

Filogenias

Jose Blanca
COMAV institute
bioinf.comav.upv.es



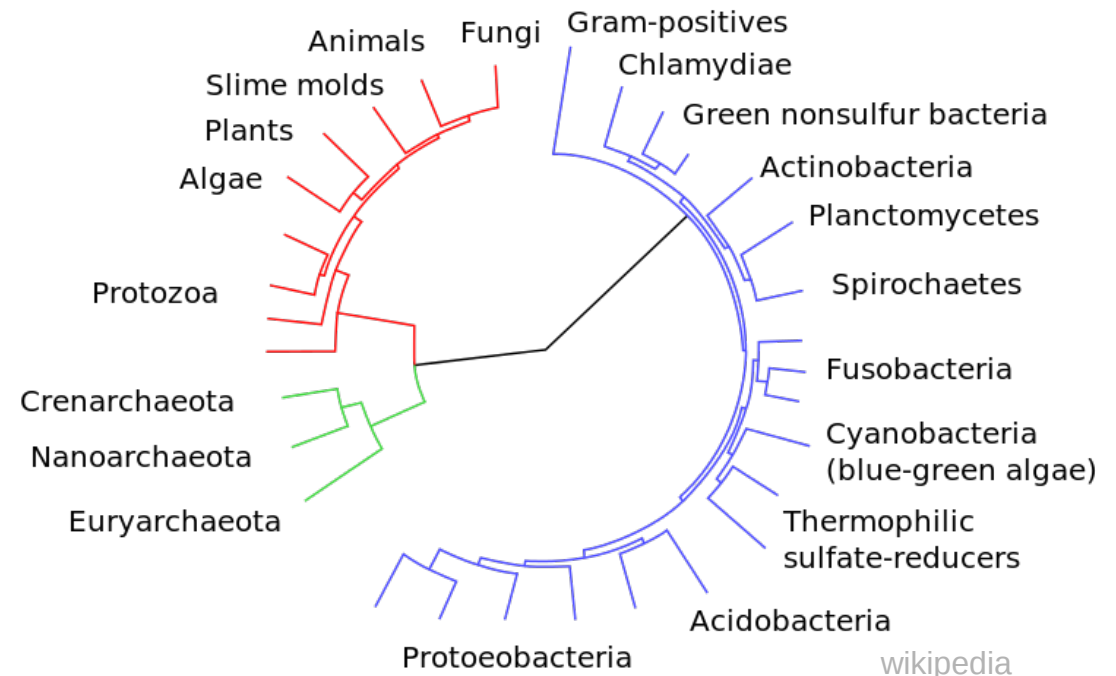
Filogenias

Estudio de las relaciones evolutivas

Asumimos:

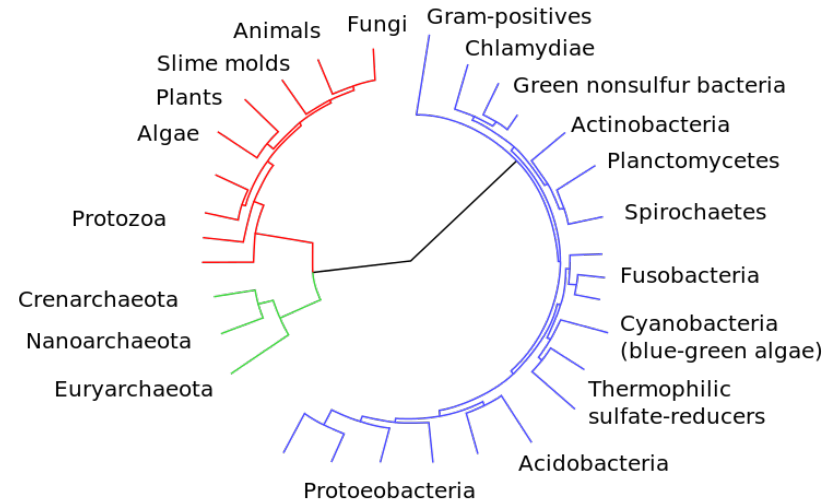
- Ancestro común
- Bifurcación

Es habitual en las poblaciones y especies cercanas que no se de la bifurcación



Reconstrucción filogenética

Evidencias



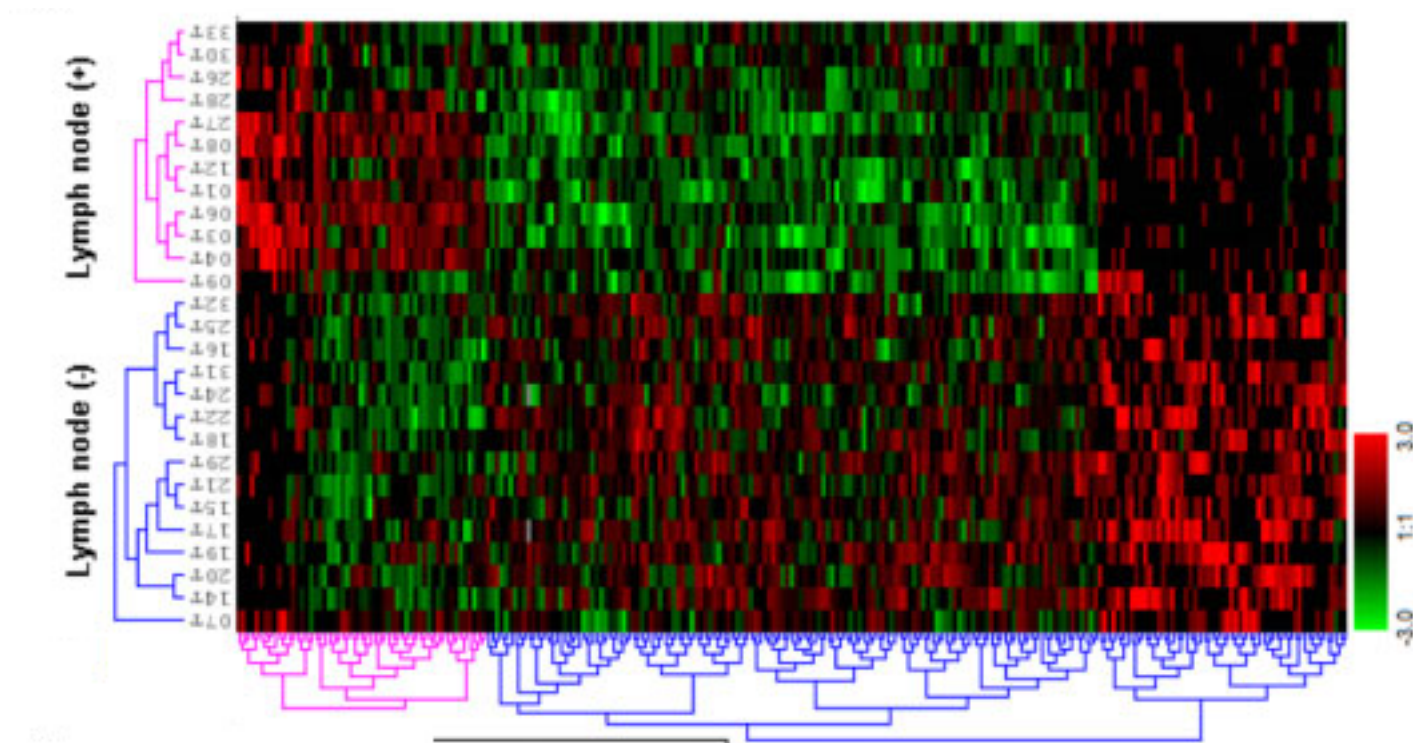
Tipos evidencia:

- Morfológicas
- Genotipos
- secuencias ADN o proteína

Filogenia vs taxonomía

En taxonomía se pretende clasificar.

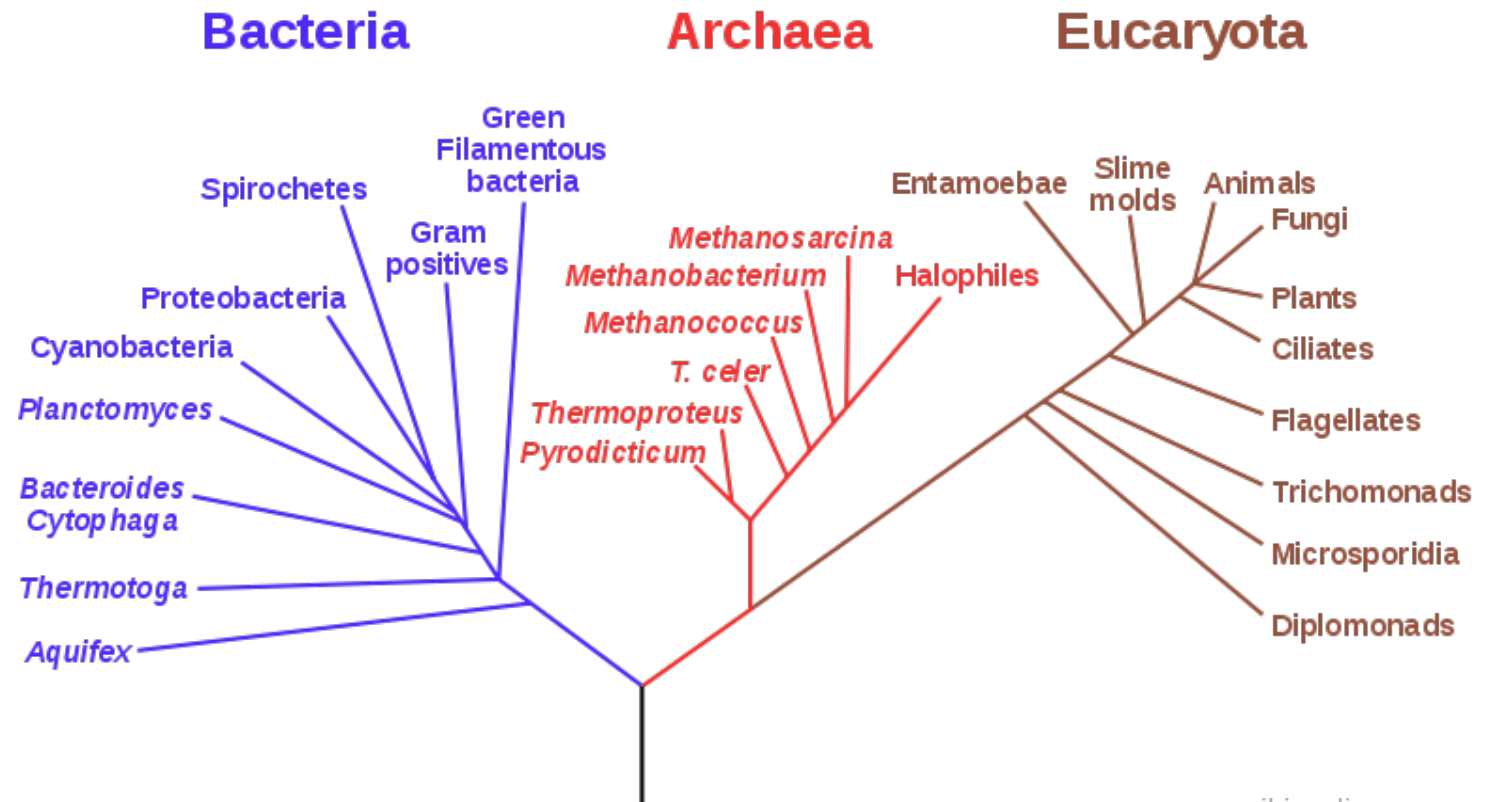
Se puede hacer una taxonomía de coches o de tipos de cánceres, pero no una filogenia.



Dendogramas

Topología → historia evolutiva

Longitud ramas → tiempos o distancias



Terminología

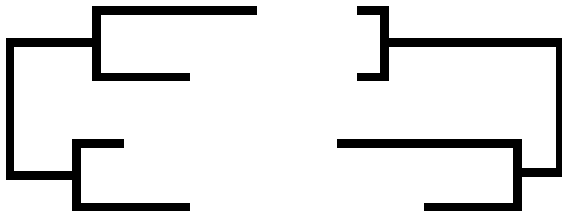
Clado: conjunto de especies emparentadas por un antepasado común.

Taxón: clado con categoría taxonómica

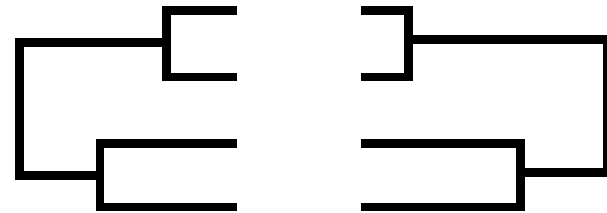
Filograma: cladograma con distancias

Árbol ultramétrico: todas las ramas miden lo mismo hasta la raíz (suele representar tiempos).

No ultramétrico



ultramétrico



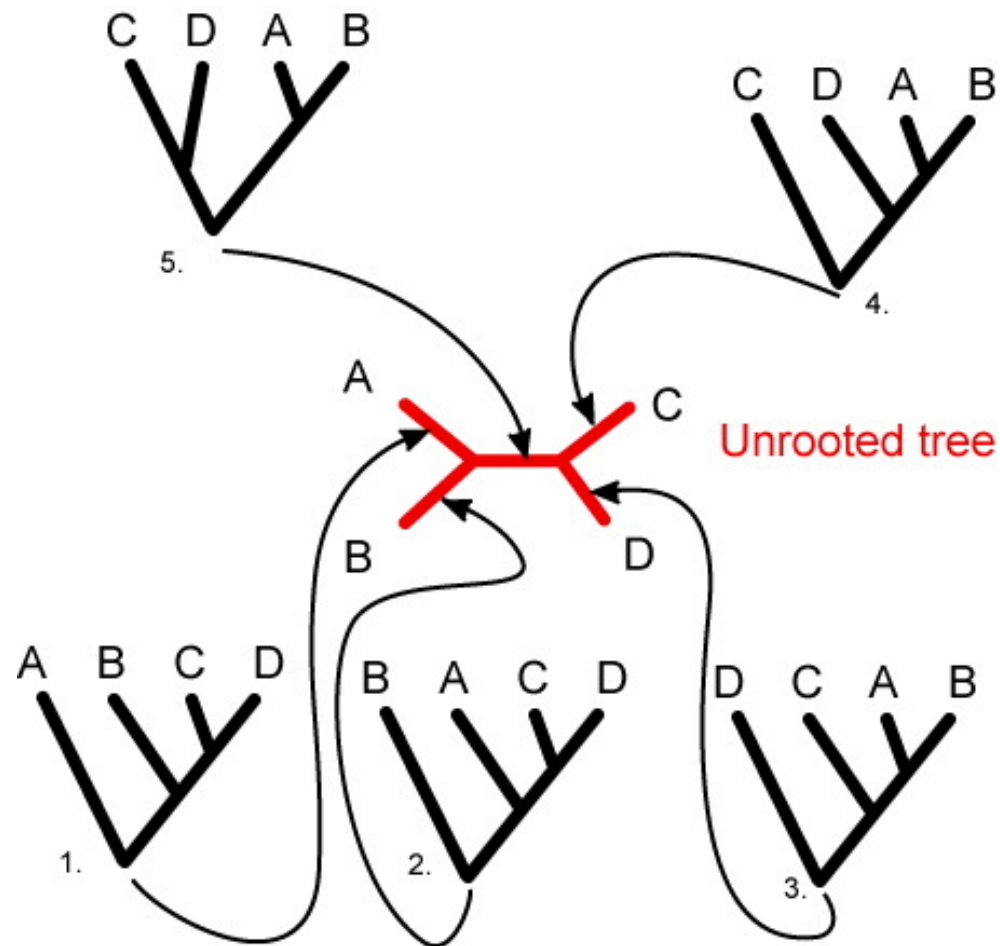
Topología

----- A
-----|
|----- B
|
-----C
-----D

----- B
-----|
|----- A
|
-----C
-----D

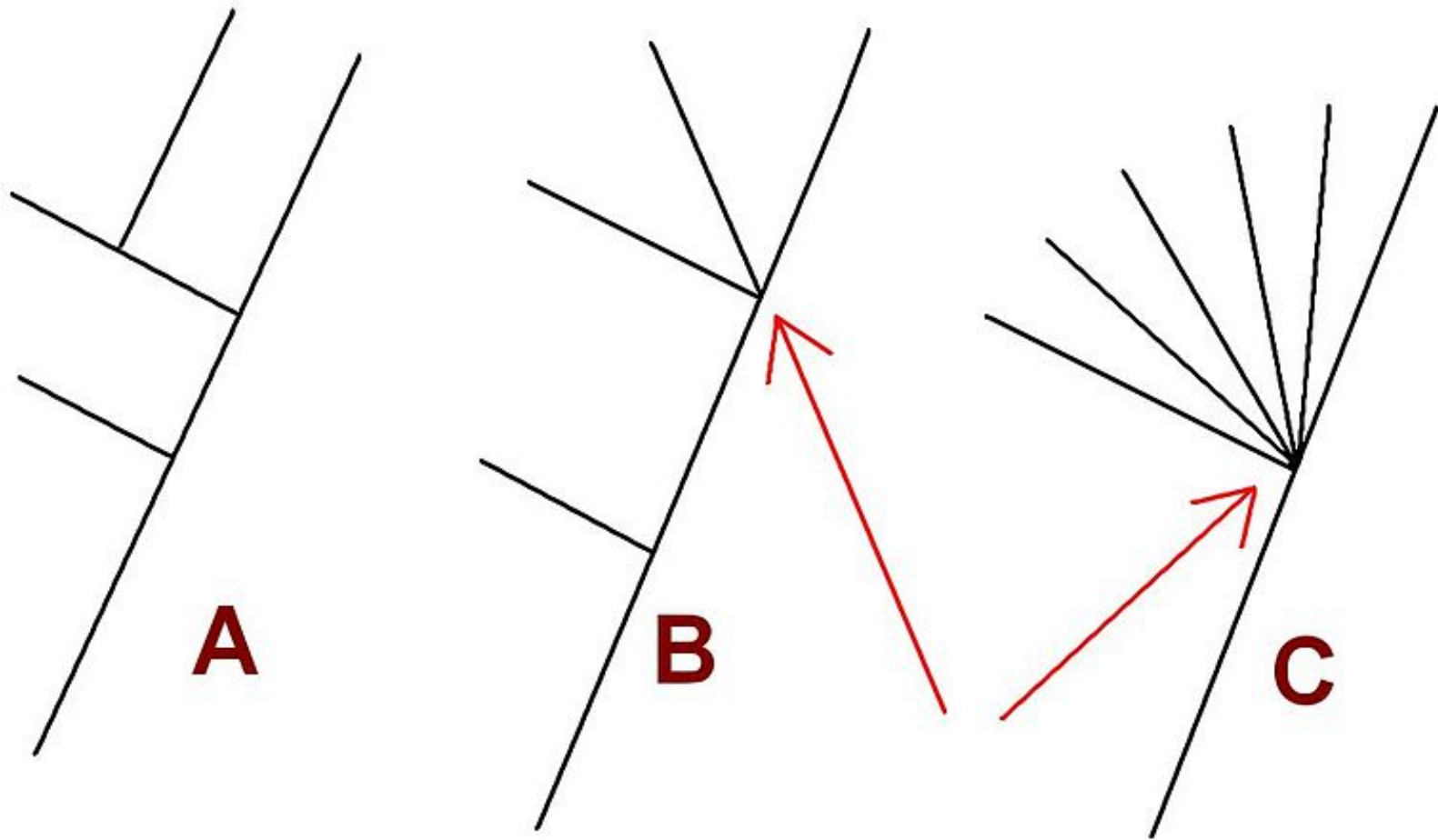
----- B
-----|
|----- A
|
-----D
-----C

Raíz



Se puede añadir una raíz incluyendo un outgroup (a priori sabemos que taxón es el más antiguo)

Politomía

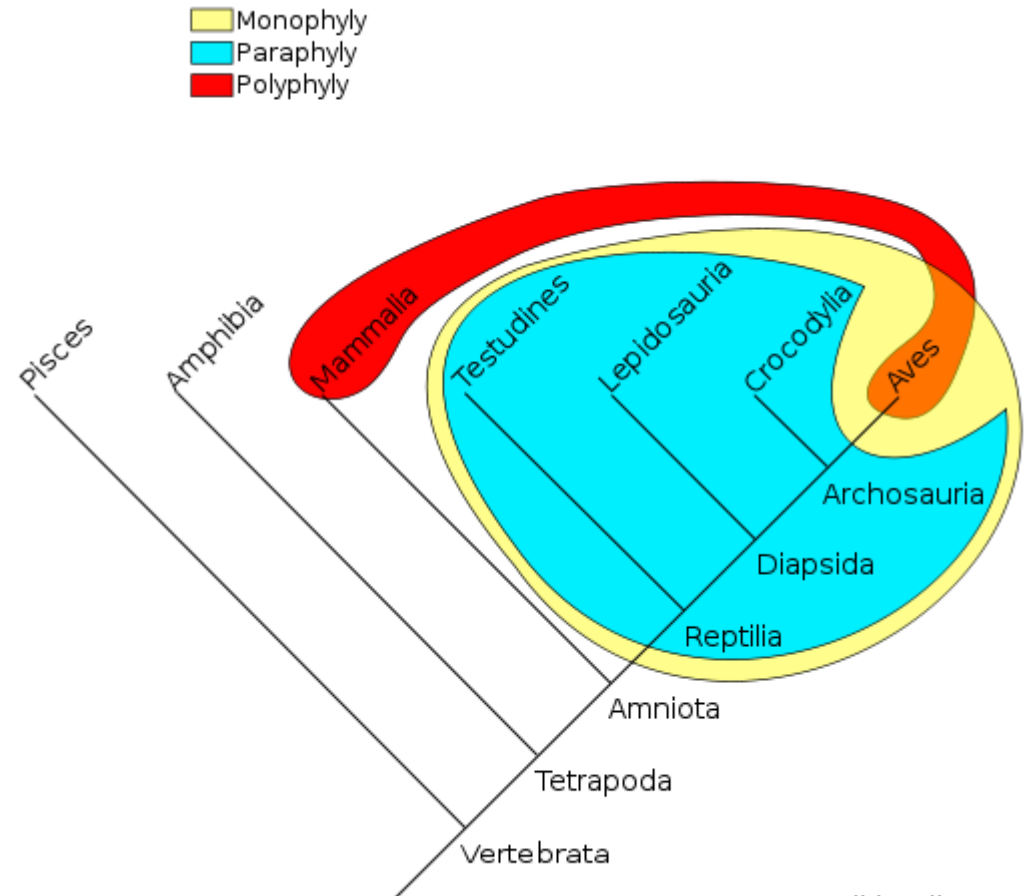


Monofilia, polifilia y parafilia

monofilético: un único antecesor

polifilético: varios antepasados (ej. gusanos)

parafiléticos: un único antecesor, pero no todos descendientes (ej. reptiles)

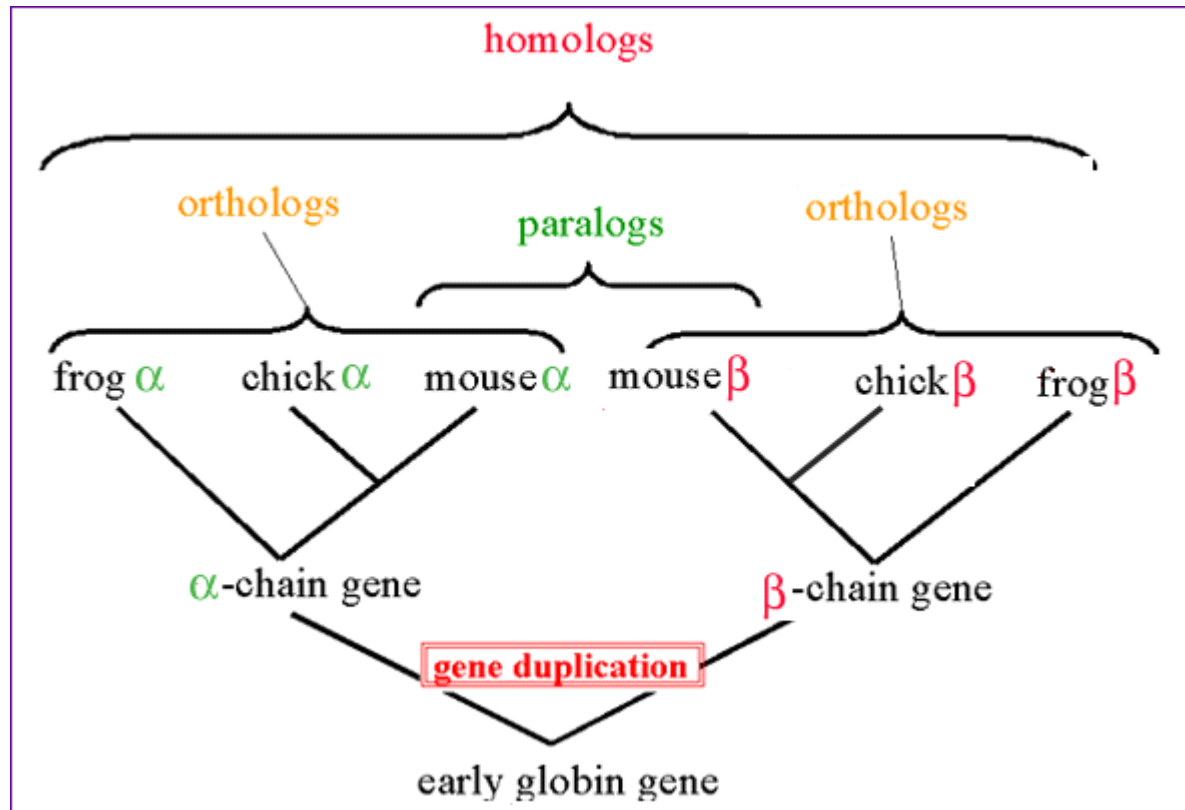


Homología

Ortólogo: Secuencias especies diferentes ancestro común

Parólogo: Secuencias misma especie ancestro común

Xenólogo: Secuencias especies diferentes ancestro común, transferencia horizontal



Premisas reconstrucción

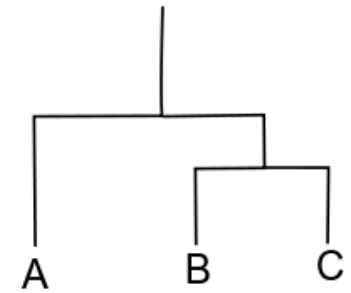
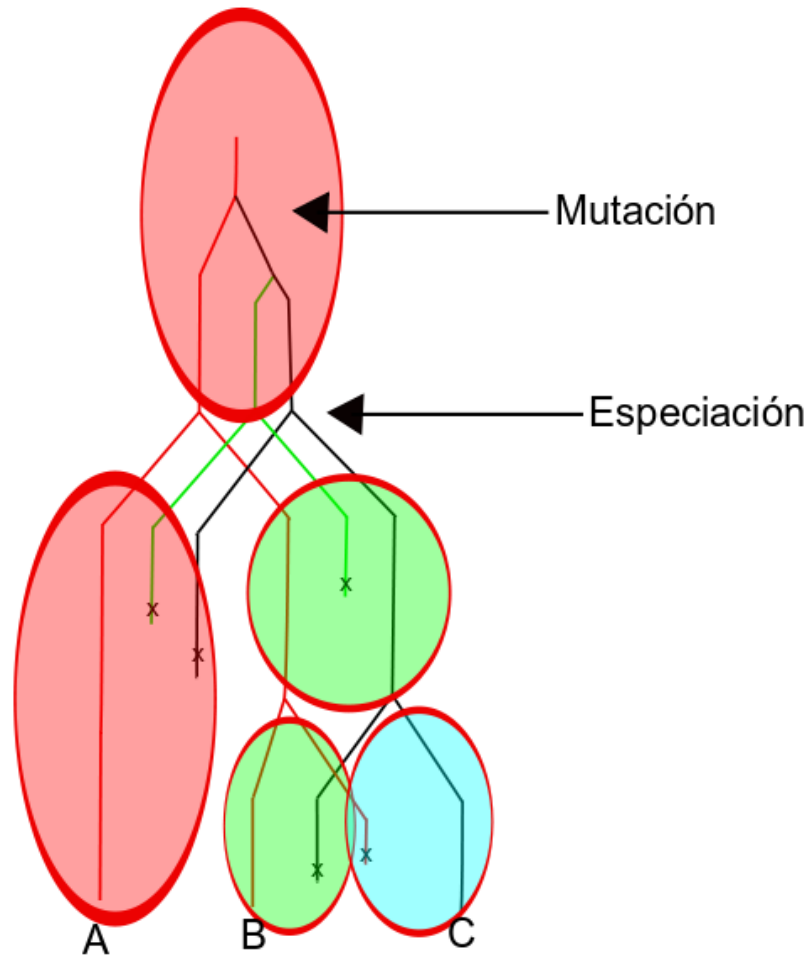
Premisas reconstrucción:

- Secuencia es correcta y pertenece a su organismo.
- Secuencias homólogas y ancestro común.
- Posición alineamiento homólogas.
- Las secuencias evolucionan al azar.
- Evolución independiente entre posiciones.
- No recombinación.

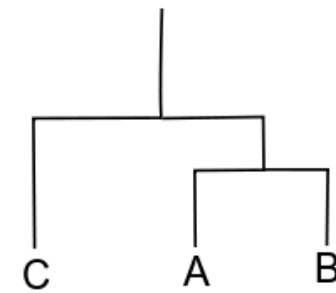
Asumimos información suficiente. Debe comprobarse.

Árbol especies vs árbol secuencias

Especies vs secuencias



Árbol especies



Árbol genes

Métodos de reconstrucción

Distancias

Parsimonia

Máxima verosimilitud

Bayesianos

Métodos basados en distancias

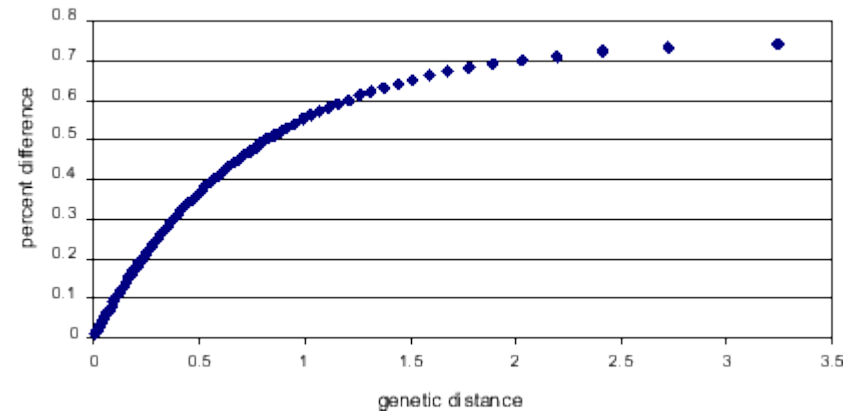
Evidencias → Matriz distancias → árbol

Distancias:

- número de cambios (depende de la longitud secuencia)
- % cambios. (saturación)
- modelos de sustitución.
 - diferencias entre las distancias tasas de sustitución: ej. transiciones vs transversiones
 - Distintas velocidades evolución entre distintos sitios.

Árboles a partir de matriz distancia:

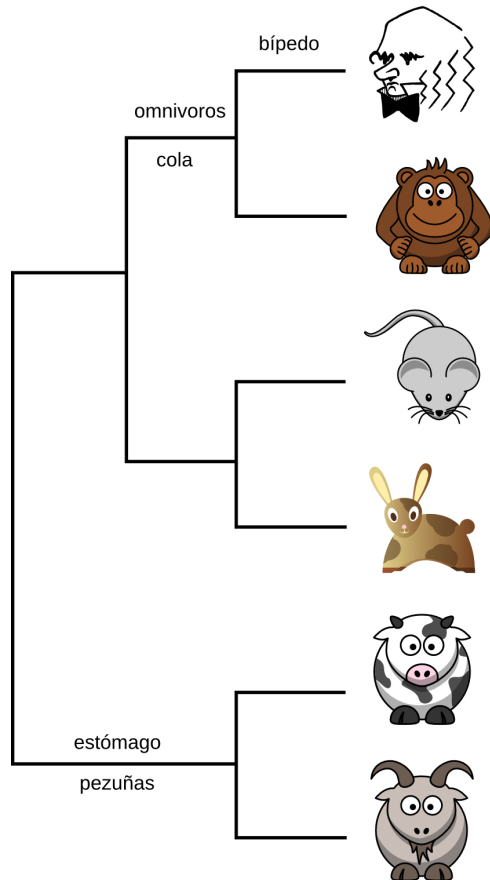
- UPGMA: ultramétrico
- Neighbor-joining: No ultramétrico



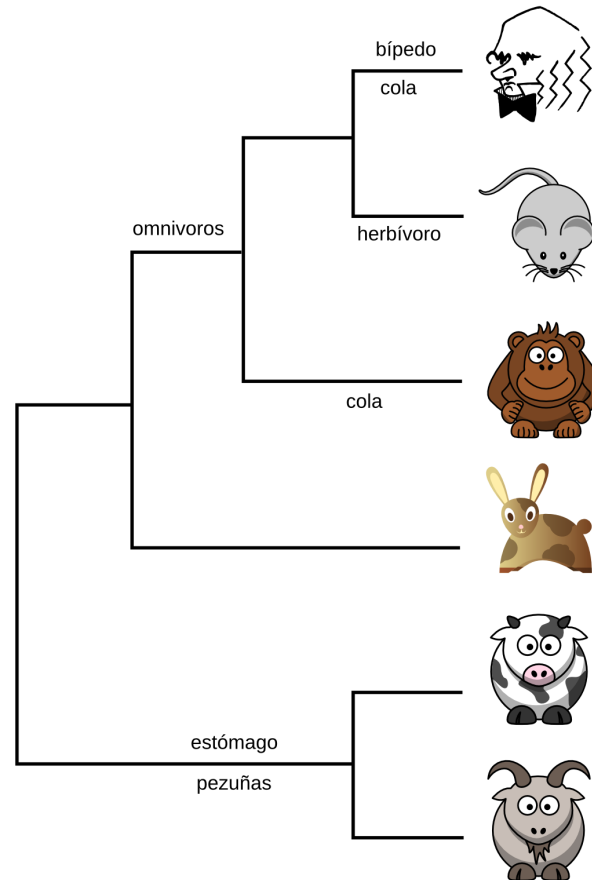
Parsimonia

	Alimentación	Estómago	Pezuñas	Bípedo	Cola
Hombre	Omnívoro	Simple	No	Sí	No
Chimpancé	Omnívoro	Simple	No	No	No
Ratón	Herbívoro	Simple	No	No	Sí
Conejo	Herbívoro	Simple	No	No	Sí
Vaca	Herbívoro	Compuesto	Sí	No	Sí
Cabra	Herbívoro	Compuesto	Sí	No	Sí

5 cambios



7 cambios



Parsimonia

La explicación correcta es la de menos cambios.

Se buscan los árboles con menos cambios.

- Métodos heurísticos

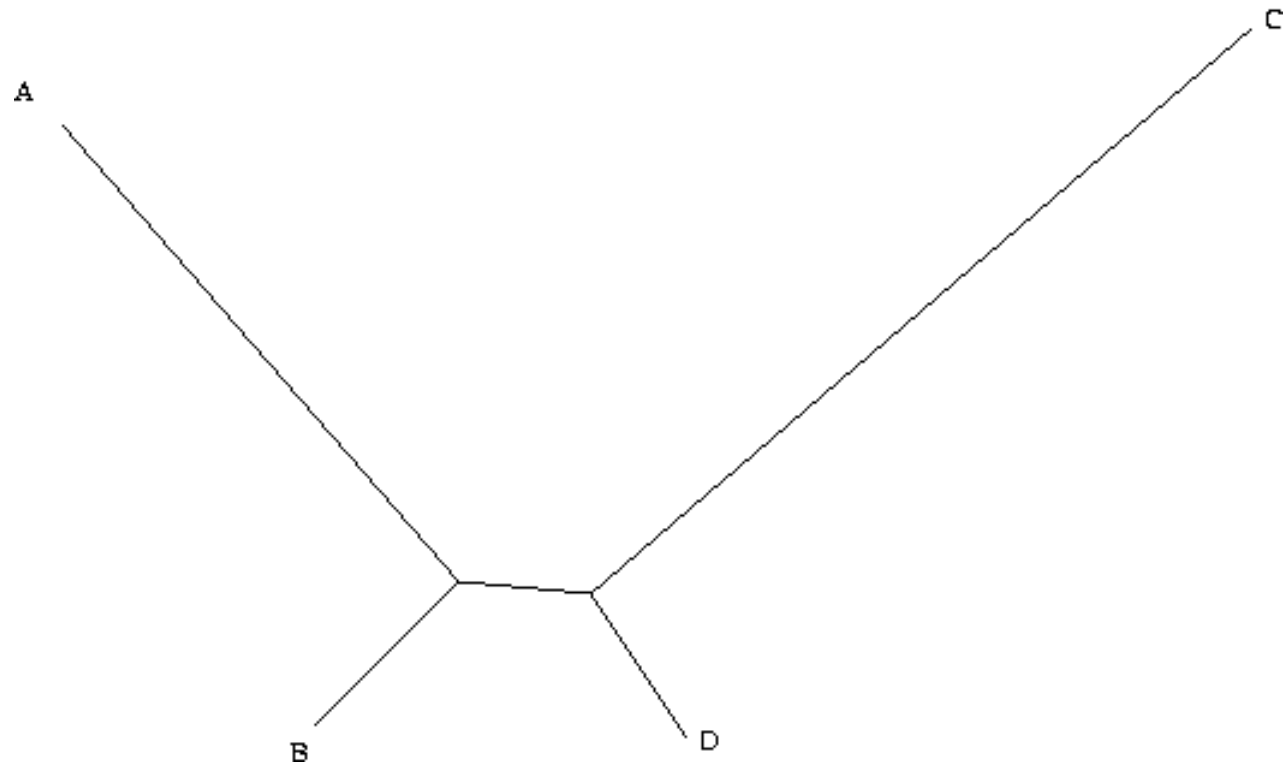
Tipos de caracteres:

- no variables :(
- variables en todos :(
- autoapomorfías: variable sólo en un taxón :(
- sinapomorfías: compartidos por varios, pero no por todos (caracteres derivados) :)

A	aat	tcg	ctt	cta	gga	atc	tgc	cta	atc	ctg
Ba	..gt.	t..a
Ca	..ct	t.a
Da	..ag	..t	...	t.t	..t	t..
	1	2	3			4				5

Problema parsimonia

Long branch attraction



Máxima verosimilitud

Búscala el árbol maximoverosimil

- Evidencias + modelo -> árbol más probable

Métodos heurísticos de búsqueda del árbol.

Elección modelo mutación:

- más parámetros, mejor ajuste, peores varianzas.
- programas para quedarse con el más adecuado a nuestros datos.

Ventajas:

- A partir de los mismos datos mejor estadísticamente que distancias y parsimonia.
- No problema con ramas largas si hay suficiente información

Desventajas:

- Coste computacional.

Métodos bayesianos

Inferencia bayesiana:

evidencias + prob a priori de los modelos \rightarrow prob posterior de los modelos

Modelo:

- parámetros modelo mutación
- árbol

Metropolis-coupled Montecarlo vía Cadenas de Markov (MCMCMC) para explorar el espacio de árboles.

Computacionalmente muy lento.

This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Jose Blanca
COMAV institute
bioinf.comav.upv.es

