

# Comparación de especies y generación de árboles evolutivos de felinos y camélidos del Perú y de África.

## Integrantes:

Reewos Erwin Talla Chumpitaz

Konrad Benjamin Trejo Chavez

### 1. Resumen ejecutivo:

El proyecto consiste en la creación y desarrollo de una aplicación, basada en lenguaje Python y utilizando librerías y software Open Source, que permita el análisis bioinformático entre especies de la fauna peruana, específicamente la clase Mammalia (mamíferos). Sin embargo, debido a la gran cantidad de animales encontrados, se decidió por analizar específicamente la familia Felidae (felinos), y compararlos con otras especies que se encuentran en otros lugares, específicamente en África, ya que es el lugar donde hay una mayor cantidad de felinos. Además, se analizará la familia Camelidae (camélidos) presentes del Perú, así como las especies que se encuentran en África para su comparación. A partir de los resultados, elaborar árboles filogenéticos y analizar las relaciones evolutivas.

### 2. Abstract:

The project consists in the creation and development of an application, based on Python language and using libraries and Open Source software, which allows bioinformatic analysis among species of Peruvian fauna, specifically the Mammalia class (mammals). However, due to the large number of animals found, it was decided to specifically analyze the Felidae family (felines), and compare them with other species found elsewhere, specifically in Africa, since it is the place where there is a greater amount of cats. From the results, develop phylogenetic trees and analyze evolutionary relationships.

### 3. Descripción del Proyecto:

Se utilizó primeramente una base de datos de animales endémicos del Perú [1] y un csv de genomas disponibles de mamíferos descargado del NCBI [2] para comparar y extraer los animales que tienen su genoma completo en la base de datos del NCBI para su análisis. Los cuales son los siguientes:

- ❖ Felidae (genoma completo):
  - Puma concolor (Puma)
  - Panthera onca (Jaguar)
- ❖ Camelidae (genoma completo):
  - Vicugna pacos (Alpaca)

Además, se optó por analizar otras especies que no tienen su genoma completo, pero que se tiene registrado proteínas o genes. Para ello se utilizó la página de Taxonomy [5] y se comprobó si existen registros de los felinos encontrados de [1]. También se analizó para el caso de los camélidos.

- ❖ Felidae (con genes y proteínas disponibles):
  - Leopardus jacobitus (Gato andino)
  - Leopardus colocolo (Gato del pajonal)
  - Leopardus pardalis (Ocelote)
  - Leopardus tigrinus (Tigrino)

- *Leopardus wiedii* (Huamburushu o Margay)
- *Puma concolor* (Puma)
- *Puma yagouaroundi* (Yahuarundi)
- *Panthera onca* (Jaguar)

❖ **Camelidae (con genes y proteínas disponibles):**

- *Lama guanicoe* (Guanaco)
- *Lama glama* (Llama)
- *Vicugna vicugna* (Vicuña)
- *Vicugna pacos* (Alpaca)

De los felinos, se analizarán lo siguiente:

- Genes:
  - CYTB

De los camélidos, se analizarán:

- Genes:
  - CYTB

A modo de comparación, se tomará en cuenta otras especies que pertenecen al continente africano. Para ello primero evaluaremos los felinos y camélidos que se encuentren en África. Se tiene entre ellos:

❖ **Felidae (genoma completo):**

- *Acinonyx jubatus* (Guepardo o Chita)
- *Panthera pardus* (Leopardo)

❖ **Camelidae (genoma completo):**

- *Camelus dromedarius* (Dromedario)

Se tomará en cuenta otras especies que no tengan genomas completos, pero que se tenga registrado genes y proteínas de la especie.

❖ **Felidae (con genes y proteínas disponibles):**

- *Leptailurus serval* (Serval)
- *Acinonyx jubatus* (Guepardo o Chita)
- *Panthera pardus* (Leopardo)
- *Panthera leo* (León)

❖ **Camelidae (con genes y proteínas disponibles):**

- *Camelus dromedarius* (Dromedario)

Para este proyecto se utilizará inicialmente solo el CTYB. El citocromo b es una proteína que se encuentra en las mitocondrias de las células eucariotas. Funciona como parte de la cadena de transporte de electrones y es la subunidad principal de los complejos de citocromo bc1 y b6f transmembrana.

Se descargó las secuencias del gen CTYB en formato fasta, el cual se separó en carpetas para distinguirlos (Felidae y Camelidae).

Para el desarrollo de la interfaz gráfica se utilizó QT, en el cual consta de:

- Toolbar: Para realizar algunas acciones adicionales.
- Título: Título del proyecto.
- Sección de selección: Donde se puede seleccionar entre especies, genes y proteínas, y en se muestra una lista de todos los elementos disponible.
- Sección de botones: En esta parte están los botones que realizan las acciones de “Alineamiento múltiple” y “Generación de árbol filogenético”.
- Sección de salida: Aparecerá mensajes los cuales muestran el avance de los procesos.



Figura 1. Interfaz gráfica desarrollada.

Se utilizó el alineamiento múltiple con la ayuda del programa CLUSTALW, el cual nos brinda dos archivos con extensiones .aln y .dnd, donde el primero es un archivo donde se encuentra el alineamiento de las secuencias. El segundo es un archivo del dendograma generado para la creación del árbol filogenético.

```

1 CLUSTAL 2.1 multiple sequence alignment
2
3
4 NC_028314.1_14922-16061 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATCAACCACTC
5 NC_028322.1_14805-15944 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATTAATCAATC
6 NC_028317.1_14883-16022 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATTAATCACTC
7 NC_028315.1_14836-15975 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATTAATCACTC
8 NC_028318.1_14898-16037 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATTAATCACTC
9 NC_028316.1_14824-15963 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATTAATCACTC
10 NC_016470.1_15085-16224 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATTAATCACTC
11 NC_028311.1_14871-16010 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATTAATCACTC
12 NC_005212.1_15061-16200 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATTAATCACTC
13 NC_028302.1_14855-15994 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATTAATCACTC
14 NC_010641.1_15127-16266 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATTAATCACTC
15 NC_022842.1_15045-16184 ATGACCAACATTGAAAAATCACACCCCTCTTATTAATAATTATTAATCACTC
16 NC_012102.1_14161-15300 ATGACCAATATCCGAAAGTCCACCCCACTACTAAAAATTGTCAATAACGC
17 NC_011822.1_14161-15300 ATGACCAATATCCGAAAGTCCACCCCACTACTAAAAATTGTCAATAACGC
18 NC_013558.1_14161-15300 ATGACCAACATCCGAAATCCACCCCGCTACTAAAAATTGTCAACACGC
19 NC_002504.1_14154-15293 ATGACCAATATCCGAAAGTCCACCCCACTACTAAAAATTGTCAATAACGC
20 NC_009849.1_14156-15295 ATGACCAACATCCGAAATCCACCCCACTCTTAAAAATTATATAACGACGC
21 *****
22
23 NC_028314.1_14922-16061 ATTCAATCGATCTACCTACCCATCCAACATCTCAGCATGATGAAACTTTG
24 NC_028322.1_14805-15944 ATTCAATCGATCTACCTACCCATCCAACATCTCAGCATGATGAAACTTTG
25 NC_028317.1_14883-16022 ATTCAATCGATCTACCTACCCATCCAACATCTCAGCATGATGAAACTTTG
26 NC_028315.1_14836-15975 ATTCAATCGATCTACCTACCCATCCAACATCTCAGCATGATGAAACTTTG
27 NC_028318.1_14898-16037 ATTCAATCGATCTACCTACCCATCCAACATCTCAGCATGATGAAACTTTG
28 NC_028316.1_14824-15963 ATTCAATCGATCTACCTACCCATCCAACATCTCAGCATGATGAAACTTTG
29 NC_016470.1_15085-16224 ATTCAATCGATCTACCTACCCATCCAACATCTCAGCATGATGAAACTTTG
30 NC_028311.1_14871-16010 ATTCAATCGATCTACCTACCCATCCAACATCTCAGCATGATGAAACTTTG
31 NC_005212.1_15061-16200 ATTCAATCGATCTACCTACCCATCCAACATCTCAGCATGATGAAACTTTG

```

Figura 2. Contenido del archivo .aln.

```

1 (
2 (
3 (
4 (
5 NC_009849.1_14156-15295:0.07926,
6 (
7 (
8 (
9 NC_012102.1_14161-15300:0.01360,
10 NC_011822.1_14161-15300:0.01447)
11 :0.01681,
12 NC_013558.1_14161-15300:0.03056)
13 :0.02217,
14 NC_002504.1_14154-15293:0.05721)
15 :0.03192)
16 :0.06927,
17 (
18 (
19 NC_028302.1_14855-15994:0.04143,
20 NC_010641.1_15127-16266:0.04191)
21 :0.00744,
22 NC_022842.1_15045-16184:0.04782)
23 :0.02524)
24 :0.00709,
25 (
26 NC_005212.1_15061-16200:0.05406,
27 (
28 NC_016470.1_15085-16224:0.05414,
29 NC_028311.1_14871-16010:0.04761)
30 :0.00515)
31 :0.00454)
32 :0.00201,
33 (
34 (
35 (
36 NC_028314.1_14922-16061:0.03735,
37 NC_028322.1_14805-15944:0.03721)
38 :0.00373.

```

Figura 3. Contenido del archivo .dnd.

Para la creación del árbol filogenético, se utilizará el archivo .dnd como base, ya que contiene de manera estructurada del árbol junto con sus distancias. Sin embargo, al estructurar el árbol filogenético solo aparece los ID de las secuencias de las especies. Para ello se implementó una función que traduzca los ID a los nombres de los animales. De esta manera, el dendrograma ya contiene correctamente los nombres. Para la visualización en otra ventana del árbol, se utilizó la función de “Draw” del módulo “Bio.Phylo” el cual se encuentra en BioPython.

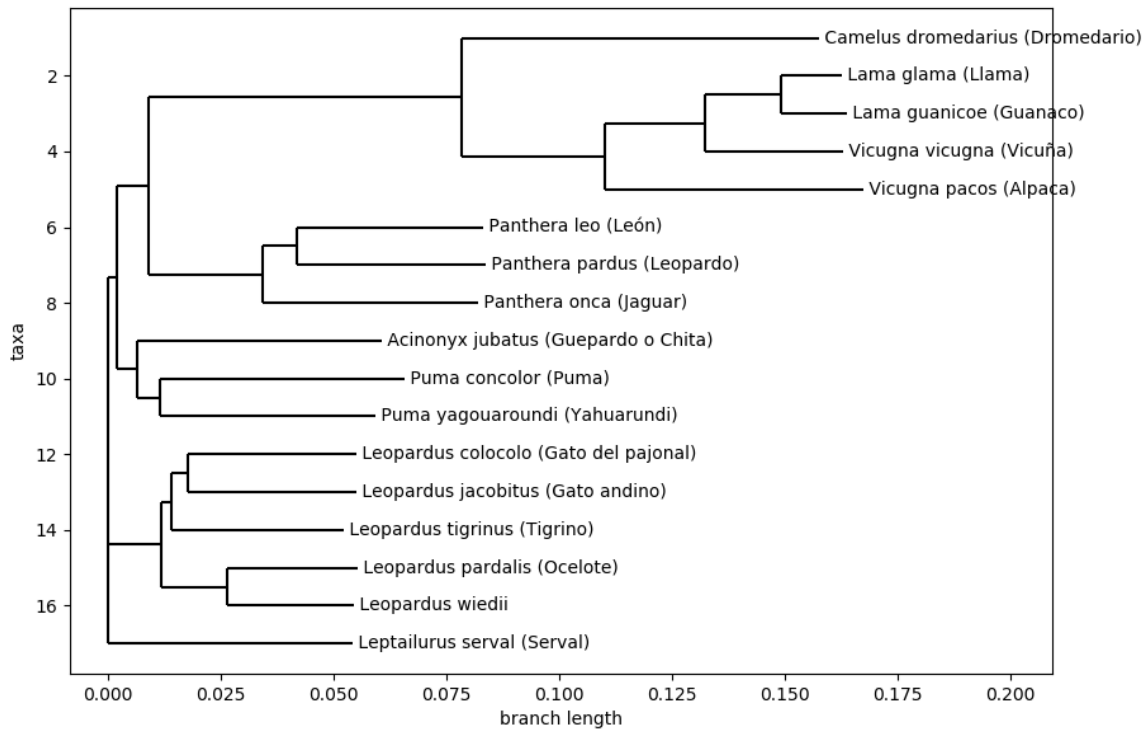


Figura 4. Árbol filogenético resultante del alineamiento del CTYB

## 2.1 Alcance del proyecto

- Con este proyecto deseamos llegar a realizar árboles filogenéticos de las principales especies del Perú, en este caso mamíferos de las familias Camelidae y Felidae.
- Llegar a clasificar las especies endémicas del Perú.

## 2.2 Cronograma

Fecha	Actividad
Lunes 23 Set - Viernes 27 set	<ul style="list-style-type: none"> <li>• Investigar las secuencias genéticas y proteínas de especies endémicas del Perú disponibles en NCBI.</li> <li>• Revisar la documentación de las Librerías y el Software a usar en el proyecto (BioPython, Clustalw, TKinter, PyQT)</li> <li>• consensuar las especies a emplear en el análisis filogenético.</li> <li>• Consensuar las proteínas a analizar.</li> </ul>

Lunes 28 Oct - Viernes 01 Nov	<ul style="list-style-type: none"> <li>• Propuesta final consensuada las proteínas y genes a usar en los camélidos y felinos.</li> <li>• Obtención de las secuencias genéticas en formato fasta de las especies endémicas y sus comparables en el mundo.</li> <li>• Desarrollo de interfaces en Tkinter y QT (Básicas) para el análisis y la generación de árboles evolutivos de especies camelidae y felidae</li> </ul>
Lunes 18 Nov - Viernes 22 Nov	<ul style="list-style-type: none"> <li>• Realización y análisis de alineamientos de genes y proteínas para la familia Camelidae</li> <li>• Realización y análisis de alineamientos de genes y proteínas para la familia Felidae</li> <li>• Integración de la UI desarrollado en QT con los controladores(clases y métodos en python que analizara las secuencias) para una mejor usabilidad del software</li> </ul>
Lunes 25 Nov - Viernes 29 Nov	<ul style="list-style-type: none"> <li>• Análisis y generación de Árboles Filogenéticos para la familia Camelidae</li> <li>• Análisis y generación de Árboles Filogenéticos para la familia Felidae</li> <li>• Pruebas e integración final del proyecto.</li> </ul>

#### 4. Algoritmos e implementación computacional:

Los códigos a utilizar son:

- Análisis y extracción de animales coincidentes: extractor\_coincidencias.py\*
- Algoritmos de Clustal:
  - Neighbor joining
  - Pairwise alignment

Herramientas computacionales a utilizar son:

- Python (librerías Biopython, etc.)
- Clustal
- QT (como interfaz gráfica)

#### 5. Resultados:

Los resultados son muy aproximados a los teóricos., ya que en el árbol filogenético se muestra que se agrupa por una parte los camélidos:

[5]: <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Root>