

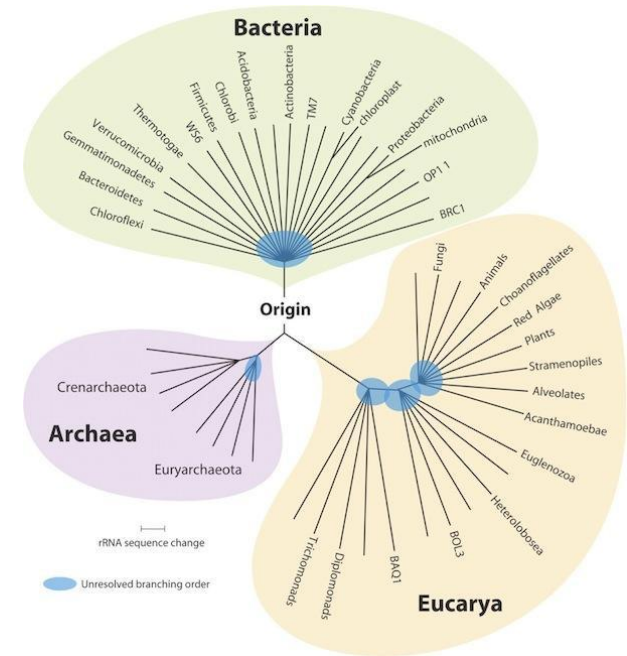
# Curso de Evolución 2021

Facultad de Ciencias

Montevideo, Uruguay

<http://evolucion.fcien.edu.uy/>

<http://eva.fcien.universidad.edu.uy/>



Tema 2. Las filogenias como contexto de análisis de la evolución. Aplicaciones del análisis filogenético. Análisis filogenético según el principio de parsimonia. Métodos basados en distancias y en modelos de evolución molecular.

# Algunos métodos de inferencia filogenética

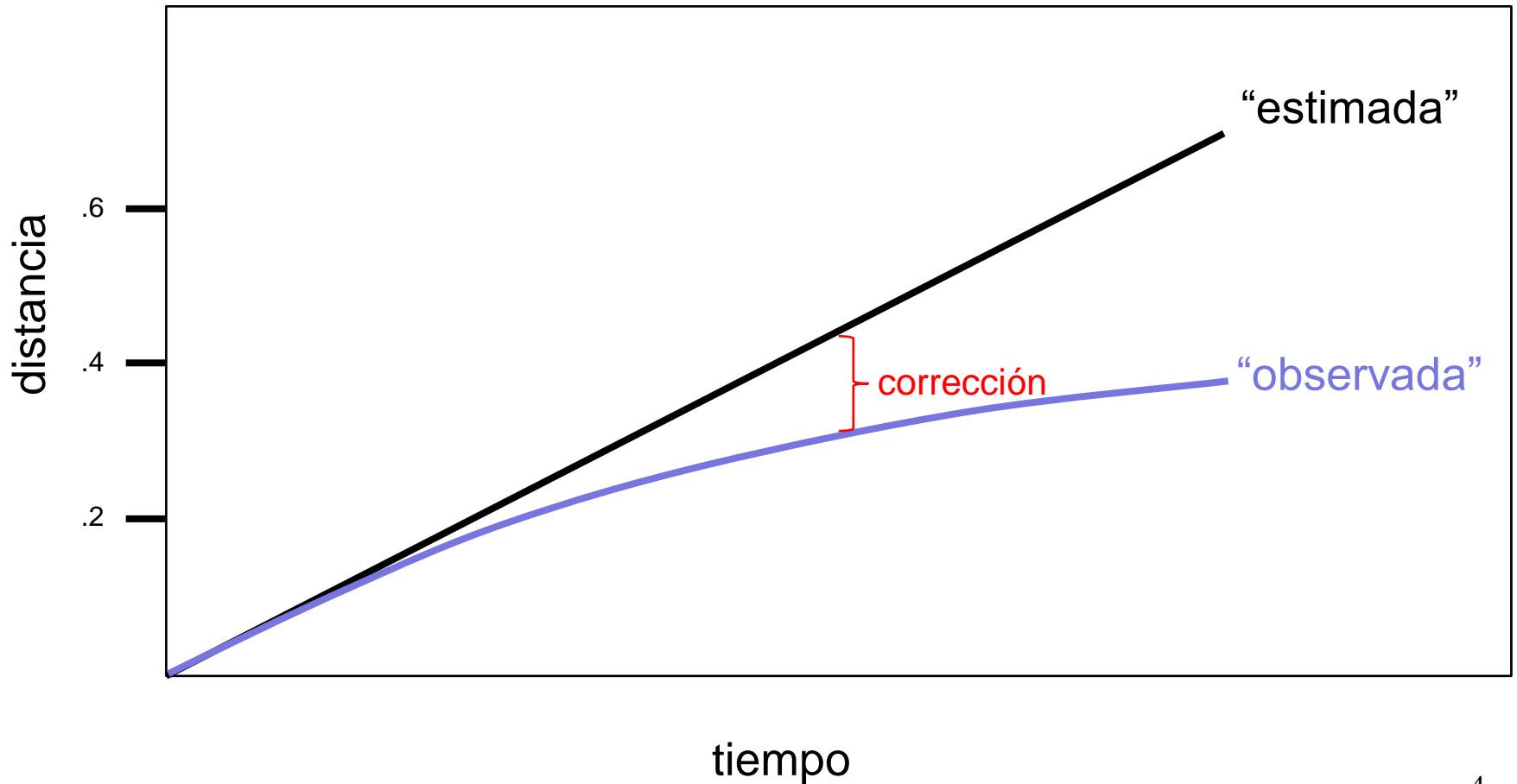
<b>Método</b>	<b>Variantes</b>	<b>Criterio de optimización</b>	<b>Uso de variación no observada</b>
Parsimonia	máxima parsimonia	minimizar el número de pasos requeridos para obtener los datos	No
Distancias	- evolución mínima	minimizar los cambios requeridos para obtener las distancias estimadas entre taxa	Sí (incorporadas en las distancias)
	- unión de vecinos (neighbor joining)	una aproximación al árbol de evolución mínima	ídem
Inferencia estadística	...		

# Optimización: parsimonia vs. distancias

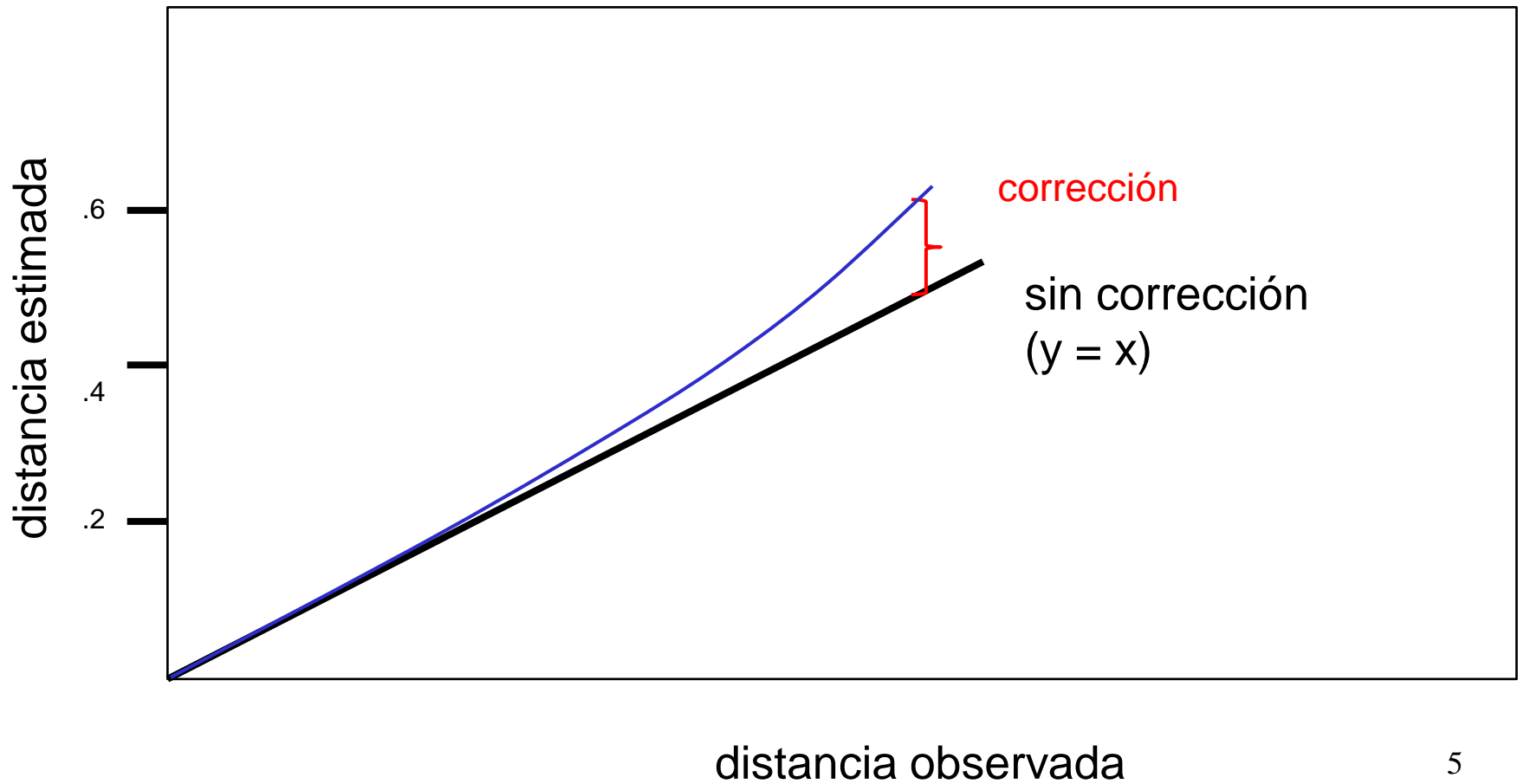
	Parsimonia	Distancias
Criterio de optimización	Minimizar la longitud (número de pasos) del árbol	Minimizar la longitud (suma de todas las ramas, medidas como distancias) del árbol
Efecto de la homoplasia*	Los mejores árboles requieren más pasos que el mínimo ideal	1) Las distancias estimadas son mayores a las observadas.  2) El árbol óptimo requiere un largo total mayor al mínimo requerido por las distancias.

Homoplasia: similitud que no heredada de un ancestro común. En contraste con homología: similitud resultante de una condición heredada de un ancestro común. Estos conceptos se aplican a caracteres, así como a estados de caracteres.

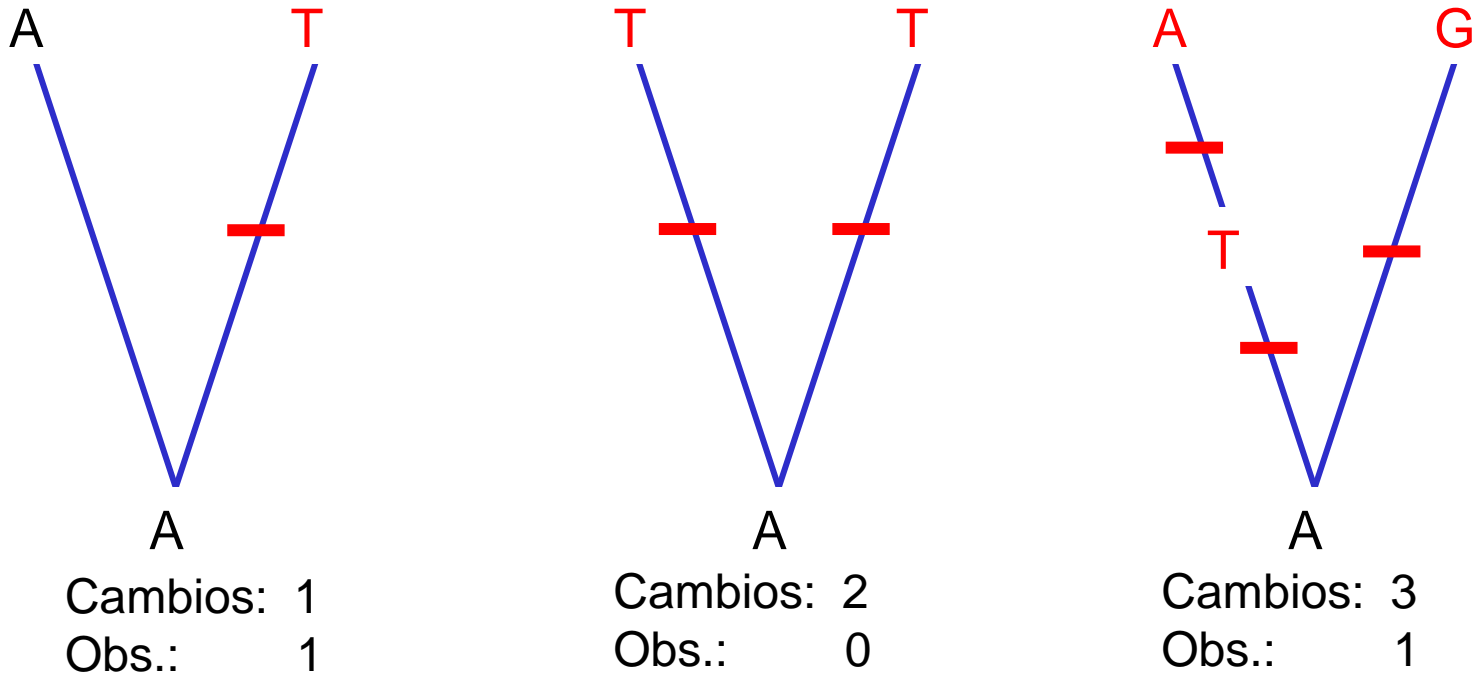
# ¿Por qué calcular distancias?



# Adicional 2021



# Distancia observada $\leq$ “Distancia real”

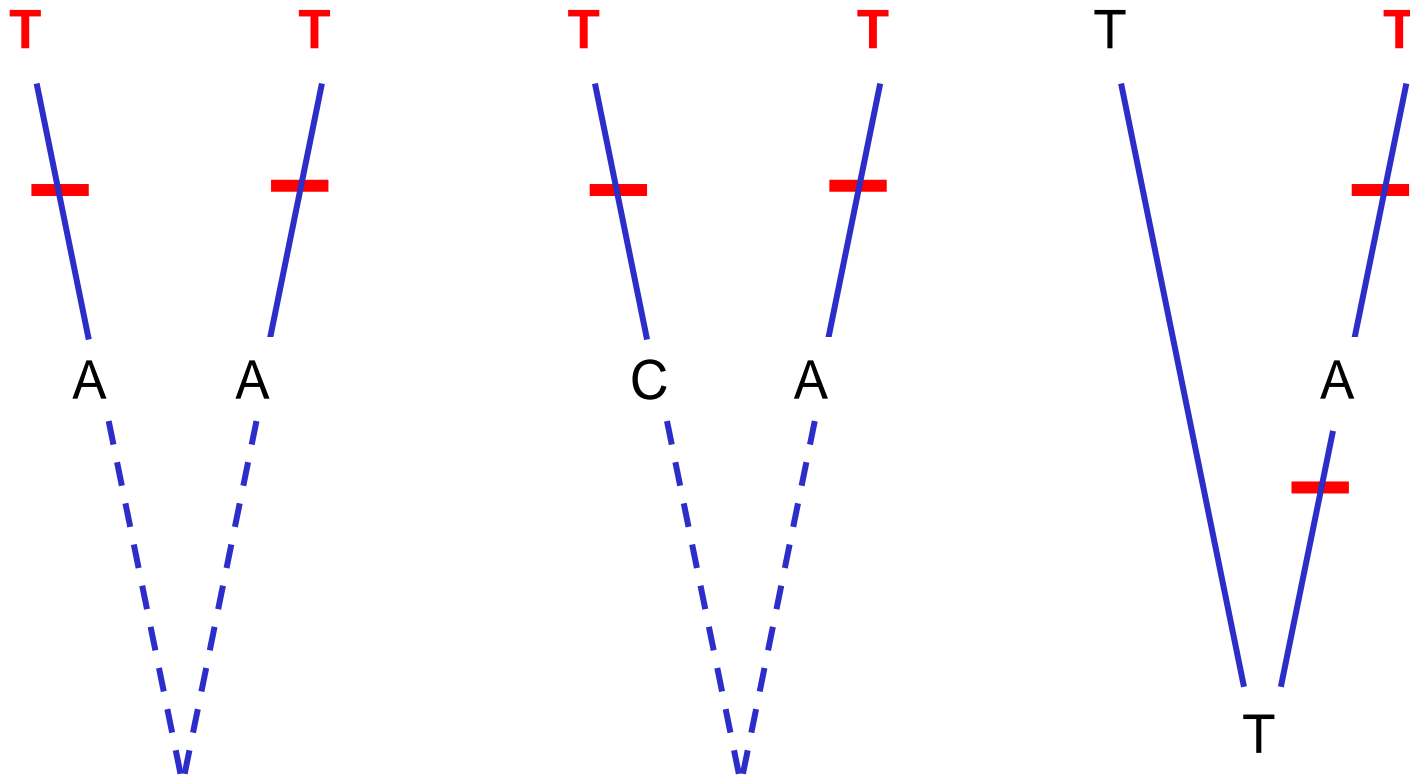


---

“distancia real”: cambios acumulados en la evolución

Distancia observada: diferencias observadas entre las secuencias finales

# Homoplasia: paralelismo, convergencia, reversión



# Administrando la homoplasia: parsimonia vs. distancias

- Parsimonia:
  - No intenta evaluar (o corregir) cambios no observados como diferencias en los datos.
  - La homoplasia resulta en cambios (pasos) adicionales, o sea árboles más largos al mínimo posible.
  - El criterio de máxima parsimonia procura minimizar la longitud del árbol.
- Distancias:
  - Las distancias corregidas (estimadas) procuran, precisamente, ajustar los cambios no observados, que resultan en homoplasia.



# Relación entre distancias y modelos de evolución molecular

- Para calcular la distancia "estimada" (corregida con una estimación de los cambios no observados), necesitamos la distancia observada y un modelo de evolución molecular.
- Ejemplos: modelo de Jukes y Cantor:
  - La distancia correspondiente es  $d = \frac{3}{4} \ln(1 - \frac{4}{3} p)$ , siendo  $p$  la distancia observada (proporción de sitios diferentes).

	G	A	T	C
G	$1-3\mu$	$\mu$	$\mu$	$\mu$
A	$\mu$	$1-3\mu$	$\mu$	$\mu$
T	$\mu$	$\mu$	$1-3\mu$	$\mu$
C	$\mu$	$\mu$	$\mu$	$1-3\mu$

# Relación entre distancias y modelos de evolución molecular

El modelo de Jukes y Cantor tiene un solo parámetro ( $\mu$ ).

	G	A	T	C
G	$1-3\mu$	$\mu$	$\mu$	$\mu$
A	$\mu$	$1-3\mu$	$\mu$	$\mu$
T	$\mu$	$\mu$	$1-3\mu$	$\mu$
C	$\mu$	$\mu$	$\mu$	$1-3\mu$

Un modelo popular de dos parámetros es el de Kimura, que distingue transiciones y transversiones:

	G	A	T	C
G	$1-\alpha-2\beta$	$\alpha$	$\beta$	$\beta$
A	$\alpha$	$1-\alpha-2\beta$	$\beta$	$\beta$
T	$\beta$	$\beta$	$1-\alpha-2\beta$	$\alpha$
C	$\beta$	$\beta$	$\alpha$	$1-\alpha-2\beta$

## Motivación del modelo de Kimura 2 parámetros

### ADN mitocondrial (sitios variables)

Humano	G	G	T	C	T	C	T	A
Chimpancé	A	A	C	T	C	T	C	T

**A, G:** purinas

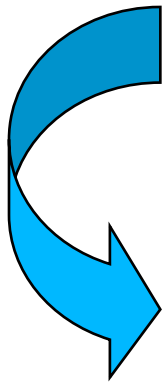
**C, T:** pirimidinas

7/8 diferencias son transiciones

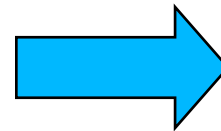
1/8 diferencias es una transversión

# Filogenias usando distancias

	Caracteres									
	1	2	3	4	5	6	7	8	9	10
Especie A	c	a	a	g	t	c	c	g	t	a
Especie B	.	.	t	.	.	t	.	a	.	.
Especie C	.	.	t	.	.	.	t	a	.	.
Especie D	t	g	.	.	c	.	.	.	.	g
Especie E	t	g	.	a	c	.	.	t	.	.



	A	B	C	D	E
A		0,3	0,3	0,4	0,5
B			0,2	0,6	0,7
C				0,7	0,7
D					0,3
E					



Árbol

# Método de unión de vecinos (NJ: “neighbor joining”)

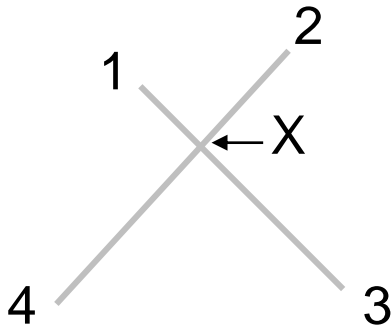
- NJ es una aproximación simple al criterio de evolución mínima:  
El árbol óptimo es aquel que requiere la menor longitud total (suma de todas las ramas, medidas como distancias moleculares).
- El algoritmo de NJ es sumamente eficiente, aunque es criticado como un algoritmo arbitrario.
- Al optar por optimizar sobre distancias,
  - Resumimos la información sobre las OTUs (caracteres y sus estados) en distancias entre pares de OTUs. (A diferencia de la parsimonia, que trabaja con los caracteres originales)
  - Las distancias pueden ser las observadas (número o fracción de sitios diferentes) o, más comúnmente, distancias estimadas (corregidas) en base a un modelo de evolución para incorporar cambios no observados. (A diferencia de la parsimonia, que no usa caracteres no observados, y solamente incorpora cambios adicionales cuando así lo requieren los árboles más parsimoniosos).

# Método de unión de vecinos (idea general 2021)

- Comenzamos con una policromía (todos los taxa unidos a un único nodo interno).
- Unimos el par de taxa más cercanos. (En el ejemplo omitimos un paso intermedio para corregir tasas desiguales de evolución molecular).
- Reemplazamos dicho par por su ancestro. Para eso, calculamos las distancias entre el ancestro del par reemplazado y los restantes taxones.
- Repetimos los pasos anteriores hasta resolver todos los nodos.
- Calculamos las longitudes de las ramas del árbol.

## Método NJ ilustrado con un ejemplo

- Punto de partida: la hipótesis nula ( $H_0$ ) para un ejemplo con 4 OTUs ( $m=4$ )
- Longitud  $L_0$  (en distancias) del árbol  $H_0$ :



$$S_0 = \sum_{i=1}^m L_{iX} = \frac{1}{m-1} \sum_{i<j}^m d_{ij}$$

Observamos que

$$d_{12} = L_{1X} + L_{2X}$$

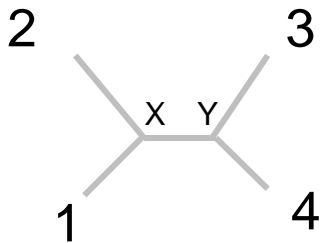
En la suma basada en  $d_{ij}$ ,  $L_{1X}$  forma parte de  $d_{12}$ ,  $d_{13}$  y  $d_{14}$ , para un total de  $m-1$  veces. Esto se aplica a todos los segmentos  $L_{iX}$ , por lo que dividimos la suma por  $m-1$ .





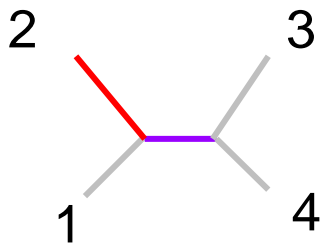
## Método NJ ilustrado con un ejemplo

- Longitud del nuevo árbol que une a 1 y 2 ( $S_{12}$ )



$$S_{12} = \underbrace{L_{1X} + L_{2X}}_{d_{12}} + L_{XY} + \underbrace{L_{3Y} + L_{4Y}}_{d_{34}}$$

$$L_{XY} = \frac{1}{4} \left( \underbrace{d_{13} + d_{14} + d_{23} + d_{24}}_{\text{sum of distances from X to tips 3 and 4}} - \underbrace{2d_{12} - 2d_{34}}_{\text{adjustment terms}} \right)$$





# Método NJ ilustrado con un ejemplo

Ejemplo: una matriz de distancias para 4 taxa:

	1	2	3
2	6		
3	11	11	
4	13	15	10

$$S_0 = 66/3 = 22$$

$$S_{12} = 6 + L_{XY} + 10$$

$$L_{XY} = [(11+13+11+15)-(2 \times 6)-(2 \times 10)]/4$$

$$L_{XY} = 4,5$$

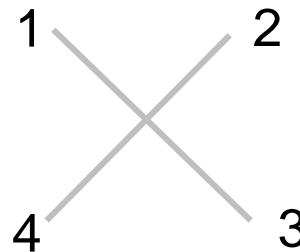
$$S_{12} = 6 + 4,5 + 10 = 20,5$$

$$S_0 = \sum_{i=1}^m = \frac{1}{m-1} \sum_{i < j}^m d_{ij}$$

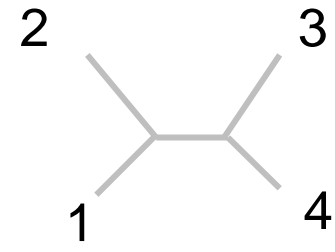
$$S_{12} = L_{1X} + L_{2X} + L_{XY} + L_{3Y} + L_{4Y}$$

$$\underbrace{\hspace{10em}}_{d_{12}} \quad \underbrace{\hspace{10em}}_{d_{34}}$$

$$L_{XY} = \frac{1}{4} (d_{13} + d_{14} + d_{23} + d_{24} - 2d_{12} - 2d_{34})$$



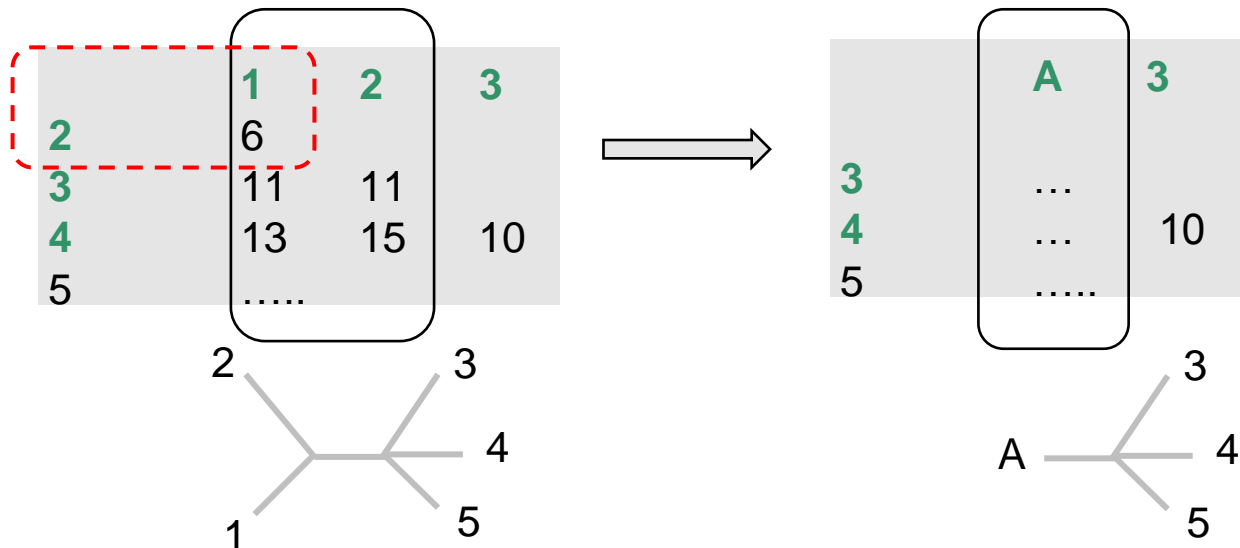
$$S_0 = 22$$



$$S_0 = 20,5$$

# Método NJ ilustrado con un ejemplo

- Para 4 taxa la topología está resuelta, y sólo resta calcular las longitudes de las ramas
- Para más taxa, tendríamos una politomía formada por todos los taxa menos 1 y 2.
- El algoritmo procede:
  - Reemplazando 1 y 2 por su ancestro común A en la matriz.
  - Calculando las distancias de A con los restantes taxa.
  - Buscando el siguiente par de vecinos más cercanos.... etc.



# Algunos métodos de inferencia filogenética

<b>Método</b>	<b>Variantes</b>	<b>Criterio de optimización</b>	<b>Uso de variación no observada</b>
Parsimonia	máxima parsimonia	minimizar el número de pasos requeridos para obtener los datos	No
Distancias	- evolución mínima	minimizar los cambios requeridos para obtener las distancias estimadas entre taxa	Sí (incorporadas en las distancias)
	- unión de vecinos (neighbor joining)	una aproximación al árbol de evolución mínima	ídem
Inferencia estadística	Máxima verosimilitud	maximizar la verosimilitud de observar los datos, dado un árbol y un modelo de evolución molecular.	Sí (considerando todos los estados posibles en los nodos).
	Inferencia bayesiana		

# Algunas comparaciones

- Parsimonia vs. otros:

- No se usan caracteres no observados.
- Se trabaja con los datos originales, sin resumirlos como distancias pareadas.
- No se utiliza un modelo de evolución (por ej., molecular)
- El uso de los datos es idéntico, independientemente de la fuente (molecular, morfológica, ...).  
(Aunque existen algunas variaciones sobre este punto, por ejemplo aquellas que dan pesos distintos a distintos caracteres o cambios de estado).

- Inferencia estadística (verosimilitud, bayesiana) vs. distancias:

- En común: modelos de evolución molecular y distancias asociadas.
- Diferencias: criterio y forma de optimización.

