

Práctica 1

Bioinformática

José Manuel Sánchez Aquilué 759267

20 de febrero de 2021

1. Introducción

En el siguiente documento se comentará el desarrollo de la primera práctica de bioinformática. Se detallará las consultas realizadas en cada punto, junto a los resultados obtenidos y, en caso de que el enunciado lo mande, el procedimiento de descarga de secuencias.

2. Bases de datos biológicas: GenBank

2.1. Buscar todas las secuencias almacenadas en GenBank procedentes de seres humanos.

Siguiendo las indicaciones de nuestro compañero, hacemos una consulta en genbank sobre el campo organismo. En esta ocasión, el organismo es una especie, los seres humanos (*Homo sapiens*). Así que lanzaremos la siguiente consulta:

```
1 Homo sapiens[Organism]
```

Este es el resultado (Figura 1). Como podemos observar, arroja una lista de 27823339 secuencias.

2.2. A partir de la búsqueda anterior, filtrar y quedarse solo con aquellas pertenecientes al cromosoma X.

Ahora sería cuestión de añadir la condición de pertenecer al cromosoma X. Las condiciones se unirían con el operador AND. Como no hay ninguna manera de filtrar por cromosomas, por genes sí pero por cromosomas no; añadimos la sentencia como texto libre.

```
1 (Homo sapiens[Organism]) AND (X chromosome)
```

Hemos pasado de tener millones de secuencias (Figura 1) a únicamente 2087 (Figura 2).

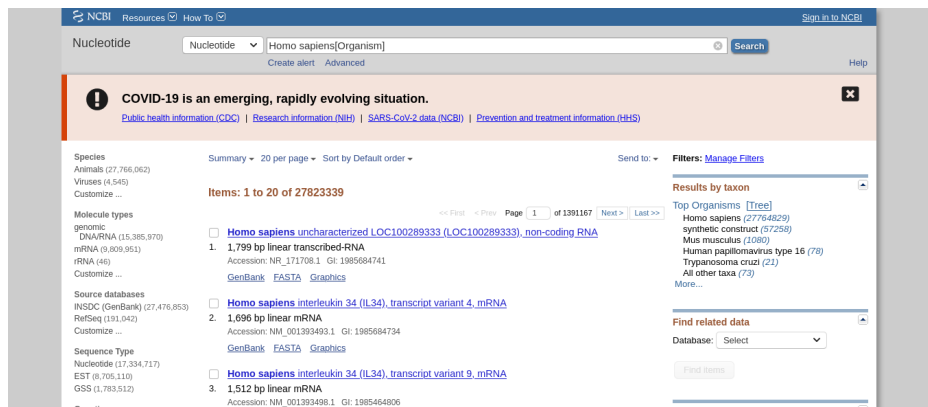


Figura 1: Resultado de la búsqueda de secuencias procedentes de seres humanos.

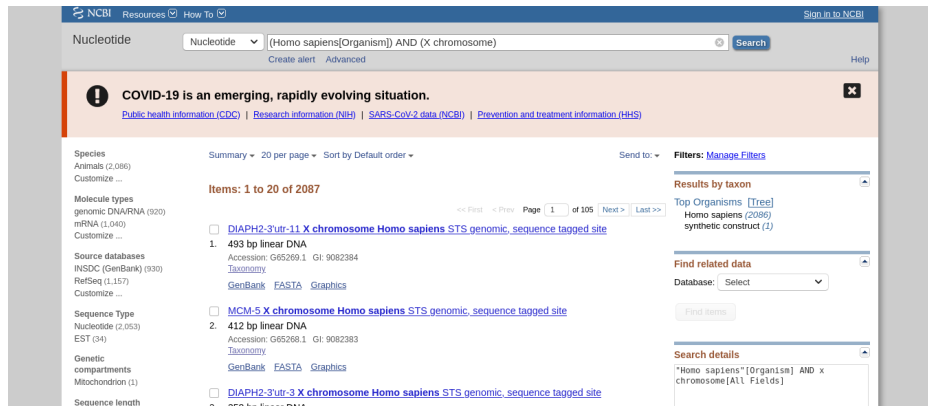


Figura 2: Resultado de la búsqueda de secuencias procedentes de seres humanos pertenecientes al cromosoma X.

2.3. Buscar todas las secuencias de DNA de referencia de mamíferos. Descargar las 10 primeras en formato FASTA.

Las búsquedas sobre clases taxonómicas, al igual que las de especie, se hacen sobre el campo de organismos. Cabe recordar que el nombre de la clase debe figurar en latín. En la barra de la izquierda, hay un filtro para el DNA referencial. En este punto, no nos piden únicamente realizar una consulta, además se pide descargar las 10 primeras secuencias. Para ello, en primer lugar, se seleccionan las secuencias. Bajo el aviso de la covid hay un botón de send to. Pulsamos ahí e indicamos la opción de fichero y el formato FASTA.

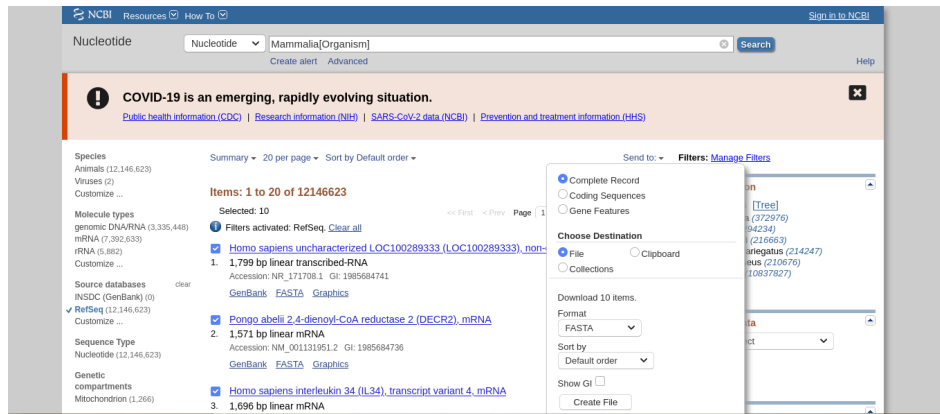


Figura 3: Búsqueda de secuencias de DNA referencial de mamíferos.

2.4. Buscar todas las secuencias de proteínas pertenecientes al cromosoma Y de primates.

La consulta no es muy diferente a las anteriores. Los primates son un orden taxonómico, así que nuevamente, se busca en el campo de organismos.

1 `Primates[Organism] AND Y chromosome`

Es importante tener en cuenta que se debe seleccionar la categoría de proteínas. La consulta nos otorga 18734 resultados (Figura 4).

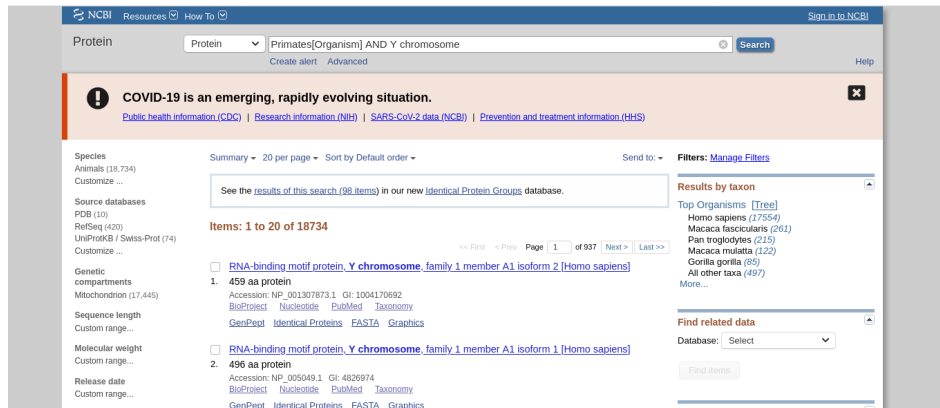


Figura 4: Búsqueda de secuencias de proteínas pertenecientes al cromosoma Y de primates.

2.5. Buscar todas las secuencias de proteínas que se han incluido en lo que llevamos de año.

Antes de nada, aclarar que los resultados que muestro (Figura 5) son de la mañana del 19 de Febrero y que, en función de cuando se realice la consulta, estos pueden cambiar. Las búsquedas temporales se indican en el campo PDAT.

1 ("2021/01/01"[PDAT] : "2021/02/19"[PDAT])

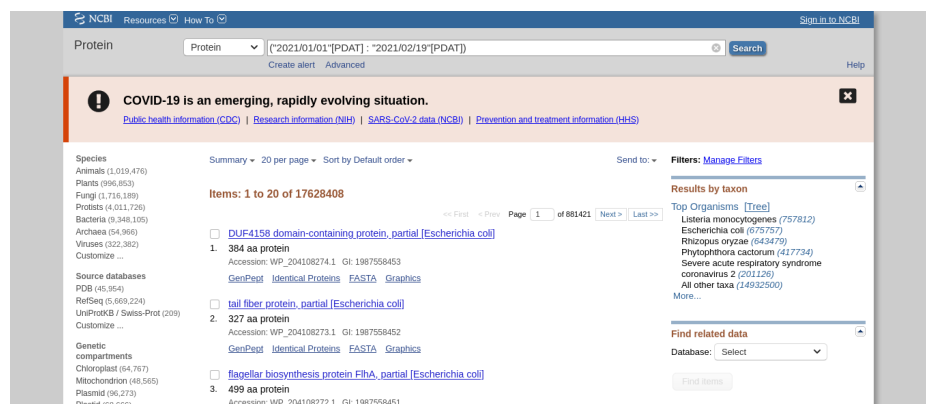


Figura 5: Búsqueda de secuencias de proteínas pertenecientes al cromosoma Y de primates.

2.6. Buscar todas las secuencias de DNA mitocondrial humano con una longitud que difiera como mucho en 1 a la de la secuencia de referencia(16569bp). Descargar las 10 primeras en formato GENBANK.

Partiendo del input del apartado 2.1, seleccionamos en la barra de la izquierda un intervalo de longitud de 16568 y 16570 y la categoría de DNA mitocondrial. Para descargarlo realizamos el mismo proceso que en el apartado 2.3, salvo que en vez de formato FASTA, elegimos GENBANK.

3. Caso Práctico: búsqueda de secuencias conectadas con un caso criminal

Tras leer el resumen del artículo [1], observamos que en el pie de la primera página se indica que todas las secuencias del paper están almacenadas en Genbank con los números de acceso AY156734 -AY156907. Por tanto, debemos ir a Genbank e introducir la siguiente consulta.

1 (AY156734[Accession] : AY156907[Accession])

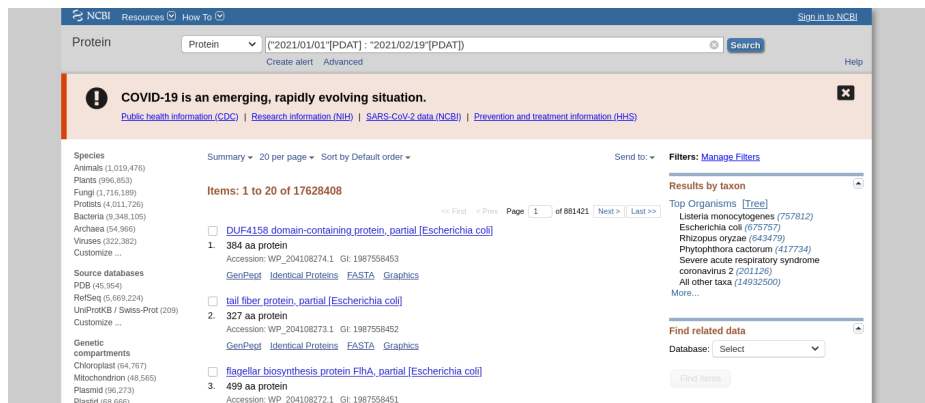


Figura 6: Secuencias de DNA mitocondrial 16568bp a 16570bp de en seres humanos.

Obtenemos 174 secuencias, que descargamos en formato FASTA (Send to → *File* → *Format : FASTA*).

4. Códigos de barras de ADN: BOLD

En BOLD el proceso de búsqueda es el siguiente. Desde la página principal seleccionamos explore the data y nos aparecerá una barra de búsqueda. Ahí introducimos el nombre de la especie en latín y arrojará un elenco con todos los códigos de barras. Entre ellos tratamos de dar con [COI-5P:995]. Se observa en pantalla un botón que permite descargar las secuencias en formato FASTA.

Una vez comprendido el procedimiento, descargaremos el código de los seres

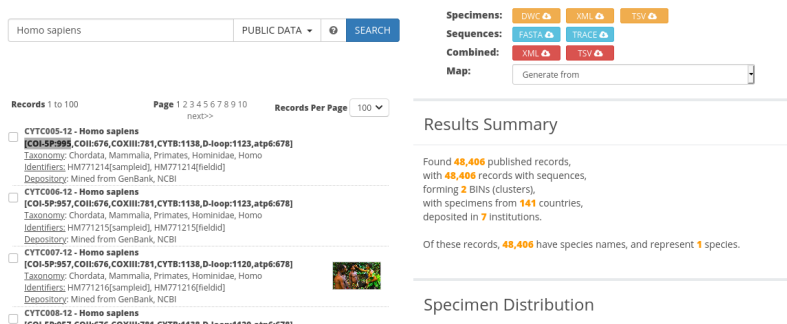


Figura 7: Resultados de la búsqueda de código de barras de seres humanos. [COI-5P:995] destacado.

humanos (*Homo sapiens*), los chimpancés (*Pan troglodytes*), el mono narigudo (*Nasalis larvatus*) y el águila solitaria (*Buteogallus solitarius*).

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.