

---

## Taller I

Inferencia de árboles de especies con \*BEAST.

---

### Descripción

Este tutorial se propone estimar un árbol de especies en un marco coalescente Bayesiano. Se utilizará BEAST v1.6, un programa ampliamente utilizado en análisis evolutivos. Además de BEAST, el paquete incluye:

- BEAUti (Bayesian Evolutionary Analysis Utility): es un programa con interfaz gráfica para crear los input para BEAST.
- LogCombiner: puede ser usado para combinar dos análisis independientes, escritos en archivos .log y .trees (output de BEAST), en un solo archivo cada uno, para que sea leído por Tracer (.log) y TreeAnnotator (.trees). NOTA: solo tiene sentido combinar análisis que fueron realizados usando el mismo input .xml.
- TreeAnnotator: es usado para resumir el muestreo de árboles realizado en el análisis de BEAST. El resultado es un árbol que maximiza la credibilidad de los clados y resume las estimaciones posteriores de los parámetros para que sean visualizadas en el árbol.

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. BEAST es libre y puede ser descargado de su web oficial.

### Inferencias de árboles de especies en un enfoque Bayesiano con \*BEAST

En los últimos años se han desarrollado una serie de programas que permiten estimar “árboles de especies” cuando se tienen más de un gen independiente. Esto surge a partir de haber reconocido que cada árbol de gen independiente (no ligados) representa una historia evolutiva particular, que no necesariamente es igual a la evolución de las especies, ni a la de otro gen. Esto es por un proceso que conoce como división de linaje incompleta (“incomplete lineage sorting”, ver Maddison 1997; Funk y Omland 2003).

Cuándo se concatenan varios genes en una sola matriz, se asume que esas secuencias se comportan como un supergen, donde todos los genes deben representar la misma historia. En cambio, las inferencias de árboles de especies, tienen incorporado en su modelo de inferencia que cada gen tiene su propia historia evolutiva independiente, lo que aparentemente los hace más eficientes (ver Leaché y Rannala 2011).

Ejercicios en \*BEAST (inferencia de árboles de especies):

1. Ubicar la matriz “12s\_Liolaemus.fas” en la carpeta Taller\_1-arboles\_especies>matrices 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 hibridación en un estudio molecular multilocus y morfológico (para detalles ver Olave et al. 2011, Mol. Phylog. Evol. 61:381-391). En este caso, se han eliminado de las matrices los

individuos que mostraban introgresión mitocondrial, ya que la inferencia de \*BEAST no modela flujo génico interespecífico (asume que toda la discordancia de los árboles de genes se debe a división de linaje incompleta).

Observar que los genes nucleares se encuentran separados en fases (una secuencia por cromosoma). Para ello se siguieron los pasos descritos en el Tutorial 0, utilizando el programa DnaSp.

2. Abrir BEAUti 1.6.
3. Hacer clic en el botón del signo + y elegir las 5 matrices fasta que se encuentran en la carpeta Taller\_1-arboles\_especies>matrices.
4. Tildar la opción que dice Use species tree ancestral reconstruction (\*BEAST). Apretar ok en la venta que aparezca. Automáticamente, BEAUti cambia a la solapa “Traits”.
5. En la solapa “Traits” se deben especificar que individuos son de las mismas especie. Las matrices de secuencias originales listan los nombres de los taxa con un número de colección y, separado por un guión bajo, una indicación de la especie. Entonces, “gra” refiere a *L. gracilis*, “bib” a *L. bibronii* y los demás corresponden a especies outgroups.

Para especificar qué individuos son parte de qué especie, se puede escribir el nombre de la especie en la columna “Value” a mano, o bien se puede pedir a BEAUti que automáticamente lo detecte. Como en este caso, siempre hay un guión bajo antes de la especie, hacer clic en el botón Guess trait value y en “with delimiter” escribir un \_

En “Define by its order” seleccionar second. Apretar ok.

6. Volver a la solapa “Partition”. Observar las columnas Sites Model, Clock Model y Partition Tree. En este caso, tenemos dos genes mitocondriales en el análisis (cyt-b y 12s). Como la mitocondria se hereda como un bloque completo, cyt-b y 12s si deberían reconstruir el mismo árbol de gen. Entonces, seleccionar ambos genes y hacer clic en el botón Link trees. Apretar ok en la ventana que aparece.
7. Ir a la solapa “Sites”. Seleccionar los modelos de evolución GTR+I para cyt-b y 12s, y dejar por defecto HKY en los otros 3 nucleares (esto depende de los resultados arrojados por jModeltest).
8. Ir a la solapa “Trees”. Observar que solo aparece uno de los genes mitocondriales. Esto es porque en el punto 5 especificamos que cyt-b y 12s comparten el mismo árbol. Seleccionar el gen mitocondrial y en Ploide type seleccionar mitochondrial. Esto es necesario porque la mitocondria no solo se hereda como bloque completo, sino que además tiene un tamaño poblacional efectivo menor respecto a los nucleares (ya que son haploides, entonces hay una sola copia por individuo, además de ser heredado solo de la madre en muchos organismos) y el programa debe contemplar esto en la inferencia filogenética.
9. Ir a la solapa MCMC. Las inferencias de árboles de especies Bayesianos suelen requerir cadenas de MCMC largas (> 300 millones), dependiendo de los datos que analicemos. Por esto, son computacionalmente muy demandantes.

En esta práctica solo vamos a correr 10 millones de generaciones, y muestrear parámetros cada mil generaciones. Especificar estas cantidades escribiendo 10000000 en Length of chain y 1000 en Echo state to screen every y Log parameter every. Hacer clic en el botón Generate BEAST file. Tener en cuenta que en general, es conveniente que el análisis cuente con 10,000 árboles y estimaciones de los parámetros muestreados. Entonces, para un este caso  $10,000,000/10,000 = 1,000$ ; finalmente Log parameters every = 1,000.

10. Abrir BEAST 1.6 y correr el .xml que generamos.

11. Observar que se generaron diferentes archivos output. \*BEAST estimó un árbol para cada gen por separado (excepto por los mitocondriales que habíamos ligado las particiones de los árboles). A medida que \*BEAST iba estimando los árboles de genes por separado, estimaba el árbol de especie más probable dados esos árboles de genes. Entonces, la estimación de árboles de especies se encuentra guardada en el archivo que tiene terminación .species.trees. Por último, hay un archivo .log que tiene los detalles de la estimación de los parámetros.
12. Abrir el archivo .log en Tracer. Examinar la validez del análisis. Sacar conclusiones.  
  
NOTA: Recordar que es importante realizar (al menos) 2 análisis independientes (usando exactamente el mismo .xml generado en BEAUti), y cargar ambos resultados .log en Tracer, para ver la convergencia del análisis en la estimación del likelihood y de los diferentes parámetros. Tener en cuenta que comprar en Tracer dos análisis distintos (en el que se varíen secuencias, número de individuos, o se cambien parámetros de los modelos), NO tiene sentido.
13. Abrir TreeAnnotator. Seleccionar como input el archivo terminado en .species.trees. Dar un nombre de output, asumir 10% de burnin y 0.5 de Posterior probability limit.
14. Abrir el árbol resultante en FigTree.