

Introducción a la Bioinformática - Trabajo Practico 1

Fecha de entrega 25 de Octubre

● **Ejercicio 1.** Escribir un script que lea una o más secuencias (de nucleótidos) de un archivo en formato GenBank, las transcriba a su secuencia de amino ácidos correspondiente y la escriba en un archivo en formato FASTA. Deben generarse su archivo GenBank de secuencias input, por ejemplo entrando al NCBI-Gene realizar una consulta por “globulin” y obtener uno o más resultados de la base de datos de nucleótidos en formato GenBank en un archivo de texto. (Si tienen algún problema en tu dir de bioperl-1.6/t/data hay varios archivos *.gbk)

- Input: Archivo de secuencias Genbank (ej. Xxxxx.gbk con una o más secuencias)
- Output : Archivo de secuencias Fasta (ej. Xxxxx.fas con una o más secuencias de amino ácidos)

Deben entregar el script Ex1.pm y su input file con una breve descripción para que lo pruebe.

● **Ejercicio 2.** Escribir un script que realice un BLAST de una secuencia (o varias) y escriba el resultado en un archivo. Nota: Pueden ejecutar BLAST de manera remota o bien localmente (si hacen ambos les doy más puntos!), para esto deben instalarse BLAST localmente del NCBI, luego bajarse la base de datos <ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/swissprot.gz> y descomprimirla en un dir por ej. ncbi-blast-2.2.27+/data/, luego usar el comando ncbi-blast-2.2.27+/bin/makeblastdb sobre el archivo swissprot (el original esta en formato FASTA) para darle formato BLAST.

- Input: Secuencia Fasta (ej. Xxxxx.fas con una o más secuencias)
- Output: Repote Blast (ej. blast.out, si deciden hacer múltiples pueden generar un único o varios archivos)

Deben entregar el script Ex2.pm y su input file con una breve descripción.

● **Ejercicio 3.** Escribir un script que analice (parsear) un reporte blast e identifique los hits que en su descripción aparezca un pattern determinado que le pidamos. El pattern puede ser una palabra. Nota: Si quieren algún punto adicional pueden parsear el ACCESSION del hit seleccionado y con el modulo Bio::DB::GenBank obtener la secuencia completa en formato FASTA y escribirla un archivo.

- Input: Reporte Blast (ej. blast.out) y un pattern (ej. Arabidopsis)
- Output: Lista de los hits que coincidan con el pattern (ej. solo los hits de Arabidopsis)

Deben entregar el script Ex3.pm y su input file con una breve descripción.

● **Ejercicio Bonus.** Instalar EMBOSS. Escribir un script BioPerl que analice el contenido de una secuencia con algunos programas EMBOSS que ustedes crean convenientes como newcpgreport, remap, etc. y generar un reporte o mostrar el resultado.