

---

## Taller 0

Alineación de secuencias de ADN usando MAFFT. Manipulación de matrices moleculares con Sequence Matrix. Búsqueda de modelos de evolución molecular usando jModeltest. Separación en fases de genes nucleares con DnaSp (Phase).

---

Para seguir esta práctica, es conveniente que pegues la carpeta “Practica” en el escritorio de tu computadora. Entonces se respetan las rutas de acceso a los datos y programas, tal como lo describimos acá. Los programas usados en este tutorial son libres y se pueden descargar de las webs oficiales.

### Alineamiento usando MAFFT

Ubicar la matriz no alineada “12s\_Liolaemus.fas” en la carpeta “datos” y abrirla en un editor de texto (como WordPad o TextWrangler). Esta matriz contiene secuencias del gen mitocondrial 12s de individuos de las especies de lagartijas patagónicas *Liolaemus gracilis* y *L. bibronii*. Entre ambas especies se han reportado evidencias de hibridización en un estudio molecular multilocus y morfológico (para detalles ver en la carpeta de papers Olave et al. 2011, Mol. Phylog. Evol. 61:381-391).

Usando MAFFT desde tu computadora personal:

1. Instalar el programa ejecutando el archivo mafft-7.037-signed.pkg.
2. Una vez instalado, la utilización es por línea de comando. Para ello, entra a “Símbolos de sistema” si tenés PC, o “Terminal” si usas una MAC. Tipea mafft y da enter.
3. Ingresar la ruta y el nombre de la matriz de datos. Para ello, escribir la ruta Escritorio/Practica/datos/12s\_Liolaemus.fas
4. Escribir el nombre del output: Escritorio/Practica/datos/12s\_Liolaemus-alineado.fas NOTA: si no se da la ruta, entonces se guarda en el directorio raíz de la computadora.
5. Elegir el formato de la matriz output. En este caso mantener el formato fasta, entonces escribir 4 y apretar enter (i.e. formato fasta manteniendo el orden de los taxa que presentamos en el input).
6. Elegir la estrategia “Fast” para un alineamiento rápido (cuando trabajes con tus datos es mejor usar una de las “accurate”, pero en esta práctica vamos a buscar resultados más rápidos). Entonces, escribir 2 y apretar enter.
7. No vamos a dar argumentos adicionales. Entonces simplemente apretar enter.
8. Finalmente, se muestra el comando final para que el usuario verifique. Apretar enter.
9. Abrir el output 12s\_Liolaemus-alineado.fas en un editor de texto y observar el contenido.

Usando MAFFT online:

1. Entrar al sitio <http://www.ebi.ac.uk/Tools/msa/mafft/>
2. Copiar el contenido de la matriz 12s\_Liolaemus.fas (versión no alineada), y pegarlo en el cuadro de dialogo del paso 1 “STEP 1 - Enter your input sequences”.

3. En el “STEP 2 – Set your parameters”, dejar por defecto en fasta el formato output expresado en el paso 2, y hacer clic en “More options”. En la opción “TREE REBUILDING NUMBER” seleccionar 100 (long run) y en “MAXITERATE” seleccionar también 100 (long run). Dado que tenemos una matriz chica, no toma demasiado tiempo.
4. En “STEP 3 – Submit your job” hacer clic en el botón “Submit”, y esperar el output.
5. Copiar la matriz alineada y pegarla en un editor de texto en blanco, bajo el nombre 12s\_Liolaemus-alineado\_online.fas.

### **Manipulación de matrices**

El programa Sequence Matrix permite:

- I. Visualizar y modificar nombres de taxa de matrices moleculares.
- II. Da información resumen (como largo de secuencias, entradas faltantes, cantidad de ambigüedades).
- III. Permite exportar diferentes formatos de matrices (phylip, nexus y fasta).
- IV. Permite concatenar diferentes genes en una sola matriz.
- V. Calcula distancias de secuencias de pares (pairwise distances).

Ejercicios:

1. Abrir las matrices fasta “12s\_Liolaemus-alineado.fas” y “cytb\_Liolaemus.nxs” en un editor de texto y observar el contenido.
2. Abrir el programa Sequence Matrix. Ir a Import>Add sequences y abrir las dos matrices de secuencias o bien arrastrar los archivos hacia el programa. Ante la pregunta “Would you like to recode external gaps as question marks?”, seleccionar “yes to all”. Esto hará que en caso de encontrar gaps al comenzar a leer una secuencia los reemplazará por entradas faltantes (“?”).
3. Observar la consola. Identificar número total de taxa (ver abajo), cantidad de genes (nombrado abajo y listado en la cantidad de columnas) y el largo total de las secuencias detalladas para cada taxón de la matriz (columna “Total length”).
4. Ubicar el individuo 10388. Observar que este individuo fue guardado con el nombre 10388 en la matriz de cyt-b, y con el nombre 10388bib en la matriz de 12s. Como el nombre del individuo en cada matriz es diferente, Sequence Matrix reconoce ese taxón como dos individuos distintos. Si se modifica el nombre de uno de ellos, el programa automáticamente une ambas secuencias en un solo individuo. Entonces, modificar el nombre de 10388 a 10388bib, haciendo clic sobre 10388 y reescribiendo el nombre. Observar el cambio.
5. Repetir el punto 4 con los individuos 6847 y 8137.
6. El individuo 5044saxa es un individuo outgroup en los análisis que vamos a hacer más adelante. Excluirlo de la matriz haciendo clic sobre 5044saxa y después el botón “Delete taxon” de arriba.
7. Exportar una matriz por cada gen por separado en formato fasta. Para eso, ir a Export>Export sequences (one file per column). Seleccionar fasta como formato de output

en “Write file with format.” y en “Browse” seleccionar el directorio donde se quiere guardar (Escritorio/Practica/datos/) y hacer clic en el botón “Write files”.

NOTA: Sequence Matrix tiene una maña algo molesta, que es que para poder grabar la matriz output en una carpeta, cuando se selecciona no se debe abrir la carpeta, sino que simplemente se debe seleccionar. De otra manera da error.

8. Ahora exportar una matriz concatenada con ambos genes cyt-b y 12s. Para eso, ir a Export>Export sequences as Nexus (non-interleave), y guardar con el nombre cytb&12s\_Liolaemus.nxs en el directorio Escritorio/Practica/datos/
9. Abrir la matriz cytb&12s\_Liolaemus.nxs en un editor de texto y observarla. Esta matriz tiene varios caracteres no necesarios para el formato Nexus que suelen traer problemas para que un programa pueda leerla. Sin embargo, al final tiene un bloque “CHARSET” que tiene información importante describiendo desde donde hasta donde va cada gen (en largo de pb).
10. Repetir el punto 8 pero con la opción Export> Export sequences as Nexus (naked e.g. for GARLI) y guardando con el nombre cytb&12s\_Liolaemus-naked.nxs.
11. Abrir la matriz cytb&12s\_Liolaemus-naked.nxs en un editor de texto y observarla.

### **Búsqueda de modelos de evolución molecular usando jModeltest**

1. Abrir el programa jModeltest.jar que está en la carpeta Practica/Programas
2. Ir a File>load DNA alignment y ubicar la matriz 12s\_Liolaemus-alineado.fas
3. Ir a Analyses>compute likelihood scores.
4. Para hacer el trabajo del programa más rápido, solo vamos a evaluar el likelihood de 3 modelos en “number of substitution schemes”. Esto sumado a los parámetros +F, +I, +G, nos permite explorar un total de 24 modelos posibles para nuestra matriz de 12s. Entonces, seleccionar 3 en “number of substitution schemes” y dejar lo demás por defecto. Hacer clic en el botón “Compute likelihood” y esperar (paciencia! Va a demorar algunos minutos. En mi caso tomó 3 minutos).
5. Una vez que haya terminado, ir a Analyses>Do AIC calculations. Tildar la opción “Use AICc correction”.
6. Repetir el punto 5 con BIC calculations.
7. Ir a Results>Show result table. En la ventana que se abre, hacer clic en el botón AICc superior. Después ordenar la tabla de mayor AICc a menor. Para eso hacer clic en la columna AICc. El primer modelo de la tabla es el seleccionado con este criterio (también se puede ver pintado de rojo)
8. Ahora comparar resultados con los obtenidos con el criterio BIC. Para eso, hacer clic en el botón BIC y después ordenar la tabla haciendo clic en la columna BIC.
9. Si los modelos seleccionados son distintos entre criterios, comparar los modelos seleccionados. ¿Con cuál te quedarías y por qué?

**Preparando matrices para la inferencia de árboles de especies en BEAST: separando genes nucleares en fases con DnaSp.**

Hasta ahora hemos trabajado con genes mitocondriales, la cual es haploide. Para el caso de trabajar con genes nucleares, hay que tener en cuenta que estos genes son diploides. Entonces, puede separarse las fases de ambos cromosomas que contiene una matriz. Por un lado esto es importante porque, de no hacerlo, la matriz probablemente contenga muchas ambigüedades, cuando los individuos son heterocigotas. Pero además, provee una gran ventaja, ya que duplicamos la cantidad de datos en la matriz. Entonces, en esta parte utilizaremos el programa DnaSp, que incluye PHASE, para separar las fases de los genes nucleares.

**Ejercicios:**

1. Abrir DnaSp e ir a file> open unphase/geneotype data file y seleccionar LPA11e\_Liolaemus.fas que esta dentro de matrices-arbol\_especies.
2. En este caso dejamos valores por default y apretar run.
3. Cuando termine, ir a file> Save/Export data as y guardar con el nombre del gen y \_phases.txt
4. Repetir con las matrices nucleares LPB4g\_Liolaemus.fas y LPB9c\_Liolaemus.fas.
5. Una vez que las tres matrices nucleares hayan sido separadas en fases, usar el Sequence Matrix para unificar nombres con los genes mitocondriales. No hay problema en que un “individuo” tenga missing data para los genes mitocondriales.