

# Práctica 2

## Bioinformática

José Manuel Sánchez Aquilué 759267

26 de febrero de 2021

### 1. Introducción.

El informe muestra como utilizar Clustal Omega para encontrar alineamientos sobre las secuencias descargadas en la práctica anterior. A lo largo de la memoria se analizaran los resultados obtenidos y se trabajará sobre distintos ficheros y con distintos parámetros. Por otro lado, los alineamientos resultantes serán visualizados a través de la herramienta Jalview.

### 2. Alineamientos de secuencias de DNA.

#### 2.1. Ejecución.

Antes de nada, nos disponemos a descargar aquellos resultados de búsqueda que no fueron pedidos expresamente en la práctica anterior. Como son demasiadas secuencias, nos quedaremos únicamente con las diez primeras de cada consulta.

En esta práctica utilizaremos Clustal Omega en local, desde una maquina con ubuntu. Se ha escrito un script en bash que ejecuta el algoritmo con todas las secuencias de la práctica anterior y con distintas iteraciones. EL algoritmo almacena por un lado la matriz de distancia y por otro el alineamiento en formato FASTA. En el caso del alineamiento con DNA mitocondrial humano, no se ha testado con un número diverso de iteraciones porque con cero iteraciones ya mostraba buenos resultados y con dos tardaba veinte minutos en ejecutarse.

Con las secuencias procedentes de seres humanos (apartado 1 de la práctica anterior), observamos que a mayor número de iteraciones, peores resultados. Las matrices de distancia se pueden consultar en el apéndice A. En general las distancias son bastante grandes, así que podemos deducir que son secuencias dispares y no hay alineamientos largos. Aunque hay casos en los que la distancia es bastante corta como NM\_001330136.3 con NM\_001393566.1 y algún otro. Si hacemos las pruebas con secuencias del cromosoma X, comprobamos, en primer lugar, una mejora sustancial en el tiempo de ejecución (figura 1). Además a

Apartados	0 iteraciones	2 iteraciones	4 iteraciones	8 iteraciones
1	4.44s	1m 9.47s	2m 21s	4m 50s
2	0.324s	2.09s	4.44s	7.49s
3	6.61s	1m 29s	3m 7s	6m 23s
6	3m 6s	-	-	-

Figura 1: Tiempos de ejecución en local.

medida que se itera los resultados mejoran ligeramente. Las distancias aquí son más uniformes, es decir, no hay dos secuencias a una distancia muy reducida en comparación con el resto.

Con las secuencias de DNA de mamíferos (apartado 3) conforme se itera, en algunos casos se reduce la distancia y en otros se agranda. A modo de ejemplo NM\_001393498.1 y NM\_001393497.1 llegan a reducir su distancia a cero, el nombre de ambas ya indica cierta similitud y por tanto facilidad a la hora de encontrar alineamientos. Por otro lado, NM\_013371.5 y NR\_171708.1 pasan de una distancia de 0.8 a 1 (disimilitud total). Por último, las secuencias de DNA mitocondrial muestran una distancia muy corta entre ellas con cero iteraciones. Lo que llama la atención es la manera en la que se dispara el tiempo de ejecución.

## 2.2. Visualización.

A través de la puntuación podemos hacernos una idea de los alineamientos. Sin embargo, lo interesante sería poder visualizarlos en un formato intuitivo. En un principio, se pensaba utilizar MSA, pero debido a los problemas que ocasionó, tanto su versión web como su versión en local, se terminó haciendo uso de Jalview para desempeñar esta tarea.

Al abrir un alineamiento, todas las secuencias aparecen en filas (Figura 2). Las letras simbolizan la base nitrogenada del nucleótido en cuestión y los guiones son los gaps. En la parte de abajo indica cual es la base que tienen en común la mayoría de secuencias en cada tramo y la cantidad de secuencias alineadas.

En secuencias con distinta longitud se ve que han introducido muchos gaps en las más pequeñas para cuadrar el alineamiento. Con el DNA mitocondrial de longitudes parecidas se aprecia un alineamiento casi total (Figura 3).

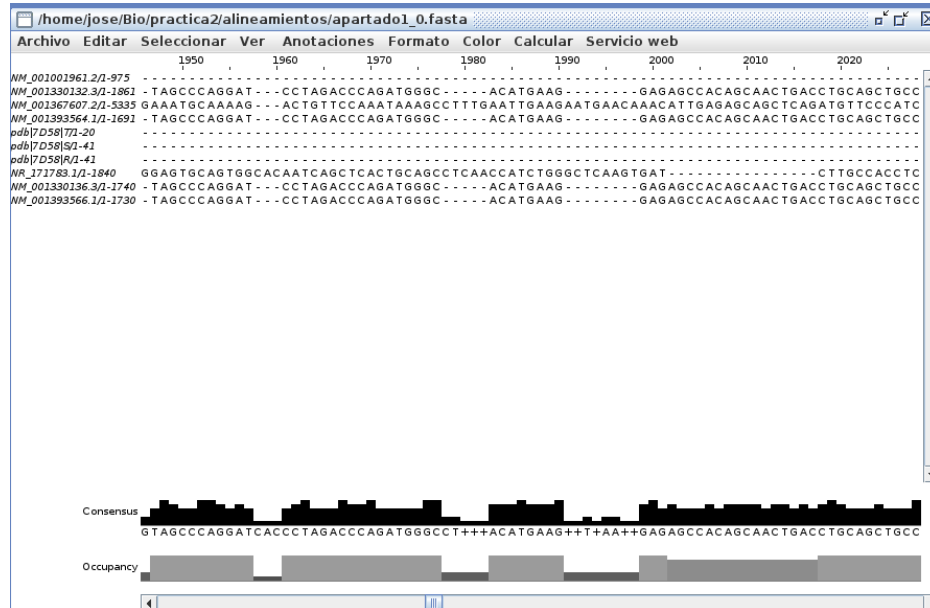


Figura 2: Visualización de alineamiento de DNA humano.

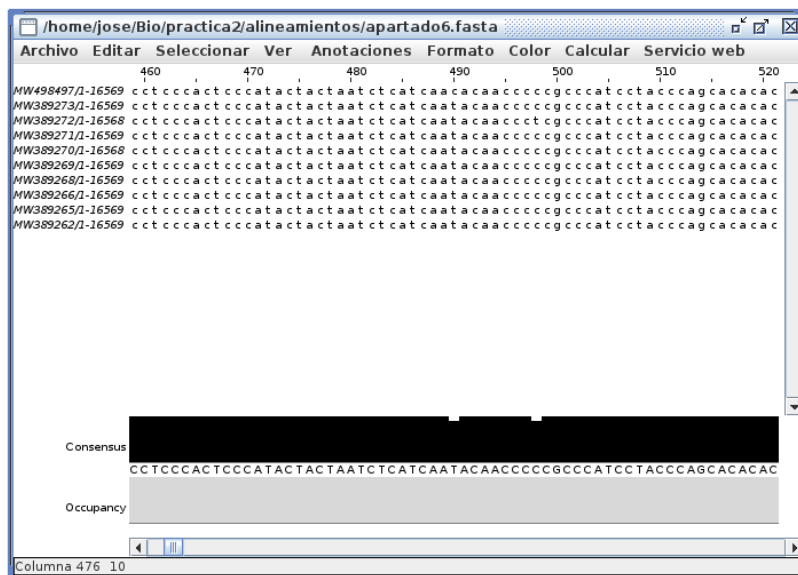


Figura 3: Visualización de alineamiento de DNA humano mitocondrial.

### 3. Alineamientos de secuencias de proteínas.

Al igual que en el apartado anterior, vamos a comenzar descargando las secuencias en formato fasta, siguiendo el procedimiento de la última práctica. Después programaremos el mismo script de antes, *mutatis mutandi*, para realizar esta tarea.

Las matrices de distancia (Apéndice B) nos indican que las proteínas pertenecientes al cromosoma Y de los primates pueden alinearse bastante bien y que cada iteración del algoritmo supone una mejora en la calidad del alineamiento. Todo lo contrario ocurre con las secuencias de proteínas descubiertas en lo que va de año, los resultados son bastante pésimos y van empeorando a medida que se itera.

En lo que concierne a tiempos de ejecución, en cuestión de segundos se han obtenido todos los resultados (Figura 4). Como es natural, el tiempo crece de manera directamente proporcional al número de iteraciones. También se aprecia que las secuencias del apartado 5 de la práctica anterior va más lento.

Como estamos trabajando con proteínas, lo primero que notamos es que las secuencias son de aminoácidos, en vez de nucleótidos. Jalview nos muestra la calidad del alineamiento basándose en la puntuación BLOSUM 62, así como la conservación. Visualmente se aprecia mejor las diferencias entre alineamientos.

Apartados	0 iteraciones	2 iteraciones	4 iteraciones	8 iteraciones
4	0.29s	1.46s	2.80s	5.17s
5	0.42s	4.71s	8.99s	17.41s

Figura 4: Tiempos de ejecución en local.

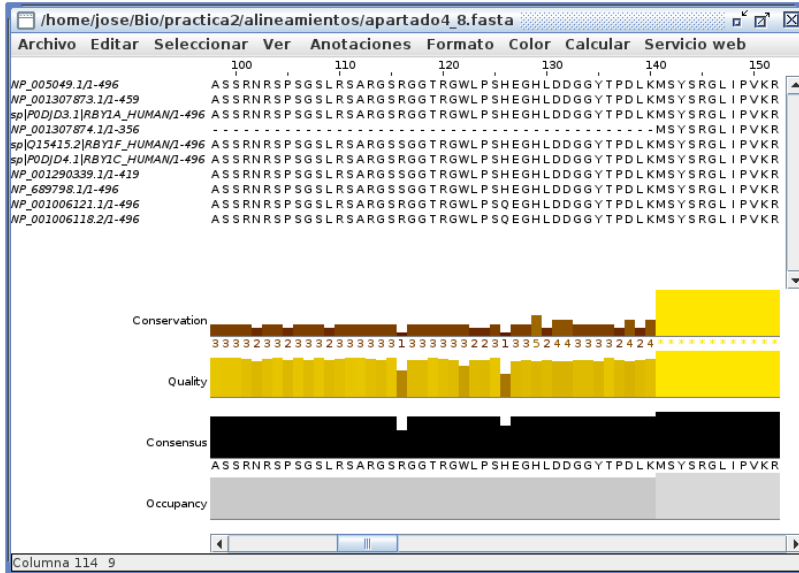


Figura 5: Alineamiento de proteínas del cromosoma Y de primates.

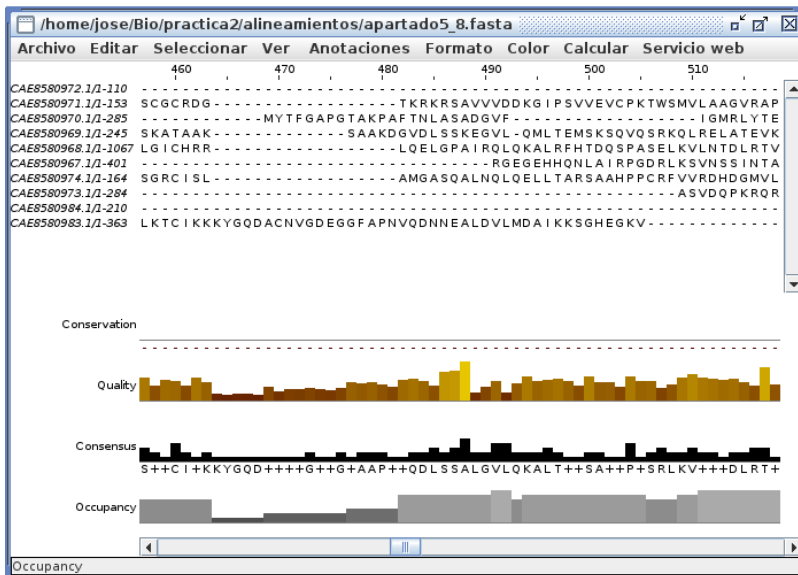


Figura 6: Alineamiento de proteínas añadidas en lo que va de año.

## 4. Mejoras de eficiencia.

La mejor forma de aumentar la eficiencia es añadiendo thread. Ahora bien, por defecto el programa añade los máximos, subir el número mediante parámetros no supondría ningún cambio. Así que tendremos que optar reducir el número de iteraciones de arboles guía y de HMM. En los apartados anteriores, hemos incluido una tabla con el tiempo en función del número de iteraciones (Figuras 1 y 4) y se han comparado los resultados. Si se desea mirar más detalladamente las distintas distancias según las iteraciones, en los apéndices se encuentran todas las matrices.

## 5. Alineamientos de secuencias del artículo de Metzker M.L., Mindell D.P. y Liu X.M. [1]

Vamos a hacer uso del fichero fasta que contiene todas las secuencias mencionadas en el paper visto en la práctica anterior.

En principio no hubo ningún problema en alinear todas las secuencias. En menos de un minuto ya se obtuvieron los resultados. No obstante, es bastante complicado interpretar una matriz de 174x174. En líneas generales podemos afirmar que las secuencias guardan una gran distancia entre ellas. Desde Jalview se puede visualizar de manera más esclarecedora los alineamientos (Figura ).

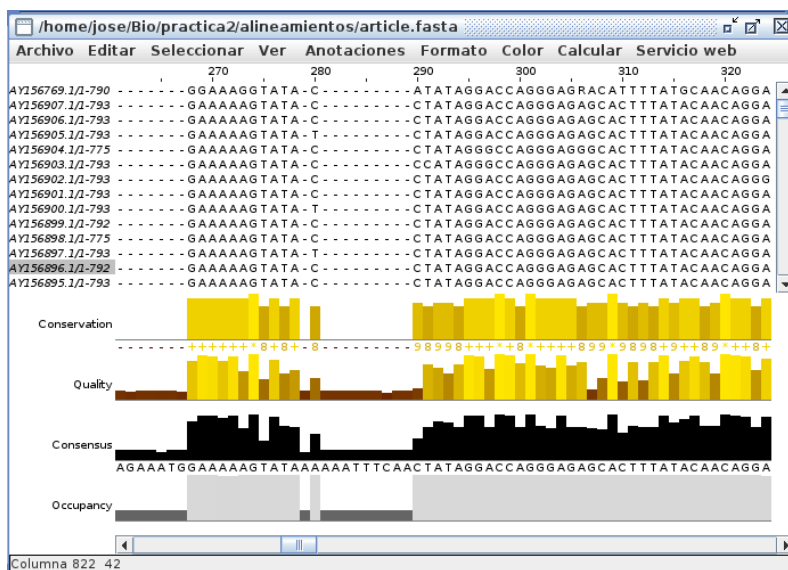


Figura 7: Alineamiento de secuencias del artículo.

## 6. Alineamientos de códigos de barras.

Para este ejercicio necesitaremos unos cuantos barcodes más de las mismas especies. Siguiendo el procedimiento detallado en la práctica anterior, entramos a BOLD y buscamos las especies que nos interesan. Preferiblemente buscaremos COI-5P:995. Como del águila solitaria había un solo código de barras y no era COI-5P:995, se ha decidido utilizar en su lugar barcodes de watusi (*Bos taurus*), del puma (*Puma concolor*) y del ganso común (*Anser anser*).

La matriz nos muestra una pequeña distancia entre barcodes de una misma especie. De hecho hay algunos ejemplos de la misma especie cuya distancia es 0. Quizás sea demasiado precipitado lanzar una hipótesis con la reducida cantidad de datos que tenemos, aparentemente se podría afirmar que con una distancia inferior a 0.01, el código pertenece a la misma especie. Por otro lado, el código de barras de los seres humanos se alinea muy bien con el de los primates. Seguramente eso se deba que pertenecemos a la misma tribu (*Hominini*). Antes de hacer los cálculos, esperaba una mayor distancia entre los gansos y el resto, ya que el resto de muestras son de mamíferos y el ganso es un ave. Sin embargo, la distancia no es tan notable. También esperaba mejores alineamientos entre el código del mono narigudo y el del resto de primates.

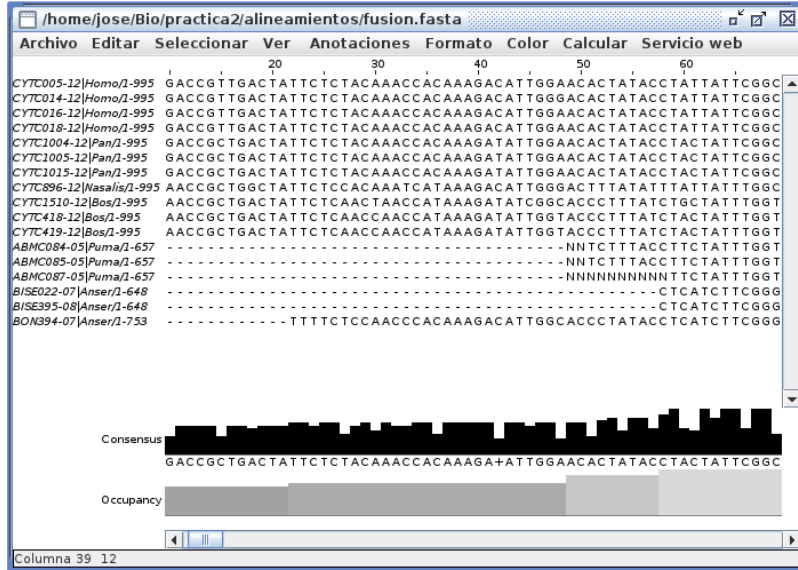


Figura 8: Alineamiento de secuencias de códigos de barras.

## A. Matrices de distancia de los alineamientos de secuencias de DNA.

$$\begin{pmatrix}
 NM_01001961,2 & 0,000000 & 0,777436 & 0,775385 & 0,777436 & 0,500000 & 0,634146 & 0,560976 & 0,791795 & 0,777436 & 0,777436 \\
 NM_01330132,3 & 0,777436 & 0,000000 & 0,793122 & 0,044944 & 0,500000 & 0,585366 & 0,634146 & 0,822826 & 0,066092 & 0,059538 \\
 NM_01367607,2 & 0,775385 & 0,793122 & 0,000000 & 0,782969 & 0,450000 & 0,560976 & 0,560976 & 0,794565 & 0,782759 & 0,784971 \\
 NM_01393564,1 & 0,777436 & 0,044944 & 0,782969 & 0,000000 & 0,500000 & 0,609756 & 0,634146 & 0,806032 & 0,002957 & 0,043761 \\
 pdb|7D58|T & 0,500000 & 0,500000 & 0,450000 & 0,500000 & 0,000000 & 0,600000 & 0,650000 & 0,500000 & 0,500000 & 0,500000 \\
 pdb|7D58|S & 0,634146 & 0,585366 & 0,560976 & 0,609756 & 0,600000 & 0,000000 & 0,658537 & 0,682927 & 0,609756 & 0,609756 \\
 pdb|7D58|R & 0,560976 & 0,634146 & 0,560976 & 0,634146 & 0,650000 & 0,658537 & 0,000000 & 0,634146 & 0,634146 & 0,634146 \\
 NR_171783,1 & 0,791795 & 0,822826 & 0,794565 & 0,806032 & 0,500000 & 0,682927 & 0,634146 & 0,000000 & 0,820115 & 0,818497 \\
 NM_01330136,3 & 0,777436 & 0,066092 & 0,782759 & 0,002957 & 0,500000 & 0,609756 & 0,634146 & 0,820115 & 0,000000 & 0,063006 \\
 NM_01393566,1 & 0,777436 & 0,059538 & 0,784971 & 0,043761 & 0,500000 & 0,609756 & 0,634146 & 0,818497 & 0,063006 & 0,000000
 \end{pmatrix}$$

Figura 9: Matriz de distancia de secuencias de DNA humano con cero iteraciones.



$$\begin{pmatrix} NM_01001961,2 & 0,000000 & 0,953846 & 0,567180 & 0,953846 & 1,000000 & 1,000000 & 1,000000 & 0,713846 & 0,953846 & 0,953846 \\ NM_01330132,3 & 0,953846 & 0,000000 & 0,599678 & 0,036073 & 1,000000 & 0,682927 & 1,000000 & 0,945652 & 0,052874 & 0,050867 \\ NM_01367607,2 & 0,567180 & 0,599678 & 0,000000 & 0,601419 & 0,450000 & 0,414634 & 0,512195 & 0,605435 & 0,608046 & 0,596532 \\ NM_01393564,1 & 0,953846 & 0,036073 & 0,601419 & 0,000000 & 1,000000 & 1,000000 & 1,000000 & 0,940863 & 0,031342 & 0,040804 \\ pdb|7D58|T & 1,000000 & 1,000000 & 0,450000 & 1,000000 & 0,000000 & 1,000000 & 1,000000 & 0,800000 & 1,000000 & 1,000000 \\ pdb|7D58|S & 1,000000 & 0,682927 & 0,414634 & 1,000000 & 1,000000 & 0,000000 & 1,000000 & 1,000000 & 1,000000 & 1,000000 \\ pdb|7D58|R & 1,000000 & 1,000000 & 0,512195 & 1,000000 & 1,000000 & 1,000000 & 0,000000 & 1,000000 & 1,000000 & 1,000000 \\ NR_171783,1 & 0,713846 & 0,945652 & 0,605435 & 0,940863 & 0,800000 & 1,000000 & 1,000000 & 0,000000 & 0,942529 & 0,942197 \\ NM_01330136,3 & 0,953846 & 0,052874 & 0,608046 & 0,031342 & 1,000000 & 1,000000 & 1,000000 & 0,942529 & 0,000000 & 0,053757 \\ NM_01393566,1 & 0,953846 & 0,050867 & 0,596532 & 0,040804 & 1,000000 & 1,000000 & 1,000000 & 0,942197 & 0,053757 & 0,000000 \end{pmatrix}$$

Figura 10: Matriz de distancia de secuencias de DNA humano con dos iteraciones.

$$\begin{pmatrix} NM_01001961,2 & 0,000000 & 0,958974 & 0,606154 & 0,958974 & 1,000000 & 1,000000 & 1,000000 & 0,711795 & 0,958974 & 0,958974 \\ NM_01330132,3 & 0,958974 & 0,000000 & 0,607738 & 0,037256 & 1,000000 & 0,682927 & 1,000000 & 0,944022 & 0,059195 & 0,048555 \\ NM_01367607,2 & 0,606154 & 0,607738 & 0,000000 & 0,607333 & 0,450000 & 0,439024 & 0,536585 & 0,617391 & 0,613218 & 0,606936 \\ NM_01393564,1 & 0,958974 & 0,037256 & 0,607333 & 0,000000 & 1,000000 & 1,000000 & 1,000000 & 0,939089 & 0,036073 & 0,044352 \\ pdb|7D58|T & 1,000000 & 1,000000 & 0,450000 & 1,000000 & 0,000000 & 1,000000 & 1,000000 & 0,800000 & 1,000000 & 1,000000 \\ pdb|7D58|S & 1,000000 & 0,682927 & 0,439024 & 1,000000 & 1,000000 & 0,000000 & 1,000000 & 1,000000 & 1,000000 & 1,000000 \\ pdb|7D58|R & 1,000000 & 1,000000 & 0,536585 & 1,000000 & 1,000000 & 1,000000 & 0,000000 & 1,000000 & 1,000000 & 1,000000 \\ NR_171783,1 & 0,711795 & 0,944022 & 0,617391 & 0,939089 & 0,800000 & 1,000000 & 1,000000 & 0,000000 & 0,940805 & 0,940462 \\ NM_01330136,3 & 0,958974 & 0,059195 & 0,613218 & 0,036073 & 1,000000 & 1,000000 & 1,000000 & 0,940805 & 0,000000 & 0,056069 \\ NM_01393566,1 & 0,958974 & 0,048555 & 0,606936 & 0,044352 & 1,000000 & 1,000000 & 1,000000 & 0,940462 & 0,056069 & 0,000000 \end{pmatrix}$$

Figura 11: Matriz de distancia de secuencias de DNA humano con cuatro iteraciones.

$$\begin{pmatrix} NM_01001961,2 & 0,000000 & 0,954872 & 0,637949 & 0,954872 & 1,000000 & 1,000000 & 1,000000 & 0,757949 & 0,954872 & 0,954872 \\ NM_01330132,3 & 0,954872 & 0,000000 & 0,611499 & 0,031934 & 1,000000 & 1,000000 & 1,000000 & 0,945652 & 0,055747 & 0,051445 \\ NM_01367607,2 & 0,637949 & 0,611499 & 0,000000 & 0,610881 & 0,300000 & 0,487805 & 0,512195 & 0,617391 & 0,618966 & 0,612139 \\ NM_01393564,1 & 0,954872 & 0,031934 & 0,610881 & 0,000000 & 1,000000 & 1,000000 & 1,000000 & 0,940863 & 0,032525 & 0,038439 \\ pdb|7D58|T & 1,000000 & 1,000000 & 0,300000 & 1,000000 & 0,000000 & 1,000000 & 1,000000 & 0,700000 & 1,000000 & 1,000000 \\ pdb|7D58|S & 1,000000 & 1,000000 & 0,487805 & 1,000000 & 1,000000 & 0,000000 & 1,000000 & 1,000000 & 1,000000 & 1,000000 \\ pdb|7D58|R & 1,000000 & 1,000000 & 0,512195 & 1,000000 & 1,000000 & 1,000000 & 0,000000 & 1,000000 & 1,000000 & 1,000000 \\ NR_171783,1 & 0,757949 & 0,945652 & 0,617391 & 0,940863 & 0,700000 & 1,000000 & 1,000000 & 0,000000 & 0,942529 & 0,942197 \\ NM_01330136,3 & 0,954872 & 0,055747 & 0,618966 & 0,032525 & 1,000000 & 1,000000 & 1,000000 & 0,942529 & 0,000000 & 0,054913 \\ NM_01393566,1 & 0,954872 & 0,051445 & 0,612139 & 0,038439 & 1,000000 & 1,000000 & 1,000000 & 0,942197 & 0,054913 & 0,000000 \end{pmatrix}$$

Figura 12: Matriz de distancia de secuencias de DNA humano con ocho iteraciones.

$$\begin{pmatrix} G65269,1 & 0,000000 & 0,750000 & 0,756983 & 0,808824 & 0,762004 & 0,762646 & 0,783613 & 0,797674 & 0,781553 & 0,786765 \\ G65268,1 & 0,750000 & 0,000000 & 0,782123 & 0,742718 & 0,754854 & 0,750973 & 0,759709 & 0,762136 & 0,771845 & 0,752451 \\ G65267,1 & 0,756983 & 0,782123 & 0,000000 & 0,745810 & 0,798883 & 0,719844 & 0,770950 & 0,756983 & 0,784916 & 0,770950 \\ G65266,1 & 0,808824 & 0,742718 & 0,745810 & 0,000000 & 0,750000 & 0,700389 & 0,773109 & 0,790698 & 0,754854 & 0,757353 \\ G65265,1 & 0,762004 & 0,754854 & 0,798883 & 0,750000 & 0,000000 & 0,747082 & 0,760504 & 0,772093 & 0,786408 & 0,767157 \\ G65264,1 & 0,762646 & 0,750973 & 0,719844 & 0,700389 & 0,747082 & 0,000000 & 0,731518 & 0,762646 & 0,774319 & 0,696498 \\ G65263,1 & 0,783613 & 0,759709 & 0,770950 & 0,773109 & 0,760504 & 0,731518 & 0,000000 & 0,755814 & 0,655340 & 0,779412 \\ G65262,1 & 0,797674 & 0,762136 & 0,756983 & 0,790698 & 0,772093 & 0,762646 & 0,755814 & 0,000000 & 0,754854 & 0,764706 \\ G65261,1 & 0,781553 & 0,771845 & 0,784916 & 0,754854 & 0,786408 & 0,774319 & 0,655340 & 0,754854 & 0,000000 & 0,794118 \\ G65260,1 & 0,786765 & 0,752451 & 0,770950 & 0,757353 & 0,767157 & 0,696498 & 0,779412 & 0,764706 & 0,794118 & 0,000000 \end{pmatrix}$$

Figura 13: Matriz de distancia de secuencias de DNA humano del cromosoma X con cero iteraciones.

$$\begin{pmatrix} G65269,1 & 0,000000 & 0,672330 & 0,712291 & 0,680672 & 0,686848 & 0,677043 & 0,707983 & 0,676744 & 0,720874 & 0,637255 \\ G65268,1 & 0,672330 & 0,000000 & 0,726257 & 0,623786 & 0,665049 & 0,700389 & 0,720874 & 0,745146 & 0,754854 & 0,666667 \\ G65267,1 & 0,712291 & 0,726257 & 0,000000 & 0,653631 & 0,670391 & 0,700389 & 0,720670 & 0,743017 & 0,715084 & 0,620112 \\ G65266,1 & 0,680672 & 0,623786 & 0,653631 & 0,000000 & 0,613445 & 0,571984 & 0,693277 & 0,665116 & 0,723301 & 0,573529 \\ G65265,1 & 0,686848 & 0,665049 & 0,670391 & 0,613445 & 0,000000 & 0,649805 & 0,701681 & 0,697674 & 0,703884 & 0,661765 \\ G65264,1 & 0,677043 & 0,700389 & 0,700389 & 0,571984 & 0,649805 & 0,000000 & 0,715953 & 0,642023 & 0,708171 & 0,540856 \\ G65263,1 & 0,707983 & 0,720874 & 0,720670 & 0,693277 & 0,701681 & 0,715953 & 0,000000 & 0,693023 & 0,514563 & 0,691176 \\ G65262,1 & 0,676744 & 0,745146 & 0,743017 & 0,665116 & 0,697674 & 0,642023 & 0,693023 & 0,000000 & 0,689320 & 0,659314 \\ G65261,1 & 0,720874 & 0,754854 & 0,715084 & 0,723301 & 0,703884 & 0,708171 & 0,514563 & 0,689320 & 0,000000 & 0,723039 \\ G65260,1 & 0,637255 & 0,666667 & 0,620112 & 0,573529 & 0,661765 & 0,540856 & 0,691176 & 0,659314 & 0,723039 & 0,000000 \end{pmatrix}$$

Figura 14: Matriz de distancia de secuencias de DNA humano del cromosoma X con dos iteraciones.

$$\begin{pmatrix} G65269,1 & 0,000000 & 0,679612 & 0,729050 & 0,695378 & 0,691023 & 0,704280 & 0,693277 & 0,679070 & 0,711165 & 0,659314 \\ G65268,1 & 0,679612 & 0,000000 & 0,715084 & 0,645631 & 0,665049 & 0,758755 & 0,740291 & 0,740291 & 0,752427 & 0,700980 \\ G65267,1 & 0,729050 & 0,715084 & 0,000000 & 0,659218 & 0,664804 & 0,696498 & 0,712291 & 0,745810 & 0,703911 & 0,606145 \\ G65266,1 & 0,695378 & 0,645631 & 0,659218 & 0,000000 & 0,613445 & 0,568093 & 0,703781 & 0,686046 & 0,708738 & 0,593137 \\ G65265,1 & 0,691023 & 0,665049 & 0,664804 & 0,613445 & 0,000000 & 0,657588 & 0,712185 & 0,704651 & 0,706311 & 0,664216 \\ G65264,1 & 0,704280 & 0,758755 & 0,696498 & 0,568093 & 0,657588 & 0,000000 & 0,743191 & 0,657588 & 0,747082 & 0,544747 \\ G65263,1 & 0,693277 & 0,740291 & 0,712291 & 0,703781 & 0,712185 & 0,743191 & 0,000000 & 0,688372 & 0,516990 & 0,720588 \\ G65262,1 & 0,679070 & 0,740291 & 0,745810 & 0,686046 & 0,704651 & 0,657588 & 0,688372 & 0,000000 & 0,689320 & 0,664216 \\ G65261,1 & 0,711165 & 0,752427 & 0,703911 & 0,708738 & 0,706311 & 0,747082 & 0,516990 & 0,689320 & 0,000000 & 0,759804 \\ G65260,1 & 0,659314 & 0,700980 & 0,606145 & 0,593137 & 0,664216 & 0,544747 & 0,720588 & 0,664216 & 0,759804 & 0,000000 \end{pmatrix}$$

Figura 15: Matriz de distancia de secuencias de DNA humano del cromosoma X con cuatro iteraciones.

$$\begin{pmatrix} G65269,1 & 0,000000 & 0,677184 & 0,726257 & 0,684874 & 0,699374 & 0,680934 & 0,678571 & 0,706977 & 0,701456 & 0,661765 \\ G65268,1 & 0,677184 & 0,000000 & 0,717877 & 0,662621 & 0,650485 & 0,750973 & 0,733010 & 0,745146 & 0,742718 & 0,723039 \\ G65267,1 & 0,726257 & 0,717877 & 0,000000 & 0,662011 & 0,681564 & 0,692607 & 0,715084 & 0,754190 & 0,701117 & 0,628492 \\ G65266,1 & 0,684874 & 0,662621 & 0,662011 & 0,000000 & 0,605042 & 0,591440 & 0,705882 & 0,679070 & 0,703884 & 0,595588 \\ G65265,1 & 0,699374 & 0,650485 & 0,681564 & 0,605042 & 0,000000 & 0,645914 & 0,712185 & 0,702326 & 0,696602 & 0,683824 \\ G65264,1 & 0,680934 & 0,750973 & 0,692607 & 0,591440 & 0,645914 & 0,000000 & 0,754864 & 0,669261 & 0,719844 & 0,540856 \\ G65263,1 & 0,678571 & 0,733010 & 0,715084 & 0,705882 & 0,712185 & 0,754864 & 0,000000 & 0,709302 & 0,524272 & 0,730392 \\ G65262,1 & 0,706977 & 0,745146 & 0,754190 & 0,679070 & 0,702326 & 0,669261 & 0,709302 & 0,000000 & 0,699029 & 0,678922 \\ G65261,1 & 0,701456 & 0,742718 & 0,701117 & 0,703884 & 0,696602 & 0,719844 & 0,524272 & 0,699029 & 0,000000 & 0,750000 \\ G65260,1 & 0,661765 & 0,723039 & 0,628492 & 0,595588 & 0,683824 & 0,540856 & 0,730392 & 0,678922 & 0,750000 & 0,000000 \end{pmatrix}$$

Figura 16: Matriz de distancia de secuencias de DNA humano del cromosoma X con ocho iteraciones.

$$\begin{pmatrix} NR_171708,1 & 0,000000 & 0,797581 & 0,780071 & 0,791005 & 0,810376 & 0,804330 & 0,772050 & 0,790995 & 0,806559 & 0,802640 \\ NM_01131951,2 & 0,797581 & 0,000000 & 0,784214 & 0,789683 & 0,806493 & 0,810991 & 0,767027 & 0,788670 & 0,802037 & 0,786799 \\ NM_01393493,1 & 0,780071 & 0,784214 & 0,000000 & 0,177249 & 0,795468 & 0,791007 & 0,003106 & 0,791863 & 0,801887 & 0,172277 \\ NM_01393498,1 & 0,791005 & 0,789683 & 0,177249 & 0,000000 & 0,815476 & 0,794338 & 0,172619 & 0,779762 & 0,796958 & 0,003307 \\ NM_013371,5 & 0,810376 & 0,806493 & 0,795468 & 0,815476 & 0,000000 & 0,786012 & 0,798758 & 0,792487 & 0,793083 & 0,809241 \\ NM_01393359,2 & 0,804330 & 0,810991 & 0,791007 & 0,794338 & 0,786012 & 0,000000 & 0,795171 & 0,793505 & 0,795171 & 0,793505 \\ NM_01393496,1 & 0,772050 & 0,767027 & 0,003106 & 0,172619 & 0,798758 & 0,795171 & 0,000000 & 0,766460 & 0,798758 & 0,177558 \\ NM_01393492,1 & 0,790995 & 0,788670 & 0,791863 & 0,779762 & 0,792487 & 0,793505 & 0,766460 & 0,000000 & 0,807651 & 0,768977 \\ NM_01393499,1 & 0,806559 & 0,802037 & 0,801887 & 0,796958 & 0,793083 & 0,795171 & 0,798758 & 0,807651 & 0,000000 & 0,802640 \\ NM_01393497,1 & 0,802640 & 0,786799 & 0,172277 & 0,003307 & 0,809241 & 0,793505 & 0,177558 & 0,768977 & 0,802640 & 0,000000 \end{pmatrix}$$

Figura 17: Matriz de distancia de secuencias de DNA de mamíferos con cero iteraciones.

$$\begin{pmatrix} NR_171708,1 & 0,000000 & 0,863781 & 0,731132 & 0,734788 & 0,838998 & 0,858451 & 0,734783 & 0,649250 & 0,668149 & 0,734653 \\ NM_01131951,2 & 0,863781 & 0,000000 & 0,824316 & 0,842593 & 0,841502 & 0,830142 & 0,841502 & 0,623806 & 0,670910 & 0,842904 \\ NM_01393493,1 & 0,731132 & 0,824316 & 0,000000 & 0,136905 & 0,797853 & 0,813489 & 0,001242 & 0,591392 & 0,658019 & 0,136634 \\ NM_01393498,1 & 0,734788 & 0,842593 & 0,136905 & 0,000000 & 0,776455 & 0,813489 & 0,138889 & 0,607143 & 0,667989 & 0,000000 \\ NM_013371,5 & 0,838998 & 0,841502 & 0,797853 & 0,776455 & 0,000000 & 0,596170 & 0,790062 & 0,737030 & 0,613596 & 0,776238 \\ NM_01393359,2 & 0,858451 & 0,830142 & 0,813489 & 0,813489 & 0,596170 & 0,000000 & 0,813489 & 0,762698 & 0,646128 & 0,813489 \\ NM_01393496,1 & 0,734783 & 0,841502 & 0,001242 & 0,138889 & 0,790062 & 0,813489 & 0,000000 & 0,590683 & 0,661491 & 0,139934 \\ NM_01393492,1 & 0,649250 & 0,623806 & 0,591392 & 0,607143 & 0,737030 & 0,762698 & 0,590683 & 0,000000 & 0,622519 & 0,606601 \\ NM_01393499,1 & 0,668149 & 0,670910 & 0,658019 & 0,667989 & 0,613596 & 0,646128 & 0,661491 & 0,622519 & 0,000000 & 0,666667 \\ NM_01393497,1 & 0,734653 & 0,842904 & 0,136634 & 0,000000 & 0,776238 & 0,813489 & 0,139934 & 0,606601 & 0,666667 & 0,000000 \end{pmatrix}$$

Figura 18: Matriz de distancia de secuencias de DNA de mamíferos con dos iteraciones.

$$\begin{pmatrix} NR_171708,1 & 0,000000 & 0,786124 & 0,749410 & 0,757936 & 1,000000 & 0,915071 & 0,754037 & 0,663702 & 0,690939 & 0,758416 \\ NM_01131951,2 & 0,786124 & 0,000000 & 0,716741 & 0,718915 & 1,000000 & 0,856786 & 0,718014 & 0,649268 & 0,684914 & 0,718152 \\ NM_01393493,1 & 0,749410 & 0,716741 & 0,000000 & 0,139550 & 1,000000 & 0,876769 & 0,001242 & 0,603774 & 0,656250 & 0,137954 \\ NM_01393498,1 & 0,757936 & 0,718915 & 0,139550 & 0,000000 & 1,000000 & 0,875937 & 0,138228 & 0,615079 & 0,673280 & 0,001323 \\ NM_013371,5 & 1,000000 & 1,000000 & 1,000000 & 1,000000 & 0,000000 & 0,958368 & 1,000000 & 1,000000 & 0,618366 & 1,000000 \\ NM_01393359,2 & 0,915071 & 0,856786 & 0,876769 & 0,875937 & 0,958368 & 0,000000 & 0,875937 & 0,835970 & 0,590341 & 0,876769 \\ NM_01393496,1 & 0,754037 & 0,718014 & 0,001242 & 0,138228 & 1,000000 & 0,875937 & 0,000000 & 0,602484 & 0,659006 & 0,141254 \\ NM_01393492,1 & 0,663702 & 0,649268 & 0,603774 & 0,615079 & 1,000000 & 0,835970 & 0,602484 & 0,000000 & 0,644533 & 0,615182 \\ NM_01393499,1 & 0,690939 & 0,684914 & 0,656250 & 0,673280 & 0,618366 & 0,590341 & 0,659006 & 0,644533 & 0,000000 & 0,672607 \\ NM_01393497,1 & 0,758416 & 0,718152 & 0,137954 & 0,001323 & 1,000000 & 0,876769 & 0,141254 & 0,615182 & 0,672607 & 0,000000 \end{pmatrix}$$

Figura 19: Matriz de distancia de secuencias de DNA de mamíferos con cuatro iteraciones.

$$\begin{pmatrix} NR_171708,1 & 0,000000 & 0,800127 & 0,764741 & 0,789021 & 1,000000 & 1,000000 & 0,760870 & 0,682046 & 0,670928 & 0,788779 \\ NM_01131951,2 & 0,800127 & 0,000000 & 0,739656 & 0,733466 & 1,000000 & 1,000000 & 0,740929 & 0,714195 & 0,666454 & 0,732673 \\ NM_01393493,1 & 0,764741 & 0,739656 & 0,000000 & 0,139550 & 1,000000 & 1,000000 & 0,000000 & 0,613797 & 0,660377 & 0,139274 \\ NM_01393498,1 & 0,789021 & 0,733466 & 0,139550 & 0,000000 & 1,000000 & 1,000000 & 0,139550 & 0,624339 & 0,670635 & 0,000000 \\ NM_013371,5 & 1,000000 & 1,000000 & 1,000000 & 1,000000 & 0,000000 & 0,836803 & 1,000000 & 1,000000 & 0,635659 & 1,000000 \\ NM_01393359,2 & 1,000000 & 1,000000 & 1,000000 & 1,000000 & 0,836803 & 0,000000 & 1,000000 & 1,000000 & 0,649459 & 1,000000 \\ NM_01393496,1 & 0,760870 & 0,740929 & 0,000000 & 0,139550 & 1,000000 & 1,000000 & 0,000000 & 0,613665 & 0,663354 & 0,141254 \\ NM_01393492,1 & 0,682046 & 0,714195 & 0,613797 & 0,624339 & 1,000000 & 1,000000 & 0,613665 & 0,000000 & 0,654998 & 0,624422 \\ NM_01393499,1 & 0,670928 & 0,666454 & 0,660377 & 0,670635 & 0,635659 & 0,649459 & 0,663354 & 0,654998 & 0,000000 & 0,669967 \\ NM_01393497,1 & 0,788779 & 0,732673 & 0,139274 & 0,000000 & 1,000000 & 1,000000 & 0,141254 & 0,624422 & 0,669967 & 0,000000 \end{pmatrix}$$

Figura 20: Matriz de distancia de secuencias de DNA de mamíferos con ocho iteraciones.

$$\begin{pmatrix} MW498497 & 0,000000 & 0,002716 & 0,003139 & 0,003380 & 0,002535 & 0,002052 & 0,002354 & 0,002173 & 0,002293 & 0,002173 \\ MW389273 & 0,002716 & 0,000000 & 0,001690 & 0,002595 & 0,001630 & 0,001267 & 0,000483 & 0,001267 & 0,001388 & 0,001267 \\ MW389272 & 0,003139 & 0,001690 & 0,000000 & 0,003199 & 0,001871 & 0,001992 & 0,001690 & 0,002113 & 0,002233 & 0,002113 \\ MW389271 & 0,003380 & 0,002595 & 0,003199 & 0,000000 & 0,002294 & 0,002052 & 0,002354 & 0,002052 & 0,002173 & 0,002052 \\ MW389270 & 0,002535 & 0,001630 & 0,001871 & 0,002294 & 0,000000 & 0,000724 & 0,001388 & 0,000604 & 0,000724 & 0,000604 \\ MW389269 & 0,002052 & 0,001267 & 0,001992 & 0,002052 & 0,000724 & 0,000000 & 0,000905 & 0,000362 & 0,000362 & 0,000362 \\ MW389268 & 0,002354 & 0,000483 & 0,001690 & 0,002354 & 0,001388 & 0,000905 & 0,000000 & 0,001026 & 0,001147 & 0,001026 \\ MW389266 & 0,002173 & 0,001267 & 0,002113 & 0,002052 & 0,000604 & 0,000362 & 0,001026 & 0,000000 & 0,000362 & 0,000241 \\ MW389265 & 0,002293 & 0,001388 & 0,002233 & 0,002173 & 0,000724 & 0,000362 & 0,001147 & 0,000362 & 0,000000 & 0,000362 \\ MW389262 & 0,002173 & 0,001267 & 0,002113 & 0,002052 & 0,000604 & 0,000362 & 0,001026 & 0,000241 & 0,000362 & 0,000000 \end{pmatrix}$$

Figura 21: Matriz de distancia de secuencias de DNA humano mitocondrial.

## B. Matrices de distancia de los alineamientos de secuencias de proteínas.

$NP_005049,1$	0,000000	0,074074	0,114919	0,000000	0,139113	0,114919	0,062053	0,139113	0,137097	0,137097
$NP_001307873,1$	0,074074	0,000000	0,054466	0,112360	0,061002	0,056645	0,157518	0,061002	0,058824	0,056645
$sp P0DJD3,1 RBY1A_HUMAN$	0,114919	0,054466	0,000000	0,000000	0,139113	0,114919	0,062053	0,139113	0,137097	0,137097
$NP_001307874,1$	0,000000	0,112360	0,000000	0,000000	0,005618	0,000000	0,283708	0,005618	0,002809	0,002809
$sp Q15415,2 RBY1F_HUMAN$	0,139113	0,061002	0,139113	0,005618	0,000000	0,110887	0,054893	0,127016	0,112903	0,110887
$sp P0DJD4,1 RBY1C_HUMAN$	0,114919	0,056645	0,114919	0,000000	0,110887	0,000000	0,064439	0,139113	0,139113	0,139113
$NP_001290339,1$	0,062053	0,157518	0,062053	0,283708	0,054893	0,064439	0,000000	0,054893	0,064439	0,064439
$NP_0089798,1$	0,139113	0,061002	0,139113	0,005618	0,127016	0,139113	0,054893	0,000000	0,112903	0,110887
$NP_001006121,1$	0,137097	0,058824	0,137097	0,002809	0,112903	0,139113	0,064439	0,112903	0,000000	0,135081
$NP_001006118,2$	0,137097	0,056645	0,137097	0,002809	0,110887	0,139113	0,064439	0,110887	0,135081	0,000000

Figura 22: Matriz de distancia de secuencias de proteínas pertenecientes al cromosoma Y de los primates con 0 iteraciones.

$$\left( \begin{array}{l} NP_005049,1 \\ NP_01307873,1 \\ sp|P0DJ3,1|RB Y1A_HUMAN \\ NP_01307874,1 \\ sp|Q15415,2|RB Y1F_HUMAN \\ sp|P0DJ4,1|RB Y1C_HUMAN \\ NP_01290339,1 \\ NP_689798,1 \\ NP_01006121,1 \\ NP_01006118,2 \end{array} \begin{array}{l} 0,000000 \ 0,000000 \ 0,000000 \ 0,000000 \ 0,008065 \ 0,002016 \ 0,052506 \ 0,008065 \ 0,004032 \ 0,004032 \\ 0,000000 \ 0,000000 \ 0,000000 \ 0,103933 \ 0,008715 \ 0,002179 \ 0,140811 \ 0,008715 \ 0,004357 \ 0,004357 \\ 0,000000 \ 0,000000 \ 0,000000 \ 0,000000 \ 0,008065 \ 0,002016 \ 0,052506 \ 0,008065 \ 0,004032 \ 0,004032 \\ 0,000000 \ 0,103933 \ 0,000000 \ 0,000000 \ 0,005618 \ 0,000000 \ 0,272472 \ 0,005618 \ 0,002809 \ 0,002809 \\ 0,008065 \ 0,008715 \ 0,008065 \ 0,005618 \ 0,000000 \ 0,010081 \ 0,045346 \ 0,000000 \ 0,008065 \ 0,012097 \\ 0,002016 \ 0,002179 \ 0,002016 \ 0,000000 \ 0,010081 \ 0,000000 \ 0,054893 \ 0,010081 \ 0,006048 \ 0,006048 \\ 0,052506 \ 0,140811 \ 0,052506 \ 0,272472 \ 0,045346 \ 0,054893 \ 0,000000 \ 0,045346 \ 0,054893 \ 0,054893 \\ 0,008065 \ 0,008715 \ 0,008065 \ 0,005618 \ 0,000000 \ 0,010081 \ 0,045346 \ 0,000000 \ 0,008065 \ 0,012097 \\ 0,004032 \ 0,004357 \ 0,004032 \ 0,002809 \ 0,008065 \ 0,006048 \ 0,054893 \ 0,008065 \ 0,000000 \ 0,004032 \\ 0,004032 \ 0,004357 \ 0,004032 \ 0,002809 \ 0,012097 \ 0,006048 \ 0,054893 \ 0,012097 \ 0,004032 \ 0,000000 \end{array} \right)$$

Figura 23: Matriz de distancia de secuencias de proteínas pertenecientes al cromosoma Y de los primates con 2 iteraciones.

$$\left( \begin{array}{l} NP_005049,1 \\ NP_01307873,1 \\ sp|P0DJ3,1|RB Y1A_HUMAN \\ NP_01307874,1 \\ sp|Q15415,2|RB Y1F_HUMAN \\ sp|P0DJ4,1|RB Y1C_HUMAN \\ NP_01290339,1 \\ NP_689798,1 \\ NP_01006121,1 \\ NP_01006118,2 \end{array} \begin{array}{l} 0,000000 \ 0,000000 \ 0,000000 \ 0,000000 \ 0,008065 \ 0,002016 \ 0,052506 \ 0,008065 \ 0,004032 \ 0,004032 \\ 0,000000 \ 0,000000 \ 0,000000 \ 0,103933 \ 0,008715 \ 0,002179 \ 0,140811 \ 0,008715 \ 0,004357 \ 0,004357 \\ 0,000000 \ 0,000000 \ 0,000000 \ 0,000000 \ 0,008065 \ 0,002016 \ 0,052506 \ 0,008065 \ 0,004032 \ 0,004032 \\ 0,000000 \ 0,103933 \ 0,000000 \ 0,000000 \ 0,005618 \ 0,000000 \ 0,272472 \ 0,005618 \ 0,002809 \ 0,002809 \\ 0,008065 \ 0,008715 \ 0,008065 \ 0,005618 \ 0,000000 \ 0,010081 \ 0,045346 \ 0,000000 \ 0,008065 \ 0,012097 \\ 0,002016 \ 0,002179 \ 0,002016 \ 0,000000 \ 0,010081 \ 0,000000 \ 0,054893 \ 0,010081 \ 0,006048 \ 0,006048 \\ 0,052506 \ 0,140811 \ 0,052506 \ 0,272472 \ 0,045346 \ 0,054893 \ 0,000000 \ 0,045346 \ 0,054893 \ 0,054893 \\ 0,008065 \ 0,008715 \ 0,008065 \ 0,005618 \ 0,000000 \ 0,010081 \ 0,045346 \ 0,000000 \ 0,008065 \ 0,012097 \\ 0,004032 \ 0,004357 \ 0,004032 \ 0,002809 \ 0,008065 \ 0,006048 \ 0,054893 \ 0,008065 \ 0,000000 \ 0,004032 \\ 0,004032 \ 0,004357 \ 0,004032 \ 0,002809 \ 0,012097 \ 0,006048 \ 0,054893 \ 0,012097 \ 0,004032 \ 0,000000 \end{array} \right)$$

Figura 24: Matriz de distancia de secuencias de proteínas pertenecientes al cromosoma Y de los primates con 4 iteraciones.

$$\left( \begin{array}{l} NP_005049,1 \\ NP_01307873,1 \\ sp|P0DJ3,1|RB Y1A_HUMAN \\ NP_01307874,1 \\ sp|Q15415,2|RB Y1F_HUMAN \\ sp|P0DJ4,1|RB Y1C_HUMAN \\ NP_01290339,1 \\ NP_689798,1 \\ NP_01006121,1 \\ NP_01006118,2 \end{array} \begin{array}{l} 0,000000 \ 0,000000 \ 0,000000 \ 0,000000 \ 0,008065 \ 0,002016 \ 0,052506 \ 0,008065 \ 0,004032 \ 0,004032 \\ 0,000000 \ 0,000000 \ 0,000000 \ 0,103933 \ 0,008715 \ 0,002179 \ 0,140811 \ 0,008715 \ 0,004357 \ 0,004357 \\ 0,000000 \ 0,000000 \ 0,000000 \ 0,000000 \ 0,008065 \ 0,002016 \ 0,052506 \ 0,008065 \ 0,004032 \ 0,004032 \\ 0,000000 \ 0,103933 \ 0,000000 \ 0,000000 \ 0,005618 \ 0,000000 \ 0,272472 \ 0,005618 \ 0,002809 \ 0,002809 \\ 0,008065 \ 0,008715 \ 0,008065 \ 0,005618 \ 0,000000 \ 0,010081 \ 0,045346 \ 0,000000 \ 0,008065 \ 0,012097 \\ 0,002016 \ 0,002179 \ 0,002016 \ 0,000000 \ 0,010081 \ 0,000000 \ 0,054893 \ 0,010081 \ 0,006048 \ 0,006048 \\ 0,052506 \ 0,140811 \ 0,052506 \ 0,272472 \ 0,045346 \ 0,054893 \ 0,000000 \ 0,045346 \ 0,054893 \ 0,054893 \\ 0,008065 \ 0,008715 \ 0,008065 \ 0,005618 \ 0,000000 \ 0,010081 \ 0,045346 \ 0,000000 \ 0,008065 \ 0,012097 \\ 0,004032 \ 0,004357 \ 0,004032 \ 0,002809 \ 0,008065 \ 0,006048 \ 0,054893 \ 0,008065 \ 0,000000 \ 0,004032 \\ 0,004032 \ 0,004357 \ 0,004032 \ 0,002809 \ 0,012097 \ 0,006048 \ 0,054893 \ 0,012097 \ 0,004032 \ 0,000000 \end{array} \right)$$

Figura 25: Matriz de distancia de secuencias de proteínas pertenecientes al cromosoma Y de los primates con 8 iteraciones.

$$\left( \begin{array}{l} CAE8580972,1 \\ CAE8580971,1 \\ CAE8580970,1 \\ CAE8580969,1 \\ CAE8580968,1 \\ CAE8580967,1 \\ CAE8580974,1 \\ CAE8580973,1 \\ CAE8580984,1 \\ CAE8580983,1 \end{array} \begin{array}{l} 0,000000 \ 0,881818 \ 0,863636 \ 0,863636 \ 0,845455 \ 0,854545 \ 0,845455 \ 0,872727 \ 0,845455 \ 0,881818 \\ 0,881818 \ 0,000000 \ 0,888889 \ 0,856209 \ 0,843137 \ 0,869281 \ 0,888889 \ 0,836601 \ 0,862745 \ 0,856209 \\ 0,863636 \ 0,888889 \ 0,000000 \ 0,885714 \ 0,863158 \ 0,884211 \ 0,890244 \ 0,901408 \ 0,895238 \ 0,880702 \\ 0,863636 \ 0,856209 \ 0,885714 \ 0,000000 \ 0,869388 \ 0,873469 \ 0,847561 \ 0,869388 \ 0,880952 \ 0,885714 \\ 0,845455 \ 0,843137 \ 0,863158 \ 0,869388 \ 0,000000 \ 0,877805 \ 0,841463 \ 0,866197 \ 0,861905 \ 0,878788 \\ 0,854545 \ 0,869281 \ 0,884211 \ 0,873469 \ 0,877805 \ 0,000000 \ 0,859756 \ 0,873239 \ 0,885714 \ 0,895317 \\ 0,845455 \ 0,888889 \ 0,890244 \ 0,847561 \ 0,841463 \ 0,859756 \ 0,000000 \ 0,853659 \ 0,871951 \ 0,871951 \\ 0,872727 \ 0,836601 \ 0,901408 \ 0,869388 \ 0,866197 \ 0,873239 \ 0,853659 \ 0,000000 \ 0,885714 \ 0,883803 \\ 0,845455 \ 0,862745 \ 0,895238 \ 0,880952 \ 0,861905 \ 0,885714 \ 0,871951 \ 0,885714 \ 0,000000 \ 0,876190 \\ 0,881818 \ 0,856209 \ 0,880702 \ 0,885714 \ 0,878788 \ 0,895317 \ 0,871951 \ 0,883803 \ 0,876190 \ 0,000000 \end{array} \right)$$

Figura 26: Matriz de distancia de secuencias de proteínas que se han incluido en lo que llevamos de año con 0 iteraciones.

$$\begin{pmatrix} CAE8580972,1 & 0,000000 & 1,000000 & 0,972727 & 1,000000 & 0,800000 & 0,954545 & 1,000000 & 1,000000 & 0,945455 & 1,000000 \\ CAE8580971,1 & 1,000000 & 0,000000 & 0,980392 & 0,934641 & 0,862745 & 0,941176 & 0,954248 & 0,888889 & 1,000000 & 0,947712 \\ CAE8580970,1 & 0,972727 & 0,980392 & 0,000000 & 0,979592 & 0,898246 & 0,905263 & 0,987805 & 0,992958 & 1,000000 & 0,929825 \\ CAE8580969,1 & 1,000000 & 0,934641 & 0,979592 & 0,000000 & 0,804082 & 0,942857 & 0,945122 & 0,946939 & 1,000000 & 0,955102 \\ CAE8580968,1 & 0,800000 & 0,862745 & 0,898246 & 0,804082 & 0,000000 & 0,810474 & 0,835366 & 0,827465 & 0,871429 & 0,837466 \\ CAE8580967,1 & 0,954545 & 0,941176 & 0,905263 & 0,942857 & 0,810474 & 0,000000 & 0,951219 & 0,957747 & 1,000000 & 0,925620 \\ CAE8580974,1 & 1,000000 & 0,954248 & 0,987805 & 0,945122 & 0,835366 & 0,951219 & 0,000000 & 0,908537 & 1,000000 & 0,939024 \\ CAE8580973,1 & 1,000000 & 0,888889 & 0,992958 & 0,946939 & 0,827465 & 0,957747 & 0,908537 & 0,000000 & 1,000000 & 0,936620 \\ CAE8580984,1 & 0,945455 & 1,000000 & 1,000000 & 1,000000 & 0,871429 & 1,000000 & 1,000000 & 1,000000 & 0,000000 & 1,000000 \\ CAE8580983,1 & 1,000000 & 0,947712 & 0,929825 & 0,955102 & 0,837466 & 0,925620 & 0,939024 & 0,936620 & 1,000000 & 0,000000 \end{pmatrix}$$

Figura 27: Matriz de distancia de secuencias de proteínas que se han incluido en lo que llevamos de año con 2 iteraciones.

$$\begin{pmatrix} CAE8580972,1 & 0,000000 & 1,000000 & 0,972727 & 1,000000 & 0,827273 & 0,954545 & 1,000000 & 1,000000 & 0,945455 & 1,000000 \\ CAE8580971,1 & 1,000000 & 0,000000 & 0,973856 & 0,947712 & 0,869281 & 0,954248 & 0,954248 & 0,908497 & 1,000000 & 0,967320 \\ CAE8580970,1 & 0,972727 & 0,973856 & 0,000000 & 0,979592 & 0,901754 & 0,901754 & 0,975610 & 0,989437 & 1,000000 & 0,933333 \\ CAE8580969,1 & 1,000000 & 0,947712 & 0,979592 & 0,000000 & 0,853061 & 0,971429 & 0,963415 & 0,955102 & 1,000000 & 0,963265 \\ CAE8580968,1 & 0,827273 & 0,869281 & 0,901754 & 0,853061 & 0,000000 & 0,825436 & 0,835366 & 0,845070 & 0,876190 & 0,851240 \\ CAE8580967,1 & 0,954545 & 0,954248 & 0,901754 & 0,971429 & 0,825436 & 0,000000 & 0,945122 & 0,943662 & 1,000000 & 0,931129 \\ CAE8580974,1 & 1,000000 & 0,954248 & 0,975610 & 0,963415 & 0,835366 & 0,945122 & 0,000000 & 0,920732 & 1,000000 & 0,951219 \\ CAE8580973,1 & 1,000000 & 0,908497 & 0,989437 & 0,955102 & 0,845070 & 0,943662 & 0,920732 & 0,000000 & 1,000000 & 0,936620 \\ CAE8580984,1 & 0,945455 & 1,000000 & 1,000000 & 1,000000 & 0,876190 & 1,000000 & 1,000000 & 1,000000 & 0,000000 & 1,000000 \\ CAE8580983,1 & 1,000000 & 0,967320 & 0,933333 & 0,963265 & 0,851240 & 0,931129 & 0,951219 & 0,936620 & 1,000000 & 0,000000 \end{pmatrix}$$

Figura 28: Matriz de distancia de secuencias de proteínas que se han incluido en lo que llevamos de año con 4 iteraciones.

$$\begin{pmatrix} CAE8580972,1 & 0,000000 & 1,000000 & 0,972727 & 1,000000 & 0,827273 & 0,954545 & 1,000000 & 1,000000 & 0,945455 & 1,000000 \\ CAE8580971,1 & 1,000000 & 0,000000 & 0,960784 & 0,921569 & 0,843137 & 0,954248 & 0,947712 & 0,921569 & 1,000000 & 0,941176 \\ CAE8580970,1 & 0,972727 & 0,960784 & 0,000000 & 0,959184 & 0,898246 & 0,905263 & 0,981707 & 0,989437 & 1,000000 & 0,929825 \\ CAE8580969,1 & 1,000000 & 0,921569 & 0,959184 & 0,000000 & 0,857143 & 0,959184 & 0,932927 & 0,967347 & 1,000000 & 0,914286 \\ CAE8580968,1 & 0,827273 & 0,843137 & 0,898246 & 0,857143 & 0,000000 & 0,825436 & 0,847561 & 0,845070 & 0,876190 & 0,851240 \\ CAE8580967,1 & 0,954545 & 0,954248 & 0,905263 & 0,959184 & 0,825436 & 0,000000 & 0,945122 & 0,926056 & 1,000000 & 0,933884 \\ CAE8580974,1 & 1,000000 & 0,947712 & 0,981707 & 0,932927 & 0,847561 & 0,945122 & 0,000000 & 0,920732 & 1,000000 & 0,957317 \\ CAE8580973,1 & 1,000000 & 0,921569 & 0,989437 & 0,967347 & 0,845070 & 0,926056 & 0,920732 & 0,000000 & 1,000000 & 0,943662 \\ CAE8580984,1 & 0,945455 & 1,000000 & 1,000000 & 1,000000 & 0,876190 & 1,000000 & 1,000000 & 1,000000 & 0,000000 & 1,000000 \\ CAE8580983,1 & 1,000000 & 0,941176 & 0,929825 & 0,914286 & 0,851240 & 0,933884 & 0,957317 & 0,943662 & 1,000000 & 0,000000 \end{pmatrix}$$

Figura 29: Matriz de distancia de secuencias de proteínas que se han incluido en lo que llevamos de año con 8 iteraciones.

## Referencias

- [1] Liu X.M. Metzker M.L., Mindell D.P. Molecular evidence of hiv-1 transmission in a criminal case. *Natl Acad Sci USA.*, 2002.