

# Archivo de Bases de Datos para Análisis Metagenómicos<sup>1</sup>

Alejandro Navas González

2023-06-16

<sup>1</sup>Andera Projekt

# Introducción a las Plataformas y Bases de Datos para la Metagenómica y la Ecología Microbiana

La metagenómica es una rama de la genómica que se centra en el estudio de muestras tomadas directamente de ambientes naturales. Esta disciplina ha revolucionado nuestra comprensión de las comunidades microbianas, permitiéndonos explorar la diversidad y la función de los microbios en una variedad de contextos, desde el cuerpo humano hasta los océanos y el suelo.

Ahora bien, el volumen de datos generados exige a su vez una forma de organizar esta información, y ahí es donde entran las bases de datos y plataformas de la metagenómica. Éstas conceden acceso a una gran cantidad de datos genéticos y metagenómicos que pueden ser utilizados para identificar y caracterizar nuevas especies microbianas, entender las interacciones entre diferentes especies microbianas y su entorno, y explorar cómo las comunidades microbianas cambian en respuesta a diferentes factores ambientales y perturbaciones.

Además, cabe mencionar que no solamente proporcionan los datos brutos necesarios para la investigación metagenómica, sino que también ofrecen herramientas y recursos para el análisis y la interpretación de estos datos. Esto incluye herramientas para la anotación de secuencias, la comparación de comunidades microbianas, la creación de grafos, la identificación de funciones metabólicas y la visualización de datos entre otras.

## NCBI (National Center for Biotechnology Information)

Es una organización gubernamental de los Estados Unidos que proporciona acceso a una gran cantidad de información sobre genómica, genética y biología molecular. En el contexto de la metagenómica y la ecología microbiana, NCBI ofrece una variedad de bases de datos y herramientas útiles, como la base de datos de secuencias genéticas GenBank y la base de datos de secuencias de lectura corta, SRA ([“National Center for Biotechnology Information,” n.d.](#)).

- NCBI SRA (Sequence Read Archive): Es una base de datos que almacena secuencias de lectura corta generadas por tecnologías de secuenciación de próxima generación. SRA es una fuente crucial de datos para la metagenómica y la ecología microbiana, ya que permite a los investigadores acceder a datos de secuenciación de una amplia variedad de proyectos y estudios ([“NCBI Sequence Read Archive,” n.d.](#)).

## European Bioinformatics Institute (EBI)

Es una organización de investigación que forma parte del Laboratorio Europeo de Biología Molecular. EBI proporciona acceso a muchas bases de datos y herramientas útiles en el campo de la bioinformática, in-

cluyendo la base de datos de secuencias de ADN y proteínas, Ensembl, y la base de datos de metagenómica ([“European Bioinformatics Institute,” n.d.](#)).

- EBI Metagenomics: Es una base de datos y una plataforma de análisis que permite a los investigadores depositar, buscar y analizar datos metagenómicos. Esta plataforma es especialmente útil para los estudios de ecología microbiana que se centran en la secuenciación de ADN ambiental ([“EBI Metagenomics,” n.d.](#)).

## HMP DACC (Human Microbiome Project Data Analysis and Coordination Center)

Es una iniciativa que tiene como objetivo caracterizar y analizar la microbiota humana para entender su papel en la salud y la enfermedad. Los datos generados por el HMP son de gran valor para los estudios de metagenómica y ecología microbiana ([“Human Microbiome Project Data Analysis and Coordination Center,” n.d.](#)).

## Metagenomic Rapid Annotations using Subsystems Technology (MG-RAST)

MG-RAST una plataforma de análisis de metagenómica que proporciona acceso a datos y herramientas para el análisis de secuencias metagenómicas. MG-RAST es útil para los investigadores en ecología microbiana que buscan entender la diversidad y la función de las comunidades microbianas ([“Metagenomic Rapid Annotations using Subsystems Technology,” n.d.](#)).

## iMicrobe

iMicrobe es una plataforma que proporciona acceso a datos y herramientas para la investigación en microbiología y metagenómica. iMicrobe incluye datos de

proyectos como el Proyecto de Microbiología Marina de la Fundación Gordon y Betty Moore ([“iMicrobe,” n.d.](#)).

## Joint Genome Institute (JGI)

Es un instituto de investigación que proporciona acceso a una variedad de recursos genómicos, incluyendo secuenciación de ADN y análisis de metagenómica. JGI es una valiosa fuente de datos para los estudios de ecología microbiana ([“Joint Genome Institute,” n.d.](#)).

## European Nucleotide Archive (ENA)

Es una base de datos que almacena secuencias de nucleótidos y información relacionada. ENA es una fuente importante de datos para la metagenómica y la ecología microbiana ([“European Nucleotide Archive,” n.d.](#)).

## DNA Data Bank of Japan (DDBJ)

Es una base de datos que almacena secuencias de ADN y ARN y anotaciones asociadas. DDBJ es una fuente importante de datos para la metagenómica y la ecología microbiana, ya que proporciona acceso a secuencias genéticas de una amplia variedad de organismos, incluyendo microbios ([“DNA Data Bank of Japan,” n.d.](#)).

# Conclusión

En resumen, estas bases de datos y plataformas son fundamentales para la metagenómica y la ecología microbiana al facilitar la exploración y el entendimiento de la diversidad y la función de las comunidades microbianas, siendo así esenciales para avanzar en nuestra comprensión de la biología microbiana.

El conocimiento de estas bases se convierte entonces un pilar de la metagenómica, si bien éste es solamente el primer paso para efectuar un proyecto. Los datos de secuenciación de estas plataformas requerirán de un procesamiento y limpieza antes de que puedan ser analizados de forma adecuada. Este procesamiento puede implicar la eliminación de secuencias de baja calidad, la agrupación de secuencias en unidades taxonómicas operativas (OTUs) y la anotación taxonómica de las OTUs. Muchas de estas tareas pueden ser realizadas con herramientas de bioinformática como **QIIME** y **Mothur**, y más adelante los datos resultantes del procesamiento podrán ser importados en R para su análisis con **Phyloseq** u otras librerías.

# Referencias

- “DNA Data Bank of Japan.” n.d. <https://www.ddbj.nig.ac.jp/>.
- “EBI Metagenomics.” n.d. <https://www.ebi.ac.uk/metagenomics/>.
- “European Bioinformatics Institute.” n.d. <https://www.ebi.ac.uk/>.
- “European Nucleotide Archive.” n.d. <https://www.ebi.ac.uk/ena>.
- “Human Microbiome Project Data Analysis and Coordination Center.” n.d. <https://www.hmpdacc.org/>.
- “iMicrobe.” n.d. <https://www.imicrobe.us/>.
- “Joint Genome Institute.” n.d. <https://www.jgi.doe.gov/>.
- “Metagenomic Rapid Annotations using Subsystems Technology.” n.d. <https://www.mg-rast.org/>.
- “National Center for Biotechnology Information.” n.d. <https://www.ncbi.nlm.nih.gov/>.
- “NCBI Sequence Read Archive.” n.d. <https://www.ncbi.nlm.nih.gov/sra>.