

Optimizando el análisis de Metagenómica por Shotgun: MeTaxFun

Isabel M. Cerezo ¹, Ana M. Cabello ², Isabel Ferrera ², Belén Delgado ¹, Miguel A. Moriñigo ³, Silvana T. Paniagua ³, Rocío Bautista ¹

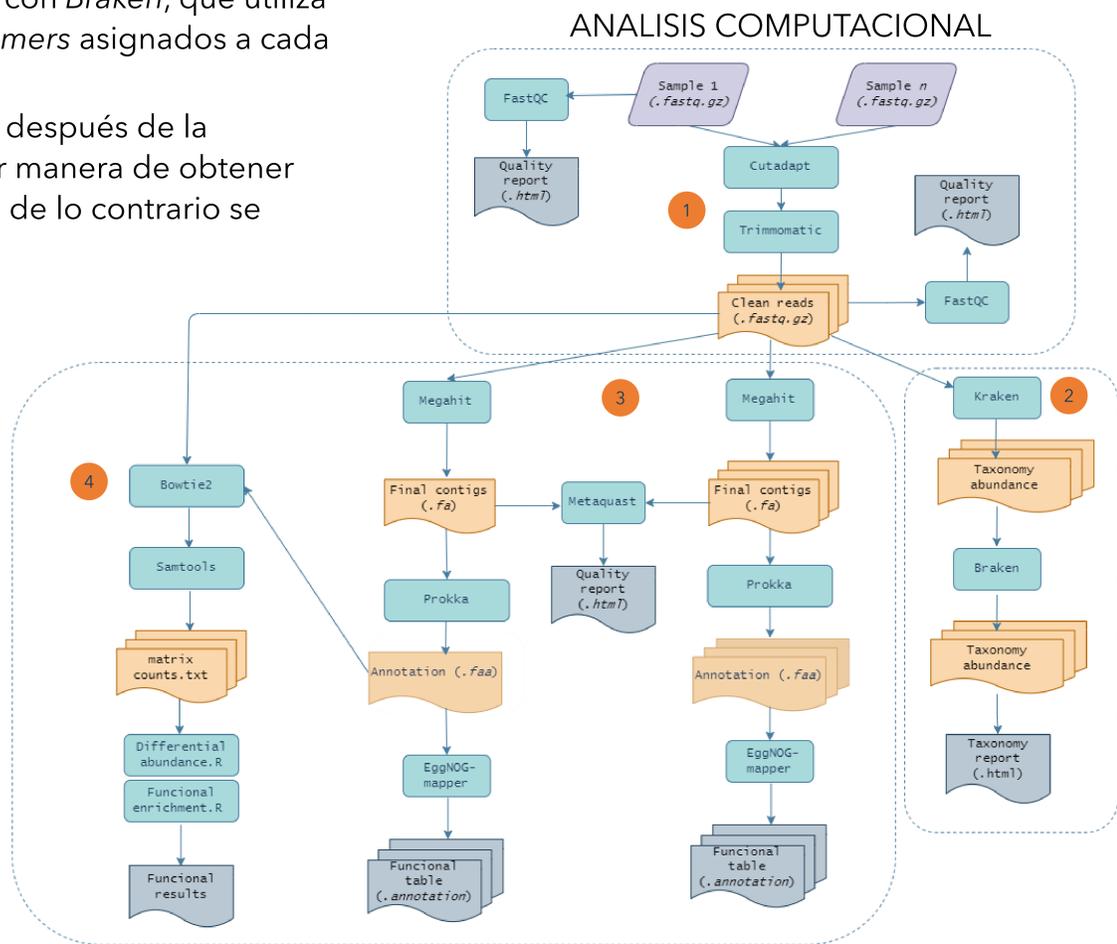
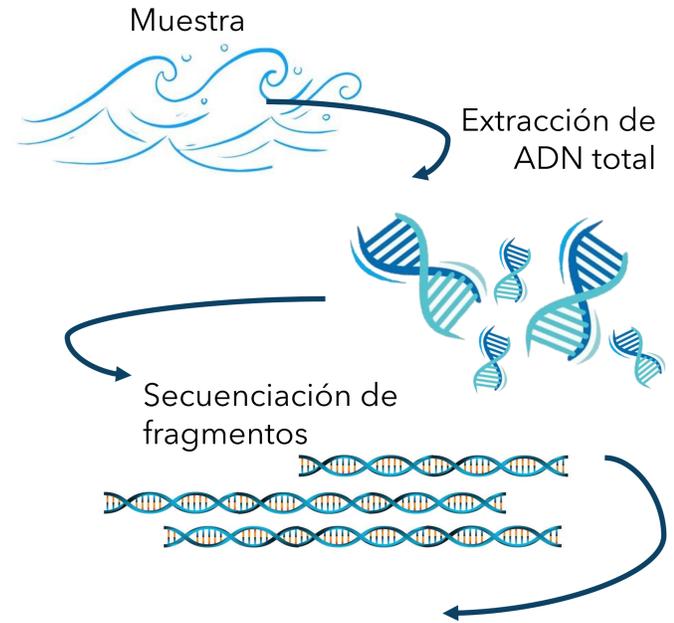
¹ Unidad de Bioinformática, SCBI, Universidad de Málaga, Málaga, 29590 España

² Centro Oceanográfico de Málaga, IEO-CSIC, Fuengirola, 29640 España

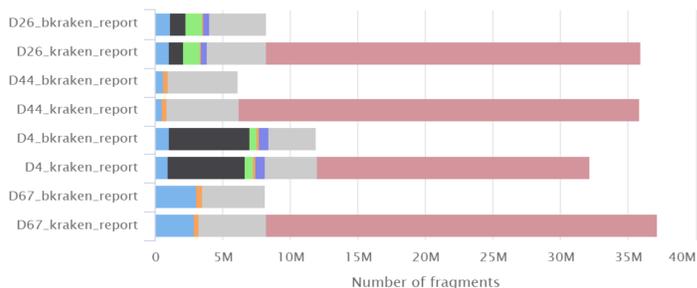
³ Departamento de Microbiología, Facultad de Ciencias, Universidad de Málaga, Málaga, 29010 España

La **metagenómica por shotgun** es una técnica utilizada para estudiar todo el material genético de **comunidades microbianas** presentes en **muestras ambientales o biológicas**. La gran cantidad de datos generados por esta técnica proporciona una comprensión profunda de la diversidad genética y funcional de los microorganismos en un ecosistema. Sin embargo, el análisis de los datos es un proceso difícil y complejo que requiere de herramientas informáticas avanzadas y una comprensión detallada de los procesos de análisis

- 1) Procesamiento de secuencias** teniendo en cuenta la longitud de las lecturas, la calidad de la secuenciación, el número de indeterminaciones o la presencia de adaptadores propios del proceso de secuenciación.
- 2) Análisis taxonómico por muestra.** Hemos mejorado el clasificador taxonómico *Kraken* mediante la incorporación de una normalización con *Braken*, que utiliza un enfoque estadístico basado en la frecuencia de los *k-mers* asignados a cada especie para estimar su abundancia relativa (Fig 1). Además, comparamos la asignación taxonómica antes y después de la reconstrucción de los *contigs* y, concluimos que la mejor manera de obtener información es a partir de las lecturas procesadas ya que de lo contrario se reduce la diversidad de la muestra (Fig 2).
- 3) Reconstrucción del metagenoma** por muestra y metagenoma común. *Megahit* divide las secuencias en *k-mers*, y construye un grafo de *Bruijn* identificando regiones repetitivas, superposición y ambigüedad, logrando *contigs* más largos y precisos (Fig 3). Tras ello, *Prokka* anotó los genes identificando las regiones codificantes y *EggNOG-mapper* asignó las funciones específicas basadas en su base de datos proporcionando información sobre la función biológica y la categoría funcional de los genes anotados.
- 4) Abundancias diferenciales.** Se utilizó *Bowtie2* para estimar el número de genes en cada muestra a partir del metagenoma común, calculando su abundancia (Fig 4). Posteriormente, se estimó la diferencia en la abundancia de los genes y en el enriquecimiento funcional utilizando *scripts* desarrollados en R.



Kraken 2: Top taxa



QUAST: Number of Contigs

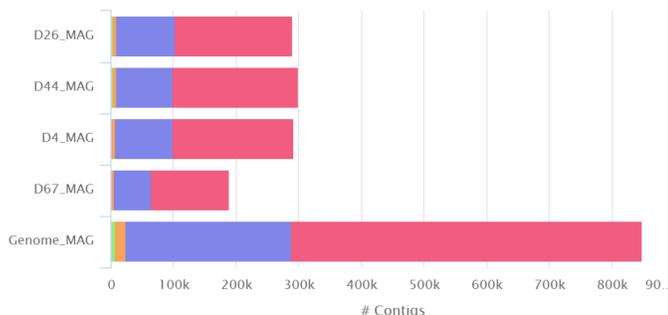


Fig 1: Taxonomía a nivel de especie usando Kraken y Braken

Fig 3: Número de contigs reconstruidos por muestra y metagenoma común tras usar megahit



Fig 2: Taxonomía general según la normalización de Braken, en Krona (interactivo)

Bowtie 2: PE Alignment Scores

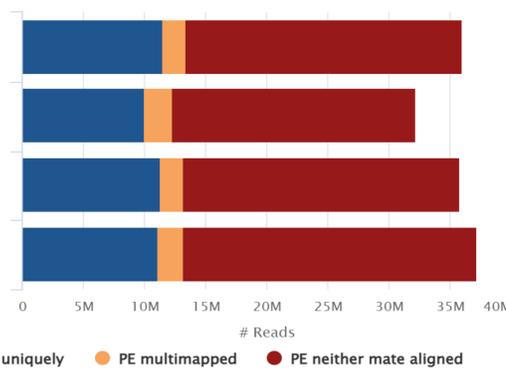


Fig 4: Lecturas mapeadas de cada una de las muestras sobre el metagenoma común usando bowtie2

MeTaxFun (Shotgun METagenomic - TAXonomic and FUNctional analysis) fue diseñado para proporcionar una solución global, eficiente y confiable para el análisis de los datos.

A partir de este flujo se puede obtener gráficos y resultados a nivel taxonómico y funcional para cada una de las muestras individuales, así como para un metagenoma ambiental

