

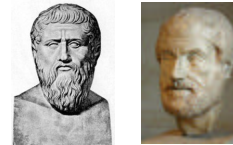
Biología Genómica y Evolución IV – Inferencia Filogenética y Evolución Molecular

Pablo Vinuesa (vinuesa@ccg.unam.mx)
Programa de Ingeniería Genómica, CCG-UNAM, México
<http://www.ccg.unam.mx/~vinuesa/index.html>
Profesor Asistente: Agustín Avila : aavila@icg.unam.mx

Tema 1: Conceptos básicos de evolución molecular y filogenética

1. Breve introducción histórica del desarrollo de la biología evolutiva y de su impacto en la sociedad occidental y en nuestra percepción del mundo
2. ¿Porqué estudiar filogenética y evolución molecular?
- 3.- Tipos de datos usados en filogenética y evolución molecular
- 4.- Protocolo básico para el análisis filogenético de secuencias moleculares
- 5.- Tasas de evolución y selección de marcadores
- 6.- Qué es un árbol filogenético y tipos de árboles : una visión del "bosque"
- 7.- Definición de homología y tipos de homología
- 8.- Presentación de los tipos de métodos de reconstrucción filogenética

Tema 1: Breve historia de la biología evolutiva y filogenética



Platón (izda.) y Aristóteles (dcha.)

1. Antes de Darwin

- La filosofía y visión occidental del mundo estuvieron dominadas durante muchos siglos por las ideas del gran filósofo griego **Platón** (428-348 AC) y su discípulo **Aristóteles** (348-322 AC).
- Uno de los pilares de la filosofía platónica es el concepto de la "eidos", la forma o la idea, una forma transcendente ideal imitada de manera imperfecta por sus encarnaciones terrenales. En el contexto de esta **filosofía esencialista** la **variación o diversidad biológica se interpretaba como imperfección accidental**. Aristóteles desarrolló el concepto platónico de inmutabilidad de las esencias proponiendo el concepto de que las especies tienen propiedades fijas.

Tema 1: Breve historia del desarrollo de la biología evolutiva y filogenética

1. Antes de Darwin



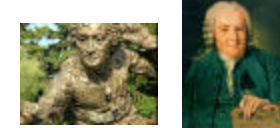
El jardín del Edén, por Jan Breughel (1568-1625).

- Más tarde los **cristianos** interpretaron literalmente el Génesis, concluyendo que cada especie fue creada individualmente "a imagen y semejanza" del Creador, en la forma en la que las conocemos hoy. A esta creencia se la conoce como "**creación especial**".
- En la creencia y filosofía cristiana se asume que el orden es un estado superior al desorden y que por lo tanto las creaciones divinas siguen un plan: una gradación que va desde objetos inanimados y formas "minimamente animadas", progresando a través de plantas e invertebrados hacia formas cada vez "más elevadas" de vida.

- Esta **scala naturae** culmina con el humano, la única criatura con naturaleza dual: física y espiritual.
- El cristianismo asume que la escala natural es permanente e inmutable, ya que la posibilidad de cambio implicaría la existencia de imperfección en la creación original.

Tema 1: Breve historia de la biología evolutiva y filogenética

Los naturalistas han tratado de **detectar, describir y explicar la diversidad biológica** desde muchos siglos atrás. Este es el **objetivo de la sistemática**.



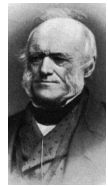
En 1758 **Carl von Linné** (Carolus Linnaeus) formalizó un sistema jerárquico de nomenclatura para clasificar a los organismos (**sistema binomial**) en su **Systema Naturae** (1735). Esta jerarquía fue concebida independientemente de la teoría evolutiva y pretendía revelar el **orden natural** o patrón implícito en el plan del Creador.

De hecho hasta el S. XVIII el papel de las ciencias naturales era catalogar y hacer manifiesto el plan del Creador para que pudiéramos apreciar su sabiduría infinita.

Tema 1: Breve historia de la biología evolutiva y filogenética



La interpretación bíblica literal comenzó a ser cuestionada y reemplazada cuando en el **S. XVII** comenzaron a surgir visiones más materialistas del mundo natural, notablemente con los **Principia Mathematica** de **Isaac Newton** (1643-1727) que describen las leyes de la gravitación universal y las leyes del movimiento, sentando las bases de la mecánica clásica. Fue el primero en demostrar que estas leyes del movimiento y gravitación aplican igual a la Tierra como al resto del universo. El poder unificador y predictivo de sus leyes fueron decisivos para permitir la culminación de la **Revolución Científica** iniciada en 1543 por **Nicolaus Copernicus** con su *De revolutionibus orbium coelestium*, sentando así pilares fundamentales de la ciencia moderna.



Fueron los descubrimientos en astronomía y geología hechos en los siglos **XVIII** y **XIX** los que sentaron las bases decisivas para el nacimiento de las ideas evolutivas. Particularmente importante fue el trabajo del geólogo escocés **Charles Lyell** (1797-1875), quien expuso el **principio del uniformitarismo**. Este se basa en reconocer que los mismos procesos geológicos reconocidos en el presente operaban en el pasado, y que por lo tanto las formaciones geológicas se pueden explicar por causas observadas en el presente.

Lyell tuvo una gran influencia en el pensamiento de Charles Darwin, quien adoptó el uniformitarismo en su pensamiento evolutivo.

Tema 1: Breve historia de la biología evolutiva y filogenética

En el **S. XVIII** varios naturalistas y filósofos franceses sugirieron que las especies biológicas habían surgido por causas naturales.



Evolucionistas tempranos como **George Louis Buffon** (1753) ya se oponían al sistema linneano y al esencialismo aristotélico en el que se basaba el orden natural lineano.



La hipótesis evolutiva pre-darwiniana más significativa fue sin duda la formulada por **Chevalier de Lamarck** en su *Philosophie Zoologique* (1809).

La teoría lamarckiana proponía que cada especie se originaba por medio de la generación espontánea a partir de materia inanimada, comenzando así en la base de la **"cadena del ser"**. El suponía que dentro de cada especie actuaba un "fluido nervioso" que la hacía progresar ascendiendo la cadena.

Asumía que las alteraciones corporales adquiridas durante el lapso de vida de un individuo se heredaban, sentando las bases del principio conocido como **"herencia de caracteres adquiridos"**.

Tema 1: Breve historia de la biología evolutiva y filogenética



Charles Darwin, 1854



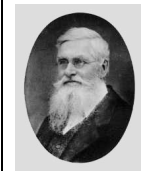
Charles Robert Darwin (1809-1882) fue un naturalista inglés que comenzó a estudiar medicina y luego la carrera de clérigo en la Univ. de Cambridge (UK) antes de embarcarse en un viaje alrededor del mundo a bordo de la **H.S.M. Beagle** (1831-36), que la marina británica enviaba para cartografiar las costas de Sudamérica. El fue invitado como naturalista y acompañante del Capitán Fitz Roy. Este viaje no sólo cambiaría la vida de Darwin, sino que revolucionó radicalmente el pensamiento occidental al presentar evidencia contundente de que las especies biológicas han evolucionado a partir de ancestros comunes.

Darwin se convenció de ello durante el viaje del Beagle. Consciente de la importancia y repercusiones de su descubrimiento, pasó 20 años recopilando evidencias antes de hacer públicas sus ideas.

El **1 de Julio de 1858** **Charles Darwin** propuso su teoría de la **evolución por medio de la selección natural** en una reunión de la Sociedad Lineana de Londres. Su tratado monumental **"EL origen de las especies"** fue publicado un año después. El día de su publicación el libro quedó agotado.

Su libro no sólo revolucionó el pensamiento biológico, sino también la política, sociología y la filosofía moral de occidente.

Tema 1: Breve historia de la biología evolutiva y filogenética



Darwin escribió una serie de ensayos privados en 1844 y en 1856 comenzó un libro que quería titular Natural Selection. Nunca lo terminó ya que en 1858 Darwin (dcha.) recibió un manuscrito de un joven naturalista, **Alfred Russel Wallace** (1823-1913) en el que quedaba claro que había descubierto de manera independiente el principio de la selección natural. Esta carta fue la que lo motivó a presentar su famosa plática ante la Soc. Linneana de Londres en 1858, basada en partes de sus ensayos y en la carta enviada por Wallace. Seguidamente pasó a escribir un "resumen o abstract" de 490 páginas de su manuscrito Natural Selection, el cual fue publicado en **1859** en Londres bajo el título de:

On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life



El punto más débil en la teoría Darwiniana era su incapacidad de explicar cómo surgían y cómo se transferían los caracteres hereditarios de generación en generación.

Tema I: Breve historia de la biología evolutiva y filogenética

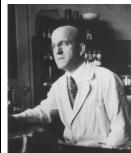


Siete años después de la publicación del Origen de las Especies (1866), el monje agustino austriaco **Gregor Mendel** describe sus trabajos con cruza de arvejas titulado "Versuche über Pflanzenhybriden", en el que **establece la existencia de "caracteres elementales de la herencia" y describe leyes estadísticas que gobiernan su transmisión de generación en generación**. El no sabía si estos caracteres elementales tenían una base material o meramente representaban "interacciones vitales"



En 1869, **Johann Friedrich Miescher**, un químico suizo, **descubre** una sustancia ácida rica en P en células espermáticas, a la que llamó **nucleína**. Llegó a especular sobre su posible relación con la herencia, pero pronto abandonó esta "idea absurda".

Darwin, Mendel y Miescher no podían imaginar lo conectados que sus respectivos descubrimientos estarían 100 años después ...



En 1944 **Oswald Avery** demuestra que el DNA es el material hereditario (y no las proteínas como se pensaba). Incluso llega a especular que los genes están hechos de DNA.

<http://profiles.nlm.nih.gov/CC/Views/Exhibit/narrative/biographical.html>
Avery, Oswald T., Colin M. MacLeod, and Maclyn McCarty. [Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types](#). J. Exp. Med. 79, 2 (February 1, 1944): 137-158.

Tema I: Breve historia de la biología evolutiva y filogenética



En las décadas de 1930-40 Ronald A. Fisher, JBS Haldane y Sewall Wright sientan las bases de la teoría de genética de poblaciones, desarrollando además muchos de los índices estadísticos usados para estimar la intensidad de fuerzas evolutivas como la deriva, selección, mutación y migración. Nace la **Síntesis Evolutiva o Síntesis Moderna**. En esencia clarifica que la mutación y selección natural actúan conjuntamente para causar evolución adaptativa.

Entre 1930 y 1950 **T. Dobzhansky, G. Simpson, E. Mayr y G. Stebbins** fusionan las teorías propias de las disciplinas de genética, paleontología, zoología y botánica bajo la luz de la teoría evolutiva en lo que se vino a conocer como **la nueva síntesis**.



Theodosius Dobzhansky



George Simpson



Ernst Mayr



G. Ledyard Stebbins

Tema I: Breve historia del desarrollo de la biología evolutiva y filogenética

Desafortunadamente, muy poco de estos avances fundamentales en el desarrollo moderno de la teoría evolutiva salió a la luz pública en USA entre los 1940-60 por la presión que ejercían los "creacionistas" en sobre los editores de libros de texto, obligándolos a evitar cualquier mención de evolución en libros de biología. Por miedo a perder negocio, las editoriales cedieron a estas presiones.



No fue hasta los 60's que se revirtió el problema. En parte debido al éxito de la ciencia soviética por el lanzamiento del primer satélite Sputnik en 1957. Ello creó una oleada de pánico a través de sectores de la ciencia americana, incluida la biología evolutiva. En 1967, la legislatura del estado de Tennessee vuelve a permitir la mención de evolución.



Algunos **creacionistas afirman** que el diluvio universal fue el responsable de la aparición de todos los fósiles y formaciones geológicas sobre la tierra. Afirman que **la tierra se formó en exactamente 6 días (un 23 de Octubre del 4004 adC, según estimó el obispo James Ussher, SXVII)**. Los geólogos abandonaron esta visión hace más de 300 años. Se trata de cristianos fundamentalistas que hacen una lectura literal del Génesis para explicar la historia de la vida en el planeta Tierra.

La cronología de la creación y otros eventos importantes recopilados en la "historia bíblica" de la Tierra, según el Arzobispo anglicano de Armagh (Irlanda), James Ussher

— J. Ussher, *The Annals of the World* iv (1658)

Fuentes: http://en.wikipedia.org/wiki/Ussher_chronology
<http://www.lhup.edu/~dsimanek/usshe.htm>

Ussher's history of the Earth:

- 4004 BC - [Creation](#) (23 de Octubre, Domingo)
- 4004 BC - Adam and Eve were driven from Paradise on Monday 10 November
- 2348 BC - [Noah's Flood](#); the ark touched down on Mt Ararat on 5 May 2348 BC (Wed)
- 1921 BC - God's call to [Abraham](#)
- 1491 BC - [The Exodus](#) from [Egypt](#)
- 1012 BC - The founding of the [Temple in Jerusalem](#)
- 586 BC - The destruction of [Jerusalem](#) by [Babylon](#) and the beginning of the [Babylonian Captivity](#)
- 4 BC - The birth of [Jesus](#)

Tema 1: Conceptos básicos de evolución molecular y filogenética

Genómica Evolutiva I, Licenciatura de Ciencias Genómicas - UNAM, México. Semestre 2009-1

Tema 1: Breve historia del desarrollo de la biología evolutiva y filogenética



En 1953 **Francis Crick** y **James Watson** publican en Nature su modelo de la estructura en **doble hélice del DNA**. Se abre con ello la vía para el nacimiento de la evolución molecular.



Emile Zuckerkandl fue junto con Linus Pauling uno de los pioneros de la disciplina de la **evolución molecular**, al descubrir que las moléculas de **DNA y las proteínas** que codifican son **"documentos de la historia evolutiva"** dada la relativa constancia con la que acumulan variaciones (mutaciones). Ambos publican un paper histórico en 1965 proponiendo formalmente la existencia de un **reloj molecular**.



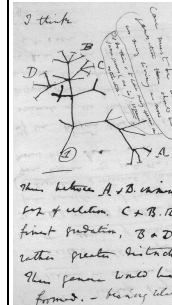
A finales de los 60's el genetista matemático **Motoo Kimura** combina los principios teóricos de la genética de poblaciones con los nuevos conocimientos de biología molecular para desarrollar la **teoría neutral de evolución molecular** (1968), en la que la deriva genética es la principal fuerza evolutiva que afecta al cambio en las frecuencias alélicas.

Tema 1: Breve historia del desarrollo de la biología evolutiva y filogenética



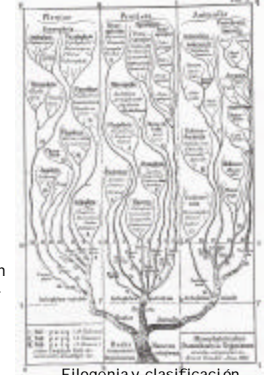
Los primeros intentos de reconstruir la historia filogenética estaban basados en pocos o ningún criterio objetivo.

Reflejaban las ideas o hipótesis plausibles generadas por expertos de grupos taxonómicos particulares.



La mayor parte de la 1a. mitad del SXX los sistemáticos estaban más preocupados por el problema de definir a las especies biológicas, descubrir mecanismos de especiación y la variación geográfica de las especies, que en entender su filogenia.

No fue hasta los 40 's y 50's que los esfuerzos de individuos como **Walter Zimmermann** y **Willi Henning** comenzaron a definir métodos objetivos para reconstruir filogenias en base a caracteres compartidos entre organismos fósiles y contemporáneos.



Filogenia y clasificación de la vida tal y como la propuso **Ernst von Haeckel** en 1866



Tema 1: Breve historia del desarrollo de la biología evolutiva y filogenética



Robert R. Sokal

En la década de los 60's estos métodos objetivos para estimar filogenias se refinan y se desarrollan criterios explícitos como el de la taxonomía numérica (**Peter Sneath** y **Robert R. Sokal**, 1963). Desarrollan métodos de agrupamiento basados en matrices de distancias como el UPGMA

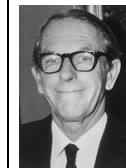


Luca Cavalli-Sforza (parado) y **Anthony Edwards** colaboraron en la década de los 60's en la Universidad de Pavia, Italia, para fundar la disciplina de filogenética numérica, concibiéndola como un problema de inferencia estadística. Introdujeron los **métodos de parsimonia, verosimilitud y de matrices de distancias** para inferir filogenias.

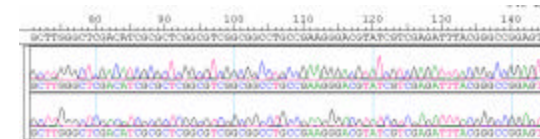
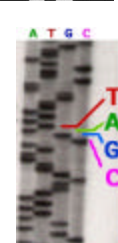


Tema 1: Breve historia del desarrollo de la biología evolutiva y filogenética

A la par que se acumulaban nuevas secuencias de proteínas y se desarrollaban métodos estadísticos para inferir filogenias y patrones de evolución molecular, la biología molecular desarrolla **métodos cada vez más eficientes de secuenciación** de proteínas y ácidos nucleicos



Frederick Sanger desarrolla técnicas muy eficientes de secuenciación de proteínas (1955) y DNA (dideoxy chain termination, 1975). Recibe **dos premios nobel en química**.



Tema 1: Conceptos básicos de evolución molecular y filogenética

Genómica Evolutiva I, Licenciatura de Ciencias Genómicas - UNAM, México. Semestre 2009-1

Tema 1: Breve historia del desarrollo de la biología evolutiva y filogenética



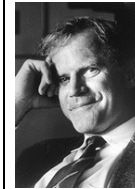
En los 60's se contaba con unas reducida base de datos de secuencias de proteínas globulares. **Margaret Dayhoff** fue la primera en comenzar a construir una base de datos de secuencias y de inferir propiedades estadísticas del proceso de sustitución. Describen las **matrices PAM**. Además introduce métodos computacionales de parsimonia para calcular árboles de secuencias de proteínas.



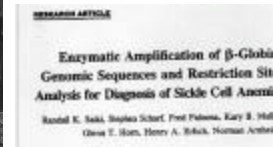
En 1967 **Walter Fitch** y **Emanuel Margoliash** publican una filogenia de proteínas de citocromo C basada en un método de matrices de distancias y un **algoritmo para su representación en forma de un árbol filogenético** usando el métodos de los mínimos cuadrados.

Walter Fitch

Tema 1: Breve historia del desarrollo de la biología evolutiva y filogenética

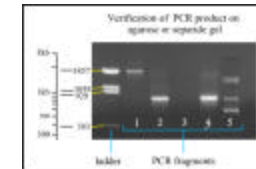
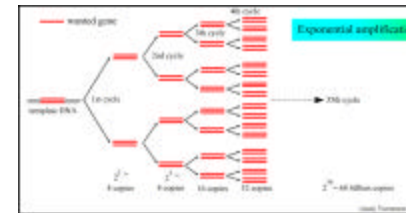


En 1993 **Kary B. Mullis** gana el premio Nobel de bioquímica por su método de amplificación enzimática de DNA (**PCR**).



Los protocolos de PCR disparan la generación de secuencias y el desarrollo de nuevas metodologías, tales como la generación de fingerprints genómicos de diversos tipos, tales como AFLPs, PCR-RFLPs, RAPDs ...

Science, 1985 Dec 20, 230(4732):1350-4.



Nuevos métodos de secuenciación se basan en tecnología de PCR.

¿Qué estudian la sistemática y evolución molecular? : conceptos básicos

- La **sistemática y evolución molecular** engloban a un conjunto muy amplio de herramientas y modelos bio-estadísticos que nos permiten estudiar la "arqueología molecular" de los organismos, es decir, el **registro evolutivo escrito en el material hereditario** (DNA) y en sus productos (proteínas y RNAs estructurales, anatomía etc.), con el fin de poder hacer inferencias sobre:
 1. Las relaciones ancestro-descendiente (**filogenéticas**) de dominios de prot., genes y organismos. Las hipótesis filogenéticas resultantes son la base para hacer **predicciones** (inferencias) sobre propiedades biológicas de los grupos revelados por la filogenia mediante el mapeo de caracteres sobre la **topología** (hipótesis evolutiva).
 2. Estimar la intensidad y papel de las fuerzas evolutivas (selección, deriva, migración, recombinación) en el modelaje de la estructura de dominios proteicos, genes, poblaciones y genomas (**evolución molecular**).

La **sistemática molecular** usa **marcadores genéticos** para hacer inferencias sobre procesos que acontecen en las poblaciones y para reconstruir su filogenia. Así se generan grandes bases de datos de secuencias de genes específicos para una gran cantidad de especies y organismos.

Los estudios de evolución molecular usan estos sets de datos para evaluar tasas, procesos y constricciones en el cambio molecular a lo largo del tiempo. Los resultados de estos estudios de evolución molecular a su vez proveen de nuevos criterios para la selección más informada de marcadores moleculares para estudios filogenéticos y de genética de poblaciones.

¿Porqué estudiar filogenética y evolución molecular? : conceptos básicos

- Buena parte de la **bioinformática y genómica** es esencialmente **biología comparada** predicciones basadas en comparación de motivos, secuencias, genomas y estructuras moleculares
- **Los organismos y sus componentes estructurales tienen una historia evolutiva => la filogenética yace en el corazón de la biología comparada**
- Comparaciones basadas en filogenias moleculares pueden proporcionar importantes descubrimientos difíciles de hacer mediante otras aproximaciones:
 - elucidación del árbol (red?) universal
 - distinción entre ortólogos, parálogos y xenólogos y descubrimiento de funcionalidades diferenciadas entre ellos; clasificación y anotación de familias (multi)genéticas
 - guía de alineamientos múltiples de nt y aa;
 - guía para modelajes estructurales de RNAs y proteínas
 - ...

La relación entre filogenética y evolución molecular:

- La **filogenética** tiene por objetivo el **trazar la relación ancestro descendiente de los organismos** (árbol filogenético) a diferentes niveles taxonómicos, incluyendo el árbol universal, haciendo una reconstrucción de esta relación en base a diversos **caracteres homólogos** (adquiridos por descendencia directa), tanto morfológicos como moleculares
- La **evolución molecular estudia los mecanismos y procesos** que han llevado a la formación de dichos caracteres, desde el nivel de posiciones de un codón hasta la organización y estructura genómica y anatómica de un organismo, en un marco de biología comparada en contextos tanto de poblaciones (**microevolución**) como de linajes. Para ello requiere de la hipótesis evolutiva de relaciones entre entidades revelada por una filogenia
- Gracias a la cantidad masiva de secuencias disponibles en las bases de datos (incluyendo centenas de genomas completos) y la disponibilidad de sofisticados modelos de evolución de secuencias y de su implementación en programas de cómputo muy eficientes, **las filogenias moleculares son indispensables para examinar todo tipo de cuestiones evolutivas**.
- El desarrollo de métodos de simulación de secuencias y rigurosos marcos de filogenética estadística, tanto frecuentistas y Bayesianos, permiten hacer **contrastos de hipótesis en un contexto evolutivo**!

Refs:

Huelsenbeck, J.P., Rannala, B., 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* 276, 227-232.
Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294, 2310-2314

¿Porqué estudiar filogenética y evolución molecular?

Corolario I:

"Nothing in biology makes sense except in the light of evolution"

- Theodosius Dobzhanski, 1973
(*The American Biology Teacher* 35:125)

Corolario II:

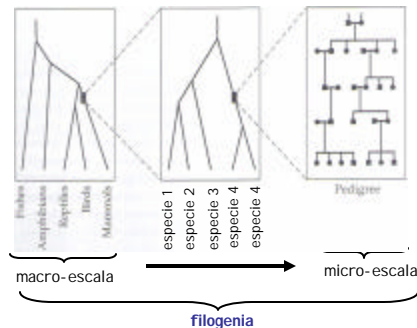
"Nothing in evolutionary biology makes sense except in the light of a phylogeny"

- Jeff Palmer, Douglas Soltis, Mark Chase, 2004
(*American J. Botany* 91: 1437-1445)



El concepto de filogenia y homología: definiciones básicas

"The stream of heredity makes phylogeny; in a sense, it is phylogeny. Complete genetic analysis would provide the most priceless data for the mapping of this stream".
G.G. Simpson (1945)



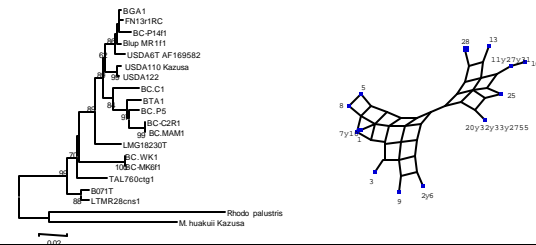
Filogenia: historia evolutiva del flujo hereditario a distintos niveles evolutivos/temporales, desde la genealogía de genes en poblaciones (micro-escala; dominio de la genética de poblaciones) hasta el árbol universal (macro-escala)

El concepto de filogenia y homología: definiciones básicas

El proceso evolutivo con sus eventos de **especiación** ha involucrado la sucesiva ramificación y algunas anastomosis de los linajes hereditarios. Los organismos actuales son el producto de esta historia y repesetan la puntas o **nodos terminales** del árbol de la vida.

El entendimiento completo de una filogenia requiere del conocimiento del patrón u orden de ramificación (**cladogénesis**) y longitudes de cada rama (**cambios anagenéticos** dentro de los linajes a lo largo del tiempo).

Eventos de transferencia horizontal de DNA entre ramas del árbol (**evolución reticulada**) son el resultado de **hibridaciones interespecíficas** mediadas por diversos mecanismos. El producto de la hibridación puede resultar en genomas 100% híbridos (eucariontes) o sólo afectar a determinados loci (**mosaicos genómicos** de los procariontes).



El concepto de filogenia y homología: definiciones básicas

Homología: es la relación entre dos caracteres que han descendido generalmente con modificación, de un ancestro común. Estrictamente se refiere a ancestría común inferida.

Analogía: es la relación existente entre dos caracteres cuando éstos, aun siendo similares, han descendido convergentemente a partir de caracteres ancestrales no relacionados en términos genealógicos.

Cenacestro: del inglés (cenacestor), es el ancestro común más reciente de los taxa bajo consideración.

El concepto de homología: definiciones básicas

Dado que filogenia es "el flujo de la herencia", sólo los caracteres genéticos o heredables son informativos desde una perspectiva genealógica.

Caracteres y estados de carácter. Los evolucionistas distinguen entre caracteres, como por ejemplo los amino ácidos, y sus estados de carácter, como pueden ser gly o trp. La homología reside en los caracteres, no en sus estados !!!

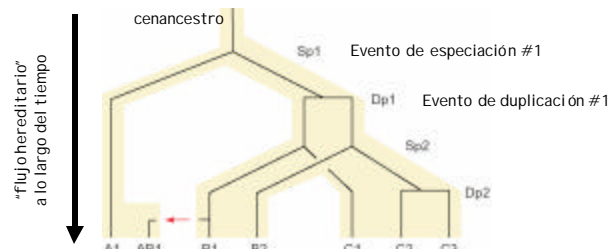
El reconocimiento de la condición de homología entre caracteres. La homología no es una cualidad cuantitativa. Sólo hay dos condiciones posibles: ser o no homólogo. No se es más o menos homólogo. Es como el embarazo. Se está o no se está en dicho estado.

Por tanto, para cuantificar el parecido entre un par de secuencias homólogas se dice que presentan globalmente un 70% y 95% de **identidad y similitud**, respectivamente (no existe algo como 95% de homología).

El concepto de homología es simplemente una abstracción sobre la relación entre caracteres, sobre su ascendencia común, relación que es indispensable determinar para poder hacer reconstrucciones filogenéticas que reflejen la historia del "flujo de la herencia".

El concepto de homología: definiciones básicas

Subtipos de homología: ortología, paralogía y xenología

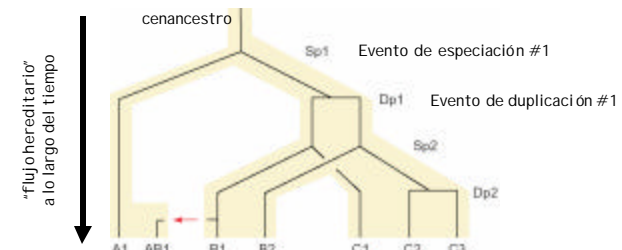


ortología: relación entre secuencias en la que la divergencia acontece tras un evento de especiación. El ancestro común es el cenacestro. La filogenia recuperada de estas secuencias refleja la filogenia de las especies.

paralogía: condición evolutiva en la que la divergencia observada acontece tras un evento de duplicación génica. La mezcla de ortólogos y parálogos en un mismo análisis filogenético recupera la filogenia correcta de los genes pero no necesariamente la de los organismos o taxa.

xenología: relación entre secuencias dada por un evento de transferencia horizontal entre linajes. Distorsiona fuertemente la filogenia de las especies.

Subtipos de homología: ortología, paralogía y xenología



La naturaleza del subtipo de relación de homología entre nodos terminales depende sólo de si la secuencia cenacestral está localizada en un punto que correspondea un evento de especiación ("Y invertida") o de duplicación (línea horizontal).

A, B y C corresponden a tres poblaciones (especies). Existen 2 eventos de especiación (Sp) y dos de duplicación (Dp). Dos genes cuya sec. cenacestral está en un nodo correspondiente a una Y invertida son ortólogos. Si dicha sec. correspondea un evento de duplicación, son parálogos. Así C2 y C3 son parálogos entre ellas pero ortólogos con respecto a B2. Ambas son parálogos respecto a B1, pero ortólogos respecto a A1. La flecha roja denota un evento de transferencia horizontal entre las especies A y B. El gen AB es xenólogo con respecto a los otros 6 genes.

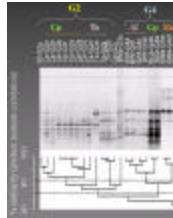
Cuando sea posible, hay que usar estos subtipos para definir la relación de homología

Marcadores moleculares usados en filogenética y evolución molecular

Polimorfismos de DNA y proteínas

I) Marcadores dominantes (? secuencias)

- RFLPs
- Fingerprints genómicos (AFLPs, RAPDs, Rep-PCR, SINES, SSCP, NSNPs ...)
- Análisis multilocus de isoenzimas
- etc ...



Los datos moleculares revelan información genética. **Sólo caracteres con una base genética son de interés en filogenética y evolución.** De ahí que los marcadores moleculares son generalmente los favorecidos para hacer inferencias filogenéticas y evolutivas a distintos niveles taxonómicos.

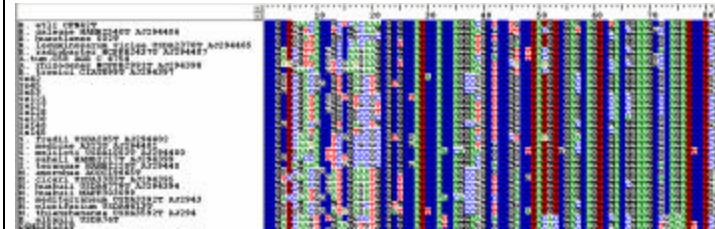
Los caracteres fenotípicos muchas veces tienen una base genética menos clara y están gobernados por las interacciones de muchos genes con el ambiente. **Muchos fenotipos presentan gran plasticidad**, es decir, que un mismo genotipo puede presentar una gradación de fenotipos. Esta variación fenotípica puede confundir las verdaderas relaciones filogenéticas y determinación de parentescos.

El uso de protocolos de PCR permite acceder a todo el mundo biológico para escrutinios genéticos, incluyendo a las comunidades de microorganismos no cultivables.

Los métodos moleculares permiten una fácil y robusta distinción entre homología y analogía y permiten hacer comparaciones de divergencia evolutiva usando métricos universales.

Marcadores moleculares usados en filogenética y evolución molecular

II) Secuencias moleculares DNA/proteína

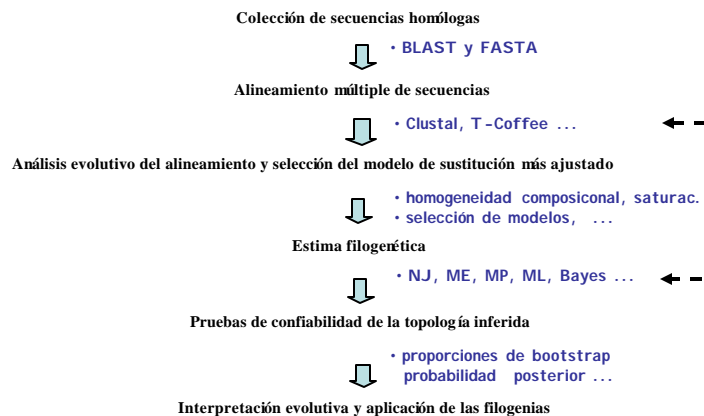


- La premisa fundamental en evol. molec. es que en las secuencias de DNA y de sus productos se encuentra escrita una buena parte de su historia evolutiva.
- Secuencias de DNA representan el "nivel anatómico" más fino de un organismo
- Buena parte de la biología moderna tiene por objetivo revelar la información contenida en secuencias moleculares.

Objetivos importantes del curso son:

- 1.- mostrar qué información evolutiva es la que se encuentra escrita en las secuencias de DNA y prot.
- 2.- proporcionar una base teórica y práctica sobre los métodos para recuperarla

Protocolo básico para un análisis filogenético de secuencias moleculares



Selección de marcadores adecuados para hacer inferencias evolutivas a distintos niveles de profundidad filogenética

Restricciones funcionales vs. tasas de sustitución:

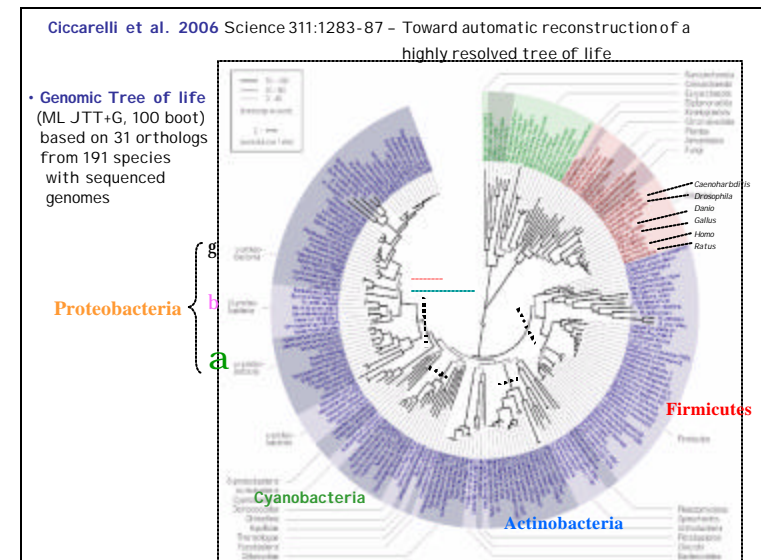
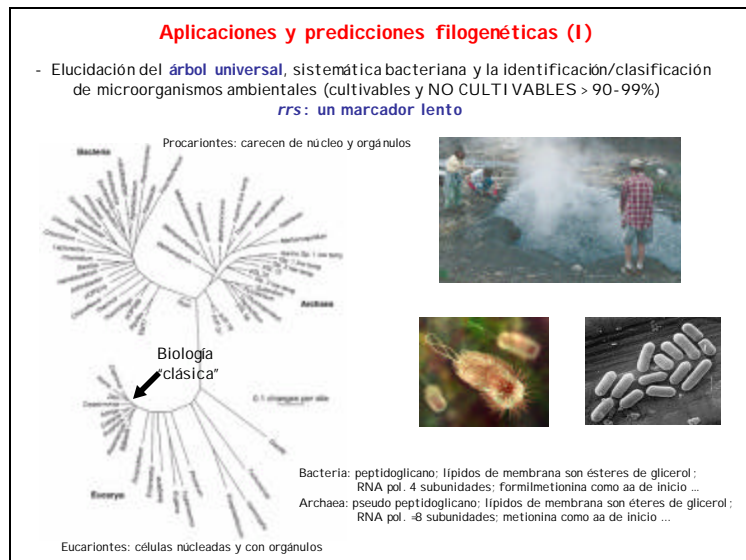
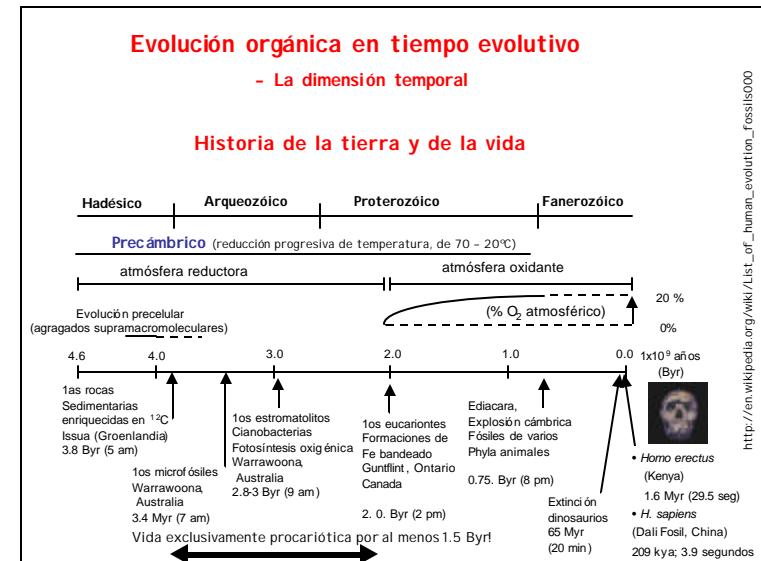
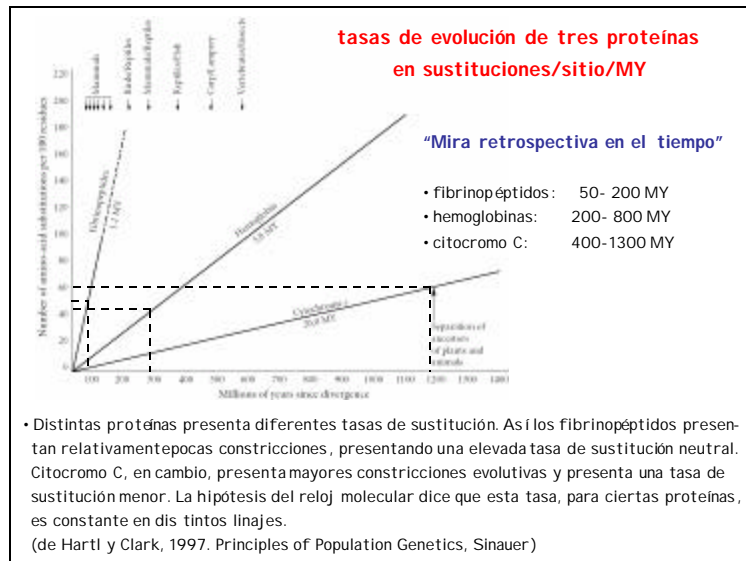
- Existe gran variabilidad en la tasa de sustitución entre genes y dominios genéticos:
 - intrones vs. exones
 - regiones codificadoras vs. regiones intergénicas o pseudogenes
 - residuos catalíticos vs. no catalíticos, dominios estructurales vs. no estructurales
 - 3as. posiciones vs. 1as y 2as en codones de secuencias codificadoras,
 - asas vs. orquillas en rRNAs y tRNAs ...
- Existen genes de evolución muy rápida o muy lenta:
 - fibrinopéptidos evolucionan una tasa $\times 900$ a la de ubiquitina y $\times 20$ a citocromo C
 - genes de HIV evolucionan a $\times 10^6$ veces la tasa de un gen humano promedio!

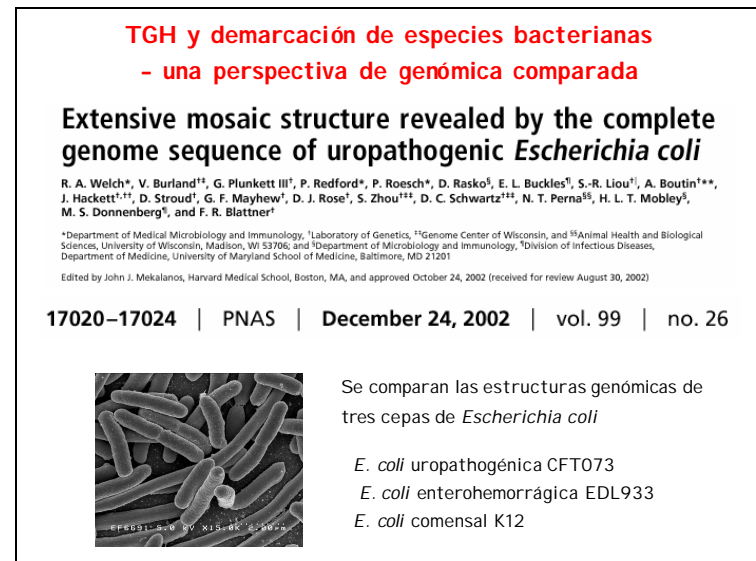
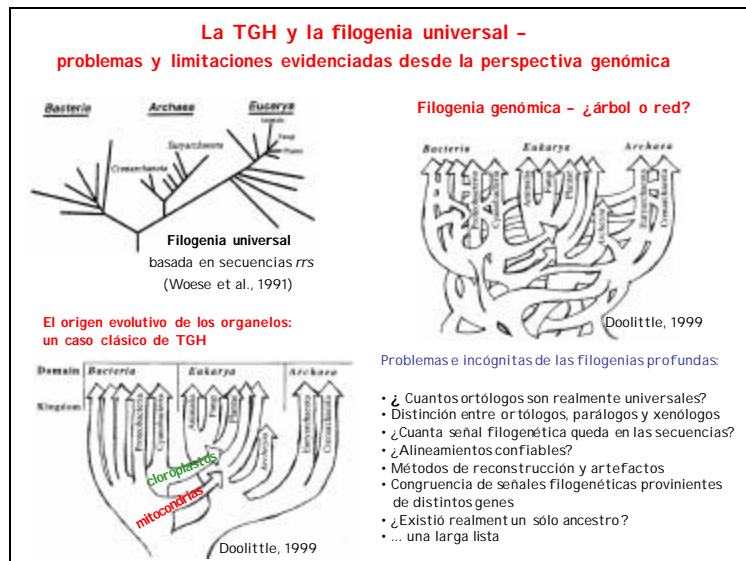
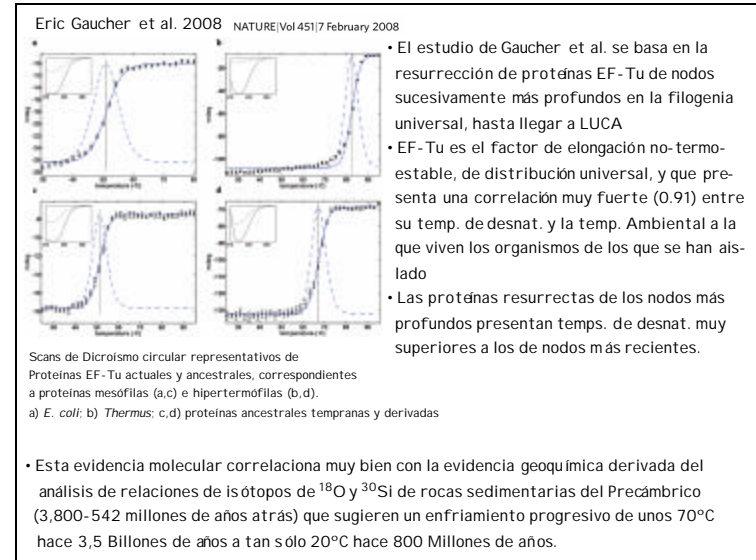
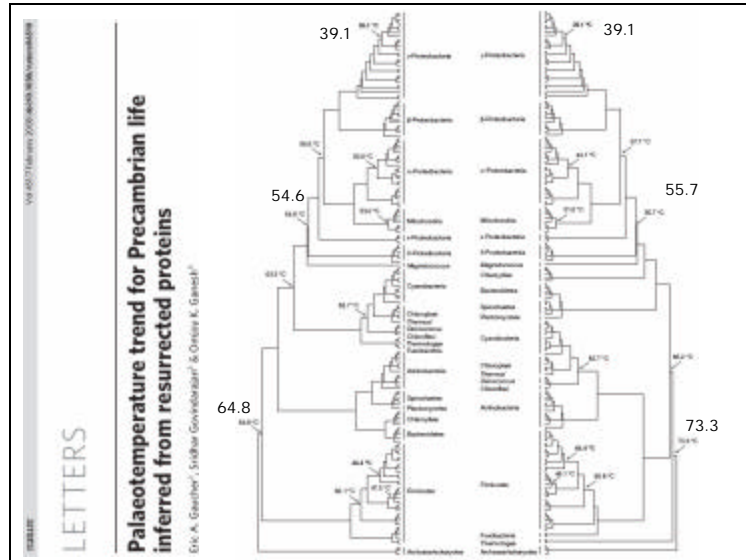
Tasas de evolución y la teoría neutral de evolución molecular:

el reloj molecular, calibración y datación de eventos de especiación/extinción de linajes y de pandemias ...

Tema 1: Conceptos básicos de evolución molecular y filogenética

Genómica Evolutiva I, Licenciatura de Ciencias Genómicas - UNAM, México. Semestre 2009-1





tomado de Welch et al. 2002

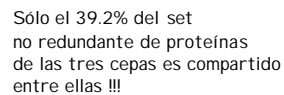


Fig. 2. Shared *E. coli* proteins. Comparison of the predicted proteins of the three *E. coli* strains shows the number of orthologs in each shared category and numbers of strain-specific proteins. Hypervariable proteins and proteins spanning island-backbone junctions were excluded from the analysis. Number of proteins counted: K-12, 4,288; CFT073, 5,016; EDL933, 5,063. In the totals for the three strains, orthologous proteins are counted *only once*. Orthologous proteins meet the same match criteria used for designation of backbone (see Materials and Methods).



- 85% (52 out of 61) of island codons have a significantly different codon usage vs. backbone
- EDL and CFT share only 10% of island genes, but >98% identity among backbone encoded proteins in the 3-strain comparison
- CFT and EDL strains contain 60 and 57 islands >4 kb, most of them at the same relative positions with respect to backbone markers, although island contents are unrelated
- 12 CFT073 and 10 EDL933 islands are closely associated to tRNA genes

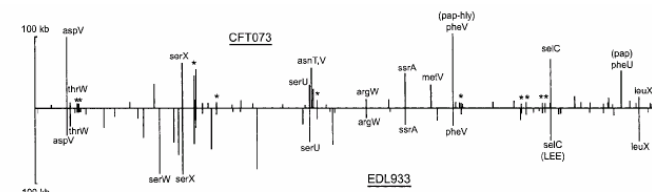


Fig. 3. Locations and sizes of CFT073 and EDL933 islands. Island size, vertical axis; position in colinear backbone, horizontal axis. All islands >4 kb are shown. Islands located at tRNAs are indicated by tRNA labels. One tmRNA (ssrA) is also an insertion target. +, CFT073 and EDL933 islands in the same backbone location but not near tRNAs.

A la luz de estos resultados queda evidenciado que ha de ejercerse un gran cuidado a la hora de seleccionar genes para la reconstrucción de filogenias de especies bacterias.

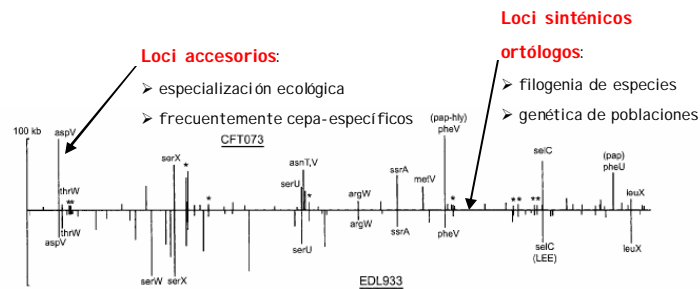
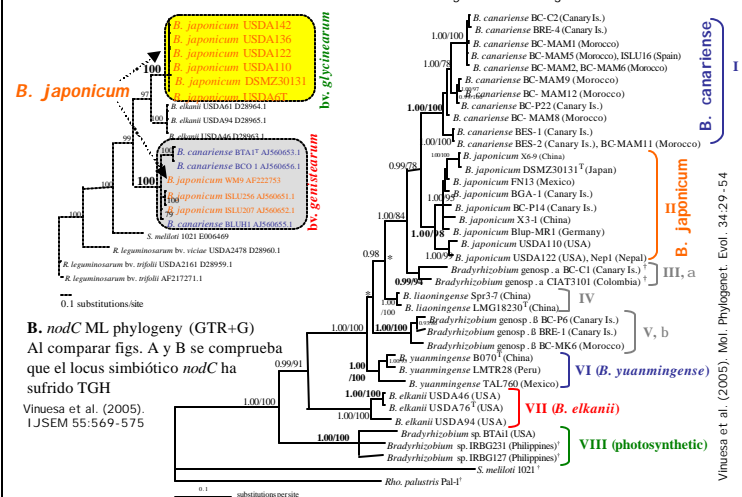


Fig. 3. Locations and sizes of CFT073 and EDL933 islands. Island size, vertical axis; position in colinear backbone, horizontal axis. All islands >4 kb are shown. Islands located at tRNAs are indicated by tRNA labels. One tmRNA (*ssrA*) is also an insertion target. *, CFT073 and EDL933 islands in the same backbone location but not near tRNAs.

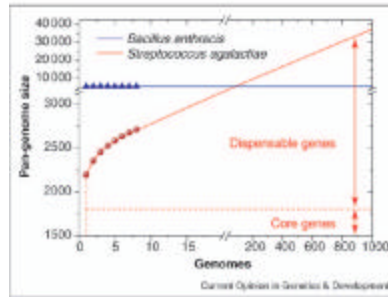
A. Filogenia Bayesiana de especies de *Bradyrhizobium* basada en ortólogos *glnI1*+*recA* congruentes



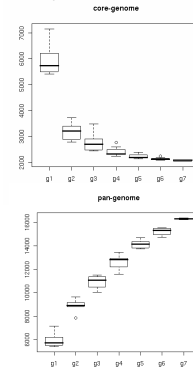
El pangenoma microbiano: pangenomas abiertos y cerrados

The microbial pan-genome
Duccio Medini¹, Claudio Donati¹, Hervé Tettelin², Vega Masignani³ and Rino Rappelli⁴

Current Opinion in Genetics & Development 2005, 15:589-594



Estima de tamaños del genoma "núcleo" (core) y pangenoma de *Rhizobiaceae* simbióticos (Vinuesa y Contreras-Moreira, 2007, unpubl.)



Aplicaciones y predicciones filogenéticas (II):

Evidencia molecular de transmisión de HIV-1 en un caso criminal usando genes de evol. rápida

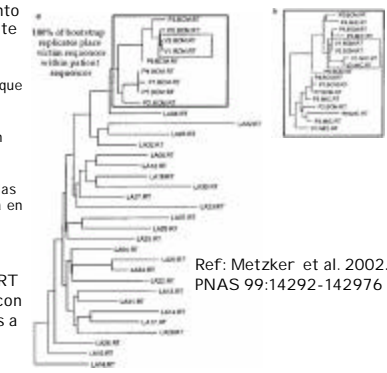
Un gastroenterólogo fue acusado del intento de asesinato en 2° grado de su novia mediante inyección de sangre contaminada con HIV-1.

Este estudio representa el primer caso en el que reconstrucciones filogenéticas de secuencias (paciente P, víctima V y controles LA de portadores en la población) fueron admitidas en una corte criminal en EUA.

Las filogenias de RT y de env mostraron que las secuencias de la V compartían ancestría directa en forma de paralogía con las de una P del gastroenterólogo.

Análisis de posiciones de codones de la RT de la V revelaron genotipos consistentes con mutaciones que confieren AZTR, similares a las presentadas en la P.

El establecimiento a priori de la P y V como posible par de transmisión del HIV-1 representó una clara hipótesis para ser evaluada en marcos de estadística filogenética.



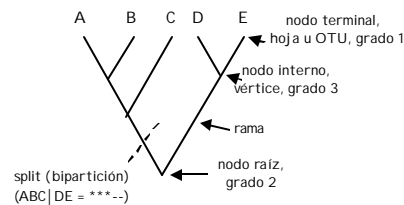
Ref: Metzker et al. 2002. PNAS 99:14292-142976

Filogenias del gen RT basadas en secuencias de la V, la P y LA, obtenidas por dos labs. independientes.
a) Baylor College of Medicine, Houston, TX (BMC)
b) Dpt. Ecology and Evol. Biol., Univ. Michigan (MI C)

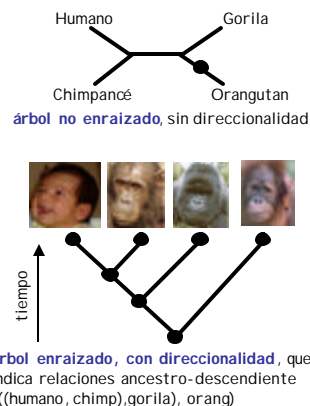
Arboles filogenéticos: una introducción al bosque (I) terminología y conceptos básicos: anatomía de un árbol

- Definición: Un árbol filogenético es una estructura matemática usada para representar la historia evolutiva (relaciones de ancestro-descendiente) entre un grupo de secuencias o organismos. Dicho patrón de relaciones históricas es la estima hecha de la filogenia o árbol evolutivo.

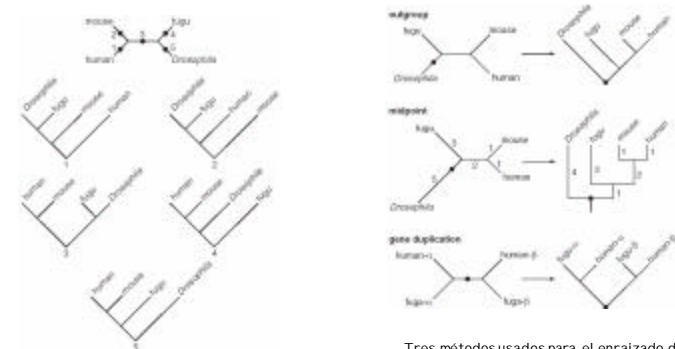
- Anatomía básica de un árbol



- reconstrucción de caracteres ancestrales
- longitud de ramas
- soportede, o confianza en biparticiones



Arboles filogenéticos: una introducción al bosque (II) enraizamiento de árboles



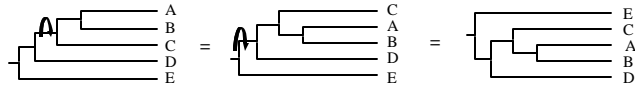
- La mayoría de los métodos de reconstrucción estiman árboles no enraizados, por lo que no disciernen entre las 5 posibles topologías enraizadas generables a partir de 4 OTUs.
- Para enraizar un árbol (decidir cual topología es la que refleja el proceso evolutivo), necesitamos información biológica adicional

Tres métodos usados para el enraizamiento de árboles:

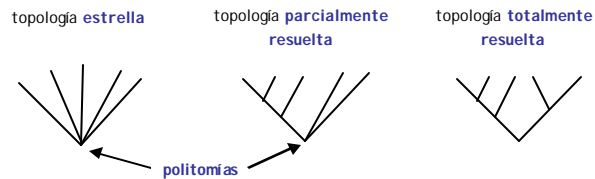
- grupo externo - (invertido) a grupo interno (vertebrados)
- punto medio - se pone la raíz en el punto intermedio del camino más largo del árbol
- duplicación génica - enraizamos en el nodo que separa a las copias paralogas

Arboles filogenéticos: una introducción al bosque (III) terminología y conceptos básicos

- Los árboles son como móviles: las ramas pueden rotarse sobre sí mismas sin afectar a las relaciones entre los OTUs; (((A,B),C),D),E) se puede representar como:

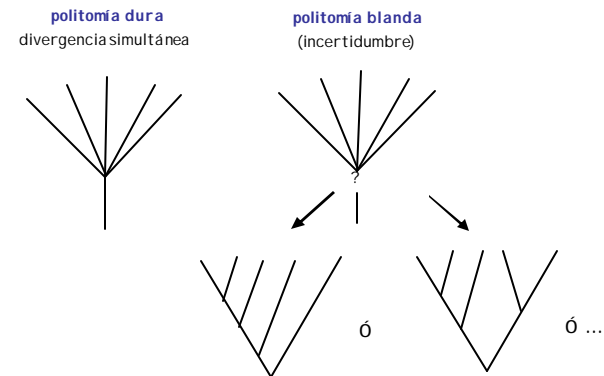


- Los árboles presentan distintos grados de resolución (árboles consenso)

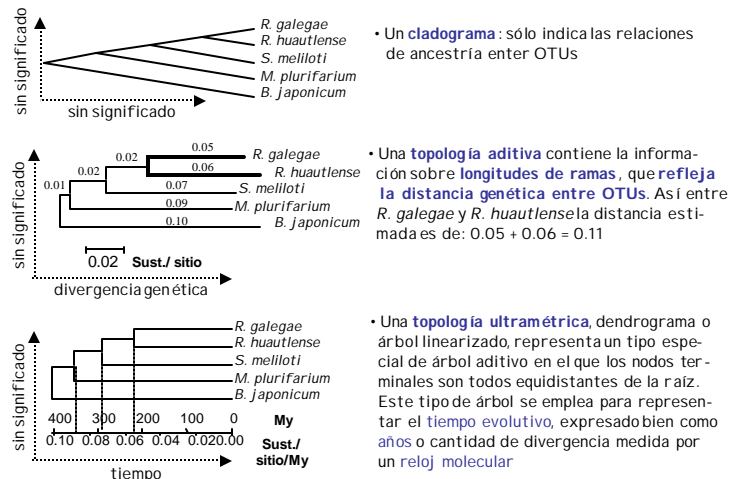


Arboles filogenéticos: una introducción al bosque (IV) terminología y conceptos básicos: tipos de politomías

- Existen distintos tipos de politomías

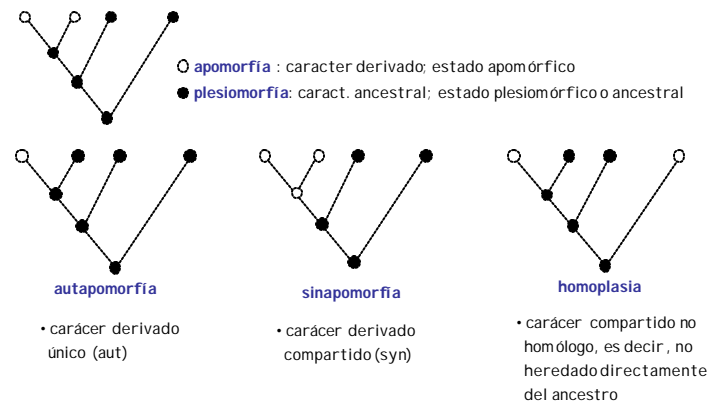


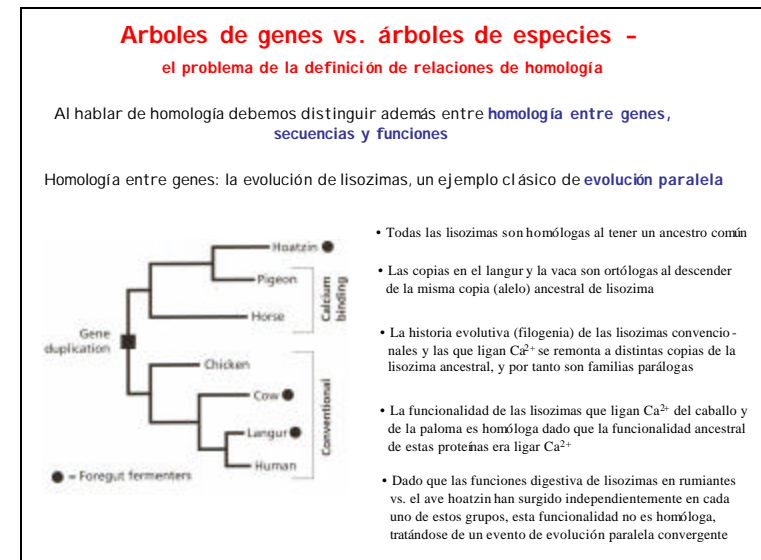
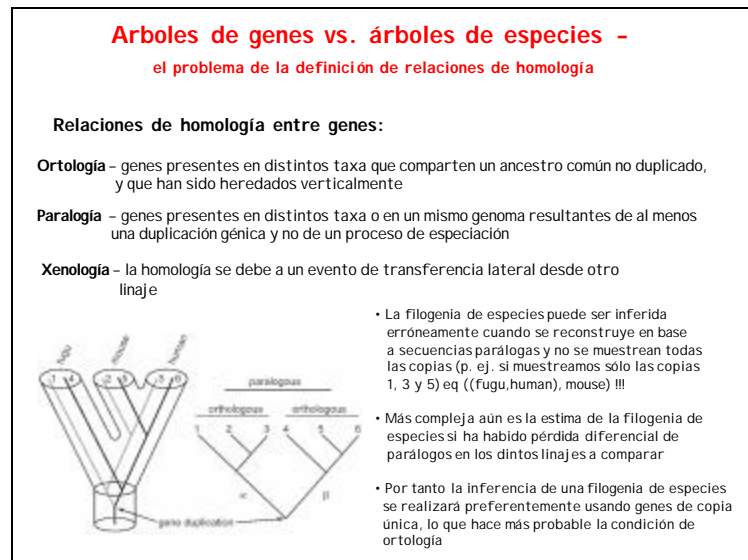
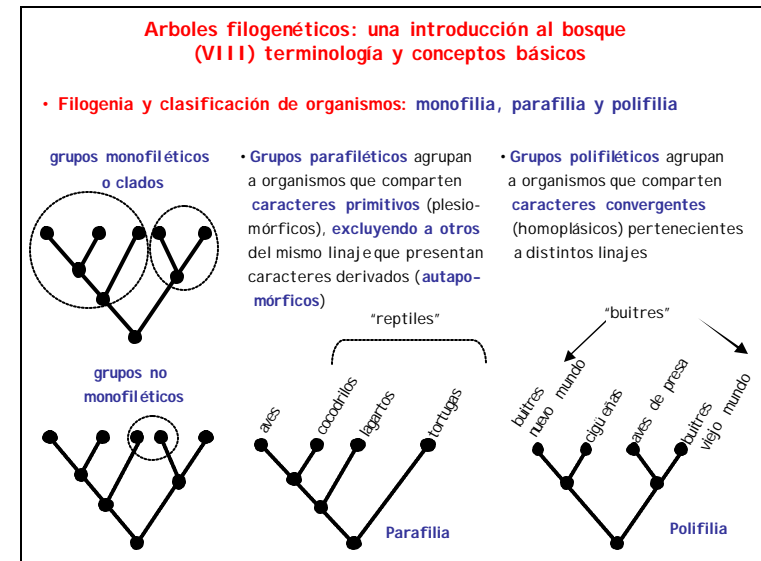
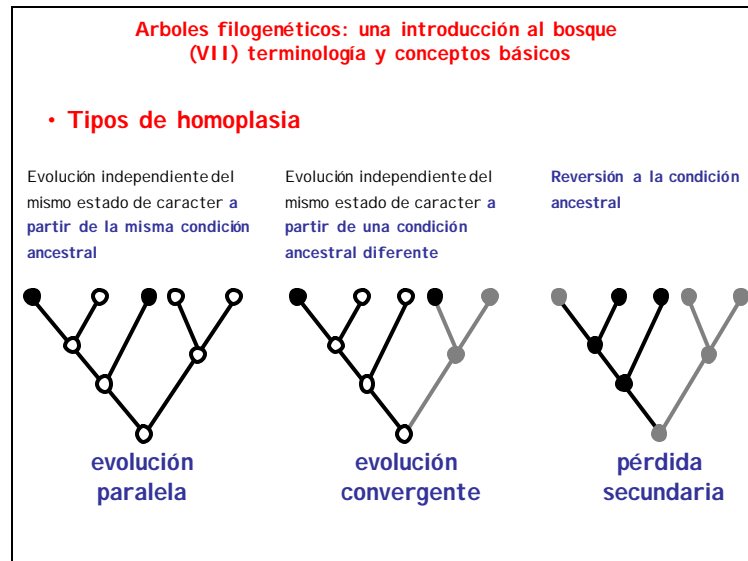
Arboles filogenéticos: una introducción al bosque (V) terminología y conceptos básicos: tipos de árboles



Arboles filogenéticos: una introducción al bosque (VI) terminología y conceptos básicos

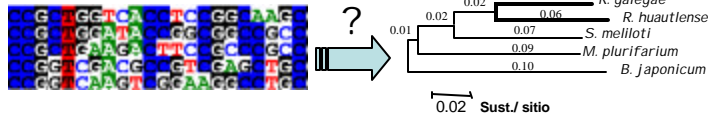
- Terminología relacionada con la reconstrucción de la **historia de cambios en estados de carácter**





Inferencia Filogenética - introducción

- La inferencia de relaciones filogenéticas a partir de secs. moleculares requiere de la selección de uno de los muchos métodos disponibles
- Con frecuencia la inferencia filogenética es considerada como una "caja negra" en la que "entran las secuencias y salen los árboles"



- Objetivos fundamentales de este curso son:**
 - desarrollar un marco conceptual para entender los fundamentos teóricos que distinguen a los distintos métodos de inferencia (clasificación de métodos)
 - presentar el uso de **modelos y suposiciones** en filogenética
 - manejo empírico de diversos paquetes de software para inferencia filogenética bajo diversos criterios

Métodos de reconstrucción filogenética - introducción

- La inferencia de una filogenia es un **proceso de estimación**; se trata de obtener la mejor estima posible de una historia evolutiva basada en la información incompleta y con frecuencia ruidosa contenida en los datos. Estos, por lo general, son moléculas y especies contemporáneas
- En principio, sería posible postular escenarios evolutivos *ad hoc* mediante los cuales cualquier filogenia tomada al azar podría haber producido los datos observados; es esencial por ello contar con un criterio estadístico y biológicamente riguroso para la selección de una o más topologías de entre todas las posibles
- Los métodos de inferencia filogenética están diseñados para este fin siguiendo una de dos estrategias computacionales:
 - mediante la definición de un **algoritmo** que determina los pasos a seguir para la reconstrucción de la topología
 - mediante la definición de un **criterio de optimización** mediante el cual poder decidir cual(es) topología(s) son las mejores (o igualmente favorecidas)

Métodos de reconstrucción filogenética: algoritmos vs. criterios de optimización

- Los **métodos algorítmicos combinan la inferencia del árbol y la definición del mejor árbol en una misma operación**. Son por ello muy rápidos
- Aquellos basados en criterios de optimización (CO) tienen en cambio dos pasos lógicos:
 - definir el **criterio de optimización** (descrito formalmente en una **función objetiva**) para evaluar cada posible topología, asignándole una puntuación con la que poder comparar cuantitativamente el mérito de cada árbol en base al criterio de optimización
 - en un segundo paso se usan **algoritmos de búsqueda** específicos para calcular el valor de la función de objetividad y para encontrar el/los árbol(es) con la mejor puntuación acorde a este criterio (un valor máximo o mínimo, según el caso)
- Los métodos basados en CO desacoplan por lo tanto los supuestos evolutivos hechos en el primer paso de las técnicas computacionales del segundo. El precio de esta claridad lógica es que estos métodos son muchísimo más lentos que los algorítmicos, debido a que tienen que hacer búsquedas en el inmenso espacio de topologías para encontrar la(s) mejor(es) topologías
- Los métodos algorítmicos tratan a los datos de diferente manera que los basados en criterios de optimización: análisis de **distancias vs. caracteres discretos**

Inferencia filogenética molecular - clasificación de métodos

- Podemos clasificar a los métodos de reconstrucción filogenética en base al tipo de datos que emplean (**caracteres discretos vs. distancias**) y si usan un **método algorítmico** o un **método de búsqueda basado en un criterio de optimización** para encontrar la topología óptima bajo el criterio seleccionado

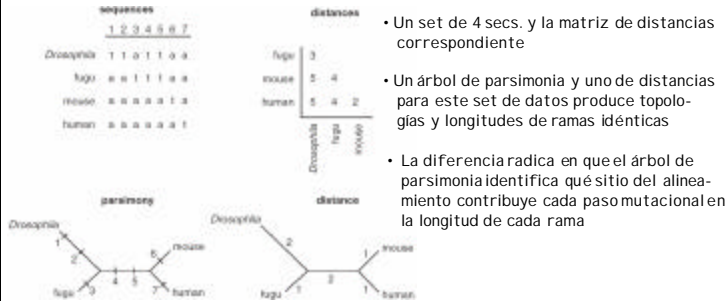
		Tipo de datos	
Método de reconstrucción	algoritmo de agrupamiento	distancias	caracteres discretos
		UPGMA y Neighbor joining	X
Método de reconstrucción	criterio de optimización	Minimos cuadrados y Evolución mínima	Máxima parsimonia y Máxima verosimilitud

Métodos de reconstrucción filogenética - una clasificación

I.- Tipos de datos: distancias vs. caracteres discretos

• Los **métodos de distancia** primero convierten los alineamientos de secuencias en una matriz de distancias genéticas en base al modelo evolutivo seleccionado, la cual es usada por el método algorítmico de reconstrucción para recuperar el árbol (**UPGMA y NJ**)

• Los **métodos discretos** (**MP, ML, Bayesianos**) consideran cada sitio del alineamiento (o una función probabilística para cada sitio) directamente



• Un set de 4 secs. y la matriz de distancias correspondiente

• Un árbol de parsimonia y uno de distancias para este set de datos produce topologías y longitudes de ramas idénticas

• La diferencia radica en que el árbol de parsimonia identifica qué sitio del alineamiento contribuye cada paso mutacional en la longitud de cada rama

Métodos de reconstrucción filogenética - una clasificación

II. Métodos algorítmicos vs. criterios de optimización

• Los **métodos algorítmicos o de agrupamiento** (clustering) siguen una serie dada de pasos o reglas computacionales previamente definidas (algoritmo) para reconstruir el árbol. Los métodos más usados son el **UPGMA y NJ**, basados todos en una **matriz de distancias**



• Estos métodos son muy rápidos, pero sensibles a parámetros tales como el orden en que se van añadiendo OTUs al árbol creciente

• Al no seguir un criterio de optimización, no se puede evaluar la bondad relativa de ajuste de topologías alternas a un set de datos particular

Métodos de reconstrucción filogenética - una clasificación

II. Métodos algorítmicos vs. criterios de optimización

• **Criterios de optimización**: reglas para decidir entre pares de topologías cual es mejor (dados los datos)

• Los métodos de reconstrucción de **MP y ML** utilizan diferentes criterios de optimización para seleccionar el/los árbol(es) entre las topologías que han de evaluar

• A cada topología se le asigna una puntuación (**score**) que es función del **ajuste existente entre la topología y los datos**

• Los métodos de optimización tienen la gran ventaja de requerir una función probabilística explícita que relaciona los datos con la topología (p. ej. un modelo de sustitución). Ello **permite evaluar la calidad de cualquier árbol** (topología), permitiendo el uso de distintas técnicas estadísticas para **evaluar la significancia con la que las distintas hipótesis evolutivas (topologías) en competición se ajustan a los datos!!!**

• Ejemplos de métodos de búsqueda de árboles por criterio de optimización son:

SCORE

- **MP: máxima parsimonia** (menor es mejor)
- **ML: máxima verosimilitud** (mayor es mejor)
- **ME: evolución mínima** (menor es mejor)
- **LS: cuadrados mínimos** (menor es mejor)

• La gran limitación de los métodos de optimización radica en que son computacionalmente muy costosos, requiriendo por lo general implementaciones heurísticas del algoritmo

Métodos de inferencia filogenética: en busca de la topología óptima

III.- Criterios de optimización y el problema del número astronómico de topologías

El número de topologías posibles incrementa exponencialmente con cada nuevo taxon o secuencia (S) que se añade al análisis

$$\text{No. de árboles no enraizados} = (2s-5)!/2^{s-3}(s-3)$$

$$\text{No. de árboles enraizados} = (2s-3)!/2^{s-2}(s-2)$$

Taxa	árboles no enraiz.	Árb. Enraiz.
4	3	15
8	10,395	135,135
10	2,027,025	34,459,425
22	3×10^{23}	...
50	3×10^{74} *	...

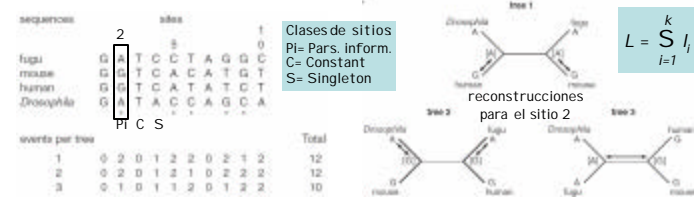
* Esto son aprox. 10,000 x el no. de átomos en el universo!!!

Por tanto se requieren de **estrategias heurísticas de búsqueda** árboles cuando $n > \sim 12$ para poder implementar métodos basados en criterios de optimización

Criterios de optimización

IV. Parsimonia: dados dos árboles, se prefiere el que requiere menos cambios en estados de carácter

- El método de parsimonia (Pars) considera cada sitio filogenéticamente informativo (PI) el alineamiento (al menos 2 pares de secuencias que compartan un polimorfismo). Los sitios constantes (C) y los singletons (S) no son considerados.
- El supuesto teórico (modelo de evolución) implícito al método es que el árbol más verosímil es aquel que requiere el mínimo número de sustituciones para explicar los datos (alineamiento). El criterio de optimización de la Pars es el de cambio o evolución mínima.
- Para cada sitio del alineamiento el objetivo es reconstruir su evolución bajo la construcción de invocar el número mínimo de pasos evolutivos. El número total de cambios evolutivos sobre un árbol (longitud en pasos evolutivos del árbol) es simplemente la suma de cambios de estados de carácter (p. ej. sustituciones) en cada sitio PI de la matriz o alineamiento



Criterios de optimización

V. Máxima verosimilitud: dadas dos topologías, la que hace los datos observados más probables ("menos sorprendentes") es la preferida

El método de máxima verosimilitud (ML) considera cada sitio variable del alineamiento (incluidos singletons). Bajo el criterio de ML se busca la topología que hace más verosímil el patrón de sustituciones de un alineamiento dado un modelo evolutivo explícito!

Así, para un set de datos D y una hipótesis evolutiva (topología) H , la verosimilitud de dichos datos viene dado por la expresión:

$L_D = \Pr(D|H)$ que es la probabilidad de obtener D dada H (una probabilidad condicional) !

Por tanto la topología que hace nuestros datos el resultado evolutivo más probable corresponde a la estima de máxima verosimilitud de la filogenia (likelihood score ó valor de verosimilitud).

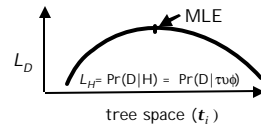
- la probabilidad está relacionada con la "sorpresividad" de los datos
- Estáramos sorprendidos de obtener este resultado, dada su bajísima probabilidad $(1/6)^{20}$ ó 1 en 3,656,158, 440,062,976!
- Pero la probabilidad depende del modelo probabilístico asumido
- En filogenética, las distintas topologías representan a los distintos modelos, y se selecciona aquel modelo que nos hace sorprendernos menos de los datos que hemos coleccionado



Criterios de optimización: la alteranativa Bayesiana

Aproximaciones tradicionales (matrices de distancia, ME, ML, MP)

- la búsqueda tiene por objetivo encontrar la topología óptima (estima puntual)
- no pueden establecer el soporte relativo de las biparticiones a partir de una única búsqueda



Aproximación Bayesiana

- no busca una solo topología óptima sino una población de árboles muestreados en función de su probabilidad posterior (algoritmos MCMC)
- la muestra de árboles obtenidos en una sola sesión de "búsqueda" es usada para valorar el soporte de cada split en términos de probabilidad posterior

