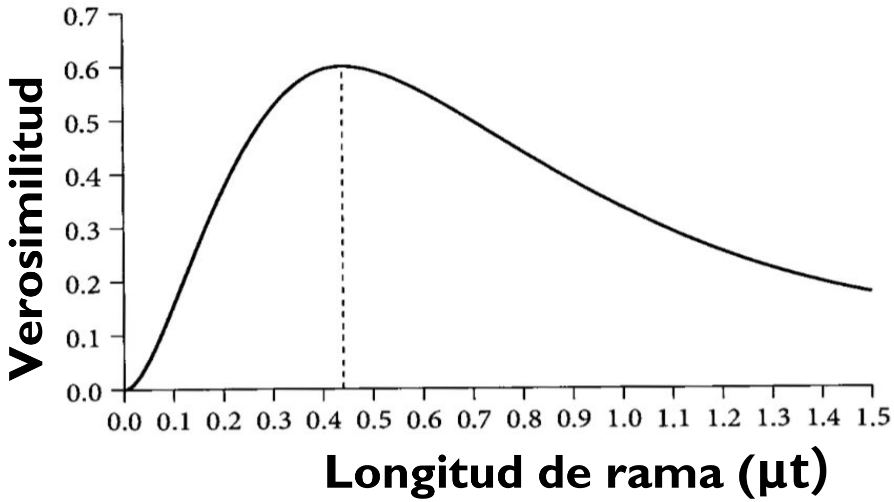


Modelo con más parámetros

Felsenstein (1981): F81

- Bases con frecuencias desiguales

Modelo F81: además de estimar µt…



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

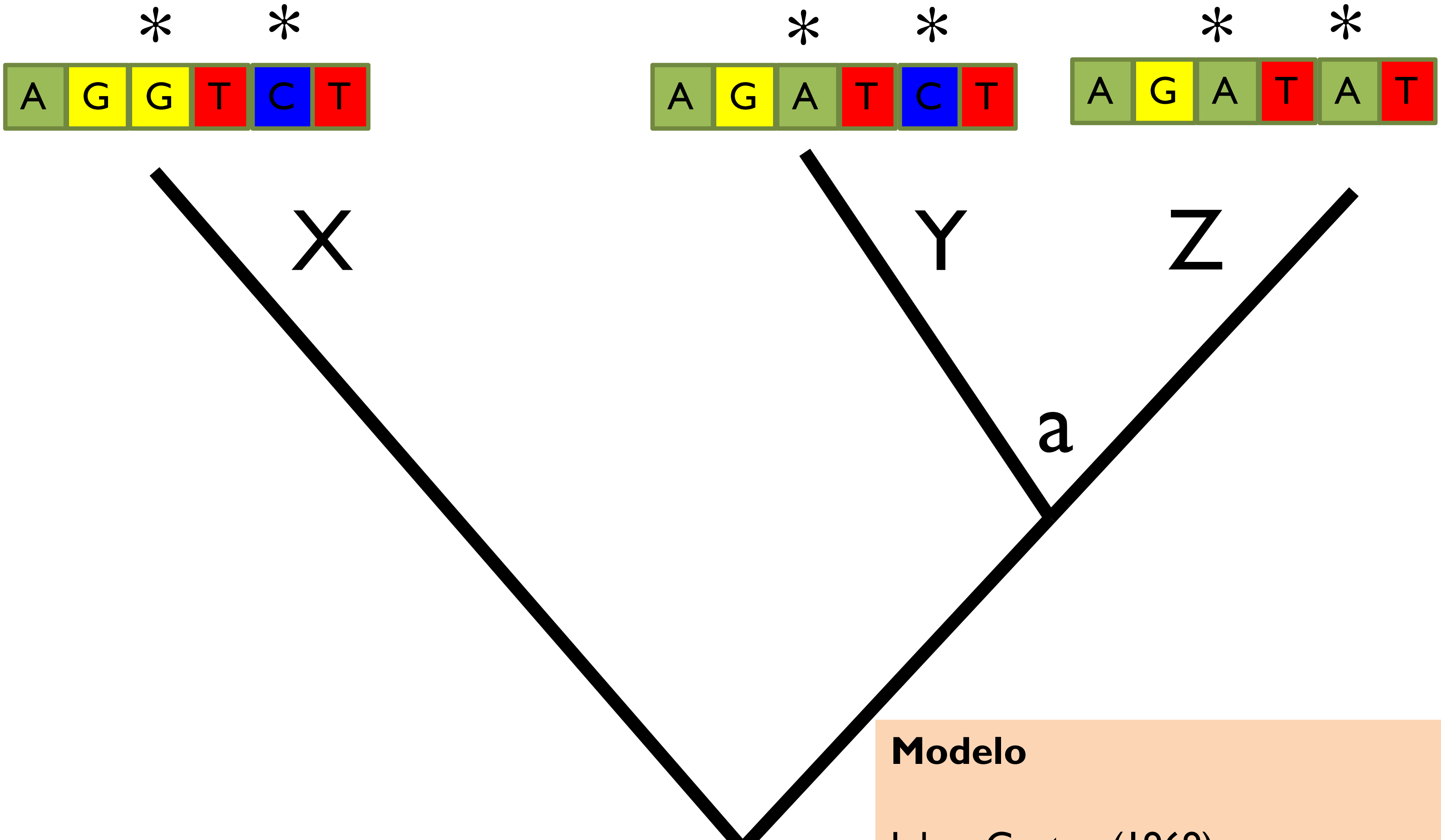
2. Se calcula la verosimilitud de cada posición (caracter)

3. Se multiplican las verosimilitudes de todas las posiciones (caracteres)

4. Se usa un algoritmo para optimizar la longitud de ramas y otros parámetros (repitiendo 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

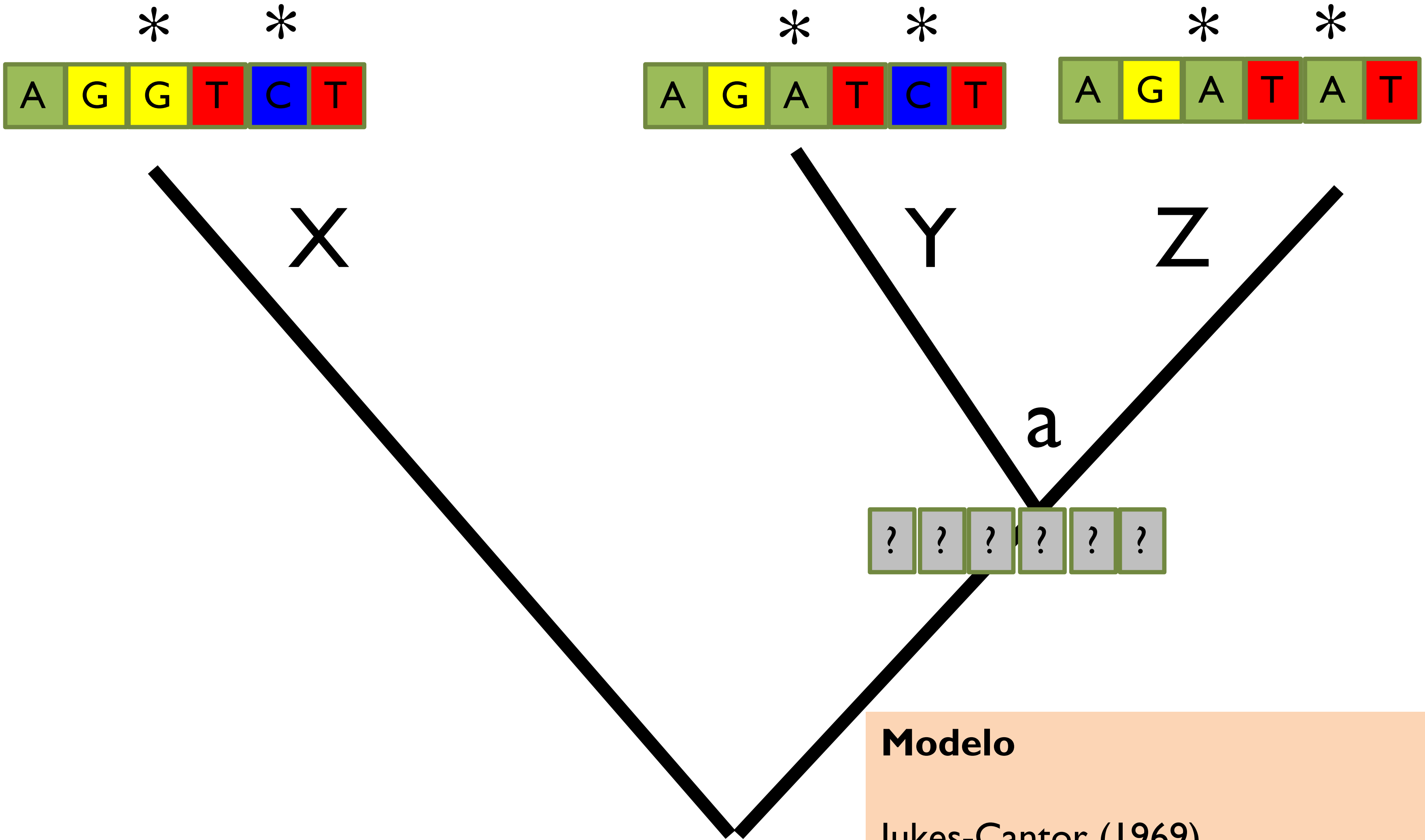
¿Y si adicionamos terminales al árbol?



Modelo

Jukes-Cantor (1969)

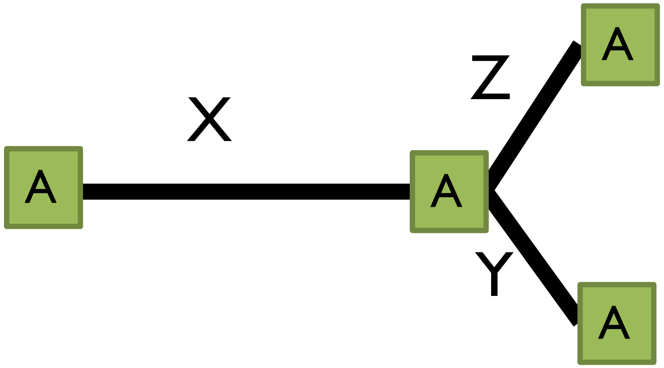
¿Y si adicionamos terminales al árbol?



Modelo

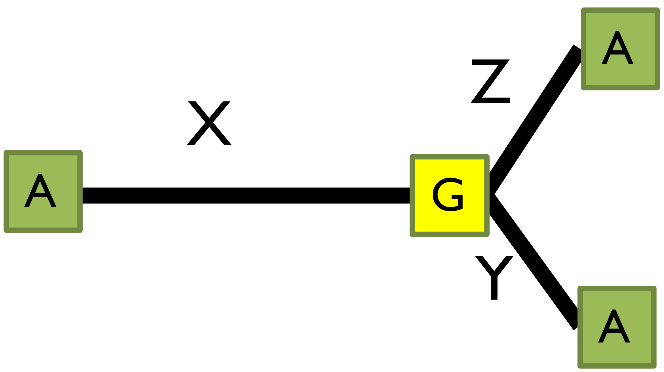
Jukes-Cantor (1969)

Verosimilitud posición # 1



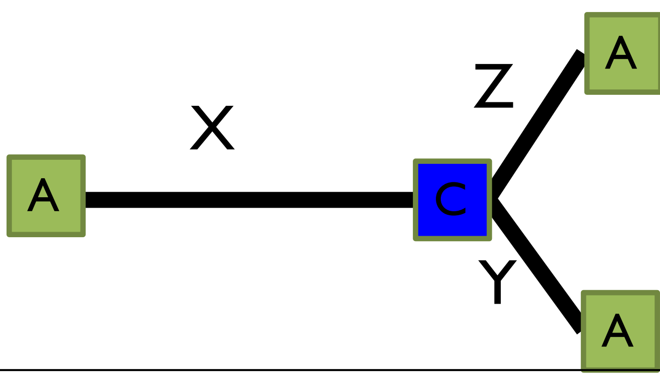
$$P(A\square A) = [\tfrac{1}{4} (\tfrac{1}{4} + \tfrac{3}{4} e^{-4/3\mu t})]^3$$

+



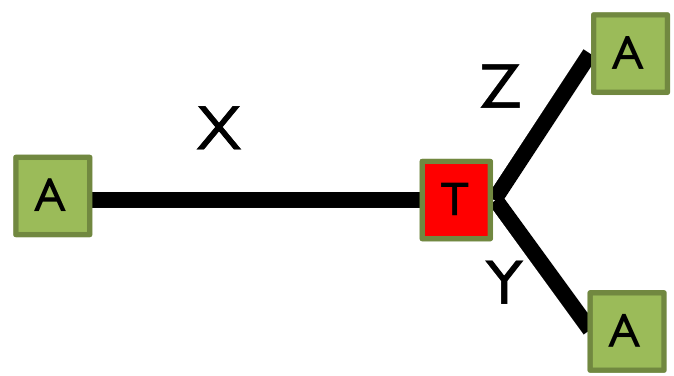
$$P(A\square G) = [\tfrac{1}{4} (\tfrac{1}{4} - \tfrac{1}{4} e^{-4/3\mu t})]^3$$

+



$$P(A\square C) = [\tfrac{1}{4} (\tfrac{1}{4} - \tfrac{1}{4} e^{-4/3\mu t})]^3$$

+



$$P(A\square T) = [\tfrac{1}{4} (\tfrac{1}{4} - \tfrac{1}{4} e^{-4/3\mu t})]^3$$