

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

**Análisis de la composición de consorcios microbianos de la
Amazonía ecuatoriana con capacidad de remediación de aguas
residuales**

Donald Alexander Terán Espinosa

Ingeniería en Biotecnología

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero Biotecnólogo

Quito, 17 de diciembre de 2021

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA

**Análisis de la composición de consorcios microbianos de la Amazonía
ecuatoriana con capacidad de remediación de aguas residuales**

Donald Alexander Terán Espinosa

Nombre del profesor, Título académico

María de Lourdes Torres, PhD

Juan José Guadalupe, MSc

Quito, 17 de diciembre de 2021

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Donald Alexander Terán Espinosa

Código: 00200751

Cédula de identidad: 1718196619

Lugar y fecha: Quito, 17 de diciembre de 2021

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

RESUMEN

La creciente cantidad de aguas residuales generadas por actividades humanas representa un serio problema de salud para las personas y el medio ambiente. En la región Amazónica, se ha encontrado que los contaminantes superan entre cien y mil veces los límites máximos permitidos en los ríos. Además, esta región cuenta con un número limitado de plantas de tratamiento, por lo que un gran porcentaje de aguas que no son tratadas regresan a los ríos. El tratamiento de aguas residuales con consorcios alga-bacteria es una alternativa interesante para reducir los niveles de contaminación, ya que estos microorganismos tienen relaciones mutualistas que les permite tener una alta tasa de remoción de los contaminantes. A pesar de su alto potencial, la información de la estructura y composición de estos consorcios microbianos es escasa. Por lo tanto, en este proyecto se realizó la caracterización de consorcios microbianos provenientes de la Amazonía ecuatoriana con potencial de remediación de aguas residuales utilizando el secuenciamiento en MinION para el análisis de metagenomas. Con esto, se determinaron las especies dominantes, su interacción, el cambio del consorcio a lo largo del tiempo y las vías metabólicas asociadas. Mediante la clasificación taxonómica se encontró que los filos más abundantes son *Proteobacteria*, *Terrabacteria*, *FCB*, *Acidobacteria*, *Chlorophyta* y *Opisthokonta*, cuya alta eficiencia en la remoción de contaminantes de aguas residuales ha sido reportado en estudios previos. Se encontró una diversidad relativamente alta en base a los índices de alfa y beta diversidad, mientras que en el estudio de las vías metabólicas se encontraron vías de degradación de los contaminantes más comunes de las aguas residuales. En general, los resultados muestran que el consorcio analizado en este estudio podría ser útil para el tratamiento de las aguas residuales de la Amazonía, por lo que se establece una base importante para estudios futuros.

Palabras clave: agua residual, Amazonía, consorcio, MinION, metagenomas, diversidad.

ABSTRACT

The increase in the amount of wastewater generated by human activities represents a serious problem for the health of people and the environment. In the Amazon region, it has been found that contaminants in rivers surpass the maximum established limits by 100 to 1000 times. Additionally, this region has a limited number of conventional wastewater treatment facilities, which causes that a considerable amount of wastewater gets poured back into rivers without being treated. Wastewater treatment with algae-bacteria consortium is an interesting alternative to reduce the contamination levels, given that these microorganisms have mutualist relationships that allows them to have a high efficiency to remove the contaminants. Although the consortium has a great potential to treat wastewater, data related to its composition and structure is scarce. Therefore, this project aims to characterize the microbial consortium collected from a water sample in the Amazon region, which has the potential to treat wastewater. For this, nanopore sequencing with MinION was used for the analysis of the metagenome. This enabled us to determine the dominant species, their interaction, their variation over time and the metabolic routes. The taxonomic classification found that the most abundant phyla were *Proteobacteria*, *Terrabacteria*, *FCB*, *Acidobacteria*, *Chlorophyta* and *Opisthokonta*, which are highly efficient in removing wastewater contaminants, as has been reported in previous studies. A relatively high diversity was found using the alpha and beta diversity indexes. Additionally, degradation routes of the most common wastewater contaminants were also found. The results show that the consortium analyzed in this study could be useful for the treatment of wastewater generated in the Amazon region, which sets an important foundation for further studies.

Key words: wastewater, Amazon, MinION, metagenome, diversity, taxonomic classification, consortium, sequencing, degradation.

TABLA DE CONTENIDO

Introducción	11
Problemas ambientales de salud asociados a las aguas residuales.....	11
Limitaciones de sistemas de tratamiento convencionales.....	11
Tratamientos basados en consorcios alga-bacteria	12
Herramientas disponibles para caracterizar las comunidades microbianas	13
Métodos	15
Extracción de ADN con kit DNeasy Power Soil®	15
Cuantificación de ADN y electroforesis en gel de agarosa	15
Preparación de librerías y secuenciamiento	16
Análisis bioinformáticos	16
Clasificación taxonómica.....	17
Alfa diversidad.....	17
Beta diversidad	17
Resultados	18
Análisis de calidad del secuenciamiento	18
Análisis de clasificación taxonómica con Kaiju	18
Alfa diversidad.....	19
Beta diversidad	20
Vías metabólicas.....	21
Discusión.....	22
Conclusiones.....	26
Tablas.....	27
Figuras	29
Referencias bibliográficas.....	32
Anexos.....	40

ÍNDICE DE TABLAS

Tabla 1: Calidad del secuenciamiento	27
Tabla 2: Lecturas clasificadas y no clasificadas	27
Tabla 3: Índices de Shannon-Wiener y Simpson para el análisis de alfa diversidad	28
Tabla 4: Índice de Bray-Curtis para el análisis de beta diversidad.....	28

ÍNDICE DE FIGURAS

Figura 1: Comparación de los porcentajes de clasificación a nivel de dominio	29
Figura 2: Clasificación taxonómica a nivel de filo.....	29
Figura 3: Clasificación taxonómica a nivel de clase	30
Figura 4: Mapa de calor para el análisis de la beta diversidad.....	30
Figura 5: Comparación de vías metabólicas encontradas en cada Time Point	31

ÍNDICE DE ANEXOS

Anexo 1: Lexturas clasificadas a nivel de dominio.....	40
Anexo 2: Lecturas clasificadas a nivel de filo	41
Anexo 3: Lecturas clasificadas a nivel de clase.....	42
Anexo 4: Lecturas generadas para turas metabólicas	43

INTRODUCCIÓN

Problemas ambientales y de salud asociados a las aguas residuales

La elevada cantidad de residuos generados por las actividades humanas está directamente relacionada con el aumento desmesurado de la población en las últimas décadas (Barreiro, 2019). Muchos de estos residuos, que en su mayoría provienen de zonas urbanas e industriales, son vertidos en los cuerpos de agua sin recibir un tratamiento previo. Esto genera aguas residuales con un alto nivel de contaminación que contienen concentraciones elevadas de materia orgánica, nutrientes, metales pesados, sales y microorganismos patógenos, lo que representa un problema grave para la salud humana y del ecosistema (Guldhe et al., 2017; WWAP, 2017). Si la población sigue en aumento y los niveles de contaminación persisten, se estima que para el año 2025 el 60% de la población global no podrá acceder de forma sencilla a una fuente de agua segura (Cuellar-Bermúdez et al., 2017). De forma similar, el vertido de aguas contaminadas al medio ambiente puede ocasionar una pérdida permanente de la biodiversidad. Esto se debe a que las aguas residuales contienen altas cantidades de nutrientes, como fósforo y nitrógeno, que son usados por la flora acuática para su proliferación. La acumulación excesiva de biomasa proveniente de esta flora causa que el oxígeno del medio se reduzca considerablemente, por lo que no podrá ser utilizado por otros organismos y causará la degradación del ecosistema en procesos como la eutrofización (Renuka et al., 2013).

Limitaciones de sistemas de tratamiento convencionales

A pesar de los problemas mencionados anteriormente, existe poco interés en implementar sistemas para tratar las aguas residuales. Un estudio realizado en 2015 estimó que el 62% de las aguas residuales generadas en países de ingresos medio-altos no eran tratadas, mientras que el porcentaje en países en vías de desarrollo oscilaba entre

el 72% y el 92% (Sato et al., 2013). Este notable desinterés se debe principalmente a los altos costos asociados a los sistemas de tratamiento convencionales, que en su mayoría están basados en la tecnología de fangos activos. Los costos de esta tecnología están asociados al oxígeno que se debe suministrar a las bacterias aerobias que se utilizan en el proceso, cuya función es degradar los contaminantes de las aguas residuales (Barreiro, 2019). La energía necesaria para suministrar oxígeno puede representar hasta el 75% del total del costo energético de una planta de tratamiento (Cuellar Bermúdez et al., 2017; Eurostat-EU, 2021). Por otro lado, la instalación de la infraestructura necesaria para esta tecnología no resulta rentable para poblaciones menores a quince mil personas, ya que los costos aumentan exponencialmente cuanto menor es la población y menor es el tamaño de la planta. Como consecuencia, las aguas residuales generadas por estas poblaciones pequeñas son vertidas al medio ambiente sin recibir tratamiento. Adicionalmente, las plantas de tratamiento basadas en la tecnología de fangos activos pueden llegar a generar gases de efecto invernadero (GEI) cuyas emisiones suponen el 0.4% de las emisiones globales (Barreiro, 2019; Eurostat-EU, 2021).

Tratamientos basados en consorcios alga-bacteria

Considerando las limitaciones de la tecnología convencional disponible, ha surgido un gran interés en desarrollar tecnologías alternativas eficientes para el tratamiento de aguas residuales. Dentro de estas alternativas prometedoras, destaca el sistema de tratamiento basado en consorcios alga-bacteria, ya que se reducen los costos energéticos, disminuyen las emisiones de GEI, y se genera biomasa que puede ser utilizada para producir biocombustible o biofertilizante (Barreiro, 2019; Guldhe et al., 2017; Pires et al., 2013; Suganya et al., 2016). Estas ventajas se deben principalmente a las relaciones mutualistas y simbióticas entre las microalgas y las bacterias. En los consorcios, las microalgas producen el oxígeno requerido por las bacterias para

metabolizar los contaminantes de las aguas residuales, lo que reduce los costos asociados a la aireación (Tiron et al., 2015). A su vez, las bacterias producen el dióxido de carbono que usan las microalgas como fuente de carbono para sintetizar biomasa, lo que elimina la contaminación de las aguas, fija el CO_2 producido y reduce las emisiones de GEI. Por otro lado, el agua residual resulta un medio de cultivo potencialmente útil para los consorcios alga-bacteria, ya que cuentan con altos niveles de nitrógeno y fósforo que son usados por los microorganismos para su desarrollo (Barreiro, 2019).

Herramientas disponibles para caracterizar las comunidades microbianas

A pesar de las ventajas mencionadas, hay una falta de información considerable con respecto a la composición de los consorcios microbianos y las rutas metabólicas asociadas en el tratamiento de aguas residuales, lo que resulta en un limitante para la implementación a gran escala de esta tecnología. Resultan escasos los estudios que determinan las especies dominantes en los consorcios, las interacciones entre los microorganismos y los genes funcionales asociados (Barreiro, 2019). Esto se debe a que el 99% de los microorganismos no pueden ser cultivados en el laboratorio, lo que significa que nuestro conocimiento sobre su diversidad está limitado al 1% del total de microorganismos del ambiente (Awasthi et al., 2020). Esto representa un obstáculo considerable, ya que los microorganismos no cultivables tienen un alto potencial de ser utilizados en diferentes ramas de la industria, como en el tratamiento de aguas residuales. No obstante, los avances en la tecnología de la metagenómica posibilitan el estudio de estos microorganismos no cultivables, ya que no se requiere el aislamiento ni el cultivo de las comunidades microbianas para analizar su material genético (Bharagava et al., 2019). La metagenómica es el estudio de todo el material genético que se encuentra en una muestra ambiental y que posibilita caracterizar las comunidades microbianas y estudiar los genomas de los microorganismos (Techtmann & Hazen, 2016). Por

consiguiente, se pueden identificar nuevas especies, estudiar la degradación microbiana de contaminantes y determinar los genes responsables de dicha degradación (Uhlik et al., 2013).

Para obtener la información de los microorganismos presentes en el consorcio se utiliza la tecnología de secuenciación de última generación (NGS, por sus siglas en inglés), que forma parte del flujo de trabajo de los estudios metagenómicos (Bharagava et al., 2019). Dentro de la tecnología NGS destaca el secuenciación con nanoporos, que fue desarrollado por Oxford Nanopore Technologies Ltd. y cuya tecnología se basa en la medición de los cambios en la corriente iónica que generan las moléculas de ADN o ARN conforme pasan por un nanoporo incorporado en una membrana. Esta señal se analiza en tiempo real y se determina la secuencia de bases en las hebras de material genético que pasan por los nanoporos (Goodwin et al., 2015). Se han desarrollado varios instrumentos de secuenciación que utilizan la tecnología de nanoporos, entre los que se incluye al MinION, el cual es una herramienta portátil capaz de secuenciar fragmentos extremadamente largos de material genético (mayores a 10 kb) sin comprometer la calidad de la secuencia (Goodwin et al., 2015; Loman et al., 2015).

Por esta razón, el presente estudio se enfoca en analizar la composición de consorcios microbianos provenientes de la Amazonía ecuatoriana, considerando que esta región presenta altos niveles de contaminación de nitrógeno, fósforo, materia orgánica y metales pesados, causados por el vertido de aguas contaminadas sin tratar a las fuentes de agua. De forma concreta, con esta investigación se pretende clasificar taxonómicamente a los microorganismos presentes en una muestra proveniente de la Amazonía, analizar la variación del consorcio microbiano a lo largo del tiempo y determinar los genes funcionales con capacidad de remediación de aguas residuales utilizando la tecnología de secuenciación del MinION y el análisis con diferentes herramientas bioinformáticas.

MÉTODOS

Para caracterizar la comunidad del consorcio microbiano proveniente de la Amazonía y evaluar cómo esta comunidad interactúa y cambia a lo largo del tiempo, se establecieron diferentes muestras para el análisis: las muestras tratamiento y Dark Control (DC). Los tratamientos fueron cultivados bajo condiciones de luz y oscuridad en periodos de 12 horas, mientras que los DC fueron cultivados únicamente bajo condiciones de oscuridad. Con el DC, se buscó determinar si las algas influyen en la capacidad de remoción de contaminantes del consorcio. Además, se establecieron los Time Points (TP) en los días 0, 9 y 12 para evaluar las muestras a lo largo del tiempo. Se seleccionaron específicamente estos días ya que el consorcio mostró porcentajes de remoción altos en estos días.

Extracción de ADN con kit DNeasy Power Soil®

Se recolectaron 15 mL de las muestras del tratamiento y DC de cada uno de los TP establecidos de los consorcios. El TP 0 fue el único que no tuvo un DC, ya que no es posible diferenciar las muestras tratamiento y DC en este punto del tiempo. El volumen recolectado de la muestra fue descongelado y centrifugado a 9 mil rpm por 8 minutos para concentrar la biomasa del consorcio. Luego, se realizó la lisis mecánica usando nitrógeno líquido y se continuó con la extracción de ADN usando el kit DNeasy PowerSoil® de Qiagen (2016), siguiendo las instrucciones del fabricante. Se adoptaron dos modificaciones al protocolo establecido de Qiagen, al cambiar el vortex de 10 minutos por uno de 30 minutos, y reemplazando los 100 μ l de la solución C6 con 30 μ l de H_2O PCR.

Cuantificación de ADN y electroforesis en gel de agarosa

La calidad y concentración del ADN de las muestras de cada TP se analizaron con el espectrofotómetro NanoDrop 2000 de Thermo Fisher Scientific y con el fluorómetro Qubit™ 3.0 de Invitrogen (Thermo Fisher Scientific, 2009; Thermo Fisher Scientific,

2014). Para el análisis en NanoDrop se utilizaron 2 μ l de ADN, mientras que con el Qubit™ se siguieron las instrucciones del fabricante. Las muestras de ADN también fueron evaluadas mediante electroforesis en gel de agarosa al 1.5% (30 minutos a 100 voltios).

Preparación de librerías y Secuenciamiento

Las muestras de ADN con un índice de calidad 260/280 entre 1.8-2.06 y concentraciones entre 22.4-91.5 ng/ μ l fueron procesadas para la preparación de las librerías usando el kit Native Barcoding Genomic DNA de Oxford Nanopore Technologies, siguiendo las instrucciones proporcionadas por el fabricante (ONT, 2019). Se realizaron lavados de las muestras de ADN usando beads magnéticas AMPure XP en relación 1.8X en el primer lavado para purificar fragmentos con tamaños variados, y 1X para los lavados posteriores (Beckman Coulter, 2021). Para el paso de “DNA repair and end-prep” se modificó el protocolo para que se incluya un solo lavado con etanol al 70%. Se utilizaron 700 ng de librería en 65 μ l de H_2O libre de nucleasas para el pool de muestras que fue cargado en FLO-MIN106 Flow Cells siguiendo las instrucciones del fabricante y se obtuvieron los resultados en archivos con formato FASTQ (Oxford Nanopore Technologies, 2019).

Análisis Bioinformáticos

Se utilizó la herramienta Porechop para realizar el demultiplexing de las secuencias generadas, en donde se retiran los barcodes y adaptadores colocados en la preparación de librerías. Para esto, se instaló Porechop en la terminal Linux y se subieron los archivos fastq_pass derivados del secuenciamiento con MinION (GitHub, 2018). También se utilizó la herramienta Nanoplot para el análisis de calidad del secuenciamiento con nanoporos de todos los Time Points. Se obtuvo información respecto a la longitud media de las lecturas, la calidad media en base al Phred score, y al estadístico Q7 (GitHub, 2020). Por otro lado, se utilizó la herramienta Flye para el ensamblaje de los

metagenomas. Para esta herramienta, se utilizó la siguiente configuración en la terminal de Linux: *--nanoraw, --meta, --threads 5* (Kolmogorov et al., 2019). Finalmente, se instaló la herramienta SqueezeMeta en Linux para identificar las vías metabólicas presentes en el metagenoma (Tamames & Puente-Sánchez, 2019). Se corrió el programa con la siguiente configuración: *-m coassembly, -extassembly, --euk*.

Clasificación taxonómica

Para todos los TP, se analizó el perfil taxonómico utilizando el *web server* del programa Kaiju (Menzel et al., 2016). Para esto, se comprimieron los archivos .fastq provenientes del Porechop (GitHub, 2018) a formato .zip y se subieron al servidor web de Kaiju. La base de datos usada fue “NCBI BLAST nr+euk” y se seleccionó la opción “Greedy” para no descartar las secuencias con mismatches en exceso. Se obtuvieron resultados en forma de un Krona Chart con respecto al número de lecturas, porcentaje de clasificación y la diversidad de microorganismos presentes (Menzel et al., 2016).

Alfa diversidad

Se utilizó el paquete Vegan en R (R Core Team, 2019) para obtener información de la alfa diversidad, que es el análisis de la diversidad de especies en un sitio a escala local. En este caso, se analizó la diversidad dentro de cada TP y se utilizaron los índices de Shannon y Simpson para obtener valores que indicaron una alta o baja diversidad (Oksanen et al., 2019).

Beta diversidad

Se instaló el programa Simka (Benoit et al., 2016) en la terminal Linux para obtener información de la beta diversidad, que es el análisis de la variación de especies entre diferentes consorcios y comparar los resultados entre los distintos TP (Benítez et al., 2018). Se obtuvieron resultados en forma de tablas en formato .csv y mapas de calor (heatmaps).

RESULTADOS

Análisis de calidad del secuenciamiento

Se usó la información generada en Nanoplot para determinar la calidad del secuenciamiento de los Time Points 0, 9 y 12. La herramienta Nanoplot generó un archivo .html con información respecto a la longitud y número de lecturas, su calidad y al estadístico Q7, que permiten determinar la calidad general del secuenciamiento. El valor de la calidad media se interpretó con el Phred Quality Score, que tiene un rango de valores entre 4 y 60 y define la probabilidad de que una base específica sea llamada correctamente (Liao, Satten & Juan, 2017). Por otro lado, el estadístico Q7 indica la probabilidad de que una base haya sido llamada incorrectamente por el secuenciador (Illumina, 2011). Como se observa en la Tabla 1, se obtuvo un número de lecturas entre 112,841 y 638,996, donde se obtuvieron menos lecturas para el tratamiento y DC del TP 12. Se encontró que la longitud media de las lecturas estuvo entre 1,914.2 y 2,855.9 bp, la calidad media entre 10.7 y 11, y el estadístico Q7 cercano al 100% para los TP 0, 9 y 12.

Análisis de clasificación taxonómica con Kaiju

Una vez que se cargaron los archivos .fastq del Porechop en el servidor web de Kaiju, se obtuvieron resultados de cada Time Point sobre el número de lecturas, el porcentaje de clasificación y la abundancia taxonómica. Esta información se presentó gráficamente en forma de un Krona Chart, de donde se extrajo la información que se presenta en la Tabla 1, y en las Figuras y Anexos 1, 2 y 3. El valor del porcentaje se calculó en base al total de lecturas clasificadas de todos los organismos celulares de la muestra.

En la Tabla 2 se puede ver que el porcentaje de lecturas que no fueron clasificadas fue mayor al 50% en todos los Time Points. El TP 0 fue el que obtuvo más lecturas totales, con 638,996 lecturas, mientras que el tratamiento del TP 12 fue el que obtuvo el mayor porcentaje de clasificación de las lecturas generadas, con un 50% de clasificación. Por

otro lado, el Anexo 1 muestra la comparación de la abundancia del dominio bacteria y eucariota en cada uno de los TP. Se observa una tendencia marcada de la mayor abundancia de las bacterias sobre las eucariotas en los tratamientos y DC de todos los TP. El Anexo 2 muestra el número de lecturas y el porcentaje relativo encontrado para los microorganismos clasificados por filo, según el sistema jerárquico linneano. De forma general, se encontró que los filos más abundantes que se presentan en todos los TP fueron *Proteobacteria*, grupo *FCB*, *Terrabacteria*, *PVC*, *Acidobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Chlorophyta* y *Opisthokonta*. Finalmente, el Anexo 3 también muestra información sobre el número de lecturas y el porcentaje relativo para los microorganismos encontrados, pero esta vez clasificados por clase de acuerdo con el sistema linneano de clasificación taxonómica. Se observa un dominio marcado por las α -*Proteobacterias*, con un porcentaje entre el 31% y 51% en relación con el total de organismos celulares clasificados.

Alfa diversidad

Se utilizaron los índices de Shannon y Simpson para analizar la alfa diversidad del consorcio en los diferentes TP y determinar la diversidad de especies a nivel local. Estos índices reflejan la heterogeneidad de una comunidad en base a dos factores: el número de especies presentes y su abundancia relativa (Allen, Kon & Bar-Yam, 2009; He & Hu, 2005). Normalmente, los valores del índice de Shannon oscilan entre 0 y 3.4, donde los valores cercanos al 0 indican menor diversidad, mientras que los valores cercanos al 3.4 indican mayor diversidad (Allen, Kon & Bar-Yam, 2009). Por otro lado, el índice de Simpson regresa valores que oscilan entre 0 y 1, donde los valores cercanos al 0 indican menor diversidad y los valores cercanos al 1 indican mayor diversidad (He & Hu, 2005). La Tabla 3 muestra los valores obtenidos para los tratamientos y DC de todos los TP analizados. El tratamiento del TP 12 fue el que obtuvo valores de los índices asociados a

una mayor diversidad de especies presentes en el consorcio, con valores de 1.413 y 0.807 para los índices de Shannon y Simpson, respectivamente. Por otro lado, las muestras tratamiento y DC de los otros TP obtuvieron valores de los índices entre 1.336-1.395 y 0.680-0.709.

Beta diversidad

Se utilizó el índice de Bray-Curtis para analizar la beta diversidad entre los TP 0, 9 y 12. Este índice se utiliza como una medida de la diferencia entre la abundancia de especies presentes entre las muestras analizadas. Los valores de este índice oscilan entre 0 y 1, donde los valores cercanos a 0 indican similitud entre las muestras, mientras que valores cercanos al 1 indican una mayor divergencia (Ricotta & Podani, 2017). Como se observa en la Tabla 3, los resultados del índice Bray-Curtis indican que la diversidad de las muestras del TP 0 y el DC del TP 9 son bastante similares, ya que obtuvieron un valor de 0.595. Por otro lado, se observa que la diversidad entre el TP 0 y el tratamiento del TP 12 es la que presenta mayor diferencia, ya que se obtuvo un valor del índice de 0.979. También se encuentra una diferencia considerable entre el Time Point 0 y el Dark Control del Time Point 12, ya que se obtuvo un valor del índice Bray-Curtis de 0.963. Por otro lado, la Figura 4 muestra un heatmap que compara la abundancia entre los consorcios de los Time Points analizados. En este heatmap, los recuadros con colores amarillos y verdes indican mayor divergencia, mientras que los recuadros con tonos de color rojo y marrón indican mayor similitud entre las muestras. Como se observa en la Figura 4, los resultados del heatmap coinciden con los resultados de la Tabla 3, donde se encuentra una mayor divergencia entre el Time Point 0 y el tratamiento del Time Point 12.

Vías metabólicas

La información sobre las vías metabólicas encontradas en el consorcio para la degradación y metabolismo de contaminantes se obtuvo mediante la herramienta SqueezeMeta. La ejecución exitosa de este programa en la terminal Linux dio como resultado archivos .csv que muestran las vías metabólicas encontradas en el consorcio y el número de lecturas generadas que corresponden a cada una de las vías metabólicas. Esta información se muestra en el Anexo 4 y en la Figura 5, donde se compara el número de lecturas de las vías metabólicas entre los diferentes Time Points. Se observa que hay una mayor cantidad de lecturas correspondientes al metabolismo de butanoato, propanoato, nitrógeno, metano, y a la degradación del benzoato.

DISCUSIÓN

Para el análisis de la clasificación taxonómica en los TP establecidos, se utilizó la información obtenida del secuenciamiento basado en nanoporos. Esta técnica permite realizar una caracterización más detallada de las comunidades microbianas en comparación con las técnicas tradicionales dependientes de cultivo, que solo permiten identificar una pequeña fracción de la totalidad de microorganismos presentes (Bharagava et al., 2019; Datta et al., 2020). Los resultados del secuenciamiento obtenidos para este estudio demuestran ser de buena calidad, en base al análisis del Phred Score para las lecturas generadas, que fue de 10.9 en promedio para todos los TP (Tabla 1), lo que es similar a estudios que utilizan lecturas generadas en MinION (Shin et al., 2018).

Por otro lado, la asignación taxonómica para cada TP se obtuvo mediante el servidor web del programa Kaiju, que tiene la ventaja de regresar porcentajes de clasificación taxonómicos más altos que otros programas disponibles (Corvelo et al., 2018). A pesar de esta ventaja, el porcentaje de clasificación taxonómica en cada TP fue bajo, con un promedio de 39.4% de lecturas clasificadas (Tabla 2). Esto pudo ser causado por la alta tasa de error del MinION del 12%, la enorme diversidad microbiana que existe en los consorcios provenientes de la Amazonía, o debido a que las bases de datos para microalgas no son completas y existe una gran cantidad de “dark matter” en estudios metagenómicos (Goodwin et al., 2015; Ordóñez, 2020). En base a las lecturas clasificadas, se encontró que los filos más abundantes fueron *Proteobacteria*, *Terrabacteria*, grupo *FCB*, *Chlorophyta*, *Cyanobacteria* y *Opisthokonta* (Figura 2, Anexo 2), en donde *Proteobacteria* y grupo *FCB* destacaron por su mayor abundancia en las muestras tratamiento y DC de todos los TP. Esto concuerda con los resultados reportados en estudios previos realizados en la Amazonía y Galápagos, donde se muestra a *Chlorophyta* como el filo más dominante de microalgas por su eficiencia en la

asimilación de nitrógeno y fósforo, que son abundantes en las aguas residuales (Ordóñez, 2020; Posso, 2019). *Proteobacteria* y *Bacteroidetes* tuvieron una abundancia constante en todos los TP, esto se debe a su relación simbiótica que tienen con las microalgas, ya que promueven su crecimiento a través de la absorción eficiente de nutrientes del medio (Krustok, 2016; Posso, 2019). Por otro lado, se observa que la abundancia de *Chlorophyta* varía en las muestras tratamiento de los TP 0, 9 y 12, y también se observa una abundancia considerable de este filo en el DC del TP 9, algo que no se esperaba debido a que estos organismos necesitan luz para su desarrollo, y las condiciones de cultivo de estas muestras fueron en ausencia de luz. Las posibles explicaciones para esto son: una fuente de luz no planificada llegó a las muestras DC, se clasificaron las lecturas de microalgas que ya no estaban vivas, o que las algas presentaron un crecimiento heterótrofo en ausencia de luz (Morales-Sánchez et al., 2014).

También se encontró que los filos abundantes reportados en este estudio coinciden con los resultados de estudios previos de clasificación taxonómica realizados en plantas de tratamiento de agua residual en Italia (Tandoi et al., 2018). Una posible explicación para esta semejanza es que el sitio de recolección de las muestras para este estudio haya estado contaminado con aguas residuales (Posso, 2019). En los estudios previos, se encontró que los filos más abundantes fueron *Actinobacteria*, *Proteobacteria*, *Firmicutes* y *Bacteroidetes*. Este descubrimiento resulta importante debido a la alta eficiencia de remoción que se reporta para estos consorcios, por lo que la similitud encontrada con el consorcio analizado en este estudio puede significar que tiene un alto potencial para el tratamiento de aguas residuales debido a que estarían aclimatados a desarrollarse en ambientes con altos niveles de contaminación.

La alfa diversidad se analizó con los índices de Shannon-Wiener y Simpson, que son ampliamente utilizados en estudios ecológicos por tomar en cuenta la riqueza,

abundancia y uniformidad de las especies presentes en la muestra (Brraich & Kaur, 2015). En base al índice de Shannon-Wiener, se obtuvieron valores entre 1.336 y 1.413 para los TP analizados, mientras que los valores en base al índice de Simpson oscilaban entre 0.680 y 0.807, como se observa en la Tabla 3. Cabe mencionar que el índice de Shannon-Wiener comúnmente otorga más importancia a las lecturas de especies poco comunes, mientras que el índice Simpson da más importancia a las especies más abundantes. Esto puede generar un cierto nivel de sesgo en los resultados, sin embargo, estos concuerdan con los resultados reportados en literatura para ambos índices en estudios realizados en la Amazonía (Miura et al., 2007; Posso, 2019). Con respecto a la beta diversidad, se utilizó el índice de Bray-Curtis para obtener información sobre la diferencia en la abundancia de especies entre dos muestras distintas. Este índice es muy utilizado en estudios ambientales debido a que las lecturas de especies comunes y poco comunes obtienen una importancia similar en el cálculo de los valores del índice, por lo que se genera un mínimo sesgo (Demircan et al., 2018). En base a este índice, se obtuvieron resultados para todos los Time Points con un rango entre 0.595 y 0.979, como se observa en la Tabla 4, los cuales son similares a los valores reportados en estudios de diversidad con muestras ecológicas (Xu et al., 2014).

En base a los resultados obtenidos para los índices de alfa y beta diversidad, se encontró una diversidad relativamente alta de especies microbianas que se mantiene constante en cada TP analizado, lo que demuestra que los microorganismos pertenecientes a los filos más abundantes pueden tener actividades esenciales en el proceso de tratamiento de aguas residuales (Qiang et al., 2017). Sin embargo, no es posible asumir que una alta diversidad de especies microbianas esté directamente relacionada con un mayor rendimiento en la remoción de contaminantes, ya que se ha encontrado que la diversidad y el rendimiento no siempre van de la mano. Múltiples

estudios han reportado que el dominio de pocas especies en el consorcio puede presentar mayores ventajas para remover los contaminantes del agua residual (Ibarbalz et al., 2013; Saavedra et al., 2016). Por otro lado, se encontró que las muestras que presentaron una mayor diferencia en cuanto a la abundancia de especies fueron entre las muestras tratamiento del TP 0 y TP 12, con un valor de 0.979 para el índice de Bray-Curtis. Esto evidencia la variación del consorcio a lo largo del tiempo, que puede deberse a la disponibilidad de sustratos disponibles, lo que causa la dominancia de microorganismos que se encuentran más adaptados a desarrollarse en ese ambiente (Morales-Sánchez et al., 2014).

Finalmente, se obtuvo información sobre las vías metabólicas presentes en el consorcio mediante el programa SqueezeMeta usando la base de datos KEGG. Los resultados obtenidos muestran rutas metabólicas que están involucradas con la degradación o metabolismo de contaminantes que se encuentran comúnmente en las aguas residuales, como nitrógeno, fósforo, nitrato, compuestos xenobióticos, hidrocarburos, benzoato, nitrotolueno y estireno. Estos resultados concuerdan con los reportados en literatura para estudios realizados en cuerpos de agua contaminada, lo que demuestra una vez más el alto potencial que tiene el consorcio analizado para remover eficientemente los contaminantes de aguas residuales (Yadav et al., 2021). A pesar de esto, se observa que el número de lecturas es similar en todas las muestras tratamiento y DC de todos los TP (Figura 5). Incluso se observa que en el tratamiento del TP 12 no se encontraron rutas metabólicas que sí se encontraron en otros TP. Esto se debe a que el método que se usó para analizar los perfiles funcionales se basa en el mapeo de genes codificantes para proteínas a una colección de pathways de referencia, como la base de datos KEGG. Se ha reportado que este método puede generar “agujeros” en los pathways reconstruidos debido a información incompleta de los genes (Ye & Doak, 2009).

CONCLUSIONES

En conclusión, se encontró que el consorcio analizado podría ser útil para remover eficientemente los contaminantes de las aguas residuales de la Amazonía debido a que se ha reportado en otros estudios que los organismos encontrados son eficientes en la remoción de estos contaminantes, y además se encontró que estos microorganismos tuvieron una presencia abundante y constante en todos los Time Points, lo que demuestra sus actividades esenciales en la biorremediación. Estos hallazgos resultan relevantes para el tratamiento de aguas residuales en la región Amazónica, que cuenta con una baja cantidad de plantas de tratamiento de aguas residuales en el país (20% del total), un bajo porcentaje de tratamiento de las aguas que ingresan (68%), y un alto porcentaje de aguas no tratadas que son vertidas nuevamente a las fuentes de agua (77%) (INEC, 2016; Orbe, 2020). Además, los resultados resultan importantes para la biorremediación de las fuentes de agua de las que dependen las comunidades de la Amazonía, ya que se pueden implementar estos consorcios en sistemas de tratamiento efectivos. Por estas razones, el consorcio analizado podría resultar útil para desarrollar un sistema de tratamiento efectivo y de bajo costo, que contribuya a la limpieza de las aguas residuales de esta región. Sin embargo, se recomienda seguir investigando el consorcio debido al alto porcentaje de no clasificación que se tuvo en Kaiju, lo que demuestra la falta de información en las bases de datos; y a las muestras en las que no se obtuvieron lecturas para las vías metabólicas, lo que demuestra que se debe utilizar un método más efectivo para analizar estos datos. Es necesario trabajar con información más completa debido a que la reconstrucción de las vías metabólicas y la clasificación taxonómica son esenciales para entender las funciones biológicas de un genoma secuenciado. Finalmente, los resultados de este estudio son preliminares, por lo que se recomienda realizar análisis más profundos de la estructura y composición del consorcio.

TABLAS

Tabla 1.- Resultados en NanoPlor de la calidad del secuenciamiento por nanoporos para los Time Points 0, 9 y 12.

	TP0	TP9_trat	TP9_DC	TP12_trat	TP12_DC
Longitud media	2,855.9	1,776.3	2,450.4	1,940.8	1,914.2
Calidad media	11	10.9	10.8	10.7	10.7
Número de reads	638,996	571,821	591,931	112,841	126,676
Longitud N50	6,813	5,941	6,753	5,382.0	5,526
Número de bases	1.82×10^9	1.01×10^9	1.45×10^9	2.19×10^8	2.42×10^8
>Q7	100%	99.9%	99.9%	100%	100%

Descripción. Resultados obtenidos con la herramienta Nanoplot para determinar la calidad del secuenciamiento de todos los Time Points.

Tabla 2.- Porcentaje y número de lecturas clasificadas y no clasificadas obtenidas en el web server Kaiju.

		TP0	TP9_trat	TP9_DC	TP12_trat	TP12_DC
No clasificadas	Porcentaje	66%	67%	69%	50%	51%
	Lecturas	420,461	79,777	95,651	34,034	63,973
Clasificadas	Porcentaje	34%	33%	31%	50%	49%
	Lecturas	218,535	38,801	42,781	34,503	62,703
Total	Lecturas	638,996	119,578	138,432	68,537	126,676

Descripción. Número de lecturas generadas del secuenciamiento para cada Time Point.

El valor del porcentaje se calculó en base al total de lecturas generadas en el secuenciamiento. La información se obtuvo con el programa Kaiju, usando la base de datos NCBI BLAST nr+euk.

Tabla 3.- Índices de Shannon-Wiener y Simpson para el análisis de alfa diversidad.

	Shannon-Wiener (H')	Simpson (D)
TP0	1.336	0.680
TP9_trat	1.377	0.693
TP9_DC	1.395	0.709
TP12_trat	1.413	0.807
TP12_DC	1.387	0.701

Descripción. Resultados de los índices de Shannon-Wiener y Simpson para cada Time Point, obtenidos con el paquete Vegan en R.

Tabla 4.- Índice de Bray-Curtis para el análisis de beta diversidad.

	TP0	TP9_trat	TP9_DC	TP12_trat	TP12_DC
TP0	0	0.83	0.595	0.979	0.963
TP9_trat	0.83	0	0.78	0.96	0.918
TP9_DC	0.595	0.78	0	0.965	0.963
TP12_trat	0.979	0.96	0.965	0	0.853
TP12_DC	0.963	0.918	0.963	0.853	0

Descripción. Resultados del índice de Bray-Curtis comparando los tratamientos y Dark Control de todos los Time Points. Resultados obtenidos con la herramienta Simka.

FIGURAS

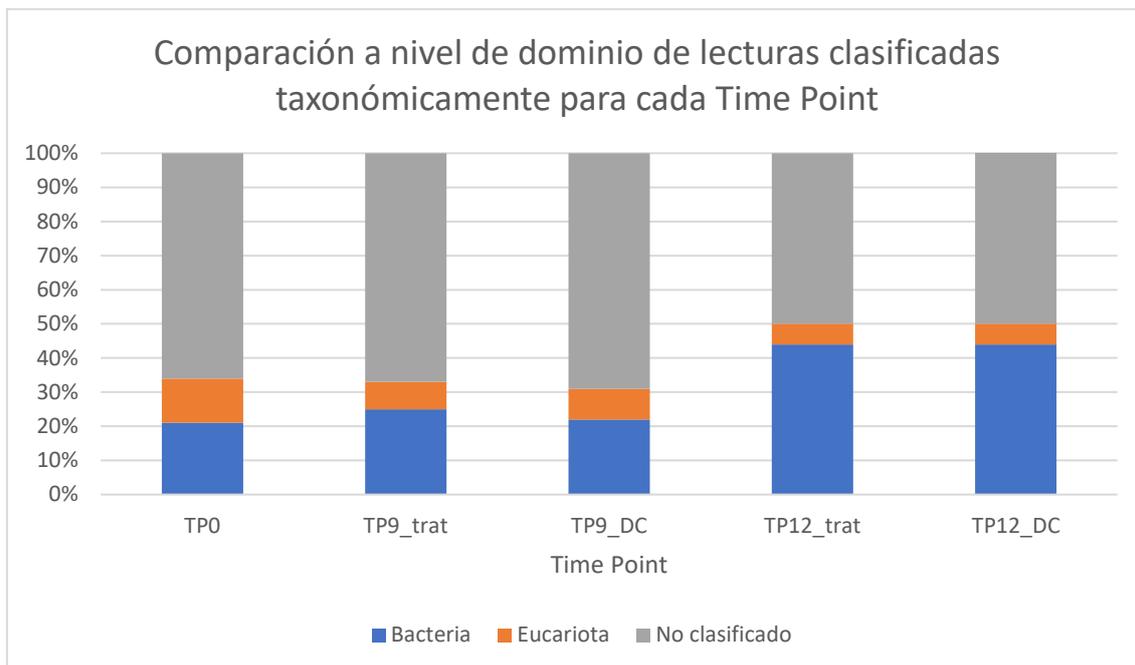


Figura 1.- Comparación a nivel de dominio de lecturas clasificadas taxonómicamente para cada Time Point.

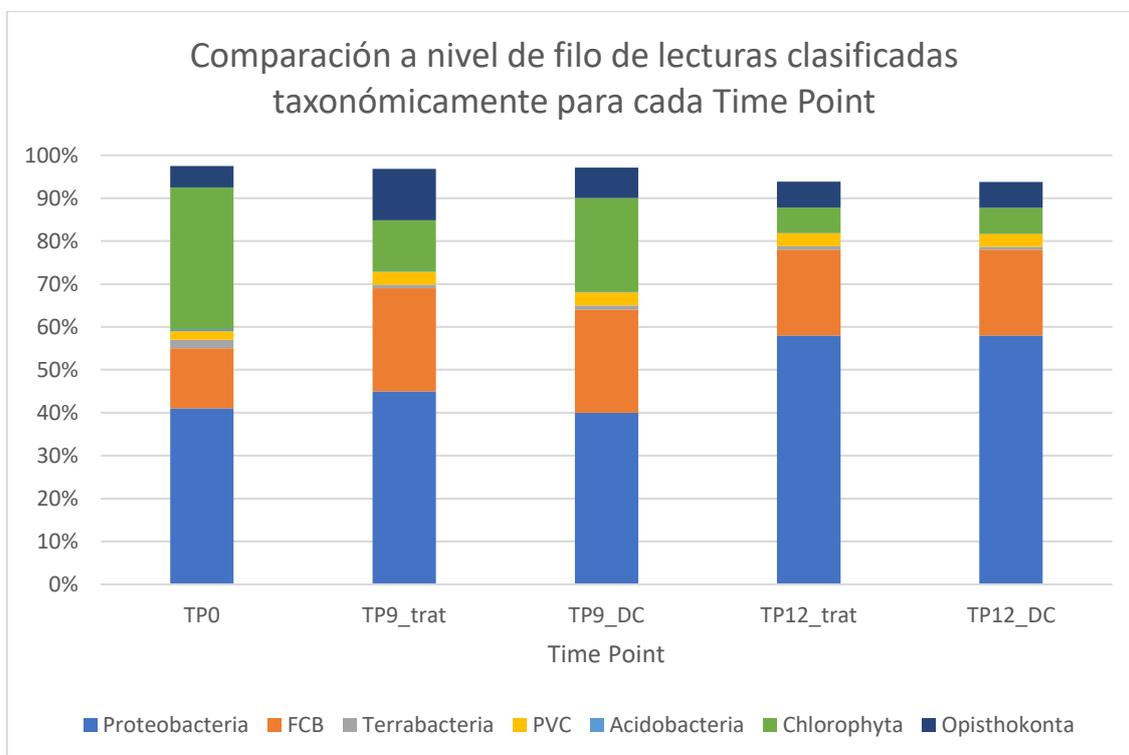


Figura 2.- Comparación a nivel de filo de lecturas clasificadas taxonómicamente para cada Time Point.

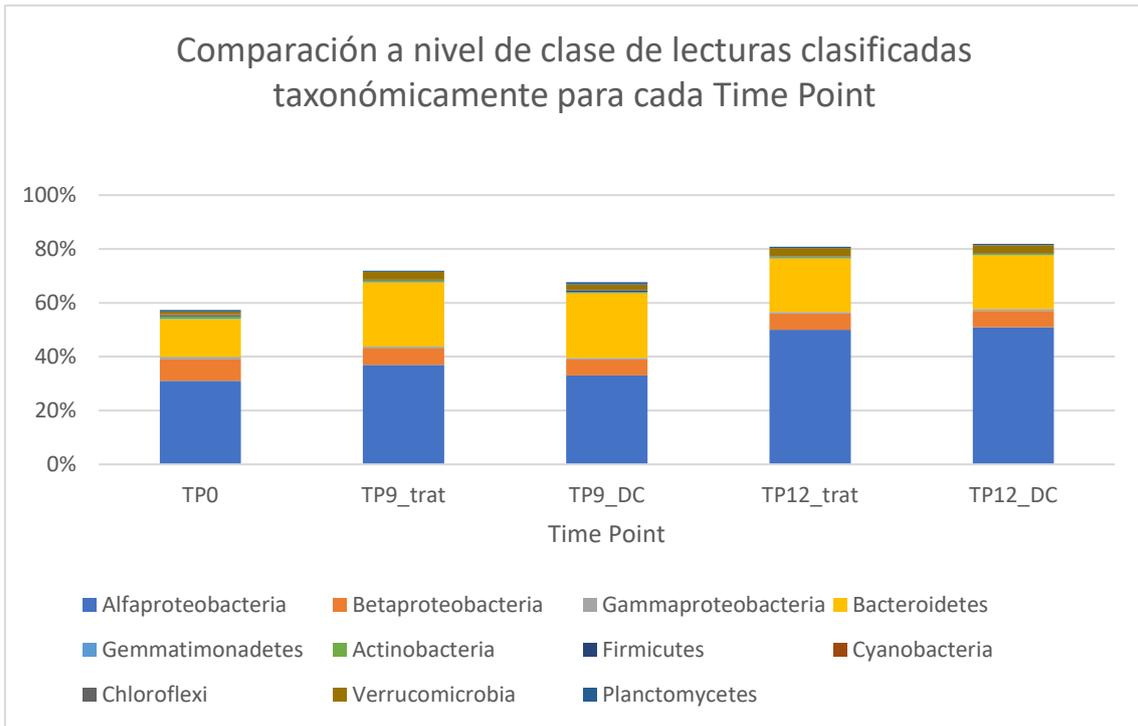


Figura 3.- Comparación a nivel de clase de lecturas clasificadas taxonómicamente para cada Time Point.

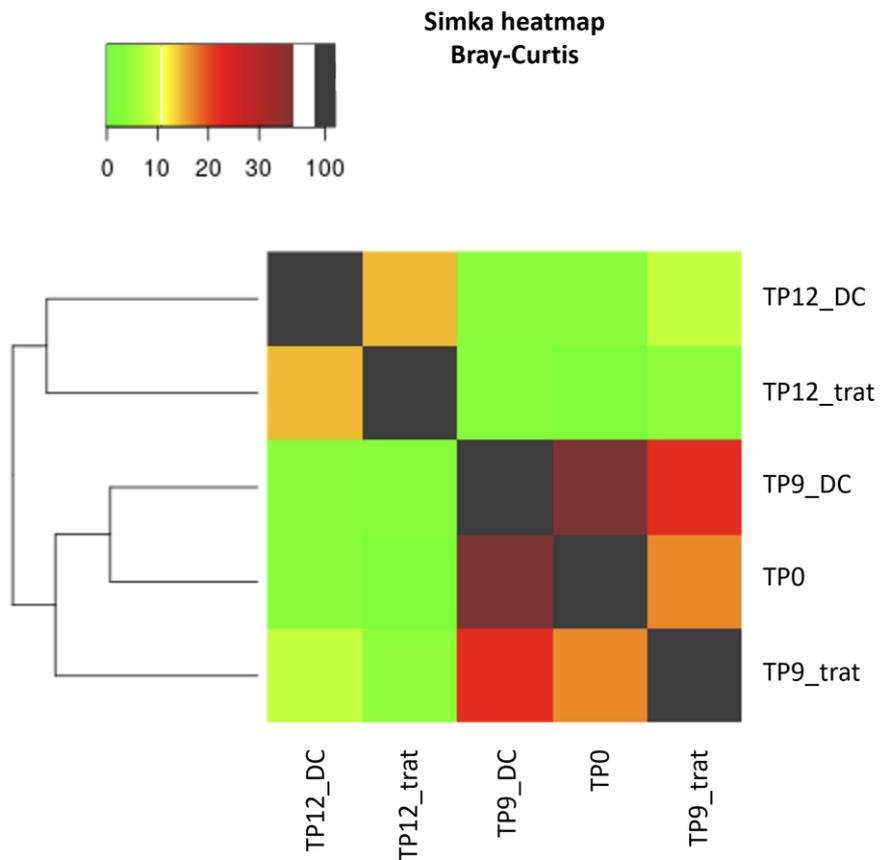


Figura 4.- Mapa de calor/heatmap para el análisis de la beta diversidad.

Descripción. Resultados de la beta diversidad entre los Time Points analizados mediante un mapa de calor. Los tonos verde-amarillo indican una mayor divergencia, mientras que los tonos rojo-marrón indican una mayor similitud entre las muestras analizadas. Los cuadrados de color negro corresponden a la comparación entre las mismas muestras.

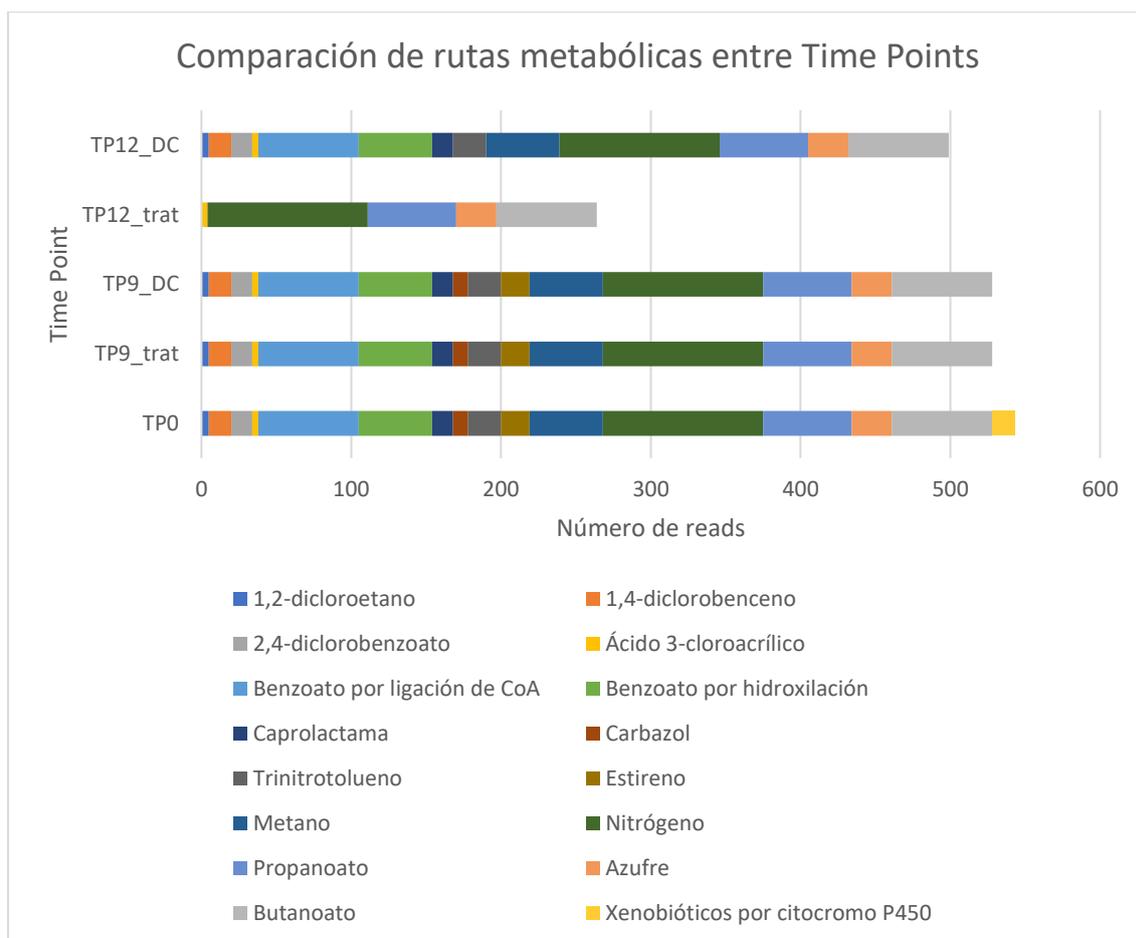


Figura 5.- Comparación de vías metabólicas encontradas en cada Time Point.

REFERENCIAS BIBLIOGRÁFICAS

- Allen, Kon & Bar-Yam (2009) Allen B, Kon M, Bar-Yam Y. *A new phylogenetic diversity measure generalizing the Shannon index and its application to phyllostomid bats*. *The American Naturalist*. 2009;174(2):236–243.
doi: 10.1086/600101.
- Beckman Coulter. (2021). *AMPure XP Performance and Data*. Obtenido de Beckman Coulter Life Sciences: <https://www.beckman.com/reagents/genomic/cleanup-and-size-selection/pcr/performance>
- Benítez, M., Champagne, P., Ramos, A., Torres, A., & Ochoa-Herrera, V. (2018). *Wastewater treatment for nutrient removal with Ecuadorian native microalgae*. *Environmental Technology*, 1-9.
- Benoit G, Peterlongo P, Mariadassou M, Drezen E, Schbath S, Lavenier D, Lemaitre C. (2016). *Multiple comparative metagenomics using multiset k-mer counting*. *PeerJ Computer Science* 2:e94 <https://doi.org/10.7717/peerj-cs.94>
- Bharagava, R. N., Purchase, D., Saxena, G., & Mulla, S. I. (2019). *Applications of Metagenomics in Microbial Bioremediation of Pollutants*. *Microbial Diversity in the Genomic Era*, 459–477. doi:10.1016/b978-0-12-814849-5.00026-5
- Braich, O. S., & Kaur, R. (2015). *Phytoplankton Community Structure and Species Diversity of Nangal*. *International Research Journal of Biological Sciences*, 4(3), 76–83.
- Corvelo, A., Clarke, W. E., Robine, N., & Zody, M. C. (2018). taxMaps: *Comprehensive and highly accurate taxonomic classification of short-read data in reasonable time*. *Genome Research*, 28(5), 751–758.
<https://doi.org/10.1101/gr.225276.117>

- Cuellar-Bermudez, S. P., Aleman-Nava, G. S., Chandra, R., Garcia-Perez, J. S., Contreras-Angulo, J. R., Markou, G., Muylaert, K., Rittmann, B., & Parra-Saldivar, R. (2017). *Nutrients utilization and contaminants removal. A review of two approaches of algae and cyanobacteria in wastewater*. *Algal Research*, 24, 438-449. <https://doi.org/10.1016/j.algal.2016.08.018>
- Datta, S., Rajnish, K. N., Samuel, M. S., Pugazhendhi, A., & Selvarajan, E. (2020). *Metagenomic applications in microbial diversity, bioremediation, pollution monitoring, enzyme and drug discovery. A review*. *Environmental Chemistry Letters*. doi:10.1007/s10311-020-01010-z
- Demircan, T., Ovezmyradov, G., Yildirim, B., Keskin, İ., İlhan, A. E., Fesçioğlu, E. C., ... Yildirim, S. (2018). *Experimentally induced metamorphosis in highly regenerative axolotl (ambystoma mexicanum) under constant diet restructures microbiota*. *Scientific Reports*, 8(1). doi:10.1038/s41598-018-29373-y
- Eurostat-EU. (2021). *Generation and discharge of wastewater in volume*. Obtenido de Eurostat:http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=env_ww_genv&lang=en
- GitHub. (2018). *Porechop*. Obtenido de GitHub: <https://github.com/rrwick/Porechop>
- GitHub. (2020). *NanoPlot*. Obtenido de GitHub: <https://github.com/wdecoster/NanoPlot>
- Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. C., & McCombie, W. R. (2015). *Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome*. *Genome research*, 25(11), 1750–1756. <https://doi.org/10.1101/gr.191395.115>
- Guldhe, A., Kumari, S., Ramanna, L., Ramsundar, P., Singh, P., Rawat, I., & Bux, F. (2017). *Prospects, recent advancements and challenges of different wastewater*

- streams for microalgal cultivation*. *Journal of Environmental Management*, 203, 299–315. doi:10.1016/j.jenvman.2017.08.012
- He, F., and X. Hu. 2005. *Hubbell's fundamental biodiversity parameter and the Simpson diversity index*. *Ecology Letters* 8:386–390.
- Ibarbalz, F., Figuerola, E., & Erijman, L. (2013). *Industrial activated sludge exhibit unique bacterial community composition at high taxonomic ranks*. *Water Research*, 11(47), 3854–3864.
- Illumina. (2011). *Quality Scores for Next-Generation Sequencing* . Obtenido de Illumina: https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf
- INEC. (2016). *Estadística de Información Ambiental Económica en Gobiernos Autónomos Descentralizados Municipales 2015 (Agua y Alcantarillado)*. Obtenido de Instituto Nacional de Estadística y Censos: https://www.ecuadorencifras.gob.ec/documentos/web-inec/Encuestas_Ambientales/Municipios_2015/Documento_Tecnico-Gestion_de_Agua_y_Alcantarillado_2015.pdf
- Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. (2019). *Assembly of long, error-prone reads using repeat graphs*. *Nature Biotechnology*.
- Krustok, I. (2016). *Microbiological analysis of municipal wastewater treating photobioreactors Mälardalens Högskola Microbiological analysis of municipal wastewater treating photobioreactors Ivo Krustok Supervisors : Monica Odlare (PhD) Jaak Truu (PhD), (January)*.
- Kumar Awasthi, M., Ravindran, B., Sarsaiya, S., Chen, H., Wainaina, S., Singh, E., Liu, T., Kumar, S., Pandey, A., Singh, L., & Zhang, Z. (2020). *Metagenomics for*

- taxonomy profiling: tools and approaches*. *Bioengineered*, 11(1), 356–374.
<https://doi.org/10.1080/21655979.2020.1736238>
- Liao, P., Satten, G., & Juan, Y. (2017). *PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies*. *Genetic Epidemiology*, 41(5), 375-387.
- Loman, N. J., Quick, J., & Simpson, J. T. (2015). *A complete bacterial genome assembled de novo using only nanopore sequencing data*. *Nature Methods*, 12(8), 733–735. doi:10.1038/nmeth.3444
- Menzel, P., Lee, K., & Krogh, A. (2016). *Fast and sensitive taxonomic classification for metagenomics with Kaiju*. *Nature Communications*, 7(11257).
- Miura, Y., Hiraiwa, M. N., Ito, T., Itonaga, T., Watanabe, Y., & Okabe, S. (2007). *Bacterial community structures in MBRs treating municipal wastewater: Relationship between community stability and reactor performance*. *Water Research*, 41(3), 627–637. <https://doi.org/10.1016/j.watres.2006.11.005>
- Morales-Sánchez, D., Martínez-Rodríguez, O. A., Kyndt, J., & Martínez, A. (2014). *Heterotrophic growth of microalgae: metabolic aspects*. *World Journal of Microbiology and Biotechnology*, 31(1), 1–9. doi:10.1007/s11274-014-1773-2
- Oksanen, J., & al, e. (2019). *Community Ecology Package*.
- Orbe, T. (2020). *Hallan metales pesados en ríos amazónicos de Ecuador*. Obtenido de SciDevNet: <https://www.scidev.net/america-latina/news/hallan-metales-pesados-en-rios-amazonicos-de-ecuador/>
- Oxford Nanopore Technologies. (2019). *Native barcoding genomic DNA (with EXP-NBD104, EXP-NBD114, and SQK-LSK109)*. Oxford.

- Pires, J. C. M., Alvim-Ferraz, M. C. M., Martins, F. G., & Simões, M. (2013). *Wastewater treatment to enhance the economic viability of microalgae culture*. *Environmental Science and Pollution Research*, 20(8), 5096–5105. doi:10.1007/s11356-013-1791-x
- Posso, S. (2019). *Explorando la biodiversidad de consorcios microbianos Amazónicos con capacidad biorremediadora de aguas residuales mediante secuenciación de última generación*. Tesis de pregrado, Universidad San Francisco de Quito USFQ.
- QIAGEN. (2016). *DNeasy PowerSoil Pro Kit Handbook*. Obtenido de QIAGEN: <https://www.qiagen.com/us/resources/resourcedetail?id=9bb59b74-e493-4aeb-b6c1-f660852e8d97&lang=en>
- Qiang, J., Kurade, M., & Jeon, B. (2017). *Can Microalgae Remove Pharmaceutical Contaminants from Water?* *Trends in Biotechnology*, 1-15.
- R Core Team (2019) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Renuka, N., Sood, A., Ratha, S.K., Prasanna, R., & Ahluwalia, A.S. (2013). *Evaluation of microalgal consortia for treatment of primary treated sewage effluent and biomass production*. *Journal of Applied Phycology*, 25, 1529-1537.
- Ricotta, C., & Podani, J. (2017). *On some properties of the Bray-Curtis dissimilarity and their ecological meaning*. *Ecological Complexity*, 31, 201–205. doi:10.1016/j.ecocom.2017.07.003
- Sardi-Saavedra, Antonella, Peña-Salamanca, Enrique J, Madera-Parra, Carlos A, & Cerón-Hernández, Víctor A. (2016). *Diversidad de las comunidades de algas asociadas a un sistema algal de alta tasa fotosintética para la biorremediación*

- de lixiviados de rellenos sanitarios*. Latin american journal of aquatic research, 44(1), 113-120. <https://dx.doi.org/10.3856/vol44-issue1-fulltext-11>
- Sato, T., Qadir, M., Yamamoto, S., Endo, T., & Zahoor, A. (2013). *Global, regional, and country level need for data on wastewater generation, treatment, and use*. Agricultural Water Management, 130, 1–13. doi:10.1016/j.agwat.2013.08.007
- Shin, H., Lee, E., Shin, J., Ko, S. R., Oh, H. S., Ahn, C. Y., ... Cho, S. (2018). *Elucidation of the bacterial communities associated with the harmful microalgae Alexandrium tamarense and Cochlodinium polykrikoides using nanopore sequencing*. Scientific Reports, 8(1), 4–11. <https://doi.org/10.1038/s41598-018-23634-6>
- Suganya, T., Varman, M., Masjuki, H. H., & Renganathan, S. (2016). *Macroalgae and microalgae as a potential source for commercial applications along with biofuels production: A biorefinery approach*. Renewable and Sustainable Energy Reviews, 55, 909–941. doi:10.1016/j.rser.2015.11.026
- Tamames, J., & Puente, F. (2019). *SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline*. Frontiers in Microbiology, 1-10.
- Tandoi, V., Levantesi, C., Eikelboom, D., van der Waarde, J., Snaidr, J., Rossetti, S., ... Lemaitre, M. (2018). *Phylogenetic analysis and in situ identification of “Nostocoida limicola”-like filamentous bacteria in activated sludge from industrial wastewater treatment plants*. Water Science and Technology, 46(1–2), 99–104. <https://doi.org/10.2166/wst.2002.0462>
- Techtmann, S. M., & Hazen, T. C. (2016). *Metagenomic applications in environmental monitoring and bioremediation*. Journal of Industrial Microbiology & Biotechnology, 43(10), 1345–1354. doi:10.1007/s10295-016-1809-8

Thermo Fisher Scientific Inc. (2009). *NanoDrop 2000/2000c Spectrophotometer* .

Obtenido de ThermoFisher Scientific: [https://www.thermofisher.com/document-connect/document-](https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FCAD%2Fmanuals%2FNanoDrop-2000-User-Manual-EN.pdf)

[connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FCAD%2Fmanuals%2FNanoDrop-2000-User-Manual-EN.pdf](https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FCAD%2Fmanuals%2FNanoDrop-2000-User-Manual-EN.pdf)

Thermo Fisher Scientific Inc. (2014). *Qubit® 3.0 Fluorometer*. Obtenido de

ThermoFisher Scientific:

https://tools.thermofisher.com/content/sfs/manuals/qubit_3_fluorometer_man.pdf

Tiron, O., Bumbac, C., Patroescu, I. V., Badescu, V. R., & Postolache, C.

(2015). *Granular activated algae for wastewater treatment*. *Water Science and Technology*, 71(6), 832–839. doi:10.2166/wst.2015.010

Uhlik, O., Leewis, M. C., Strejcek, M., Musilova, L., Mackova, M., Leigh, M. B., & Macek, T. (2013). *Stable isotope probing in the metagenomics era: a bridge towards improved bioremediation*. *Biotechnology advances*, 31(2), 154–165. <https://doi.org/10.1016/j.biotechadv.2012.09.003>

WWAP (United Nations World Water Assessment Programme). 2017. *The United Nations World Water Development Report 2017: Wastewater, The Untapped Resource*. Paris, UNESCO

Xu, Z., Asser , M., Hansen, L., Jacquiod, S., & Sørensen, S. (2014). *Bioinformatic Approaches Reveal Metagenomic Characterization of Soil Microbial Community*. *PLoS One*, 9(4).

Yadav, R., Rajput, V., & Dharne, M. (2021). *Functional metagenomic landscape of polluted river reveals potential genes involved in degradation of xenobiotic*

pollutants. Environmental Research, 192,

110332. doi:10.1016/j.envres.2020.110332

Ye, Y., & Doak, T. G. (2009). *A Parsimony Approach to Biological Pathway*

Reconstruction/Inference for Genomes and Metagenomes. PLoS Computational

Biology, 5(8), e1000465. doi:10.1371/journal.pcbi.1000465

ANEXO 1: LECTURAS CLASIFICADAS A NIVEL DE DOMINIO

		TP0	TP9_trat	TP9_DC	TP12_trat	TP12_DC
Bacteria	Porcentaje	61%	75%	70%	88%	88%
	Lecturas	132,108	29,709	29,884	30,339	55,119
Eucariota	Porcentaje	38%	25%	30%	12%	12%
	Lecturas	83,918	9871	12,644	4,044	7,364

Descripción. El porcentaje para cada dominio se calculó en base a las lecturas clasificadas de todos los organismos celulares en cada Time Point. El número de lecturas se obtiene en base al total de lecturas generadas. Se usó el programa Kaiju y la base de datos NCBI BLAST nr+euk.

ANEXO 2: LECTURAS CLASIFICADAS A NIVEL DE FILO

		TP0	TP9_trat	TP9_DC	TP12_trat	TP12_DC
Proteobacteria	Porcentaje	41%	45%	40%	58%	58%
	Lecturas	89,838	17,782	17,289	20,010	36,566
FCB	Porcentaje	14%	24%	24%	20%	20%
	Lecturas	29,906	9,440	10,145	6,908	12,242
Terrabacteria	Porcentaje	2%	0.80%	1%	0.8%	0.7%
	Lecturas	4,709	329	533	271	452
PVC	Porcentaje	2%	3%	3%	3%	3%
	Lecturas	3,773	1,231	1,088	1,108	1,921
Acidobacteria	Porcentaje	0.50%	0.05%	0.09%	0.04%	0.04%
	Lecturas	1,053	19	38	14	28
Chlorophyta	Porcentaje	33%	12%	22%	6%	6%
	Lecturas	71,510	4,799	9,259	1,986	3,550
Opisthokinta	Porcentaje	5%	12%	7%	6%	6%
	Lecturas	10,131	4,874	3,063	1,958	3,621
Bacteroidetes	Porcentaje	14%	24%	24%	20%	20%
	Lecturas	29,601	9,429	10,136	6,902	12,234
Cyanobacteria	Porcentaje	0.30%	0.10%	0.10%	0.09%	0.06%
	Lecturas	731	53	54	30	36

Descripción. El porcentaje para cada filo se calculó en base a las lecturas clasificadas de todos los organismos celulares en cada Time Point. El número de lecturas se obtiene en base al total de lecturas generadas. Se usó el programa Kaiju y la base de datos NCBI BLAST nr+euk. Se observa que hay una mayor abundancia del filo Proteobacteria en las muestras de todos los Time Points.

ANEXO 3: LECTURAS CLASIFICADAS A NIVEL DE CLASE

		TP0	TP9_trat	TP9_DC	TP12_trat	TP12_DC
α-Proteobacteria	Porcentaje	31%	37%	33%	50%	51%
	Lecturas	67,735	14,820	14,172	17,397	31,748
β-Proteobacteria	Porcentaje	8%	6%	6%	6%	6%
	Lecturas	18,056	2,257	2,448	1,979	3,579
γ-Proteobacteria	Porcentaje	0.90%	0.70%	0.60%	0.6%	0.7%
	Lecturas	1,918	259	259	215	408
Gemmatimonadetes	Porcentaje	0.10%	0.02%	0.01%	0.01%	0.006%
	Lecturas	288	7	5	4	4
Actinobacteria	Porcentaje	1%	0.40%	0.40%	0.4%	0.5%
	Lecturas	2,703	140	159	146	291
Firmicutes	Porcentaje	0.30%	0.30%	0.60%	0.2%	0.1%
	Lecturas	748	110	267	65	85
Chloroflexi	Porcentaje	0.20%	0.05%	0.10%	0.06%	0.04%
	Lecturas	453	18	44	21	26
Verrucomicrobia	Porcentaje	1%	3%	2%	3%	3%
	Lecturas	2,721	1,077	731	915	1,581
Planctomycetes	Porcentaje	0.50%	0.40%	0.80%	0.5%	0.5%
	Lecturas	1,001	146	342	183	318

ANEXO 4: LECTURAS GENERADAS PARA RUTAS METABÓLICAS

	TP0	TP9_trat	TP9_DC	TP12_trat	TP12_DC
1,2-dicloroetano	5	5	5	0	5
1,4-diclorobenceno	15	15	15	0	15
2,4-diclorobenzoato	14	14	14	0	14
Ácido 3-cloroacrílico	4	4	4	4	4
Benzoato por ligación de CoA	67	67	67	0	67
Benzoato por hidroxilación	49	49	49	0	49
Caprolactama	14	14	14	0	14
Carbazol	10	10	10	0	0
Trinitrotolueno	22	22	22	0	22
Estireno	19	19	19	0	0
Metano	49	49	49	0	49
Nitrógeno	107	107	107	107	107
Propanoato	59	59	59	59	59
Azufre	27	27	27	27	27
Butanoato	67	67	67	67	67
Xenobióticos por citocromo P450	15	0	0	0	0

Descripción. Número de lecturas generadas para las rutas metabólicas relevantes en biorremediación. Información obtenida con el uso de la herramienta SqueezeMeta.