

Máxima Verosimilitud

Datos Moleculares en Biodiversidad II

Nelson R. Salinas

Septiembre 19, 2020

Repositorio sesión:

https://github.com/nrsalinas/datos_moleculares_ii/tree/master/Taller_5.

En la sesión de hoy se reanalizará una parte del estudio de [Nickrent y colaboradores](#) (2004. Phylogenetic inference in Rafflesiales: the influence of rate heterogeneity and horizontal gene transfer. BMC Evol Biol 4, 40). En dicho trabajo los autores realizan la inferencia filogenética de un grupo diverso de angiospermas, con el objetivo principal de identificar las relaciones filogenéticas de varios taxa parásitos. Antiguamente varios linajes de angiospermas parásitas se clasificaban en el orden Rafflesiaceae, que contenía cuatro familias: Apodanthaceae, Cytinaceae, Mitrastemonaceae y Rafflesiaceae. Actualmente los miembros de este orden se clasifican en órdenes muy diferentes.

El conjunto de datos que manipularán durante el taller contiene representantes de las cuatro familias que componían Rafflesiales: *Pilostyles* (Apodanthaceae), *Rafflesia*, *Rhizanthus*, *Sapria* (Rafflesiaceae), *Bdallophyton*, *Cytinus* (Cytinaceae) y *Mitrastema* (Mitrastemonaceae).

El taller será evaluado, para lo cual deben describir las actividades realizadas en forma de miniartículo, incluyendo las partes clásicas que componen un manuscrito científico: introducción, métodos, resultados y discusión. A lo largo de la guía de laboratorio encontrarán varias preguntas, cuya finalidad es orientarlos en la elaboración del documento.

1. Suponga que ud. es el primer botánico que realizará la inferencia filogenética de este grupo de plantas parásitas. ¿Cuál cree que es la hipótesis nula de su investigación? ¿Que todas las plantas parásitas conforman un grupo monofilético? ¿Cuál es la hipótesis alternativa? ¿Por qué?
2. Descargue los archivos `18s.fasta`, `rbcL.fasta` y `atpB.fasta` del repositorio de la clase. Se recomienda que éste y todos los demás archivos y programas sean descargados al mismo directorio de su computador.
3. Instalar [IQTREE](#). La mayoría de computadores son de arquitectura 64 bits, pero si no está seguro confirme en la utilidad del sistema operativo que despliega la información básica del procesador y/o computador.

Una vez descargado, el programa puede ser ejecutado a través de la terminal de comando (“Símbolo del sistema” o “Command prompt”). Tenga en cuenta que debe utilizar los comandos apropiados para acceder al directorio donde guardó el binario de IQTREE, como `dir` para listar los elementos de la ubicación actual y `cd` para cambiar de directorio.

La instalación será confirmada ejecutando la opción de ayuda:

```
iqtree --help
```

4. Seleccione el modelo de evolución molecular más apropiado para cada partición (o gen) usando el criterio de información de Akaike (*Akaike Information Criterion*, AIC). Para ello debe ejecutar el programa usando las opciones que indican el archivo de entrada (`-s`) y desactivan la inferencia filogenética (`-m`). Ejemplo:

```
iqtree -s 18s.fasta -m TESTONLY.
```

Los resultados serán desplegados en pantalla y guardados en un archivo con el mismo nombre del archivo de entrada con el sufijo `.log` (p.e., `18s.fasta.log`). En dicho contenido busque la línea que contiene el modelo seleccionado por el AIC.

5. Instalar [RAxML](#). Los binarios para Windows se pueden descargar del sitio en [Github](#) del proyecto. Para seleccionar el tipo de binario es necesario conocer el número de procesadores del computador y la clase de instrucciones vectoriales que soportan. Casi todos los procesadores actuales tienen múltiples núcleos o hilos, por lo cual lo apropiado sería instalar la versión PTHREADS del binario. Si se dispone de una máquina muy antigua con un solo procesador entonces se debe instalar alguno de los demás binarios. Si tiene dudas respecto al número de núcleos, busque la información del modelo del procesador en internet o en la utilidad de información básica del procesador y/o computador.

Por otro lado, si su procesador es Intel entonces es sencillo saber las instrucciones vectoriales: la página de Intel tiene información detallada de los modelos y bajo el campo *Instruction Set Extensions* se encuentran las opciones soportadas. Si el procesador es AMD es más complicado saberlo y es recomendable instalar la versión más sencilla (SSE3).

La instalación será confirmada ejecutando la opción de ayuda. Ejemplo:

```
raxmlHPC-PTHREADS-AVX -help
```

6. Con base a los resultados del paso 4 y la [documentación](#) de RAxML, busque la opción más apropiada del programa para especificar el modelo de evolución molecular (el listado de modelos para secuencias de ADN se encuentra en la página 29).
7. Ejecute una búsqueda simple del árbol más óptimo (bajo el criterio de optimalidad ML) en RAxML. Para ello descargue los archivos `concatenado.phy` y `concatenado.part` del repositorio, los cuales corresponden a la matriz concatenada y al archivo de configuración requerido por RAxML, respectivamente. Ábralos en un editor de texto e indique a qué formato corresponde el primero y qué información despliega el segundo.

Una búsqueda sencilla en RAxML requiere especificar las siguientes opciones:

- (a) `-p`: Número de inicio para seleccionar algunos parámetros de ejecución. Puede seleccionar cualquier número entero mayor a 0.
- (b) `-s`: Archivo que contiene la matriz de caracteres.

- (c) **-q**: Archivo de configuración para matrices con múltiples particiones.
- (d) **-m**: Modelo de evolución molecular.
- (e) **-n**: Texto que será empleado para nombrar los archivos de salida.
- (f) **-o**: Taxa a emplear como grupos externos. En nuestro caso, debemos usar las secuencias de *Ceratophyllum*, *Cinnamomum*, *Laurus*, *Piper* y *Saururus*.

Considerando la información anterior, ¿cuál es el comando completo necesario para ejecutar la búsqueda? Puede encontrar más información en el [tutorial](#) oficial del programa o el mencionado manual.

8. Descargue [FigTree](#). Seleccione la distribución más apropiada para su sistema operativo (zip es la opción más usual para Windows). Después de descomprimir el archivo abra el ejecutable **FigTree v1.4.4**. Si no lo puede abrir es posible que necesite instalar el [Java Runtime Environment](#).
9. Visualize en FigTree el árbol filogenético producido por el punto 7 en Figtree (el nombre del archivo debe comenzar con “RAxML_bestTree”).
10. Compare los resultados con aquellos producidos por un análisis de parsimonia. Descargue el archivo **consenso_parsimonia.newick** del repositorio, que corresponde al consenso estricto de los árboles más parsimoniosos para el conjunto de datos en cuestión. ¿Existen diferencias topológicas entre los árboles producidos por ambos métodos? ¿Cuales son las relaciones filogenéticas de los taxa parásitos? ¿Cree que este es un ejemplo del fenómeno “atracción de ramas largas”?