

Desarrollo de pipelines

Tutorial adaptado de Chen et. al. (2025). Comparative analysis of the mitochondrial genomes of the softshelled turtles *Palea steindachneri* and *Pelodiscus axenaria* and phylogenetic implications for Trionychia.

[Acceder aqui](#)

⚠ Antes de iniciar recuerde definir su carpeta de trabajo.

Requerimientos

Puede trabajarse con WSL o Ubuntu. El código se puede adaptar a Mac. Software

- VS Code Studio (Extensiones a instalar: Markdown, WSL (solo si está en windows), Display PDF in VSCode (Tomoki))

Software en terminal

- Conda, miniconda and miniforge
- Dos2Unix (this requires your Ubuntu/Linux password)

```
sudo apt install dos2unix
```

- MAFFT

```
conda install bioconda::mafft
```

Ejecutables

Debe definir una carpeta para guardar los programas.

```
mkdir programas
```

- Gblocks

```
#it creates a default folder
wget https://ponce.cc/slackware/sources/repo/Gblocks_Linux64_0.91b.tar.Z
uncompress Gblocks_Linux64_0.91b.tar.Z
tar xvf Gblocks_Linux64_0.91b.tar
rm Gblocks_Linux64_0.91b.tar
```

- FASTconGENE

```
git clone https://github.com/PatrickKueck/FASconCAT-G.git
```

- FigTree

```
cd programas/  
wget  
https://github.com/rambaut/figtree/releases/download/v1.4.4/FigTree.v1.4.4.zip  
unzip FigTree.v1.4.4.zip  
rm FigTree.v1.4.4.zip
```

- IQTREE-3

```
wget https://github.com/iqtree/iqtree3/releases/download/v3.0.1/iqtree-3.0.1-  
Linux.tar.gz  
tar xvf iqtree-3.0.1-Linux.tar.gz  
rm iqtree-3.0.1-Linux.tar.gz
```

Descarga y procesamiento de las secuencias

La lista de secuencias está disponible en el artículo de Chen et. al. (2025). Utilizaremos solo los genes codantes de proteínas (PCGs) mitocondriales.

🔍 Revisar el artículo para encontrar la lista de secuencias. ¿Cuántas secuencias son? ¿Cuántos genes mitocondriales espera encontrar por especie?

1. Descargar las secuencias del GenBank. Recordar que solo vamos a trabajar con PCGs. Copiar su archivo a su carpeta de trabajo y si quiere renombrar.

¿Qué comando utilizaría para saber cuántas secuencias son? ¿Qué formato tiene el archivo?

2. *Pasar a un sola línea*

Usar el script `onerow_mitogenome.sh`

3. *Renombrar el archivo fasta*

Usar el script `rename.sh`

Verificar que el número de genes por especie, corresponda al archivo original. ¿Qué comando usaría?

```
while read -r gene || [[ -n "$gene" ]]; do count=$(grep -ow "$gene" renamed.fasta  
| wc -l); printf "%s\t%s\n" "$gene" "$count"; done < input/mito_genes.txt
```

- Siempre realizar una revision manual.

Nota: Cuando parece no poder ejecutarse, convertir el archivo a Unix. Usualmente los archivos generados en windows tienen caracteres innecesarios no visibles para el usuario. Puede usar `dos2unix` o `sed -i 's/\r/'`

your_file.txt

4. Separar cada PCG a un solo archivo Usar el script `separate_files.sh`. Luego, para eliminar toda informacion que este despues del ".GEN" usar.

```
sed -i 's/\.*\\.*/' *.fasta
```

¿Cómo verificar que tenemos el número de archivos esperado?

```
find . -type f | wc -l
```

5. Uniformizar la longitud del nombre a 9 caracteres *¿por qué?*

```
for file in genes/*.fasta; do sed -E '/^>/ {s/^(>)([^\s:]{8})$/\1\2_/;}'  
"$file" > temp && mv temp "$file"; done
```

Los siguientes pasos debe realizarse a cada uno de los archivos.

Alineamiento

Alinearemos con MAFFT.

```
#Create a folder to include the aligned files  
mkdir aligned  
#Run MAFFT  
for i in genes/*.fasta; do mafft --quiet "$i" > "aligned/${basename  
"${i%.fasta}").aligned.fasta"; done
```

Usar el script `onerow.sh` para dar formato a una sola linea

Eliminación de caracteres ambiguos

Usaremos el programa `GBLOCKS`.

```
for i in gblocks/*.fasta; do programas/Gblocks_0.91b/Gblocks $i -t=d -b5=n -p=y;  
done
```

Vamos a reorganizar la carpeta, ingresar a la carpeta online.

```
cd gblocks  
mkdir fasta-gb html
```

```
mv *.fasta-gb fasta-gb/  
mv *.htm html/
```

Vamos a transformar de `.fasta-gb` a `.fasta`. Usaremos el script `fasta-gb2fasta.sh`.

Concatenar

Usaremos `FASconCAT-G`. Debe ubicarse en la carpeta que contiene a los genes alineados y trimeados con GBLOCKS.

```
cd gblocks/onerow/  
perl ../../programas/FASconCAT-G/FASconCAT-G_v1.06.1.pl -p -p -s
```

Mover los archivos generados a la carpeta `concatenated/`

Análisis filogenético

Vamos a correr el archivo en IQTREE3.

```
#help for iqtree3  
programas/iqtree-3.0.1-Linux/bin/iqtree3 -h  
#comand for iqtree3  
programas/iqtree-3.0.1-Linux/bin/iqtree3 -s concatenated/FcC_supermatrix.phy -m  
MFP -B 1000
```

Visualizar el árbol filogenético

Podemos usar el ejecutable FigTree. El archivo a abrir termina en `.treefile`. Podemos renombrar los nombres en el archivo usando `sed`.

```
sed -i 's/code1/long_name1/g' yourfilename.T1;
```

Siguientes pasos

💬 ¿Qué puntos podrían mejorarse? ¿Qué software alternativos podrían usarse?