
TEMA 3.6. Obtención de árboles mediante el método de Máxima Verosimilitud

Contacto: Virginia Valcárcel (virginia.valcarcel@uam.es)

PROGRAMAS NECESARIOS

- (1) RAxML (Randomized Axelerated Maximum Likelihood). Este programa realiza inferencias filogenéticas basadas en Maximum Likelihood y permite el análisis rápido de grandes matrices. Se puede descargar de forma gratuita gratuita desde la web <http://sco.h-its.org/exelixis/software.html>. En esta misma página tienes la posibilidad de estimar la potencia de cálculo que necesitas para los análisis que quieres realizar. Para ello, incluye el número de muestras que tienes ("n" individuos, en nuestro caso 2 por especie, 17 en total) y el número de posiciones de la matriz (m).

Manual del programa: [RAxML](#)

Tutorial para RAxML: <http://sco.h-its.org/exelixis/hands-On.html>

- (2) RAxMLGUI. Esta es una interface gráfica para RAxML. Puede descargarse de forma gratuita desde la web <http://sourceforge.net/projects/raxmlgui/>

Manual de la interface gráfica: [RAxMLGUI](#)

Programas alternativos:

- (1) GARLIC (Generic Algorithm for Rapid Likelihood). Este programa se puede descargar de forma gratuita desde la web <http://code.google.com/p/garli/>. Además, podéis seguir un curso online para el aprendizaje (http://bodegaphylo.wikispot.org/Maximum_Likelihood_%28GARLI%29)

METODOLOGÍA Y PRÁCTICA

El objetivo de esta práctica es realizar un análisis de Maximum Likelihood combinando de las regiones que tenemos (ITS y *trnL-F*).

Paso 1. Descargar el programa RAxML.

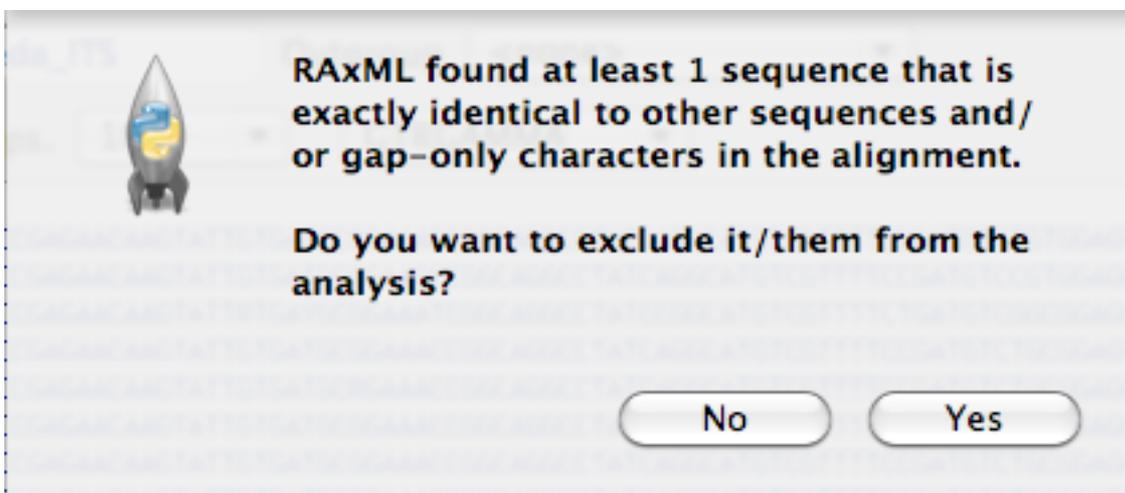
Para descargar el programa ve a la página web <http://sco.h-its.org/exelixis/software.html>, y avanza -al inicio aparecen otros programas: un editor de árboles, versiones de RAxML para matrices complejas, etc.- hasta que llegues al epígrafe RAxML. Descárgate la versión recomendada para el sistema operativo que tengas. RAxML es especialmente rápido en comparación con otros programas que realizan también búsquedas basadas en ML, como PAUP, ya que incorpora: (1) la construcción de un árbol inicial mediante parsimonia usando un algoritmo de random stepwise addition (véase tema 2.1), (2) para realizar el reajuste de las ramas utiliza Lazy Subtree rearrangement, (3) implementa un nuevo modelo de substitución además de los tradicionalmente usados en ML (GTR+GAMMA), que permite estimar las tasa por cada posición (este modelo es sólo recomendable para matrices muy grandes) y

además mejora la búsqueda ya que a pesar de usar un algoritmo Hill-climbing, permite la realización de pasos hacia atrás en el proceso.

Para este curso vamos a usar una interface gráfica de las que se han desarrollado. Puedes descargarla avanzando en la misma página hasta el epígrafe “Graphical User Interfaces”, la que vamos a usar en este curso es la desarrollada por Silvestri & Michalak.

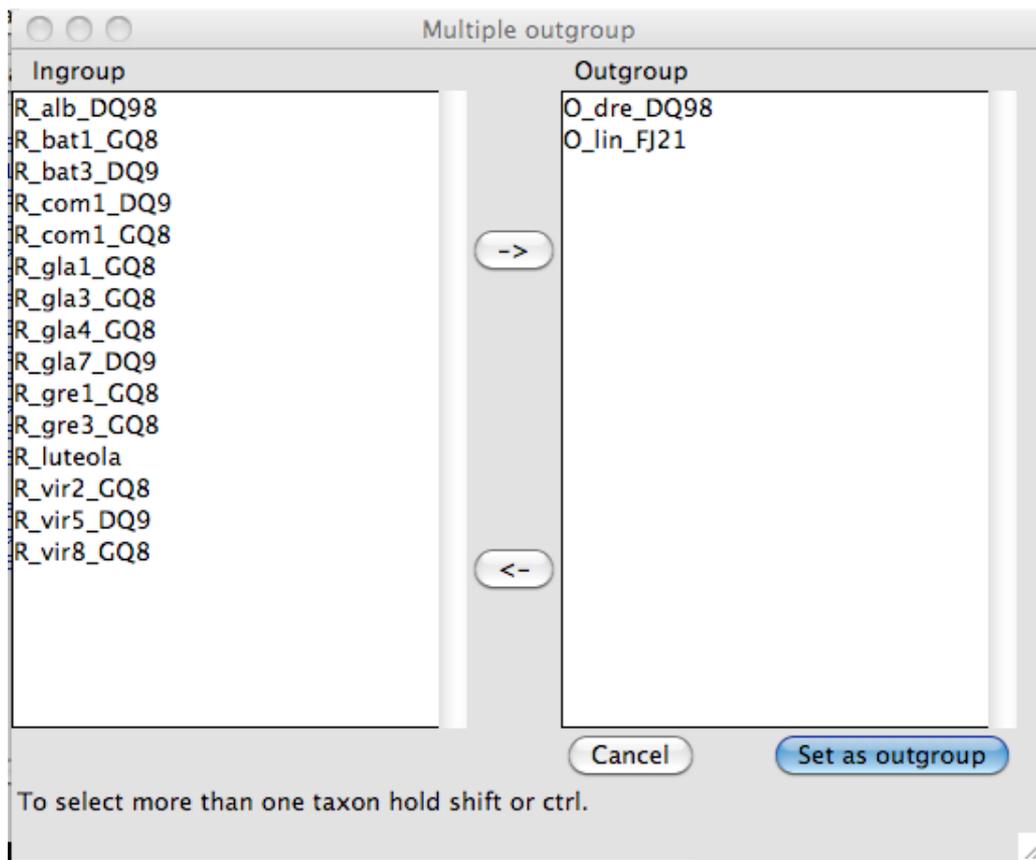
Una gran parte del software disponible para realizar análisis filogenéticos funcionan mediante comandos a través de la terminal. Dado que éste es un curso de iniciación hemos decidido minimizar el empleo de la terminal. No obstante, os sugerimos que una vez familiarizados con las herramientas básicas del análisis filogenético que aquí os proponemos, profundicéis en vuestra formación mediante el uso de estos programas desde la terminal. Para poder iniciaros en el manejo de RAxML desde la terminal podéis seguir el curso online de *Bodega Phylogenetics Wiki* (http://bodegaphylo.wikispot.org/RAxML_Tutorial)

- Paso 2. Vamos a realizar un análisis combinado de varias regiones, no sólo el espaciador ITS del ADN ribosómico y del espaciador trnL-F del genoma plastidial, sino que también acuérdate que la propia región ITS presenta tres regiones distintas: 2 espaciadores (ITS-1 e ITS-2) separados entre sí por el gen 5.8S. Por ello para hacer la combinación de la matriz podemos hacer varias alternativas. Una RAxML lee matrices con formato Phylip o Newick. Por ello, el primer paso es convertir la matriz (Glaucoseda_ITS_LF) a formato phylip y guardarla (véase tema 3.2).
- Paso 3. Abre la *Interface* gráfica del programa RAxML (raxmlGUI095-Mac/PC/Linux) carga la matriz “Glaucoseda_ITS_LF.phy” que has construido en “Add alignment”. Si el formato de la matriz es correcto en la pantalla aparecerá la matriz en formato phylip. Si en la matriz hay 2 o más secuencias completamente iguales te aparecerá un mensaje avisándote y dándote la opción de eliminar aquellas secuencias que sean idénticas dejando solamente una. Dile que no para que no elimine aquellas muestras iguales (ya que al ir a encadenar más genes es posible que dos individuos presenten la misma secuencia para una de las regiones pero diferente para otra región)

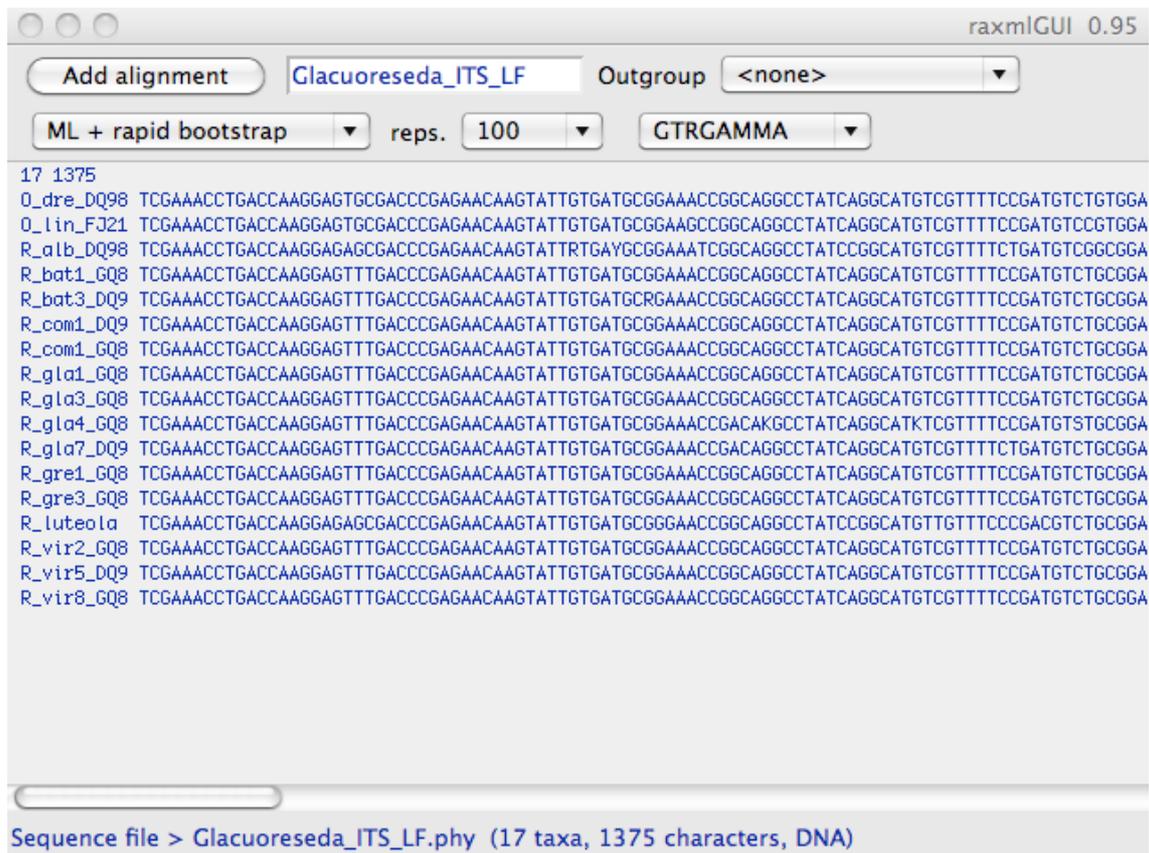


Paso 4. Para seleccionar el outgroup tienes dos opciones. Si sólo vas a seleccionar una muestra como outgroup entonces puedes seleccionarlo directamente en la pantalla en la opción "Outgroup". Si quieres seleccionar más de una muestra como outgroup, entonces ve al menú principal y en "Options" selecciona la opción "Select multiple outgroups...". Se abrirá una nueva ventana

Dale a la opción "Set as outgroups". Inmediatamente te saldrá una pantalla de aviso en la que se advierte de que si los outgroups no son monofiléticos, el primero de la lista será el utilizado como outgroup.



Paso 5. Las búsquedas pueden hacerse con varias opciones, deja señalada la que aparece por defecto en la que se realiza una búsqueda de ML y un análisis de remuestreo tipo *bootstrap* (véase tema 2.1.) rápido. Podemos hacer una prueba inicial realizando únicamente 100 réplicas [Ten en cuenta que para una publicación deberías utilizar al menos 1.000.000 de réplicas]. Para set de datos pequeños se recomienda usar el modelo GTRGAMMA. Dale a Run



Paso 6. Terminado el análisis se generarán automáticamente los siguientes archivos en la carpeta donde estuviese la matriz original:

- (1) RAxML_bestTree.Glaucoreseda_ITS_LF.tre
Ábrelo con Figtree. Este archivo contiene el mejor árbol.
- (2) RAxML_bipartitions.Glaucoreseda_ITS_LF.tre
Ábrelo con Figtree. Este archivo contiene el mejor árbol con los valores de bootstrap incluidos.
- (3) RAxML_bootstrap.Glaucoreseda_ITS_LF.tre
Ábrelo con Figtree (o si tarda mucho con un editor de texto). Este archivo contiene todos los árboles que se han conseguido en cada una de las réplicas de Bootstrap que hayas hecho. En nuestro caso deberían aparecer 1000 árboles.
- (4) RAxML_bipartitionsBranchLLLabels.Glaucoreseda_ITS_LF.tre
Ábrelo con Figtree. Este archivo contiene el mejor árbol con las longitudes de ramas.
- (5) RAxML_info.Glaucoreseda_ITS_LF
Ábrelo con un editor de texto (WordPad para PC, o TextWrangler para MAC; OJO es muy importante que no lo abráis con Word) en él podrás encontrar un resumen del análisis realizado, particiones, además de los Likelihood, el número de reajustes de las ramas del árbol.