

## PRÁCTICA INFOMÁTICA FILOGENIA

### INFERENCIA FILOGENÉTICA

---

**Programa:** Para esta práctica se utilizarán los siguientes programas:

Geneious

<http://www.geneious.com/>

Geneious es un software comercial y por lo tanto esta practica se deberá hacer utilizando la versión de prueba que te deja utilizar el programa durante 15 días de forma gratuita. Es muy fácil de instalar y de obtener la licencia. Para proceder, primero instalar la ultima versión del software Geneious en los ordenadores del aula de tal manera que cuando se vaya a hacer la actividad ya esté el software instalado. Se puede bajar la ultima versión desde esta dirección:

<https://www.geneious.com/download/>

Cuando se empiece la practica abrir el programa y solicitar una licencia de prueba. Esto se tiene que hacer individualmente en cada ordenador y requiere de un nombre y una dirección de email para que te envíen los códigos de la licencia. Una vez activada la licencia el programa se podrá abrir y cerrar tantas veces como se quiera y utilizarlo durante los siguientes 15 días. Una misma licencia no sirve para varios ordenadores y por lo tanto se tiene que activar en cada ordenador. La activación del programa es muy simple ya que pones tus datos de contacto y te envían un email con las claves automáticamente y por lo tanto en pocos minutos ya puedes empezar a utilizar el programa.

**Ficheros de datos:** El fichero de datos lo tendréis que hacer vosotros mediante secuencias de GenBank <http://www.ncbi.nlm.nih.gov/>

**Objetivo:** Aprender a bajarse secuencias de GenBank, manipularlas, alinearlas e inferir un árbol filogenético de la **Superfamilia Hominoidea** utilizando datos moleculares. Las relaciones filogenéticas de Hominoidea son muy bien conocidas y por lo tanto esto nos permite saber si realmente el árbol filogenético obtenido ha salido bien. Junto con el tutorial se incluye el árbol filogenético de Hominoidea para ver que especies están incluidas, cual es el grupo externo de Hominoidea que utilizaremos para enraizar nuestra filogenia y cuales son las relaciones filogenéticas entre los diferentes géneros.

Para inferir la filogenia, nos bajaremos de GenBank los genomas mitocondriales completos accesibles de los 5 géneros de la Familia Hominidae (*Pan*, *Pongo*, *Gorilla* y *Homo*), incluyendo 6 de las 7 especies actuales (únicas excepciones Eastern Gorilla, *Gorilla beringei*, y la nueva especie de orangután de Sumatra, *Pongo tapanuliensis*). También incluiremos en los análisis una especie extinta de *Homo* (*Homo neanderthalensis*); dos especies de la Familia Hylobatidae (dos especies del Género *Hylobates*) y un representante de la Familia Cercopithecidae (una especie del Género *Papio*) que utilizaremos para enraizar el árbol filogenético (será nuestro grupo externo o **outgroup**). Se aconseja mirar el árbol esquema proporcionado con el fichero de prácticas como referencia.

1.- Abrir el programa Geneious y crear una carpeta dentro del directorio Local (*File>New Folder*) que se denomine **Filogenia**

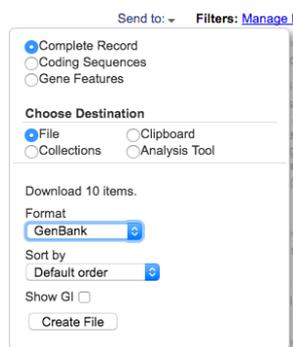
2.- A continuación, nos bajaremos los genomas mitocondriales para trabajar con ellos. A continuación tenéis un listado con los individuos y los códigos GenBank de los genomas que utilizaremos:

<b>Taxon</b>	<b>código de GenBank</b>
<u>Hominidae</u>	
<i>Gorilla gorilla gorilla</i> / Western Gorilla	NC_001645
<i>Gorilla gorilla gorilla</i> / Western Gorilla	NC_011120
<i>Pan paniscus</i> / pygmy chimpanzee	NC_001644
<i>Pan troglodytes</i> / chimpanzee	NC_001643
<i>Pongo abelii</i> / Sumatran orangutan	NC_002083
<i>Pongo pygmaeus</i> / Bornean orangutan	NC_001646
<i>Homo sapiens</i> / human	NC_012920
<i>Homo neanderthalensis</i> / Neandertal	NC_011137
<u>Hylobatidae</u>	
<i>Hylobates lar</i> / common gibbon	NC_002082
<i>Hylobates agilis</i> / agile gibbon	NC_014042
<u>Cercopitheidae</u>	
<i>Papio hamadryas</i> / hamadryas baboon	NC_001992

Lo más fácil es bajarlas de GenBank directamente. Ir a: <https://www.ncbi.nlm.nih.gov/> arriba a la izquierda, donde pone "all databases" seleccionar y buscar en el menú "nucleotides". A continuación, copiar el listado de los 11 genomas en la casilla y darle a "search". Podéis copiar y pegar el listado que se os pone a continuación para no tener que escribir todos los códigos de GenBank:

NC\_001645 NC\_011120 NC\_001644 NC\_001643 NC\_002083 NC\_001646 NC\_012920  
NC\_011137 NC\_002082 NC\_014042 NC\_001992

Una vez nos aparezcan en pantalla el listado con los 11 genomas mitocondriales (contadlos) los tenéis que bajar a vuestro ordenador. Arriba a la derecha de la pantalla buscad "send to" y al seleccionarlo os aparecerá un menú que debéis modificar para bajaros todas las 11 secuencias en formato GenBank a un único fichero. Seguid el ejemplo de la foto insertada y dadle a "create file"



Ir al programa Geneious e importad las secuencias haciendo: *File>Import> From file*. Dejad en “autodetect format” y dadle al “OK”

**¡¡ATENCIÓN!!** Cuando os pregunte si queréis mantener las secuencias juntas o separadas tenéis que seleccionar “keep separated”

3.- Si no habéis importado las secuencias desde dentro de la carpeta Filogenia, arrastrad las secuencias dentro de la carpeta. Una vez en nuestra carpeta de trabajo, podemos abrir un genoma y nos aseguraremos que la ventana *annotations* (a la derecha de la pantalla, dentro del menú *general*) esté seleccionada para poder ver todas las anotaciones de las diferentes regiones del genoma mitocondrial. Mirad los genomas mitocondriales con los que trabajaréis. Todos los genomas menos uno (*Gorilla gorilla gorilla* – NC\_011120) tienen un icono de un círculo verde que quiere decir que es un genoma circular y que está completo. El genoma NC\_011120 aparece con un icono de una doble hélice que quiere decir que no se ha podido cerrar el círculo. Esto es debido a que no se ha secuenciado la región control y no se ha podido ensamblar el genoma circular entero y por este motivo os parecerá como una secuencia lineal.

4.- Ahora como ejercicio queremos inferir una filogenia con el gen mitocondrial Subunidad 4 de la NADH Deshidrogenasa, más conocido como ND4. Es un gen que codifica una proteína.

5.- Para extraer del genoma mitocondrial que os habéis bajado un gen en concreto es muy fácil. En primer lugar, cread una carpeta llamada ND4 dentro de la carpeta Filogenia (*File>New Folder*).

A continuación, podéis utilizar dos sistemas para extraer el gen:

5.1.- En las opciones del visualizador del genoma seleccionar “sequence viewer”. Veréis el genoma mitocondrial anotado. Seleccionar el gen a extraer con el cursor, poniéndose encima de la zona anotada en verde (ND4 gene), clicar la zona y una vez veáis que esta seleccionada en el cuadro de dialogo de la pantalla clicar en *Extract*. Os pedirá un *extraction name*, decidle OK por defecto y ya esta.

5.2.- En las opciones del visualizador del genoma seleccionar *annotations* y veréis una tabla con el listado de todos los genes. Seleccionar el gen a extraer y en el cuadro de dialogo de la pantalla clicar en *Extract*. Os pedirá un *extraction name*. Os pedirá un *extraction name*, decidle OK por defecto y ya esta.

Una vez extraído el gen y cambiado el nombre arrastrar el gen a la carpeta ND4

6.- Alinear las 11 secuencias del gen ND4 de la siguiente manera:

Seleccionar las 11 secuencias de ND4 e ir a: *Tools>Align/Assemble>Multiple Align*. Seleccionar "MUSCLE Alignment" y dejar todos los parámetros por defecto. Dadle a OK y lo alineará muy rápido. Mirad el resultado que es un fichero llamado "Nucleotide alignment". Veréis las diferentes mutaciones entre las especies y podréis observar ya a simple vista que por ejemplo los especímenes del género *Homo* son más parecidos entre ellos que con el resto.

ESTE ES NUESTRO ALINEAMIENTO! Cada posición es un carácter y cada letra (A, T, C, G) es un estado de carácter.

A parte de ver las mutaciones en el alineamiento también podéis aprovechar para traducir a aminoácidos el alineamiento entero. Para hacer esto seleccionad el fichero del alineamiento (Nucleotide alignment) y elegir "translate". Os abrirá un cuadro de diálogo donde podréis cambiar si queréis (no es necesario para la practica) el nombre del fichero, escoger el código genético (recordad que existen muchos códigos genéticos diferentes dependiendo de los grupos animales). En nuestro caso, como trabajamos con vertebrados escogeremos el "Vertebrate mitocondrial", pero por ejemplo si estuviésemos trabajando con invertebrados escogeríamos el código genético del "Invertebrate mitocondrial".

7.- Inferir filogenias con *bootstrap* (una medida de soporte) para el gen ND4. Se incluye junto con la práctica un pequeño tutorial sobre qué son los modelos evolutivos y los diferentes métodos de inferencia filogenética. De todos modos no es necesario profundizar sobre estos temas, pues según el nivel y la práctica se puede seguir perfectamente con las instrucciones que se dan a continuación.

El objetivo es inferir un árbol filogenético utilizando el método de la máxima verosimilitud con el software de inferencia filogenética que viene instalado por defecto con el programa y que se llama Geneious Tree Builder. Es un método basado en distancias genéticas. Si queremos utilizar métodos de máxima verosimilitud (*maximum-likelihood*) podéis instalar el *plugin* del software PhyML haciendo lo siguiente: Dentro de Geneious ir a: *Tools>Preferences>Plugins and Features* y una vez allá seleccionar "install" para el programa PHYML. Instalarlos. Como veréis hay muchos otros plugins para hacer muchas otras cosas (MAFFT para alinear, etc...).

Para inferir el árbol haced lo siguiente:

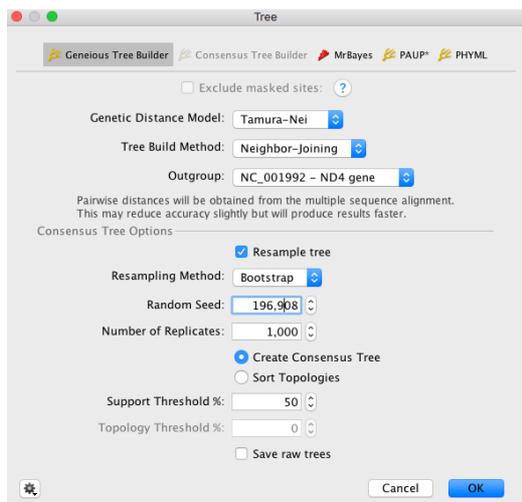
7.1.- Seleccionar el fichero Nucleotide alignment

7.2.- Ir al icono amarillo que pone Tree y seleccionarlo. Se os abrirá un cuadro de diálogo

7.3.- Escoged como método Geneious Tree Builder

7.4.- Aseguraros que esté todo como el pantallazo de abajo con la diferencia del "Rabdom seed" que se puede dejar por defecto el que os salga a vosotros (el numero no ha de coincidir, pero si queréis lo podéis cambiar). Importante escoger como "outgroup" (grupo externo para en raizar el árbol filogenético) el gen ND4 del *Papio hamadryas* que habéis extraído del genoma mitocondrial (NC\_001992 – ND4).

7.5.- Dadle a "OK"



El resultado es un árbol filogenético enraizado con el *Papio hamadryas*.

8.- En el cuadro de diálogo en la parte derecha de la pantalla del árbol seleccionar:

- *Show tip labels*>*Display: Common name*. Se os cambiaran los códigos de las secuencias por los nombres de las especies y esto hará que el árbol sea más visual
- *Show Branch Labels*>*Display: Consensus support(%)*

## EJERCICIOS

Mirar la filogenia y los valores de soporte (*consensus support (%)*) y rellenar la siguiente tabla poniendo los valores correspondientes a cada grupo.

	ARBOL FILOGENÉTICO
<b>CLADOS (grupos monofiléticos)</b>	<b><i>Consensus support(%)</i></b>
HOMINIDAE	
GORILLA+PAN+HOMO	
PAN+HOMO	
PAN	
HOMO	