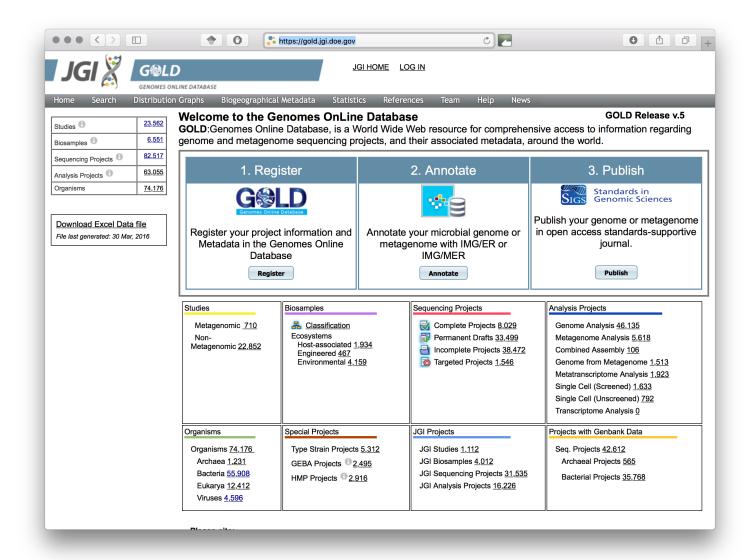
Laboratorio 03 - Ensamblaje de genomas y predicción de genes

En este laboratorio vamos a familiarizarnos con publicaciones de genomas, y como fueron ensamblados. También vamos a explorar dos métodos de predicción de genes

Parte 1: El artículo genoma

Comencemos por ir a la Genomes on Line Database (GOLD) y escoger un genoma de interés

• Ve a la base de datos GOLD y busca un genoma eucarionte de interés.

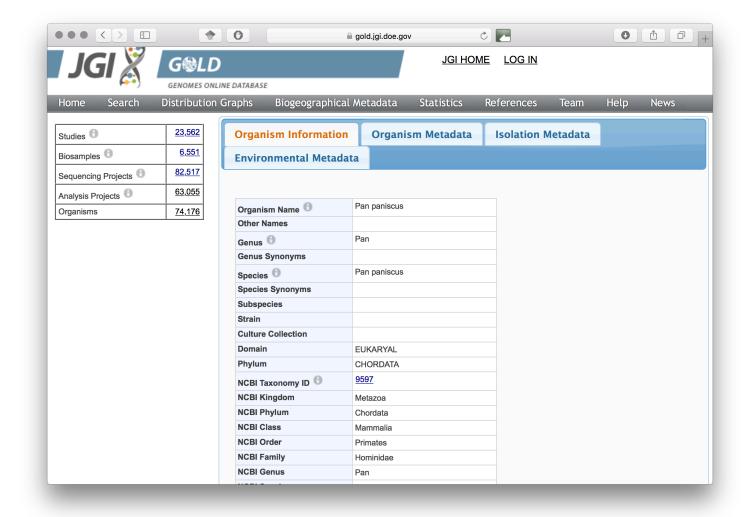


 Navega dentro del portal siguiendo vínculos a, por ejemplo, "Complete projects", "Eukarya" o la casilla de búsqueda.

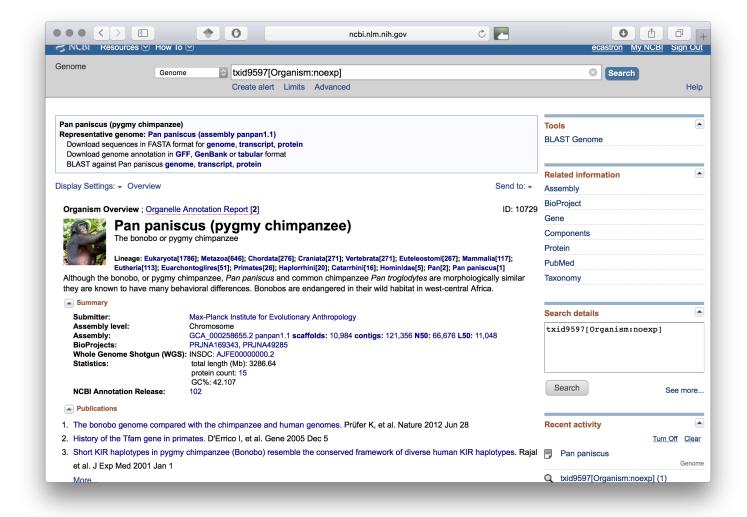
Responde:

¿Cuántos genomas han sido depositados en GOLD? ¿Son los mismos de GENBANK? ¿Cuál es la distribución de procariontes y eucariontes secuenciados?

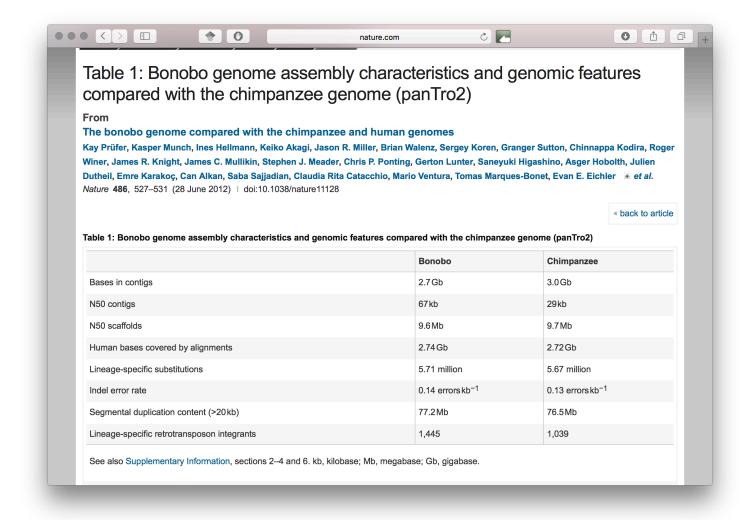
En este ejemplo continuaremos con el genoma del Bonobo *Pan paniscus*



- A partir de la ficha GOLD de tu organismo, busca la publicación de su genoma (pista: sigue el Taxonomy ID)
- Busca el vínculo a la base de datos de genomas de NCBI y ubica la referencia al artículo original



- Ahora solo tienes que conseguir el artículo y buscar en los materiales y métodos par responder lsa siguientes proguntas. Existen una serie de métricas o estadísticas que se usan para evaluar un ensamblaje de genomas.
- Sigue este vínculo y revisa las siguientes N50, L50, y NG50.



Responde:

```
¿Qué es el N50, L50, NG50?
¿Cuál es el propósito de calcular estas estadísticas?
¿Cuál es el genoma que escogiste? Adjunta la referencia.
¿Cuál es el N50 del genoma que escogiste? ¿Y el NG50?
¿Qué tipo de tecnología se uso para secuenciar el genoma que escogiste?
¿Cuántos cromosomas tiene tu organismo y cuál es su tamaño?
```

Parte 2: Predicción de genes

En esta parte vamos a utilizar dos programas clásicos para predecir genes ORFfinder y GLIMMER.

- Tu misión para esta parte es descargar una secuencia y predecir qué genes están presentes usando ambas herramientas.
- Descarga la secuencia problema <u>aquí</u> o copia y pega desde:

ATGAGTATCAAAATCTTATCTGAATCAGAAATTAAACAAGTAGCAAATTCATATCAAGCCCCAGCGGTTTTATTTGCCAAT ATCGAGCCACTAAATGCCAAAACCTTTAAGAGAAACAGTATCTGGCGTGAATACTTAACAGAAATTCTTGATGAAATAAAG CCCAAAGCTAACGAGCAAATTGCTGCAACAATTGAATTTCTTGAAAAAGCCTCTTCCGCTGAATTAGAAGAAATGGCAAAT AAACTCTTAGCACAAGAATTTAACTTAGTCAGCAGTGATAAAGCCGTCTTTATTTGGGCTGCACTTTCCCTTTATTGGTTA ${\tt CCTGTGGCAAGTATGGTACAATTGGTACATCACAAGGTTTACGCTACTTACATTGTAATTTATGTGAAAGTGAATGGAAT}$ TTGGTACGCGCACAATGCACCAATTGTAATAGTCATGACAAACTCGAAATGTGGTCACTAAATGAAGAACTTGCGCTTGTT CGTGCCGAAACCTGTGGTAGTTGTGAAAGTTACTTGAAAATGATGTTCCAAGAAAAAGATCCTTACGTAGAACCTGTAGCC CCTGCAGAAGAAGCATAAAAATATAGCCTAGAAATATCTAGGCTAATTATTTAAATCTATAGATAACACGCCATTACCACT GCATTTTCTCGTCCACCACTGGGTTTCGGATAATAATTTTTCCGAATATCCACTTCATTAAAACCAATTTTTTCATACAAG AAACGAGCAGAGTTAGACTCTCTGACTTCTAGCCATAAGGTTTGGACACCTTTTTTCCTTTAATTGAAAGATTAATTTTCCT AACAATAATTTGCCAAATCCACAACCTTGATAAGTAGGCAAAATCGCAATATTAAACAAAGTCGCTTCATCCAATACTGTT TGGCAAATAGCAAAACCGATAATCTGATTATTCTCTATTAATTTTAAATTGAGATAACGCTCCCCTTGATTATTTTTAAC GTACCAAATGACCAAGGCACCAGATGGGCTTGCTGTTCGATTTCATACAATCGCTCGAAATCACAGGCTTCAATTTGAGAA ATAATAGACATAATTAAGGCTGCTGAATTTGTTGCCACAACGCTCGTTTGGCTTGATGATTAGATTGAAATTGCTGCCAAC TTGGCGAGCGATAAACCTGCTCAGCCTGCTTGCAAAATGGCAAAGTGCGGTCAATTTGGTCGCTATTTTCTGATAGTAACC AATAACGAATAGGCTGTTTACATTCCATATGCTGGATTTGATCGTAATTCAAACATAAACAATTTTCTTTTTAAGATTAA GGCTTAACAGCACATCAGCCAACAAAGGCGAGCTACTGATATTTTCATCGGAAACAGTGATAAGGCGAATATTCTCTGCCA CACTAATTCCTACTGAACCTTGCAGTACCTCGGGGCGATATAATTCCCACTGGGAAATGCCCATTTCTTGTAAAAGAAGAT CGCGTCTGTTCATAAGTGAATTTTTTAAAAACGTTGCTTGAAAAATAGTGTTAATTTCTTTTATTAATCTTTGCTAAAATGG TGCGTAATTTTAACCAATAACCCACAGGATTCAAAGCG

Responde:

¿Cuántos ORF o genes encontró ORFfinder? ¿Cuántos ORF o genes encontró Glimmer? ¿Alguno de los genes predichos por estas herramientas coinciden? ¿En qué hebra están codificados? ¿Qué tipo de programa es GLIMMER? ¿Ab initio o por homología?

Trabajo de laboratorio para la próxima semana

El trabajo de laboratorio para la próxima semana consta de dos partes. La primera parte ya la tienes lista. Simplemente tienes que responder las preguntas que aparecen a través de esta guía y enviar un informe a bioinformatica.unab2016@gmail.com. Envien a este correo los informes hasta el jueves de la semana siguiente a la realización del práctico. La hora límite es las 23:59. En el Asunto del correo pongan Informe de Laboratorio 0x así nosotros podemos clasificar automáticamente los informes. Para la próxima semana el Asunto del correo debería ser Informe de Laboratorio 03.

Profesor	Nombre	Correo electrónico
Profesor responsable	Dr. Eduardo Castro Nallar	eduardo.castro@unab.cl
Profesor ayudante sección 1	Ingrid Araya Durán	ingrid.araya.duran@gmail.com
Profesor ayudante sección 1	Sandro Valenzuela	sandrolvalenzuelad@gmail.com
Profesor ayudante sección 2	Javier Cáceres	ja.caceresmolina@gmail.com
Profesor ayudante sección 2	Consuelo Bello	consuelobelloz@gmail.com

La segunda parte tiene que ver con leer un artículo científico sobre este tema: Mihai Pop. 2013. *Sequence assembly demystified*. **Nature Reviews Genetics** 14, 157-167 l doi:10.1038/nrg3367.

Puedes acceder al artículo <u>aquí</u>. Recuerda que el contenido de este artículo es el material para el próximo control de entrada.