

Análisis de la composición y funcionalidad de las comunidades microbianas asociadas al compost producido a partir de residuos verdes y residuos de alimentos con la adición de biochar

Johan Alexander Remolina Angulo

Trabajo de Grado para Optar al Título de Ingeniero Químico

Directora

Viviana Sánchez Torres

PhD en Ingeniería Química

Codirectora

Daniela Cristina Rey Romero

MSc. En Planificación y Desarrollo de Recursos Hidráulicos

Tutor

German Alexis Zafra Sierra

Dr. En Ciencias en Biotecnología

Universidad Industrial de Santander

Facultad de Ingenierías Físicoquímicas

Escuela de Ingeniería Química

Ingeniería Química

Bucaramanga

2024

### **Dedicatoria**

*Todo el esfuerzo realizado a lo largo de la carrera y en la realización de este trabajo es dedicado completamente para las dos únicas personas que me han mostrado su amor y apoyo incondicional durante toda mi vida, y que realmente conocen las situaciones que he pasado.*

*Mi mamá, que siempre se esforzó por darme todo lo mejor, mostrando un amor incondicional. No sabría donde estaría de no ser por ella. Este título se debe únicamente a todos los sacrificios que realizó para apoyarme.*

*También, para mi única hermana Nicolle, que me ve como modelo a seguir y me motiva a ser una mejor persona para seguirla ayudando.*

### **Agradecimientos**

*Quiero agradecer principalmente a la profesora Viviana Sánchez, al profesor Germán Zafra y a Daniela Rey por hacer posible la realización de este trabajo, cada uno de ellos siempre se mostró con la mejor disposición de ayudarme a resolver cualquier duda que haya tenido durante el tiempo de investigación.*

*Por otra parte, quiero agradecerle también a la familia Mora Ortíz, especialmente a Ester y Jose Pedro, por recibirme en su hogar durante más de 4 años y por el trato recibido como un miembro más de su familia.*

*Y, por último, mencionar a mis amigos de la carrera que hice a lo largo de estos cinco años y aportaron significativamente en la realización de este trabajo: A Derly, Danna, Valentina, Andrés, Camila, Diana, Diego, Alan, Katherine, Valeria y Karen.*

## Tabla de Contenido

	<b>Pág.</b>
Introducción .....	14
1. Marco conceptual .....	18
1.1 Etapas del co-compostaje .....	18
1.2 Metagenómica .....	19
1.3 Análisis de comunidades microbianas por secuenciación masiva del ADN .....	19
1.4 FastQC (Quality Control) .....	21
1.5 QIIME2 .....	21
1.6 PICRUS2 .....	21
2. Objetivos .....	22
2.1 Objetivo General .....	22
2.2 Objetivos Específicos .....	22
3. Descripción Metodológica .....	23
3.1 Revisión bibliográfica .....	23
3.2 Caracterización de las comunidades microbianas .....	24
3.2.1 Análisis de calidad en FastQC .....	24
3.2.1.1 Calidad de secuencia por base. ....	24
3.2.1.2 Gráfica de puntuación de calidad por secuencia .....	25
3.2.2 Trimming de muestras en DADA2 .....	26
3.2.2.1 Denoising de muestras en DADA2 .....	27
3.2.3 Clasificación taxonómica con la base de datos GreenGenes .....	27
3.2.4 Análisis estadístico en STAMP .....	27
3.2.5 Análisis de diversidad alfa .....	28
3.2.5.1 Índice de Chao1 .....	28
3.2.5.2 Índice de Shannon (o Entropía de Shannon) .....	28
3.3 Inferencia funcional de las comunidades microbianas .....	29
3.3.1 Predicción de rutas MetaCyc con PICRUS2 .....	29

3.3.2 Principales rutas metabólicas predichas para ambos tratamientos.....	29
3.3.3 Variación de la funcionalidad entre tratamientos .....	30
3.4 Formulación de protocolos guía para QIIME2 y PICRUS2 .....	30
4. Resultados y análisis .....	30
4.1 Análisis de calidad en FastQC .....	30
4.1.1 Calidad de secuencia por base .....	30
4.1.2 Estadísticas básicas para cada muestra .....	31
4.2 Análisis de diversidad alfa en QIIME2.....	33
4.2.1 Trimming y denoising de muestras en DADA2.....	33
4.2.2 Clasificación taxonómica con GreenGenes .....	34
4.2.3 Análisis estadístico en STAMP .....	38
4.2.4 Análisis de diversidad alfa .....	39
4.3 Inferencia funcional de las comunidades microbianas .....	40
4.3.1 Predicción de rutas MetaCyc con PICRUS2 .....	40
4.3.2 Principales rutas metabólicas predichas para ambos tratamientos.....	40
4.3.3 Variación de funcionalidad entre tratamientos .....	42
4.4 Formulación de protocolos guía para QIIME2 y PICRUS2 .....	46
5. Conclusiones .....	47
6. Recomendaciones .....	48
Referencias Bibliográficas .....	49
Apéndices.....	62

**Lista de Tablas**

	<b>Pág.</b>
Tabla 1. Estadísticas básicas y parámetros de FastQC para el conjunto de muestras .....	32
Tabla 2. Principales familias identificadas taxonómicamente en QIIME2 para cada tratamiento	35
Tabla 3. Resumen taxonomía de especies identificadas en QIIME2.....	39
Tabla 4. Familias identificadas en QIIME2 para cada tratamiento relacionadas con actividades de descomposición y fijación de nutrientes. ....	46

**Lista de Figuras**

	<b>Pág.</b>
Figura 1. Tecnología de secuenciación de nueva generación Illumina .....	20
Figura 2. Diagrama de flujo con las etapas metodológicas de la investigación .....	23
Figura 3. Gráfico de calidad de secuencia por base en FastQC.....	25
Figura 4. Gráfico interactivo para el conjunto de secuencias en qiime2view. ....	26
Figura 5. Gráfico de calidad promedio del conjunto de muestras en qiime2view.....	31
Figura 6. Estimado de longitud de cortes para las muestras resultantes en qiime2view. ....	34
Figura 7. Gráfico de barras apiladas familias identificadas en QIIME2 para cada tratamiento ...	36
Figura 8. Gráfico de barras de error en STAMP para familias identificadas con poca diferencia entre proporciones ( $p>0,05$ ).....	38
Figura 9. Gráfico de barras principales rutas metabólicas en STAMP.....	40
Figura 10. Gráfico de barras con perfiles en STAMP para rutas metabólicas con valores de $p=0$ relacionadas con A) descomposición de compuestos orgánicos, B) fijación de Nitrógeno, y C) fertilización del compost.....	43
Figura 11. Diagrama de flujo resumen contenido del protocolo para QIIME2 y PICRUST2 .....	47

## Glosario

**Adaptador:** secuencia de nucleótidos pequeña que se une a los extremos de ADN para su identificación en las plataformas de secuenciación.

**Artefacto:** archivos producidos por QIIME2 (en formato qza) que contienen datos específicos del microbioma. No son visualizables.

**Amplión:** Fragmento de ADN amplificado a través de técnicas de laboratorio como el PCR.

**Árbol filogenético:** herramienta gráfica que muestra las relaciones y la historia evolutiva de un grupo de diferentes organismos a partir de un único ancestro en común o relaciones entre especies.

**ARNr 16S (Subunidad 16S del ARN ribosomal):** componente del ribosoma que se encuentra en las células procariotas. El gen que lo codifica permite la identificación y caracterización taxonómica de las bacterias (del Rosario Rodicio & del Carmen Mendoza, 2004).

**Biochar:** carbón vegetal que se obtiene a partir de biomasa pirolizada a temperaturas entre 400 y 700 °C. Su adición a los suelos mejora la fertilidad y aumenta capacidad de almacenamiento de carbono atmosférico (González-Marquetti et al., 2020).

**Bioinformática:** ciencia interdisciplinaria que se encarga de la obtención, almacenamiento, manejo y análisis computacional de todas las clases de información biológica (Franco et al., 2008).

**Cebadores (primers):** secuencia corta de ADN o ARN sintetizada artificialmente para iniciar la amplificación de una región específica del ADN por medio de PCR. Se diseñan para que solamente se unan a los extremos del ADN de interés e inicien la síntesis de nuevas hebras complementarias (Innis et al., 2012).

**Co-compostaje:** compostaje de sustratos de diferentes fuentes.

**Denoising (eliminación de ruido):** eliminación de datos de lecturas de secuencias repetidas e identificación de secuencias representativas (ASVs) con valores de frecuencia para optimizar tamaños de archivos y costos computacionales.

**Diversidad alfa:** resumen cuantitativo de la estructura de una comunidad microbiana con respecto a su riqueza (cantidad de grupos taxonómicos), uniformidad (distribución y abundancia de los grupos) o ambos (Willis et al., 2019).

**Diversidad beta:** cuantifica la similitud o distancia entre conjuntos de comunidades microbianas. Utiliza principalmente enfoques estadísticos o geométricos.

**Familia:** quinto nivel taxonómico ubicado entre el orden y el género. Se caracteriza por un nivel intermedio de especificidad que permite agrupar organismos con características similares.

**Filo:** segundo nivel taxonómico ubicado entre el reino y la clase. Utilizado para subdivisiones en los reinos animal y protista.

**Forward:** dirección de lectura de la secuencia de ADN desde el extremo 5' a 3'.

**Funciones metabólicas:** comprende todas las actividades bioquímicas y transformaciones que lleva a cabo la comunidad microbiana. Estas actividades son esenciales para la funcionalidad de la comunidad en sí y del ecosistema que habite (Rowland et al., 2018)

**Glucólisis:** ruta metabólica de degradación de azúcares simples.

**Ilumina:** tecnología de secuenciación de nueva generación (NGS). Se utiliza para determinar el orden de los nucleótidos en genomas completos o regiones específicas de ADN o ARN.

**Inferencia funcional:** métodos de predicción de las posibles funciones biológicas que se espera desempeñe la comunidad microbiana basado en la presencia o ausencia de genes relacionados taxonómicamente que desarrollen actividades metabólicas.

**Mesófila (o mesofílica):** fase inicial del proceso de compostaje donde predomina la actividad de microorganismos mesófilos que descomponen los compuestos orgánicos más simples, también se eleva rápidamente la temperatura y disminuye el pH (Bueno et al., 2008).

**Metagenómica:** comprende el estudio de las secuencias de ADN de toda una comunidad microbiana de muestras ambientales, vegetales, animales y humanas por medio de técnicas modernas independientes del cultivo. (Ghazanfar et al., 2010).

**Microbiota:** comunidad microbiana que habita un ecosistema.

**PCR (reacción en cadena de la polimerasa):** técnica de la biología molecular que permite amplificar millones de copias de un segmento específico de ADN a partir de una única muestra.

**Reverse:** dirección complementaria a la lectura forward. Se lee desde el extremo 3' de una hebra hasta el extremo 5'. Ambas cadenas complementarias se unen conformando la estructura de doble hélice del ADN.

**Secuencia:** es la cuantificación lineal de la cadena de nucleótidos (A, C, G y T) que conforman el ADN.

**Secuencia representativa:** secuencia seleccionada de un conjunto de secuencias similares para representar la totalidad de un grupo. Es un proceso útil para optimizar el tamaño de los datos y facilitar los análisis.

**Taxonomía:** ciencia de la clasificación que se aplica en la biología para la ordenación sistemática y jerarquizada de los grupos de organismos.

**Termófila (o termofílica):** fase que sucede a la mesófila en el proceso de compostaje. Proliferan los microorganismos termotolerantes y termófilos, y la microbiota mesófila es inhibida por la temperatura, que aumenta lentamente hasta valores entre 65 y 70°C debido al aumento en la actividad microbiana que descompone material orgánico más complejo (Bueno et al., 2008).

**Trimming:** recorte de secuencias. Consiste en la eliminación de regiones de baja calidad de las lecturas de secuencias con el fin de mejorar la calidad general de los datos para realizar análisis posteriores.

**Unidades taxonómicas operacionales (OTUs):** definición utilizada para seleccionar y clasificar grupos de bacterias estrechamente relacionadas, generalmente basada en datos moleculares como secuencias de ADN o ARN (Pérez, 2016).

**Variantes de secuencias de amplicón (ASVs):** método de agrupación de lecturas de secuenciación por medio de herramientas bioinformáticas que identifican secuencias similares y las agrupa en unidades individuales, asumiendo que todas las secuencias de ADN están libres de errores. Muestran una mayor sensibilidad y especificidad que los métodos que utilizan OTUs (Callahan et al., 2017).

## Resumen

**Título:** Análisis de la composición y funcionalidad de las comunidades microbianas asociadas al compost producido a partir de residuos verdes y residuos de alimentos con la adición de biochar\*

**Autor:** Johan Alexander Remolina Angulo\*\*

**Palabras Clave:** Comunidades microbianas, secuenciación, metagenómica, bioinformática, compost, biochar.

**Descripción:** En este estudio se analizó la estructura y funcionalidad de las comunidades microbianas presentes en el compost, a partir de los datos de secuenciación masiva obtenidos con el sistema Illumina MiSeq junto con herramientas bioinformáticas. Se realizó un análisis de composición y una predicción de perfiles funcionales de las comunidades microbianas asociadas al compost producido a partir de residuos verdes y residuos de alimentos con la adición de biochar en las fases mesófila y termófila. Se obtuvieron lecturas de secuencias de amplicones del gen 16S RNA y se realizó una clasificación taxonómica en QIIME2 a nivel de familia; seguido de un análisis estadístico y comparación de índices de diversidad alfa. Para la predicción de perfiles funcionales se utilizó PICRUSt2 para mostrar las principales rutas metabólicas asociadas al proceso de compostaje e identificar rutas de posible interés que presenten diferencias de abundancia en cada tratamiento. Se observó que, el compost tratado con biochar en la fase termófila presenta una mayor diversidad a nivel de familia, identificando un total de 126 para este tratamiento y 96 en el tratamiento en la fase mesófila. Las familias más dominantes para el compostaje tratado en la fase mesófila fueron *Streptomyetaceae*, *Streptosporangiaceae*, *Thermoactinomycetaceae*, *Xanthomonadaceae* y *Paenibacillaceae*, mientras que para el otro tratamiento dominaron familias como *Streptomyetaceae*, *Paenibacillaceae*, *Rhizobiaceae* y otras sin clasificar. Además, el tratamiento en la fase termófila presenta valores mayores en los índices de Chao1 y Shannon (497 y 8) con respecto al tratamiento en la fase mesófila (341 y 7). Funcionalmente, el tratamiento en la fase mesófila obtuvo mayores abundancias en 16 de 18 rutas metabólicas identificadas y asociadas a procesos de descomposición de compuestos orgánicos, fijación de Nitrógeno y fertilización del compost, en donde se destacan principalmente las actividades de bacterias que pertenecen a las familias *Pseudomonadaceae* y *Enterobacteriaceae*.

---

\* Trabajo de Grado

\*\* Facultad de Ingenierías Físicoquímicas. Escuela de Ingeniería Química. Directora: Viviana Sánchez Torres. PhD en Ingeniería Química. Codirectora: Daniela Cristina Rey Romero MSc. En Planificación y Desarrollo de Recursos Hidráulicos. Tutor: German Alexis Zafra Sierra Dr. En Ciencias en Biotecnología

### Abstract

**Title:** Analysis of the composition and functionality of microbial communities associated with compost produced from green waste and food waste with the addition of biochar\*

**Author:** Johan Alexander Remolina Angulo\*\*

**Key Words:** Microbial communities, sequencing, metagenomics, bioinformatics, compost, biochar.

**Description:** In this study, the structure and functionality of microbial communities present in compost were analyzed using massive sequencing data obtained with the Illumina MiSeq system and bioinformatics tools. A composition analysis and a prediction of functional profiles of microbial communities associated with compost produced from green waste and food waste with the addition of biochar were performed in the mesophilic phase and the thermophilic phase. Amplicon sequence reads of the 16S rRNA gene were obtained and a taxonomic classification was performed in QIIME2 at the family level, followed by statistical analysis and comparison of alpha diversity indices. For the prediction of functional profiles, PICRUSt2 was used to show the main metabolic pathways associated with the composting process and identify potential pathways of interest that present differences in abundance in each treatment. It was observed that the compost treated with biochar in the thermophilic phase presents a higher diversity at the family level, identifying a total of 126 for this treatment and 96 in the mesophilic phase treatment. The most dominant families for compost with biochar added in the mesophilic phase were *Streptomyetaceae*, *Streptosporangiaceae*, *Thermoactinomycetaceae*, *Xanthomonadaceae*, and *Paenibacillaceae*, while for the other treatment, families such as *Streptomyetaceae*, *Paenibacillaceae*, *Rhizobiaceae*, and others unclassified dominated. Additionally, the thermophilic phase treatment presents higher values in the Chao1 and Shannon indices (497 and 8) compared to the mesophilic phase treatment (341 and 7). Functionally, the mesophilic phase treatment obtained higher abundances in 16 out of 18 identified metabolic pathways associated with organic compound decomposition processes, nitrogen fixation, and compost fertilization, where activities of bacteria belonging to the families *Pseudomonadaceae* and *Enterobacteriaceae* stand out.

---

\* Degree Work

\*\*Faculty of Physicochemical Engineering. Chemical Engineering School. Director: Viviana Sánchez Torres. PhD in Chemical Engineering. Co-director: Daniela Cristina Rey Romero. MSc in Water Resources Planning and Development. Tutor: German Alexis Zafra Sierra. Doctor of Science in Biotechnology

## Introducción

Los métodos tradicionales de cultivo para el análisis de comunidades microbianas tienen una utilidad limitada e inadecuada en el campo de la microbiología actual (Ritz, 2007), en donde sólo un porcentaje comprendido entre el 0.1 y el 10% de las bacterias presentes en el ambiente pueden ser cultivadas debido al desconocimiento de los requerimientos nutricionales y condiciones fisicoquímicas necesarias para el desarrollo de microorganismos desconocidos en un ambiente controlado (Lozada et al., 2004).

La diversidad de microorganismos que participan en el compostaje constituye un ecosistema que se va haciendo más complejo a medida que se va degradando la materia orgánica. Entre más diversa sea la comunidad microbiana, mayor es su grado funcional, que es la capacidad de utilizar una gran variedad de sustratos para llevar a cabo una amplia gama de procesos (Nkongolo & Narendrula-Kotha, 2020). La adición de biochar durante el compostaje puede mejorar la eficiencia del proceso, favoreciendo las condiciones fisicoquímicas para el crecimiento y actividad enzimática de los microorganismos, alterando la dinámica, composición y funciones de la comunidad microbiana (Hernández et al., 2020).

Una investigación realizada por Galezo y Díaz (2024) en el Grupo de Investigación en Recursos Hídricos y Saneamiento Ambiental (GPH) de la Universidad Industrial de Santander (UIS) (2024), analizó la adición de 2% de biochar en el proceso de co-compostaje de residuos verdes, alimentos procesados y no procesados en la fase mesófila (T1 y RT1, agregado en el día 0) y termófila (T2 y RT2, agregado en el día 5) mediante la medición y comparación de parámetros fisicoquímicos como pH, conductividad eléctrica, temperatura, humedad, sólidos volátiles, índice de germinación y relación C/N. Se concluyó que no se obtuvieron diferencias significativas en los

parámetros fisicoquímicos entre los dos tratamientos; sin embargo, se obtuvieron temperaturas máximas mayores (5 °C más) y una mayor duración de la fase termófila (4 días) en el compostaje con biochar añadido en la fase mesófila; mientras la duración de la fase de enfriamiento fue menor en el tratamiento con biochar en la fase mesófila. Las autoras infieren que el aumento de la temperatura en el compost tratado en la fase mesófila se pudo dar debido a un mayor crecimiento y actividad de los microorganismos en la descomposición del material orgánico, pero no se realiza un estudio de las comunidades microbianas que se encuentran en el compost producido.

Una alta diversidad microbiana en el compost es un indicador de su calidad y aumenta las interacciones benéficas que pueden ocurrir entre la microbiota y las plantas; por lo que el monitoreo a través de las tecnologías de secuenciación permite una mejor comprensión de su estructura y establecer relaciones con sus propiedades fisicoquímicas (Bravo, 2021). La diversidad de la comunidad microbiana que participa en el compostaje depende de la materia prima y sus propiedades, donde destacan principalmente la relación Carbono-Nitrógeno y la humedad, que puede afectar de manera positiva el crecimiento de especies de microorganismos más específicos que varían dependiendo del tipo de compostaje (Wang et al., 2015).

Las investigaciones realizadas para el análisis de diversidad microbiana en el compost se enfocan principalmente en la comprensión de cómo se ve afectada la comunidad microbiana durante las etapas dinámicas del compostaje y de la variación de la estabilidad del compost final en función en la diversidad microbiana. También se destacan aplicaciones como la optimización en el proceso de compostaje mediante la incorporación de especies exógenas (Castro & Chanamé, 2021) y la observación de microorganismos eficientes que mejoren el rendimiento en otras actividades de tipo agrícola (Tanya Morocho et al., 2019).

Silva (2016) realizó una comparación entre dos muestras de compost comercial de residuos sólidos con adición de material orgánico que se procesaron durante un mismo tiempo (6 meses). A pesar de que el compost final de ambas muestras se obtuvo con la misma materia prima, los parámetros de propiedades fisicoquímicas (relación C/N, pH, nivel de aireación y humedad) que determinaban la estabilidad del producto final variaban considerablemente, por lo que se realizó una caracterización microbiológica para inferir posibles relaciones entre estas propiedades y la estructura de sus comunidades bacterianas. Se realizó un aislamiento, amplificación por PCR y análisis del ADN de las secuencias de cada muestra de compost en la herramienta bioinformática *Bionumerics*, y se determinó que la muestra que posee una mayor diversidad microbiana muestra una mayor estabilidad debido a una mayor presencia de las familias *Streptomycetaceae*, *Pseudonocardaceae*, *Bacillaceae* y *Brevibacteriaceae* que promueven la producción de sustancias húmicas.

Otro estudio realizado por Storey (2015) analizó la dinámica de las comunidades microbianas del compost de residuos verdes triturados y granos de cerveceras, en donde se evalúan los efectos de la adición de desechos de madera tamizada como sustituto del nitrato amónico cálcico como fuente de nitrógeno, mediante la extracción del ADN, amplificación por PCR y análisis de datos por el software de bioinformática *Mothur* de las secuencias de compost. Se mostró que los desechos de madera pueden ser una opción viable como sustituto, la calidad del producto final no muestra variación en cuanto a sus variables fisicoquímicas y promueve cambios positivos en la diversidad bacteriana, como la presencia de filos como *Spirochaetes* y *Synergistetes* que promueven el ciclo de nutrientes, y de clases de microorganismos no cultivables como *Saccharibacteria*.

El estudio e investigación de las comunidades microbianas que conforman el compost a partir de técnicas de secuenciación independientes del cultivo que utilizan el principio de la metagenómica, junto con herramientas bioinformáticas y bases de datos que permiten el análisis del material genético han revolucionado la microbiología, reduciendo costos y tiempos de análisis (Hernández et al., 2020), y permitiendo una mayor comprensión de la microbiota presente en el compost y prediciendo las funciones que cumple cada grupo bacteriano.

El presente trabajo cuenta con los archivos de las secuencias del ADN de las muestras de compost producido en la investigación realizada por Galezo y Díaz (2024), en donde se comparó el efecto de la adición del biochar en diferentes fases del proceso de co-compostaje de residuos verdes y de alimentos, a través de la medición de variables fisicoquímicas. Estas muestras fueron secuenciadas en la plataforma Illumina (2024) en una colaboración con el Kyushu Institute of Technology en enero del 2024, las cuales no habían sido procesadas con herramientas bioinformáticas.

Con base a lo anterior, se plantea realizar un análisis de diversidad microbiana e inferencia funcional de las comunidades microbianas, a las lecturas de secuencias de ADN del compost producido en la investigación realizada por Galezo y Díaz (2024), utilizando las herramientas bioinformáticas QIIME2 y PICRUST2 con el fin de responder la siguiente pregunta: ¿Cómo afecta la adición del biochar en las fases mesófila y termófila a la composición de la microbiota y funcionalidad del compost producido por residuos verdes y residuos de alimentos?

## 1. Marco conceptual

### 1.1 Etapas del co-compostaje

El co-compostaje es la descomposición aeróbica de materia orgánica proveniente de diferentes fuentes, como residuos vegetales, de alimentos o heces animales o humanas. Se desarrolla en tres fases en relación con la temperatura: fase mesófila inicial, fase termófila y fase de enfriamiento y maduración (Bueno et al., 2008).

- **Fase mesófila inicial:** Es la fase más dinámica del proceso de compostaje, en donde ocurre un rápido incremento de la temperatura (de 10 a 40 °C) el pH disminuye y se degradan los compuestos orgánicos más simples. La actividad de microorganismos mesófilos y algunos termotolerantes que degradan sustancias carbonadas solubles como azúcares y aminoácidos, ocasionan el aumento en la temperatura y la disminución del pH ocurre como consecuencia de la producción de ácidos orgánicos. Esta fase tiene una duración aproximada entre 2 y 4 días (Velázquez Martí, 2024).

- **Fase termófila:** En esta fase los microorganismos mesófilos de la fase inicial se inhiben por el aumento de la temperatura, y ocurre una transición en la que domina la microbiota termófila (40-60 °C); la temperatura sigue aumentando hasta en un rango máximo entre 65 y 70 °C que se mantiene por varios días y provoca la eliminación de agentes patógenos y una reducción de la actividad microbiana (Laich, 2011).

- **Fase de enfriamiento y maduración:** En esta última fase ocurre un descenso en la temperatura debido a la disminución en la fuente de carbono, que disminuye la actividad microbiana. La temperatura disminuye a valores cercanos a los del ambiente y se produce una recolonización de microorganismos mesófilos y aparecen hongos filamentosos capaces de

degradar compuestos orgánicos complejos como la lignina y la hemicelulosa (Bohórquez Santana, 2019)

## **1.2 Metagenómica**

Comprende el estudio de las secuencias de ADN de toda una comunidad microbiana de muestras ambientales, vegetales, animales y humanas por medio de técnicas modernas independientes del cultivo. Esta metodología se ha desarrollado como una herramienta importante para el descubrimiento de nuevas bacterias y funciones microbianas (Ghazanfar et al., 2010). La predicción de posibles vías metabólicas por medio de comparaciones con genes de bases de datos es una de las principales ventajas con respecto a los métodos de cultivo ya que ofrecen una mayor comprensión sobre la importancia de estas comunidades en los ciclos biogeoquímicos globales (Turnbaugh et al., 2009).

## **1.3 Análisis de comunidades microbianas por secuenciación masiva del ADN**

La secuenciación del ADN se realiza con el fin de conocer la secuencia precisa de nucleótidos de los organismos presentes. Un experimento de secuenciación de nueva generación (NGS) para el análisis de comunidades microbianas se puede resumir en las siguientes etapas: preparación y amplificación del ADN, secuenciación y análisis bioinformático (Mordoh, 2019).

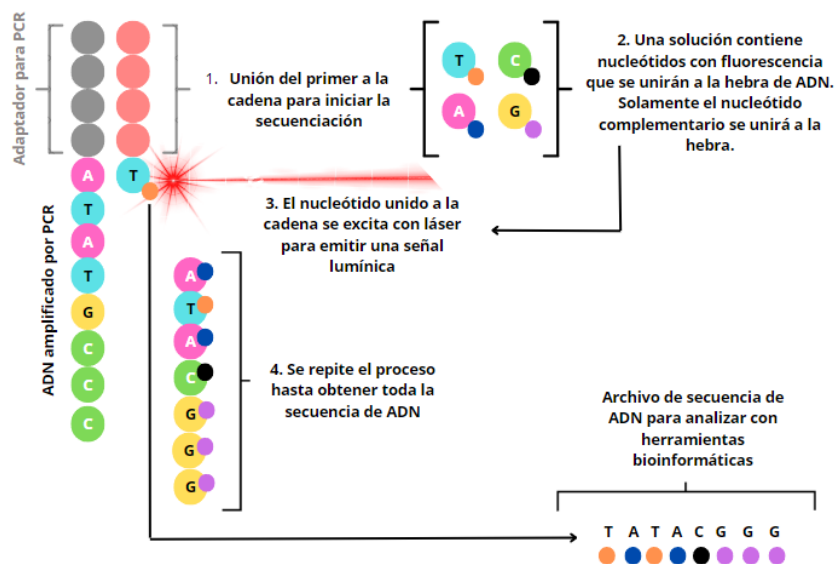
La preparación consiste en la extracción del ADN ambiental y la amplificación por PCR del gen del ARN ribosomal 16S (ARNr 16S), región objetivo para estudios de comunidades bacterianas debido a que contiene información que permite la identificación de bacterias hasta niveles de especie, y también tiene una alta capacidad de conservación, que permite el estudio y diseño de cebadores específicos (*primers*) (Cai et al., 2003).

La plataforma de secuenciación Illumina es la más utilizada para llevar a cabo la secuenciación. Se introducen los productos de PCR en una placa con una solución enriquecida con

un solo tipo de nucleótido (A, C, G o T) modificados con etiquetas fluorescentes que emiten una señal lumínica cuando se incorporan a la cadena de ADN, especificando su identificación y ubicación en la hebra. Esto se repite hasta completar la secuenciación de la cadena (Rubio et al., 2020). La figura 1 ilustra el proceso de identificación de nucleótidos llevado a cabo por la plataforma Illumina MiSeq:

### Figura 1

#### *Tecnología de secuenciación de nueva generación Illumina*



*Nota.* Tomado y adaptado de Kchouk et al. (2017)

Una vez terminada la secuenciación se obtiene un archivo de datos que se puede analizar de forma primaria (control de calidad), secundaria (comparación con bases de datos) o terciaria (generación de datos a partir del análisis secundario) con apoyo de herramientas bioinformáticas como QIIME2 y PICRUST2.

#### **1.4 FastQC (Quality Control)**

FastQC (2015) es una herramienta para el control de calidad de datos de secuencias crudas (*raw data*). Proporciona un conjunto de elementos gráficos para la elaboración de reportes y otorga información para poder determinar si los datos ingresados presentan problemas antes de realizar otros análisis adicionales. Los informes de análisis generados en FastQC se han convertido en el estándar y la base para la creación de otras herramientas automáticas de control de calidad (de Sena Brandine & Smith, 2021).

#### **1.5 QIIME2 (Quantitative Insights Into Microbial Ecology – Comprensión Cualitativa de la Ecología Microbiana)**

QIIME2 (Bolyen et al., 2019) es un conjunto de herramientas de libre acceso y código abierto que se pueden utilizar para el análisis de secuencias de comunidades microbianas. Ofrece herramientas para el filtrado de secuencias, análisis de calidad y diversidad, asignación taxonómica y reconstrucción filogenética (Brenes-Guillén, 2018), también proporciona elementos de visualización que facilitan las comparaciones entre muestras y la generación de informes de resultados.

#### **1.6 PICRUSt2 (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States – Investigación filogenética de comunidades por reconstrucción de estados no observados)**

PICRUSt2 (Douglas et al., 2020) es una herramienta bioinformática que predice funciones del metagenoma y genera datos de abundancia relativa de genes funcionales a partir de secuencias de genes de ARNr 16S con la base de datos KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa et al., 2023) que proporcionan información sobre reacciones enzimáticas y vías metabólicas al ingresar únicamente datos de variantes de secuencias de amplicones (ASVs).

## **2. Objetivos**

### **2.1 Objetivo General**

Analizar la composición y los perfiles funcionales de las comunidades microbianas asociadas a compost producido a partir de residuos verdes y residuos de alimentos con la adición de biochar.

### **2.2 Objetivos Específicos**

Caracterizar las comunidades microbianas del compost con adición de biochar en términos de composición taxonómica y diversidad microbiana, empleando la herramienta bioinformática QIIME2.

Inferir las funciones metabólicas potenciales de las comunidades microbianas del compost con adición de biochar a través de la herramienta PICRUST2.

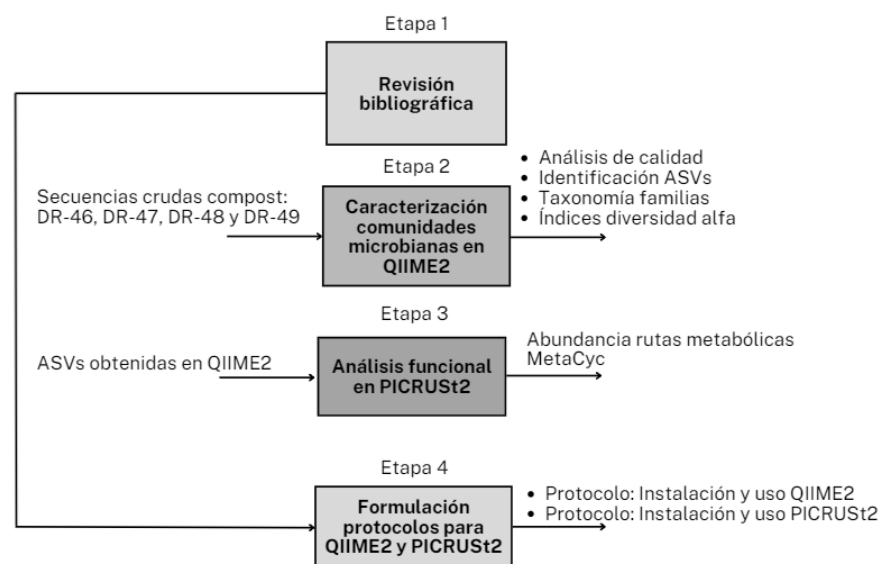
Formular protocolos para el manejo de las herramientas bioinformáticas QIIME2 y PICRUST2, para el estudio de comunidades microbianas por secuenciación masiva.

### 3. Descripción Metodológica

A continuación, en la figura 2 se muestra un diagrama de flujo que resume cada una de las etapas realizadas en el presente trabajo de investigación:

#### Figura 2

*Diagrama de flujo con las etapas metodológicas de la investigación*



#### 3.1 Revisión bibliográfica

Se llevó a cabo una revisión bibliográfica durante todo el tiempo de realización de la investigación, enfocada en los siguientes temas: Análisis de comunidades microbianas, secuenciación masiva, inferencia funcional y estudios metagenómicos en la microbiota del compost. La revisión se hizo mediante la búsqueda de artículos científicos y trabajos de investigación en motores de búsqueda como *Google Scholar*, *ScienceDirect*, *Web of Science* y la Base de datos de la Biblioteca UIS.

## 3.2 Caracterización de las comunidades microbianas

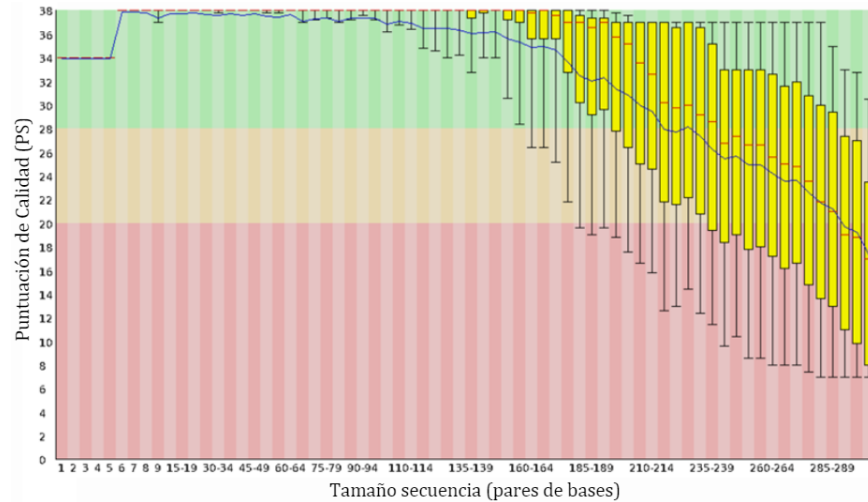
### 3.2.1 *Análisis de calidad en FastQC*

Se realizó un análisis de calidad para las cuatro secuencias de ADN obtenidas a partir de los productos finales del compostaje de residuos verdes y residuos de alimentos con adición de biochar en las fases mesófila (muestras DR-46 y DR-47) y termófila (muestras DR-48 y DR-49) realizado por Galezo y Díaz (2024). Se recopiló la información de estadísticas básicas y se resumieron en una tabla para observar secuencias de baja calidad que puedan afectar el análisis posterior. El principal parámetro para realizar el control es el gráfico de calidad por base:

**3.2.1.1 Calidad de secuencia por base.** Proporciona información sobre la distribución de las puntuaciones de calidad a lo largo de la secuencia con el apoyo de diagramas de caja y bigotes, permitiendo identificar posibles regiones de baja calidad que pudieron haber sido afectadas por errores de lectura o la presencia de adaptadores. En el eje vertical de la gráfica se muestra el indicador de calidad (*Phred Score* – PS o Q), que a su vez se divide en tres regiones de distintos colores: la zona superior verde posee valores de calidad mayores a 30 indican una alta calidad; la zona intermedia en amarillo comprende valores entre 20 y 30 de calidad media, y la zona roja con valores menores a 20 denotan una baja calidad (datos no confiables). Las cajas amarillas que se muestran a lo largo de la gráfica representan el rango intercuartil (25-75%) de los puntajes de calidad de las secuencias, y los bigotes superiores e inferiores representan el 10 y el 90% de estos puntajes. Adicionalmente otorga información sobre la calidad promedio a lo largo de la secuencia (línea azul) y media de calidad de la secuencia leída (líneas rojas) (Malhotra et al., 2021). A continuación, en la figura 3 se muestra un ejemplo de un gráfico de calidad por base obtenido en FastQC:

**Figura 3**

*Gráfico de calidad de secuencia por base en FastQC*



**3.2.1.2 Gráfica de puntuación de calidad por secuencia.** Otorga información sobre la calidad promedio de la secuencia, reportando un valor de calidad (PS) relacionado con la precisión de lectura. El Puntaje Phred es una escala logarítmica que posee una relación inversa con el porcentaje de error (O’Rawe et al., 2015).

$$PS = -10\log_{10}(\%Error) \quad (Ec. 1)$$

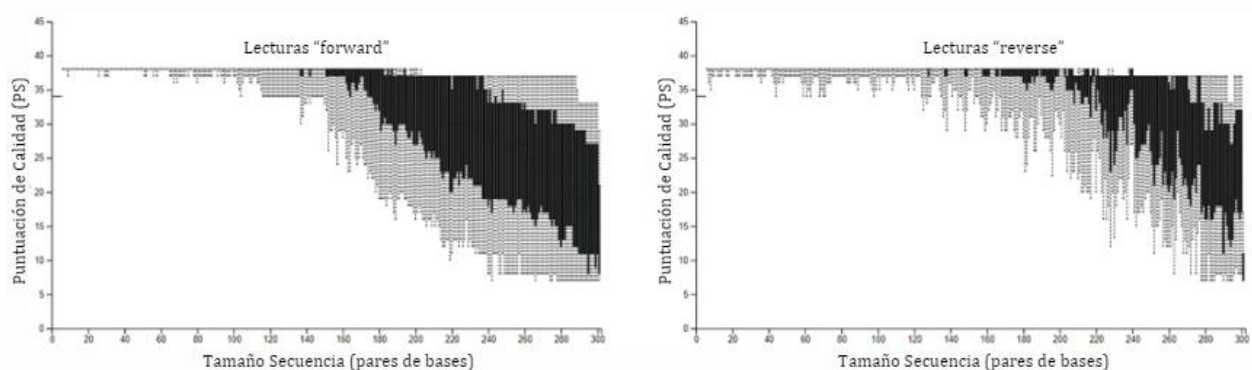
Determinar un puntaje mínimo aceptable de calidad puede variar dependiendo del tamaño de las muestras y la precisión del estudio a realizar. Secuencias de lecturas cortas que contienen valores de PS menor a 30 se consideran poco confiables (1 error cada 1000 lecturas) (Liao et al., 2017); sin embargo, para lecturas de gran tamaño como secuencias de muestras de compostaje se puede perder una gran cantidad de información al aplicar este criterio, por lo que se suele utilizar un umbral de PS menor a 20 (1 error cada 100 bases). Utilizar secuencias con PS menores a 20 puede afectar negativamente el análisis metagenómico (MacManes, 2014).

### 3.2.2 *Trimming de muestras en DADA2*

Las muestras que pasaron el análisis de calidad se ingresaron como un único archivo conjunto en *qiime2view*, una interfaz web que permite la visualización de artefactos y visualizaciones generadas en QIIME2 (formato qza y qzv respectivamente), y ofrece una gráfica interactiva que resume la calidad de todo el conjunto de muestras, como la que se muestra en la figura 3. Al archivo con las muestras se les realizó *trimming* con el software complementario DADA2 (Callahan et al., 2016) con el objetivo de aumentar la calidad de las lecturas, eliminando las regiones de baja calidad. DADA2 es la herramienta por defecto que viene incluida con QIIME2 para el recorte de lecturas de secuencias. Para determinar si los valores de corte utilizados fueron los adecuados, se observó el número de secuencias que se conservaron después de realizar el *trimming*; con el fin de obtener un valor óptimo de secuencias de tamaños similares para realizar una clasificación taxonómica.

#### Figura 4

Gráfico interactivo para el conjunto de secuencias en *qiime2view*.



**3.2.2.1 Denoising de muestras en DADA2.** De manera simultánea al proceso anterior, se realizó un *denoising* o eliminación de ruido al archivo de muestras. El *denoising* elimina las secuencias repetidas y agrupa las secuencias similares en grupos representativos (ASVs) junto a un valor de abundancia y error asociados (Callahan et al., 2016) para que posteriormente sean identificadas.

### **3.2.3 Clasificación taxonómica con la base de datos GreenGenes**

El archivo de ASVs se clasificó taxonómicamente con la base de datos GreenGenes (DeSantis et al., 2006) en QIIME2. GreenGenes contiene datos de 408.135 secuencias de genes de la región ARN 16S de bacterias y arqueas que permiten su identificación en cada uno de los siete niveles taxonómicos (McDonald et al., 2012). Las ASVs ingresadas son alineadas con las secuencias de la base de datos dependiendo de su nivel de proximidad, si una ASV no es idéntica a una secuencia de referencia, se asocia al dato de referencia que más se asemeje (Can, 2019).

### **3.2.4 Análisis estadístico en STAMP**

Para determinar si existen diferencias significativas en las abundancias de las familias de cada muestra, se utilizó la prueba exacta de Fisher bilateral utilizando la herramienta estadística STAMP (*Statistical Analysis of Metagenomic Profiles*) (Parks et al., 2014). STAMP es una herramienta estadística diseñada específicamente para archivos de perfiles metagenómicos, los cuales incluyen perfiles de taxonomía como los generados en QIIME2. Se utilizó la prueba exacta de Fisher bilateral (Sprent, 2011), que asume inicialmente una hipótesis nula como verdadera; en este caso, que no existen diferencias entre las abundancias de las familias de cada conjunto de muestras.

### 3.2.5 *Análisis de diversidad alfa*

Este análisis otorga información sobre la cantidad de especies diferentes (riqueza) y de cómo están distribuidos (uniformidad) (Willis, 2019). Se calcularon los siguientes índices en QIIME2 ingresando el archivo de taxonomía generado anteriormente, y se realizó una comparación entre el grupo de muestras a partir de los índices obtenidos.

**3.2.5.1 Índice de Chao1.** Estima la diversidad potencial de una comunidad microbiana, teniendo en cuenta la información de especies menos abundantes (Cassol et al., 2023). Se basa en el principio de que las especies más raras pueden otorgar más información sobre el número de especies potenciales que puede haber en una comunidad microbiana y se calcula utilizando las secuencias que hayan sido detectadas una vez (*singletons* – F1) y dos veces (*doubletons* – F2) (B.-R. Kim et al., 2017).

$$S_{chao1} = F_{obs} + \frac{F_1(F_1 - 1)}{2(F_2 + 1)} \quad (Ec. 3)$$

En donde  $S_{chao1}$  corresponde al Índice de Chao1 y  $F_{obs}$  es la cantidad de especies observadas en cada muestra.

**3.2.5.2 Índice de Shannon (o Entropía de Shannon).** Tiene en cuenta la riqueza y uniformidad de una comunidad microbiana. Mide el grado de incertidumbre al predecir a qué especie pertenecerá un individuo elegido al azar de una muestra que contiene una cantidad determinada de especies (Lemos et al., 2011). Entre más diversa y uniforme sea una comunidad microbiana, será más difícil predecir a qué especie puede pertenecer un individuo al azar de la muestra.

$$H = - \sum_{i=1}^s (p_i \ln p_i) \quad (Ec. 4)$$

Donde  $H$  es el Índice de Shannon,  $s$  es la cantidad de ASVs y  $p_i$  es la abundancia relativa de la ASV en la muestra.

### **3.3 Inferencia funcional de las comunidades microbianas**

#### **3.3.1 Predicción de rutas MetaCyc con PICRUSt2**

Se utilizó PICRUSt2 (Douglas et al., 2020) para llevar a cabo la predicción de rutas metabólicas a partir de los datos de ASVs generados con QIIME2. PICRUSt2 utiliza un único comando que genera abundancias de vías MetaCyc (Caspi et al., 2016) a través de un proceso de asignación filogenética utilizando el algoritmo SEPP (*SATé-Enabled Phylogenetic Placement*) (Mirarab et al., 2012) en donde las ASVs se organizan en un árbol filogenético de referencia para identificar genes y determinar abundancias. Luego, se predicen las posibles reacciones enzimáticas (números EC) y funciones comunitarias (números KO) que cada comunidad microbiana puede realizar, basándose en los perfiles genómicos de los parientes más cercanos en el árbol filogenético; esto se lleva a cabo con la base de datos KEGG (*Kyoto Encyclopedia of Genes and Genomes*) (Kanehisa et al., 2023) que se encuentra incluida en PICRUSt2. Por último, para determinar las rutas metabólicas se utiliza el algoritmo de inferencia MinPath (Ye & Doak, 2009) que calcula e identifica el mínimo de rutas metabólicas posibles que realicen todas las reacciones enzimáticas posibles encontradas en los EC.

#### **3.3.2 Principales rutas metabólicas predichas para ambos tratamientos**

Las rutas metabólicas generadas para el conjunto de muestras se identificaron utilizando la base de datos MetaCyc (Caspi et al., 2016), y fueron observadas en STAMP (Parks et al., 2014) con el objetivo de mostrar las rutas más abundantes y describir las que estén asociadas a procesos de descomposición de compuestos orgánicos, fijación de carbono y nitrógeno.

### **3.3.3 Variación de la funcionalidad entre tratamientos**

Se observaron en STAMP las rutas metabólicas que presenten estadísticamente una mayor diferencia significativa entre cada tratamiento al aplicar la prueba de Fisher bilateral ( $p=0$ ); y se agruparon las rutas metabólicas asociadas con procesos de descomposición de compuestos orgánicos, fijación de Nitrógeno y fijación de otros compuestos que beneficien el crecimiento de plantas, para determinar cuál tratamiento realizado en el compostaje tiene mayor potencial en aplicaciones de agricultura.

## **3.4 Formulación de protocolos guía para QIIME2 y PICRUST2**

Se crearon protocolos detallados para la instalación y uso de las herramientas QIIME2 y PICRUST2 para el estudio de comunidades microbianas por secuenciación masiva. La recopilación de información para la creación de estos protocolos se llevó a cabo a lo largo de la realización de la investigación. Se consideró necesaria la creación de estos protocolos para el apoyo de proyectos futuros que se lleven a cabo por el Grupo de Investigaciones en Minerales, Biohidrometalurgia y Ambiente (GIMBA) (2021).

## **4. Resultados y análisis**

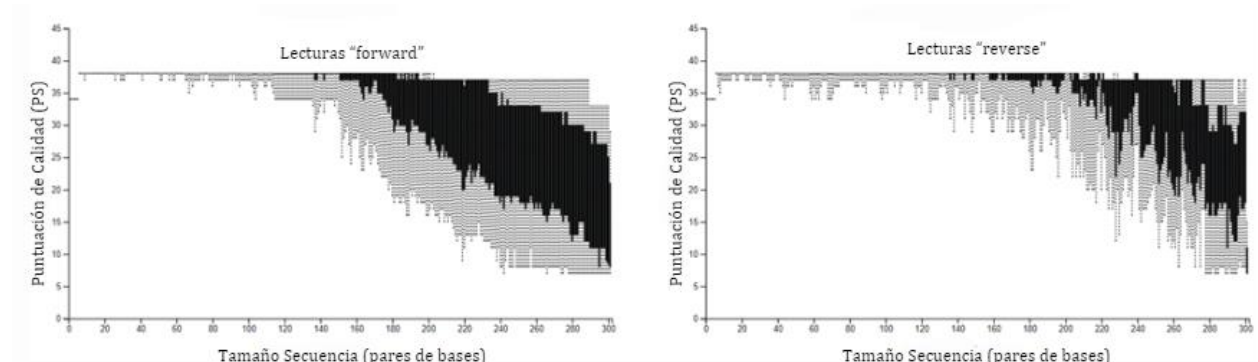
### **4.1 Análisis de calidad en FastQC**

#### **4.1.1 Calidad de secuencia por base**

La figura 5 muestra el comportamiento promedio para todo el conjunto de muestras analizadas en FastQC, visualizado como un archivo conjunto en la interfaz web *qiime2view*:

**Figura 5**

Gráfico de calidad promedio del conjunto de muestras en qiime2view.



El gráfico promedio para todas las muestras secuenciadas presenta un comportamiento de disminución de calidad a lo largo de la posición de la base, debido a la contaminación lumínica durante la detección de señales anteriores, un error que aumenta su frecuencia con el largo de la longitud de lectura (Kappelmann-Fenzl, 2021). Las lecturas *forward* de las muestras de compost disminuyen su calidad a partir de la posición 180, mientras que las lecturas *reverse* se degradan a partir de la posición 225. Generalmente los archivos de secuencias que presentan varias bases en la zona de baja calidad son descartados completamente (Malhotra et al., 2021); sin embargo, se les puede realizar *trimming* a los archivos de muestras que pasen el control de calidad para eliminar estas partes y poder realizar los análisis posteriores.

#### 4.1.2 Estadísticas básicas para cada muestra

La información básica y parámetros de los gráficos obtenidos en FastQC para cada una de las muestras se sintetizó en la tabla 1:

**Tabla 1***Estadísticas básicas y parámetros de FastQC para el conjunto de muestras*

Tratamiento	Muestra	Dirección	Secuencias Totales	Pares de bases totales (Millones)	Longitud Secuencias (pb)	Calidad promedio (PS)
Biochar fase mesófila	DR-46	forward	13707	4.1	181-301	35
		reverse	13707	4.1	181-301	36
Biochar fase mesófila – réplica	DR-47	forward	8489	2.5	158-301	35
		reverse	8489	2.5	124-301	36
Biochar fase termófila	DR-48	forward	18812	5.6	72-301	35
		reverse	18812	5.6	72-301	36
Biochar fase termófila - réplica	DR-49	forward	16884	5	35-301	35
		reverse	16884	5	35-301	36

Las secuencias ingresadas en FastQC presentan una diferencia significativa entre pares de bases y secuencias totales de las muestras DR-46 y DR-47, que corresponden al tratamiento con biochar en la fase mesófila y su réplica respectivamente. DR-47 presenta un tamaño menor comparado a todas las demás muestras y a DR-46 en particular, teniendo en cuenta que provienen del mismo tratamiento.

La muestra DR-49 que corresponde a la réplica del tratamiento con biochar en la fase termófila posee un mayor rango de longitud de secuencias de 35-301 pares de bases, en comparación con la muestra DR-48 de la misma naturaleza. Tener lecturas de longitudes muy cortas puede tener implicaciones en el análisis metagenómico posterior, ya que se pueden no detectar o identificar de forma errónea algunos homólogos de genes microbianos (Wommack et al., 2008), generando resultados de análisis taxonómicos de baja resolución (Zaura, 2012).

Con base a lo mostrado anteriormente, para realizar un análisis más preciso de las comunidades microbianas presentes, se descartaron las muestras DR-47 (tratamiento con biochar en la fase mesófila - réplica) y DR-49 (tratamiento con biochar en la fase termófila – réplica);

además de la realización de un pretratamiento para las muestras restantes DR-46 y DR-48 que consiste en un recorte (*trimming*) para eliminar las regiones de baja calidad que aún se encuentren presentes en el conjunto de archivos de lectura.

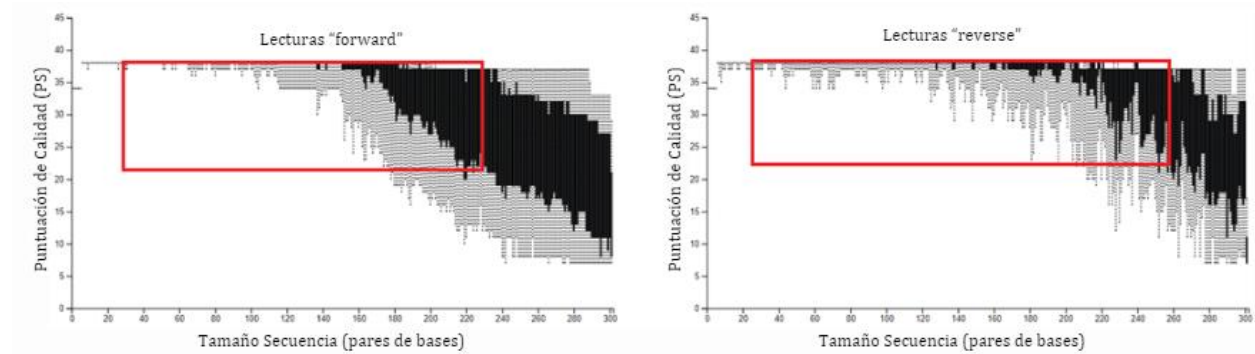
## **4.2 Análisis de diversidad alfa en QIIME2**

### **4.2.1 *Trimming y denoising de muestras en DADA2***

DADA2 requiere como entrada cuatro valores de longitud de entrada que corresponden al número de bases que se desean eliminar al principio de cada lectura (*forward* y *reverse*), que generalmente es un valor comprendido entre 20 y 30 para eliminar posibles adaptadores; y otro par de valores que corresponde a la longitud hasta la cual se desea conservar cada una de las lecturas. Para el archivo de muestras promediado que contiene las muestras DR-46 y DR-48 se realizaron cortes desde un valor inicial de 30 para ambas lecturas, y un corte final hasta la base 228 en las lecturas *forward* y 256 para las lecturas *reverse*; descartando las muestras con un  $Q < 22$ . DADA2 de manera simultánea también realizó el *denoising* y agrupación de ASVs del archivo de muestras ingresado; este segundo archivo contiene una tabla de datos que muestra la cantidad de secuencias únicas (ASVs) que se conservaron por el tratamiento realizado, la cual se utilizó como indicador para determinar si se realizaron cortes de una longitud adecuada.

**Figura 6**

*Estimado de longitud de cortes para las muestras resultantes en qiime2view.*



Para realizar una clasificación taxonómica adecuada, se recomienda una cantidad de secuencias únicas mayor o igual a 400 y unas longitudes de secuencias mayores a 100 pares de bases (Deagle et al., 2013), que además no presenten una gran variación en sus tamaños ya que pueden requerir de la implementación de métodos analíticos más sensibles (Zhou et al., 2011). Una vez realizados los cortes, se obtuvieron un total de 624 secuencias únicas con longitud promedio de 395 pares de bases y una desviación estándar de 10,54.

#### **4.2.2 Clasificación taxonómica con GreenGenes**

Se utilizó la base de datos GreenGenes (DeSantis et al., 2006) que tiene un 99% de umbral de similitud y requiere un costo computacional menor con respecto a otras bases de datos como SILVA (Quast et al., 2013) y RDP (*Ribosomal Database Project*) (Maidak et al., 1997). En este trabajo se analizó la clasificación a nivel de familia, debido a su nivel intermedio de especificidad y agrupación de bacterias que cumplan funciones similares (Parks et al., 2018).

La tabla 2 muestra las principales familias identificadas con el clasificador de GreenGenes, que presentaron una abundancia relativa mayor al 2% en cada muestra. En total, de 133 familias diferentes que fueron identificadas del conjunto de ASVs que contienen ambos tratamientos, el tratamiento con biochar en la fase mesófila registró 96 familias, de las cuales 30 no se pudieron

clasificar; por otro lado, el tratamiento en la fase termófila (DR-48) registró 126 familias, y resultaron sin clasificar 40.

**Tabla 2**

*Principales familias identificadas taxonómicamente en QIIME2 para cada tratamiento*

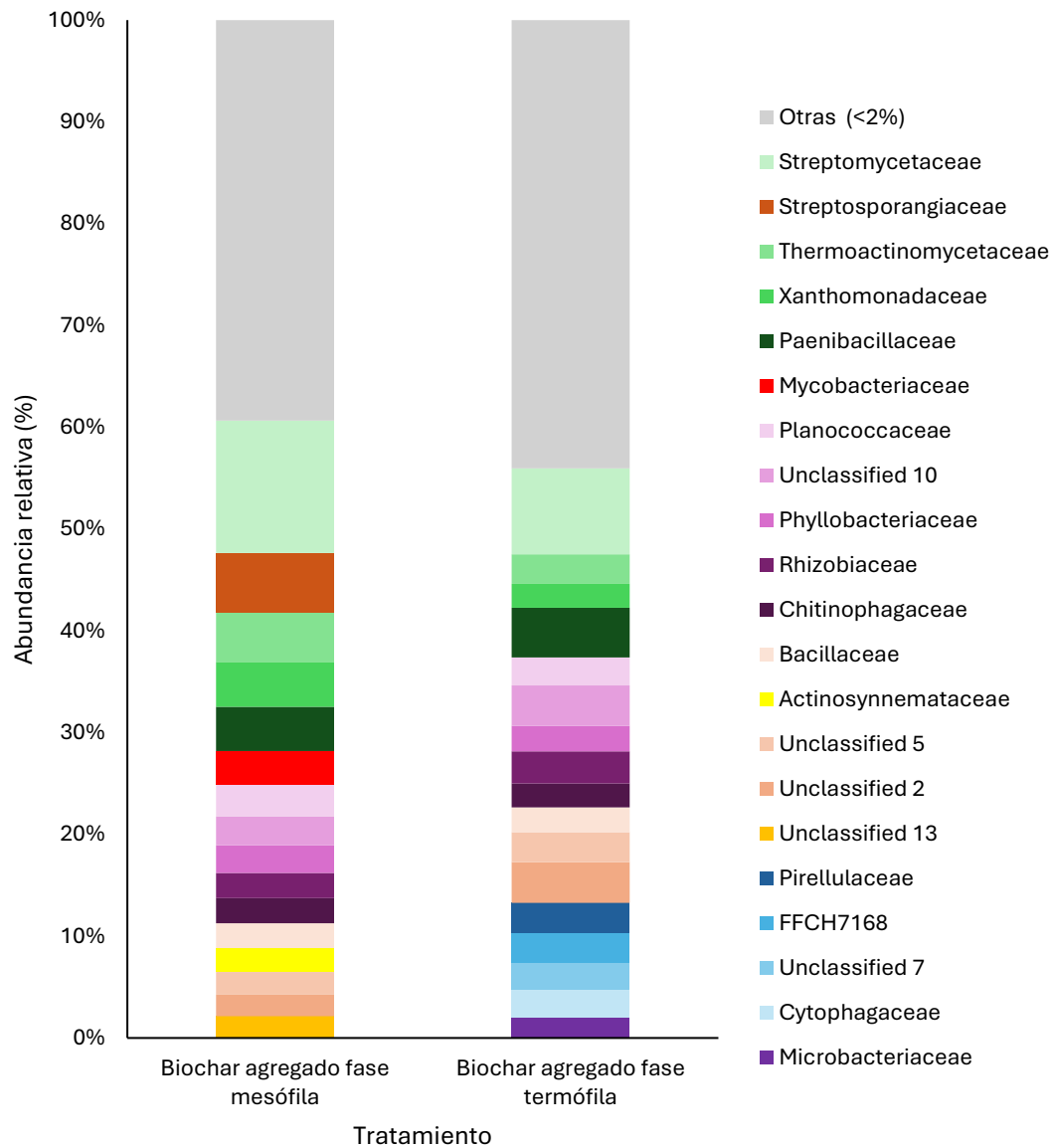
Biochar agregado en la fase mesófila		Biochar agregado en la fase termófila	
Familia	Abundancia relativa (%)	Familia	Abundancia relativa (%)
Streptomycetaceae	12,9	Streptomycetaceae	8,4
Streptosporangiaceae <sup>a</sup>	5,8	Paenibacillaceae	4,8
Thermoactinomycetaceae	4,8	Unclassified 2	4,0
Xanthomonadaceae	4,3	Unclassified 10	3,9
Paenibacillaceae	4,3	Rhizobiaceae	3,1
Mycobacteriaceae <sup>a</sup>	3,3	Microbacteriaceae <sup>b</sup>	2,9
Planococcaceae	3,1	Unclassified 5	2,9
Unclassified 10	2,8	Thermoactinomycetaceae	2,9
Phyllobacteriaceae	2,7	Cytophagaceae <sup>b</sup>	2,8
Rhizobiaceae	2,4	Planococcaceae	2,7
Bacillaceae	2,4	Unclassified 7 <sup>b</sup>	2,7
Chitinophagaceae	2,4	FFCH7168 <sup>b</sup>	2,6
Actinosynnemataceae <sup>a</sup>	2,3	Phyllobacteriaceae	2,5
Unclassified 5	2,2	Bacillaceae	2,4
Unclassified 2	2,1	Xanthomonadaceae	2,3
Unclassified 13 <sup>a</sup>	2,1	Chitinophagaceae	2,3
Otras <sup>c</sup> (80, UC=30)	40,1	Pirellulaceae <sup>b</sup>	2,0
Total	100	Otras <sup>c</sup> (109, UC=40)	44,8
		Total	100

*Nota.* <sup>a</sup>: familias con mayor abundancia en el tratamiento con biochar en la fase mesófila.

<sup>b</sup>: familias con mayor abundancia en el tratamiento con biochar en la fase termófila. <sup>c</sup>: Otras familias que presentaron una abundancia por debajo del umbral.

**Figura 7**

Gráfico de barras apiladas familias identificadas en QIIME2 para cada tratamiento



La familia más abundante identificada para ambos tratamientos es la *Streptomyetaceae*, coincidiendo con los resultados del trabajo realizado por Silva et al. (2016) en donde esta familia fue una de las más abundantes identificadas en el compost producido a partir de residuos sólidos,

y se destaca su contribución en la producción de sustancias húmicas, que ayudan a enriquecer el compost.

Para la muestra tratada con biochar en la fase mesófila, la segunda familia más abundante fue *Streptosporangiaceae* y se encuentra en mayor cantidad que en el tratamiento con biochar en la fase termófila. Esta familia de bacterias aerobias se suelen aislar para la producción de antibióticos (Pudjiraharti et al., 2011). Otra familia que aparece en un mayor porcentaje para este tratamiento es *Mycobacteriaceae*, que contiene algunas bacterias patogénicas que pueden enfermar de tuberculosis a personas, ganado y otros animales (Gunasingam, 2023); se encuentra en varios tipos de suelos y se ha probado en aplicaciones de degradación de microplásticos (X. Sun et al., 2022). Por último, la familia *Thermoactinomycetaceae*, se ha reportado que su crecimiento se puede ver favorecido por la adición de fuentes de carbono externas (Goodfellow & Jones, 2015).

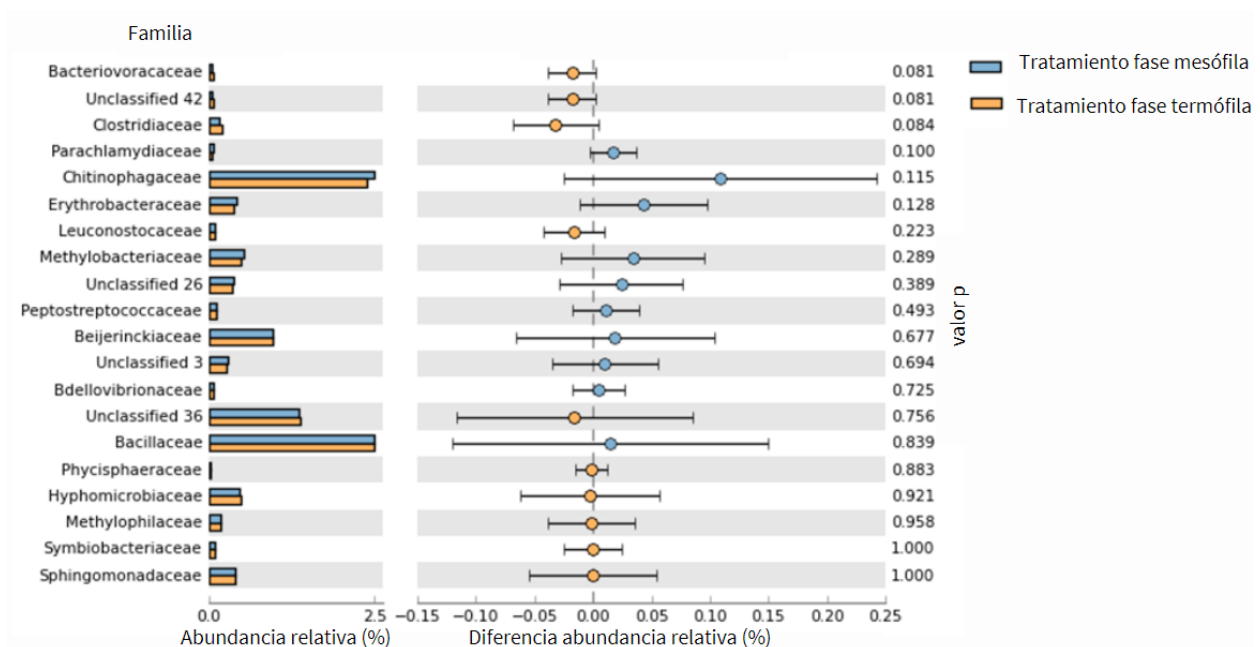
En cuanto a familias que se encuentran presentes en un mayor porcentaje de abundancia en el tratamiento con biochar en la fase termófila con respecto al tratamiento en la fase mesófila, se identificó la familia *FFCH7168* (del filo *Chloroflexi*) son predominantes en las fases termófilas de los procesos de compostaje, y son indicadores de la calidad del compost (Ren et al., 2016); la familia *Pirellulaceae* se ha relacionado con actividades de degradación de celulosa y otras sustancias, y su abundancia aumenta durante las etapas de enfriamiento y maduración del compostaje (C. Zhang et al., 2020); también se observa una mayor abundancia de las familias *Cytophagaceae* y *Microbacteriaceae*, que realizan funciones de degradación de celulosa, y tienen una mayor actividad en los primeros días del compostaje (Li et al., 2019).

### 4.2.3 Análisis estadístico en STAMP

Se filtraron en STAMP las familias con valores de  $p$  menores a 0,05 que permitan rechazar la hipótesis nula. De 133 familias diferentes ingresadas para cada tratamiento, 113 presentan un valor de  $p$  menor a 0,05, es decir, presentan diferencias estadísticamente significativas. Por otra parte, al observar en la figura 8 las abundancias relativas de las otras 20 familias con valores de  $p$  que soportan la hipótesis nula ( $p > 0,05$ , no existe diferencia en las abundancias de las familias de cada tratamiento), se observaron valores de abundancia relativa no mayores al 1% para 17 de 20 familias en cada tratamiento. Las familias con abundancias mayores al 1% identificadas taxonómicamente fueron *Bacillaceae*, *Chitinophagaceae* y una sin clasificar. En general, no se cumple la hipótesis nula, y existen diferencias estadísticamente significativas en las abundancias de las familias para cada tratamiento.

#### Figura 8

Gráfico de barras de error en STAMP para familias identificadas con poca diferencia entre proporciones ( $p > 0,05$ ).



#### 4.2.4 *Análisis de diversidad alfa*

Se compararon los índices de Shannon (riqueza y uniformidad) y Chao1 (riqueza) obtenidos en QIIME2 para cada tratamiento. El compost tratado con biochar en la fase termófila obtuvo valores mayores tanto para el índice de Chao1 como para Shannon (497 y 8 respectivamente) respecto al tratamiento con biochar adicionado en la fase mesófila (341 y 7). Ambos tratamientos presentaron índices de diversidad alfa mayores respecto al tratamiento realizado con adición de biochar en la fase termófila por Sun et al. (2021) al compost de desechos de cerdos, donde se reportan valores de 3,93 para Shannon y 305 para Chao1; pero tienen valores menores que en el tratamiento realizado por Zhang et al. (2021) con biochar de bambú al compost de desechos de cerdos, en donde se obtuvieron valores para Shannon y Chao1 de 634.3 y 62.1 respectivamente. Esta mayor abundancia a nivel de especies en el tratamiento con biochar en la fase termófila se puede observar en la tabla 3.

**Tabla 3**

*Resumen taxonomía de especies identificadas en QIIME2.*

<b>Taxonomía especies en QIIME2</b>			
Especies totales detectadas		237	
Especies sin clasificar (UC)		161	
<b>Biochar fase mesófila</b>		<b>Biochar fase termófila</b>	
Cantidad especies (>0%)	162	Cantidad especies (>0%)	215
Especies sin clasificar (UC)	105	Especies sin clasificar (UC)	149

### 4.3 Inferencia funcional de las comunidades microbianas

#### 4.3.1 Predicción de rutas MetaCyc con PICRUST2

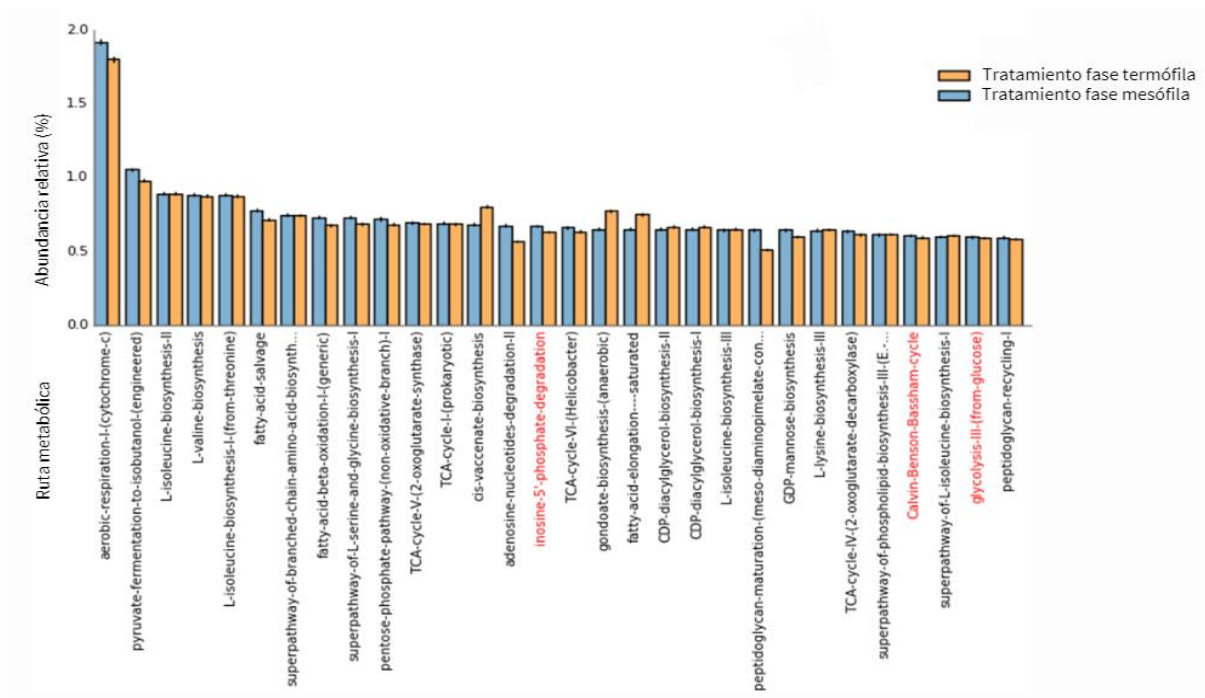
Se detectaron un total de 393 rutas metabólicas diferentes identificadas con IDs de MetaCyc. La muestra tratada con biochar en la fase mesófila registró valores de abundancia en 373 rutas, mientras que el tratamiento en la fase termófila registró 389 rutas metabólicas. Esta cantidad de rutas obtenidas es cercana a las obtenidas en la investigación realizada por Zhang et al. (2021), en donde se analizó el compost producido con biochar como aditivo para mejorar la productividad del suelo en la producción de cítricos, donde se predijeron un total de 402 rutas.

#### 4.3.2 Principales rutas metabólicas predichas para ambos tratamientos

En la figura 9 se presenta un gráfico de barras en STAMP con las rutas metabólicas obtenidas con mayor porcentaje de proporción para ambos tratamientos:

**Figura 9**

Gráfico de barras principales rutas metabólicas en STAMP.



La mayoría de las rutas principales predichas que se muestran corresponden a procesos de respiración, biosíntesis de reacciones esenciales y mantenimiento de los microorganismos que habitan las muestras de compost. La principal ruta asociada a la descomposición de compuestos orgánicos corresponde a “*glycolysis III*”, una ruta del metabolismo de azúcares simples que ha sido estudiada en especies como *Escherichia coli* de la familia *Enterobacteriaceae*, y en géneros como *Bacillus* de la familia *Bacillaceae* (Han et al., 2024); ambas familias se encuentran en la taxonomía de cada tratamiento anteriormente expuestos, presentando valores cercanos de abundancia.

La ruta “*inosine 5'-phosphate degradation*” corresponde a una ruta intermedia que produce urato a partir de la degradación del IMP (inosina monofosfato - purina), regulando el metabolismo de nitrógeno en la mayoría de los organismos (Smith & Atkins, 2002); el urato posteriormente participa en otras vías metabólicas para convertirse en amoníaco que puede ser absorbido por las plantas. Esta ruta se muestra como la más abundante relacionada con actividades de fijación de nitrógeno en el compost producido, y se ha estudiado extensamente en especies como *E. coli* (Bennett et al., 2003).

En cuanto a la fijación de carbono, la ruta más abundante asociada a esta actividad fue “*Calvin-Benson-Bassham cycle*”, una de las rutas más importantes y extensas que realizan todos los organismos de metabolismo autótrofo para fijar el dióxido de carbono inorgánico en el suelo (Li et al., 2020). Esta ruta se ha asociado y estudiado extensamente en especies de la familia *Cyanobacteriaceae* (Amichay et al., 1993); sin embargo, esta familia no se encuentra identificada en la clasificación taxonómica para ninguna muestra. Una explicación posible es que sea parte de una de las familias sin clasificar en la identificación taxonómica; otra explicación es que esta ruta de fijación de carbono sea llevada a cabo también por otras familias como *Enterobacteriaceae*, a

la cual se han realizado estudios a la especie *E. coli* de aislamiento de variantes que llevan a cabo estas actividades (Haley et al., 2014).

#### **4.3.3 Variación de funcionalidad entre tratamientos**

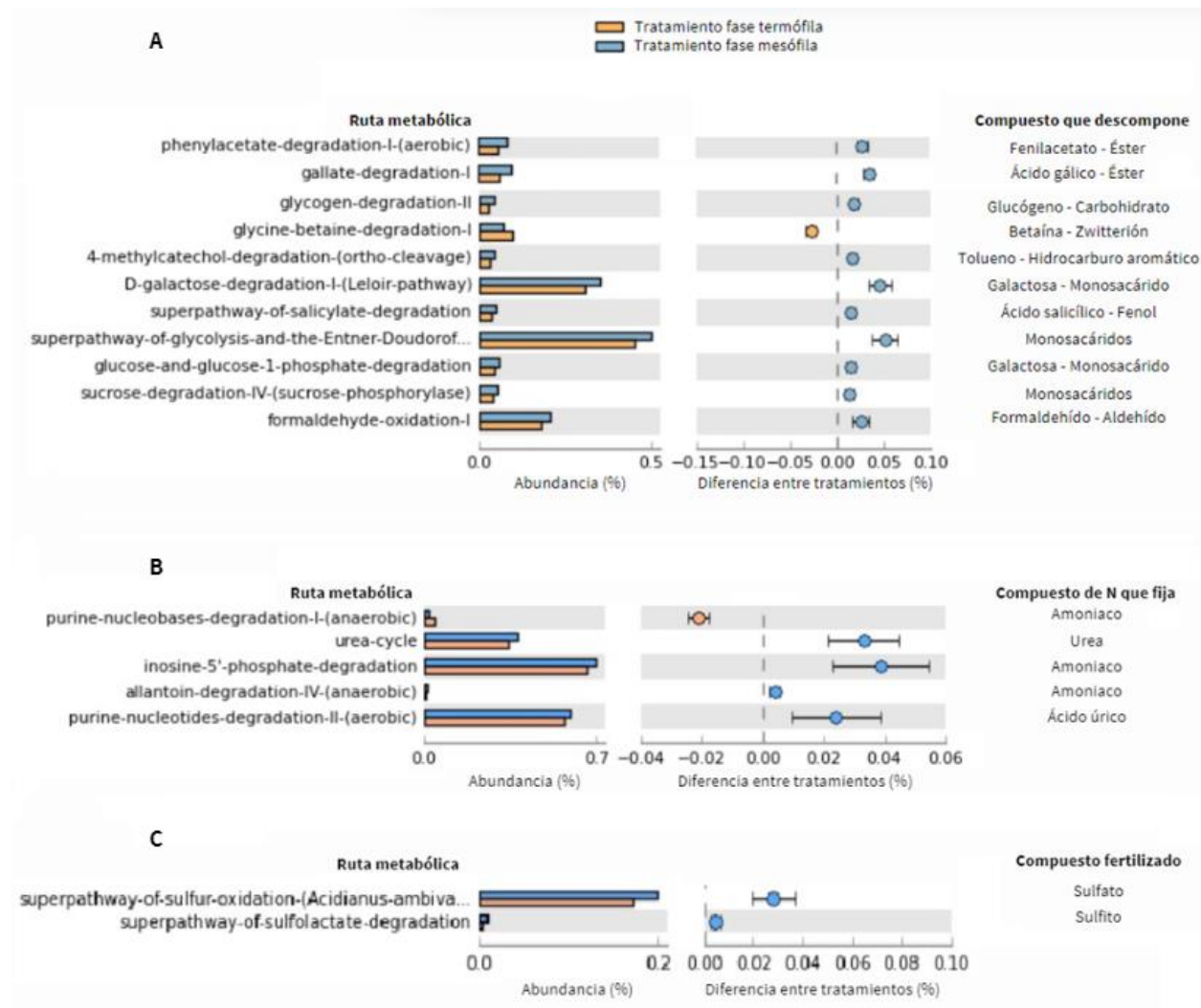
Se filtraron en STAMP un total de 185 rutas metabólicas con valores de p iguales a cero ( $p < 10^{-5}$ ) al realizar la prueba exacta de Fisher, para visualizar las rutas con mayor diferencia de abundancias entre cada tratamiento. De estas rutas filtradas, se identificaron un total de 18 relacionadas con actividades de degradación de compuestos orgánicos, fijación de nitrógeno y fertilización.

El tratamiento con biochar en la fase mesófila obtuvo una mayor abundancia en 16 de 18 rutas, a pesar de tener una diversidad menor a nivel de familia y especie, y presentar una cantidad menor de rutas metabólicas totales predichas en PICRUSt2 con respecto al tratamiento con biochar en la fase termófila. Estudios realizados en la microbiota presente en el suelo de pastizales, indican que una menor diversidad microbiana puede afectar algunas actividades específicas llevadas a cabo por la comunidad (Singh et al., 2014), y puede afectar otros roles importantes en el ecosistema como el ciclo de nitrógeno (Philippot et al., 2013); sin embargo, el incremento en la competencia e interacciones antagónicas debido a una mayor riqueza de especies, pueden reducir la funcionalidad de la comunidad microbiana (Becker et al., 2012).

A continuación, en la figura 10 se muestran rutas metabólicas identificadas asociadas a procesos de descomposición de materia y fijación de nutrientes en el compost junto con el compuesto químico que participa en cada ruta:

**Figura 10**

Gráfico de barras con perfiles en STAMP para rutas metabólicas con valores de  $p=0$  relacionadas con A) descomposición de compuestos orgánicos, B) fijación de Nitrógeno, y C) fertilización del compost.



Se identificaron un total de 11 rutas metabólicas relacionadas con procesos de descomposición de compuestos orgánicos, de las cuales 4 corresponden a descomposición de monosacáridos, 2 rutas corresponden a descomposición de ésteres y se observaron otras rutas que descomponen compuestos como tolueno, formaldehído, ácido salicílico y glucógeno. Todas las

rutas mencionadas anteriormente presentan una mayor abundancia en el tratamiento con biochar en la fase mesófila, a excepción de la ruta “*glycine-betaine-degradation-I*” que muestra mayor proporción en el tratamiento con biochar en la fase termófila, debido a una mayor abundancia en las familias *Microbacteriaceae* y *Lachnospiraceae* que contienen géneros como *Rothia* y *Blautia* respectivamente, relacionados con la degradación de la betaína que es producida en otras rutas metabólicas realizadas por bacterias del género *Bacillus* y es utilizada por la comunidad microbiana como fuente de carbono durante el compostaje durante la fase mesófila (Ryznar-Luty et al., 2015).

Con respecto a las rutas relacionadas con la degradación de compuestos orgánicos, específicamente en actividades de degradación de monosacáridos como las hexosas, que se pueden degradar en las rutas “*superpathway of glycolysis and the Entner Doudoroff pathway*”, “*glucosa and glucose-1-phosphate degradation*” y “*D-galactose degradation I (Leloir pathway)*”, se han asociado a especies como *E.coli*, pertenecientes a la familia *Enterobacteriaceae* debido a que estas rutas producen energía y permiten realizar otras actividades de biosíntesis (Y. E. Kim et al., 2022). La ruta de Entner Doudoroff es una de las principales realizadas por especies de la familia *Pseudomonadaceae* (Munoz-Elias & McKinney, 2006). Otras rutas relacionadas con la degradación de ésteres como el fenilacetato y ácido gálico se han descrito en especies como *Pseudomonas putida U* y *E. coli* pertenecientes también a las familias *Pseudomonadaceae* y *Enterobacteriaceae* (Navarro-Llorens et al., 2005). El mecanismo de la ruta “*glycogen degradation IP*” que descompone glucógeno, se ha estudiado en géneros como *Mycobacteria*, *Pseudomonas* y *Xanthomonas* (Cifuentes et al., 2019) pertenecientes a las familias *Pseudomonadaceae*, *Mycobacteriaceae* y *Xanthomonadaceae* respectivamente.

Se encontraron 5 rutas relacionadas con actividades de fijación de nitrógeno presentes en cada tratamiento, de las cuales 4 presentan una mayor abundancia en el tratamiento con biochar en la fase mesófila a excepción de la ruta “*purine nucleobases degradation I (anaerobic)*”, que tiene mayor abundancia en el tratamiento en la fase termófila, a la cual se le ha estudiado su mecanismo de reacción para la descomposición de purinas llevado a cabo por la especie *E.coli* (Zrenner et al., 2006). Las rutas de fijación de amoníaco “*inosine 5'-phosphate degradation*” y “*allantoin degradation IV (anaerobic)*” se han asociado a las familias *Pseudomonadaceae* y *Enterobacteriaceae* en donde se destaca su importancia como fijadores de nitrógeno en la salud de los suelos y la productividad de cultivos (Jalloh et al., 2024). Se identificó la ruta “*urea cycle*” que lleva a cabo el proceso de fertilización de la urea en el ciclo del Nitrógeno, y se muestra la participación del orden *Burkholderiales*, que contienen a las familias *Beijerinckiaceae*, *Brucellaceae*, *Phyllobacteriaceae* y *Rhizobiaceae*, *Alcaligenaceae*, *Comamonadaceae* y *Oxalobacteraceae* (C. Zhang et al., 2019), que se identificaron para ambos tratamientos en la identificación taxonómica

Por último, se encontraron 2 rutas relacionadas con la fertilización de otros compuestos que enriquecieron el compost producido para cada tratamiento, ambas rutas muestran valores de abundancia mayores en el tratamiento con biochar en la fase mesófila, y también fijan compuestos químicos que contienen azufre. Las rutas “*superpathway of sulfolactate degradation*” y “*superpathway of sulfur oxidation (Acidianus ambivalens)*” liberan sulfito y sulfato respectivamente, siendo el sulfato la principal fuente de azufre de las plantas, necesario para su crecimiento; y el sulfito es un intermediario importante en la asimilación del sulfato en la síntesis de aminoácidos esenciales (Kurmanbayeva et al., 2017), y se ha estudiado en especies del género *Pseudomonas* (Guo et al., 2019).

A continuación, en la tabla 4, se muestran las abundancias de las familias anteriormente mencionadas en las actividades de degradación de compuestos orgánicos, fijación de nitrógeno y fertilización identificadas en QIIME2 para cada tratamiento.

**Tabla 4**

*Familias identificadas en QIIME2 para cada tratamiento relacionadas con actividades de descomposición y fijación de nutrientes.*

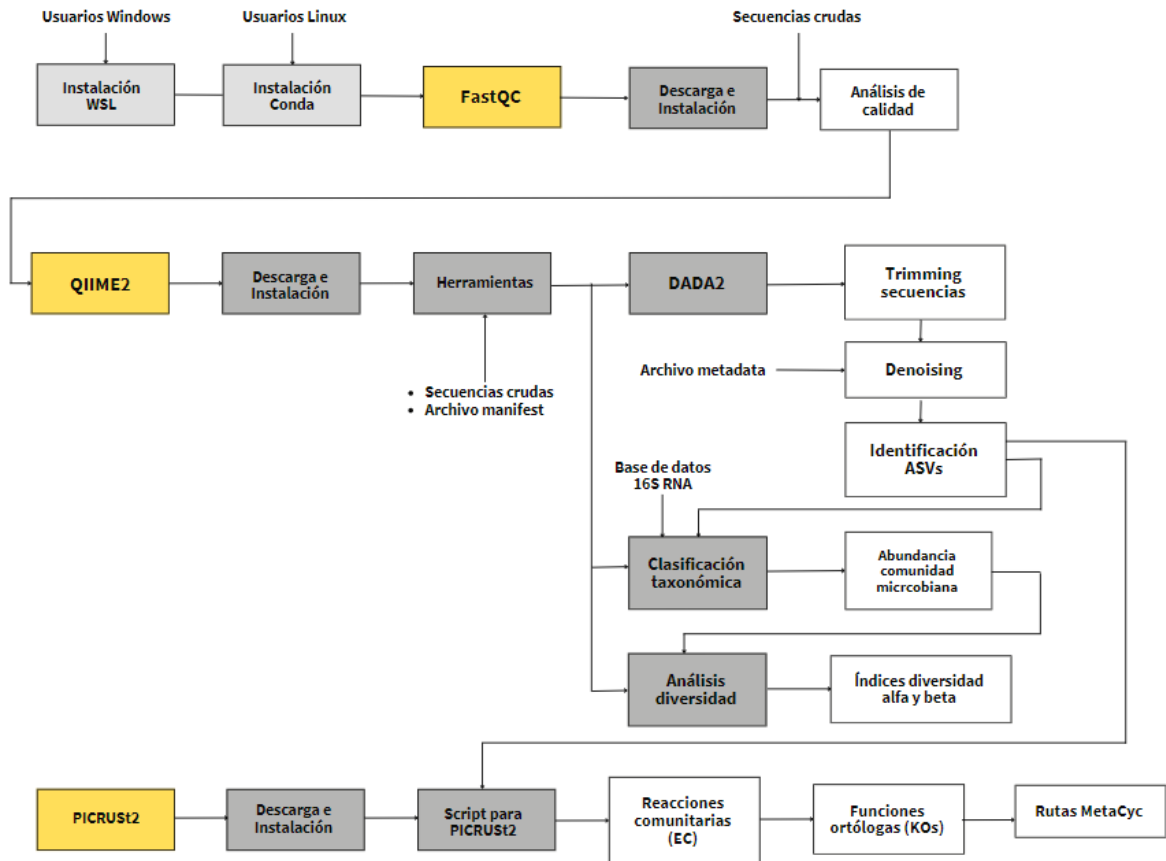
Familia	Abundancia Relativa (%)		Actividad relacionada
	Tratamiento biochar fase mesófila	Tratamiento biochar fase termófila	
Xanthomonadaceae	4,4	2,3	Degradación C
Mycobacteriaceae	3,3	1,9	Degradación C
Phyllobacteriaceae	2,7	2,5	Fijación N
Rhizobiaceae	2,5	3,2	Fijación N
Comamonadaceae	1,1	1,2	Fijación N
Beijerinckiaceae	1,0	0,9	Fijación N
Alcaligenaceae	0,8	0,6	Fijación N
Microbacteriaceae	0,8	3,0	Degradación C
Pseudomonadaceae	0,6	0,4	Degradación C, Fijación N, Fertilización
Brucellaceae	0,5	0,4	Fijación N
Lachnospiraceae	0,2	0,3	Degradación C
Enterobacteriaceae	0,1	0,1	Degradación C, Fijación N
Oxalobacteraceae	0,0	0,2	Fijación N

#### 4.4 Formulación de protocolos guía para QIIME2 y PICRUST2

Se crearon guías detalladas para la instalación y uso básico de las herramientas QIIME2 y PICRUST2, utilizadas para el presente trabajo de investigación. Este material fue entregado a los miembros del grupo de investigación GIMBA y se puede observar un resumen del contenido en el diagrama de flujo que se muestra en la figura 11.

**Figura 11**

Diagrama de flujo resumen contenido del protocolo para QIIME2 y PICRUST2



## 5. Conclusiones

El compost tratado con la adición de biochar en la fase termófila obtuvo una mayor cantidad de familias y especies (126 y 215 respectivamente) identificadas taxonómicamente en QIIME2 con respecto al tratamiento realizado en la fase mesófila (96 y 162), además obtuvo mayores índices de diversidad alfa como Shannon y Chao1 relacionados con la riqueza y

uniformidad de la microbiota presente en el compost (497 y 8 tratamiento fase termófila y 341 y 7 tratamiento fase mesófila).

El compost tratado con biochar en la fase termófila obtuvo una mayor cantidad de rutas metabólicas predichas en PICRUSt2 (389) que el tratamiento en la fase mesófila (373). Sin embargo, el tratamiento en la fase mesófila obtuvo mayores valores de abundancia en rutas asociadas a procesos de descomposición de compuestos orgánicos como monosacáridos, carbohidratos, ésteres y compuestos como el glucógeno; también obtuvo mayor abundancia en rutas de fijación de nitrógeno para compuestos como amoníaco y urea, y se observaron diferencias en las abundancias de rutas asociadas a fertilización de compuestos que contienen azufre. Estas actividades fueron asociadas principalmente a las familias *Pseudomonadaceae* y *Enterobacteriaceae* que se encontraron en mayor proporción para el tratamiento en la fase mesófila, debido a una menor competencia entre especies.

Se formularon protocolos detallados para el uso de las herramientas bioinformáticas QIIME2 y PICRUSt2, que apoyen el estudio de comunidades microbianas con técnicas de secuenciación masiva.

## 6. Recomendaciones

Realizar la caracterización y comparación de comunidades microbianas y predicción de rutas metabólicas para determinar diferencias en el compost producido con la adición y sin la adición de biochar.

**Referencias Bibliográficas**

- Amichay, D., Levitz, R., & Gurevitz, M. (1993). Construction of a *Synechocystis* PCC6803 mutant suitable for the study of variant hexadecameric ribulose biphosphate carboxylase/oxygenase enzymes. *Plant Molecular Biology*, 23(3), 465-476. <https://doi.org/10.1007/BF00019295>
- Becker, J., Eisenhauer, N., Scheu, S., & Jousset, A. (2012). Increasing antagonistic interactions cause bacterial communities to collapse at high diversity. *Ecology Letters*, 15(5), 468-474. <https://doi.org/10.1111/j.1461-0248.2012.01759.x>
- Bennett, E. M., Li, C., Allan, P. W., Parker, W. B., & Ealick, S. E. (2003). Structural basis for substrate specificity of *Escherichia coli* purine nucleoside phosphorylase. *The Journal of Biological Chemistry*, 278(47), 47110-47118. <https://doi.org/10.1074/jbc.M304622200>
- Bohórquez Santana, W. (2019). El proceso de compostaje (Vol. 1). Universidad de la Salle. [https://books.google.es/books?hl=es&lr=&id=X\\_1DwAAQBAJ&oi=fnd&pg=PA7&dq=fase+%22enfriamiento%22+compostaje&ots=0JkT6Q4KCd&sig=cXAZPD19bUpSUjBRr5rLoJION44](https://books.google.es/books?hl=es&lr=&id=X_1DwAAQBAJ&oi=fnd&pg=PA7&dq=fase+%22enfriamiento%22+compostaje&ots=0JkT6Q4KCd&sig=cXAZPD19bUpSUjBRr5rLoJION44)
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), 852-857. <https://doi.org/10.1038/s41587-019-0209-9>

- Bravo, A. (2021). Caracterización de la Diversidad Microbiana de Suelos Destinados a Cultivos de Ananas comosus (piña), Tratados con Distintos Plaguicidas en el Municipio de Lebrija, Santander.
- Brenes-Guillén, L. (2018). Herramientas bioinformáticas para el análisis de diversidad de comunidades microbianas: Qiime, Mothur y DADA2. *Revista de Biología Tropical, Blog-Blog*.
- Bueno, P., Díaz, M., & Cabrera, F. (2008). Factores que afectan al proceso de compostaje. *Compostaje*, 1, 101-102.
- Cai, H., Archambault, M., & Prescott, J. F. (2003). 16S Ribosomal RNA Sequence—Based Identification of Veterinary Clinical Bacteria. *Journal of Veterinary Diagnostic Investigation*, 15(5), 465-469. <https://doi.org/10.1177/104063870301500511>
- Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, 11(12), 2639-2643. <https://doi.org/10.1038/ismej.2017.119>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581-583. <https://doi.org/10.1038/nmeth.3869>
- Can, M. (2019). Annotation of Bacteria by Greengenes Classifier Using 16S rRNA Gene Hyper Variable Regions. *Southeast Europe Journal of Soft Computing*, 8(2), Article 2. <https://doi.org/10.21533/scjournal.v8i2.181>
- Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., & Mueller, L. A. (2016). The MetaCyc database of

- metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research*, 44(D1), D471-D480.
- Cassol, I., Ibañez, M., & Bustamante, J. P. (2023). Unraveling Key features of Microbial Alpha-Diversity Metrics and their practical applications. <https://www.researchsquare.com/article/rs-2595260/latest>
- Castro, E. R. V., & Chanamé, C. E. M. (2021). Una revisión sobre la diversidad microbiana y su rol en el compostaje aerobio. *Aporte Santiaguino*, ág. 253-275. <https://doi.org/10.32911/as.2021.v14.n2.822>
- Cifuentes, J. O., Comino, N., Trastoy, B., D'Angelo, C., & Guerin, M. E. (2019). Structural basis of glycogen metabolism in bacteria. *Biochemical Journal*, 476(14), 2059-2092.
- de Sena Brandine, G., & Smith, A. D. (2021). Falco: High-speed FastQC emulation for quality control of sequencing data. *F1000Research*, 8, 1874. <https://doi.org/10.12688/f1000research.21142.2>
- Deagle, B. E., Thomas, A. C., Shaffer, A. K., Trites, A. W., & Jarman, S. N. (2013). Quantifying sequence proportions in a DNA -based diet study using Ion Torrent amplicon sequencing: Which counts count? *Molecular Ecology Resources*, 13(4), 620-633. <https://doi.org/10.1111/1755-0998.12103>
- del Rosario Rodicio, M., & del Carmen Mendoza, M. (2004). Identificación bacteriana mediante secuenciación del ARNr 16S: Fundamento, metodología y aplicaciones en microbiología clínica. *Enfermedades infecciosas y microbiología clínica*, 22(4), 238-245.
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., & Andersen, G. L. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene

- Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, 72(7), 5069-5072. <https://doi.org/10.1128/AEM.03006-05>
- Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. (2020). PICRUSt2 for prediction of metagenome functions. *Nature biotechnology*, 38(6), 685-688.
- FastQC, A. (2015). Quality control tool for high throughput sequence data. BibSonomy. [(accessed on 17 March 2022)]. Available online: <https://www.bibsonomy.org/bibtex/f230a919c34360709aa298734d63dca3>, 1.
- Franco, M. L., Cediél, J. F., & Payán, C. (2008). Breve historia de la bioinformática. *Colombia Médica*, 39(1), 117-120.
- Galezo, A., & Diaz, J. (2024). Comparación del efecto de la adición de biochar en diferentes fases del proceso en el co-compostaje de residuos verdes y de alimentos. Universidad Industrial de Santander.
- Ghazanfar, S., Arbabi, A., Ghazanfar, M., Anjum, M., & Begum, I. (2010). Metagenomics and its application in soil microbial community studies: Biotechnological prospects. *J. Anim. Plant Sci.*, 6, 611-622.
- González-Marquetti, I., Rodríguez, M. G., Delgado-Oramas, B. P., & Schmidt, H.-P. (2020). Biochar y su contribución a la nutrición, crecimiento y defensa de las plantas. *Revista de Protección Vegetal*, 35(2). <https://censa.edicionescervantes.com/index.php/RPV/article/view/1090>
- Goodfellow, M., & Jones, A. L. (2015). *Bergey's Manual of Systematics of Archaea and Bacteria*. En W. B. Whitman (Ed.), *Bergey's Manual of Systematics of Archaea and Bacteria* (1.a ed., pp. 1-18). Wiley. <https://doi.org/10.1002/9781118960608.fbm00119>

- Grupo de investigación en Recursos Hídricos y Saneamiento Ambiental (GPH). (2024). <http://albatros.uis.edu.co/eisi/grupo/gph/#views/gm1/inicio>
- Grupo de Investigaciones en Minerales, Biohidrometalurgia y Ambiente (GIMBA). (2021). Universidad Industrial de Santander. <https://uis.edu.co/ffq-gruinv-gimba-es/>
- Gunasingam, N. (2023). An Introduction to the Pathogenesis of Mycobacteriaceae Family. *Mycobacterial Diseases*, 13(3), 1-1. <https://doi.org/10.35248/2161-1068.23.13.335>
- Guo, H., Chen, C., & Lee, D.-J. (2019). Nitrogen and sulfur metabolisms of *Pseudomonas* sp. C27 under mixotrophic growth condition. *Bioresource Technology*, 293, 122169. <https://doi.org/10.1016/j.biortech.2019.122169>
- Haley, R., Fruchtl, M., Brune, E. M., Ataa, M., Henry, R., & Beitle, R. (2014). A redesigned *Escherichia coli* triosephosphate isomerase restores growth properties in a bacterial strain useful for Immobilized Metal Affinity Chromatography (IMAC). *Journal of Biotechnology*, 188, 48-52. <https://doi.org/10.1016/j.jbiotec.2014.07.432>
- Han, Y., Zhang, Y., Yang, Z., Zhang, Q., He, X., Song, Y., Tian, L., & Wu, H. (2024). Improving Aerobic Digestion of Food Waste by Adding a Personalized Microbial Inoculum. *Current Microbiology*, 81(9), 277. <https://doi.org/10.1007/s00284-024-03796-5>
- Hernández, M., Quijada, N. M., Rodríguez-Lázaro, D., & Eiros, J. M. (2020). Aplicación de la secuenciación masiva y la bioinformática al diagnóstico microbiológico clínico. *Revista Argentina de Microbiología*, 52(2), 150-161. <https://doi.org/10.1016/j.ram.2019.06.003>
- Illumina Inc. (2024). MiSeq System. <https://www.illumina.com/systems/sequencing-platforms/miseq/specifications.html>
- Innis, M. A., Gelfand, D. H., Sninsky, J. J., & White, T. J. (2012). PCR protocols: A guide to methods and applications. Academic press.

<https://books.google.es/books?hl=es&lr=&id=Z5jwZ2rbVe8C&oi=fnd&pg=PP1&dq=PCR+Protocols:+A+Guide+to+Methods+and+Applications&ots=ICGZJiWQ6w&sig=QFzujhUAM6qoAxdRZorUIMzmlPI>

- Jalloh, A. A., Khamis, F. M., Yusuf, A. A., Subramanian, S., & Mutyambai, D. M. (2024). Long-term push–pull cropping system shifts soil and maize-root microbiome diversity paving way to resilient farming system. *BMC Microbiology*, 24(1), 92. <https://doi.org/10.1186/s12866-024-03238-z>
- Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*, 51(D1), D587-D592. <https://doi.org/10.1093/nar/gkac963>
- Kappelmann-Fenzl, M. (2021). NGS Data. En M. Kappelmann-Fenzl (Ed.), *Next Generation Sequencing and Data Analysis* (pp. 79-104). Springer International Publishing. [https://doi.org/10.1007/978-3-030-62490-3\\_7](https://doi.org/10.1007/978-3-030-62490-3_7)
- Kchouk, M., Gibrat, J.-F., & Elloumi, M. (2017). Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine*, 09. <https://doi.org/10.4172/0974-8369.1000395>
- Kim, B.-R., Shin, J., Guevarra, R. B., Lee, J. H., Kim, D. W., Seol, K.-H., Lee, J.-H., Kim, H. B., & Isaacson, R. E. (2017). Deciphering diversity indices for a better understanding of microbial communities. *Journal of Microbiology and Biotechnology*, 27(12), 2089-2093.
- Kim, Y. E., Cho, K. H., Bang, I., Kim, C. H., Ryu, Y. S., Kim, Y., Choi, E. M., Nong, L. K., Kim, D., & Lee, S. K. (2022). Characterization of an Entner–Doudoroff pathway-activated *Escherichia coli*. *Biotechnology for Biofuels and Bioproducts*, 15(1), 120. <https://doi.org/10.1186/s13068-022-02219-6>

- Kurmanbayeva, A., Brychkova, G., Bekturova, A., Khozin, I., Standing, D., Yarmolinsky, D., & Sagi, M. (2017). Determination of Total Sulfur, Sulfate, Sulfite, Thiosulfate, and Sulfolipids in Plants. En R. Sunkar (Ed.), *Plant Stress Tolerance: Methods and Protocols* (pp. 253-271). Springer. [https://doi.org/10.1007/978-1-4939-7136-7\\_15](https://doi.org/10.1007/978-1-4939-7136-7_15)
- Laich, F. (2011). El papel de los microorganismos en el proceso de compostaje. *Jornada Técnica: Fertilidad y Calidad del Suelo. Experiencias de fertilización orgánica en platanera*. Instituto Canario de Investigaciones Agrarias. ICIA, 21, 1-7.
- Lee, M. (2019). Happy Belly Bioinformatics: An open-source resource dedicated to helping biologists utilize bioinformatics. *Journal of Open Source Education*, 2(19), 53. <https://doi.org/10.21105/jose.00053>
- Lemos, L. N., Fulthorpe, R. R., Triplett, E. W., & Roesch, L. F. W. (2011). Rethinking microbial diversity analysis in the high throughput sequencing era. *Journal of Microbiological Methods*, 86(1), 42-51. <https://doi.org/10.1016/j.mimet.2011.03.014>
- Li, Z., Xin, X., Xiong, B., Zhao, D., Zhang, X., & Bi, C. (2020). Engineering the Calvin–Benson–Bassham cycle and hydrogen utilization pathway of *Ralstonia eutropha* for improved autotrophic growth and polyhydroxybutyrate production. *Microbial Cell Factories*, 19(1), 228. <https://doi.org/10.1186/s12934-020-01494-y>
- Li, Z., Yang, Y., Xia, Y., Wu, T., Zhu, J., Wang, Z., & Yang, J. (2019). The succession pattern of bacterial diversity in compost using pig manure mixed with wood chips analyzed by 16S rRNA gene analysis. *BioRxiv*, 674069.
- Liao, P., Satten, G. A., & Hu, Y.-J. (2017). PhredEM: A Phred-Score-Informed Genotype-Calling Approach for Next-Generation Sequencing Studies. *Genetic epidemiology*, 41(5), 375-387. <https://doi.org/10.1002/gepi.22048>

- Lozada, A. E., Lagarda, G. G., Jiménez, A. M., & Zapata, F. B. (2004). Diversidad bacteriana del suelo: Métodos de estudio no dependientes del cultivo microbiano e implicaciones biotecnológicas. *Agrociencia*, 38(6), 583-592.
- MacManes, M. D. (2014). On the optimal trimming of high-throughput mRNA sequence data. *Frontiers in genetics*, 5, 13.
- Maidak, B. L., Olsen, G. J., Larsen, N., Overbeek, R., McCaughey, M. J., & Woese, C. R. (1997). The RDP (Ribosomal Database Project). *Nucleic Acids Research*, 25(1), 109-110.  
<https://doi.org/10.1093/nar/25.1.109>
- Malhotra, A., Das, S., & Rai, S. N. (2021). Supplementary Materials of Analysis of Single-Cell RNA-Sequencing Data: A Step-By-Step Guide.  
[https://www.researchgate.net/profile/Samarendra-Das-4/publication/357394783\\_Supplementary\\_Materials\\_of\\_Analysis\\_of\\_Single-Cell\\_RNA-Sequencing\\_Data\\_A\\_Step-By-Step\\_Guide/links/61cc2a4fb8305f7c4b0b804a/Supplementary-Materials-of-Analysis-of-Single-Cell-RNA-Sequencing-Data-A-Step-By-Step-Guide.pdf](https://www.researchgate.net/profile/Samarendra-Das-4/publication/357394783_Supplementary_Materials_of_Analysis_of_Single-Cell_RNA-Sequencing_Data_A_Step-By-Step_Guide/links/61cc2a4fb8305f7c4b0b804a/Supplementary-Materials-of-Analysis-of-Single-Cell-RNA-Sequencing-Data-A-Step-By-Step-Guide.pdf)
- McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., Andersen, G. L., Knight, R., & Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal*, 6(3), 610-618.
- Mirarab, S., Nguyen, N., & Warnow, T. (2012). SEPP: SATé-enabled phylogenetic placement. *Pacific Symposium on Biocomputing*. Pacific Symposium on Biocomputing, 247-258.  
[https://doi.org/10.1142/9789814366496\\_0024](https://doi.org/10.1142/9789814366496_0024)

- Mordoh, A. (2019). Secuenciación masiva de ADN: La próxima generación. *Dermatología Argentina*, 25(1), Article 1.
- Munoz-Elias, E. J., & McKinney, J. D. (2006). Carbon metabolism of intracellular bacteria. *Cellular Microbiology*, 8(1), 10-22. <https://doi.org/10.1111/j.1462-5822.2005.00648.x>
- Navarro-Llorens, J. M., Patrauchan, M. A., Stewart, G. R., Davies, J. E., Eltis, L. D., & Mohn, W. W. (2005). Phenylacetate Catabolism in *Rhodococcus* sp. Strain RHA1: A Central Pathway for Degradation of Aromatic Compounds. *Journal of Bacteriology*, 187(13), 4497-4504. <https://doi.org/10.1128/JB.187.13.4497-4504.2005>
- Nkongolo, K. K., & Narendrula-Kotha, R. (2020). Advances in monitoring soil microbial community dynamic and function. *Journal of Applied Genetics*, 61(2), 249-263. <https://doi.org/10.1007/s13353-020-00549-5>
- O'Rawe, J. A., Ferson, S., & Lyon, G. J. (2015). Accounting for uncertainty in DNA sequencing data. *Trends in Genetics*, 31(2), 61-66. <https://doi.org/10.1016/j.tig.2014.12.002>
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarszewski, A., Chaumeil, P.-A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature biotechnology*, 36(10), 996-1004.
- Parks, D. H., Tyson, G. W., Hugenholtz, P., & Beiko, R. G. (2014). STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics*, 30(21), 3123-3124. <https://doi.org/10.1093/bioinformatics/btu494>
- Pérez, M. H. (2016). Seguridad alimentaria: El bueno, el feo y el malo, de las UFC a las OTUs. <http://crinoidea.semimicrobiologia.org/pdf/actualidad/62/28-MHernandez-62.pdf>

- Philippot, L., Spor, A., Hénault, C., Bru, D., Bizouard, F., Jones, C. M., Sarr, A., & Maron, P.-A. (2013). Loss in microbial diversity affects nitrogen cycling in soil. *The ISME journal*, 7(8), 1609-1619.
- Pudjiraharti, S., Takesue, N., Katayama, T., Lisdiyanti, P., Hanafi, M., Tanaka, M., Sone, T., & Asano, K. (2011). Actinomycete *Nonomuraea* sp. Isolated from Indonesian soil is a new producer of inulin fructotransferase. *Journal of Bioscience and Bioengineering*, 111(6), 671-674. <https://doi.org/10.1016/j.jbiosc.2011.02.001>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590-D596. <https://doi.org/10.1093/nar/gks1219>
- Ren, G., Xu, X., Qu, J., Zhu, L., & Wang, T. (2016). Evaluation of microbial population dynamics in the co-composting of cow manure and rice straw using high throughput sequencing analysis. *World Journal of Microbiology and Biotechnology*, 32(6), 101. <https://doi.org/10.1007/s11274-016-2059-7>
- Ritz, K. (2007). The plate debate: Cultivable communities have no utility in contemporary environmental microbial ecology. *FEMS Microbiology Ecology*, 60(3), 358-362.
- Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., & Tuohy, K. (2018). Gut microbiota functions: Metabolism of nutrients and other food components. *European Journal of Nutrition*, 57(1), 1-24. <https://doi.org/10.1007/s00394-017-1445-8>
- Rubio, S., Pacheco-Orozco, R. A., Milena Gómez, A., Perdomo, S., & García-Robles, R. (2020). Secuenciación de nueva generación (NGS) de ADN: Presente y futuro en la práctica clínica. *Universitas Médica*, 61(2), 49-63.

- Ryznar-Luty, A., Cibis, E., Krzywonos, M., & Miśkiewicz, T. (2015). Efficiency of aerobic biodegradation of beet molasses vinasse under non-controlled pH: Conditions for betaine removal. *Archives of Environmental Protection*, 41(1). <http://archive.sciendo.com/AEP/aep.2015.41.issue-1/aep-2015-0001/aep-2015-0001.pdf>
- Silva, M. E. F., Lopes, A. R., Cunha-Queda, A. C., & Nunes, O. C. (2016). Comparison of the bacterial composition of two commercial composts with different physicochemical, stability and maturity properties. *Waste Management*, 50, 20-30. <https://doi.org/10.1016/j.wasman.2016.02.023>
- Singh, B. K., Quince, C., Macdonald, C. A., Khachane, A., Thomas, N., Al-Soud, W. A., Sørensen, S. J., He, Z., White, D., Sinclair, A., Crooks, B., Zhou, J., & Campbell, C. D. (2014). Loss of microbial diversity in soils is coincident with reductions in some specialized functions. *Environmental Microbiology*, 16(8), 2408-2420. <https://doi.org/10.1111/1462-2920.12353>
- Smith, P. M. C., & Atkins, C. A. (2002). Purine biosynthesis. Big in cell division, even bigger in nitrogen assimilation. *Plant Physiology*, 128(3), 793-802. <https://doi.org/10.1104/pp.010912>
- Storey, S., Chualain, D. N., Doyle, O., Clipson, N., & Doyle, E. (2015). Comparison of bacterial succession in green waste composts amended with inorganic fertiliser and wastewater treatment plant sludge. *Bioresource Technology*, 179, 71-77. <https://doi.org/10.1016/j.biortech.2014.11.107>
- Sun, L., Long, M., Li, J., Wu, R., Ma, L., Tang, D., Lu, Y., & Wang, Z. (2021). Different Effects of Thermophilic Microbiological Inoculation With and Without Biochar on Physicochemical Characteristics and Bacterial Communities in Pig Manure Composting. *Frontiers in Microbiology*, 12. <https://doi.org/10.3389/fmicb.2021.746718>

- Sun, X., Chen, Z., Kong, T., Chen, Z., Dong, Y., Kolton, M., Cao, Z., Zhang, X., Zhang, H., Liu, G., Gao, P., Yang, N., Lan, L., Xu, Y., & Sun, W. (2022). Mycobacteriaceae Mineralizes Micropolyethylene in Riverine Ecosystems. *Environmental Science & Technology*, 56(22), 15705-15717. <https://doi.org/10.1021/acs.est.2c05346>
- Tanya Morocho, M., Leiva-Mora, M., Tanya Morocho, M., & Leiva-Mora, M. (2019). Microorganismos eficientes, propiedades funcionales y aplicaciones agrícolas. *Centro Agrícola*, 46(2), 93-103.
- Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., & Gordon, J. I. (2009). The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Science Translational Medicine*, 1(6), 6ra14-6ra14. <https://doi.org/10.1126/scitranslmed.3000322>
- Velázquez Martí, B. (2024). Plantas de compostaje. <https://riunet.upv.es/handle/10251/201587>
- Wang, X., Cui, H., Shi, J., Zhao, X., Zhao, Y., & Wei, Z. (2015). Relationship between bacterial diversity and environmental parameters during composting of different raw materials. *Bioresource Technology*, 198, 395-402. <https://doi.org/10.1016/j.biortech.2015.09.041>
- Willis, A. D. (2019). Rarefaction, Alpha Diversity, and Statistics. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.02407>
- Wommack, K. E., Bhavsar, J., & Ravel, J. (2008). Metagenomics: Read Length Matters. *Applied and Environmental Microbiology*, 74(5), 1453-1463. <https://doi.org/10.1128/AEM.02181-07>
- Ye, Y., & Doak, T. G. (2009). A Parsimony Approach to Biological Pathway Reconstruction/Inference for Genomes and Metagenomes. *PLOS Computational Biology*, 5(8), e1000465. <https://doi.org/10.1371/journal.pcbi.1000465>

- Zaura, E. (2012). Next-generation Sequencing Approaches to Understanding the Oral Microbiome. *Advances in Dental Research*, 24(2), 81-85. <https://doi.org/10.1177/0022034512449466>
- Zhang, C., Gao, Z., Shi, W., Li, L., Tian, R., Huang, J., Lin, R., Wang, B., & Zhou, B. (2020). Material conversion, microbial community composition and metabolic functional succession during green soybean hull composting. *Bioresource Technology*, 316, 123823. <https://doi.org/10.1016/j.biortech.2020.123823>
- Zhang, C., Song, Z., Zhuang, D., Wang, J., Xie, S., & Liu, G. (2019). Urea fertilization decreases soil bacterial diversity, but improves microbial biomass, respiration, and N-cycling potential in a semiarid grassland. *Biology and Fertility of Soils*, 55(3), 229-242. <https://doi.org/10.1007/s00374-019-01344-z>
- Zhang, M., Liang, W., Tu, Z., Li, R., Zhang, Z., Ali, A., & Xiao, R. (2021). Succession of bacterial community during composting: Dissimilarity between compost mixture and biochar additive. *Biochar*, 3(2), 229-237. <https://doi.org/10.1007/s42773-020-00078-8>
- Zhang, M., Zhang, L., Riaz, M., Xia, H., & Jiang, C. (2021). Biochar amendment improved fruit quality and soil properties and microbial communities at different depths in citrus production. *Journal of Cleaner Production*, 292, 126062. <https://doi.org/10.1016/j.jclepro.2021.126062>
- Zhou, Y.-H., Xia, K., & Wright, F. A. (2011). A powerful and flexible approach to the analysis of RNA sequence count data. *Bioinformatics*, 27(19), 2672-2678.
- Zrenner, R., Stitt, M., Sonnewald, U., & Boldt, R. (2006). PYRIMIDINE AND PURINE BIOSYNTHESIS AND DEGRADATION IN PLANTS. *Annual Review of Plant Biology*, 57(Volume 57, 2006), 805-836. <https://doi.org/10.1146/annurev.arplant.57.032905.105421>

## Apéndices

## Apéndice A. Total de familias identificadas taxonómicamente en QIIME2

Tratamiento biochar fase mesófila - DR-46		Tratamiento biochar fase termófila - DR-48	
Familia	Abundancia Relativa (%)	Familia	Abundancia Relativa (%)
Streptomycetaceae	12,989	Streptomycetaceae	8,433
Streptosporangiaceae	5,850	Paenibacillaceae	4,887
Thermoactinomycetaceae	4,864	Unclassified 2	4,001
Xanthomonadaceae	4,393	Unclassified 10	3,962
Paenibacillaceae	4,337	Rhizobiaceae	3,177
Mycobacteriaceae	3,340	Microbacteriaceae	2,965
Planococcaceae	3,093	Unclassified 5	2,942
Unclassified 10	2,835	Thermoactinomycetaceae	2,910
Phyllobacteriaceae	2,712	Cytophagaceae	2,856
Rhizobiaceae	2,488	Planococcaceae	2,738
Bacillaceae	2,454	Unclassified 7	2,722
Chitinophagaceae	2,454	FFCH7168	2,675
Actinosynnemataceae	2,331	Phyllobacteriaceae	2,503
Unclassified 5	2,208	Bacillaceae	2,440
Unclassified 2	2,163	Xanthomonadaceae	2,346
Unclassified 13	2,129	Chitinophagaceae	2,346
Nocardiodaceae	1,995	Pirellulaceae	2,048
Unclassified 7	1,849	Unclassified 23	1,961
Thermomonosporaceae	1,692	Mycobacteriaceae	1,898
Sphingobacteriaceae	1,625	Nocardiodaceae	1,851
Promicromonosporaceae	1,502	Polyangiaceae	1,569
C111	1,334	[Chthoniobacteraceae]	1,428
Unclassified 36	1,334	Streptosporangiaceae	1,404
Verrucomicrobiaceae	1,311	Unclassified 13	1,357
Pseudonocardiaceae	1,222	Unclassified 36	1,349
Alteromonadaceae	1,132	Unclassified 11	1,185
Cytophagaceae	1,098	Comamonadaceae	1,169
Comamonadaceae	1,053	Caldilineaceae	1,145
Haliangiaceae	0,975	Verrucomicrobiaceae	1,145
Pirellulaceae	0,964	Planctomycetaceae	1,043
Beijerinckiaceae	0,953	C111	1,028
FFCH7168	0,953	Nannocystaceae	1,028
Planctomycetaceae	0,807	Beijerinckiaceae	0,934
Alcaligenaceae	0,796	Promicromonosporaceae	0,910
[Chthoniobacteraceae]	0,796	Alteromonadaceae	0,902
Microbacteriaceae	0,751	Unclassified 9	0,863

Caldilineaceae	0,628	Rhodospirillaceae	0,832
Unclassified 9	0,594	Actinosynnemataceae	0,832
Unclassified 23	0,594	Isosphaeraceae	0,800
Polyangiaceae	0,594	Sphingobacteriaceae	0,706
Pseudomonadaceae	0,560	Solirubrobacteraceae	0,667
Rhodospirillaceae	0,527	Unclassified 14	0,659
Unclassified 11	0,516	Thermomonosporaceae	0,612
Brucellaceae	0,516	Alcaligenaceae	0,588
Methylobacteriaceae	0,504	Opitutaceae	0,541
Hyphomicrobiaceae	0,459	Unclassified 16	0,479
Flavobacteriaceae	0,459	Cystobacteraceae	0,471
Nannocystaceae	0,448	Methylobacteriaceae	0,471
Isosphaeraceae	0,437	Hyphomicrobiaceae	0,463
Erythrobacteraceae	0,403	Acetobacteraceae	0,447
Sphingomonadaceae	0,392	Brucellaceae	0,439
Rhodobacteraceae	0,392	Pseudomonadaceae	0,408
OM60	0,392	Unclassified 30	0,400
Unclassified 16	0,381	Sphingomonadaceae	0,392
Unclassified 26	0,370	Unclassified 17	0,384
Patulibacteraceae	0,359	Erythrobacteraceae	0,361
Acetobacteraceae	0,336	Unclassified 26	0,345
Opitutaceae	0,314	A4b	0,337
Nocardiopsaceae	0,303	OM60	0,337
Unclassified 1	0,291	Haliangiaceae	0,329
Unclassified 14	0,280	Unclassified 35	0,322
Unclassified 3	0,269	Flavobacteriaceae	0,306
Unclassified 30	0,247	Micromonosporaceae	0,306
Solirubrobacteraceae	0,235	Lachnospiraceae	0,298
Unclassified 35	0,224	Flammeovirgaceae	0,298
Cystobacteraceae	0,202	Unclassified 37	0,290
Unclassified 37	0,202	Unclassified 27	0,275
Aurantimonadaceae	0,179	Unclassified 3	0,259
Methylophilaceae	0,179	Rhodobacteraceae	0,251
Lachnospiraceae	0,168	Unclassified 22	0,251
A4b	0,168	Unclassified 1	0,243
Clostridiaceae	0,157	Armatimonadaceae	0,235
Unclassified 33	0,157	Unclassified 8	0,188
Enterobacteriaceae	0,146	Pseudonocardiaceae	0,188
Glycomycetaceae	0,134	Clostridiaceae	0,188
Actinomycetaceae	0,123	Methylophilaceae	0,180
Peptostreptococcaceae	0,112	[Weeksellaceae]	0,180
[Weeksellaceae]	0,112	Gemmataceae	0,180
Micromonosporaceae	0,101	Unclassified 39	0,173
Unclassified 40	0,101	Oxalobacteraceae	0,157

Cyclobacteriaceae	0,090	Aeromonadaceae	0,133
Symbiobacteriaceae	0,078	Enterobacteriaceae	0,126
Leuconostocaceae	0,078	Cyclobacteriaceae	0,126
Unclassified 22	0,078	Peptostreptococcaceae	0,102
Unclassified 27	0,078	Unclassified 12	0,102
Bdellovibrionaceae	0,067	Rhodocyclaceae	0,102
Rhodocyclaceae	0,067	Caulobacteraceae	0,102
Parachlamydiaceae	0,056	Aurantimonadaceae	0,094
Unclassified 34	0,056	Leuconostocaceae	0,094
Nocardiaceae	0,045	Iamiaceae	0,094
[Chromatiaceae]	0,045	Unclassified 38	0,094
Bacteriovoracaceae	0,045	Bradyrhizobiaceae	0,086
Unclassified 42	0,045	Nocardiopsaceae	0,086
Hahellaceae	0,034	Symbiobacteriaceae	0,078
Unclassified 6	0,022	Hahellaceae	0,078
Sinobacteraceae	0,022	[Chromatiaceae]	0,078
Trueperaceae	0,022	Conexibacteraceae	0,078
Phycisphaeraceae	0,022	Unclassified 31	0,078
Unclassified 4	0,000	Bdellovibrionaceae	0,063
MVS-65	0,000	Sinobacteraceae	0,063
Oxalobacteraceae	0,000	Unclassified 18	0,063
Unclassified 12	0,000	Unclassified 19	0,063
Unclassified 15	0,000	Unclassified 24	0,063
Bradyrhizobiaceae	0,000	Bacteriovoracaceae	0,063
Unclassified 17	0,000	Unclassified 42	0,063
Unclassified 8	0,000	Unclassified 4	0,055
Unclassified 18	0,000	[Pedosphaeraceae]	0,055
Moraxellaceae	0,000	Ellin6075	0,055
PAUC26f	0,000	Unclassified 29	0,039
Aeromonadaceae	0,000	Parachlamydiaceae	0,039
Unclassified 19	0,000	Unclassified 40	0,039
Unclassified 20	0,000	Unclassified 41	0,039
AKIW659	0,000	PAUC26f	0,031
[Pedosphaeraceae]	0,000	MVS-65	0,024
Legionellaceae	0,000	Unclassified 15	0,024
Neisseriaceae	0,000	Moraxellaceae	0,024
Unclassified 21	0,000	AKIW659	0,024
Unclassified 24	0,000	Neisseriaceae	0,024
Unclassified 25	0,000	Unclassified 21	0,024
Conexibacteraceae	0,000	Unclassified 25	0,024
Unclassified 28	0,000	Unclassified 32	0,024
Unclassified 29	0,000	Unclassified 34	0,024
Flammeovirgaceae	0,000	Phycisphaeraceae	0,024
Unclassified 31	0,000	Unclassified 20	0,016

Unclassified 32	0,000	Legionellaceae	0,016
Ellin6075	0,000	Unclassified 28	0,016
Iamiaceae	0,000	Patulibacteraceae	0,000
Caulobacteraceae	0,000	Unclassified 6	0,000
Unclassified 38	0,000	Nocardiaceae	0,000
Unclassified 39	0,000	Actinomycetaceae	0,000
Gemmataceae	0,000	Glycomycetaceae	0,000
Armatimonadaceae	0,000	Trueperaceae	0,000
Unclassified 41	0,000	Unclassified 33	0,000
<b>TOTAL</b>	<b>100</b>	<b>TOTAL</b>	<b>100</b>