

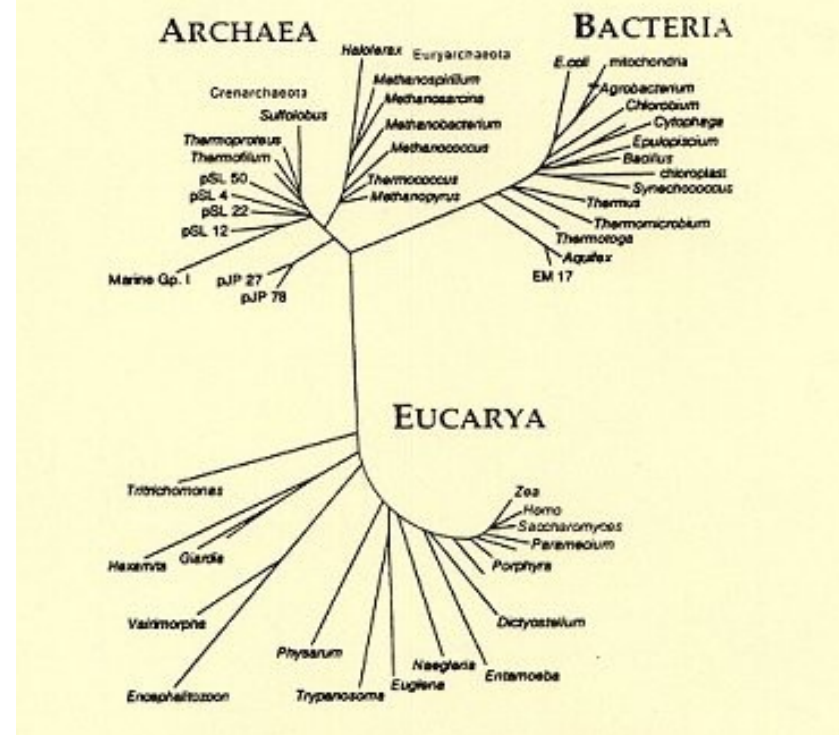
Curso de Evolución 2020

Facultad de Ciencias

Montevideo, Uruguay

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

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



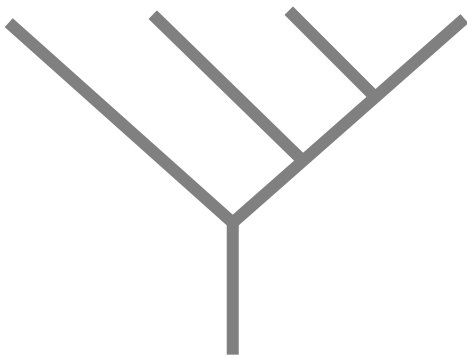
Tema 2. Las filogenias como contexto de análisis de la evolución. Aplicaciones del análisis filogenético. Análisis filogenético según el principio de parsimonia. Métodos basados en distancias y en modelos de evolución molecular.

## Método de Hennig

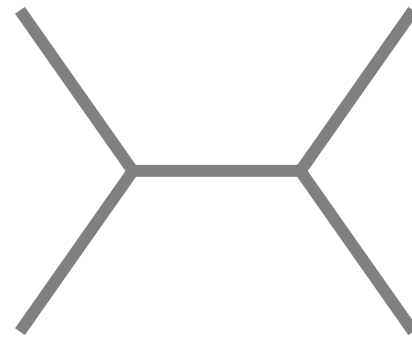
- 1) Definir la **raíz** (escogiendo grupo externo)
- 2) Establecer la **polaridad** de los caracteres
- 3) Usar las **sinapomorfías** para encontrar el árbol (o para comparar árboles)
- 4) Elegir el árbol más corto  
= más parsimonioso  
= con menor número de pasos

## Método “alternativo”

- 1) Identificar caracteres informativos
- 2) Construir árboles **sin raíz**
- 3) Elegir el árbol más corto
- 4) Eventualmente colocar la raíz, y con ello establecer la polaridad

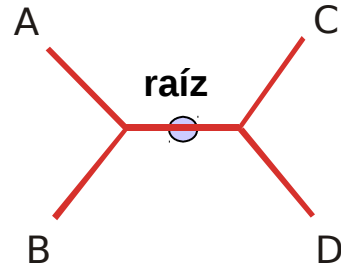


**árbol con raíz**

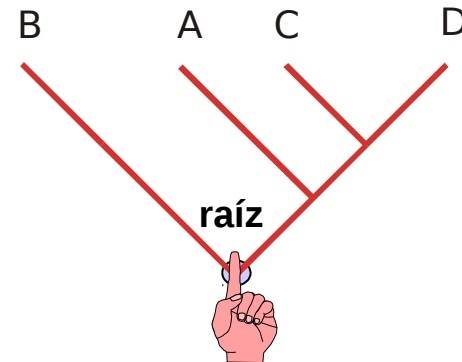
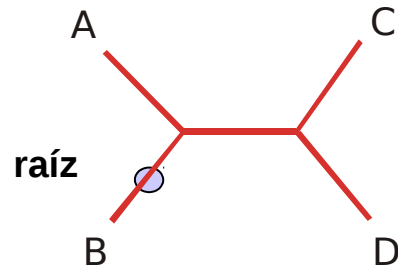
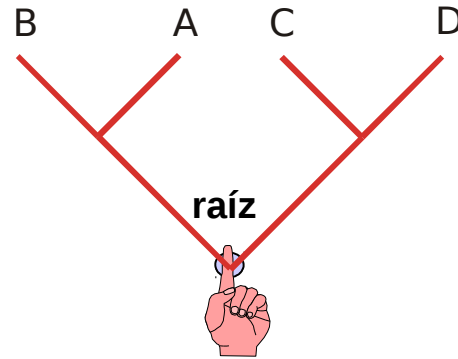


**árbol sin raíz**

### Árbol sin raíz



### Árbol con raíz

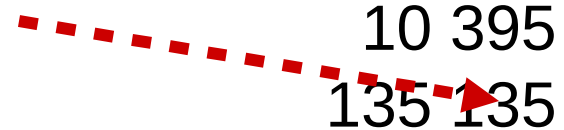


Se pueden obtener varios árboles con raíz a partir de uno sin raíz

# Número de árboles posibles con y sin raíz

Nº de OTUs	Nº de árboles con raíz	Nº de árboles sin raíz
------------	------------------------	------------------------

2	1	1
3	3	1
4	15	3
5	105	15
6	954	105
7	10 395	954
8	135 135	10 395
9	2 027 025	135 135
10	<b>34 459 425</b>	2 027 025



## Caracteres informativos:

Concepto:

Un carácter informativo es aquel para el cual hay al menos dos conjuntos de árboles que implican distinto número de pasos (en términos de parsimonia).

Criterio operativo:

Un carácter es informativo cuando presenta al menos dos estados para los que se cumple que cada uno es compartido por al menos dos de los taxa en estudio.

# Caracteres informativos:

Concepto:

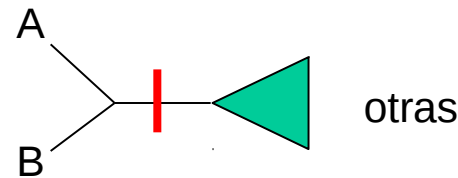
Un carácter informativo es aquel para el cual hay al menos dos conjuntos de árboles que implican distinto número de pasos (en términos de parsimonia).

Criterio operativo:

Un carácter es informativo cuando presenta al menos dos estados para los que se cumple que cada uno es compartido por al menos dos de los taxa en estudio.

Ejercicio: identificar caracteres informativos, e ilustrar cómo existen árboles que implican distinto número de pasos para dichos caracteres.

Especie A	0	0	0	0	0
Especie B	0	0	1	0	0
Especie C	1	0	1	0	1
Especie D	1	0	1	1	1
Especie E	1	0	1	2	2
Especie F	1	1	2	3	2



# Análisis de un caso

 Sitio informativo

		sitios						
Taxa	1	2	3	4	5	6	7	
1	A	A	G	A	G	T	G	
2	T	G	C	C	G	T	G	
3	A	G	A	T	A	T	C	
4	T	G	A	G	A	T	C	

Número mínimo de pasos\*  
 1    1    2    3    1    0    1

\* Si existe al menos un árbol óptimo para un carácter, el número mínimo de pasos requerido para ajustar el carácter al árbol es el número de estados del carácter menos 1.

Por lo tanto, el número mínimo de pasos (= la longitud del árbol) de el/los árboles más parsimoniosos es:

$$L = \sum_{i=1}^k E_i - 1$$

Donde L es la longitud del árbol, k es el número de caracteres, y E<sub>i</sub> es el número de estados del carácter i



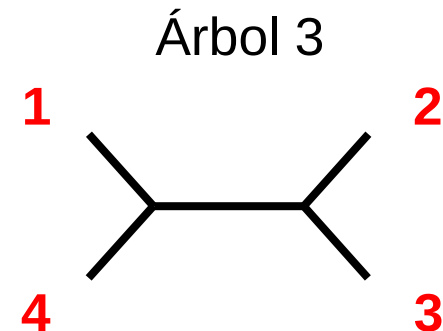
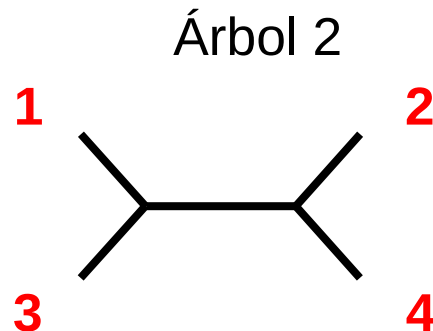
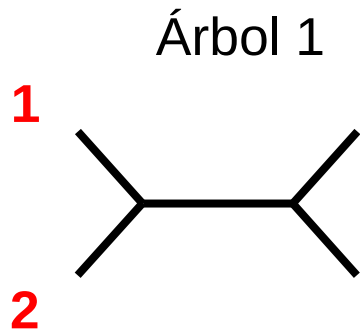
# Análisis de un caso

 Sitio informativo

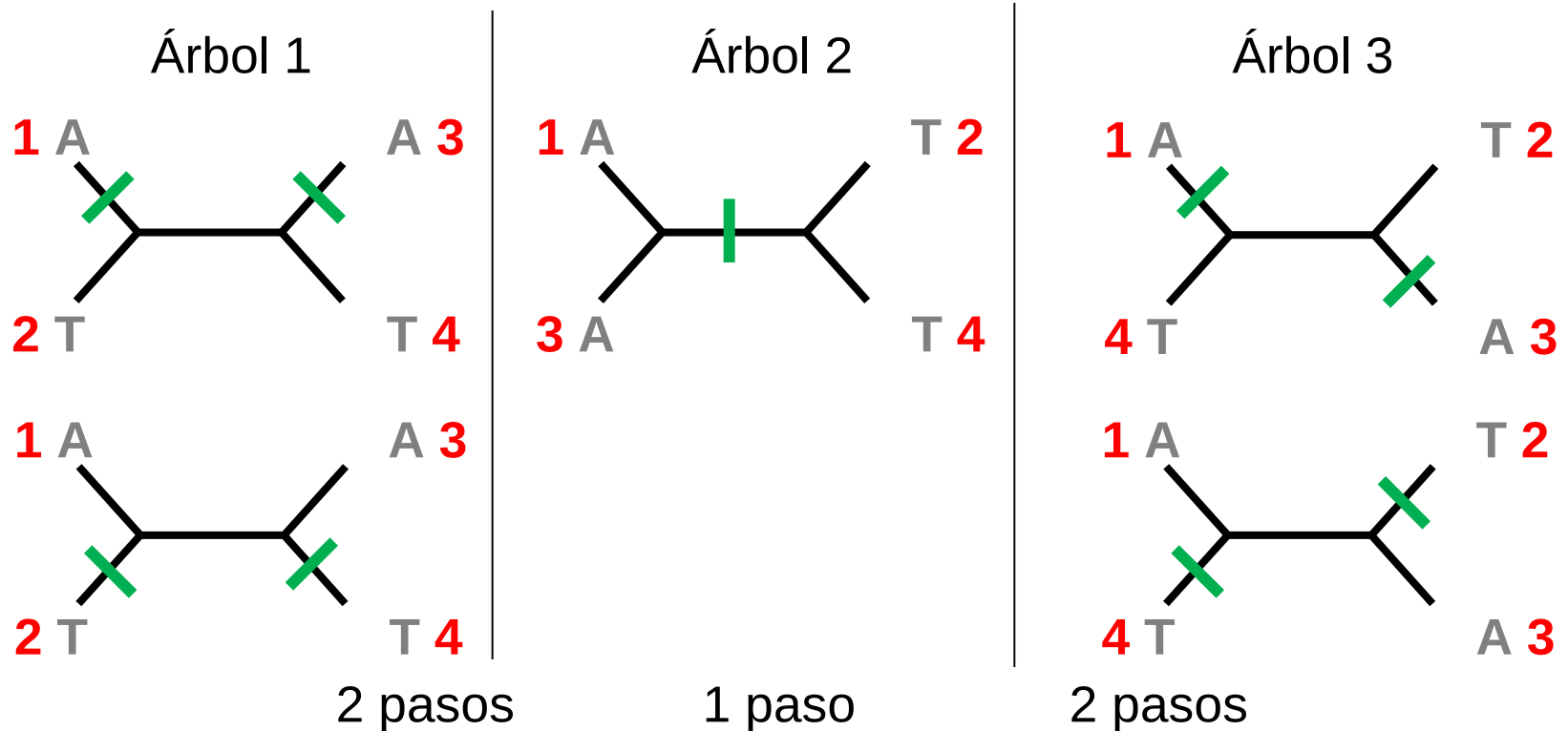
	sitios						
Taxa	1	2	3	4	5	6	7
1	A	A	G	A	G	T	G
2	T	G	C	C	G	T	G
3	A	G	A	T	A	T	C
4	T	G	A	G	A	T	C

## Análisis de un caso: 3 árboles sin raíz posibles

Mapear cada carácter en cada árbol de el/los modos más parsimoniosos (en algunos casos hay dos formas igualmente parsimoniosas)



## Ejemplo resuelto: sitio 1



- Notas:**
1. el carácter requiere 1 paso en el árbol 2, y 2 pasos en los demás;
  2. en los árboles 1 y 3, hay dos formas igualmente parsimoniosas de ajustar el carácter, pero las dos requieren 2 pasos
  3. en los árboles sin raíz, no se puede definir la polaridad (dirección de los cambios), pero sí el número y ubicación (o ubicaciones posibles) de los cambios en el árbol.

	Arbol 1	Arbol 2	Arbol 3
<b>Sitio</b>			
<b>1</b>	2	1	2
<b>5</b>	1	2	2
<b>7</b>	1	2	2
Número de pasos (caracteres informativos) ↑	4	5	6

El árbol 1 es el elegido como el más parsimonioso:

**requiere el mínimo número de cambios**

sumando para todos los sitios informativos (los sitios no informativos agregan el mismo número de pasos a todos los árboles).<sup>11</sup>

## Método de Hennig

- 1) Determinar la **raíz**  
(escogiendo grupo externo)
- 2) **Polaridad** de los caracteres
- 3) Usar las **sinapomorfías**  
para encontrar el árbol (o  
para comparar árboles)
- 4) Elegir el árbol más corto  
= más parsimonioso  
= con menor número de  
pasos

## Método “alternativo”

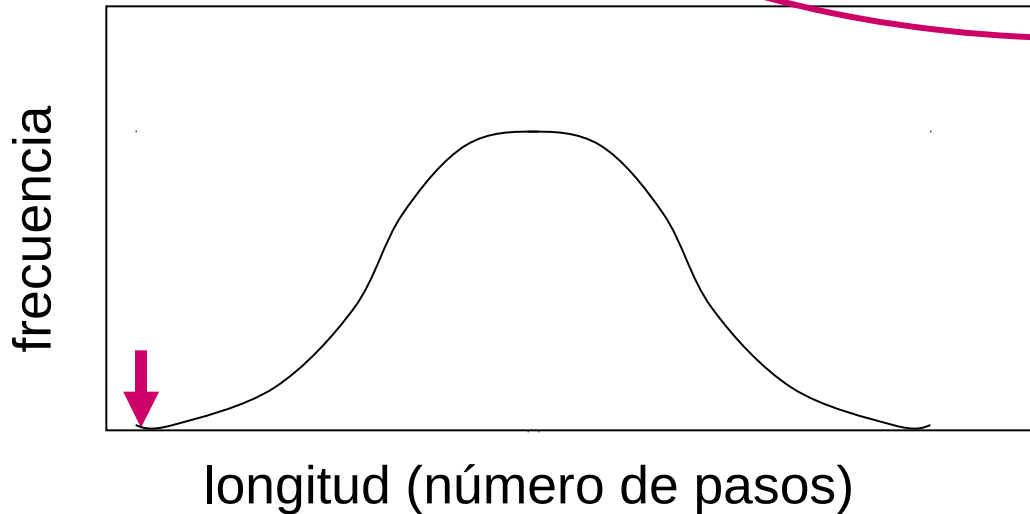
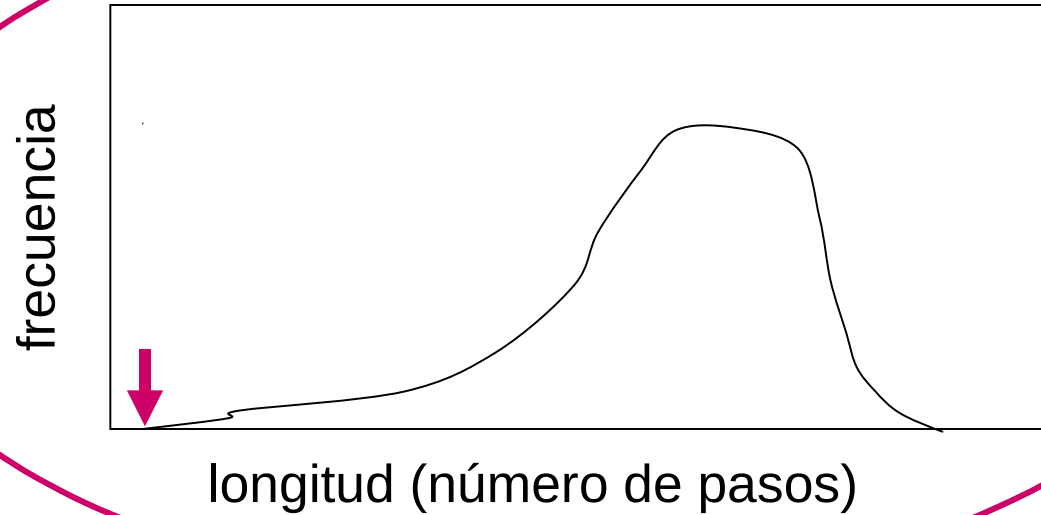
- 1) Determinar sitios
- 2) Construir árboles **sin raíz**
- 3) Elegir el **árbol** (o árboles)  
**más corto**
- 4) Eventualmente colocar la  
raíz, y con ello establecer  
la polaridad

# Evaluando los árboles

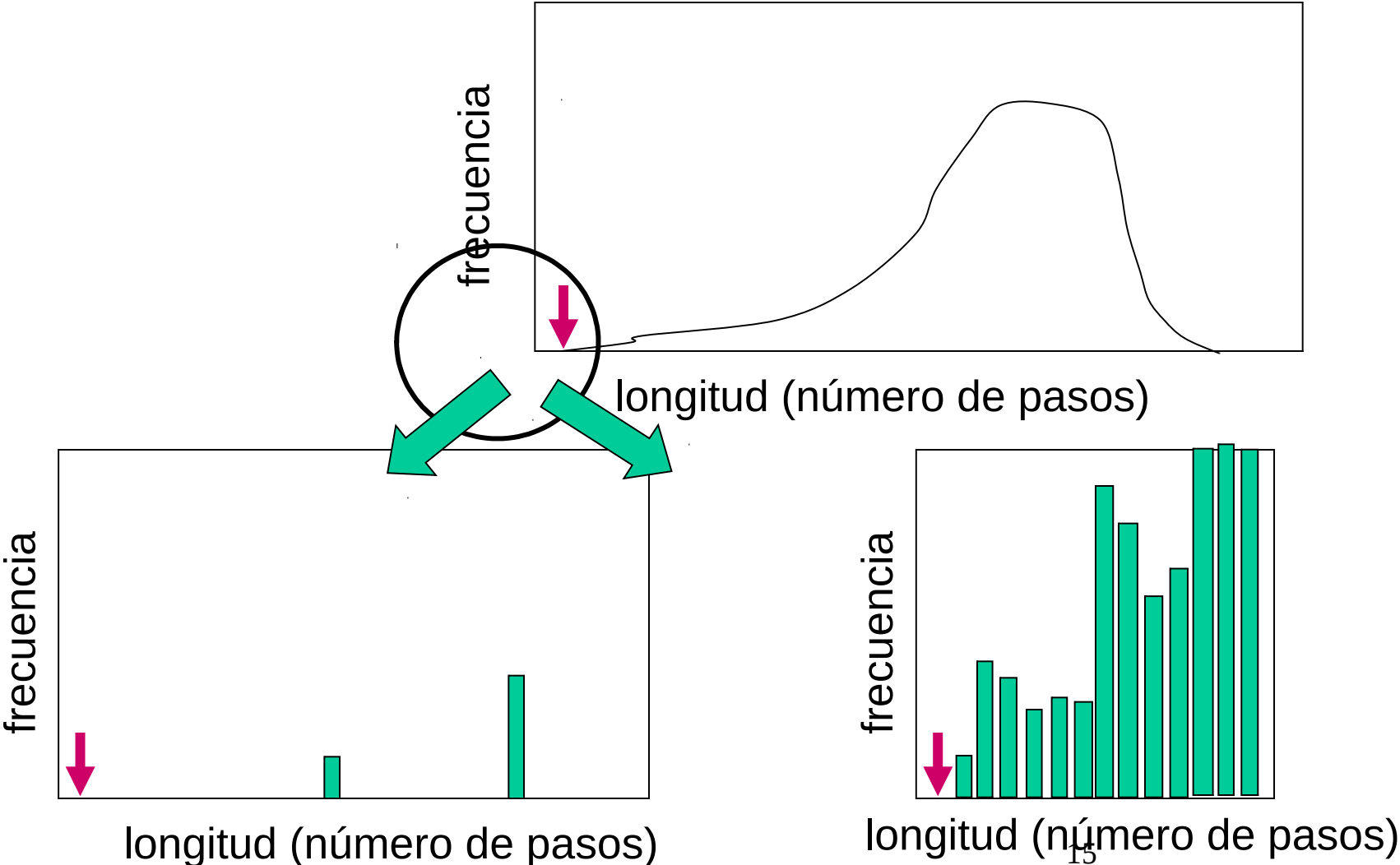
Criterios “globales”

Criterios “por nodo”

# Distribución de frecuencias de todos los árboles posibles, o de una muestra de árboles al azar



# Distribución de frecuencias de todos los árboles posibles, o de una muestra de árboles al azar



# BOOTSTRAP

## 1 Matriz de caracteres



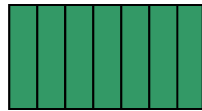
x N

## N Pseudorréplicas

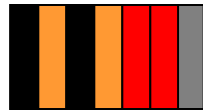
(muestreo al azar con reposición de la matriz de caracteres)



1



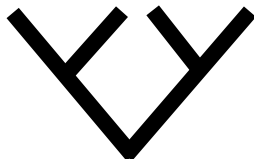
2



3



N



ÁRBOL DE CONSENSO  
por mayoría



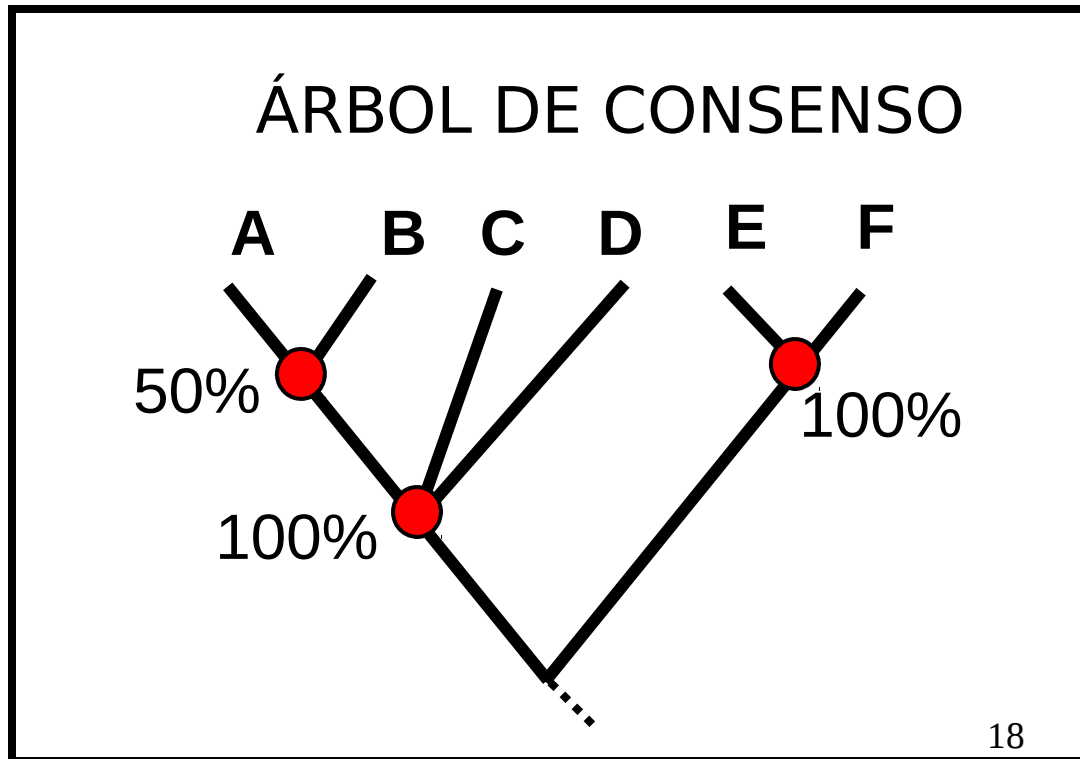
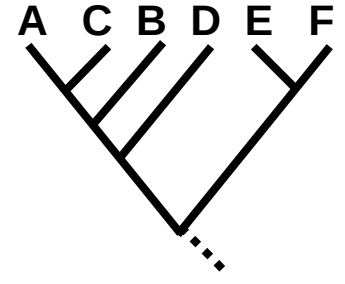
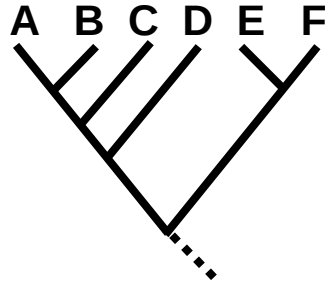
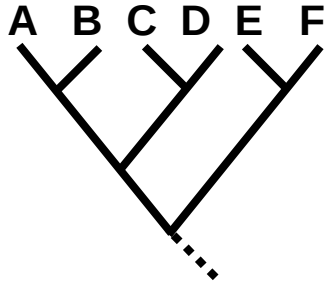
- El árbol de consenso es una especie de “árbol promedio”
- Sólo se mantienen aquellos agrupamientos que se repiten un cierto número de veces (predeterminado).

Por ejemplo:

Si el umbral predeterminado es 50%, en el árbol consenso sólo permanecerán aquellos agrupamientos que hayan aparecido en al menos el 50% de los árboles construidos a partir de las  $N$  pseudorréplicas.

- El bootstrap no evalúa la topología global, sino que proporciona el *apoyo estadístico* de los NODOS individuales

# Árboles a partir de pseudorréplicas



# Comentarios adicionales sobre los árboles de consenso

Hay varios tipos de consensos; vimos el uso del criterio de la mayoría

Otro muy usado es el **consenso estricto**: incluir solamente los nodos presentes en el 100% de las pseudorréplicas (o, más en general, de los árboles bajo consideración)

## Importante:

Comúnmente del análisis de una matriz se obtienen más de un árbol más corto:



necesidad de obtener árbol de consenso

## Algunas conclusiones

- la parsimonia ilustra aspectos comunes a todos los métodos
- la magnitud del problema (muchos árboles posibles), la inevitable complejidad de los procesos evolutivos, la parcialidad y “ruido” de los datos, hacen que...
- ningún método puede garantizar la obtención del árbol correcto (excepto bajo condiciones muy particulares)
- deberíamos retener una cuota de cautela al hablar de nuestras hipótesis filogenéticas
- deberíamos considerar y evaluar hipótesis alternativas

## Algunas conclusiones (cont.)

- en general, es deseable disponer de conjuntos de datos independientes (genes y morfología, loci no ligados, etc.)
- debemos recordar que las discordancias entre distintas clases de datos pueden tener múltiples causas:
  - ruido (homoplasia) en una o varias clases de datos
  - discordancia superficial: todos los datos son compatibles con un único árbol (aunque distintos conjuntos resulten en árboles diferentes)
  - historias diferentes para las distintas clases de datos (árbol de genes versus árbol de especies, introgresión, etc.)