

PRÁCTICA 4: construcción de árboles con parsimonia y distancias, dos casos prácticos sencillos

Elvira Mayordomo y Francisco Merino

20/Marzo/2019

1 Introducción

En esta cuarta sesión se comenzarán a utilizar herramientas de construcción de árboles filogenéticos, centrándonos en los métodos de máxima parsimonia y distancias. La duración aproximada de esta sesión será de 2 horas.

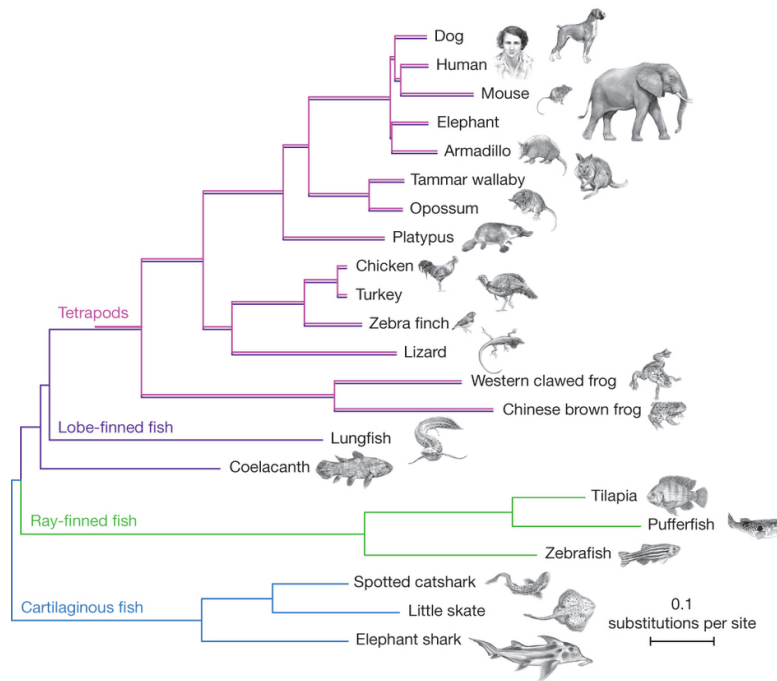
Al principio de la sesión dos estudiantes harán una introducción a las herramientas MEGA, ETE treeview y itol.

2 Filogenética computacional

En las últimas 4 clases se han presentado los conceptos básicos de filogenética computacional que pueden encontrarse en el material de la asignatura.

Una filogenia o árbol filogenético es una representación gráfica sencilla del proceso de evolución que puede seguir un conjunto de especies o secuencias biológicas. Para ello se utiliza una representación basada en una estructura de árbol (normalmente binario) como las que se explica en el grado de informática, donde los nodos representan individuos o secuencias, y las ramas una mutación o conjunto de mutaciones que sufre un nodo para *evolucionar* al siguiente. Las hojas del árbol suelen representar individuos pertenecientes a la época actual, mientras que los nodos internos representan a los ancestros del conjunto.

Los problemas relacionados con construir filogenias han sido categorizados como NP-duros, por lo que todas las herramientas software existentes se basan en heurísticas, obteniendo un resultado aceptado con un coste de



<http://www.nature.com/nature/journal/v496/n7445/images/nature12027-f1.2.jpg>

tiempo razonable. Dejaremos las herramientas que aplican el método de máxima verosimilitud, el más utilizado, para la siguiente práctica. En esta práctica nos centraremos en los métodos llamados de máxima parsimonia y los métodos basados en distancias, utilizando para ello el paquete MEGA MEGA <https://www.megasoftware.net/>

3 Uso de MEGA, ETE treeview ó itol

El primer objetivo de la práctica es familiarizarse con el uso de MEGA (sólo los métodos antes señalados) y algún visualizador de árboles como ETE treeview <http://etetoolkit.org/treeview/> o itol <https://itol.embl.de/upload.cgi>. Deben realizarse las instalaciones necesarias además de buscar las opciones disponibles para usar al menos un método de construcción de árboles por máxima parsimonia (MP) y los métodos NJ y UPGMA de construcción de árboles por distancias.

Para los métodos de distancias es suficiente utilizar como **distancia** el

número de diferencias o bien la p-distancia (ambas incluidas en MEGA). Opcionalmente se pueden probar otras distancias disponibles.

Al menos para el método de parsimonia es conveniente realizar **bootstrap** para conseguir resultados más robustos. El número de iteraciones o copias a realizar se personaliza según los casos de uso, para los casos de hoy podéis usar valores entre 10 y 100.

Es complicado comparar los resultados obtenidos por diferentes métodos sin información externa adicional, por lo que puede utilizarse también bootstrap con el método de distancias para tener un dato sobre **fiabilidad estadística** que permita comparar los métodos de distancias con el de parsimonia.

4 Caso 1: El tigre de Tasmania

El tigre de Tasmania es un mamífero nativo de Australia y Nueva Guinea que se extinguió en el siglo XX. Se trataba del último miembro viviente de su género (Thylacinus), cuyos otros miembros vivieron en tiempos prehistóricos.



<https://commons.wikimedia.org/wiki/File:Thylacinus.jpg>

Podéis leer la triste historia de su extinción en http://www.naturalworlds.org/thylacine/captivity/Benjamin/Benjamin_1.htm, o quizás no se extinguió, hay teorías alternativas <http://www.naturalworlds.org/thylacine/>.

El trabajo de hoy consiste en comparar el tigre de Tasmania con los marsupiales y con los felinos (o si se prefiere con los lobos) para clasificarlo adecuadamente. Se utilizarán **secuencias de referencia** de ADN mitocondrial obtenidas de Genbank siguiendo el proceso habitual de multialineamiento y construcción de filogenia.

5 Caso 2: Resolviendo un caso criminal

Volvemos al artículo ya utilizado en las prácticas 1 y 2:

Metzker M.L., Mindell D.P., Liu X.M., et al. (2002). Molecular evidence of HIV-1 transmission in a criminal case. *Proc Natl Acad Sci U S A*. 99(22):14292-7.

Utilizando las secuencias ya descargadas en la práctica 1 y alineadas en la práctica 2 (se pueden volver a alinear con mafft si se desea) se trata de resolver el caso criminal que motivó el citado artículo.

Entre las 174 secuencias del artículo nos quedamos sólo con las 42 en cuya definición Genbank aparece “(pol) gene” y .RT. Algunas de ellas son de la víctima (tienen la forma **Vx.xxx.RT**, por ejemplo **V1.BCM.RT**) otras del paciente del novio (tienen la forma **Px.xxx.RT**, por ejemplo **P1.BCM.RT**) y el resto de otros pacientes de sida de Lafayette (tienen la forma **LAppx.RT**, por ejemplo **LA23.RT**).

Se trata de construir el árbol filogenético de estas 42 secuencias y visualizar si se puede concluir mayor relación entre la Víctima y el Paciente que con el resto de los pacientes portadores de HIV de Lafayette.

6 Detalles del trabajo a realizar

En cada uno de los dos casos prácticos, el alumno debe obtener los datos necesarios, realizar la comparación de los mismos mediante multialineamiento y por último el análisis filogenético utilizando al menos los métodos de distancias y parsimonia.

Se deberá adjuntar a la memoria los multialineamientos utilizados además de los mejores árboles evolutivos obtenidos junto con la herramienta y parámetros usados.

Además debe incluirse una mínima discusión sobre las conclusiones del estudio realizarlo para el problema planteado en el enunciado.

7 Entrega de la práctica

La práctica debe ser entregada antes de la siguiente sesión (27 de marzo) mediante envío por correo electrónico de un fichero comprimido (preferiblemente .zip) a elvira@unizar.es