



Caso de Estudio Complejivo

Fecha de inicio:	27 de enero del 2026
Fecha de finalización	03 de febrero del 2026
Nombre del caso	Análisis metataxonómico
Número Caso	1
Data para el caso	Secuencias.rar

Problemática:

La contaminación por explosivos puede generar una alteración importante de la vegetación y la microbiota en suelos afectados. Los microorganismos y plantas capaces de sobrevivir en suelos contaminados por explosivos podrían tener un rol importante en la atenuación natural. El problema aquí planteado es parte de un análisis más amplio en el que se comparó los perfiles taxonómicos de microorganismos de raíces y suelos contaminados y no contaminados con explosivos.

Como resultado de un análisis metataxonómico se busca determinar si existen diferencias significativas entre bacterias asociadas con plantas presentes en suelos contaminados por explosivos y suelos no contaminados del mismo sector. También se espera identificar molecularmente a los microorganismos presentes en las muestras, especialmente los que están presentes en suelos contaminados. Para obtener los datos se secuenció un fragmento del gen 16S del ADN total obtenido de raíces utilizando la tecnología Illumina. El set de datos provisto consiste en secuencias crudas de tres réplicas de cada tratamiento

Tratamiento 1: suelos no contaminados
1MR1, 1MR2, 1MR3

Tratamiento 2: suelos contaminados
3MR1, 3MR2, 3MR3

Objetivos:

- Generar variantes de secuencia de amplicón (ASVs) a partir de los datos crudos
- Determinar la abundancia y la diversidad alfa de los microorganismos en cada muestra
- Determinar si existen diferencias en diversidad beta entre tratamientos
- Comparar la abundancia de los principales géneros bacterianos presentes en los diferentes tratamientos
- Encontrar ASVs con abundancia diferencial

Proceso:

1. Utilizando el programa dada2 se procederá a limpiar las secuencias, generar los ASVs, eliminar quimeras, empalmar secuencias, crear tabla de secuencias, asignar taxonomía a las secuencias
2. Con Phyloseq se determinará la abundancia de géneros, la diversidad alfa y la diversidad beta
3. Con ampvis2 se creará un mapa de calor
4. Con deseq2 se determinarán los ASVs con abundancia diferencial

CONSIDERAR:

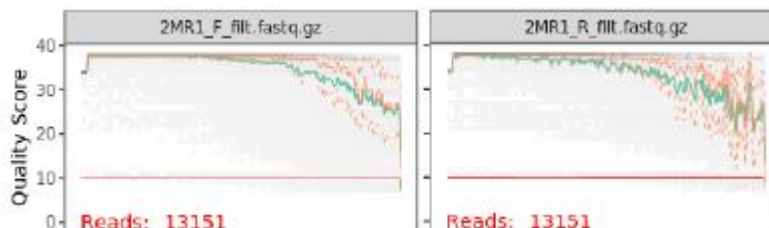
- El pipeline de dada2, complementado con el uso de phyloseq, ampvis2 y deseq2 se revisó en clase pero también existen recursos en línea que pueden ser consultados, como por ejemplo: https://benjjneb.github.io/dada2/tutorial_1_8.html
- Se recomienda el uso de una computadora con por lo menos 8Gb en RAM para el análisis de datos
- Se entregará al estudiante un script como referencia. Este script contiene los códigos necesarios para cumplir con varias tareas, incluyendo las que se plantean en este caso. El estudiante deberá generar su propio script que contenga **únicamente** los códigos utilizados para este caso.

Retos:

- Procesamiento de datos de secuenciación masiva: La cantidad de secuencias a analizar es abundante y requiere de una capacidad de cómputo elevada
- Uso de lenguaje R: Los programas propuestos para el análisis requieren que el estudiante pueda ejecutarlos con comandos de R. Se recomienda el uso de RStudio

Entregables:

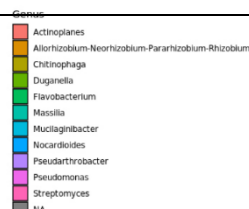
- Informe final que contenga al menos (las imágenes incluidas a continuación servirán de referencia para el estudiante):
 - Gráficos de calidad de las secuencias crudas (basta con mostrar 2 gráficos)



- Tabla que indique cuántas secuencias se fueron perdiendo durante el procesamiento en dada2 (ejemplo en la primera fila).

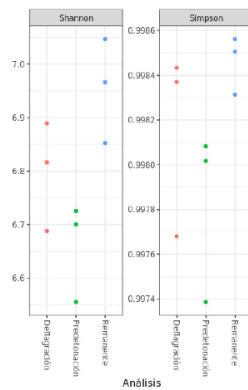
Muestra	Input	Filtered	Denoisedf	Denoisedr	Merged	Nonchim
1MR1	42630	13151	12766	12708	10025	6370
1MR2						
1MR3						
3MR1						
3MR2						
3MR3						

- Gráficos de barras que indiquen la abundancia de los principales géneros

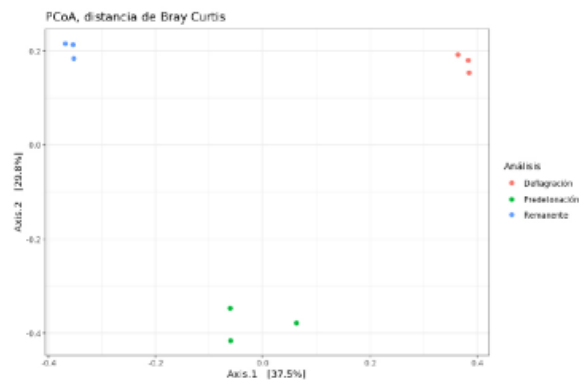




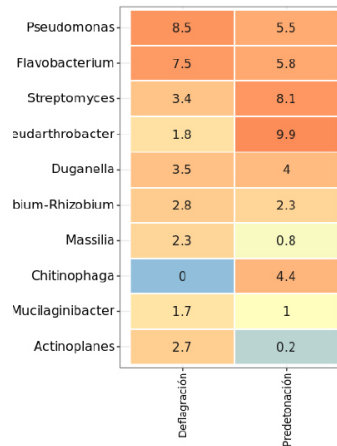
- Gráfico indicando la diversidad alfa de las muestras con las métricas de Shannon y Simpson



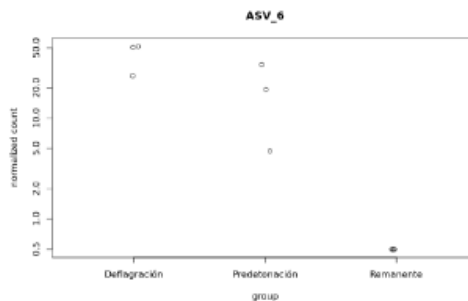
- Análisis de coordenadas principales mostrando la diversidad beta con distancia Bray Curtis



- Mapa de calor mostrando los principales géneros



- Gráficos que muestren cuales son los dos principales ASVs con abundancia diferencial



- Script con los comandos utilizados para procesar los datos, incluyendo comentarios básicos que indiquen para que sirve cada bloque de comandos (e.g. #empalme de secuencias, #eliminación de quimeras, #asignación taxonómica)

Decano Facultad de Ingeniería	Coordinador Maestría en BC	Director Tribunal Complejivo