Predicción Filogenética

Rodrigo Santamaría

Predicción Filogenética

Introducción

Trasfondo biológico

Árboles

Análisis

Métodos

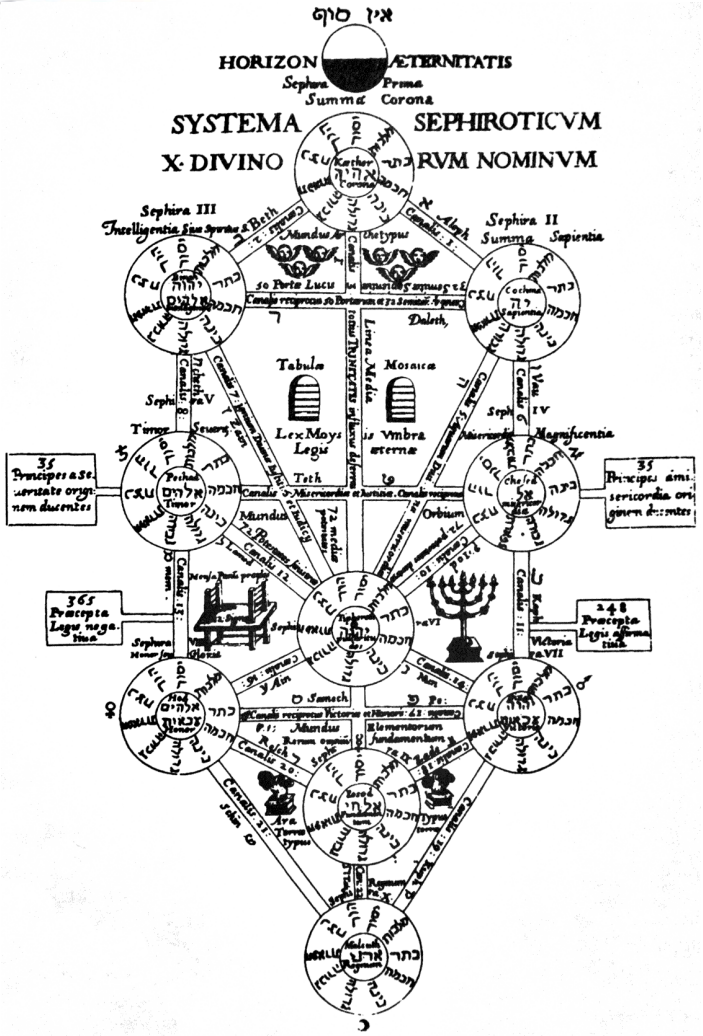
Introducción

- **Teoría de la evolución**: los organismos cambian con el tiempo, de manera que los descendientes difieren funcional y estructuralmente respecto a su ancestro
  - Los organismos pueden clasificarse según sus relaciones ancestrales
- **Filogenética o filogenia**: reconstrucción de las relaciones ancestrales entre los organismos
  - Representación: "El Árbol de la Vida"
  - Principio: agrupar los seres vivos de acuerdo a su nivel de similitud

Introducción

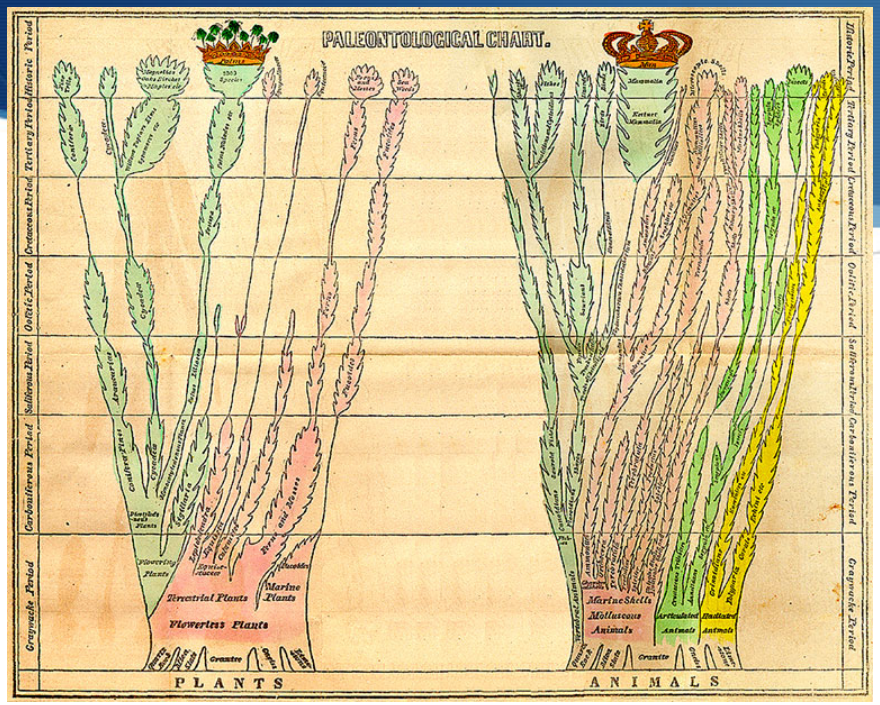
- Las comparación entre organismos se puede abordar de dos maneras
  - **Filogenética Tradicional**: a través de sus fenotipos
    - P.ej. "presencia o ausencia de alas"
  - **Filogenética Molecular**: a través de sus secuencias
    - Es en la que nos centraremos en el ámbito de la bioinformática
- **Árbol verdadero**: representa los eventos de diferenciación reales ocurridos durante la evolución. Imposible de generar
- **Árbol inferido**: representa una serie de eventos evolutivos inferidos a partir de los datos disponibles, basándonos en algún modelo

Yggdrasil
(mitología nórdica)



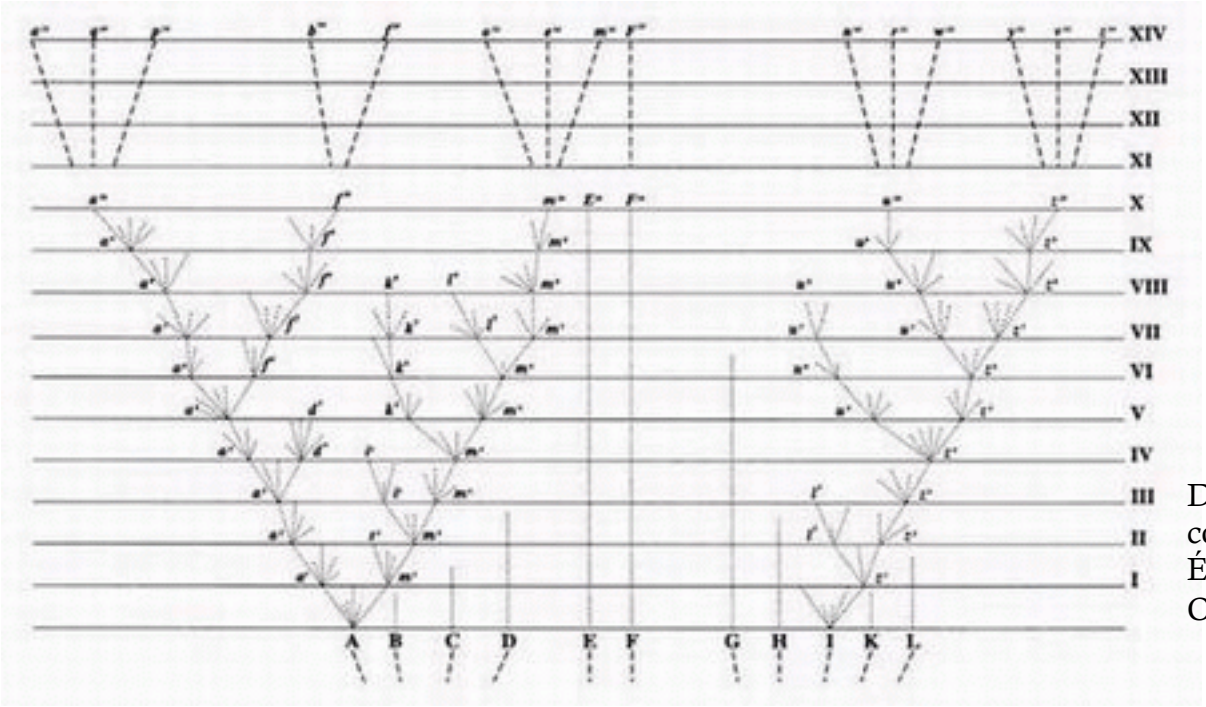
Sepher Yetsirah
(mitología hebrea)

El árbol de la vida tiene históricamente un componente filosófico y cosmogónico (S XIII o anterior)

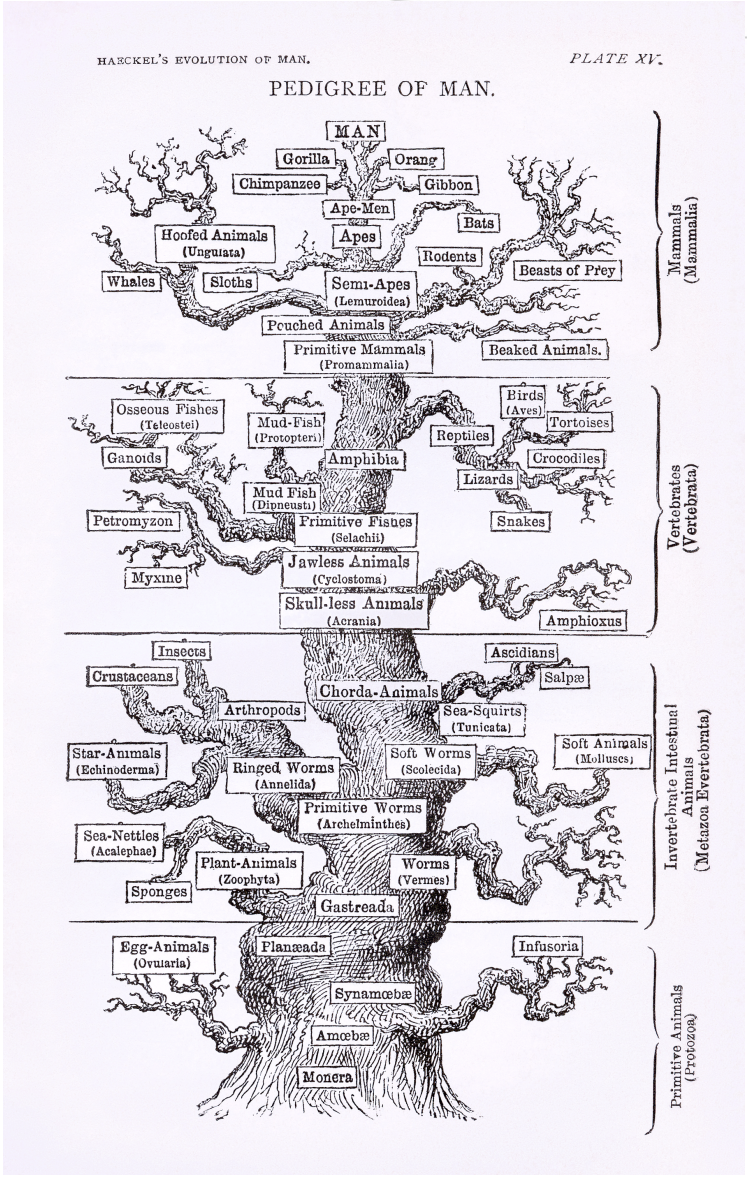


Los primeros árboles de la vida en términos de filogenética tradicional (S XVIII) no tenían en cuenta un ancestro común

Hitchcock 1840, separa animales y plantas



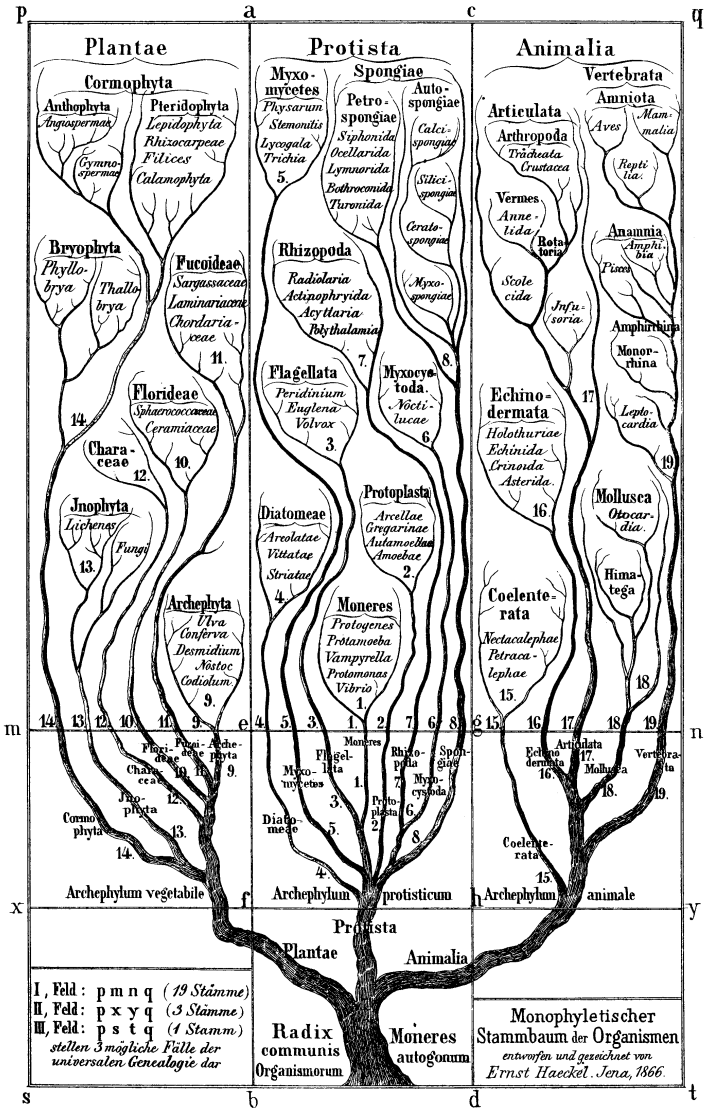
Darwin (1859) intuye ancestros comunes.
Ésta es la única ilustración de "El Origen de las Especies"

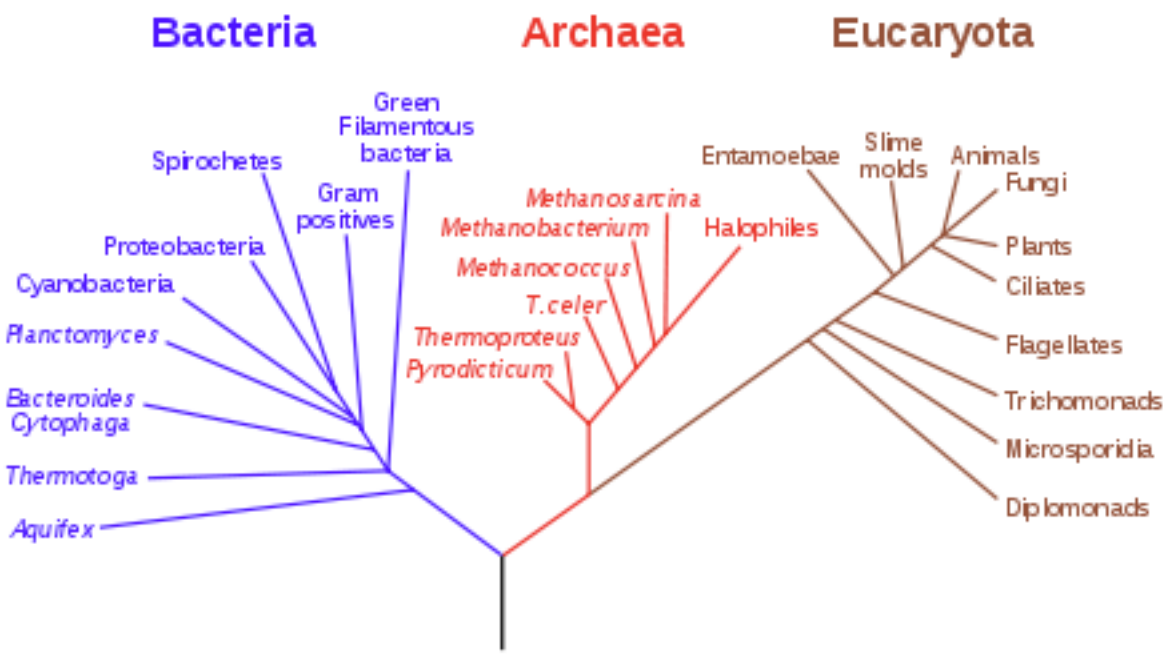


Los árboles de Haeckel ya incluyen una "raíz"

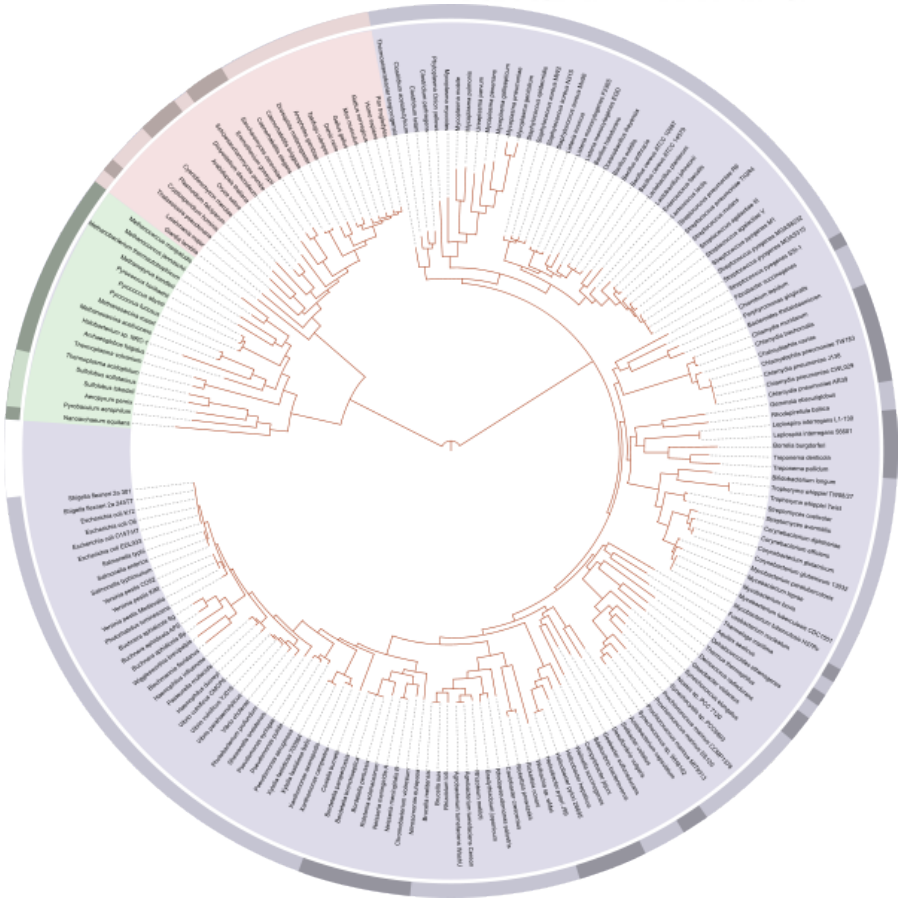
← Su primer árbol (1866) insinúa el ancestro del hombre

Su segundo árbol (1879) abandona el antropocentrismo →

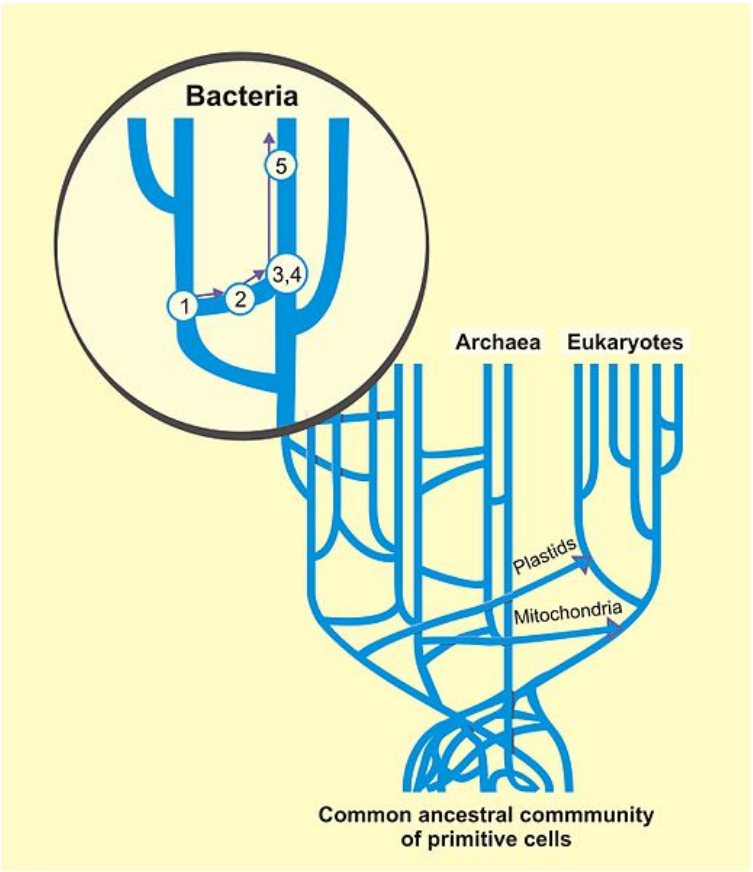




Árbol filogenético, inferido por la comparación de genes ribosómicos
Tres ramas principales: bacterias, arqueas y eukariotas



Representación gráfica del Tree of Life Web Project



Árbol de la vida mostrando los mecanismos de transferencia genética horizontal

Gracias a la filogenética molecular podemos comparar y añadir más organismos y eventos evolutivos

Predicción Filogenética

Introducción

Trasfondo biológico

- Reloj Molecular
- Selección Negativa y Positiva
- Teoría Neutral

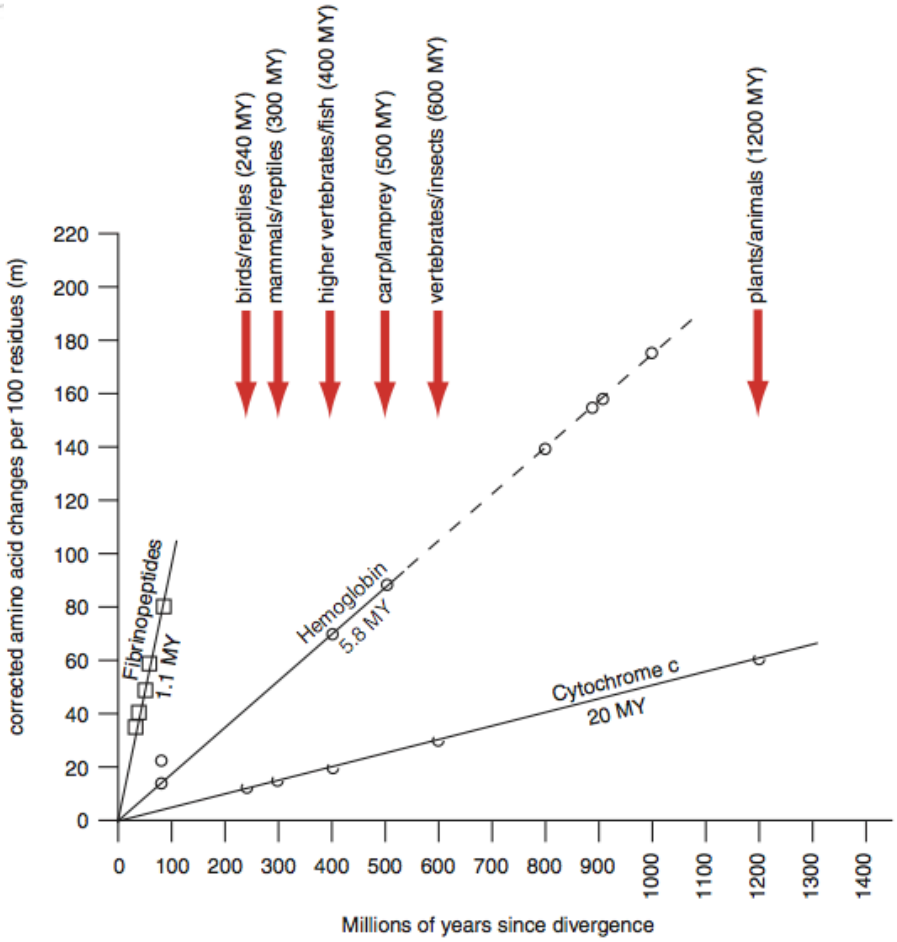
Árboles

Análisis

Métodos

Hipótesis del reloj molecular

- "Para cada gen o proteína, la tasa de evolución molecular es aproximadamente constante"
  - Hipótesis propuesta por Zuckerland y Pauling (1962)
- Soportada por el estudio de Dickerson (1971) sobre la divergencia en tres proteínas
  - Representa el nº de cambios en sus aminoácidos, en distintos organismos, contra el tiempo de divergencia (en millones de años, MY) entre dichos organismos



Reloj molecular

- Dickerson calcula el nº de sustituciones reales (*m*) a partir del nº de sustituciones observadas (*n*) por cada 100 residuos

$$\frac{m}{100} = -\ln(1 - \frac{n}{100})$$

- Conclusiones respecto a las tasas de sustitución
  - Son lineales para cada proteína
  - Varían para proteínas distintas
  - Esta variación entre proteínas responde a limitaciones funcionales impuestas por la selección natural

Reloj molecular

- Tasa de sustitución: número de cambios en una proteína por unidad de tiempo
  - "Frecuencia" del reloj molecular
- Las tasas de sustitución NO son tasas de mutación
  - Las mutaciones son el proceso bioquímico de cambio en una secuencia, y ocurren a un ritmo constante (p.ej. la tasa de error de la polimerasa)
  - La sustitución es el cambio observado en la secuencia, y se debe tanto a la mutación como a la selección
    - Teniendo en cuenta que la tasa de mutación es relativamente constante, la sustitución se debe a selección positiva o negativa

Protein	Rate	Protein	Rate
Fibrinopeptides	9.0	Histone H3	0.014
Growth hormone	3.7	Ubiquitin	0.010
Immunoglobulin (Ig) kappa chain C region	3.7	Histone H4	0.010

Tasa de sustitución de aminoácidos cada 10^9 años

Reloj molecular
Test de Tajima

- Test de tasas relativas de Tajima (1993): determina si las secuencias de dos organismos *A* y *B* evolucionan al mismo ritmo
  - Es un test de sus relojes moleculares: la hipótesis nula es que evolucionan al mismo ritmo
    - Si la rechazamos es que los organismos evolucionan a ritmos distintos
  - Para realizar el test se necesita un tercer organismo C que sirva de control o comparación con ambos
    - Debería ser el organismo más cercano a ambos pero que no sea más cercano a uno que a otro → su elección es difícil
      - Si comparamos humano y chimpancé, elegir el bonobo no es adecuado (es más cercano al humano) y elegir el ratón es demasiado lejano. Una opción adecuada sería el orangután o el gorila

Reloj molecular
Test de Tajima

- Sea m_1 el nº de residuos en *A* que difieren de los de *B* y *C*
  - Análogamente, sean m_2 los de *B* que son distintos a los de *A* y *C*
- Dado que *C* es un grupo externo, se espera que *A* y *B* sean iguales respecto a *C*: $m_1 \sim m_2$
- La igualdad se prueba con un análisis chi-cuadrado: $X^2 = \frac{(m_1 - m_2)^2}{m_1 + m_2}$
- Se observa el p-valor asociado a X^2, si es menor que, p. ej. 0.05, indicará que rechacemos que los organismos evolucionan a la par

Selección positiva y negativa

- Los atributos que mejoran la adaptación son seleccionados (selección positiva) y los que la reducen descartados (selección negativa)
  - Esto ocurre también a nivel molecular con las secuencias de ADN
- Por ejemplo, el gen de la lisozima, una enzima que sirve como proteína antimicrobiana en la leche, saliva y lágrimas
  - Hace 25MY se duplicó para asumir la misma función pero en el estómago del ancestro de los bovinos, y de forma independiente lo hizo también hace 15MY en los primates.

Teoría neutral de la evolución molecular

- "La mayoría de las sustituciones de ADN observadas deben ser neutrales o casi neutrales" (Kimura, 1968, 1983)
  - Asumiendo esta teoría la selección darwiniana tiene un papel secundario (fenotípico), mientras que la deriva genética gana peso a nivel molecular
- Se basa en la observación de que la tasa media de sustitución es de 1 cambio cada 28MY, para proteínas de 100 residuos
  - Lo cual implica una tasa de sustitución en ADN muy alta (1bp cada 2 años)
    - La mayoría de ellas deben ser inocuas, o se observarían más mutaciones

Predicción Filogenética

Introducción

Trasfondo biológico

Árboles

- Características
- Tipos

Análisis

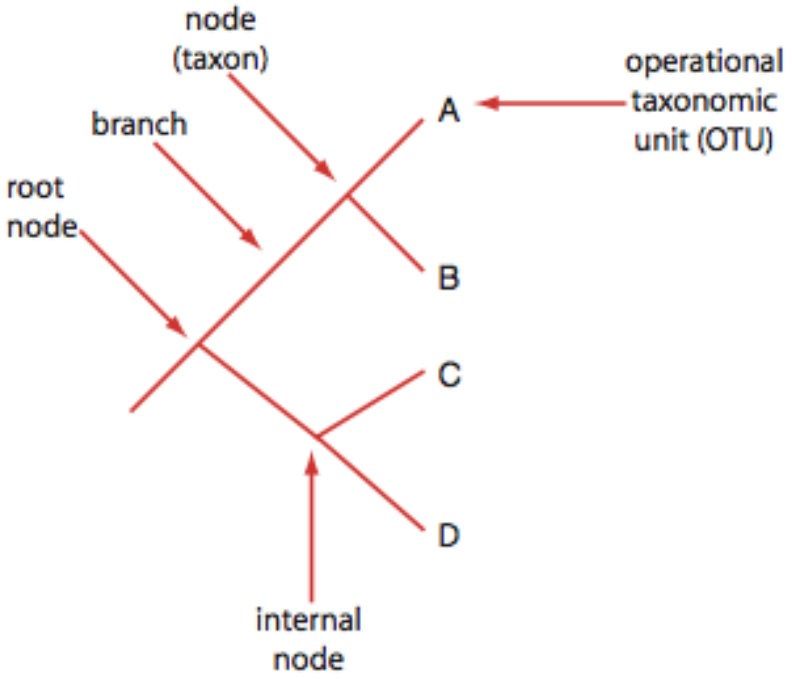
Métodos

Árboles

- La filogenética molecular estudia las relaciones evolutivas, desde distintos campos (morfología, anatomía, fisiología, paleontología)
  - Nos centraremos en su estudio mediante la construcción de árboles filogenéticos a partir de secuencias
- **Árbol**: grafo en el que dos nodos sólo están conectados por un camino de relaciones ancestro-descendiente
  - Nodo: representa una unidad taxonómica
  - Rama: conecta dos nodos

Árboles

- Nodo interno (o punto de divergencia)
  - Representa ancestros hipotéticos de los taxones
    - HTU: Hypothetical Taxonomic Unit
  - Nodo raíz: último nodo interno
    - Ancestro común más reciente de todos los taxones
- OTU: nodo hoja o externo
  - Representan las secuencias que estamos analizando

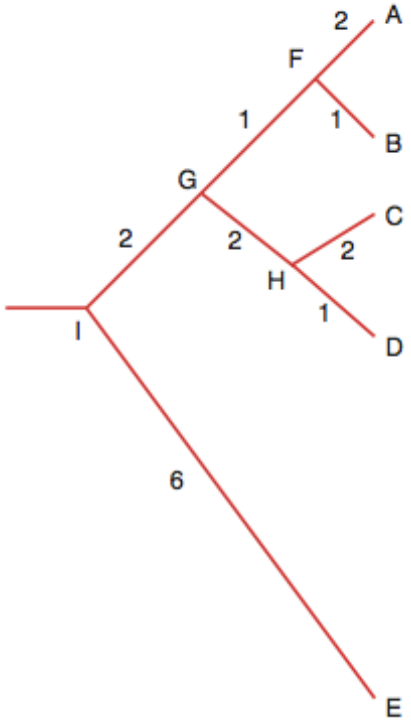


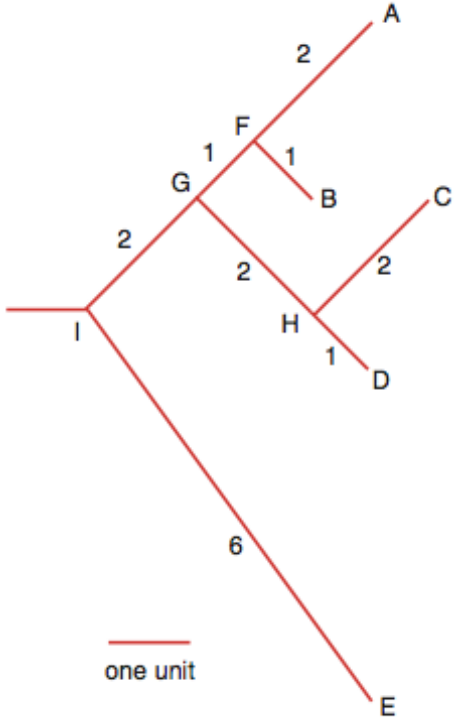
Árboles

- Características fundamentales
  - Topología: relaciones establecidas por los nodos internos
    - Determinan la clasificación de las secuencias
    - En algunos casos, las posiciones son intercambiables
  - Longitud de las ramas
    - Cuantifican el nivel de similitud entre secuencias
    - Puede también modelarse su anchura en función del bootstrapping
      - Cuantifica el nivel de consenso de la inferencia

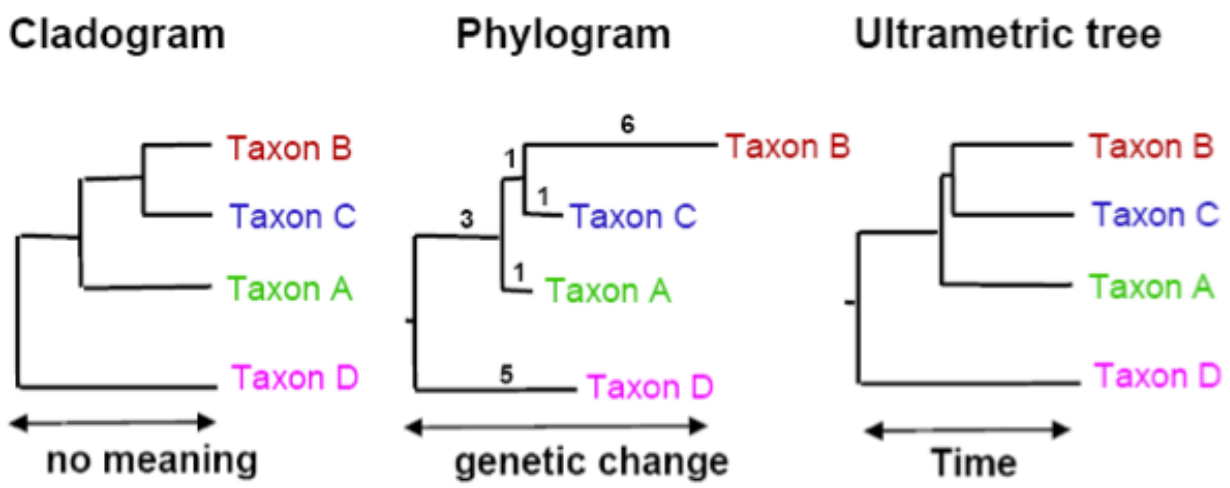
Tipos de árbol

- Árbol no escalado
  - La longitud de las ramas es constante
- Árbol escalado
  - La longitud de las ramas es proporcional al nº de cambios en la secuencia





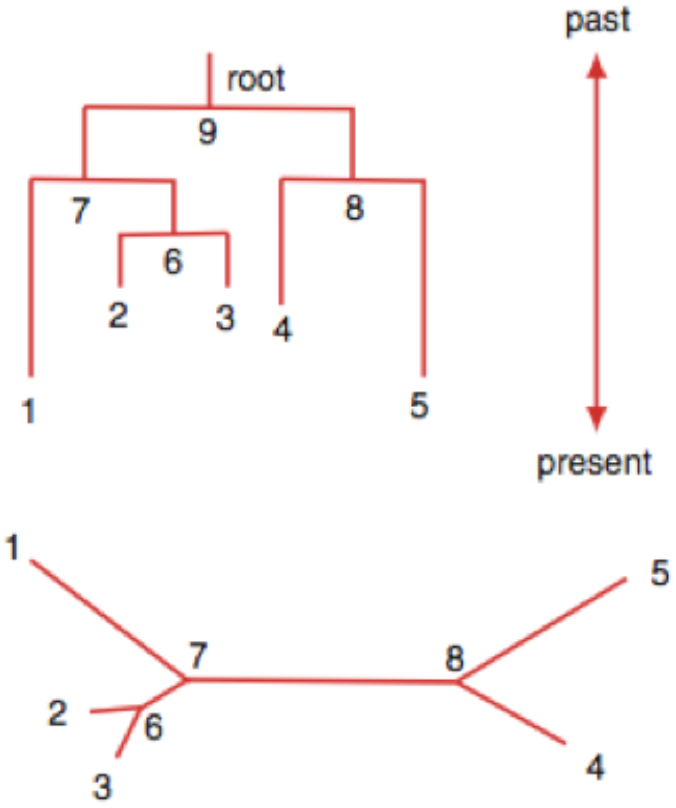
Tipos de árbol



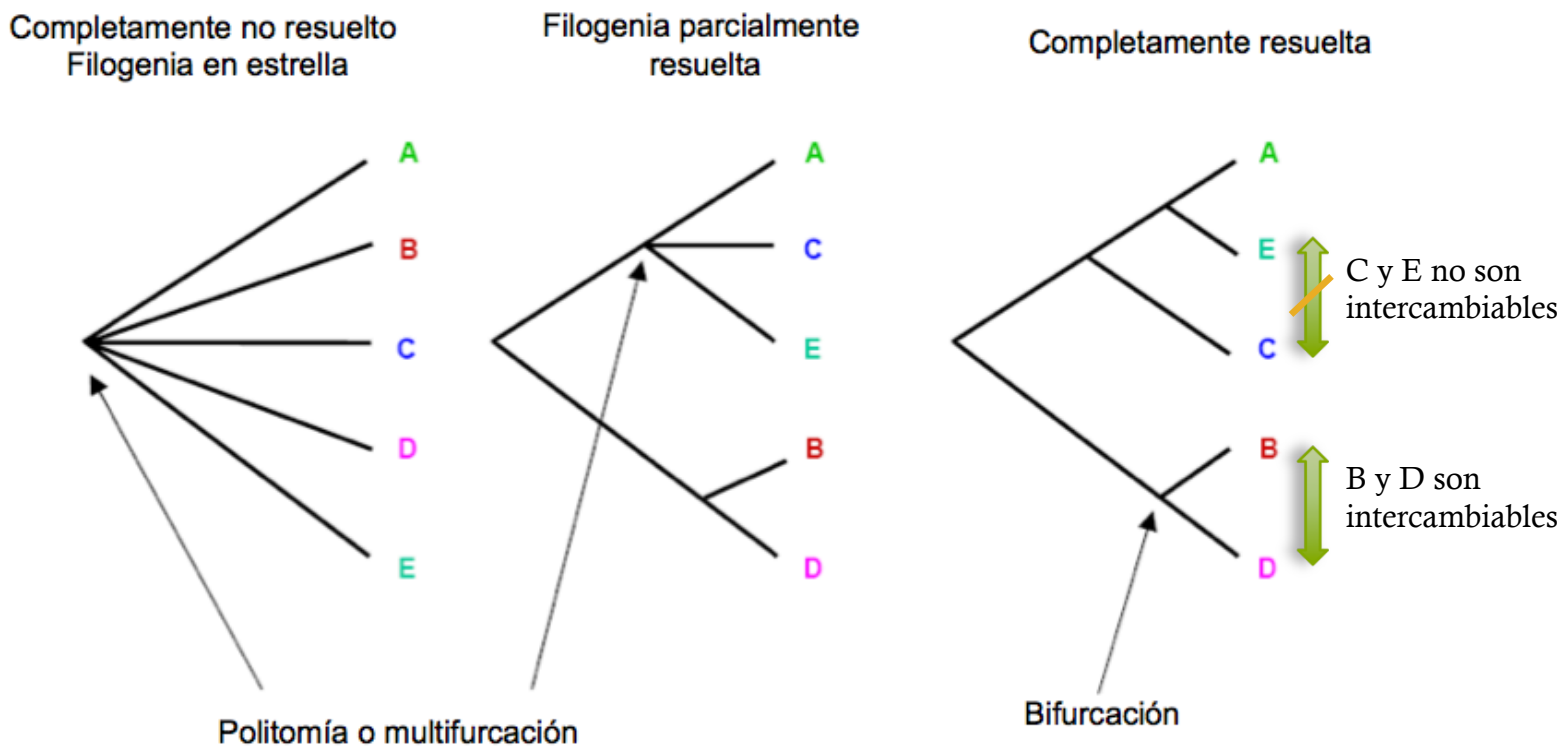
Los tres representan las mismas relaciones evolutivas, pero algunos aprovechan la escala para cuantificarlas

Tipos de árbol

- Árbol enraizado: tiene nodo raíz
  - Hay un ancestro común
  - Dirección temporal definida
- Árbol no enraizado
  - A veces el nodo raíz no es de interés o es difícil de localizar
  - Misma información de relaciones pero sin ancestro común ni dirección temporal



Tipos de árbol



Árboles y complejidad

- Nº de posibles topologías para n nodos finales:
  - Para árboles enraizados:
    - $N_r=(2n-5)!/2^{n-3}(n-3)$
  - Para árboles no enraizados:
    - $N_u=(2n-3)!/2^{n-2}(n-2)$
- A partir de n=12, es obligatorio usar heurísticas
  - Imposible calcular todos los árboles posibles

No. of OTUs	No. of Rooted Trees	No. of Unrooted Trees
2	1	1
3	3	1
4	15	3
5	105	15
6	945	105
7	10,395	945
8	135,135	10,395
9	2,027,025	135,135
10	34,489,707	2,027,025
15	213,458,046,676,875	8×10^{12}
20	8×10^{21}	2×10^{20}
50	2.8×10^{76}	3×10^{74}

Predicción Filogenética

Introducción

Trasfondo biológico

Árboles

Análisis

- Fases
- Modelos de sustitución
- Creación del árbol
- Evaluación

Análisis filogenético

- A partir de secuencias moleculares, construir un árbol filogenético que refleje sus relaciones
  - Desde un punto de vista de usuario final, se puede ver como una caja negra: “entran secuencias y salen árboles”
- Objetivo:
  - Comprender los distintos métodos de análisis filogenético
  - Saber manejar algunas herramientas para realizar análisis filogenéticos

Análisis filogenético
Fases

1. Selección de las secuencias a analizar
   - A partir de una de las BBDD vistas, en formato fasta
2. Análisis múltiple de secuencias
   - Mediante uno de los métodos o herramientas vistas
3. Elección de un modelo de sustitución
4. Construcción del árbol (inferencia filogenética)
5. Evaluación del árbol

Elección de secuencias y MSA

- La calidad de los datos de entrada es crítica
  - Si no, tendremos una solución GIGO (Garbage In, Garbage Out)
- En el caso de construcción filogenética implica:
  - Elegir secuencias que tenga sentido analizar evolutivamente
    - Asegurarse de que las secuencias son homólogas
  - Maximizar la bondad del MSA elegido
    - Probar distintos algoritmos y parámetros (matrices, huecos, etc.)
    - La información en el MSA debe ser consistente con el árbol construido

Modelos de sustitución

- Definición matemática de la distancia entre dos secuencias de longitud N
- Distancia de Hamming: cuenta el número de cambios (p)
- Distancia de Hamming normalizada: $p' = p/N$

A. Sequences

sequence A	ACGCGTTGGGCGATGGCAAC
sequence B	ACGCGTTGGGCGACGGTAAT
sequence C	ACGCATTGAATGATGATAAT
sequence D	ACACATTGAGTGATAATAAT

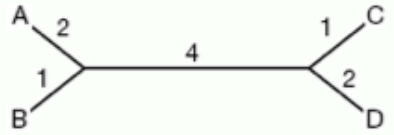
B. Distances between sequences, the number of steps required to change one sequence into the other.

n_{AB}	3
n_{AC}	7
n_{AD}	8
n_{BC}	6
n_{BD}	7
n_{CD}	3

C. Distance table

	A	B	C	D
A	–	3	7	8
B	–	–	6	7
C	–	–	–	3
D	–	–	–	–

D. The assumed phylogenetic tree for the sequences A-D showing branch lengths. The sum of the branch lengths between any two sequences on the trees has the same value as the distance between the sequences.



Modelos de sustitución

- Problema: sustituciones observadas vs reales

G	A	C	C	T	T	C	A	A	T	C	A	C	G	G	G	A	C	T
T	T	C	C	T	T	C	A	A	T	C	A	C	G	G	G	A	C	T
T	T	C	C	T	T	C	A	A	T	C	A	C	G	G	G	A	C	T
T	T	C	C	T	T	C	A	A	T	C	A	C	G	G	G	A	C	T
T	T	C	C	T	T	C	A	A	T	C	A	C	C	G	G	A	C	T
T	T	C	C	T	T	C	A	A	T	C	T	C	C	G	G	A	C	T
C	A	C	C	T	T	C	A	A	T	C	T	C	C	G	G	A	C	T
1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
2	2	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0

Observada: 3

Real: 6

Corrección de Jukes-Cantor

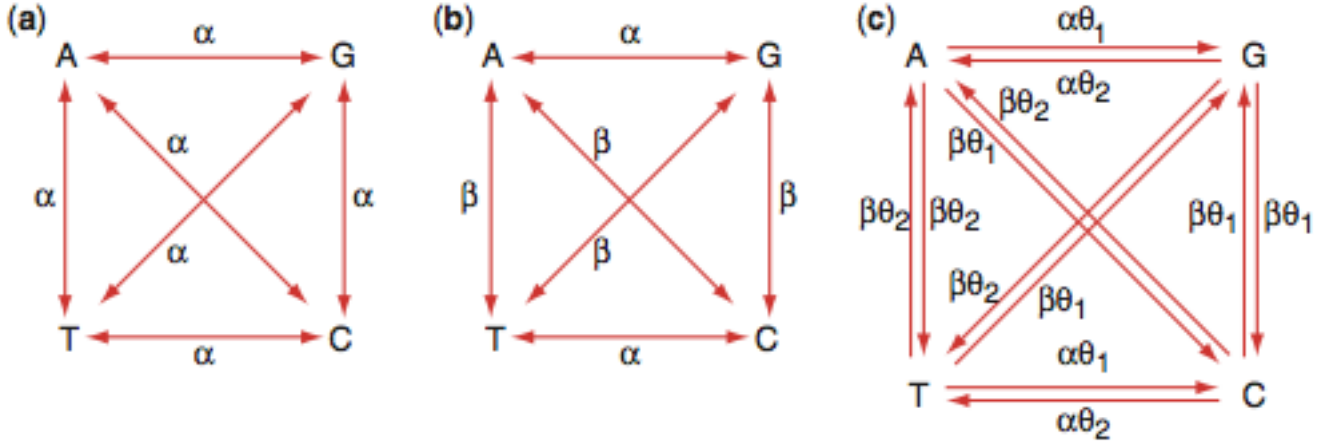
- Corrección de Jukes-Cantor (1969)
  - Sea *p'* la distancia de Hamming normalizada y *s* el número de residuos distintos (4 para nucleótidos, 20 para aminoácidos)

$$d = -\frac{s-1}{s}\ln\left(1-\frac{s}{s-1}p'\right)$$

  - *d* es una estimación del número de cambios reales
    - Considera que la probabilidad de sustitución es igual para todas las combinaciones de nucleótidos/aminoácidos

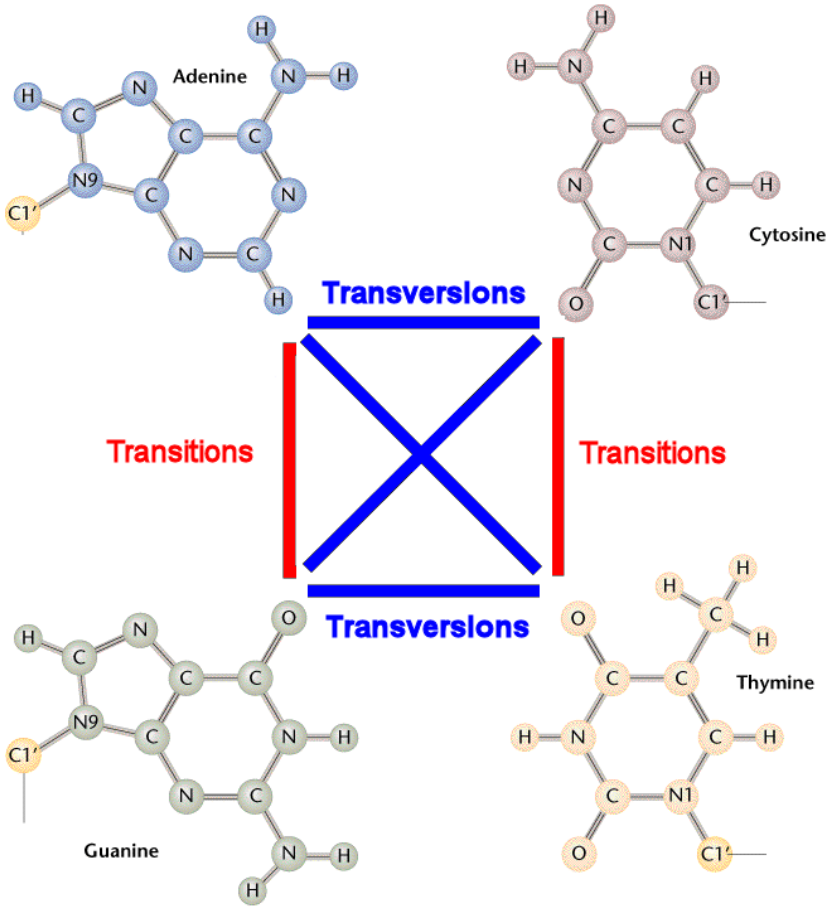
Modelo de Kimura

- Modelos de Kimura (1980) para nucleótidos
  - Asigna distintas probabilidades de sustitución
  - Modelo de dos parámetros (b): distinta probabilidad a transversión que a transición
    - Transición (α): cambio de purina a purina (o de pirimidina a pirimidina)
    - Transversión (β): cambio de purina a pirimidina (o viceversa)
  - Modelos más complejos (c): distintas probabilidades para cada sustitución



Modelo de Kimura

- Aunque hay cuatro tipos de transversiones y sólo dos de transiciones, por las propiedades químicas de las bases, la transición es mucho más común
  - Debido a la diferencia en anillos

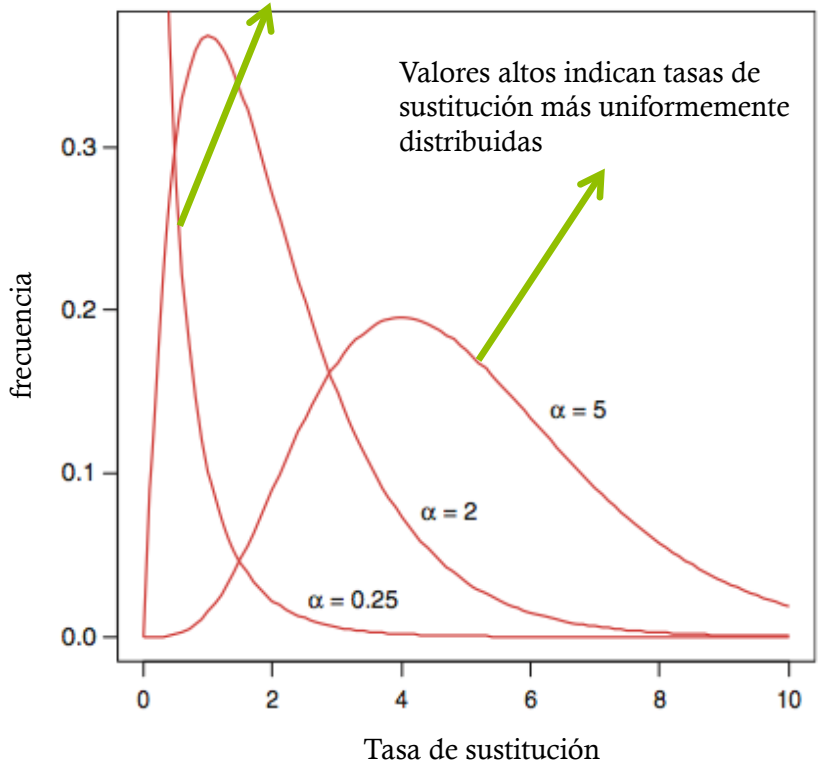


Modelo Gamma

- Algunas posiciones dentro de la proteína varían mucho y otras muy poco
  - La tercera posición de un codón suele tener una tasa de sustitución más alta que los dos primeros (código degenerado)
  - Algunas regiones de las proteínas tienen dominios conservados

- Para ello se asocia una tasa de sustitución distinta a cada posición, usando una distribución gamma
  - El parámetro α modula la forma de la distribución
  - Proteínas que evolucionan rápidamente tienen una α pequeña

Valores muy pequeños indican que casi todas las posiciones tienen la misma tasa de sustitución. Casi toda la variación se puede atribuir a unos pocos nucleótidos que varían mucho



Inferencia filogenética

- Existen varias aproximaciones para construir el árbol
  - Métodos basados en distancias
  - Métodos de maximización de la parsimonia
  - Métodos de maximización de la similitud
  - Inferencia bayesiana

  Métodos basados en caracteres

- Los métodos basados en distancias calculan la distancia entre secuencias completas para calcular el árbol
  - Descartan información sobre los residuos puntuales (caracteres)
- Los métodos basados en caracteres tienen esa información en cuenta
  - Aún así, a menudo ambos métodos generan árboles muy parecidos

Métodos de distancia

- Se calculan las distancias entre las secuencias, dos a dos
  - Generando una matriz de distancias
- Se van uniendo las secuencias con nodos internos según las distancias observadas
- Son métodos muy rápidos, particularmente útiles si tenemos un gran número de secuencias (>50)
- Veremos UPGMA y Neighbor-Joining

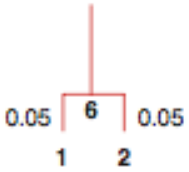
UPGMA

(a)

	1	2	3	4	5
1	—				
2	0.1	—			
3	0.8	0.8	—		
4	0.8	1	0.3	—	
5	0.9	0.9	0.3	0.2	—

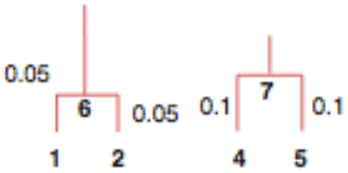
(b)

	(1,2)	3	4	5
(1,2)	—			
3	0.8	—		
4	0.9	0.3	—	
5	0.9	0.3	0.2	—



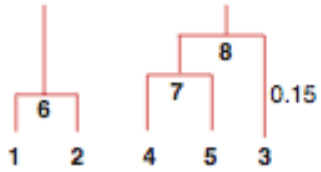
(c)

	(1,2)	3	(4,5)
(1,2)	—		
3	0.8	—	
(4,5)	0.9	0.3	—

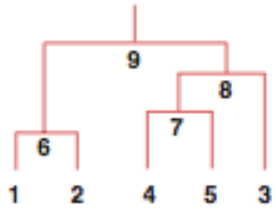


(d)

	(1,2)	[3,(4,5)]
(1,2)	—	
[3,(4,5)]	0.85	—



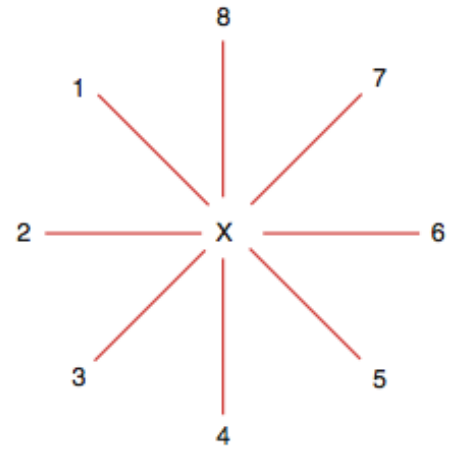
(e)

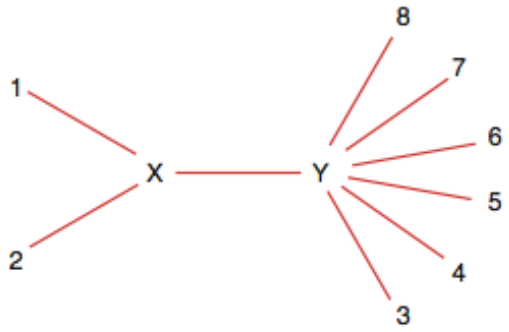


- Es un método sencillo que se basa en agrupar las secuencias más cercanas en base a su distancia
- El proceso es el siguiente:
  a) Calculamos la matriz de distancias, elegimos la menor distancia: $d_{1,2}$
  b) Unimos las secuencias 1 y 2, siendo la longitud de la rama la $0.5 \cdot d_{1,2}$. Calculamos las distancias al nuevo nodo (1,2) y seleccionamos la menor distancia ahora: $d_{4,5}$
  c) Calculamos las distancias al nuevo nodo (4,5), la longitud de las ramas, y elegimos de nuevo la menor: $d_{3,45}$
  d) Continuamos hasta terminar de unir nodos
- UPGMA asume que el reloj molecular de todos los nodos es igual
- Es un método muy utilizado en análisis de microarrays, pero para análisis filogenéticos suele ser bastante menos preciso que el método de Neighbor-Joining

Neighbor joining

- Se definen dos nodos como vecinos si existe un nodo interno X que los conecta directamente
  - Para N OTUs, podemos tener N-2 pares de nodos vecinos
- Método
  1. Comenzamos con todos los OTUs unidos directamente en un árbol de estrella (todos son vecinos)
  2. Se hacen las N(N-1)/2 comparaciones entre OTUs vecinos para determinar cuál es la pareja más cercana
  3. Esos OTUs se unen mediante un nuevo nodo interno y volvemos al paso dos, decrementando en 1 el valor de N
- El algoritmo minimiza la longitud de una rama en cada paso, así que no asegura una longitud mínima global



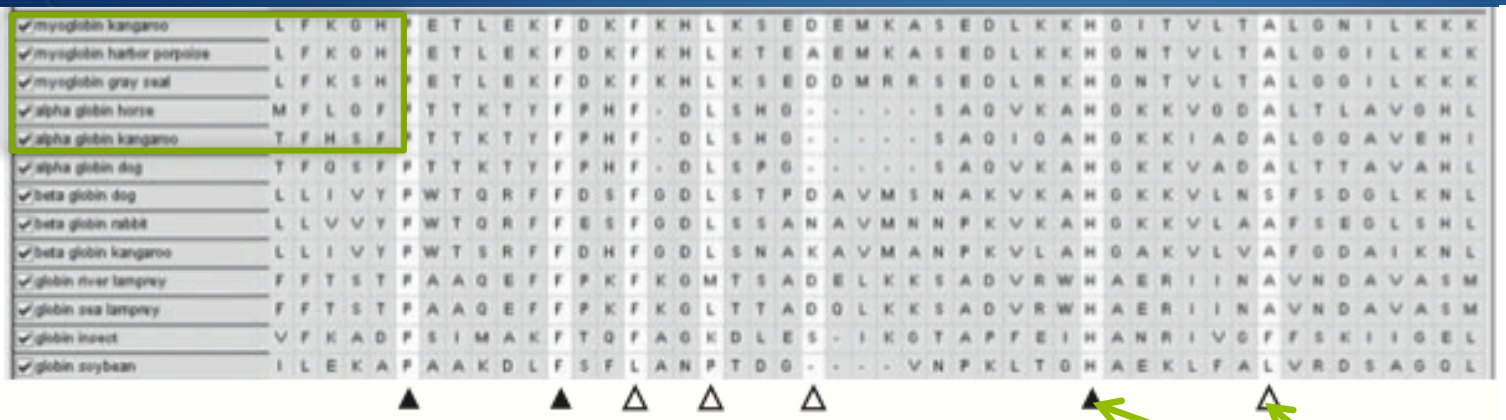


Máxima parsimonia

- *parsimonia.(Del lat. parsimonĭa).*
  - *1. f. Lentitud y sosiego en el modo de hablar o de obrar; flema, frialdad de ánimo.*
  - *→ 2. f. Frugalidad y moderación en los gastos.*
- Parte de la asunción de que el árbol que mejor explica las relaciones evolutivas es aquél que tiene las ramas más cortas a nivel global
  - El más simple de todos

Máxima parsimonia: método

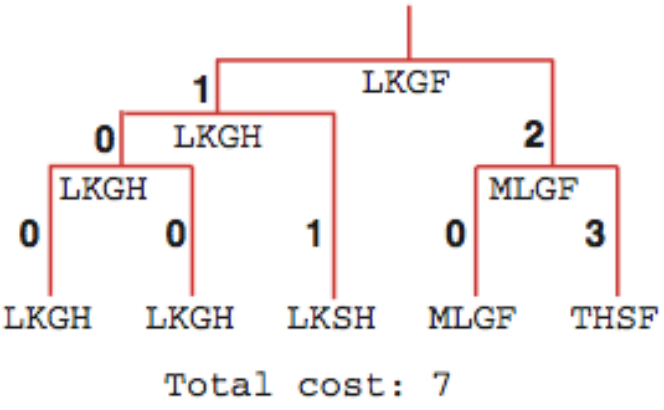
- Identificar residuos informativos
  - No son informativos aquellos que no tienen al menos dos nucleótidos distintos para dos o más secuencias
    - "informativo" significa que varía bastante en el alineamiento
- Se construyen árboles con distintas topologías.
  - Se les asigna un coste y se elige aquél de menor coste
  - Si hay muchos árboles se usan heurísticas para reducir la complejidad

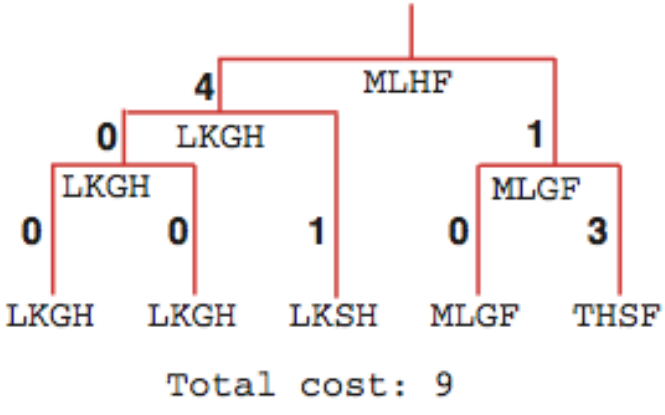


a) Los residuos con flecha se descartan por ser poco informativos (demasiado consenso)

kangaroo	LKGH
porpoise	LKGH
gray seal	LKSH
horse α	MLGF
kangaroo α	THSF

b) Tomemos un ejemplo con 4 de los 5 primeros aminoácidos para 5 secuencias





d) Construimos árboles a partir de posibles secuencias ancestrales, contando el número de cambios en cada rama. En este caso elegiríamos el de la izquierda

Máxima similitud

- Construye un árbol con una topología y longitud de ramas que maximiza la probabilidad de ser el generador de las secuencias observadas.
- Es uno de los métodos computacionalmente más costosos, pero también de los más flexibles
  - Permite variar el modelo entre distintas ramas o subfamilias, algo que los algoritmos de máxima parsimonia no hacen
  - De esta manera modelan mejor los casos en los que hay gran diferencia evolutiva entre distintas ramas

Máxima similitud

- Método de los cuartetos (Schmidt et al. 2002)
  - Para *n* secuencias, calculamos todas las topologías posibles de cuartetos de secuencias
    - Para cada cuarteto, habrá 3 topologías posibles
    - Para 12 secuencias, esto significa 495 cuartetos a probar

$$\binom{n}{4} = \binom{12}{4} = \frac{12!}{4!(12-4)!} = \frac{12!}{4!(8)!} = 495$$

    - Para cada cuarteto, se estima cuál de las tres topologías es mejor, y se le asigna
    - Los cuartetos se van ensamblando en el árbol final

Métodos Bayesianos

- Aproximación estadística basada en la teoría de Bayes
- Se calcula la probabilidad de que nuestro árbol sea correcto condicionada por los datos que tenemos: *P*(árbol | datos)
  - Lo contrario a otros métodos, que calculan la probabilidad de que nuestros datos se adapten al árbol: *P*(datos | árbol)
- Como en los de máxima probabilidad y máxima parsimonia, son métodos complejos y no entraremos en mayores detalles
  - Para más información, recurrir a:
    - Pevsner, 2009: Ch 7 *Molecular Phylogeny and Evolution*

Evaluación de los árboles

- Que un programa informático produzca un árbol filogenético no significa que sea correcto
  - Recordad GIGO (Garbage In, Garbage Out)
- En muchos casos puede ser globalmente correcto pero tener inexactitudes en algunas ramas
- Evaluación: bootstrapping o remuestreo
  - Verificación del significado biológico de un árbol evaluando su robustez

Bootstrapping (I)

- Primero, seleccionamos columnas del MSA original de forma aleatoria, hasta tener tantas como en el MSA original
  - Se permiten repeticiones (muestreo con reemplazamiento)
  - Es un alineamiento artificial, pero que conserva las características del MSA original
  - Se realizan muchos de estos muestreos aleatorios (100 a 1000)

Initial Alignment

Column	**1**	2	3	4	5	6	7	8	9
seq1	**A**	B	C	D	E	F	G	H	I
seq2	**A**	A	B	B	C	B	A	C	A
seq3	**C**	C	A	C	B	A	C	A	B

Bootstrap Alignment 1

1	**1**	8	**1**	2	5	**1**	8	2
A	**A**	H	**A**	B	E	**A**	H	B
A	**A**	C	**A**	B	C	**A**	C	A
C	**C**	A	**C**	C	B	**C**	A	C

Bootstrap Alignment 2

1	4	5	6	6	3	4	**1**	7
A	D	E	F	F	C	D	**A**	G
A	B	B	C	C	B	B	**A**	A
C	C	B	A	A	A	C	C	C

Bootstrapping (II)

- A cada MSA aleatorio se le aplica el algoritmo a evaluar, obteniendo un árbol
- Se construye un árbol de consenso con todos los árboles obtenidos
  - El porcentaje de veces que una ramificación aparece es el valor de bootstrap
    - Valores de bootstrap > 70% suelen tomarse como suficientemente robustos (equivalen a un nivel de significatividad $p < 0.05$)

Programas

- PAUP: Philogenetic Analysis Using Parsimony
  - Es el programa más usado de inferencia filogenética
  - A pesar de su nombre, permite inferencia mediante otros métodos
  - Es un programa de pago (http://paup.csit.fsu.edu/)
- MEGA: Molecular Evolutionary Genetic Analysis
  - http://www.megasoftware.net
  - Realiza MSAs e inferencia filogenética de muchos tipos
    - Distancia (UPGMA y NJ), máxima parsimonia y máxima similitud
- Tree-Puzzle: http://www.tree-puzzle.de/
  - Programa para inferencia por el método de máxima similitud
- MrBayes: http://mrbayes.csit.fsu.edu/
  - Programa para inferencia por el método de inferencia bayesiana

Resumen

- La filogenética molecular es clave para entender la evolución y las relaciones entre secuencias de aminoácidos o proteínas
- Un árbol filogenético es la representación gráfica de un alineamiento múltiple de secuencias
- Existen varios métodos para construir estos árboles (inferencia filogenética), basados en distancias, máxima parsimonia, máxima probabilidad e inferencia Bayesiana.
- Para cada una de estas aproximaciones existen herramientas que permiten construirlos. Al contrario que con el alineamiento, no son herramientas web si no de escritorio.
- A nivel de usuario, es vital partir de secuencias y alineamientos correctos. Es recomendable probar distintas herramientas y métodos de construcción de árboles
- Aún no hay consenso sobre cuál es el mejor método, ni datos de benchmarks, por ello nuestra capacidad de análisis crítico del árbol es esencial.

Ejercicio

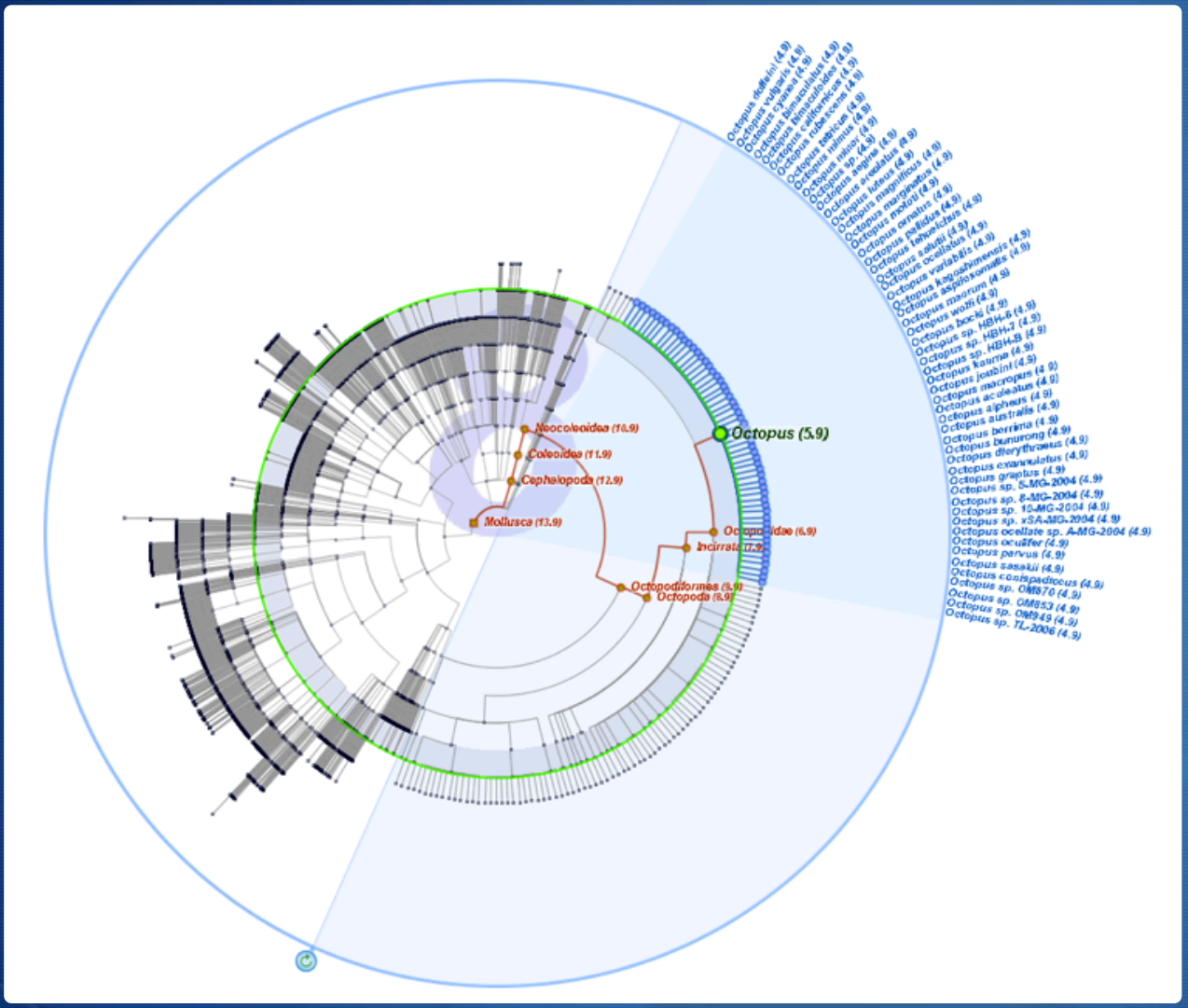
- Continuamos examinando nuestro gen “nuevo” y su “familia”, esta vez reconstruyendo su filogenia, mediante MEGA, como en el ejercicio anterior
  - Construid distintos árboles (parámetros, algoritmos, etc.)
  - Evaluadlos mediante bootstrapping
  - Comparadlos con el alineamiento múltiple
  - Tratad de extraer conclusiones sobre dominios conservados, indels …

Preguntas a debate

- Imagina un alineamiento que tiene una región claramente incorrecta. ¿Cuál es la consecuencia más probable de utilizar este alineamiento para inferir un árbol filogenético?
- ¿La teoría neutral (la mayoría de las sustituciones son neutrales) te parece compatible con las implicaciones de la teoría de Zuckerkandl y Pauling (las sustituciones se explican sobre todo debido a la selección natural)?

Lecturas adicionales

- Pevsner, 2009: Ch 7 *Molecular Phylogeny and Evolution*
- Dickerson R.E. *The cytochrome fold and the evolution of bacterial energy metabolism.* J Mol Evol 1: 26-45 (1971)
- Kimura, M. *Evolutionary rate at molecular level*. Nature 217: 624-626 (1968). PMID 5637732.



Treevolution es una herramienta para la visualización de árboles filogenéticos desarrollada en la Universidad de Salamanca

http://vis.usal.es/treevolution/