

Conceptos básicos de filogenética molecular

Dr. Eduardo A. RODRÍGUEZ TELLO

CINVESTAV-Tamaulipas

18 de julio del 2013



1 Conceptos básicos de filogenética molecular

- Introducción
- Evolución molecular y filogenética molecular
- Terminología
- Filogenia de genes y de especies
- Formas de representación de árboles
- Dificultad de encontrar un árbol real
- Procedimiento



1 Conceptos básicos de filogenética molecular

- **Introducción**

- Evolución molecular y filogenética molecular
- Terminología
- Filogenia de genes y de especies
- Formas de representación de árboles
- Dificultad de encontrar un árbol real
- Procedimiento



Cinvestav

Introducción

- Como hemos visto en las clases anteriores el análisis de secuencias biológicas está fundamentado en los principios de la evolución
- Las Similitudes y divergencias entre secuencias biológicas relacionadas (reveladas con alineamiento) a menudo requieren ser analizadas y visualizadas en el contexto de árboles filogenéticos
- Por esta razón la filogenética molecular es un área de gran importancia dentro de la bioinformática
- En esta clase introduciremos algunos conceptos básicos necesarios para comprender los métodos de construcción de árboles filogenéticos



1 Conceptos básicos de filogenética molecular

- Introducción
- **Evolución molecular y filogenética molecular**
- Terminología
- Filogenia de genes y de especies
- Formas de representación de árboles
- Dificultad de encontrar un árbol real
- Procedimiento



Evolución molecular y filogenética molecular

- Para iniciar a estudiar los conceptos de filogenética molecular necesitamos entender qué es evolución
- Generalmente se denomina *evolución* a cualquier proceso de cambio en el tiempo
- En el contexto biológico, la *evolución* es un cambio en el perfil genético de una población de individuos, que puede llevar a la aparición de nuevas especies, a la adaptación a distintos ambientes o a la aparición de novedades evolutivas



Evolución molecular y filogenética molecular

- La fuerza motriz de la evolución es la selección natural
 - Formas "no aptas" son eliminadas mediante cambios de las condiciones ambientales
 - Sólo se reproducen los más aptos debido a la selección sexual
- El mecanismo detrás de la evolución son las mutaciones genéticas que se producen espontáneamente
- Las mutaciones en el material genético proporcionan la diversidad biológica dentro de una población
- Estas determinan la viabilidad de los individuos para sobrevivir con éxito en un entorno determinado



Evolución molecular y filogenética molecular

- En otras palabras, la diversidad genética provee la fuente de materia prima para que la selección natural actúe
- La *filogenia* (del griego phylon: “tribu, raza” y geneá: “nacimiento, origen, procedencia”) es la determinación de la historia evolutiva de los organismos
- La *filogenética* es el estudio de la filogenia utilizando diagramas tipo árbol para representar los ancestros de esos organismos



Evolución molecular y filogenética molecular

- La filogenética puede ser estudiada de diversas maneras. A menudo ha sido estudiada utilizando registros fósiles, que contienen información sobre la morfología de los antepasados de especies actuales y la cronología de divergencias



Evolución molecular y filogenética molecular

- Sin embargo, el uso de registros fósiles tienen muchas limitaciones:
 - Pueden estar disponibles sólo para determinadas especies
 - Los datos existentes de fósiles puede estar fragmentados
 - La recolección de datos está limitada por la abundancia, hábitat, rango geográfico, y otros factores
 - Las descripciones de los rasgos morfológicos son a menudo ambiguas (múltiples factores genéticos)
- Por todo esto utilizar registros fósiles para determinar relaciones filogenéticas puede producir sesgos



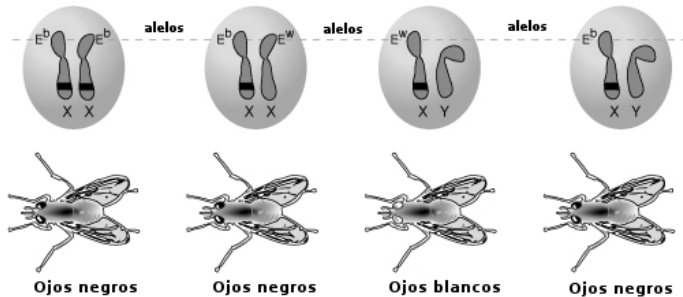
Evolución molecular y filogenética molecular

- Además los fósiles de microorganismos son prácticamente inexistentes, imposibilitando el uso de este enfoque
- Afortunadamente, los datos moleculares que están en la forma de secuencias de ADN o de proteínas pueden ser también muy útiles para proporcionar una perspectiva de la evolución de los organismos
- Esto se debe a que a medida en que los organismos evolucionan el material genético (genotipo) acumula las mutaciones que causan los cambios fenotípicos (expresión del genotipo, rasgos físicos y conductuales)



Evolución molecular y filogenética molecular

- Determinación del carácter color de ojos; genotipo(arriba), y fenotipo (abajo)



Evolución molecular y filogenética molecular

- Debido a que los genes son el medio para registrar las mutaciones acumuladas, éstos pueden servir como “fósiles moleculares”
- A través del análisis comparativo de secuencias de ADN de una serie de organismos relacionados, la historia evolutiva de los genes e incluso de los organismos puede ser revelada
- La ventaja de la utilización de datos moleculares es que son más numerosos que los registros fósiles y más fáciles de obtener
- Además no hay ningún sesgo de muestreo, como el que hay en los registros fósiles reales



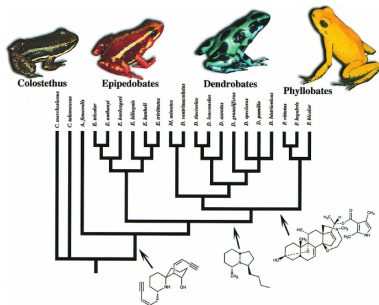
Evolución molecular y filogenética molecular

- Por lo tanto es posible construir árboles filogenéticos más precisos y robustos utilizando datos moleculares
- Debido a que ahora es posible contar con enormes cantidades de datos de secuencias moleculares la filogenética molecular ha tenido un rápido desarrollo



Evolución molecular y filogenética molecular

- En resumen la filogenética molecular es un área de gran importancia dentro de la bioinformática
- Estudia las mutaciones en las secuencias biológicas (ADN, ARN o proteínas) de una población de individuos con la finalidad de reconstruir su historia evolutiva usando árboles



- Relaciones filogenéticas entre 21 especies de ranas venenosas (Dendrobatidae) utilizando secuencias de ADN
- Revela los orígenes de algunas toxinas representativas

Figura tomada de: K.Summers and M. E. Clough, The evolution of coloration and toxicity in the poison frog family (Dendrobatidae), PNAS 2001 98 (11) 6227-6232

Evolución molecular y filogenética molecular

- Para utilizar los datos moleculares para reconstruir la historia de la evolución se requiere hacer una serie de suposiciones
 - Las secuencias moleculares utilizadas en la construcción filogenética son homólogas, i.e., comparten un origen común y posteriormente divergen al paso del tiempo
 - La divergencia filogenética se bifurca, i.e., una rama padre se divide en dos ramas hijas en cualquier punto dado
 - Cada posición en una secuencia evoluciona de manera independiente
 - La variabilidad entre secuencias es lo suficientemente informativa para construir árboles filogenéticos precisos



1 Conceptos básicos de filogenética molecular

- Introducción
- Evolución molecular y filogenética molecular
- **Terminología**
- Filogenia de genes y de especies
- Formas de representación de árboles
- Dificultad de encontrar un árbol real
- Procedimiento

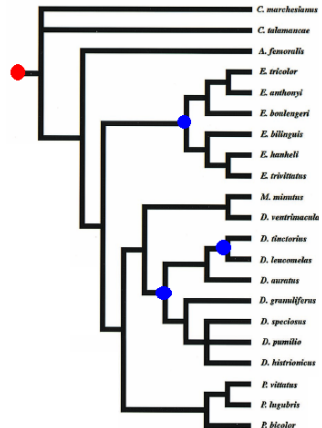


Cinvestav

Terminología

Árbol filogenético

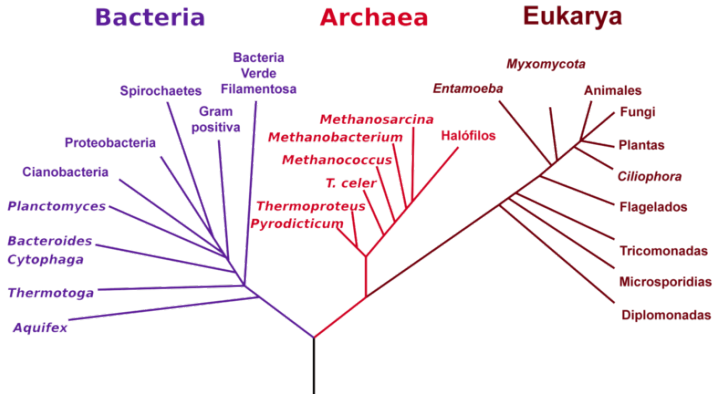
- Árboles generalmente binarios compuestos por raíz, nodos internos y hojas
- Las líneas son llamadas **ramas**
- En las hojas están representadas las secuencias de especies actuales, conocidas como **taxones**
- Los nodos internos representan un **ancestro común inferido** (evento de **especiación**)



Cinvestav

Terminología

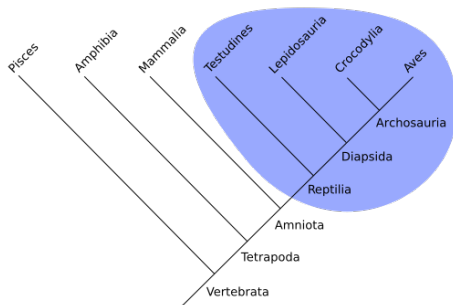
Árbol Filogenético de la Vida



Cinvestav

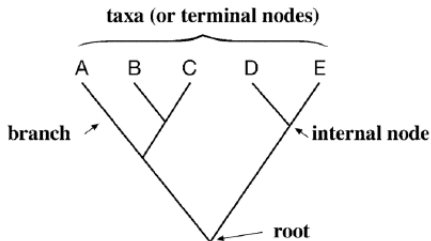
Terminología

- Un *clado* (griego klados: rama) o *grupo monofilético* es un grupo formado por un solo ancestro común y todos sus descendientes
- En un grupo monofilético, dos taxones son *hermanos* si comparten un ancestro común no compartido por ningún otro taxón (e.g., taxones crocodylia y Aves)



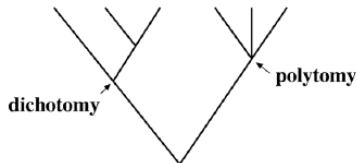
Terminología

- Una rama que indica una relación ancestro-descendiente en un árbol es un *linaje*
- Cuando un número de taxones comparten más de un ancestro común, no forman un clado, en este caso son referidos como *taxones parafiléticos* (e.g., taxones B, C, y D)



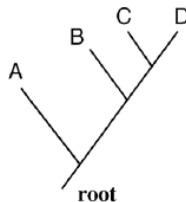
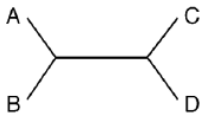
Terminología

- *Topología* es la forma en que se ramifica un árbol
- Cuando todas las ramas se bifurcan en un árbol filogenético, éstas son denominadas como una *dicotomía*
- Por el contrario si una rama tiene más de dos descendientes entonces se denomina multifurcación. El árbol con multifurcaciones se denomina *politomía*



Terminología

- Un *árbol filogenético sin raíz* no asume conocimiento de un ancestro común, sólo posiciones de los taxones para mostrar sus relaciones relativas (no hay dirección de un camino evolutivo)
- Para definir la dirección de la evolución se necesita un *árbol filogenético con raíz* donde todas las secuencias bajo estudio tienen un ancestro o nodo raíz común (más informativo)



Terminología

- El *reloj molecular* es una suposición por la cual las secuencias moleculares evolucionan a tasas constantes de manera que la cantidad de mutaciones acumuladas es proporcional al tiempo de evolución
- Así la longitud de las ramas del árbol pueden ser usadas para estimar el tiempo de divergencia
- Sin embargo, esta suposición raramente es válida en la realidad



1 Conceptos básicos de filogenética molecular

- Introducción
- Evolución molecular y filogenética molecular
- Terminología
- **Filogenia de genes y de especies**
- Formas de representación de árboles
- Dificultad de encontrar un árbol real
- Procedimiento



Cinvestav

Filogenia de genes y de especies

- Estrictamente, la filogenia de un gen (inferida de secuencias de genes o proteínas) sólo describe la evolución de ese gen (o proteína) particular codificado
- Esta secuencia puede evolucionar más o menos rápidamente que otros genes o puede tener una historia evolutiva diferente del resto del genoma debido a eventos horizontales de transferencia de genes
- La evolución de una secuencia particular no necesariamente corresponde con el camino evolutivo de las especies



Filogenia de genes y de especies

- La evolución de las especies es el resultado combinado de la evolución de múltiples genes
- En un árbol de especies, el punto de bifurcación en un nodo interno representa el evento de *especiación*
- Mientras que en un árbol de genes, un nodo interno indica un evento de *duplicación* de genes



Filogenia de genes y de especies

- Los dos eventos pueden o no coincidir
- Por esta razón, para obtener una filogenia de especies, es necesario construir árboles filogenéticos de una variedad de familias de genes para ofrecer una comparación global de la evolución de las especies



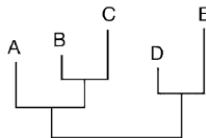
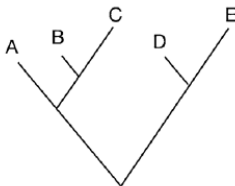
1 Conceptos básicos de filogenética molecular

- Introducción
- Evolución molecular y filogenética molecular
- Terminología
- Filogenia de genes y de especies
- **Formas de representación de árboles**
- Dificultad de encontrar un árbol real
- Procedimiento



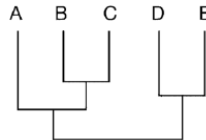
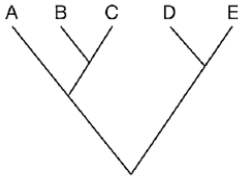
Formas de representación de árboles

- En un *filograma*, las longitudes de las ramas representan la cantidad de divergencia evolutiva (están a escala)
- Tienen la ventaja de mostrar tanto las relaciones evolutivas como la información sobre el tiempo relativo de divergencia de las ramas



Formas de representación de árboles

- En un *cladograma*, sin embargo, los taxones externos se alinean claramente
- Las longitudes de sus ramas no son proporcionales al número de cambios evolutivos y, por tanto, no tienen ningún significado filogenético



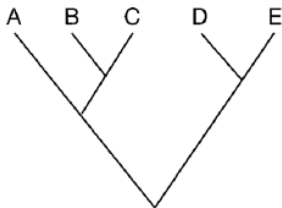
Formas de representación de árboles

- Para proporcionar información de topología de árbol a los programas de computadora sin tener que dibujar el árbol en sí, se desarrolló una representación especial en texto conocida como el *formato Newick*
- En este formato, los árboles están representados por los taxones incluidos entre paréntesis anidados
- En esta representación lineal, cada nodo interno está representado por un par de paréntesis que encierran todos los miembros de un grupo monofilético separados por una coma

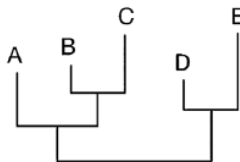


Formas de representación de árboles

- Para un filograma (ramas a escala), las longitudes de las ramas en unidades arbitrarias se colocan inmediatamente después del nombre del taxón separado por dos puntos



$((((B,C),A),(D,E)))$



$((((B:1,C:2),A:2),(D:1.2,E:2.5)))$

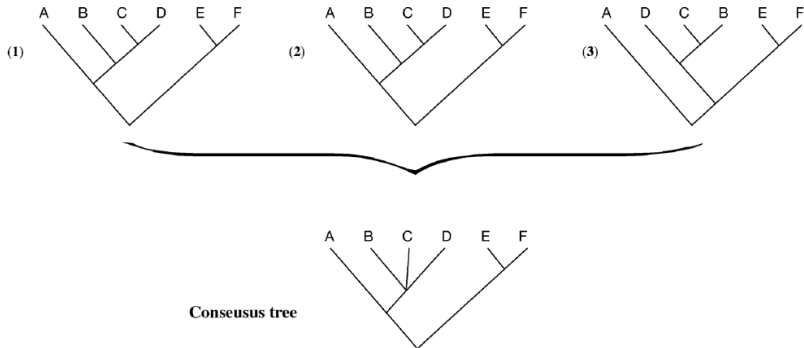
Formas de representación de árboles

- A veces, los métodos de construcción de árboles pueden producir varios árboles igualmente óptimos
- Un árbol de consenso puede ser construido mostrando las porciones de bifurcación resueltas comúnmente y colapsando aquellas que no concuerdan entre los árboles
- La combinación de los nodos se puede hacer ya sea por consenso o por la estricta regla de la mayoría
- En un árbol de consenso estricto, todos los nodos en conflicto son colapsados
- En un árbol de consenso por regla de la mayoría, los nodos en conflicto que están de acuerdo en más del 50 % de los nodos se mantienen, el resto son colapsados en una multifurcación



Formas de representación de árboles

- Un árbol de concenso es derivado de 3 árboles (regla de mayoría). Los nodos en conflicto se representan con una multifurcación



1 Conceptos básicos de filogenética molecular

- Introducción
- Evolución molecular y filogenética molecular
- Terminología
- Filogenia de genes y de especies
- Formas de representación de árboles
- **Dificultad de encontrar un árbol real**
- Procedimiento



Dificultad de encontrar un árbol real

- El objetivo de la filogenética molecular es reconstruir correctamente la historia evolucionaria basada en la divergencia de secuencias observadas entre organismos
- Esto significa encontrar una topología de árbol correcta con longitudes correctas de las ramas
- Esta tarea puede ser extremadamente difícil y computacionalmente demandante



Dificultad de encontrar un árbol real

- La razón es que el número potencial de topologías de árboles puede enorme aún con un número moderado de taxones
- El número de árboles con raíz (N_R) para n taxones está dado por la siguiente fórmula:

$$N_R = (2n - 3)! / 2^{n-2} (n - 2)! \quad (1)$$

- El número de topologías para árboles sin raíz (N_U) es:

$$N_U = (2n - 5)! / 2^{n-3} (n - 3)! \quad (2)$$

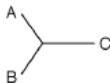
- Para 10 taxones, puede haber 2,027,025 árboles sin raíz y 34,459,425 con raíz



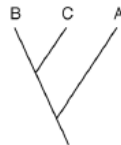
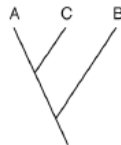
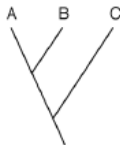
Dificultad de encontrar un árbol real

3 taxa

Unrooted

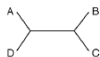
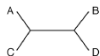
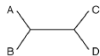


Rooted



Dificultad de encontrar un árbol real

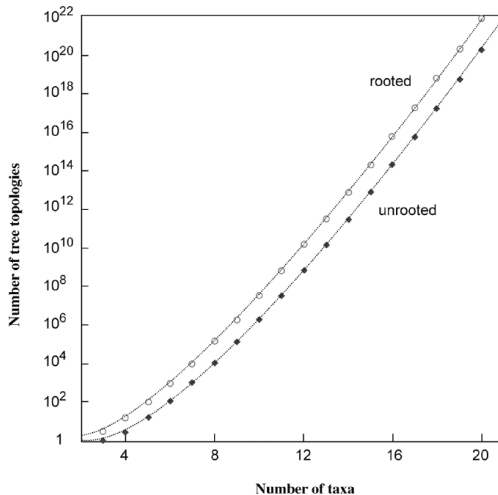
4 taxa



Cinvestav

Dificultad de encontrar un árbol real

- Eje y en escala log



1 Conceptos básicos de filogenética molecular

- Introducción
- Evolución molecular y filogenética molecular
- Terminología
- Filogenia de genes y de especies
- Formas de representación de árboles
- Dificultad de encontrar un árbol real
- **Procedimiento**

Procedimiento

- El procedimiento para construir árboles filogenéticos se divide en 5 pasos:
 - 1 Elección de los marcadores moleculares
 - 2 Alineamiento múltiple de secuencias
 - 3 Elección de un modelo de evolución
 - 4 Determinación de un método de construcción de árboles
 - 5 Verificación de la fiabilidad del árbol construido

