Desarrollo

1. **Consulta de referencias bibliográficas en PubMed.**

1. Buscando la frase “hemoglobin sequence” se encontraron 8071 referencias

¿Considera apropiado el número de referencias encontradas? Si

¿Considera que las referencias encontradas para el problema son relevantes para el problema en estudio? No, ya que existe mucha variedad en las áreas de estudio.

2. Búsqueda con operaciones lógicas

Búsqueda “hemoglobin AND sequence”: resultados obtenidos: 8 071

Búsqueda “hemoglobin OR sequence”: resultados obtenidos: 173 0847

Búsqueda “hemoglobin NOT sequence”: resultados obtenidos: 192 439

¿Alguna de las búsquedas resultó equivalente a la que se realizó sin empleo

de operadores lógicos? Si, utilizando el operador AND.

¿Por qué existe una diferencia en el número de publicaciones encontradas

con los tres operadores? Porque cada sentencia lógica implica una cierta búsqueda.

Con el operador AND se debe encontrar hemoglobin y sequence en la referencia.

Con el operador OR cualquiera de los dos, incluyendo los dos juntos.

Con el operador NOT solo los que tienen hemoglobin menos sequence.

3. Emplear una herramienta del NCBI

Búsqueda “hemoglobin [title] AND sequence”: resultados obtenidos: 2104

Búsqueda “hemoglobin [title] AND sequence [title]”: resultados obtenidos: 162

En comparación con los resultados anteriores estos suelen ser más específicos.

¿Cuál de las búsquedas ha resultado ser la más eficiente hasta el momento?

La búsqueda de “hemoglobin [title] AND sequence [title]” con 162 referencias ya que son las que debe contener el título.

4.

Búsqueda “hemoglobin[Title] AND sequence[Title] AND Hill RJ[Author]”:

resultados obtenidos: 4

5.

Búsqueda “(hemoglobin [title] AND sequence [title]) OR (glucose [title] AND oxidase[title])”

resultados obtenidos: 1920

6. Búsqueda con filtros

¿Qué tipo de refinamiento se puede realizar con dichas herramientas?

Existen filtros como el más reciente, la mejor referencia, resultados por el año de publicación, y otros datos relacionados

7.

Búsqueda "American Journal of Hematology"[Journal]: resultados obtenidos: 9297

Búsqueda ("American Journal of Hematology"[Journal]) AND ("2019/01/01"[Date - Create]: "3000"[Date – Create]): resultados obtenidos 67

1. **Consulta de secuencias biológicas mediante ENTREZ.**

1.

Búsqueda “hemoglobin homo sapiens”, 17084

Búsqueda “(hemoglobin[Title]) AND homo sapiens[Organismo]”, 2030

¿Cuál de ellas ha resultado ser la más adecuada?

La más específica es la que se especifica como organismo al homo sapiens

2.

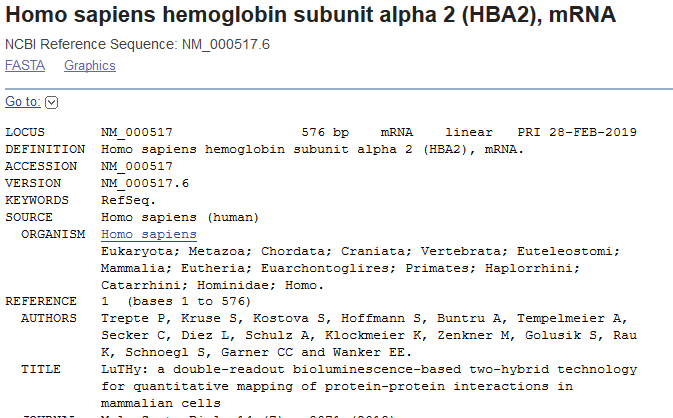
¿Cuántos registros para la cadena alfa puede observar? 485

3.

¿Cuántas secuencias se reportan en este caso para la hemoglobina? 11

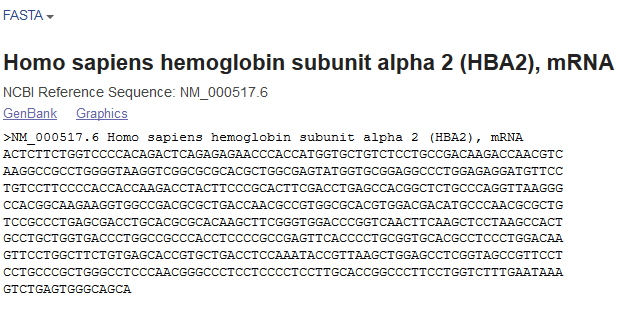
(hemoglobin[Title] AND alpha[Title]) AND homo sapiens[Organism] AND **srcdb\_refseq[PROP]**

4.

****

Guardar en registros en el formato de texto: **sequence.gb**

5.

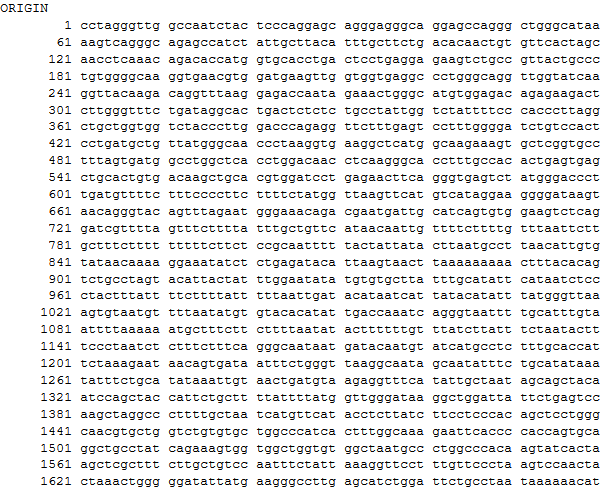
Cambie el formato del registro a “FASTA”, archivo **sequence.fasta**

1441565460 es el gi que se encontró

Cada versión posee la misma clave de acceso para el gen meintras que el gi vendría ser el identificador de cada versión para cuando se ha hecho alguna modificación de este.

7.

“(hemoglobin[title]AND Homo sapiens [Organism])”, 14313 referencias.



8. La base de datos “Gene”

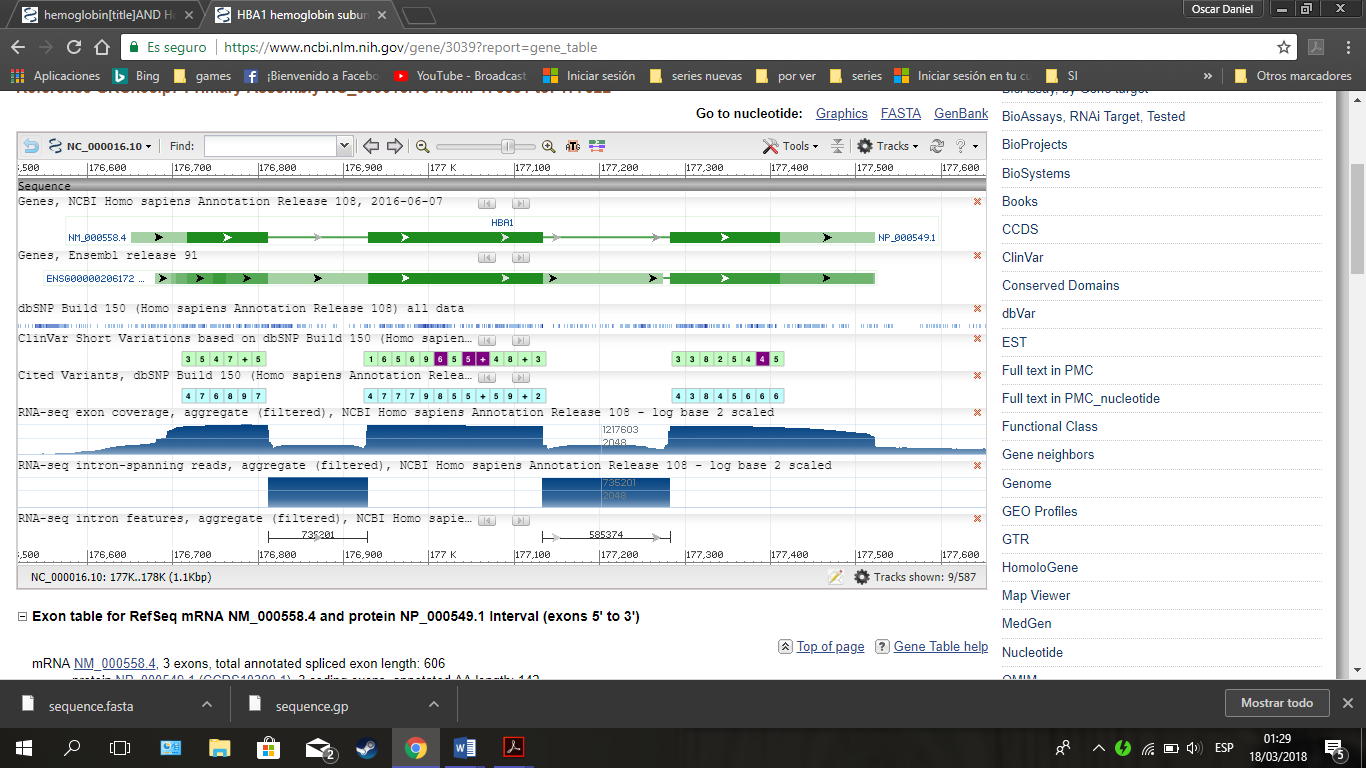
i. El símbolo oficial del gen: HBA1

ii. Otras bases de datos donde se le puede localizar: GenBank y RefSeq

iii. El cromosoma y la región cromosómica donde se localiza el gen: Se localiza en el cromosoma 16, NC\_000016.10

iv. Las versiones del ensamblado del genoma humano en las que se basa la secuencia de este gen. [Ensembl:ENSG00000206172](http://www.ensembl.org/id/ENSG00000206172) [MIM:141800;](https://www.ncbi.nlm.nih.gov/omim/141800) [Vega:OTTHUMG00000060138](http://vega.sanger.ac.uk/id/OTTHUMG00000060138)

v. El Mapa del gen mostrando la estructura del gen, en la que pueden identificarse (entre otras), exones, intrones, regiones codificantes (CDS), variaciones genéticas (SNPs), datosde ensamblado, datos de expresión (RNAseq)



vi. Exones, intrones, posición del gen en el cromosoma, posición de intrones y exones, localización de regiones codificantes y no codificantes.

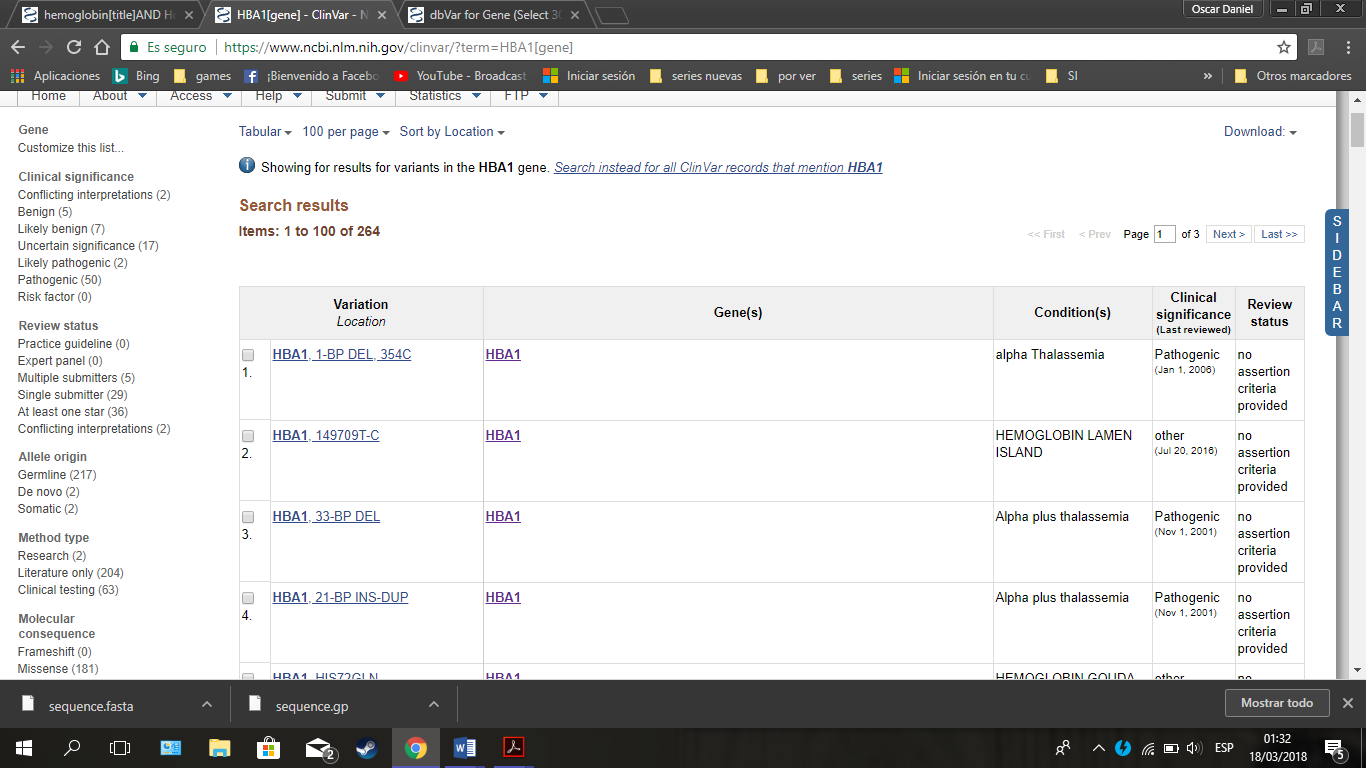
vii. Claves de acceso de la proteína, el RNA mensajero, claves de acceso relativas al ensamblado.

viii. Referencias bibliográficas relevantes.

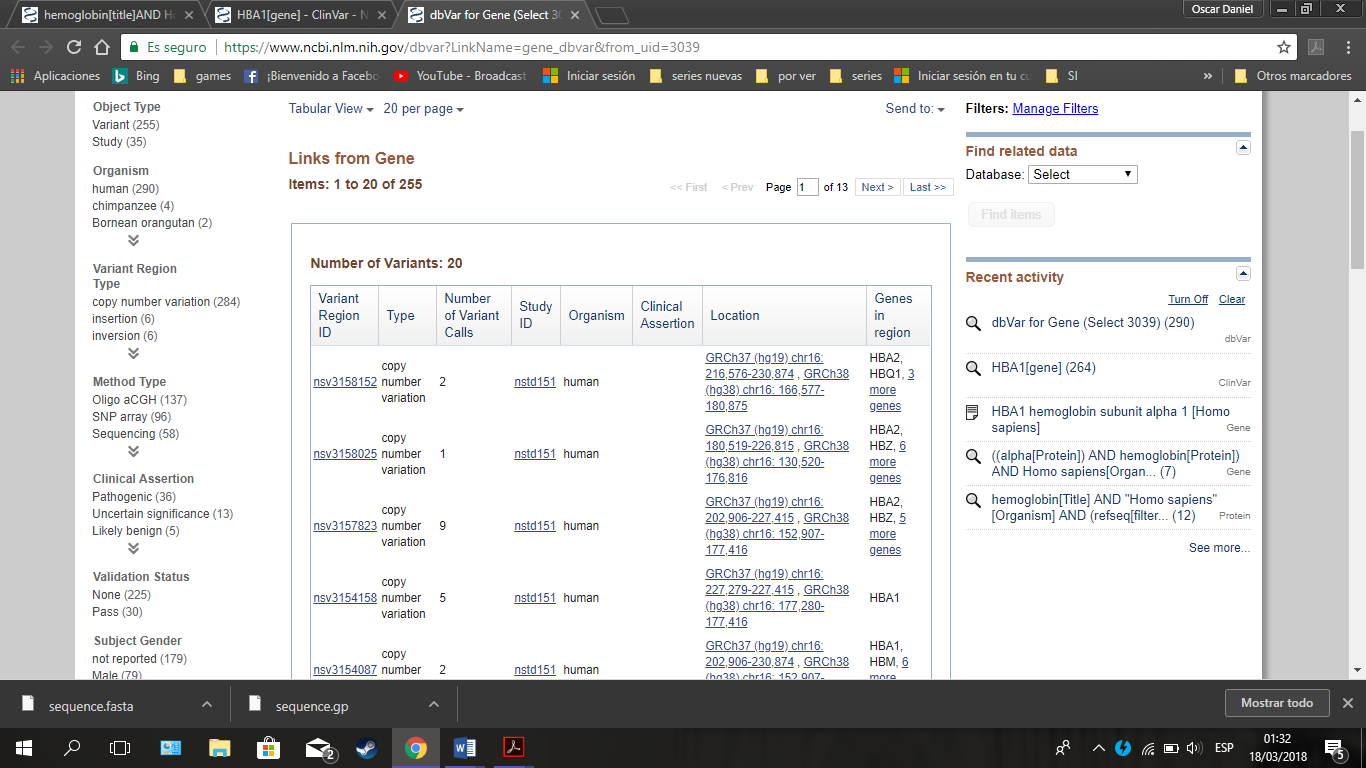
ix. Referencias bibliográficas relacionadas con la función del gen (GeneRIFs: Gene References Into Functions).

x. Fenotipos asociados con enfermedades humanas.

xi. Variaciones del gen (ClinVar, dbVar, SNPs).

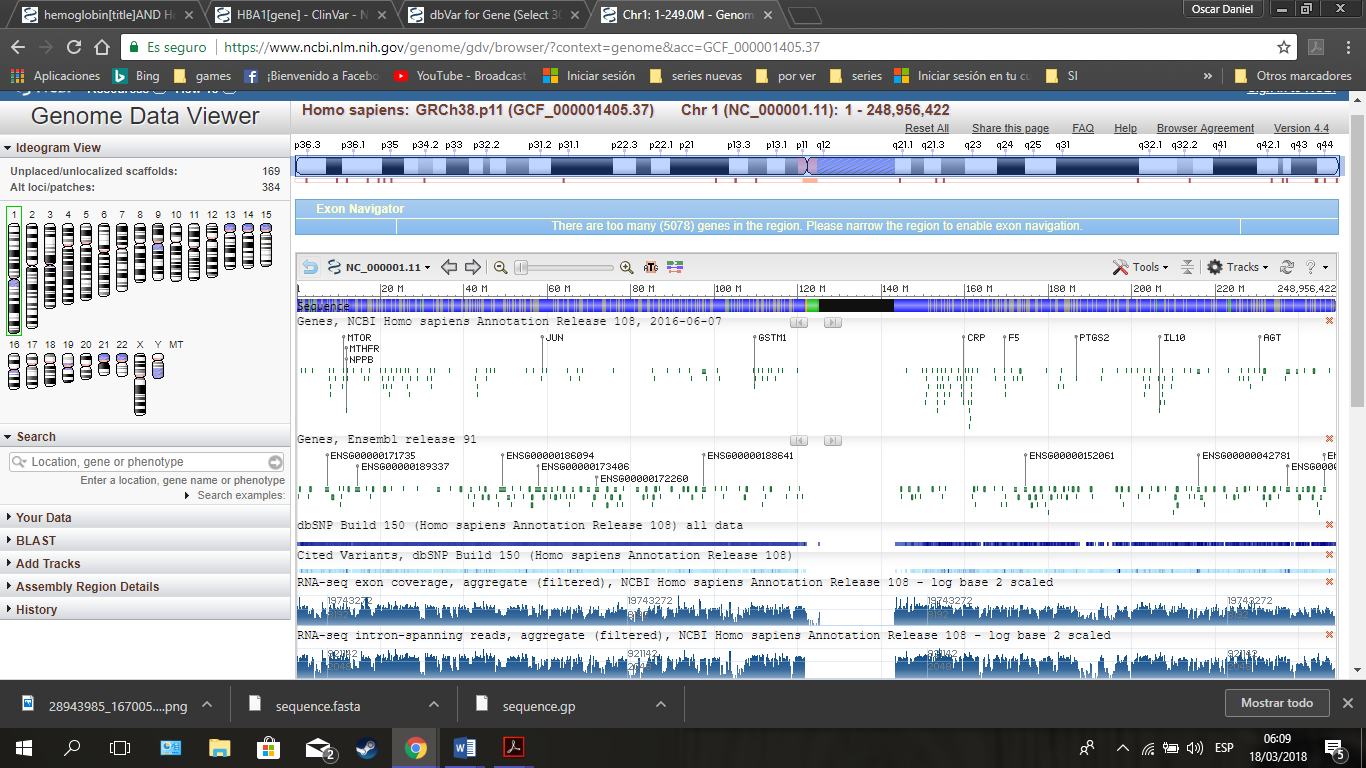


ClinVar



dbVar

9. Exploración del genoma humano.



**EJERCICIOS DE EVALUACION.**

Elaborar un reporte de texto con la siguiente información:

**E1. Las secuencias de proteínas en formato FASTA de las cadenas alfa y beta de la hemoglobina humana descargadas en esta práctica y las secuencias de nucleótidos de los mRNA completas que las codifican también en formato FASTA.**

hemoglobin subunit alpha [Homo sapiens]

>NP\_000549.1

MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNA

VAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSK

YR

hemoglobin subunit beta [Homo sapiens]

>NP\_000509.1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLG

AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVAN

ALAHKYH

Homo sapiens hemoglobin subunit alpha 1 (HBA1), mRNA

>NM\_000558.4 Homo sapiens hemoglobin subunit alpha 1 (HBA1), mRNA

CATAAACCCTGGCGCGCTCGCGGCCCGGCACTCTTCTGGTCCCCACAGACTCAGAGAGAACCCACCATGGTGCTGTCTCCTGCCGACAAGACCAACGTCAAGGCCGCCTGGGGTAAGGTCGGCGCGCACGCTGGCGAGTATGGTGCGGAGGCCCTGGAGAGGATGTTCCTGTCCTTCCCCACCACCAAGACCTACTTCCCGCACTTCGACCTGAGCCACGGCTCTGCCCAGGTTAAGGGCCACGGCAAGAAGGTGGCCGACGCGCTGACCAACGCCGTGGCGCACGTGGACGACATGCCCAACGCGCTGTCCGCCCTGAGCGACCTGCACGCGCACAAGCTTCGGGTGGACCCGGTCAACTTCAAGCTCCTAAGCCACTGCCTGCTGGTGACCCTGGCCGCCCACCTCCCCGCCGAGTTCACCCCTGCGGTGCACGCCTCCCTGGACAAGTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAATACCGTTAAGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCCTGCACCCGTACCCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGCAAAAAAAAAAAAAAAAAAAAAA

**E2. Elaborar en cuadro en el cual se resuma la siguiente información para las proteínas estudiadas:**

**Clave de acceso del cromosoma, la proteína y el mRNA; número de intrones y exones, longitud del gen, longitud de la proteína, longitud del mRNA, posiciones en el cromosoma de cada gen, posiciones de los intrones y posiciones de las regiones UTR.**

**E3. En un cuadro enumerar las diferencias existentes entre la anotación de los genes de la hemoglobina alfa y beta en el NCBI y en Ensembl.**

**E4. Utilizando la sección de genomas del GenBank realizar la búsqueda de genes para 16S rRNA s en el genoma de Escherichia coli K-12.**

**PREGUNTAS** (CONTESTAR BREVEMENTE):

1. ¿Cuántos genomas eucarióticos y procarióticos se han secuenciado de manera completa en la actualidad?

Se han secuenciado 5376 genomas eucarióticos y 134147 genomas procarióticos

2. Explique el concepto de redundancia de la base de datos y explique en que consiste la base de datos RefSeq del NCBI.

Refseq excluye artículos de un mismo tema, pero con diferente versión o muy parecidos para que las respuestas sean menores.

3. Investigue en que consiste la base de datos Uniprot.

Es una base de datos dedicada al secuenciamiento de proteínas.

4. ¿Cuál es la diferencia entre la clave de acceso (accession) y el GenInfo Identifier (gi) que aparece en los registros de secuencias del NCBI? ¿Cuál es la utilidad de estas claves?

Las claves de acceso identifican a una proteína, nucleótido, etc. Mientras que la clave Gi es una identificación en la base de datos según su versión.