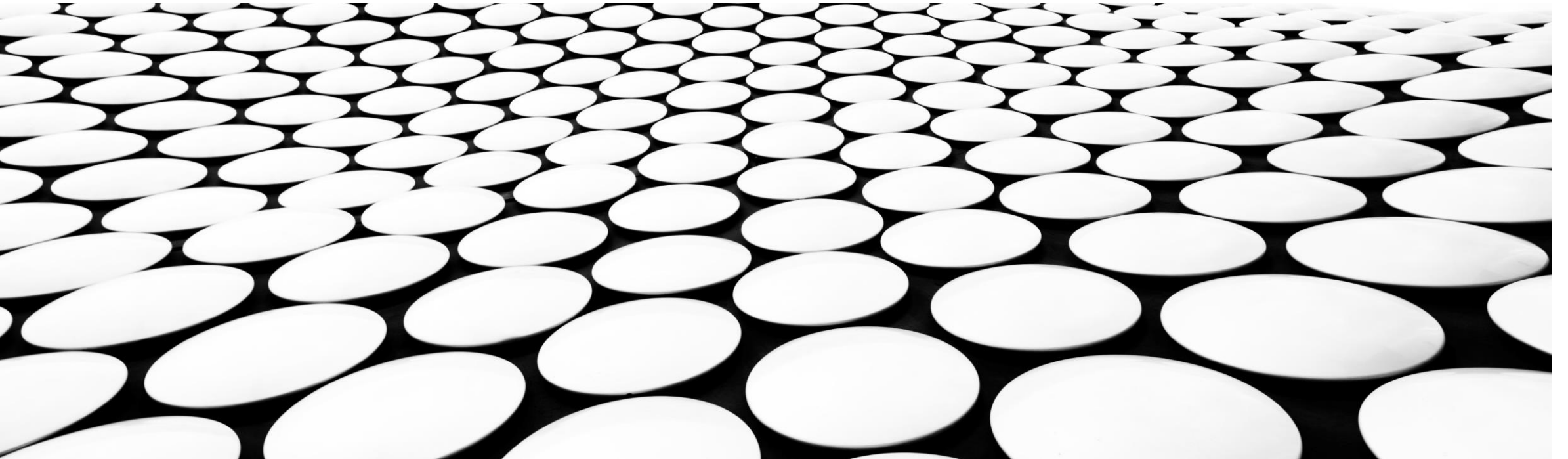

ARBOLES FILOGENÉTICOS

DRA. GUADALUPE DEL ROSARIO QUISPE SAJI



ARBOLES FILOGENETICOS

El ADN de cualquier tipo de ser vivo muta con el tiempo y parte de esas mutaciones son mantenidas

Cuando dos especies surgen a partir de un ancestro común, ellas van acumulando mutaciones distintas

- El número de mutaciones acumuladas tiende a ser proporcional a la divergencia de dos especies
- De esta manera el análisis de estas mutaciones nos permitiría inferir el proceso evolutivo de los organismos que son comparados

ARBOLES FILOGENETICOS

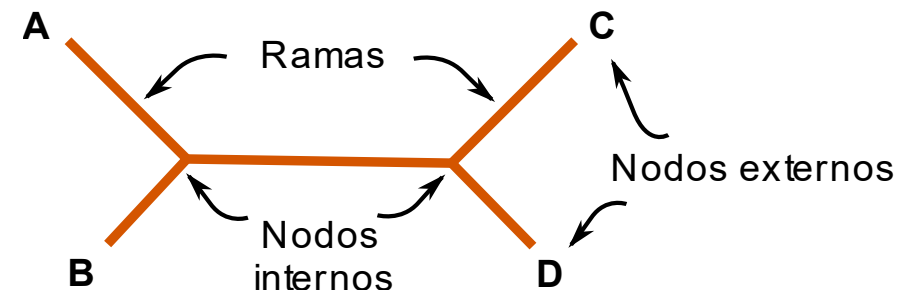
- Cuando analizamos secuencias de organismos distantemente relacionados, el ADN no tiene suficiente similitud para permitir una alineación confiable, debido a esto utilizamos secuencias de proteínas para el análisis evolutivo.
- Cabe recordar sin embargo que a la hora de analizar la evolución de las secuencias proteicas debemos tener en cuenta que las mutaciones que se producen en las proteínas son reflejo de mutaciones en el ADN y por tanto cualquier modelo para el estudio de la evolución proteica debe tener en cuenta el código genético.

ARBOLES FILOGENETICOS

Una forma de representar el proceso evolutivo de una familia de proteínas (o genes) es a través de un árbol filogenético

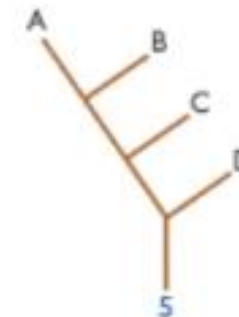
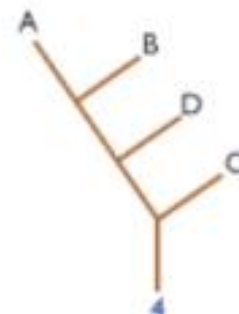
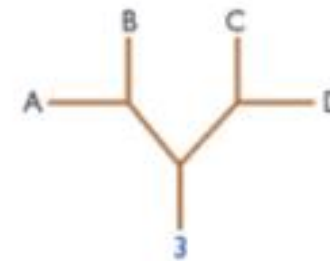
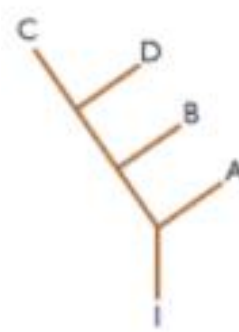
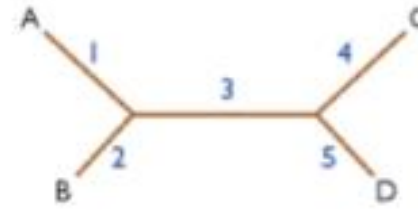
El árbol esta compuesto de nodos externos que representan los organismos o secuencias estudiadas. Las ramas son las líneas que conectan los nodos y tenemos ramas internas que conectan estas ramas.

El árbol se puede representar en la forma no enraizada, que presenta la evolución pero no determina el punto donde estaba el ancestro común.



ARBOLES FILOGENETICOS

En el ejemplo que se muestra al lado hay 5 puntos diferentes para la raíz del árbol y por lo tanto se pueden deducir 5 árboles enraizados, cada uno con diferentes implicaciones evolutivas



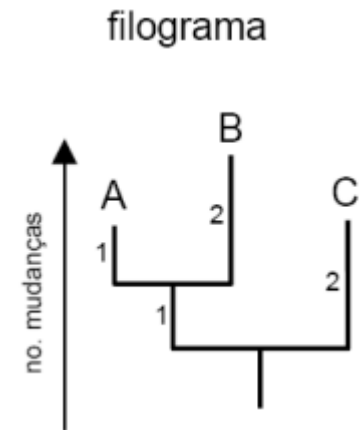
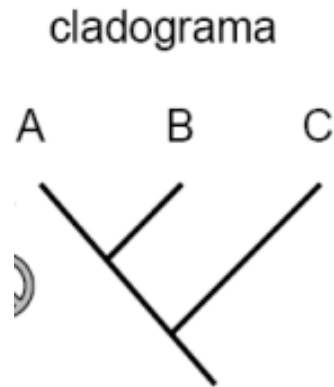
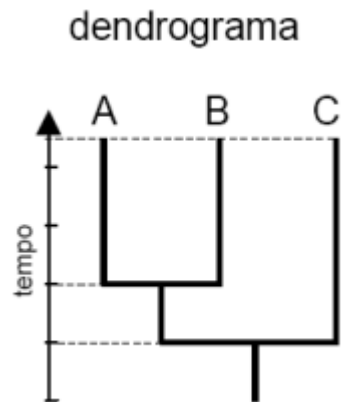
ARBOLES FILOGENETICOS

Enfoques para determinar donde se encuentra la raíz del árbol

Uno de los enfoques posibles es colocar la raíz en el punto medio del árbol (punto equidistante lejos de los nodos terminales), sin embargo, este tipo de enfoque asume que las tasas de evolución de los genes son uniformes, lo que a menudo no es cierto.

Un segundo enfoque ampliamente utilizado es la elección de un “*outgroup*”, una proteína que se sabe que es más distante del resto del grupo y el árbol ahora está enraizado en la rama que une esta proteína a los otros grupos.

ARBOLES FILOGENETICOS

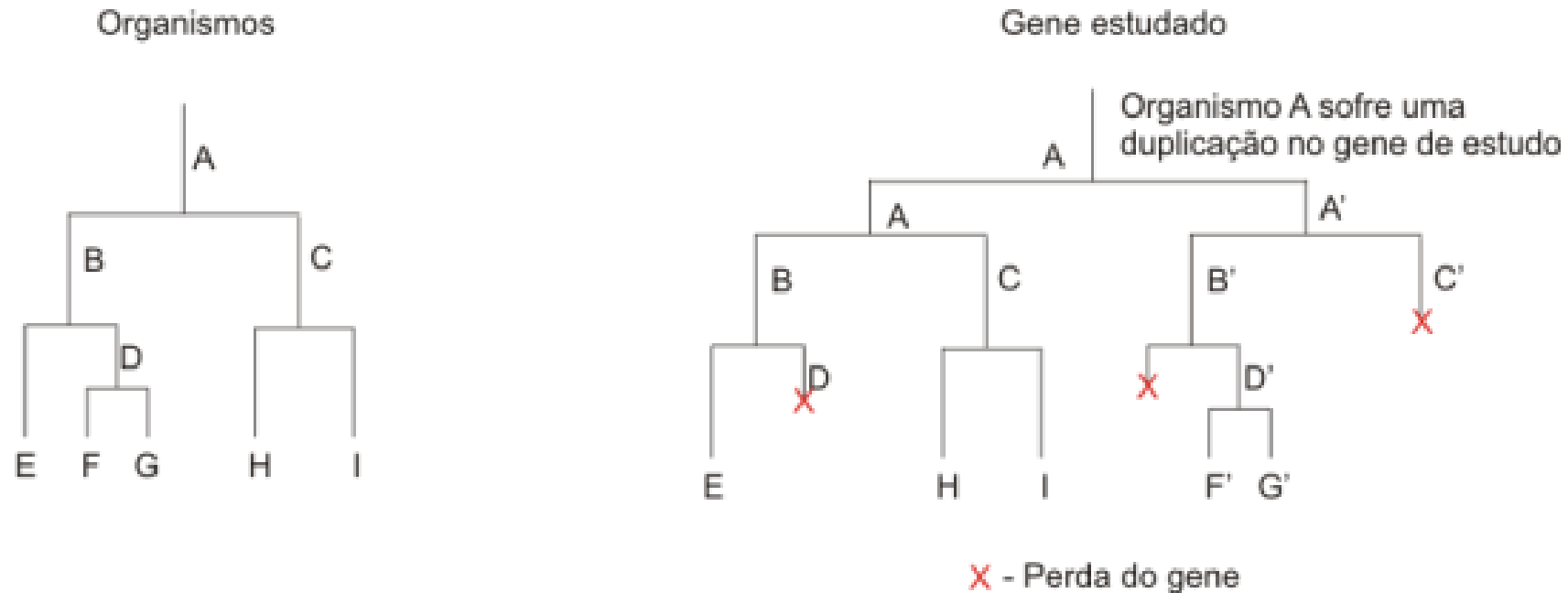


Dendrograma : Muestra relaciones, no da informacion sobre cuando se produce un evento, ni cual seria el antepasado comun

Cladograma: une con lineas especies o filos, dejando evidente el punto de divergencia donde se produce una ramificación a partir de un antepasado comun

Arbol filogenético es un filograma (tipo especifico de cladograma), Da informacion especifica del tiempo donde se produce un evento

ARBOLES FILOGENETICOS



Evolución génica: No siempre es posible asociar la filogenia detectada para un gen con la filogenia de los organismos. Fenómenos de duplicación, supresión y recombinación puede alterar significativamente la filogenia obtenida.

ARBOLES FILOGENETICOS

Fuerzas evolutivas dan forma a los genomas

En el caso de los genomas bacterianos, su estado evolutivo depende de dos procesos: a nivel mutacional y a nivel poblacional (*Ochman y Davalos 2006*).

- **Las mutaciones** pueden ocasionar interrupciones y degradar las regiones funcionales
- **A nivel poblacional** : Selección natural se contrapone a la erosión de los genes conservando en el genoma las regiones funcionales.

La selección natural actúa en la descendencia vertical de los genomas bacterianos, dirigiendo la especialización del organismo.

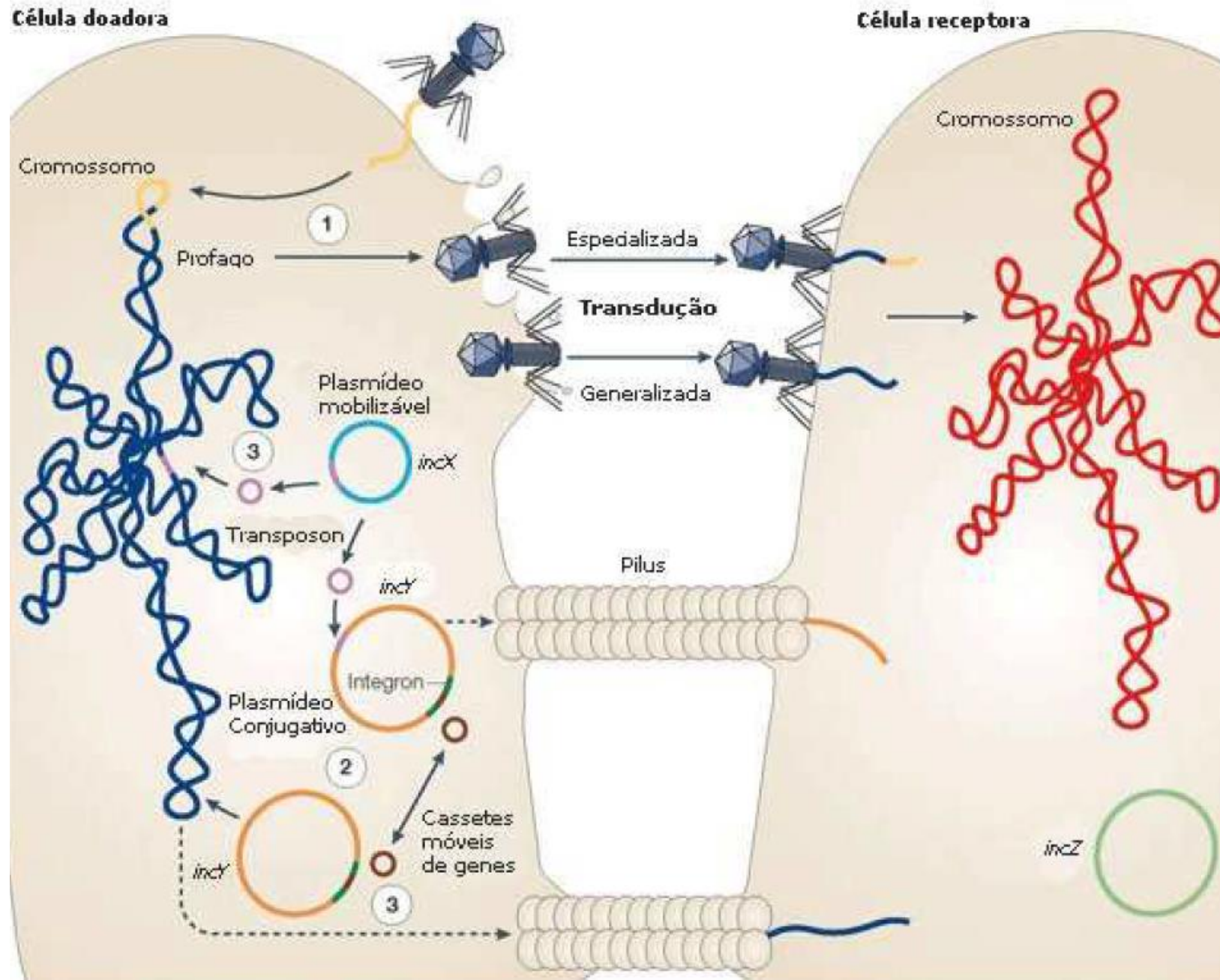
Otras fuerzas que actúan son: la pérdida de genes, duplicaciones y transferencia horizontal de genes

ARBOLES FILOGENETICOS

Fuerzas evolutivas dan forma a los genomas

Transferencia horizontal de genes (HGT Horizontal Gene Transfer)

- Genes externos son adquiridos por un determinado genoma (Koonin et al 2001)
- La transferencia es unidireccional, no sigue los patrones de un linaje parental (descendencia vertical) (Pevsner 2003).
 - **Transformación:** DNA liberado en el ambiente (bacteria donadora) es adquirido por una bacteria receptora. Ejem. *Neisseria gonorrhoeae*.
 - **Transducción:** DNA de una célula donadora es introducido en otra bacteria por un bacteriófago
 - **Conjugación** Existe contacto entre las células donadoras y receptoras por medio de un canal especializado que se forma entre ellas (sistema secretorio tipo IV). DNA puede ser transferido del cromosoma o por un plasmideo móvil,



ARBOLES FILOGENETICOS: HOMOLOGIA

Para reconstruir un árbol de especies debe considerarse que las secuencias elegidas deben ser representativas de las especies incluidas en el análisis.

En otro caso podemos generar un árbol correcto de secuencias, pero no de especies.

DEFINICION DE HOMOLOGIA

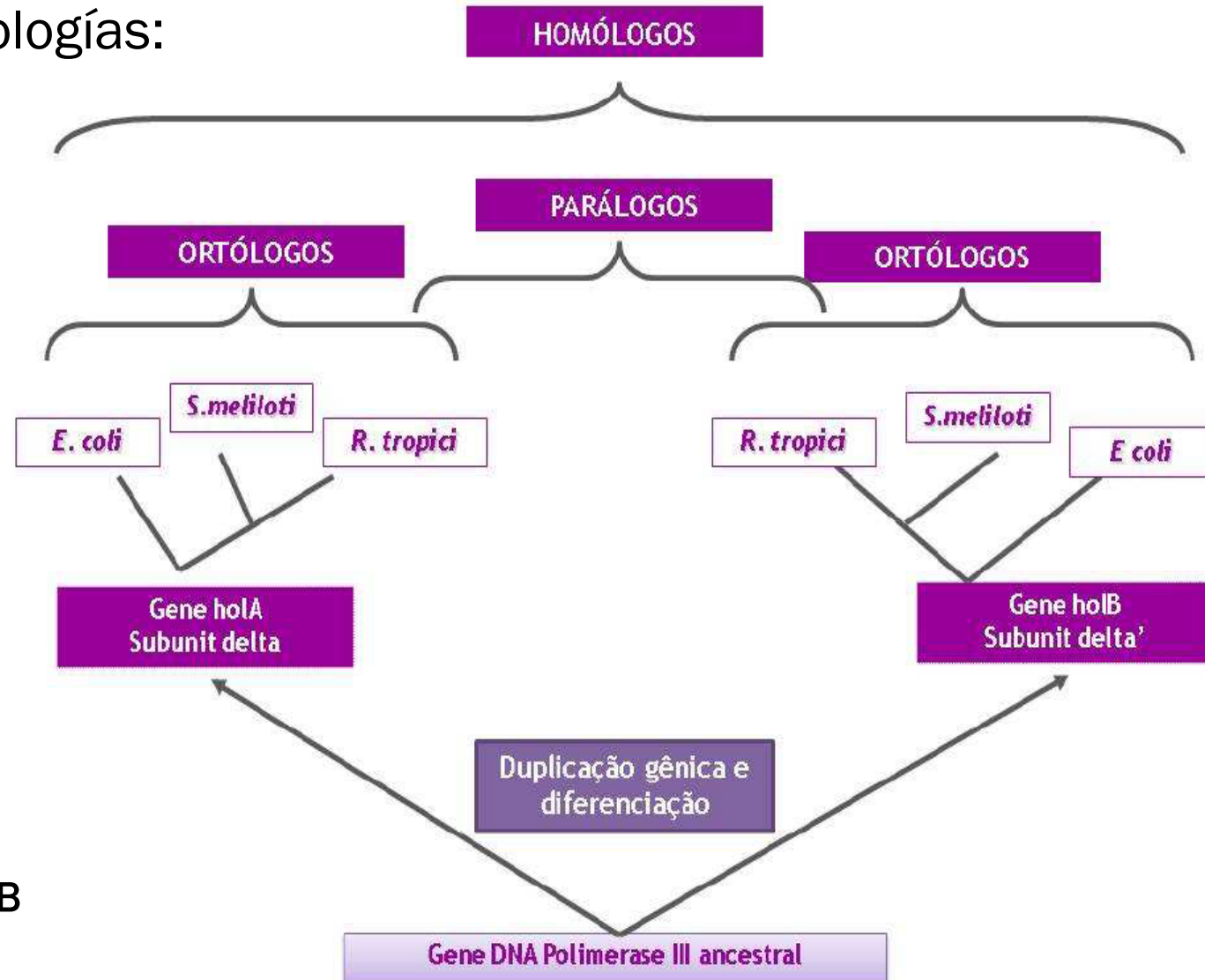
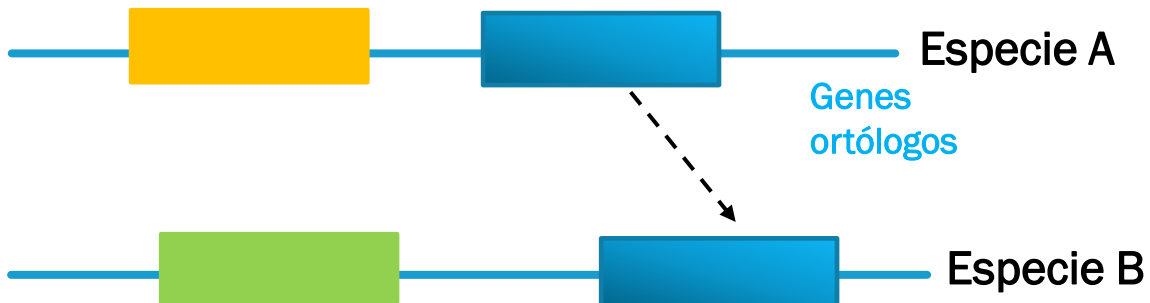
La palabra homologo designa un ancestral común (una línea genética o evolutiva de descendiente), sin ninguna especificación adicional del escenario evolutivo

- Se trata de una hipótesis
- Es una propiedad presente o ausente
- Para evaluar experimentalmente si dos genes son homólogos precisamos de evidencias indirectas. Siendo el criterio de similaridad o semejanza el mas usado:
 - Similaridad en la estructura 3D y sitios activos
 - Similaridad en la secuencia (nucleótidos o aminoácidos)

ARBOLES FILOGENETICOS : HOMOLOGIA

Recordando los distintos tipos de homologías:

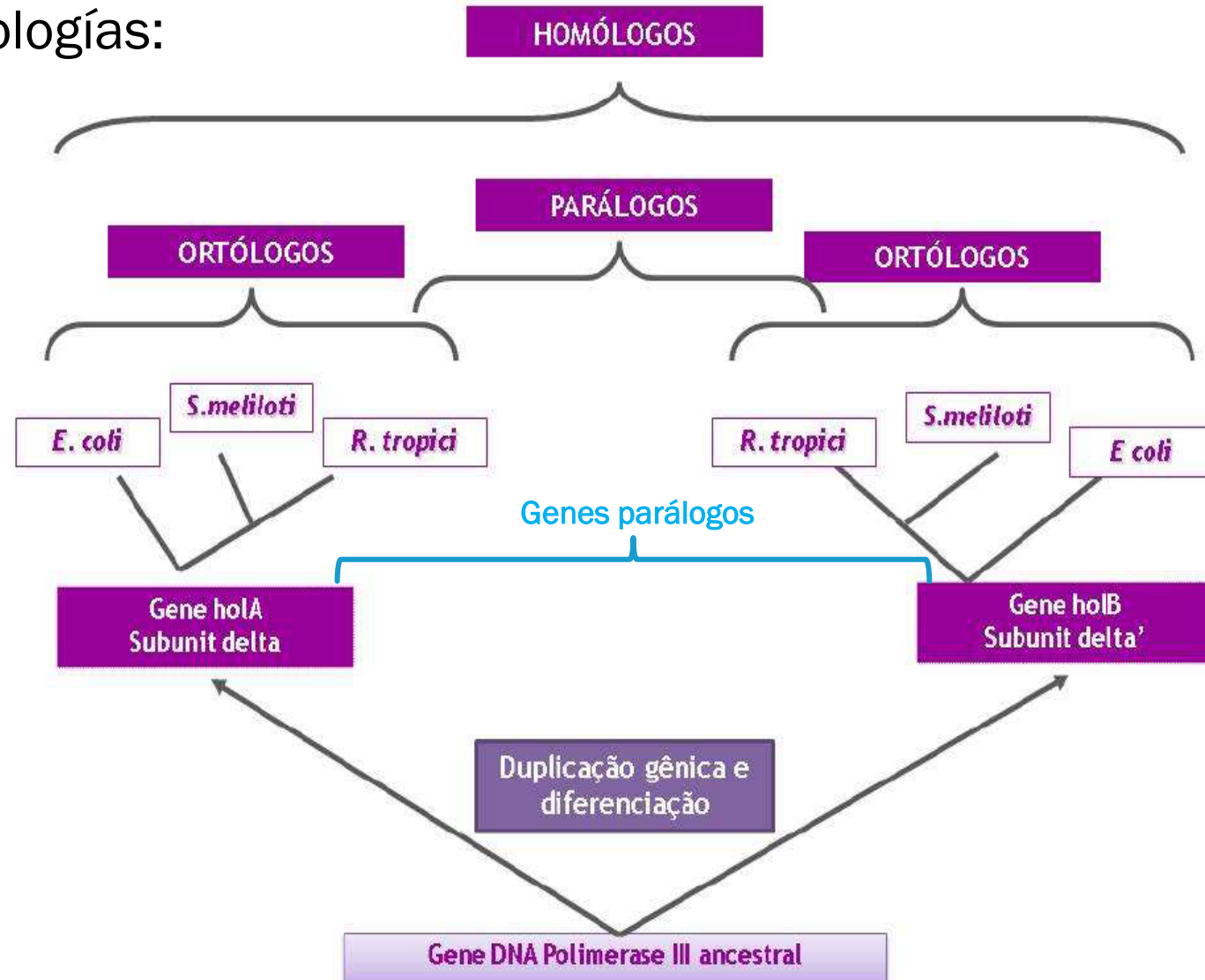
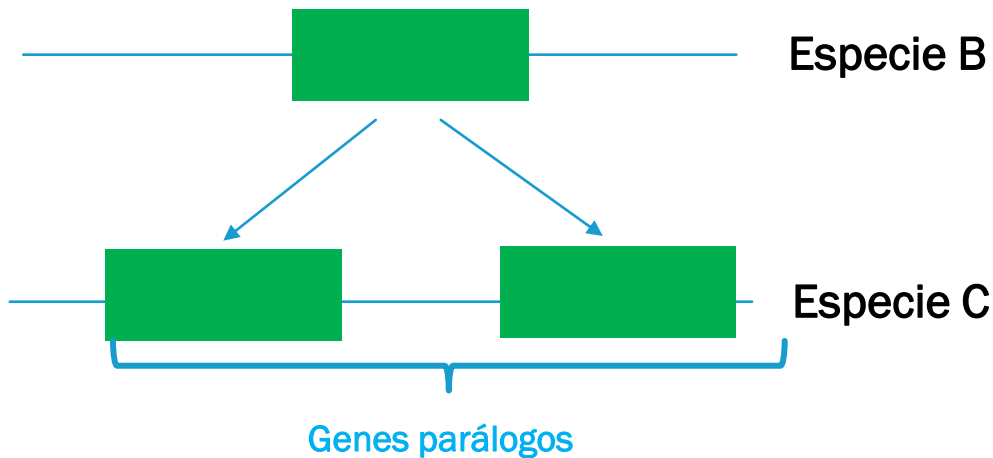
Ortólogo: Genes homólogos que proceden de un único gen ancestral en el último ancestral común de las especies comparadas



ARBOLES FILOGENETICOS: HOMOLOGIA

Recordando los distintos tipos de homologías:

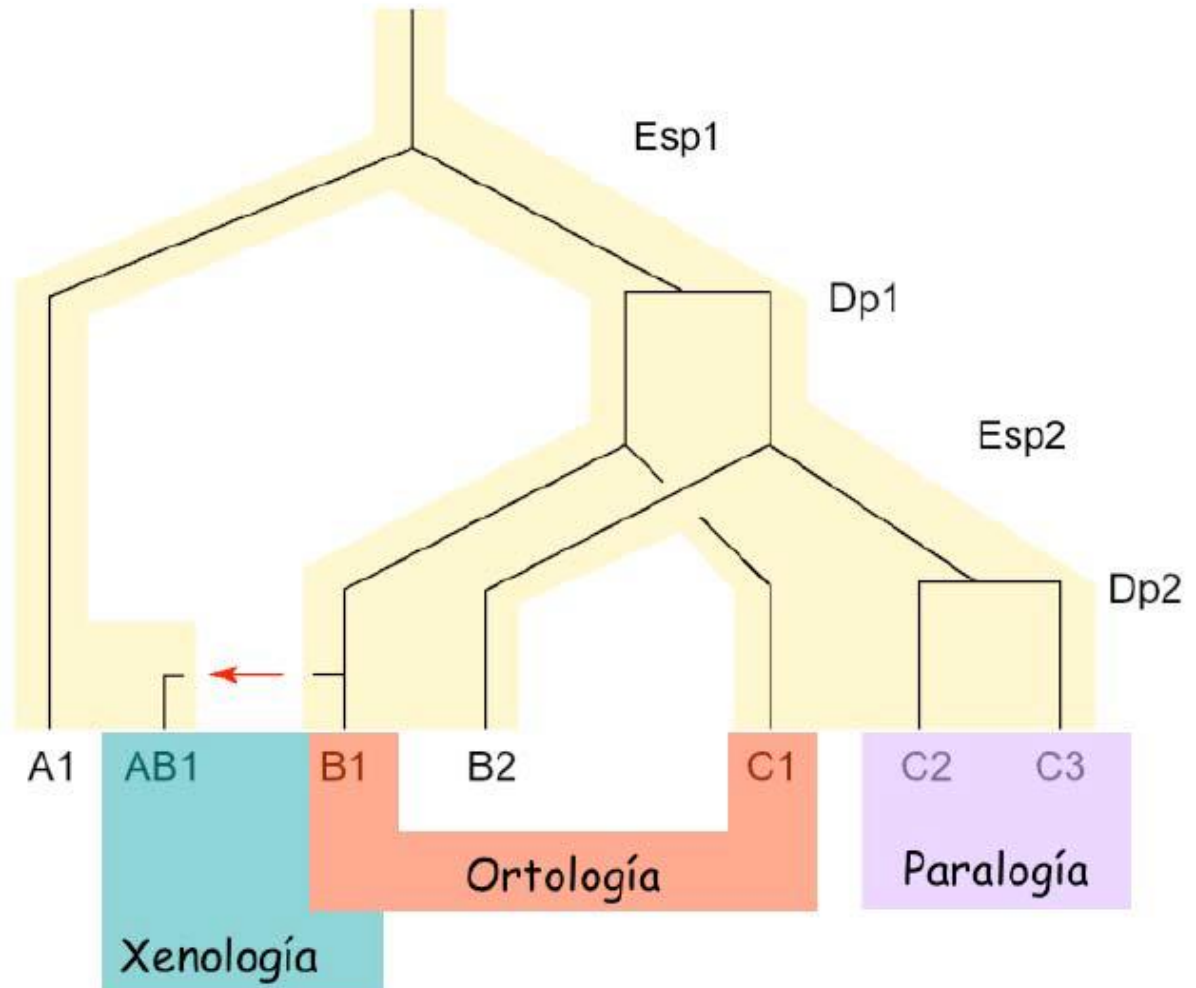
Parálogo: Genes homólogos producto de duplicación en el mismo organismo.



ARBOLES FILOGENETICOS: HOMOLOGIA

Xenólogo:

Se identifica un tercer grupo de genes, donde un gene de una fuente distante desplaza a un ortólogo residente, este es un mecanismo de transferencia horizontal.



METODOS PARA RECONSTRUCCION DE ARBOLES FILOGENETICOS

- **Máxima Parsimonia** : Busca el árbol que implica el menor número de eventos de mutación.
- **Máxima probabilidad**: similar al método de máxima parsimonia, pero usa modelos evolutivos que calculan diferentes posibilidades de mutación dependiendo de la base que mutó
- **Métodos de distancia** utilizan comparaciones par a par entre las distancias (número de mutaciones) de las diferentes secuencias estudiadas para construir un árbol que reproduzca esta distancia del modo mas fiel posible.

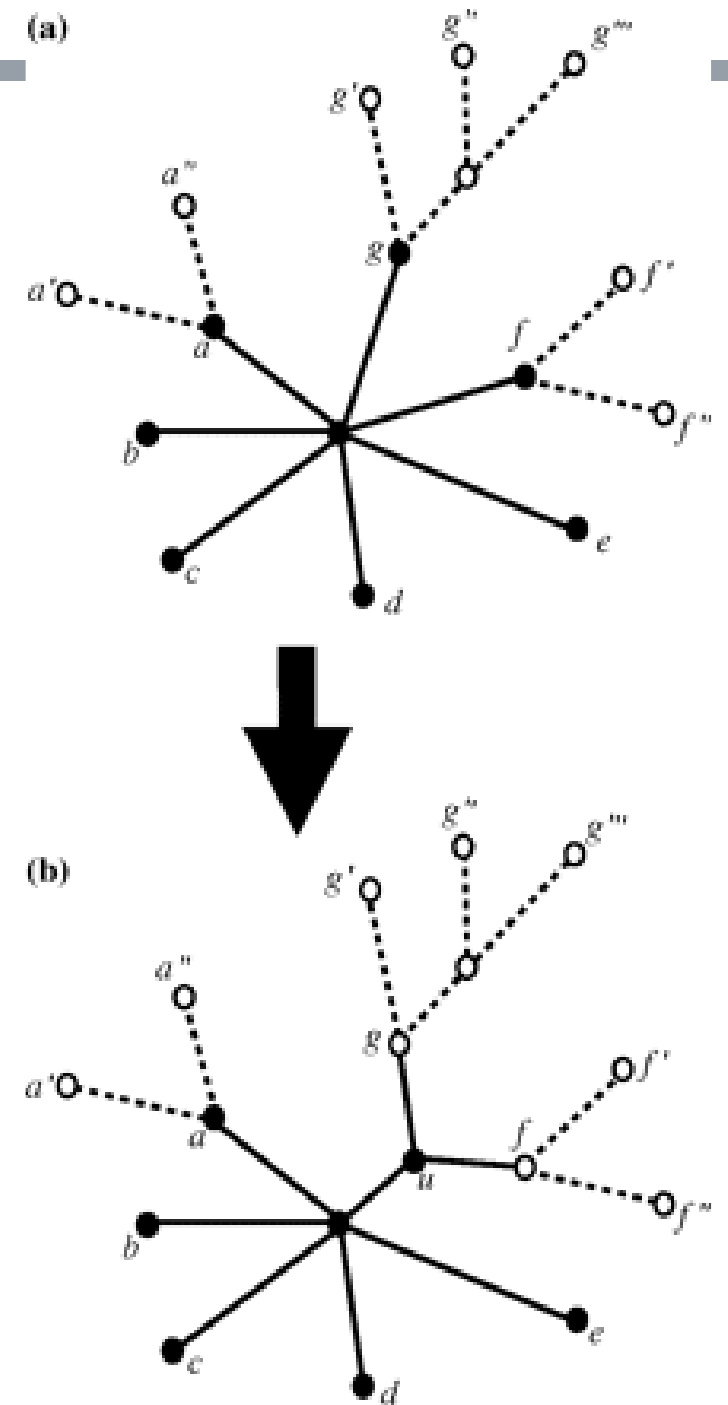
METODOS PARA RECONSTRUCCION DE ARBOLES FILOGENETICOS

- Inicialmente vimos que para 5 especies podíamos tener diferentes tipos de estructura que permitan establecer diferentes raíces.
- La figura muestra el número de árboles posibles según el número de especies (o proteínas) diferentes analizadas

Species	Number of trees
1	1
2	1
3	3
4	15
5	105
6	945
7	10,395
8	135,135
9	2,027,025
10	34,459,425
11	654,729,075
12	13,749,310,575
13	316,234,143,225
14	7,905,853,580,625
15	213,458,046,676,875
16	6,190,283,353,629,375
17	191,898,783,962,510,625
18	6,332,659,870,762,850,625
19	221,643,095,476,699,771,875
20	8,200,794,532,637,891,559,375
30	4.9518×10^{38}
40	1.00985×10^{57}
50	2.75292×10^{76}

METODOS DE NEIGHBOR JOINING

- Se calcula la distancia entre los diversos puntos (genes o proteínas) de un árbol filogenético y se establecen vecinos, que serian un par de secuencias que tienen un parámetro Q mínimo
- Este par de vecinos origina un nodo interno que representará este par de secuencias
- Luego el parámetro Q será recalculado considerando el nodo interno



METODOS DE NEIGHBOR JOINING : CALCULO DEL PARAMETRO Q

	A	B	C	D	E
A	—	22	39	39	41
B	—	—	41	41	43
C	—	—	—	18	20
D	—	—	—	—	10
E	—	—	—	—	—

1. Se calcula un factor u , cuyo valor es igual a la suma de las distancias del nodo de interés (terminal) en relación a los otros nodos o puntos dividido por el numero de puntos (número de puntos-2)

- $u_a = (22+39+39+41)/(5-2)=47$
- $u_b = (22+41+41+43)/(5-2)=49$
- $u_c = (39+41+18+20)/(5-2)=39.3$
- $u_d = (39+41+18+10)/(5-2)=36$
- $u_e = (41+43+20+10)/(5-2)=38$

METODOS DE NEIGHBOR JOINING : CALCULO DEL PARAMETRO Q

	A	B	C	D	E
A	—	22	39	39	41
B	—	—	41	41	43
C	—	—	—	18	20
D	—	—	—	—	10
E	—	—	—	—	—

Para cada par de nodos
terminales i, j se calcula Q
definido por: $D_{i,j} - u_i - u_j$
Luego es seleccionado el
menor valor

$$AB \rightarrow 22 - 47 - 49 = -74$$

$$AC \rightarrow 39 - 47 - 39.3 = -47.3$$

$$AD \rightarrow 39 - 47 - 36 = -44$$

$$AE \rightarrow 41 - 47 - 38 = -44$$

$$BC \rightarrow 41 - 49 - 39.3 = -47.3$$

$$BD \rightarrow 41 - 49 - 36 = -44$$

$$BE \rightarrow 43 - 49 - 38 = -44$$

$$CD \rightarrow 18 - 39.3 - 36 = -57.3$$

$$CE \rightarrow 20 - 39.3 - 38 = -57.3$$

$$DE \rightarrow 10 - 36 - 38 = -64$$

METODOS DE NEIGHBOR JOINING : CALCULO DEL PARAMETRO Q

	A	B	C	D	E
A	—	22	39	39	41
B	—	—	41	41	43
C	—	—	—	18	20
D	—	—	—	—	10
E	—	—	—	—	—

Para cada par de nodos
terminales i, j se calcula Q
definido por: $D_{i,j} - u_i - u_j$
Luego es seleccionado el
menor valor

$$AB \rightarrow 22 - 47 - 49 = -74$$

$$AC \rightarrow 39 - 47 - 39.3 = -47.3$$

$$AD \rightarrow 39 - 47 - 36 = -44$$

$$AE \rightarrow 41 - 47 - 38 = -44$$

$$BC \rightarrow 41 - 49 - 39.3 = -47.3$$

$$BD \rightarrow 41 - 49 - 36 = -44$$

$$BE \rightarrow 43 - 49 - 38 = -44$$

$$CD \rightarrow 18 - 39.3 - 36 = -57.3$$

$$CE \rightarrow 20 - 39.3 - 38 = -57.3$$

$$DE \rightarrow 10 - 36 - 38 = -64$$

METODOS DE NEIGHBOR JOINING : CALCULO DEL NUEVO NODO

La distancia de los puntos A y B al nuevo punto o nodo AB se calculada como:

$$v_a = \frac{1}{2} D_{AB} + \frac{1}{2} (u_a - u_b)$$

$$v_a = 11 + \frac{1}{2} (47 - 49) = 10$$

$$v_b = \frac{1}{2} D_{AB} + \frac{1}{2} (u_b - u_a)$$

$$v_b = 11 + \frac{1}{2} (49 - 47) = 12$$

- Se reemplazan los nodos terminales A y B por el nodo (AB) que se considera como un nuevo punto
- Las distancias de este nodo a los otros nodos terminales (k) se calculan con la siguiente fórmula: $D_{(AB),k} = (D_{ak} + D_{bk} - D_{ab})/2$

A	(AB)	C	D	E
(AB)	—	29	29	31
C	—	—	18	20
D	—	—	—	10
E	—	—	—	—

METODOS DE NEIGHBOR JOINING : CALCULO DEL *BOOTSTRAP*

El cálculo de *bootstrap* para árboles filogenéticos permite tener un parámetro que refleja la robustez del análisis filogenético producido.

El *Bootstrap* se genera mediante la creación de múltiples conjuntos de secuencias en las que las columnas de alineación múltiple se eligen al azar

Se generan nuevos análisis para cada uno de los nuevos conjuntos generados.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
A T A G C C A T A G C A A C C T
A T A C C C A T G A C A A C G A
A T A C C C A T A G C A A C C A
A T A G C C A T A G C A A C G A
A T C C C C A T A G C A A C C T

The real multiple alignment

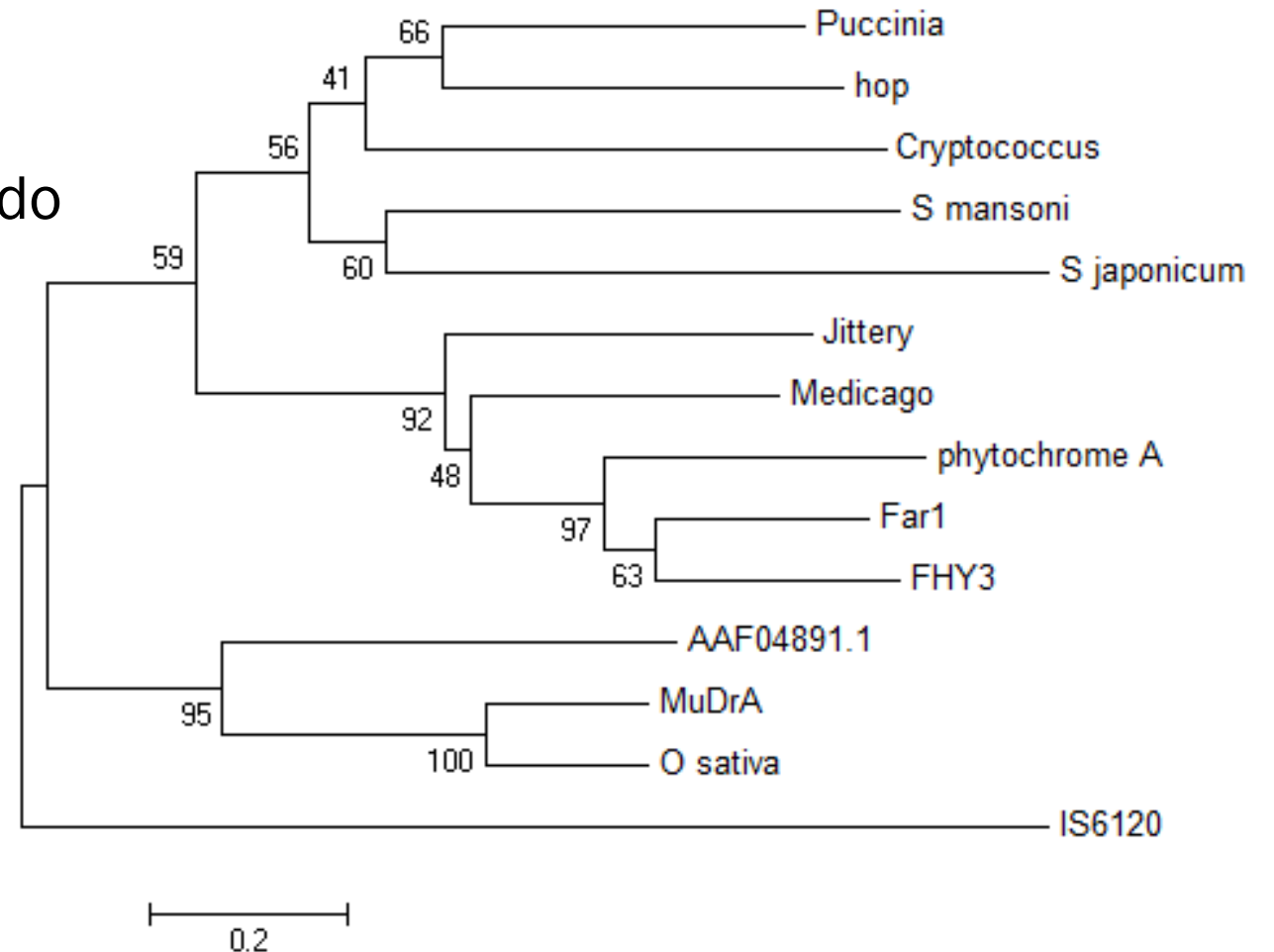
↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
2 7 4 9 11 4 16 5
T A G A C G T C
T A C G C C A C
T A C A C C A C
T A G A C G A C
T A C A C C T C

New alignment

METODOS DE NEIGHBOR JOINING : CALCULO DEL *BOOTSTRAP*

Los valores en los nodos internos representan los valores del Bootstrap, es decir el número de apariciones de este nodo interno en 100 réplicas

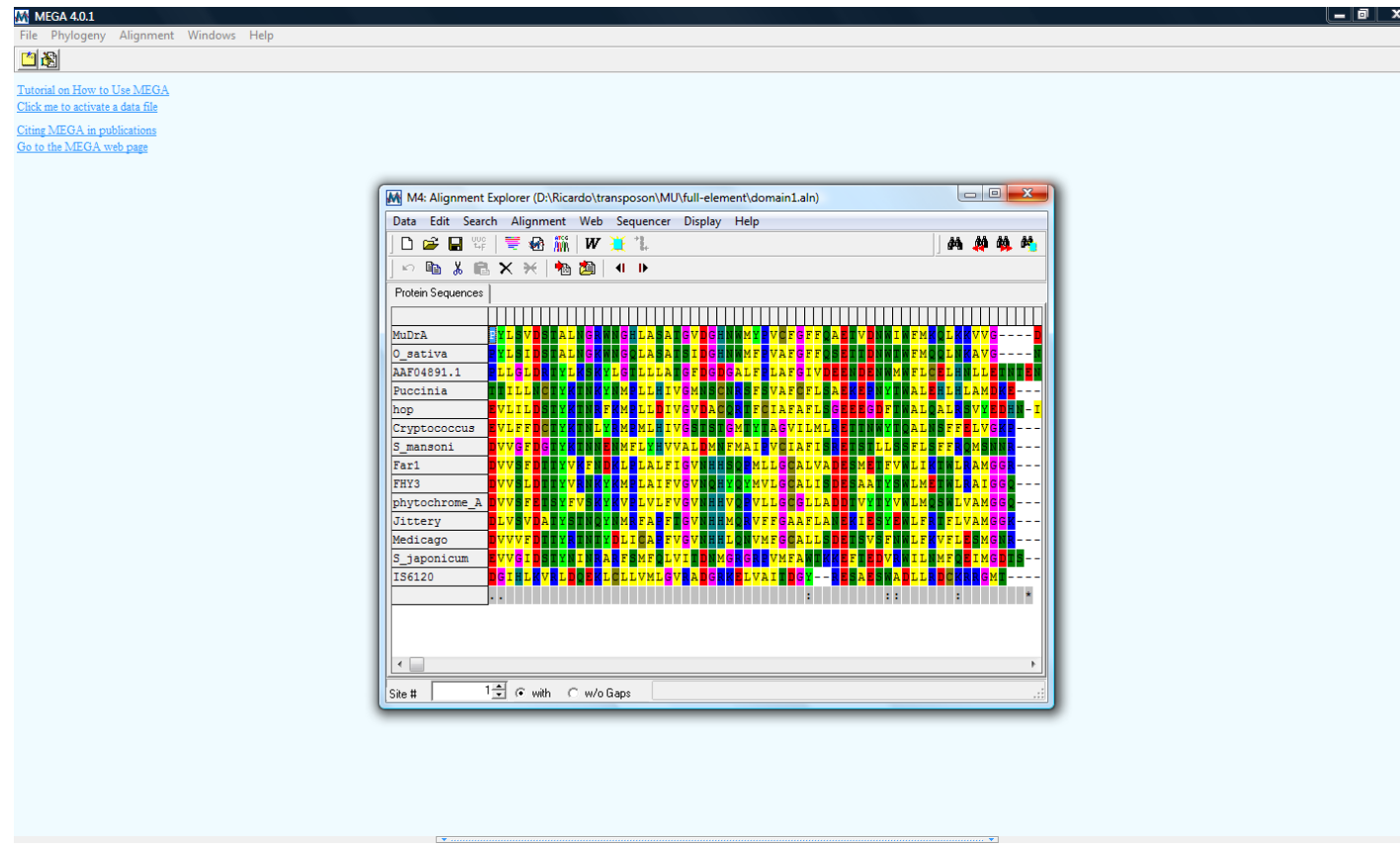
El valor bootstrap indica que en 95/100 muestreos las secuencias AAF04891.1, MuDrA y O sativa fueron agrupadas en una rama que contiene solo las tres secuencias.



MEGA

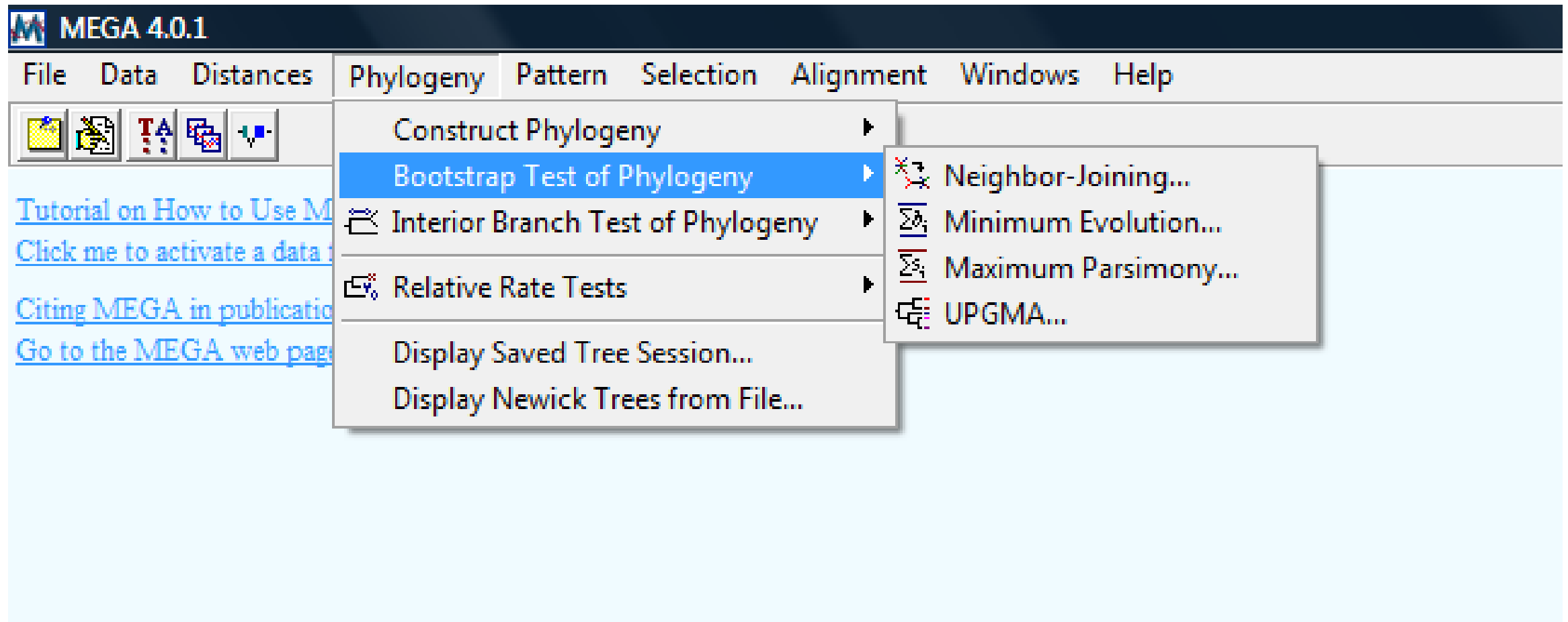
El programa MEGA realiza una serie de análisis evolutivos basados en múltiples alineamientos múltiples de secuencias.

Es posible importar secuencias de aln resultantes del alineamiento obtenido con el clustal



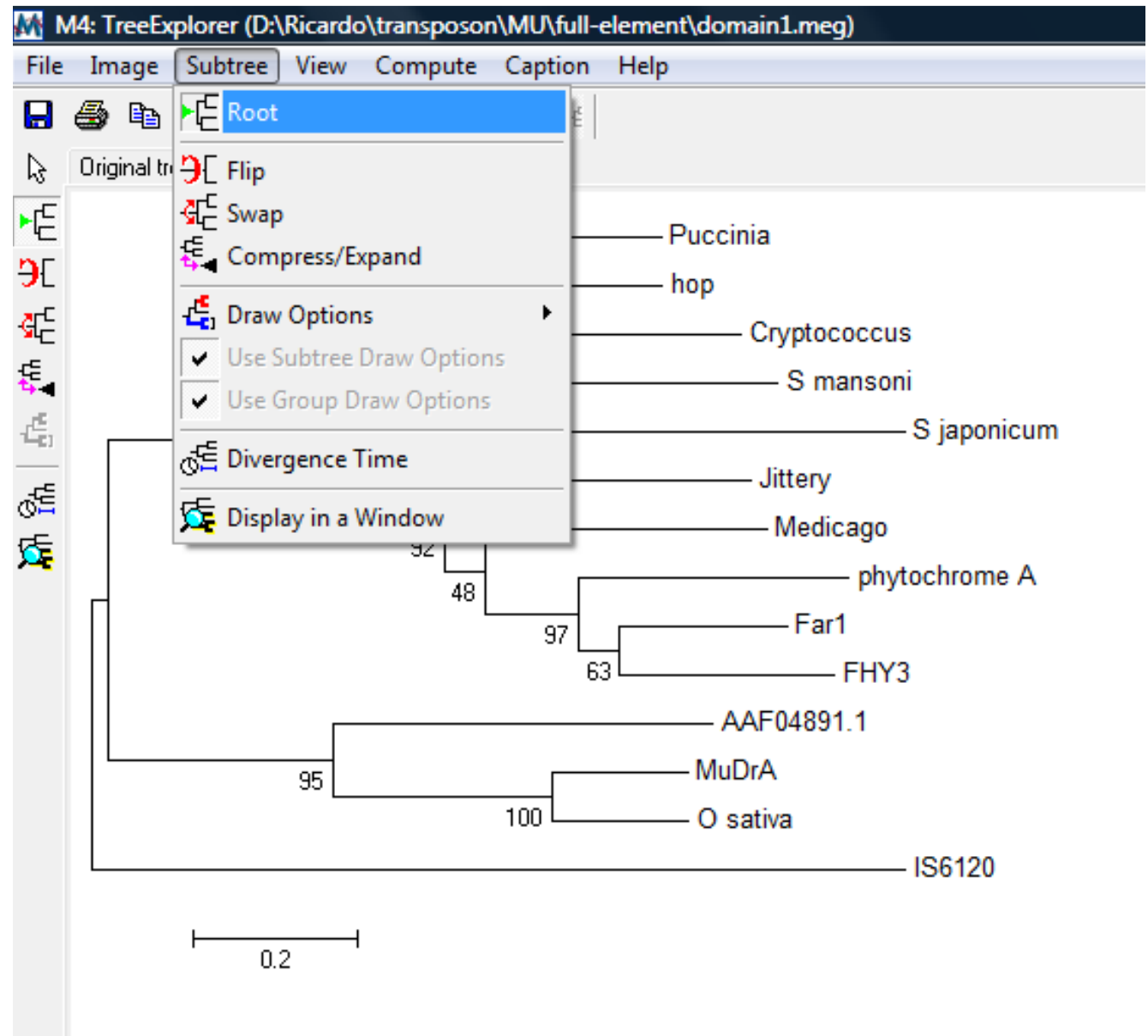
MEGA

El programa MEGA permite realizar el análisis filogenetico de alineamientos por diferentes metodos



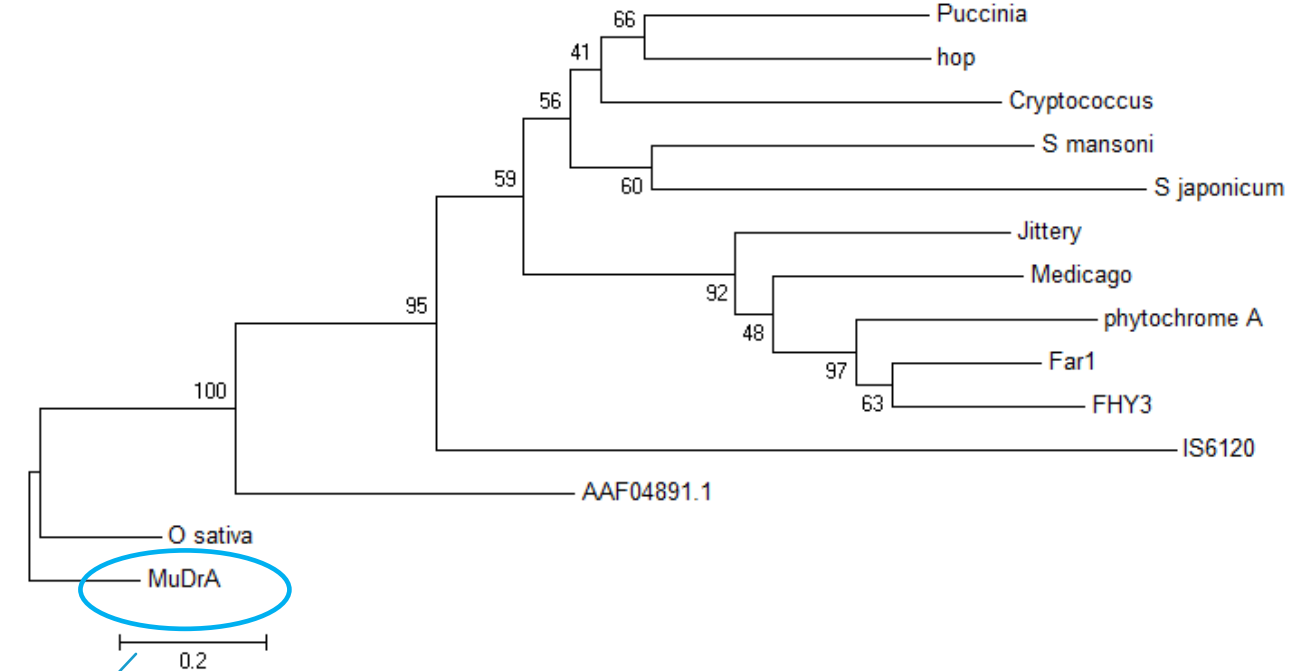
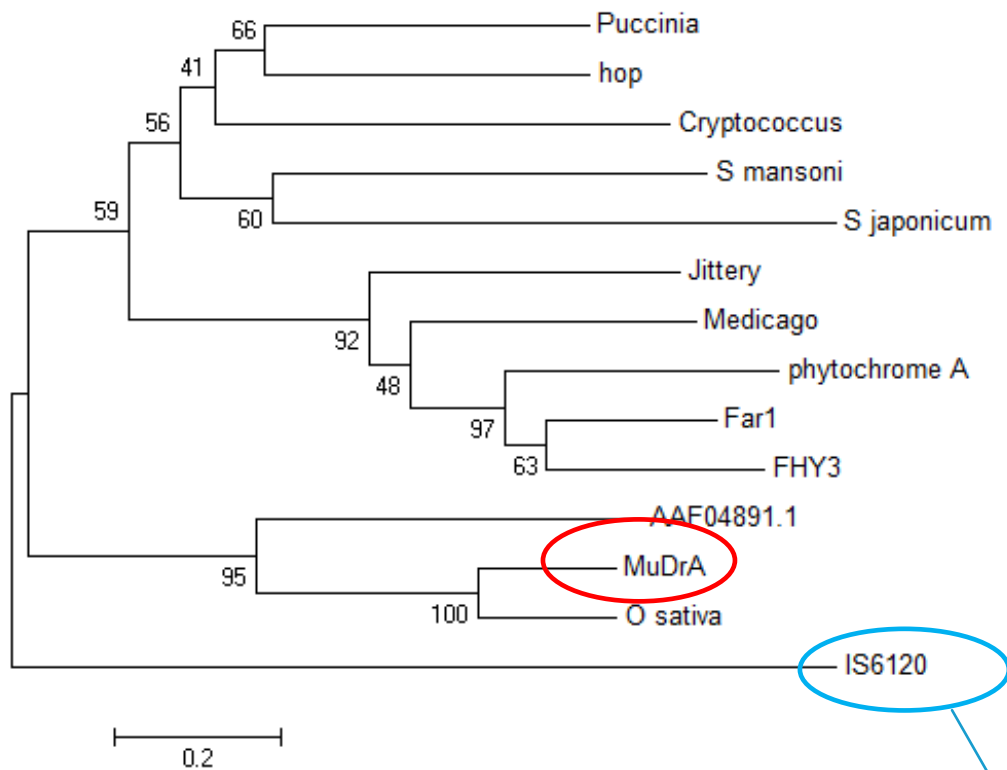
MEGA

Permite escoger la rama a partir de la cual se desea enraizar el árbol usando la opción “root”



MEGA

Selección de la rama marcada en rojo



outgroup