

# PRÀCTICA INFOMÀTICA - FILOGÈNIA

## INFERÈNCIA FILOGENÈTICA

### Programa informàtic

---

Per aquesta pràctica s'utilitzarà el següent programa informàtic:

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

Geneious és un programari comercial i per tant, aquesta pràctica es realitzarà amb la versió de prova que et deixa utilitzar el programa durant 15 dies de forma gratuïta. És molt fàcil d'instal·lar i d'obtenir la llicència. Per procedir, primer instal·lar la darrera versió del programari Geneious en els ordinadors de l'aula de tal manera que quan es vagi a fer l'activitat tinguem el programari instal·lat. Es pot baixar la darrera versió des d'aquesta adreça: <https://www.geneious.com/download/>

Quan comenci la pràctica, obrir el programa i sol·licitar una llicència de prova. Això s'ha de fer individualment a cada ordinador i requereix d'un nom i una adreça de correu electrònic perquè t'enviïn els codis de la llicència. Un cop activada la llicència el programa es podrà obrir i tancar tantes vegades com es vulgui i utilitzar-lo durant els següents 15 dies. Una mateixa llicència no serveix per a diversos ordinadors i per tant, s'ha d'activar en cada ordinador de forma independent. L'activació del programa és molt simple ja que poses les teves dades de contacte i t'envien un correu electrònic amb les claus automàticament i per tant, en pocs minuts ja pots començar a utilitzar el programa.

### Fitxers de dades

---

El fitxer de dades l'haureu de fer vosaltres mitjançant seqüències de GenBank:

<http://www.ncbi.nlm.nih.gov/>

### Objectiu

---

L'objectiu d'aquesta pràctica és aprendre a baixar seqüències del GenBank, manipular-les, alinear i inferir un arbre filogenètic de la Superfamília **Hominoidea** utilitzant dades moleculars. Les relacions filogenètiques d'Hominoidea són molt ben conegudes i per tant això ens permet saber si realment l'arbre filogenètic obtingut ha sortit bé. Juntament amb el tutorial s'inclou l'arbre filogenètic d'Hominoidea per veure quines espècies estan incloses, qui és el grup extern d'Hominoidea que utilitzarem per arrelar la nostra filogènia, i quines són les relacions filogenètiques entre els diferents gèneres.

## Procediment

Per inferir la filogènia, ens baixarem de GenBank els genomes mitocondrials complets de representants de tots els 4 gèneres de la Família Hominidae (*Pongo*, *Gorilla*, *Pan*, i *Homo*), incloent 7 de les 8 espècies actuals (única excepció la nova espècie d'orangutan de Sumatra, *Pongo tapanuliensis*). També inclourem en les anàlisis una espècie extinta d'*Homo* (*Homo neanderthalensis*); dues espècies de la Família Hylobatidae (dues espècies del Gènere *Hylobates*) i un representant de la Família Cercopithecidae (una espècie del Gènere *Papio*) que utilitzarem per arrelar l'arbre filogenètic (serà el nostre grup extern o *outgroup*). S'aconsella mirar l'arbre esquema proporcionat amb el fitxer de pràctiques. Seguir els següents passos:

### A- Baixar les seqüències dels genomes

- 1.- Obrir el programa Geneious i crear una carpeta dins del directori Local (*File > New Folder*) que es denomini **Filogenia**
- 2.- A continuació, ens baixarem els genomes mitocondrials per treballar amb ells. Aquí teniu un llistat amb els codis GenBank dels genomes que utilitzarem:

Taxó	Codi de GenBank
<b><u>Hominidae</u></b>	
<i>Gorilla gorilla gorilla</i> / Western gorilla	NC_001645
<i>Gorilla beringei graueri</i> / Eastern lowland gorilla	KF914213
<i>Pan paniscus</i> / Pygmy chimpanzee (Bonobo)	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
<b><u>Hylobatidae</u></b>	
<i>Hylobates lar</i> / Common gibbon	NC_002082
<i>Hylobates agilis</i> / Agile gibbon	NC_014042
<b><u>Cercopithecidae</u></b>	
<i>Papio hamadryas</i> / Hamadryas baboon	NC_001992

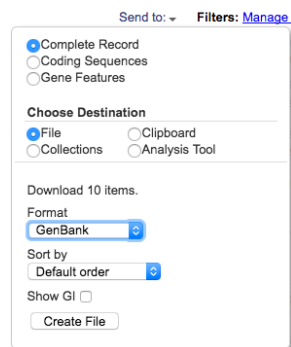
El més fàcil és baixar-se les seqüències de GenBank directament:

- Anar a: <https://www.ncbi.nlm.nih.gov/>
- A dalt a l'esquerra, on posa "all databases" seleccionar i buscar al menú "nucleotides"
- A continuació, copiar el llistat dels 11 genomes separats per un espai i donar-li a "search"
- Podeu copiar i enganxar el llistat que se us posa a continuació per no haver d'escriure tots els codis de GenBank:

NC\_001645 KF914213 NC\_001644 NC\_001643 NC\_002083 NC\_001646 NC\_012920  
NC\_011137 NC\_002082 NC\_014042 NC\_001992

- Un cop ens apareguin en pantalla el llistat amb els 11 genomes mitocondrials (compteu-los o mireu el número que us dona a dalt) els heu de baixar al vostre ordinador:

A dalt a la dreta de la pantalla busqueu "send to" i us apareixerà un menú que heu de modificar per baixar-vos totes les 11 seqüències en format GenBank a un únic fitxer. Seleccionem "File" i seguim l'exemple de la foto inserida i a continuació seleccionar "create file" i guardem el fitxer.



- Anar al programa Geneious i importeu les seqüències fent: *File>Import> From file*. Quan us aparegui un missatge deixeu en "autodetect format" i dir "OK".

ATENCIÓ!! Quan us preguntin si voleu mantenir les seqüències juntes o separades, heu de seleccionar "keep separate"

3.- Si no heu importat les seqüències des de dins de la carpeta **Filogenia**, arrossegueu les seqüències dins de la carpeta. Un cop a la nostra carpeta de treball, podem obrir un genoma i ens assegurarem que la finestra "annotations" (a la dreta de la pantalla, dins del menú general) seleccionada per poder veure totes les anotacions de les diferents regions del genoma mitocondrial. Mireu els genomes mitocondrials amb què treballareu. Tots els genomes tenen una icona d'un cercle verd que vol dir que és un genoma circular i que està complet.

## B- Extreure un gen del genoma i alinear

Ara com a exercici volem inferir una filogènia amb el gen mitocondrial Subunitat 4 de la NADH deshidrogenasa, més conegut com ND4. És un gen que codifica una proteïna.

1.- Per extreure del genoma mitocondrial que us heu baixat **un gen en concret** és molt fàcil. En primer lloc, creeu una carpeta anomenada **ND4** dins de la carpeta **Filogènia** (*File> New Folder*).

2.- A continuació, podeu utilitzar aquest sistema visual per extreure el gen:

A les opcions del visualitzador del genoma seleccionar "Sequence View". Veureu el genoma mitocondrial anotat. Seleccionar el gen a extreure amb el cursor, posant-sobre de la zona anotada en verd (ND4 gene), cliqueu la zona i un cop veieu que aquesta està seleccionada, al quadre de diàleg de la pantalla clicar a "Extract". Us demanarà un "extraction name", digueu-li OK al que us dona per defecte i ja està. Repetir per la resta de genomes el mateix procés.

3.- Un cop extret el gen, arrossegau els gens a la carpeta ND4.

4.- Alinear les 11 seqüències del gen ND4 de la següent manera:

Seleccionar les 11 seqüències de ND4 i anar a: *Tools> Align / Assemble> Multiple Align*. Seleccionar "MUSCLE Alignment" i deixar tots els paràmetres per defecte. Doneu-li a OK i ho alinearà molt ràpid. Mireu el resultat que és un fitxer anomenat "Nucleotide alignment". Veureu les diferents mutacions entre les espècies i podreu observar ja a simple vista que ara els espècimens del Gènere *Homo* són més semblants entre ells que amb la resta.

AQUEST ÉS EL NOSTRE ALINIAMENT! Cada posició és un caràcter i cada lletra (A, T, C, G) és un estat de caràcter.

A part de veure les mutacions en l'alineament també podeu aprofitar per traduir a aminoàcids l'alineament sencer. Per fer això, seleccioneu el fitxer de l'alineament (Nucleotide alignment) i triar "translate". Us obrirà un quadre de diàleg on podreu canviar si voleu (no cal per a la pràctica) el nom del fitxer, escollir el codi genètic (recordeu que hi ha molts codis genètics diferents depenent dels grups animals). En el nostre cas, com que treballem amb vertebrats escollirem el "Vertebrate mitochondrial", però per exemple, si treballéssim amb invertebrats escolliríem el codi genètic del "Invertebrate mitochondrial".

## C- Inferir filogènies

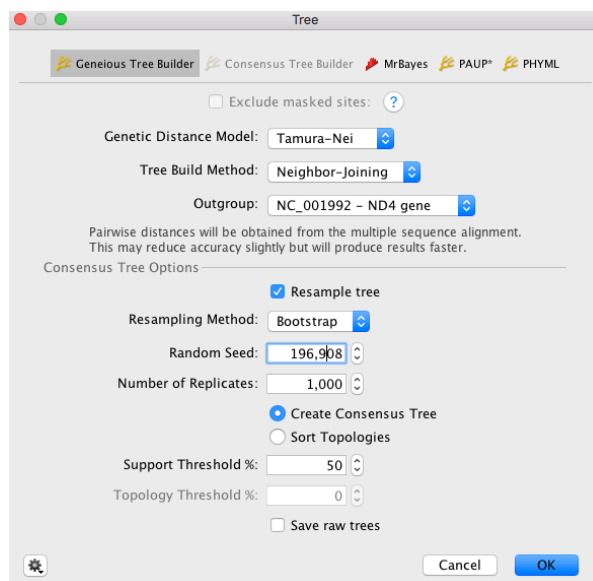
Inferirem filogènies amb *Bootstrap* (una mesura de suport) per al gen ND4.

S'inclou juntament amb la pràctica un petit tutorial sobre què són els models evolutius i els diferents mètodes d'inferència filogenètica. De totes maneres no cal aprofundir sobre aquests temes segons el nivell i la pràctica es pot seguir perfectament amb les instruccions que es donen a continuació.

L'objectiu és inferir un arbre filogenètic utilitzant el programa d'inferència filogenètica que ve instal·lat per defecte amb el programa i que es diu "Geneious Tree Builder". És un mètode basat en distàncies genètiques. Si volem utilitzar mètodes de màxima versemblança (*màxim-likelihood*) podeu instal·lar el *plugin* de programari PhyML fent el següent: Dins de Geneious anar a: *Tools*> *Preferences*> *Plugins and Features* i un cop allà seleccionar "install" per al programa PHYML. Instal·lar-los. Com veureu hi ha molts altres *plugins* per fer moltes altres coses (MAFFT per alinear, etc ...).

Per inferir l'**arbre filogenètic** feu el següent:

- 1.- Seleccionar el fitxer Nucleotide alignment
- 2.- Anar a la icona groga que posa "Tree" i apreteu la icona. S'obrirà un quadre de diàleg
- 3.- Escolliu com a mètode "Geneious Tree Builder"
- 4.- Assegureu que estigui tot com la captura de pantalla de sota amb la diferència del "Random seed" que es pot deixar per defecte el que us surti a vosaltres (el nombre no ha de coincidir, si voleu el podeu canviar). Important escollir com "outgroup" (grup extern per arrelar l'arbre filogenètic) el gen ND4 del Hamadryas baboon (*Papio hamadryas*) que heu extret del genoma mitocondrial (NC\_001992 - ND4).
- 5.- Doneu-li a "OK"



6.- El resultat es un arbre filogenètic arrelat Hamadryas baboon (*Papio hamadryas*).

7.- Un cop us mostri l'arbre, al quadre de diàleg de la part dreta de la pantalla de l'arbre seleccionar (si no està seleccionat ja per defecte):

- *Show tip labels*>*Display: Common name*. Se us canviaran els codis de les seqüències pels noms comuns de les espècies i això farà que l'arbre sigui més visual.
- *Show Branch Labels*>*Display: Consensus support(%)*. Mostrarà el suport de les diferents relacions filogenètiques de 0 (no hi ha suport) a 100 (màxim suport). Es considera que una agrupació està ben suportada quan el valor de *Bootstrap* és igual o superior a 70.

## EXERCICI

Mirar la filogènia i els valors de suport (consensus support (%)) i omplir la següent taula posant els valors corresponents a cada grup.

	ARBRE FILOGENÈTIC
CLADES (grups monofilètics)	<i>Consensus support</i> (%)
HOMINIDAE	
GORILLA+PAN+HOMO	
PAN+HOMO	
PAN	
HOMO	