

Bioinformática y análisis de datos ómicos

© Ignacio Varela Egocheaga

Este material se publica bajo licencia Creative Commons CC BY-NC-SA 4.0



EJERCICIOS TEMA 8

Alumno:

Realiza las distintas actividades de abajo copiando debajo de cada actividad los comandos necesarios. Para ello descárgate la [carpeta para realizar los ejercicios del Tema 8](#). Finalmente, convierte este documento a PDF y entrégalo a través del moodle.

1.- En la carpeta correspondiente al Tema 8 encontrarás dos archivos de reads de Illumina correspondientes a una secuenciación metagenómica. Realiza un análisis de calidad utilizando Fastqc. Pega aquí el gráfico en el que se puede ver la calidad de los reads.

2.- Utiliza la aplicación SICKLE para limpiar los reads y eliminar aquellos que no pasan los criterios de calidad. Utiliza como criterio de calidad una precisión del 99.9% en el base calling y una longitud mínima de 50. La calidad está expresada de acuerdo al software de Illumina CASAVA 1.5. Genera la salida en formato comprimido con gzip.

3.- Vuelve a comprobar la calidad de los reads procesados usando de nuevo FastQC. ¿Cuántos reads han pasado el corte de calidad?

4.- Ensambla los reads en contigs utilizando MEGAHIT usando los parámetros por defecto.

5.- Alinea los archivos fastq limpios frente a los contigs generados en el archivo final.contigs.fa usando BWA

7.- Convierte el archivo de alineamiento de .sam a .bam, ordénalo e indexalo con SAMTOOLS

8.- Predice los MAGs (metagenome-assembled genomes) utilizando METABAT2 utilizando un tamaño mínimo de contig de 1500 nt. Para ello utiliza primero la herramienta `jgi_summarize_bam_contig_depths` para generar el archivo de abundancias (depths) a partir del bam y luego utiliza este archivo para correr metabat2. Sigue las instrucciones en este enlace: <https://gensoft.pasteur.fr/docs/MetaBAT/2.15/>

9.- Extrae los ORFs de tus MAGs utilizando PRODIGAL del primero de tus MAGs/genomas. Identifica la primera proteína identificada usando BLAST.