

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR
FACULTAD DE HÁBITAT, INFRAESTRUCTURA Y CREATIVIDAD**

MAESTRÍA EN BIOLOGÍA COMPUTACIONAL

Identificación de genes de resistencia a antibióticos en bacterias presentes en vinaza de
caña de azúcar (*Saccharum officinarum*) mediante análisis metagenómico

Trabajo previo a la obtención del título de:

Magíster en Biología Computacional

Autora: Mishelle Nathaly Culcay García

Directora: Mgtr. Yadira Salguero Salas

Quito, 2025

DERECHOS DE AUTOR

Yo, Mishelle Nathaly Culcay García, con documento de identidad Nro. 1500894694, expreso que soy la autora del presente trabajo de titulación y consulté las referencias indicadas en el mismo. Este trabajo no fue presentado de forma previa para la obtención de ningún grado académico.

La Pontificia Universidad Católica del Ecuador puede utilizar los derechos del trabajo de titulación de acuerdo con la Ley de Propiedad Intelectual y su normativa institucional.

A handwritten signature in blue ink, appearing to read 'Mishelle Culcay', is positioned above a horizontal line.

Ing. Mishelle Culcay

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN

Certifico que el trabajo de titulación: “Identificación de genes de resistencia a antibióticos en bacterias presentes en vinaza de caña de azúcar (*Saccharum officinarum*) mediante análisis metagenómico”, para la obtención del título de Magíster en Biología Computacional fue realizado por Mishelle Nathaly Culcay García, bajo mi dirección.

El mismo se llevó a cabo bajo la normativa y reglamentación institucional y puede ser presentado para su calificación.

Mgtr. Yadira Salguero Salas

HOJA DE EVIDENCIA ANTIPLAGIO

Trabajo de Titulación Mishelle Culcay

ORIGINALITY REPORT

9 %	8 %	3 %	3 %
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repositorio.puce.edu.ec Internet Source	2 %
2	hdl.handle.net Internet Source	1 %
3	Submitted to Pontificia Universidad Catolica del Ecuador - PUCE Student Paper	<1 %
4	Submitted to UNIBA Student Paper	<1 %
5	eprints.ucm.es Internet Source	<1 %
6	buleria.unileon.es Internet Source	<1 %
7	www.colibri.udelar.edu.uy Internet Source	<1 %
8	www.frontiersin.org Internet Source	<1 %
9	repositorio.uisek.edu.ec Internet Source	<1 %

DEDICATORIA

A mi querida Tía Fanny, a ti te debo este logro.

ÍNDICE GENERAL

DERECHOS DE AUTOR.....	2
APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN	3
HOJA DE EVIDENCIA ANTIPLAGIO.....	4
DEDICATORIA.....	5
ÍNDICE GENERAL.....	6
ÍNDICE DE FIGURAS	9
ÍNDICE DE TABLAS.....	10
ÍNDICE DE ANEXOS	11
RESUMEN.....	12
ABSTRACT	13
1. INTRODUCCIÓN.....	14
1.1. Formulación del problema.....	14
1.2. Justificación	16
1.3. Objetivos.....	18
1.3.1. Objetivo general	18
1.3.2. Objetivos específicos	18
2. REVISIÓN DE LA LITERATURA.....	19
2.1. Producción de caña de azúcar y vinaza.....	19
2.2. Impacto ambiental de la vinaza	20
2.3. Genes de resistencia a antibióticos (ARGs).....	20
2.4. Transferencia horizontal de genes (THG).....	23

2.5.	Metagenómica y su aplicación ambiental	23
3.	METODOLOGÍA.....	25
3.1.	Obtención de datos de secuenciación	26
3.2.	Control de calidad de lecturas.....	27
3.3.	Filtrado y recorte de lecturas	27
3.4.	Clasificación taxonómica.....	28
3.5.	Estimación de abundancia.....	29
3.6.	Ensamblaje metagenómico	30
3.7.	Agrupamiento de contigs en bins y evaluación de bins.....	31
3.8.	Identificación de genes de resistencia a antibióticos	33
3.9.	Anotación del genoma	34
4.	RESULTADOS Y ANÁLISIS DE RESULTADOS.....	36
4.1.	Obtención de datos de secuenciación	36
4.2.	Control de calidad de lecturas.....	36
4.3.	Filtrado y recorte de lecturas	38
4.4.	Clasificación taxonómica.....	40
4.5.	Estimación de abundancia.....	51
4.6.	Ensamblaje metagenómico	59
4.7.	Agrupamiento de contigs en bins y evaluación de bins.....	62
4.8.	Identificación de genes de resistencia a antibióticos	64
4.9.	Anotación del genoma	76
5.	CONCLUSIONES Y RECOMENDACIONES	79

5.1. Conclusiones	79
5.2. Recomendaciones	80
REFERENCIAS	82
ANEXOS	95

ÍNDICE DE FIGURAS

Figura 1. Flujo selectivo de ARG del medio ambiente a la clínica (Skandalis et al., 2021)	21
Figura 2. Calidad promedio por posición de base en las secuencias filtradas.....	38
Figura 3. Cantidad de lecturas únicas y duplicadas por archivo	39
Figura 4. Distribución del contenido GC por secuencia.....	40
Figura 5. Diagrama circular de Krona del metagenoma SRX3743992.....	42
Figura 6. Diagrama circular de Krona del metagenoma SRX3743987.....	44
Figura 7. Diagrama circular de Krona del metagenoma SRX3743986.....	46
Figura 8. Diagrama circular de Krona del metagenoma SRX3743982.....	48
Figura 9. Diagrama circular de Krona del metagenoma SRX3743980.....	49
Figura 10. Diagrama circular de Krona del metagenoma SRX3743975.....	50
Figura 11. Comparación del ensamblaje del metagenoma SRX3743975 con MEGAHIT y metaSPAdes mediante QUAST.....	60
Figura 12. Comparación de distribuciones de contigs ensamblados del metagenoma SRX3743975 con MEGAHIT y metaSPAdes.....	61
Figura 13. Evaluación de calidad de los bins generados por MaxBin2 mediante CheckM	63
Figura 14. Evaluación de calidad de los bins generados por MetaBAT2 mediante CheckM	63
Figura 15. Distribución de puntuaciones de los bins generados con DAS Tool	76
Figura 16. Evaluación de la calidad los bins obtenidos con DAS Tool	77

ÍNDICE DE TABLAS

Tabla 1. Descripción de los datos de los metagenomas de vinaza	26
Tabla 2. Métricas de calidad de las secuencias crudas obtenidas para los metagenomas analizados	37
Tabla 3. Estimación de abundancia a nivel de especie para el metagenoma SRX3743992	52
Tabla 4. Estimación de abundancia a nivel de especie para el metagenoma SRX3743987	53
Tabla 5. Estimación de abundancia a nivel de especie para el metagenoma SRX3743986	54
Tabla 6. Estimación de abundancia a nivel de especie para el metagenoma SRX3743982	55
Tabla 7. Estimación de abundancia a nivel de especie para el metagenoma SRX3743980	56
Tabla 8. Estimación de abundancia a nivel de especie para el metagenoma SRX3743975	57
Tabla 9. Genes de resistencia detectados en los bins de los 18 metagenomas por ResFinder y CARD.....	70
Tabla 10. Evaluación de linaje de CheckM.....	77

ÍNDICE DE ANEXOS

Anexo 1. Tabla resumen de las estadísticas generales de MultiQC	95
Anexo 2. Parámetros utilizados en el Trimmomatic	96
Anexo 3. Diagrama circular de Krona del metagenoma SRX3743991	97
Anexo 4. Diagrama circular de Krona del metagenoma SRX3743990	97
Anexo 5. Diagrama circular de Krona del metagenoma SRX3743989	98
Anexo 6. Diagrama circular de Krona del metagenoma SRX3743988	98
Anexo 7. Diagrama circular de Krona del metagenoma SRX3743985	99
Anexo 8. Diagrama circular de Krona del metagenoma SRX3743984	99
Anexo 9. Diagrama circular de Krona del metagenoma SRX3743983	100
Anexo 10. Diagrama circular de Krona del metagenoma SRX3743981	100
Anexo 11. Diagrama circular de Krona del metagenoma SRX3743979	101
Anexo 12. Diagrama circular de Krona del metagenoma SRX3743978	101
Anexo 13. Diagrama circular de Krona del metagenoma SRX3743977	102
Anexo 14. Diagrama circular de Krona del metagenoma SRX3743976	102

RESUMEN

La vinaza de caña de azúcar, subproducto del proceso para la obtención de etanol, es utilizada a gran escala como fertilizante debido a su alto contenido de nutrientes. No obstante, su aplicación recurrente en suelos agrícolas plantea ciertas implicaciones para el ambiente y la salud, entre ellas la posible diseminación de genes de resistencia a antibióticos (ARGs). Este trabajo realizó un análisis metagenómico *in silico* sobre 18 metagenomas de vinaza proveniente de una planta productora de alcohol ubicada en São Paulo, Brasil, con la finalidad de caracterizar la diversidad microbiana presente e identificar ARGs asociados.

Se utilizaron herramientas bioinformáticas para analizar y procesar los datos, entre ellas Kraken2 y Bracken para la clasificación taxonómica, MEGAHIT para el ensamblaje de secuencias, MetaBAT2 para el agrupamiento de contigs en bins y ABRicate para la detección de ARGs, apoyada en las bases de datos ResFinder y CARD. El estudio reveló comunidades bacterianas dominadas por Firmicutes, con especies frecuentes como *Megasphaera elsdenii* y *Lactobacillus amylovorus*, las cuales han sido reportadas como portadoras de genes de resistencia antimicrobiana.

Entre los genes identificados se encuentran *lnu(A)*, *lnu(B)*, *lnu(C)*, *erm(B)*, *blaACI-1*, *lsaE* y *vat(E)*, asociados a resistencias frente a lincosamidas, macrólidos, betalactámicos y estreptograminas. Estos genes se detectaron con altos niveles de identidad y cobertura, lo que respalda su relevancia en la matriz analizada. Los resultados obtenidos indican que la vinaza representa un reservorio activo de ARGs en ambientes agrícolas, con implicaciones potenciales para la propagación de resistencia antimicrobiana.

ABSTRACT

Sugarcane vinasse, a byproduct of ethanol production, is widely used as a fertilizer due to its high nutrient content. However, its repeated application to agricultural soils poses potential environmental and health risks, including the potential dissemination of antibiotic resistance genes (ARGs). This study conducted an *in silico* metagenomic analysis of 18 vinasse metagenomes from an alcohol production plant located in São Paulo, Brazil, with the aim of characterizing the microbial diversity present and identifying associated ARGs.

Bioinformatic tools were employed for data processing and analysis, including Kraken2 and Bracken for taxonomic classification, MEGAHIT for sequence assembly, MetaBAT2 for contig binning, and ABRicate for ARG detection, supported by the ResFinder and CARD databases. The analysis revealed bacterial communities dominated by Firmicutes, with frequent species such as *Megasphaera elsdenii* and *Lactobacillus amylovorus*, which have been reported as carriers of antimicrobial resistance genes.

Among the identified genes were *lnu(A)*, *lnu(B)*, *lnu(C)*, *erm(B)*, *blaACI-1*, *lsaE*, and *vat(E)*, associated with resistance to lincosamides, macrolides, beta-lactams, and streptogramins. These genes were detected with high identity and coverage, supporting their relevance in the analyzed matrix. The results indicate that vinasse represents an active reservoir of ARGs in agricultural environments, with potential implications for the spread of antimicrobial resistance.

1. INTRODUCCIÓN

1.1. Formulación del problema

La resistencia a los antibióticos representa una seria amenaza sanitaria a nivel mundial, comprometiendo la eficacia de tratamientos antimicrobianos y facilitando la transmisión de infecciones resistentes (Organización Mundial de la Salud, 2020). Según Tien et al. (2017), los genes de resistencia a antibióticos (ARGs, por sus siglas en inglés) se encuentran presentes en una amplia variedad de ecosistemas, incluidos los ambientes agrícolas. En estos entornos, la diseminación de ARGs puede verse favorecida por prácticas como el uso de residuos industriales en la fertilización de suelos, aumentando la probabilidad de transferencia horizontal de estos genes entre bacterias.

Entre estos residuos destaca la vinaza, un derivado de la caña de azúcar (*Saccharum officinarum*), resultante del proceso de producción de etanol, la cual destaca por su alto contenido de materia orgánica, nutrientes y minerales como calcio y potasio, además de fósforo y nitrógeno en cantidades moderadas, lo que ha promovido su uso como fertilizante en sistemas de fertirrigación (Christofolletti et al., 2013). Sin embargo, la vinaza también posee un alto potencial contaminante, debido a la presencia de compuestos como fenoles, polifenoles, metales pesados y residuos antibacterianos, los cuales pueden afectar negativamente tanto a microorganismos como a plantas, generando inquietudes respecto a su impacto en el equilibrio microbiano del suelo (Ospina León et al., 2022).

Braga et al. (2017) evidencia que su aplicación en suelos puede alterar la dinámica del resistoma, aumentando la abundancia de ARGs y modificando sus interacciones, sugiriendo un impacto directo en la evolución de la resistencia bacteriana, lo que respalda la hipótesis de que la vinaza no es un simple fertilizante orgánico, sino un vector en

potencia de genes de resistencia que pueden ser transferidos a bacterias ambientales y patógenas.

En el Ecuador, la caña de azúcar representa un cultivo de alto valor económico. En 2020, la producción alcanzó aproximadamente 11 millones de toneladas métricas, representando el 0,19 % del Producto Interno Bruto (PIB) nacional (Asobanca, 2022). Este cultivo es vital para la obtención de azúcar y etanol. Se calcula que la producción anual de etanol en el país alcanza los 100 millones de litros (Antony Cristofer et al., 2023), lo que implica una generación proporcional de vinaza que bordea los 1,200 millones de litros anuales, ya que por cada litro de etanol pueden generarse hasta 12 litros de vinaza (Del M. Chaile et al., 2024). Ingenios como Valdez reportan un uso anual de hasta 468,000 m³ de vinaza en fertirrigación (Suárez Valles & Garzón Avilés, 2024).

El riesgo asociado a la fertirrigación con vinaza es significativo, ya que las bacterias del suelo pueden actuar como reservorios de ARGs, los cuales podrían ser transferidos a patógenos humanos a través del contacto con el suelo, el consumo de cultivos contaminados o la infiltración en fuentes de agua (Forsberg et al., 2012). La dispersión de ARGs en sistemas agrícolas puede comprometer la seguridad alimentaria y contribuir a la crisis global de la resistencia antimicrobiana, disminuyendo la efectividad de los antibióticos en el combate contra infecciones (Lemos et al., 2021).

Pese a lo antes descrito, la diversidad microbiana presente en la vinaza y su rol en la propagación de la resistencia antimicrobiana han sido poco estudiados, lo que deja una brecha de conocimiento fundamental. Por tal motivo, este trabajo busca identificar genes de resistencia a antibióticos en bacterias presentes en vinaza de caña de azúcar (*Saccharum officinarum*) mediante análisis metagenómico para comprender el papel de la vinaza como reservorio y vector potencial de ARGs.

1.2. Justificación

La progresiva preocupación global por la propagación de genes de resistencia a antibióticos (ARGs) ha impulsado la necesidad de generar estudios que aborden este fenómeno desde una perspectiva amplia e integradora. En este sentido, se ha propuesto adoptar un enfoque integral que contemple los mecanismos de resistencia en distintos ambientes, para tratar de comprender los elementos genéticos y ecológicos que regulan la transmisión y evolución de dicha resistencia. Esta visión integral es promovida por el marco conceptual de “One Health”, el cual recomienda investigaciones multidisciplinarias en distintos niveles (local, nacional y mundial), para abordar los riesgos compartidos entre la salud humana, animal y ambiental (Durso & Cook, 2019).

Desde esta perspectiva, resulta prioritario identificar y caracterizar posibles fuentes de diseminación de ARGs. Entre ellas, los residuos agroindustriales utilizados como fertilizantes destacan como vectores relevantes. La vinaza, es ampliamente utilizada en fertirrigación debido a su alto contenido de materia orgánica y nutrientes (Christofoletti et al., 2013). Sin embargo, su aplicación intensiva ha sido señalada como un potencial factor de riesgo en la diseminación de ARGs y en la alteración del equilibrio microbiano natural del suelo. Estudios previos han demostrado que bacterias portadoras de ARGs, mediante mecanismos de transferencia horizontal, pueden transmitir estos genes a otros microorganismos (Forsberg et al., 2012).

No obstante, existe una importante brecha de conocimiento respecto a la caracterización microbiana de la vinaza y su rol específico en la propagación de la resistencia antimicrobiana. En Ecuador, los procesos asociados a la industria azucarera generan grandes volúmenes de vinaza, alcanzando hasta 12 litros por cada litro de etanol producido; esto se traduce en millones de litros anuales aplicados en suelos agrícolas, especialmente en zonas como Guayas, Los Ríos y Cañar (Antony Cristofer et al., 2023;

Del M. Chaile et al., 2024). A pesar de esta alta exposición, la caracterización de su microbiota y su potencial como reservorio de ARGs ha sido escasamente abordada en el país, lo que evidencia la premura de generar investigaciones científicas en este campo, para facilitar investigaciones futuras en el Ecuador.

La pertinencia de esta investigación, gira en torno a la necesidad de generar conocimiento sobre la presencia y diversidad de ARGs en bacterias presentes en la vinaza, lo cual contribuirá a entender su posible papel en la diseminación de resistencia antimicrobiana en sistemas agrícolas. En este contexto, el presente trabajo integra herramientas metagenómicas y bioinformáticas, mediante las cuales se ensamblará el metagenoma de muestras de vinaza y se identificarán genes asociados a la resistencia de antibióticos.

La metagenómica ha demostrado ser una potente herramienta para el análisis de comunidades microbianas complejas y la detección de ARGs en matrices ambientales sin necesidad de cultivo (Quince et al., 2017). A su vez, esta investigación permitirá aplicar estrategias de ensamblaje a partir de datos metagenómicos, aportando al conocimiento y desarrollo de metodologías bioinformáticas aplicadas al análisis de microbiomas ambientales.

Si bien los datos metagenómicos empleados en este trabajo provienen de muestras recolectadas en Brasil, su análisis es metodológicamente aplicable para el caso ecuatoriano, dado que la vinaza en ambos países proviene de procesos industriales similares (CINCAE, 2018; Rodrigues Reis & Hu, 2017). Esto otorga al estudio un valor extrapolable, capaz de generar insumos relevantes para la toma de decisiones en contextos locales donde aún no existen datos propios.

El impacto potencial de este trabajo se traduce en beneficios tanto científicos como ecológicos. La identificación de bacterias portadoras de ARGs en la vinaza podría servir

como una base para futuras regulaciones sobre el manejo de residuos agroindustriales y para el diseño de estrategias de control que minimicen la dispersión de resistencia antimicrobiana en ambientes agrícolas. De este modo, este estudio contribuirá al avance del conocimiento en microbiología ambiental y biotecnología aplicada, con implicaciones directas en la salud pública y la sostenibilidad del sector agrícola.

1.3. Objetivos

1.3.1. Objetivo general

Identificar genes de resistencia a antibióticos (ARGs) en bacterias presentes en vinaza de caña de azúcar (*Saccharum officinarum*) mediante análisis metagenómico

1.3.2. Objetivos específicos

- Preprocesar los datos metagenómicos de vinaza, asegurando la calidad y eliminación de secuencias contaminantes y de baja calidad.
- Ensamblar el metagenoma de la comunidad microbiana presente en la vinaza de caña de azúcar
- Caracterizar estructuralmente la comunidad microbiana presente en la vinaza de caña de azúcar
- Identificar genes de resistencia a antibióticos (ARGs) en bacterias presentes en la vinaza de caña de azúcar.
- Ensamblar el genoma de una bacteria de la familia *Lactobacillaceae* a partir de los datos metagenómicos
- Anotar estructuralmente el genoma ensamblado de una bacteria de la familia *Lactobacillaceae*.

2. REVISIÓN DE LA LITERATURA

2.1. Producción de caña de azúcar y vinaza

La caña de azúcar (*Saccharum officinarum*) es uno de los cultivos económicamente más trascendentes de América Latina, incluyendo a Ecuador, donde constituye un motor productivo en provincias como Guayas, Cañar, Santa Elena y Los Ríos (Antony Cristofer et al., 2023; Suárez Valles & Garzón Avilés, 2024). Este cultivo se emplea principalmente para la producción de azúcar y obtención de bioetanol, un biocombustible en creciente demanda, generado a través de la fermentación de azúcares. Por cada litro de etanol obtenido, se derivan en promedio 12 litros de vinaza residuo (Del M. Chaile et al., 2024).

Además del impacto económico, la producción de caña y bioetanol implica una gestión intensiva de residuos industriales. En Ecuador, ingenios como el Ingenio Valdez aplican anualmente más de 468,000 m³ de vinaza en fertirrigación (Suárez Valles & Garzón Avilés, 2024), lo que repercute en desafíos ambientales relacionados con la acumulación de materia orgánica, sales y otros compuestos químicos en los suelos agrícolas.

La alta carga orgánica y mineral de la vinaza, que incluye potasio, calcio, nitrógeno y fósforo, le otorga propiedades fertilizantes valiosas (Christofolletti et al., 2013). No obstante, su uso indiscriminado y sin tratamiento adecuado puede provocar efectos adversos como la salinización del suelo, la alteración de su pH y la contaminación de cuerpos de agua por lixiviación. Estos impactos negativos demandan una evaluación integral de su uso agrícola, considerando también su papel como reservorio de microorganismos y posibles ARGs (Ospina León et al., 2022).

2.2. Impacto ambiental de la vinaza

La aplicación de vinaza como fertilizante ha demostrado mejorar ciertas propiedades pedológicas, como la capacidad de retención de agua y el contenido de materia orgánica. Sin embargo, múltiples estudios alertan sobre sus efectos ecológicos adversos, particularmente en suelos mal drenados o con aplicación excesiva (Leal G. et al., 2003). Uno de los efectos más documentados es la eutrofización de ecosistemas acuáticos, causada por la escorrentía de materia orgánica (Zúñiga Cerón & Gandini Ayerbe, 2013). Asimismo, la disposición continua de vinaza afecta la microbiota del suelo, alterando la diversidad y abundancia de especies microbianas. Estos impactos se agravan en climas tropicales y suelos intensivamente cultivados, como los del litoral ecuatoriano. Allí, la aplicación repetida de vinaza sin medidas de mitigación podría derivar en un efecto acumulativo que comprometa la calidad del suelo a largo plazo y fomente la selección positiva de microorganismos resistentes, afectando tanto la salud del ecosistema como la seguridad alimentaria (CINCAE, 2018).

2.3. Genes de resistencia a antibióticos (ARGs)

Los ARGs son secuencias genéticas que permiten a las bacterias sobrevivir a la exposición a antibióticos y se encuentran presentes de forma natural en el ambiente. No obstante, actividades humanas como el uso de antibióticos en agricultura e industria alimentaria han incrementado su abundancia y diseminación (Lemos et al., 2021). Estos genes pueden provenir de bacterias patógenas a través de transferencia horizontal, exacerbando la crisis de salud pública asociada a infecciones resistentes (von Wintersdorff et al., 2016).

La literatura ha reportado ARGs en diversos ambientes, incluyendo aguas residuales, suelos agrícolas, y residuos industriales como la vinaza (Jian et al., 2021). Su presencia en estos sistemas plantea la preocupación de que puedan ser transferidos a bacterias

patógenas humanas, aumentando la prevalencia de infecciones intratables (Jian et al., 2021; Skandalis et al., 2021). El estudio de ARGs en ambientes agrícolas es, por tanto, esencial para prevenir su entrada en la cadena alimentaria tal y como se aprecia en la Figura 1.

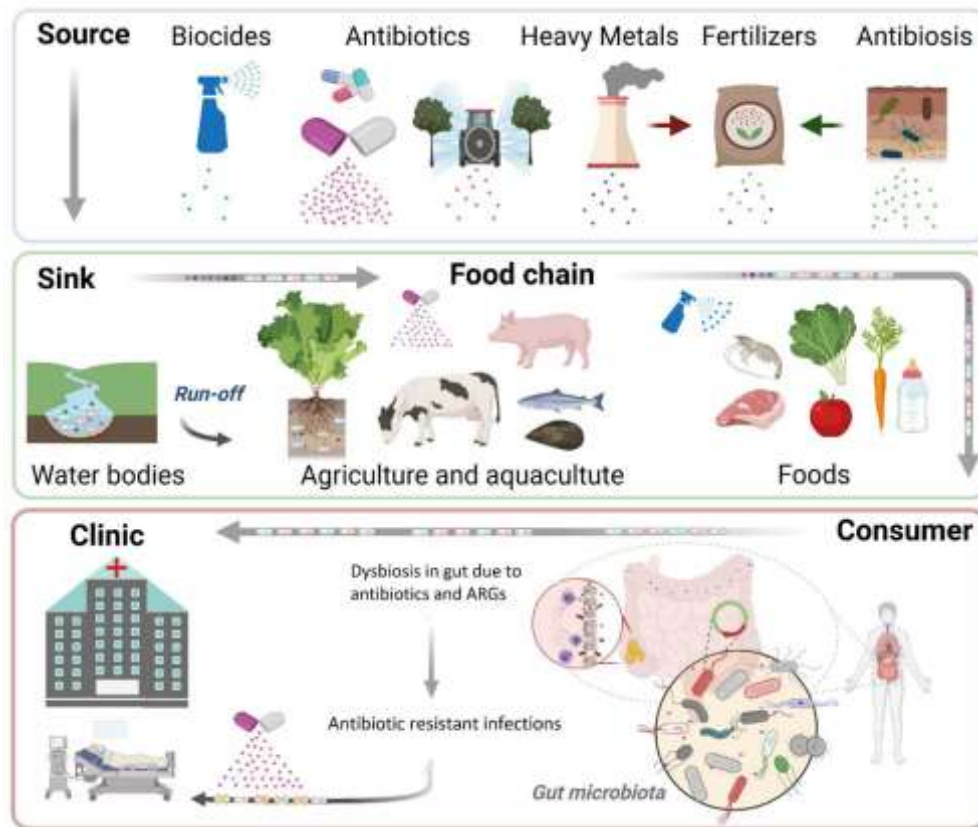


Figura 1. Flujo selectivo de ARG del medio ambiente a la clínica (Skandalis et al., 2021)

La resistencia a antibióticos de amplio espectro es particularmente preocupante, como las betalactamasas de espectro extendido (ESBLs) y los genes *mcr*, que confieren resistencia a colistina, un antibiótico de último recurso (Mondal et al., 2024). Estos genes han sido hallados en ambientes relacionados con la producción agrícola y en residuos como la vinaza, lo que refuerza la necesidad de vigilancia activa y estudios genómicos detallados (Luo et al., 2020).

Entre otros genes de resistencia que demandan especial atención se encuentra el gen *lnu(A)*, el cual codifica una lincosamida nucleotidiltransferasa capaz de inactivar antibióticos como la lincomicina mediante adenilación, impidiendo su unión a la subunidad ribosomal 50S (Chang et al., 2022). Este gen ha sido hallado principalmente en plásmidos de estafilococos como *Staphylococcus aureus* y *Staphylococcus lugdunensis*, aislados tanto de humanos como de animales, y se ha asociado a entornos hospitalarios y ganaderos, donde la presión antibiótica puede favorecer su selección y diseminación (Chang et al., 2022; Lozano et al., 2012). Su localización en elementos genéticos móviles como plásmidos multirresistentes sugiere un alto potencial para la transferencia horizontal entre especies bacterianas, incluyendo bacterias ambientales que podrían estar presentes en matrices como la vinaza. Pese a no haber sido reportado directamente en vinaza, la documentación de *lnu(A)* en bacterias asociadas al ambiente agrícola y animal apunta a un riesgo plausible de diseminación en suelos fertilizados con residuos orgánicos, especialmente si estos contienen microbiotas resistentes no caracterizadas (Liu et al., 2024).

De igual manera, el gen *lnu(C)* representa otra variante de lincosamida nucleotidiltransferasa que ha sido identificada en bacterias como *Brachyspira hyodysenteriae*, *Campylobacter coli* y *Streptococcus uberis*, todas asociadas a entornos ganaderos y avícolas (De Luca et al., 2018; W. Li et al., 2023; Petinaki et al., 2008). Este gen ha sido localizado en transposones como MTnSag1, capaces de integrarse en el cromosoma bacteriano, lo que aumenta su estabilidad y potencial de diseminación bajo presión selectiva de antibióticos como la lincomicina, usada comúnmente en medicina veterinaria (W. Li et al., 2023). Diversos estudios han demostrado que bacterias que albergan *lnu(C)* presentan incrementos significativos en las concentraciones mínimas inhibitorias (MIC) de lincomicina, superando los 64 mg/L en algunos casos, lo que indica

una resistencia clínicamente relevante (De Luca et al., 2018). La identificación de *lnu(C)* en bacterias presentes en animales destinados a la producción alimentaria refuerza su relevancia en sistemas agroindustriales. En consecuencia, su posible presencia en matrices residuales como la vinaza, aunque aún no reportada directamente, debe ser considerada un riesgo potencial dada su relación con ambientes agrícolas y su capacidad de diseminación horizontal.

2.4. Transferencia horizontal de genes (THG)

La transferencia horizontal de genes (THG) es un proceso mediante el cual los microorganismos adquieren genes de otros sin necesidad de reproducción. Esta transferencia ocurre a través de mecanismos como la conjugación, transducción y transformación, y es responsable de la rápida propagación de ARGs en poblaciones bacterianas diversas (von Wintersdorff et al., 2016).

La THG está facilitada por elementos genéticos móviles como plásmidos, transposones e integrones. Estos elementos permiten que ARGs se diseminen eficientemente entre bacterias ambientales y patógenas, lo que agrava la problemática de la resistencia antimicrobiana (van Hoek et al., 2011).

Estudios recientes han sugerido que los hotspots ambientales de THG, como los sistemas de tratamiento de aguas servidas o suelos fertilizados con residuos agroindustriales, deben ser monitoreados con tecnologías de vanguardia, como la metagenómica y la secuenciación de genomas completos, para detectar de manera oportuna nuevos eventos de transferencia y evolución de resistencia (Forsberg et al., 2012).

2.5. Metagenómica y su aplicación ambiental

La metagenómica permite estudiar comunidades microbianas completas a partir del ADN total extraído de una muestra ambiental, sin necesidad de cultivo (Thomas et al., 2012).

Esta metodología se ha posicionado como una herramienta fundamental para explorar la biodiversidad microbiana, identificar genes funcionales y rastrear ARGs en diversos ambientes.

Aplicaciones metagenómicas han sido empleadas para analizar matrices complejas como suelos agrícolas, residuos industriales y aguas residuales. En el contexto de la vinaza, estudios como el de Cassman et al. (2018) han utilizado metagenómica para identificar taxones dominantes y rutas metabólicas clave, proporcionando más información de este subproducto agroindustrial.

Adicionalmente, la metagenómica se ha expandido hacia enfoques como el metagenoma ensamblado de genomas (MAGs), que permiten reconstruir genomas completos a partir de muestras ambientales, facilitando una comprensión más detallada de los organismos que portan ARGs y sus características funcionales (Quince et al., 2017).

3. METODOLOGÍA

Para desarrollar la presente investigación se empleó datos metagenómicos en formato FASTQ provenientes del estudio "Genome-resolved metagenomics of sugarcane vinasse bacteria" realizado por Cassman et al. (2018). Dicho estudio recolectó seis muestras de vinaza entre noviembre de 2013 y noviembre de 2014 en una planta productora de alcohol ubicada en São Paulo, Brasil. A partir de estas muestras, se efectuó la extracción y cuantificación del ADN utilizando el kit MoBio PowerSoil, seguido de la secuenciación con Illumina MiSeq, lo que permitió obtener 18 metagenomas shotgun en formato FASTQ (tres réplicas por muestra) cargados en la plataforma MG-RAST y NCBI. Estos metagenomas representaron el punto de partida para identificar genes de resistencia a antibióticos (ARGs) en bacterias observadas en vinaza de caña de azúcar.

El estudio de Cassman et al. (2018) recolectó muestras de vinaza en distintos momentos a lo largo de un año, como se logra ver en la Tabla 1, lo que permite capturar una mayor diversidad microbiana y genética, dado que la vinaza es un subproducto con variabilidad en su composición debido a los cambios en las condiciones de fermentación y procesamiento del bioetanol (Christofolletti et al., 2013). La secuenciación mediante la plataforma Illumina MiSeq proporciona datos de alta precisión y profundidad, facilitando un ensamblaje genómico robusto y la identificación precisa de genes de interés (L. Wu et al., 2015), que involucra genes de resistencia a antibióticos (ARGs). Además, el formato FASTQ preserva la información de calidad de las lecturas, lo que resulta esencial para optimizar el filtrado y mejorar la precisión en los análisis metagenómicos (Saremi et al., 2021).

Tabla 1. Descripción de los datos de los metagenomas de vinaza

Date Sampled	Accession SRA	DNA Weight (g)	DNA Conc. (ng/ul)	# reads	Forward # bases (Mbp)	Reverse # bases (Mbp)	Reads mapped to cross-configs (%)	Percent merged	# merged reads	Avg. merged read length	# merged bases (Mbp)
Nov. 2013	SRX3743992	0.297	32.5	461,784	131	117	92.77	87.98%	236896	592,17	85
	SRX3743991	0.250	39.6	454,721	130	115	92.69	88.72%	230245	592,38	83
	SRX3743990	0.295	20.4	469,527	130	115	92.19	87.40%	245213	590,59	83
Dec. 2013	SRX3743989	0.300	39.2	417,625	110	94	73.20	93.07%	263198	582,17	86
	SRX3743988	0.295	53.1	517,039	142	131	73.67	95.62%	359330	580,67	116
	SRX3743987	0.292	38.5	542,208	150	139	74.67	95.15%	376765	581,71	124
July. 2014	SRX3743986	0.301	13.9	362,499	100	89	92.38	88.74%	193318	590,16	64
	SRX3743985	0.267	14.2	501,511	138	123	92.53	91.90%	275242	589,16	91
	SRX3743984	0.291	14.4	432,207	119	107	92.12	91.96%	252621	588,15	83
Aug. 2014	SRX3743983	0.295	17.2	489,336	138	132	95.12	95.97%	402629	581,89	139
	SRX3743982	0.291	18.6	280,161	77	74	95.33	95.88%	235413	577,5	79
	SRX3743981	0.294	21.0	351,407	971	935	95.3	96.49%	300148	576,57	100
Oct. 2014	SRX3743980	0.294	6.19	363,382	981	914	91.30	94.29%	258649	584,54	84
	SRX3743979	0.280	5.57	434,111	117	108	91.18	92.39%	293786	585,17	96
	SRX3743978	0.290	7.22	472,732	135	124	91.47	91.44%	305881	589,63	111
Nov. 2014	SRX3743977	0.294	7.29	444,056	122	114	91.14	93.26%	320201	586,67	109
	SRX3743976	0.285	6.64	500,636	136	128	91.29	95.37%	383266	583,43	126
	SRX3743975	0.292	5.53	323,376	883	794	91.66	84.59%	269124	575,92	94

Obtenida de: Cassman et al. (2018).

Este estudio se desarrolló bajo un enfoque descriptivo y exploratorio utilizando análisis metagenómico *in silico* para identificar genes de resistencia a antibióticos (ARGs) presentes en vinaza de caña de azúcar (*Saccharum officinarum*). El análisis se realizó mediante una serie de herramientas bioinformáticas alojadas en la plataforma Galaxy Australia (<https://usegalaxy.org.au/>), reconocida por su capacidad para realizar análisis reproducibles y accesibles de datos genómicos y metagenómicos (Afgan et al., 2018).

3.1. Obtención de datos de secuenciación

Se emplearon 18 conjuntos de datos metagenómicos disponibles abiertamente en el repositorio [NCBI Sequence Read Archive \(SRA\)](https://www.ncbi.nlm.nih.gov/sra/). Cada conjunto fue gestionado de manera individual en la plataforma Galaxy, creando un historial independiente por metagenoma, identificado con su correspondiente número de acceso SRA. Para la descarga de los datos, se utilizó la herramienta "Download and Extract Reads in FASTQ format from NCBI SRA", seleccionando el formato comprimido FASTQ Sanger (.fastq.gz). Este proceso permitió recuperar las secuencias pareadas (paired-end reads),

separadas en archivos correspondientes a las lecturas forward y reverse, generando las bibliotecas necesarias para los análisis metagenómicos posteriores.

3.2. Control de calidad de lecturas

Para llevar a cabo el control de calidad de las secuencias metagenómicas, se empleó la herramienta FastQC (v0.74+galaxy1), la cual permite evaluar parámetros críticos como la calidad base por posición, contenido de GC, presencia de adaptadores y distribución de longitud de lecturas (Babraham Institute, 2020). El análisis se aplicó de forma individual a las secuencias forward y reverse. Dado que la herramienta no requiere configuración avanzada, bastó con seleccionar como entrada los archivos FASTQ correspondientes a las secuencias forward y reverse previamente descargadas desde el NCBI SRA. Los resultados se visualizaron mediante reportes interactivos en formato HTML, que contienen gráficos detallados por módulo de evaluación.

3.3. Filtrado y recorte de lecturas

Se empleó la herramienta Trimmomatic (v0.36.6) para realizar el filtrado y recorte de lecturas, eliminando bases de baja calidad y asegurando la integridad de los datos metagenómicos. Se seleccionó el modo de procesamiento "Paired-end (two separate input files)", utilizando como entrada los archivos FASTQ correspondientes a las lecturas forward (R1) y reverse (R2). Se desactivó la etapa inicial de recorte de adaptadores ("ILLUMINACLIP"), dado que las secuencias no incluían adaptadores residuales (Anexo 2). Se aplicó la operación "SLIDINGWINDOW", configurada con una ventana de 4 bases y un umbral de calidad promedio mínima de 30, lo que garantiza la retención de regiones de alta fidelidad en las secuencias (Bolger et al., 2014). No se generaron archivos de log ni trimlog adicional. Para asegurar la calidad de los datos downstream, solo se conservaron las secuencias emparejadas ("R1 paired" y "R2 paired"), descartando aquellas no alineadas.

Posteriormente, se realizó una nueva evaluación de calidad sobre los archivos de salida generados por Trimmomatic, utilizando la herramienta FastQC (v0.74+galaxy1) bajo los mismos parámetros previamente descritos para los archivos crudos. Este análisis se realizó tanto para las lecturas forward como reverse de cada uno de los 18 metagenomas y permitió verificar la eficacia del proceso de recorte y confirmar la mejora de los parámetros de calidad en las secuencias depuradas.

Los resultados individuales de FastQC fueron integrados y visualizados mediante la herramienta MultiQC (v1.27+galaxy3), lo que facilitó una revisión consolidada de los 36 informes generados (18 muestras, cada una con lecturas forward y reverse) (Ewels et al., 2016). MultiQC fue ejecutado especificando FastQC como herramienta fuente y Raw data como tipo de entrada.

3.4. Clasificación taxonómica

Para asignar taxonómicamente las secuencias metagenómicas filtradas, se hizo uso de la herramienta Kraken2 (v2.1.3+galaxy1), la cual clasifica lecturas a partir del análisis de k-mers exactos en comparación con una base de datos de referencia, utilizando un enfoque basado en árboles de decisión para determinar la taxonomía más probable (Wood et al., 2019). En este análisis se trabajó con lecturas emparejadas (paired-end), correspondientes a los archivos de salida de Trimmomatic, y se utilizó la base de datos MiniKraken2 v2 (8 GB, versión 2020-02-05).

Entre los parámetros establecidos, se mantuvieron los valores predeterminados para la calidad mínima por base (0), el número mínimo de grupos de aciertos (minimum hit groups: 2) y se desactivaron funciones adicionales como la división de lecturas clasificadas/no clasificadas o el cálculo de confianza probabilística, priorizando la rapidez de ejecución. Además, se configuró la opción para mostrar únicamente identificadores taxonómicos (taxIDs) en lugar de nombres científicos.

La salida de Kraken2 fue posteriormente transformada mediante la herramienta Convert Kraken data to Galaxy taxonomy representation (v1.2+galaxy0), que permite adaptar los resultados a una representación jerárquica interpretable dentro de la plataforma Galaxy (Wood & Salzberg, 2014). En este paso, se seleccionó explícitamente la base de datos NCBI Taxonomy (versión 2024-06-05) como referencia taxonómica.

Una vez obtenida la clasificación taxonómica con Kraken2, se utilizó la herramienta Kraken report (v1.3.1) para generar un informe estructurado de abundancias taxonómicas jerárquicas (Wood & Salzberg, 2014). En esta etapa, se empleó como entrada la salida de clasificación de Kraken2 y se mantuvo la base de datos seleccionada como MiniKraken2.

Para facilitar la exploración interactiva de los resultados, se utilizó la herramienta Krona (v2.7.1+galaxy0), que genera gráficos circulares jerárquicos navegables en formato HTML (Ondov et al., 2011). En este paso, se indicó que los datos de entrada correspondían a "taxonomy", se seleccionó el archivo generado por la conversión previa con Kraken2, y se especificó la visualización hasta el nivel de subespecie mediante la opción "show ranks from root to: Subspecies". Además, se estableció el nombre "Root" como categoría basal del árbol taxonómico.

3.5. Estimación de abundancia

Para refinar las estimaciones de abundancia taxonómica obtenidas tras la clasificación con Kraken2, se utilizó la herramienta Bracken (Bayesian Reestimation of Abundance with Kraken) (v3.1+galaxy0). En lugar de procesar directamente el archivo de salida cruda de Kraken2, Bracken requiere como entrada el archivo estructurado generado por Kraken Report, el cual resume los conteos taxonómicos jerárquicamente por niveles taxonómicos (Lu et al., 2017).

Este análisis, se efectuó a partir de la base de datos Prebuilt Refseq indexes: Standard-Full (Version: 2022-06-07, K-mer: 35, Read: 100), una referencia precompilada ampliamente reconocida por su cobertura integral de secuencias genómicas completas de bacterias, arqueas y virus, lo que permite una clasificación taxonómica más precisa y específica a nivel de especie, además es particularmente adecuada para estudios metagenómicos basados en lecturas shotgun, ya que incluye genomas completos bien anotados (O’Leary et al., 2016; Wood et al., 2019). El análisis se realizó a nivel taxonómico de especie, permitiendo hasta 10 desajustes en la coincidencia de k-mers, lo cual ofrece un equilibrio adecuado entre sensibilidad y precisión.

3.6. Ensamblaje metagenómico

Partiendo del análisis taxonómico exploratorio utilizando las herramientas Kraken 2, Bracken y Krona (Galaxy), se seleccionó una muestra representativa para validar metodológicamente el flujo de ensamblaje y binning. En función de la abundancia relativa de taxones, se eligió el metagenoma con mayor presencia de la familia *Lactobacillaceae* para llevar a cabo un análisis posterior dentro de la plataforma KBase, que ofrece herramientas de evaluación y refinamiento no disponibles en Galaxy Australia.

En este análisis, se realizó un ensamblaje del metagenoma seleccionado en la plataforma Kbase utilizando las herramientas MEGAHIT (v1.2.9) y metaSPAdes (v3.15.3). Posteriormente, se aplicó QUASt (v4.4) para evaluar métricas de calidad de los ensamblajes (Gurevich et al., 2013), y adicional se utilizó la herramienta Compare Assembled Contig Distributions (v1.1.2) (Arkin et al., 2018) con el fin de comparar gráficamente la distribución de longitudes de contigs generados por cada ensamblado y con base en estos resultados se definió cuál sería el ensamblador a utilizar de forma estandarizada para las 18 muestras.

A partir de esta validación metodológica, el ensamblaje de los 18 metagenomas se realizó en Galaxy Australia, utilizando MEGAHIT (v1.2.9+galaxy2), diseñada específicamente para el ensamblaje eficiente de metagenomas de gran escala mediante la construcción de grafos de Bruijn optimizados (D. Li et al., 2015). Se seleccionó la opción de ensamblaje con lecturas "paired-end", utilizando como entradas las secuencias forward (R1 paired) y reverse (R2 paired) previamente depuradas mediante Trimmomatic.

En cuanto a los parámetros técnicos, se definió un conjunto de tamaños de k-mers específicos (21, 29, 39, 59, 79, 99, 119, 141), lo que permite capturar variaciones estructurales en fragmentos de diferentes longitudes y complejidades y se estableció una multiplicidad mínima de k-mers de 2, filtrando así secuencias poco representadas que podrían corresponder a errores de secuenciación (D. Li et al., 2015). Además, se configuró una intensidad de fusión de burbujas en el valor 2 y una longitud mínima de burbujas complejas de 20 con una cobertura mínima del 95 %, con el fin de mejorar la resolución de estructuras ambiguas en el grafo.

Para el filtrado posterior al ensamblaje, se definió una longitud mínima de contigs de 200 pb, asegurando que únicamente fragmentos con potencial informativo fueran retenidos para el análisis. Asimismo, se aplicaron estrategias automáticas de poda y desconexión de unitigs basadas en cobertura local, con el fin de eliminar ramas ambiguas o artefactos derivados de errores de secuenciación o baja cobertura, lo que mejora la calidad del grafo de ensamblaje (D. Li et al., 2015).

3.7. Agrupamiento de contigs en bins y evaluación de bins

Previo al binning metagenómico de los 18 metagenomas, se llevó a cabo una comparación metodológica del desempeño de distintas herramientas de agrupamiento. Esta validación se realizó en la plataforma KBase, donde se aplicaron las herramientas MetaBAT2 y MaxBin2 sobre el ensamblaje representativo generado con el metagenoma seleccionado

en el paso anterior, con el objetivo de obtener el algoritmo más adecuado para su posterior aplicación en el procesamiento global de muestras.

La evaluación de la calidad de los bins generados en esta etapa se efectuó mediante CheckM, herramienta que permite estimar métricas como la completitud y la contaminación de los bins a partir de la detección de genes marcadores y su redundancia (Parks et al., 2015). A partir de este análisis metodológico, se estableció el uso de MetaBAT2 como herramienta estándar para el binning de los 18 ensamblajes metagenómicos, procesados en la plataforma Galaxy Australia.

Los contigs previamente ensamblados con MEGAHIT (v1.2.9+galaxy2) fueron utilizados como entrada para MetaBAT2 (v2.17+galaxy0), la cual permite realizar el binning metagenómico, es decir, la agrupación de contigs en unidades genómicas tentativas (bins) basadas en parámetros como el contenido de GC, la covarianza de cobertura entre muestras y patrones tetranucleotídicos (TNF) (Kang et al., 2019).

Para la ejecución, se seleccionó como entrada el archivo de ensamblaje generado previamente, manteniendo los parámetros establecidos por defecto. Un tamaño mínimo de contig de 1500 pb, con el objetivo de filtrar secuencias demasiado cortas que pudieran afectar la precisión del agrupamiento. Un porcentaje del 95 % de contigs considerados "buenos" para su evaluación, y se utilizó un umbral de score mínimo de 60 para considerar conexiones significativas entre nodos. Además, se configuraron límites de profundidad mínima de cobertura promedio (1x) tanto por muestra como en promedio global.

Los bins generados serán útiles para la identificación de genes de resistencia a antibióticos.

3.8. Identificación de genes de resistencia a antibióticos

Para efectuar la detección de genes de resistencia antimicrobiana en los bins genómicos obtenidos mediante MetaBAT2 de los 18 metagenomas, se empleó la herramienta ABRicate (v1.0.1), una plataforma ampliamente utilizada para el análisis masivo de secuencias genómicas frente a múltiples bases de datos especializadas en resistencia (Seemann, 2016). Como entrada, se seleccionaron las secuencias ensambladas en formato FASTA provenientes de la salida “BIN SEQUENCES” de MetaBAT2.

La búsqueda se realizó en dos fases, la primera utilizando ResFinder y la segunda utilizando CARD.

ResFinder, desarrollada por el Center for Genomic Epidemiology, misma que está orientada específicamente a la identificación de genes de resistencia adquiridos en bacterias clínicamente relevantes. Esta base de datos se renueva periódicamente y ha demostrado alta sensibilidad en estudios epidemiológicos y ambientales (Florensa et al., 2022).

CARD (Comprehensive Antibiotic Resistance Database), un repositorio curado y respaldado por evidencia experimental, incluye tanto genes adquiridos como mutaciones cromosómicas que confieren resistencia. Su enfoque combina anotaciones computacionales con pruebas fenotípicas validadas, lo cual la convierte en una base sólida para el análisis genómico comparativo (Alcock et al., 2023).

Tanto ResFinder como CARD fueron seleccionadas por su robustez y complementariedad: mientras ResFinder proporciona un enfoque centrado en bacterias patógenas conocidas, CARD amplía la cobertura a mecanismos emergentes de resistencia y estructuras móviles.

Los parámetros utilizados en ABRicate se configuraron de manera conservadora para minimizar falsos positivos. Se estableció un margen mínimo de 90 % de similitud y 90 % de cobertura del gen, lo cual garantiza la especificidad en la asignación de los genes detectados (Seemann, 2016). No se suprimió el encabezado de salida (“Suppress header = false”), de forma que cada reporte incluyera metadatos relevantes sobre el análisis.

3.9. Anotación del genoma

A partir del metagenoma con mayor presencia de bacterias de la familia *Lactobacillaceae* se obtuvo los bins. Dichos bins fueron obtenidos en la plataforma KBase, tras el proceso de binning aplicado con las herramientas MetaBAT2 y MaxBin2, seguido de un refinamiento mediante DAS Tool (Sieber et al., 2018). Este proceso permitió obtener unidades genómicas tentativas optimizadas, cuya calidad fue evaluada mediante CheckM (Parks et al., 2015).

Una vez definidos los bins de interés, se procedió a su extracción individual mediante la herramienta BinUtil: Extract Bins as Assemblies from BinnedContigs (v1.0.2) (Arkin et al., 2018), disponible en KBase. Posteriormente, se seleccionó el bin perteneciente a la familia *Lactobacillaceae* y se realizó la anotación estructural de cada uno utilizando la herramienta Annotate Multiple Microbial Assemblies with RASTtk (v1.073) (Brettin et al., 2015), implementada en la plataforma KBase.

La herramienta Annotate Multiple Microbial Assemblies with RASTtk (v1.073) permite identificar y caracterizar elementos codificantes y no codificantes en genomas bacterianos mediante un enfoque modular y automatizado (Brettin et al., 2015). Para la ejecución, se seleccionó el dominio biológico Bacteria y se utilizó el código genético 11, correspondiente a arqueas, la mayoría de bacterias y mitocondrias. Se habilitaron los predictores estándar para genes codificantes (glimmer3 y prodigal), así como la detección de elementos como ARNr, ARNt, selenoproteínas, proteínas con pirrolisina, repeticiones

y regiones CRISPR. La anotación funcional se realizó mediante algoritmos basados en similitud proteica y k-mer (v1 y v2).

4. RESULTADOS Y ANÁLISIS DE RESULTADOS

4.1. Obtención de datos de secuenciación

Se recuperaron un total de 18 metagenomas crudos desde la base de datos pública del NCBI Sequence Read Archive (SRA), organizados en seis muestras con tres réplicas cada una. Cada conjunto de datos consistió en secuencias paired-end en formato .fastq.gz, correspondientes a lecturas forward y reverse obtenidas mediante tecnología de secuenciación Illumina.

4.2. Control de calidad de lecturas

La herramienta FastQC generó un total de 36 informes correspondientes a las secuencias forward y reverse de los 18 metagenomas procesados. En la Tabla 2 se resumen los principales indicadores de calidad de las estadísticas básicas para cada archivo, incluyendo el número total de secuencias, el número de bases, la longitud de la secuencia, el porcentaje de contenido GC y la cantidad de secuencias marcadas con baja calidad.

El número de secuencias por archivo varió entre 287,232 y 554,542, con un rango de bases totales entre 79.8 Mbp y 157.9 Mbp. En todos los casos, no se detectaron secuencias catalogadas con una calidad baja. El tamaño de lectura se mantuvo constante entre 35 y 301 nucleótidos, con la excepción de dos archivos (SRX3743979 forward y reverse), que presentaron lecturas a partir de 1 nt. El contenido GC mostró una variación entre 39 % y 53 %.

Tabla 2. Métricas de calidad de las secuencias crudas obtenidas para los metagenomas analizados

Muestra	Accession SRA	Tipo de lectura	Secuencias totales	Bases totales	Secuencias marcadas de baja calidad	Longitud de la secuencia	%GC
Nov. 2013	SRX3743992	Forward	477589	139.3 Mbp	0	35-301	51
		Reverse	477589	140.3 Mbp	0	35-301	53
	SRX3743991	Forward	468948	136.9 Mbp	0	35-301	52
		Reverse	468948	138.0 Mbp	0	35-301	53
	SRX3743990	Forward	484423	137.3 Mbp	0	35-301	52
		Reverse	484423	138.4 Mbp	0	35-301	53
Dic. 2013	SRX3743989	Forward	425155	119.3 Mbp	0	35-301	49
		Reverse	425155	120.8 Mbp	0	35-301	50
	SRX3743988	Forward	528640	147.7 Mbp	0	35-301	48
		Reverse	528640	149.5 Mbp	0	35-301	49
	SRX3743987	Forward	554542	156.2 Mbp	0	35-301	47
		Reverse	554542	157.9 Mbp	0	35-301	48
Jul. 2014	SRX3743986	Forward	374644	105.5 Mbp	0	35-301	51
		Reverse	374644	106.5 Mbp	0	35-301	53
	SRX3743985	Forward	517219	145.9 Mbp	0	35-301	52
		Reverse	517219	147.4 Mbp	0	35-301	53
	SRX3743984	Forward	444597	125.2 Mbp	0	35-301	52
		Reverse	444597	126.3 Mbp	0	35-301	53
Ago. 2014	SRX3743983	Forward	503293	142.7 Mbp	0	35-301	39
		Reverse	503293	143.7 Mbp	0	35-301	39
	SRX3743982	Forward	486616	141 Mbp	0	35-301	45
		Reverse	486616	142 Mbp	0	35-301	46
	SRX3743981	Forward	456271	127 Mbp	0	35-301	45
		Reverse	456271	127.8 Mbp	0	35-301	45
Oct. 2014	SRX3743980	Forward	512965	141.1 Mbp	0	35-301	44
		Reverse	512965	142 Mbp	0	35-301	45
	SRX3743979	Forward	326760	91.2 Mbp	0	1-301	44
		Reverse	326760	89.5 Mbp	0	1-301	45
	SRX3743978	Forward	287232	79.8 Mbp	0	35-301	39
		Reverse	287232	80.4 Mbp	0	35-301	39
Nov. 2014	SRX3743977	Forward	359268	100 Mbp	0	35-301	39
		Reverse	359268	100.6 Mbp	0	35-301	39
	SRX3743976	Forward	372983	102.6 Mbp	0	35-301	46
		Reverse	372983	103.5 Mbp	0	35-301	47
	SRX3743975	Forward	446246	122.5 Mbp	0	35-301	46
		Reverse	446246	123.7 Mbp	0	35-301	47

4.3. Filtrado y recorte de lecturas

Los resultados de una nueva evaluación de calidad utilizando FastQC y MultiQC evidenciaron que el procesamiento mejoró y conservó de forma adecuada la calidad general de las secuencias (Anexo 1).

En la Figura 2 se muestra la métrica de calidad promedio por posición de base de los 36 archivos. Se observa que la mayoría de las posiciones a lo largo de las secuencias mantienen puntuaciones Phred superiores a 34, dentro del rango de alta calidad (zona verde), confirmando una mejora significativa en la calidad general. Esta etapa es crítica para evitar errores en la clasificación taxonómica y reducir la ambigüedad en los ensamblajes (Bolger et al., 2014).

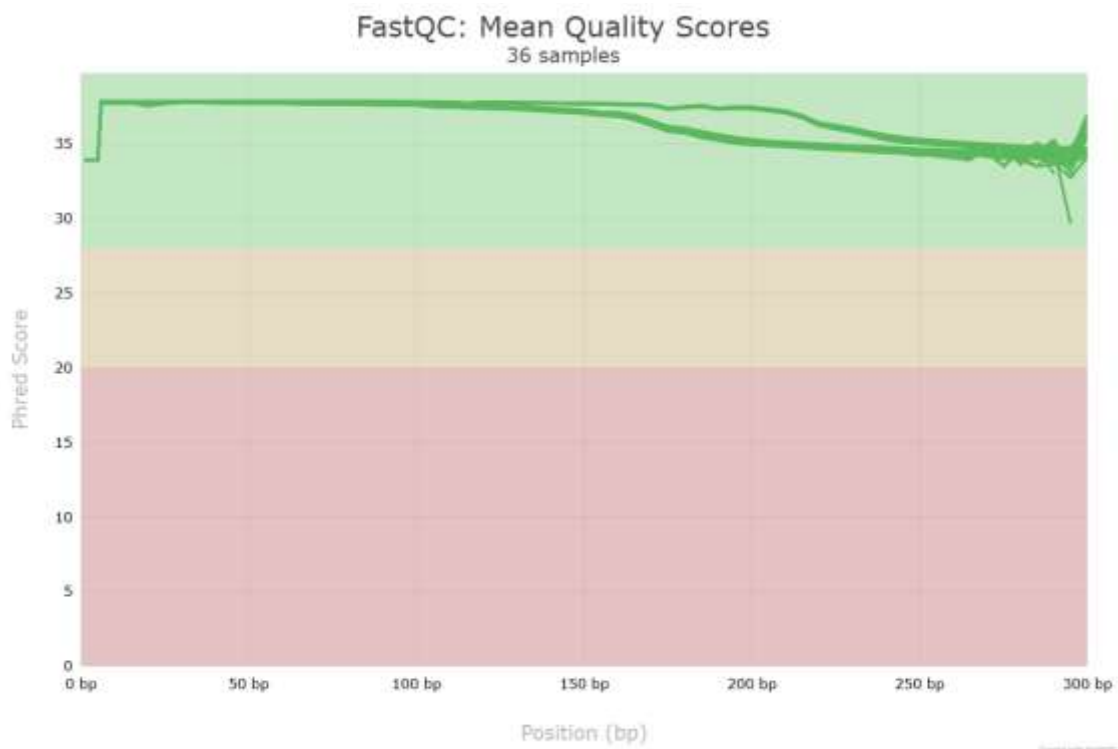


Figura 2. Calidad promedio por posición de base en las secuencias filtradas

En la Figura 3 se aprecia una aproximación del número de secuencias por archivo y el nivel de duplicación. Las muestras presentaron un número de lecturas que varió entre aproximadamente 280,000 y 500,000, manteniéndose dentro del rango esperado tras el

filtrado. La gran mayoría de las lecturas son únicas (en azul claro), mientras que la fracción de lecturas duplicadas (en negro) es baja, lo que representa que hay un sesgo mínimo en la amplificación o secuenciación (Thomas et al., 2012).

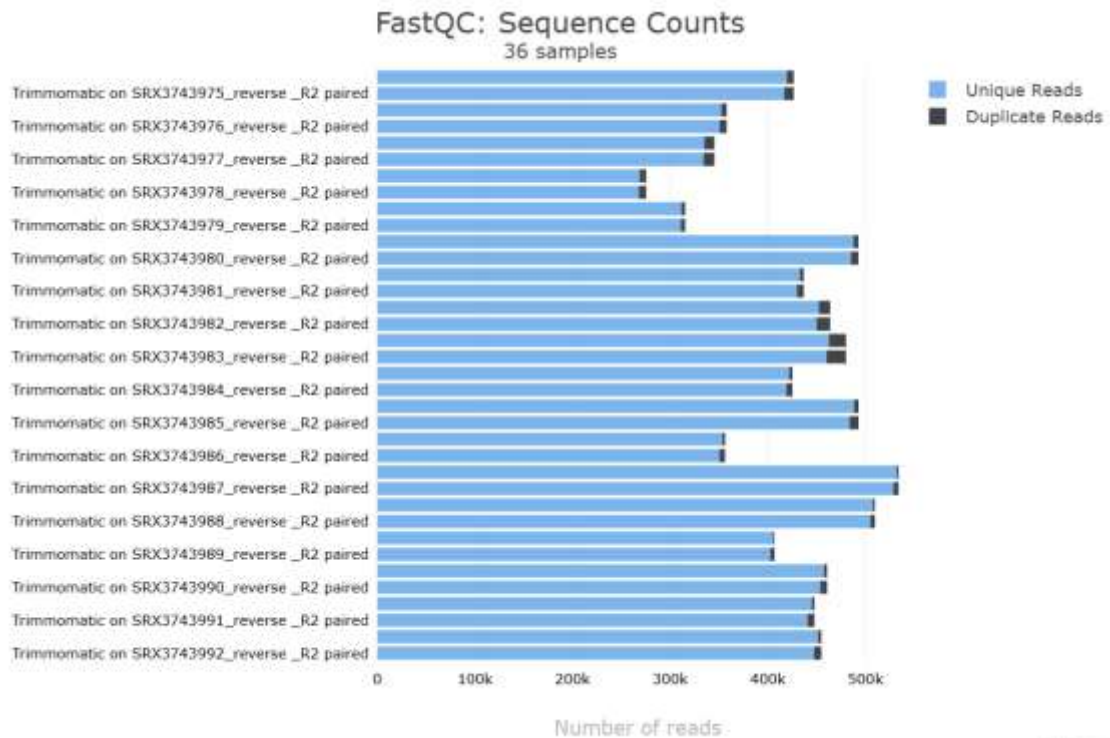


Figura 3. Cantidad de lecturas únicas y duplicadas por archivo

En la Figura 4 se muestra el contenido de GC por secuencia. Se pueden identificar varios perfiles de distribución, con picos principales entre el 40 % y el 55 % de GC. La distribución del contenido de GC observada en las muestras está dentro del rango común reportado para comunidades bacterianas ambientales (20 % - 70 %) y las variaciones entre muestras pueden reflejar diferencias potenciales en la composición taxonómica (Hildebrand et al., 2010).

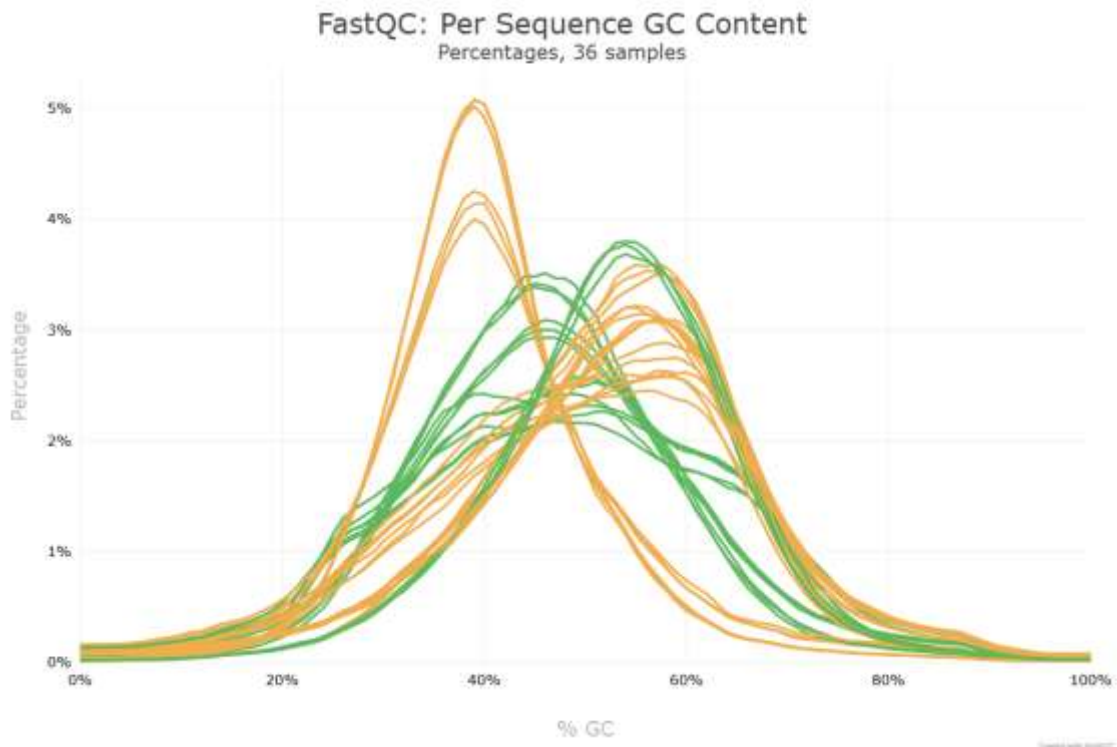


Figura 4. Distribución del contenido GC por secuencia

En la figura, las curvas verdes representan distribuciones de contenido GC consideradas normales según FastQC, caracterizadas por ser unimodales, simétricas y centradas, lo que indica una composición microbiana relativamente homogénea. En cambio, las curvas naranjas muestran desviaciones respecto a la distribución esperada, interpretadas como advertencias que podrían reflejar una mayor diversidad taxonómica, presencia de contaminantes o artefactos técnicos. La morfología de las curvas también ofrece pistas sobre la estructura comunitaria: distribuciones leptocúrticas (estrechas y concentradas) suelen asociarse a comunidades dominadas por pocos taxones, mientras que curvas platicúrticas (anchas y aplanadas) sugieren mayor heterogeneidad microbiana (Babraham Institute, 2020; HBC Training, n.d.).

4.4. Clasificación taxonómica

Para la representación de los perfiles taxonómicos derivados del análisis con Kraken2, se seleccionó una muestra representativa por cada una de las seis muestras analizadas,

considerando entre las tres repeticiones técnicas de cada muestra aquella que presentó mayor riqueza observable en los gráficos generados por Krona. No obstante, para un análisis más detallado, en el área de Anexos se encuentran los gráficos de Krona correspondientes a cada metagenoma.

En la Figura 5 se muestra el diagrama circular de Krona correspondiente a la primera muestra representativa (SRX3743992), que pertenece a la repetición 1 de la muestra 1. Se identificó una clara prevalencia del dominio *Bacteria*, sin presencia detectable de otros dominios. Dentro de este grupo, el filo *Firmicutes* fue el más abundante, abarcando aproximadamente el 42 % de las lecturas clasificadas. En su interior, destacó la clase *Negativicutes*, orden *Selenomonadales*, familia *Veillonellaceae*, donde predominó el género *Megasphaera*, siendo *Megasphaera elsdenii* la especie con mayor representación en esta muestra.

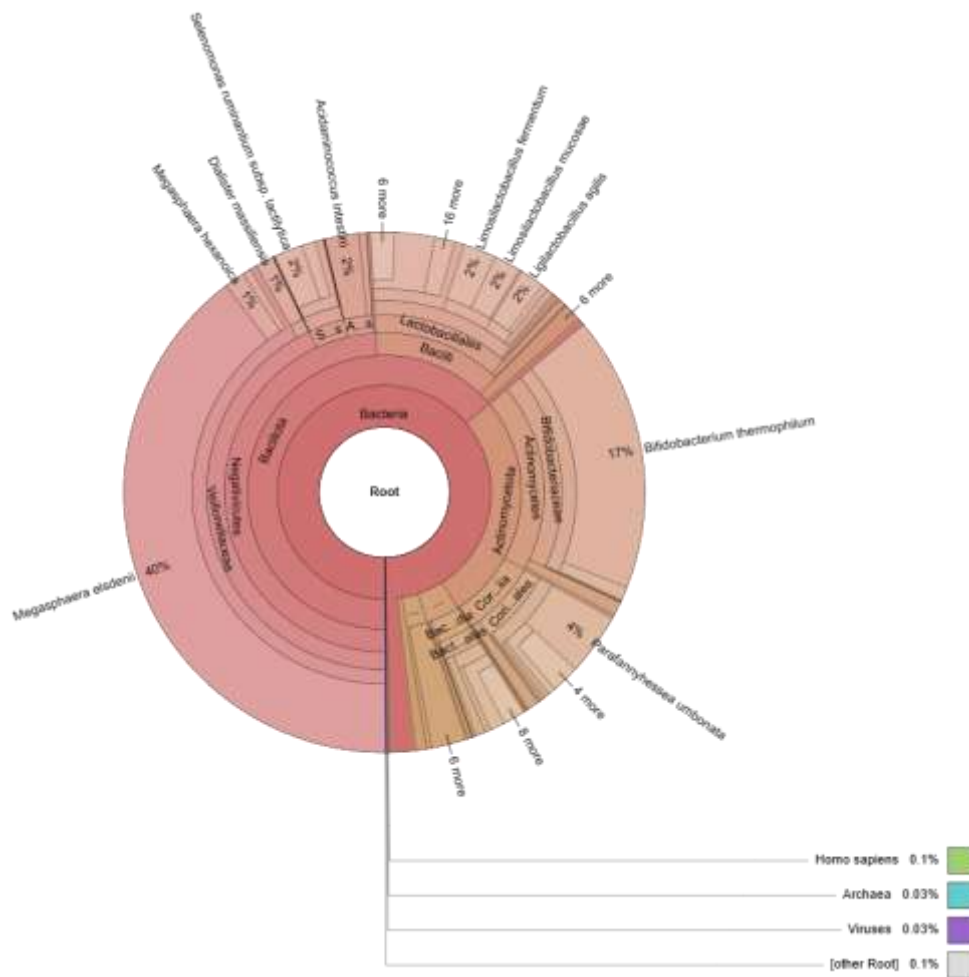


Figura 5. Diagrama circular de Krona del metagenoma SRX3743992

La presencia de *Megasphaera elsdenii* en las muestras de vinaza resulta microbiológica y ecológicamente relevante. Esta especie, es ampliamente reconocida por su papel funcional en el rumen como fermentadora de lactato y estabilizadora del pH en dietas de alta carga energética (Cabral & Weimer, 2024), sin embargo, presenta características genéticas que ameritan seguimiento desde la perspectiva de la resistencia antimicrobiana. Algunas cepas han demostrado portar plásmidos pequeños con genes de resistencia a antibióticos de uso frecuente en producción animal, como tetraciclinas y macrólidos. En ambientes como la vinaza, un residuo orgánico denso en biomasa microbiana, estas bacterias pueden integrarse a comunidades edáficas tras su aplicación al suelo, contribuyendo a la persistencia y circulación de genes de resistencia a través de

mecanismos de transferencia horizontal. Según ha señalado (Topp et al., 2013), la exposición prolongada de suelos agrícolas a residuos con carga antimicrobiana puede acelerar la adaptación microbiana y favorecer la diseminación de resistencia en ecosistemas productivos. De esta manera, *Megasphaera elsdenii* además de su valor funcional en la industria agropecuaria, adquiere un interés particular como microorganismo modelo para el estudio de las interacciones entre salud animal, prácticas agrícolas y diseminación ambiental de resistencia antimicrobiana.

El segundo filo más abundante fue *Actinobacteria*, con aproximadamente un 18 % de las secuencias, dentro del cual se identificó principalmente el género *Bifidobacterium*, perteneciente a la familia *Bifidobacteriaceae*. A nivel de especie, sobresalió *Bifidobacterium pseudolongum*, bacteria que conforma la microbiota intestinal habitual de animales de producción, especialmente ganado vacuno y porcino, y se ha detectado que *B. pseudolongum* pueden portar resistencias atípicas a antibióticos como clindamicina, eritromicina y tetraciclina, apuntando a que podrían actuar como reservorios silenciosos de genes de resistencia (Mayrhofer et al., 2007).

Se registraron también otros filos en menor proporción, como *Proteobacteria*, *Bacteroidetes* y *Cyanobacteria*. Otras especies detectadas en baja abundancia incluyeron *Oscillibacter valericigenes*, *Clostridium beijerinckii*, *Lactobacillus reuteri*, *Selenomonas sputigena* y *Acidaminococcus fermentans*, todas con representaciones inferiores al 5 % del total.

En el estudio de Cassman et al. (2018), también se reportó de forma inesperada la presencia de bacterias típicamente asociadas al intestino de rumiantes y bovinos, pertenecientes a los géneros *Megasphaera*, *Mitsuokella* y *Bifidobacterium*, en los análisis de metagenomas de vinaza. Estos microorganismos no habían sido identificados previamente como contaminantes comunes en tuberías de fermentación y aún se

desconoce si estos taxones interactúan funcionalmente entre sí en la matriz de vinaza, ni se ha determinado la dirección o naturaleza de tales interacciones.

La Figura 6 corresponde al metagenoma SRX3743987, la réplica 3 de la muestra 2, en este diagrama de Krona se observa a nivel general un dominio exclusivo del reino *Bacteria*, dentro del cual se distribuyen múltiples taxones. Entre los filos predominantes destacan *Proteobacteria*, que representa el 34 % del total, seguido por *Firmicutes* (24 %) y en menor proporción *Actinobacteria*.

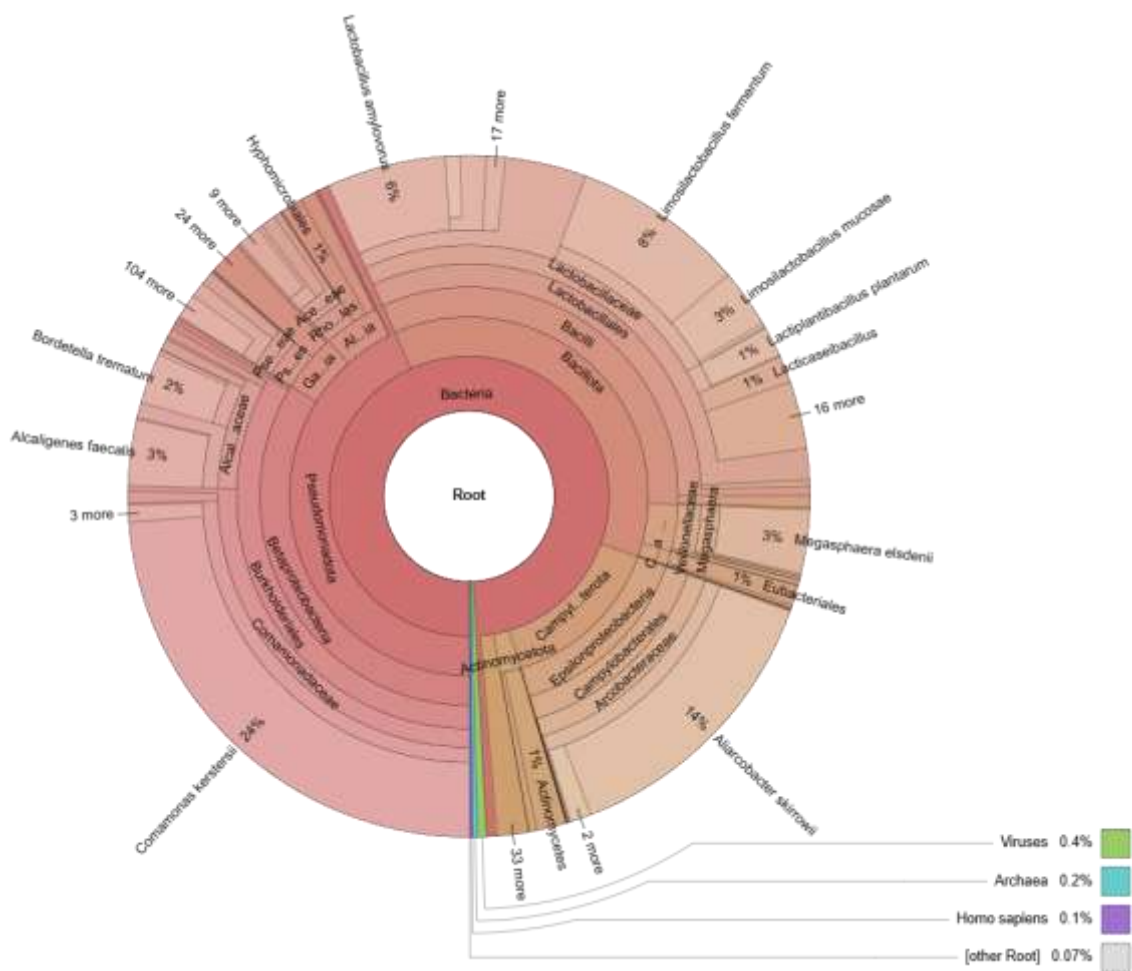


Figura 6. Diagrama circular de Krona del metagenoma SRX3743987

En cuanto a la composición taxonómica interna, dentro de *Proteobacteria*, la clase *Betaproteobacteria* es la más representativa, seguida por diversos órdenes como

Burkholderiales. Dentro de *Firmicutes*, se evidencian principalmente miembros de la clase *Bacilli*, especialmente del orden *Lactobacillales*.

En este metagenoma, la especie con mayor representación fue *Comamonas kerstersii*, perteneciente a la familia *Comamonadaceae* del filo *Proteobacteria*. La detección de *Comamonas kerstersii* en vinaza plantea importantes consideraciones clínicas y ambientales. Esta especie ha sido históricamente subestimada, sin embargo, casos recientes la han implicado en infecciones graves como peritonitis y apendicitis perforada, especialmente en pacientes inmunocomprometidos (Farfán-Cano et al., 2020). Su creciente reconocimiento como patógeno coincide con una alta adaptabilidad ecológica, lo que explica su presencia en residuos orgánicos como la vinaza, rica en nutrientes y posible carga antimicrobiana. Aplicada al suelo mediante fertirrigación, esta matriz podría facilitar la incorporación de *C. kerstersii* en comunidades microbianas edáficas, favoreciendo la persistencia de determinantes de resistencia a antibióticos. Estudios genómicos han revelado que el género *Comamonas* posee genes asociados a adaptación ambiental, metabolismo versátil y posibles factores de virulencia (Y. Wu et al., 2018).

En el metagenoma SRX3743986, correspondiente a la primera repetición de la Muestra 3, se identificó un predominio del filo *Firmicutes*, el cual representó más de la mitad de las lecturas clasificadas. Dentro de este grupo, el orden *Selenomonadales* mostró una alta representación, destacándose particularmente la especie *Megasphaera elsdenii* como el taxón más abundante del conjunto.

En la Figura 7 se puede observar también la presencia del filo *Actinobacteria*, el cual concentró una proporción considerable de las secuencias, principalmente dentro de la familia *Bifidobacteriaceae*. Asimismo, el orden *Lactobacillales*, perteneciente también a *Firmicutes*, estuvo representado por diferentes géneros de la familia *Lactobacillaceae*, lo que refleja cierta diversidad dentro de este grupo.

Otros filios como *Proteobacteria* y *Bacteroidetes* estuvieron presentes en menor proporción, con representación dispersa en varios géneros y órdenes, pero sin alcanzar una dominancia significativa en comparación con los grupos mencionados anteriormente.

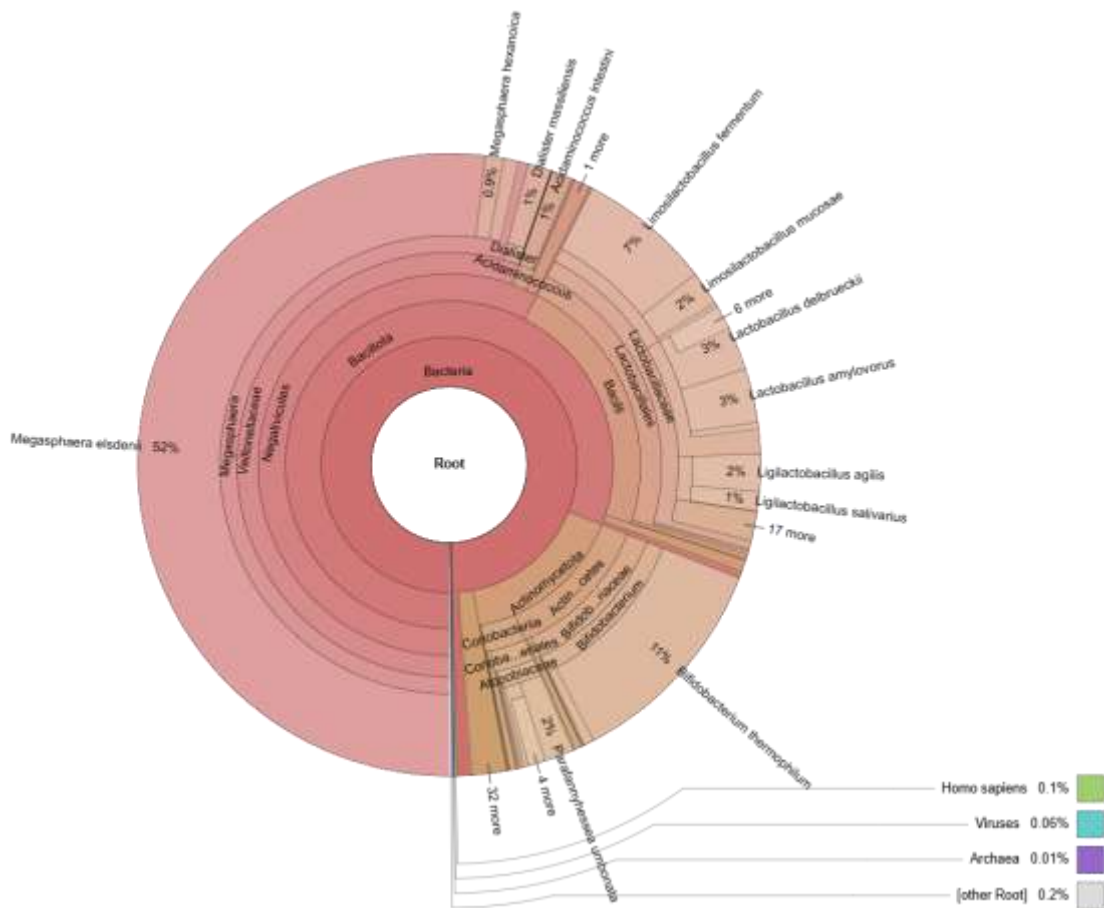


Figura 7. Diagrama circular de Krona del metagenoma SRX3743986

El análisis del metagenoma SRX3743982, correspondiente a la segunda repetición de la Muestra 4, reveló una clara predominancia del filo *Firmicutes*, con una representación destacada del orden *Lactobacillales*. Dentro de este grupo, la especie *Lactobacillus amylovorus* fue la más abundante en esta muestra, sobresaliendo respecto a las demás. Esta especie es conocida por su aplicación como probiótico en producción porcina, se caracteriza por su capacidad para fermentar almidones y sobrevivir en ambientes ácidos, lo que explica su presencia en matrices como la vinaza, rica en compuestos fermentables y materia orgánica. Aunque no se considera un patógeno, investigaciones como las de

Ammor et al. (2007) han documentado la presencia de genes de resistencia a antibióticos en cepas de *L. amylovorus*, particularmente frente a tetraciclinas, eritromicina y cloranfenicol, lo cual, aunque de bajo riesgo clínico directo, plantea interrogantes sobre su papel como reservorio de genes de resistencia en ambientes abiertos. Aplicada como fertirrigante, la vinaza podría facilitar la incorporación de esta bacteria a la microbiota del suelo, donde su actividad fermentativa podría alterar las condiciones edáficas y favorecer interacciones horizontales con otros microorganismos. En este contexto, *L. amylovorus* actúa como un agente microbiano ambientalmente activo que, si bien tiene un rol funcional positivo en su origen, debe ser evaluado críticamente cuando forma parte de matrices orgánicas aplicadas al suelo.

También se puede identificar en la Figura 8 una presencia importante de *Megasphaera elsdenii*, perteneciente al filo *Firmicutes*, específicamente al orden *Selenomonadales*. Aunque en menor proporción que *L. amylovorus*, su abundancia fue considerable y cercana.

Además, se observó una participación relevante de secuencias correspondientes al filo *Actinobacteria*, con la familia *Bifidobacteriaceae* bien representada, y del grupo *Proteobacteria*, aunque en menor proporción. También se detectó una fracción de lecturas clasificadas como virus, específicamente dentro de la familia *Siphoviridae*, aunque estas no alcanzaron niveles comparables a los de los filos bacterianos dominantes.

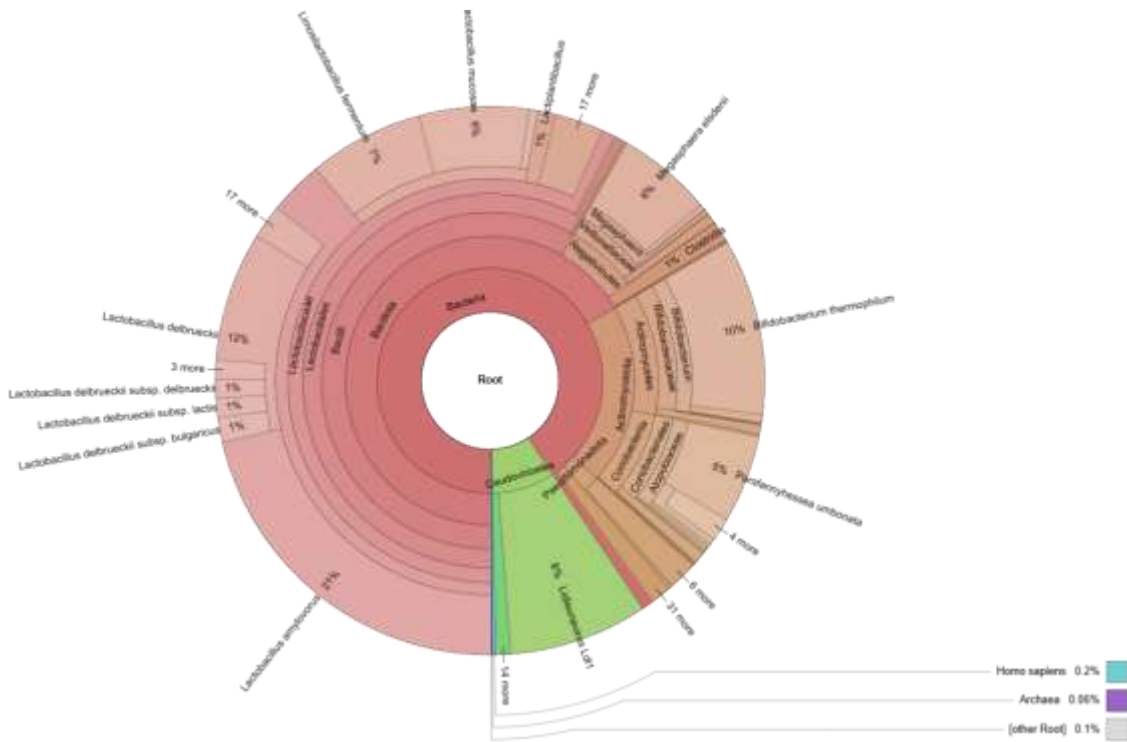


Figura 8. Diagrama circular de Krona del metagenoma SRX3743982

En la Figura 9 se observa el diagrama de Krona del metagenoma SRX3743980, correspondiente a la primera réplica de la Muestra 5, en este se puede notar que existe una marcada predominancia de la especie *Megasphaera elsdenii*, la cual representó aproximadamente la mitad de las lecturas clasificadas.

La segunda especie más abundante fue *Lactobacillus amylovorus*, también dentro del filo *Firmicutes*, pero perteneciente al orden *Lactobacillales*. Dentro de este mismo orden se identificaron otras especies del género *Lactobacillus*, con participaciones menores pero constantes.

A nivel de filo, además de la clara dominancia de *Firmicutes*, se identificaron secuencias correspondientes a *Actinobacteria*, *Proteobacteria* y *Bacteroidetes*, aunque estas estuvieron representadas en proporciones considerablemente menores. Dentro de *Actinobacteria*, se destacan miembros de la familia *Bifidobacteriaceae*, mientras que en

4.5. Estimación de abundancia

A continuación, se presentan los resultados de la herramienta Bracken, los cuales permitieron confirmar los resultados de Kraken 2 visualizados mediante Krona en la sección 4.4. Bracken permitió refinar los conteos iniciales, proporcionando una aproximación más cercana a la abundancia real de cada taxón en la muestra, y confirmando los perfiles observados en las representaciones jerárquicas (Lu et al., 2017).

Para facilitar la comprensión de los perfiles taxonómicos generados por Bracken, en esta sección se muestran únicamente las cinco especies con mayor número de lecturas estimadas para el metagenoma más representativo de cada muestra, seleccionadas entre un total aproximado de 160 especies detectadas por réplica. No obstante, para un análisis más detallado, se proporciona un enlace de acceso a las tablas completas correspondientes a cada metagenoma.

Link: <https://drive.google.com/drive/folders/1RXA0tbwKFslfOGzyJ2-6Bys2bG8Ahd3?usp=sharing>

En el caso del metagenoma SRX3743992, correspondiente a la primera réplica de la muestra 1, se puede observar en la Tabla 3 que la especie más abundante fue *Megasphaera elsdenii*, con una representación estimada del 45,1 % del total de lecturas. Esta especie dominó la comunidad microbiana de esta muestra, lo que concuerda con su prevalencia visualizada en el análisis con Krona.

Le siguió *Bifidobacterium thermophilum*, con una abundancia del 19,5 %, y posteriormente *Lactobacillus delbrueckii*, con un 4,5 %. Estos taxones también estuvieron entre los más destacados en la clasificación taxonómica jerárquica previa.

Finalmente, se identificaron *Olsenella umbonata* (4,3 %) y *Acidaminococcus intestini* (2,4 %) como parte del conjunto de especies más representativas, estas especies también fueron encontradas en estudios como el de Cassman et al., (2018).

Tabla 3. Estimación de abundancia a nivel de especie para el metagenoma SRX3743992

Nombre	Taxonomía ID	Nivel de Taxonomía	Lecturas asignadas por Kraken	Lecturas añadidas	Nuevas lecturas estimadas	Fracción de lecturas totales
<i>Megasphaera elsdenii</i>	907	S	34378	780	35158	0,4231
<i>Bifidobacterium thermophilum</i>	33905	S	14856	55	14911	0,17944
<i>Lactobacillus delbrueckii</i>	1584	S	3462	915	4377	0,05267
<i>Olsenella umbonata</i>	604330	S	3331	247	3578	0,04306
<i>Acidaminococcus intestini</i>	187327	S	1843	172	2015	0,02425

En el metagenoma SRX3743987, correspondiente a la tercera réplica de la muestra 2, las cinco especies con mayor abundancia relativa estimada fueron *Comamonas kerstersii*, *Arcobacter skirrowii*, *Lactobacillus fermentum*, *Lactobacillus amylovorus* y *Alcaligenes faecalis*.

Como se puede observar en la Tabla 4, la especie más representativa fue *Comamonas kerstersii*, con un total estimado de 42.396 lecturas (new_est_reads), lo que equivale al 28,1 % del total de secuencias asignadas en este metagenoma. Le siguió *Arcobacter skirrowii* con 24.444 lecturas estimadas (16,2 %), mientras que *Lactobacillus fermentum* y *Lactobacillus amylovorus* aportaron 14.591 (9,7 %) y 10.056 (6,7 %) lecturas respectivamente. Finalmente, *Alcaligenes faecalis* alcanzó un total de 5.592 lecturas, representando un 3,7 % del total. Estos resultados concuerdan con el análisis visual previo realizado con Krona, en el que *Comamonas kerstersii* se destacó como la especie dominante dentro del perfil taxonómico del metagenoma analizado.

El valor de *new_est_reads* corresponde a la estimación refinada del número de lecturas que se asignan a cada especie tras el reajuste bayesiano aplicado por Bracken, tomando como base la distribución de los k-mers y la estructura taxonómica del informe de Kraken (Lu et al., 2017). Esta métrica proporciona una aproximación más precisa de la abundancia real, en comparación con los conteos brutos generados en la etapa inicial de clasificación.

Tabla 4. Estimación de abundancia a nivel de especie para el metagenoma SRX3743987

Nombre	Taxonomía ID	Nivel de Taxonomía	Lecturas asignadas por Kraken	Lecturas añadidas	Nuevas lecturas estimadas	Fracción de lecturas totales
<i>Comamonas kerstersii</i>	225992	S	42396	851	43247	0,25126
<i>Arcobacter skirrowii</i>	28200	S	24444	25	24469	0,14216
<i>Lactobacillus fermentum</i>	1613	S	14591	704	15295	0,08886
<i>Lactobacillus amylovorus</i>	1604	S	10056	3618	13674	0,07945
<i>Alcaligenes faecalis</i>	511	S	5592	1161	6753	0,03923

En la muestra correspondiente al metagenoma SRX3743986 se identificaron las cinco especies más representativas a partir de los datos procesados por Bracken, según el número de lecturas asignadas inicialmente por Kraken2 (*kraken_assigned_reads*). En la Tabla 5 se observa que *Megasphaera elsdenii* fue la especie con mayor número de lecturas, con un total de 66417, representando el 56,07 % del total estimado para esta muestra. Le sigue *Bifidobacterium thermophilum*, con 14.591 lecturas asignadas (12,32 %), mientras que *Lactobacillus fermentum* registró 8.856 lecturas, correspondiente al 7,48 % del total.

Las otras dos especies con mayor abundancia relativa fueron *Lactobacillus delbrueckii* y *Lactobacillus amylovorus*, con 4.461 (3,76 %) y 3.609 lecturas (3,05 %) respectivamente.

Esta distribución sugiere una alta participación de especies del género *Lactobacillus* en la muestra, las cuales en conjunto representan más del 14 % de las lecturas estimadas.

El alto número de lecturas asignadas por Kraken2 a estas especies fue mantenido tras la estimación corregida por Bracken (*new_est_reads*), indicando coherencia en la clasificación y reforzando la predominancia de *Megasphaera elsdenii* en esta réplica. Estos datos son consistentes con los obtenidos a través de la visualización Krona, donde también se observó una alta proporción de dicha especie en la muestra analizada.

Tabla 5. Estimación de abundancia a nivel de especie para el metagenoma SRX3743986

Nombre	Taxonomía ID	Nivel de Taxonomía	Lecturas asignadas por Kraken	Lecturas añadidas	Nuevas lecturas estimadas	Fracción de lecturas totales
<i>Megasphaera elsdenii</i>	907	S	66417	1146	67563	0,54027
<i>Bifidobacterium thermophilum</i>	33905	S	14591	32	14623	0,11693
<i>Lactobacillus fermentum</i>	1613	S	8856	273	9129	0,073
<i>Lactobacillus delbrueckii</i>	1584	S	4461	1338	5799	0,04637
<i>Lactobacillus amylovorus</i>	1604	S	3609	625	4234	0,03386

En la Tabla 6 se observa la estimación de abundancia para el metagenoma SRX3743982, la réplica número 2 de la muestra 4. La especie con mayor número de lecturas estimadas fue *Lactobacillus amylovorus*, con un total de 23.810 lecturas según el valor de *new_est_reads*, lo que representa aproximadamente el 23,84 % del total de lecturas asignadas. Le sigue *Lactobacillus delbrueckii*, con 13.856 lecturas (13,87 %), y *Bifidobacterium thermophilum* con 11.575 lecturas (11,59 %).

Un aspecto destacable en esta muestra es la presencia de *Lactobacillus phage Ldl1*, un bacteriófago asociado al género *Lactobacillus*, con 9.121 lecturas estimadas, lo que equivale al 9,13 % de la muestra, posicionándose como el cuarto elemento más abundante

en la tabla. Finalmente, *Lactobacillus fermentum* también aparece entre las especies dominantes, con un total de 7.498 lecturas (7,51 %). La detección del bacteriófago *Lactobacillus phage Ldl1* en entornos industriales fermentativos, y su posible presencia en matrices orgánicas como la vinaza, representa un componente poco explorado pero relevante en el análisis microbiológico de residuos aplicados al ambiente. Este virus infecta cepas específicas de *Lactobacillus delbrueckii subsp. lactis* y ha sido caracterizado por su capacidad de replicarse dentro de estas bacterias y destruirlas al final del ciclo, liberando nuevas partículas virales (Casey et al., 2015). Este comportamiento, que interrumpe activamente el crecimiento bacteriano, adquiere importancia en residuos fermentativos como la vinaza, donde pueden coexistir bacterias ácido-lácticas sensibles y fagos funcionales. Aunque estos virus no representan un riesgo directo para la salud humana o animal, su introducción al suelo mediante fertirrigación podría alterar la estructura de las comunidades microbianas edáficas, afectando especialmente a bacterias beneficiosas como *Lactobacillus spp.*

Cabe recalcar que la mayoría de las especies más abundantes en esta réplica pertenecen al género *Lactobacillus*, lo cual se refleja en la alta proporción de lecturas asociadas a esta agrupación.

Tabla 6. Estimación de abundancia a nivel de especie para el metagenoma SRX3743982

Nombre	Taxonomía ID	Nivel de Taxonomía	Lecturas asignadas por Kraken	Lecturas añadidas	Nuevas lecturas estimadas	Fracción de lecturas totales
<i>Lactobacillus amylovorus</i>	1604	S	23810	1607	25417	0,23508
<i>Lactobacillus delbrueckii</i>	1584	S	13856	2714	16570	0,15326
<i>Bifidobacterium thermophilum</i>	33905	S	11575	14	11589	0,10719
<i>Lactobacillus phage Ldl1</i>	1552735	S	9121	0	9121	0,08436
<i>Lactobacillus fermentum</i>	1613	S	7498	166	7664	0,07088

En el metagenoma SRX3743980, correspondiente a la primera réplica de la muestra 5, se identificó un perfil taxonómico dominado por la especie *Megasphaera elsdenii*, la cual representó el 54,65 % del total de las lecturas. En segundo lugar, se posicionó *Lactobacillus amylovorus*, con una representación del 7,84 %, seguida por *Lactobacillus mucosae* (6,49 %), *Lactobacillus delbrueckii* (5,77 %) y *Bifidobacterium thermophilum* (4,61 %).

Como se puede apreciar en la Tabla 7, tres de las cinco especies más abundantes corresponden al género *Lactobacillus*, lo cual indica una importante presencia de bacterias ácido-lácticas en esta muestra. Esta abundancia sugiere condiciones fermentativas activas o residuales, típicas de ambientes ricos en carbohidratos y ácidos orgánicos como la vinaza. Asimismo, la detección de *Bifidobacterium thermophilum*, una especie adaptada al tracto intestinal de animales y conocida por su papel en la fermentación de oligosacáridos y polisacáridos complejos refuerza la presencia de bacterias con funciones fermentativas potenciales (Ammor et al., 2007).

Tabla 7. Estimación de abundancia a nivel de especie para el metagenoma SRX3743980

Nombre	Taxonomía ID	Nivel de Taxonomía	Lecturas asignadas por Kraken	Lecturas añadidas	Nuevas lecturas estimadas	Fracción de lecturas totales
<i>Megasphaera elsdenii</i>	907	S	53286	858	54144	0,50822
<i>Lactobacillus amylovorus</i>	1604	S	7647	730	8377	0,07863
<i>Lactobacillus delbrueckii</i>	1584	S	5629	1290	6919	0,06494
<i>Lactobacillus mucosae</i>	97478	S	6332	68	6400	0,06007
<i>Bifidobacterium thermophilum</i>	33905	S	4493	6	4499	0,04223

En el metagenoma SRX3743975, correspondiente a la tercera réplica de la muestra 6, se observó un perfil taxonómico caracterizado por un predominio notable de especies pertenecientes a la familia *Lactobacillaceae*. En Tabla 8 se puede observar que la especie más abundante fue *Lactobacillus amylovorus*, la cual representó el 21,85 % del total de lecturas estimadas. Le siguieron *Lactobacillus delbrueckii* (18,43 %), *Lactobacillus fermentum* (8,05 %) y *Lactobacillus mucosae* (7,70 %), lo que muestra que cuatro de las cinco especies más frecuentes corresponden a este mismo linaje bacteriano.

La única especie no perteneciente a *Lactobacillus* dentro de las más representativas fue *Bifidobacterium thermophilum*, con una proporción del 12,42 %, también frecuente en ambientes fermentativos. Este patrón de dominancia de *Lactobacillus* se presentó de forma consistente en las otras dos réplicas analizadas de la muestra 6 (metagenomas SRX3743977 y SRX3743976), en las cuales *Lactobacillus amylovorus* también fue la especie con mayor abundancia. La presencia significativa de *Bifidobacterium thermophilum*, una bacteria intestinal típica de animales de sangre caliente, refuerza el perfil fermentativo de la comunidad microbiana, al tratarse de un organismo capaz de metabolizar oligosacáridos complejos en ambientes pobres en oxígeno (Ammor et al., 2007; Ríos-Covián et al., 2016).

Tabla 8. Estimación de abundancia a nivel de especie para el metagenoma SRX3743975

Nombre	Taxonomía ID	Nivel de Taxonomía	Lecturas asignadas por Kraken	Lecturas añadidas	Nuevas lecturas estimadas	Fracción de lecturas totales
<i>Lactobacillus amylovorus</i>	1604	S	22066	1515	23581	0,21475
<i>Lactobacillus delbrueckii</i>	1584	S	18614	3141	21755	0,19812
<i>Bifidobacterium thermophilum</i>	33905	S	12546	15	12561	0,11439
<i>Lactobacillus fermentum</i>	1613	S	8131	176	8307	0,07565
<i>Lactobacillus mucosae</i>	97478	S	7775	70	7845	0,07144

La especie más frecuente en múltiples metagenomas fue *Megasphaera elsdenii*, especialmente dominante en SRX3743992, SRX3743986 y SRX3743980, con representaciones superiores al 45 %. Esta bacteria es una fermentadora estricta de lactato, común en el rumen de rumiantes, y ha sido propuesta como estabilizadora del pH en ambientes ricos en carbohidratos (Cabral & Weimer, 2024). Estudios como el de os Reis et al. (2015), identifica a *Megasphaera spp* directamente en vinaza de caña de azúcar sometida a digestión anaeróbica, participando activamente en la producción de metabolitos fermentativos como ácido butírico, etanol y propionato, así como en la generación de hidrógeno.

Otras especies recurrentes son *Bifidobacterium thermophilum* y *Lactobacillus amylovorus*, las cuales han sido asociadas en la literatura con la portación de genes de resistencia a antibióticos como tetraciclinas, eritromicina y clindamicina, especialmente en contextos agroindustriales (Ammor et al., 2007). Además de *Megasphaera elsdenii*, géneros como *Lactobacillus* y *Mitsuokella* han sido identificados como componentes clave en metagenomas de vinaza de caña, junto con representantes de los filos *Proteobacteria* y *Bacteroidetes* (Callejas et al., 2022; Cassman et al., 2018; Mota et al., 2024).

El análisis metagenómico realizado por Cassman et al. (2018) reportó una diversidad microbiana relativamente baja, destacando como géneros más representativos a *Lactobacillus*, *Megasphaera* y *Mitsuokella*, todos pertenecientes al filo *Firmicutes*, que en conjunto abarcaron entre el 35 % y el 97 % de las lecturas en las distintas muestras. Además, también se detectaron otros géneros de interés, como *Arcobacter* y *Alcaligenes* (*Proteobacteria*), *Dysgonomonas* (*Bacteroidetes*) y *Bifidobacterium* (*Actinobacteria*).

En contextos fermentativos acidogénicos con vinaza se han identificado microorganismos funcionalmente activos como *Clostridium ljungdahlii* y *Pectinatus frisingensis*, los cuales desempeñan un papel fundamental en la modulación de metabolitos y se ha descrito que la formación de acetato a partir de la degradación de la vinaza se atribuye principalmente a *Megasphaera* y *Clostridium* (Mota et al., 2024).

4.6. Ensamblaje metagenómico

Con base en los resultados de los análisis taxonómicos previos (ver sección 4.5), se identificó que el metagenoma SRX3743975 presentaba la mayor abundancia relativa de la familia *Lactobacillaceae*, según los perfiles generados por Kraken2 y Bracken. Por ende, se seleccionó dicho metagenoma como modelo representativo para llevar a cabo la validación metodológica del ensamblaje y las etapas posteriores de binning y anotación en la plataforma KBase.

En la Figura 11 se presenta el reporte generado por QUAST, donde se comparan las métricas de calidad de ensamblaje obtenidas para las herramientas MEGAHIT y metaSPAdes, aplicadas exclusivamente al metagenoma SRX3743975.

El reporte generado por QUAST muestra que el ensamblaje con MEGAHIT presentó una mayor longitud del contig (61637 pb) y una mayor suma de longitud total en contigs ≥ 1000 pb (9938654 pb) respecto a metaSPAdes (57281 pb; 9504175 pb). Aunque el valor N50 fue levemente superior en metaSPAdes (1059 vs. 1026), MEGAHIT tuvo un número nulo de bases ambiguas (N's), mientras que metaSPAdes presentó 4829, lo cual representa una ventaja en términos de limpieza y precisión del ensamblaje.

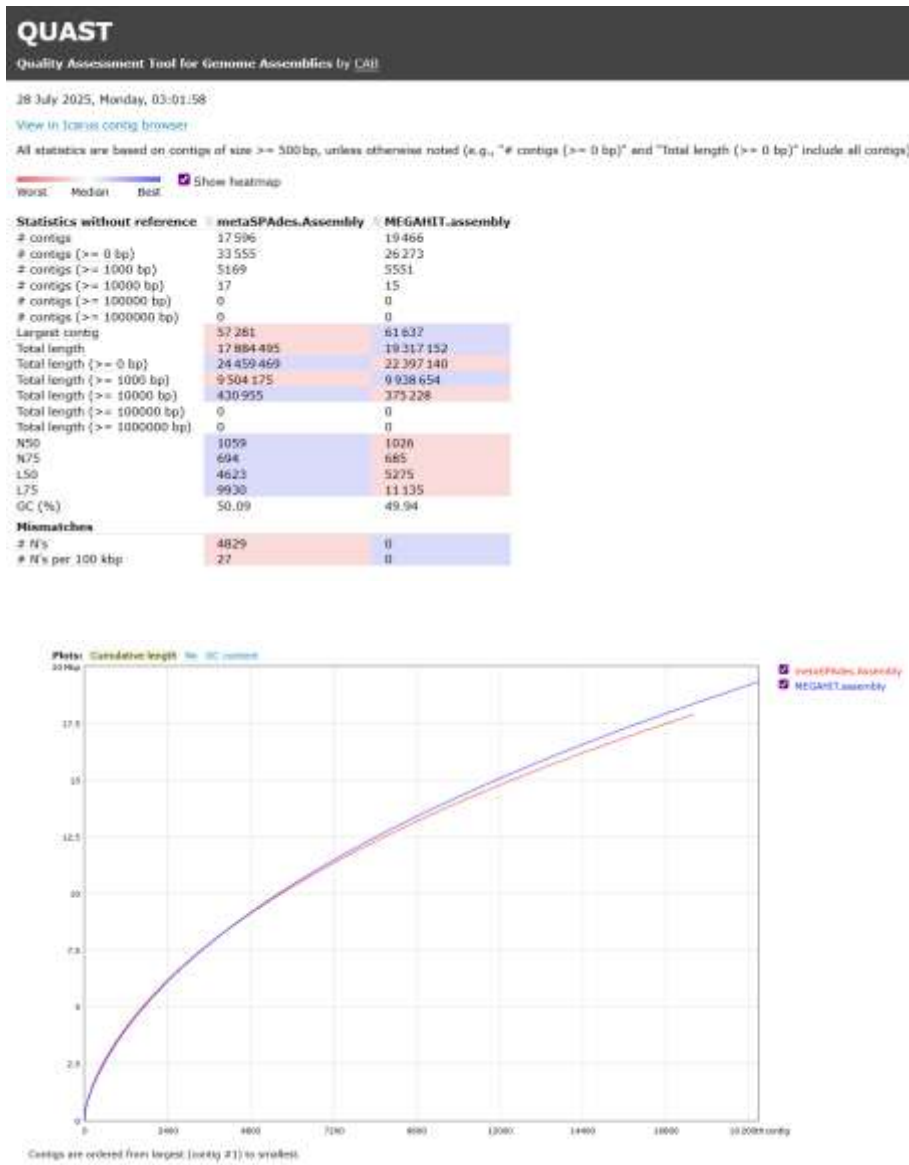


Figura 11. Comparación del ensamblaje del metagenoma SRX3743975 con MEGAHIT y metaSPAdes mediante QUAST

En la Figura 12, generada mediante la herramienta Compare Assembled Contig Distributions, se puede observar las métricas de calidad clave, destacando nuevamente el mejor desempeño del ensamblaje MEGAHIT en cuanto a longitud máxima de contig, número total de contigs ≥ 10 kb, y valores Nx favorables (N75, N90). Los histogramas de distribución muestran que MEGAHIT tiene una mayor proporción de contigs intermedios y largos, lo cual es deseable para una posterior reconstrucción genómica y análisis funcional. Además, el ensamblaje con MEGAHIT muestra una curva ligeramente más

pronunciada, indicando una mejor capacidad para generar secuencias continuas con menos fragmentos.

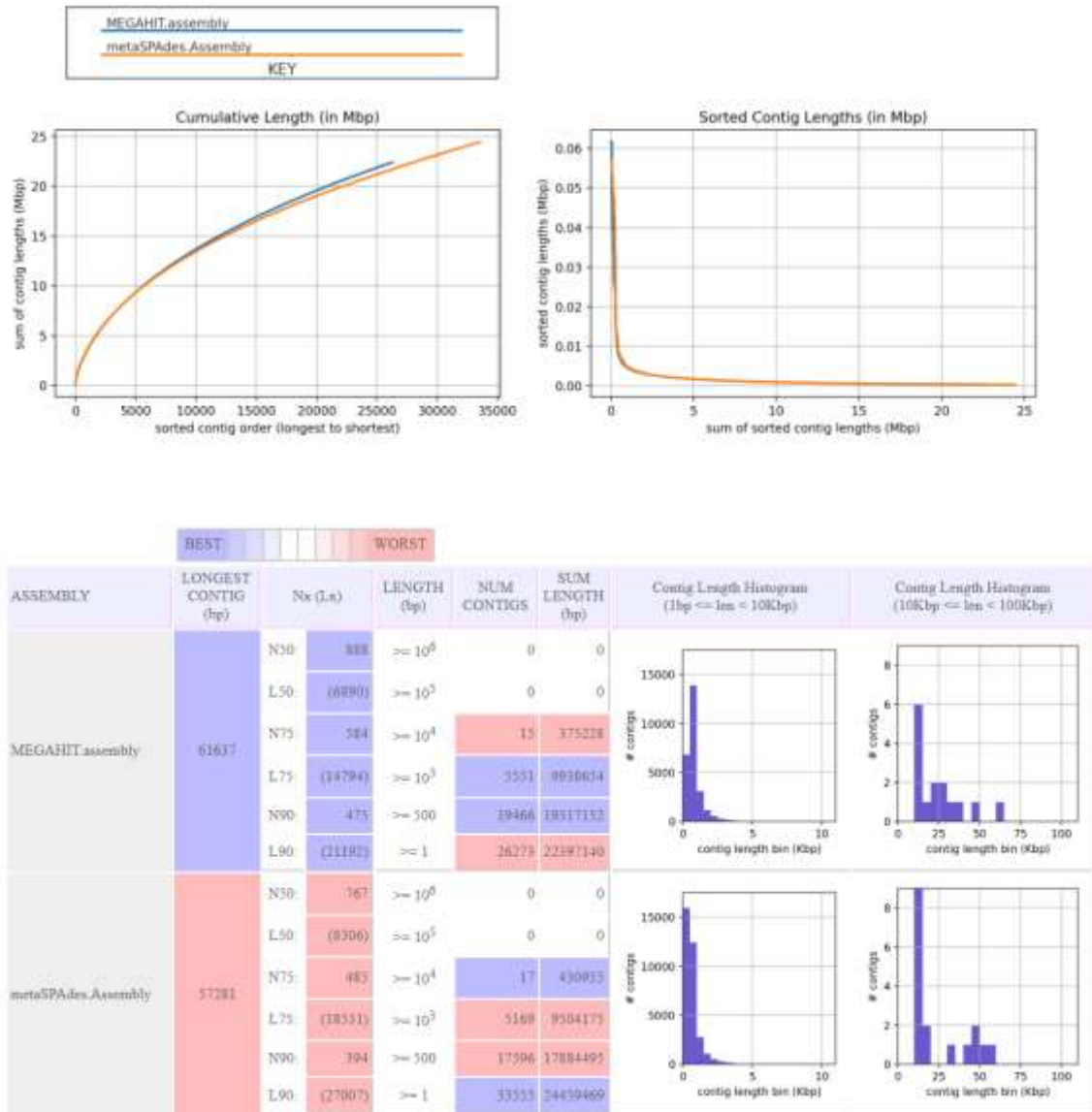


Figura 12. Comparación de distribuciones de contigs ensamblados del metagenoma SRX3743975 con MEGAHIT y metaSPAdes

De acuerdo con los resultados obtenidos mediante QUASt y la herramienta Compare Assembled Contig Distributions, se determinó que MEGAHIT presentó un mejor desempeño en métricas clave de ensamblaje en comparación con metaSPAdes. Por este motivo, se seleccionó MEGAHIT como ensamblador estandarizado para el resto de las

muestras, y sus resultados se emplearon como base para las etapas posteriores. La elección del ensamblador más adecuado no es un proceso al azar, ya que su rendimiento puede fluctuar notablemente en dependencia del tipo de muestra, la calidad de las secuencias y los objetivos del estudio. Tal como indican Murigneux et al. (2020), incluso al trabajar con tecnologías de secuenciación avanzadas, resulta fundamental comparar diferentes estrategias aplicadas sobre una misma muestra con el fin de identificar el enfoque más preciso, eficiente y ajustado a las necesidades biológicas del análisis. La elección de MEGAHIT no solo se basó en estos resultados, sino que además permitió una comparación directa con los ensamblajes generados por metaSPAdes, como los evidenciados en Cassman et al. (2018), aportando una validación cruzada que respalda la decisión metodológica adoptada.

4.7. Agrupamiento de contigs en bins y evaluación de bins

Como parte del análisis comparativo para definir la herramienta de binning metagenómico más adecuada, se procesó el ensamblaje del metagenoma SRX3743975 utilizando los algoritmos MaxBin2 y MetaBAT2.

La herramienta MaxBin2 generó 4 bins a partir de un total de 5110 contigs clasificados, lo que representó el 19.4 % de los 26273 contigs de entrada. No obstante, un 78.9 % de estos contigs fueron descartados por tener longitudes inferiores al umbral mínimo establecido (1500 pb) . A pesar de esta limitación, los contigs agrupados por MaxBin2 sumaron 9151816 pb, equivalente al 40.9 % de la longitud total del ensamblaje. La evaluación de calidad mediante CheckM (Figura 13) evidenció variabilidad entre los bins, con algunos presentando niveles moderados de contaminación (zonas en tonos naranjas y rojos) y heterogeneidad, además de pérdida de genes marcadores esenciales en ciertos casos.

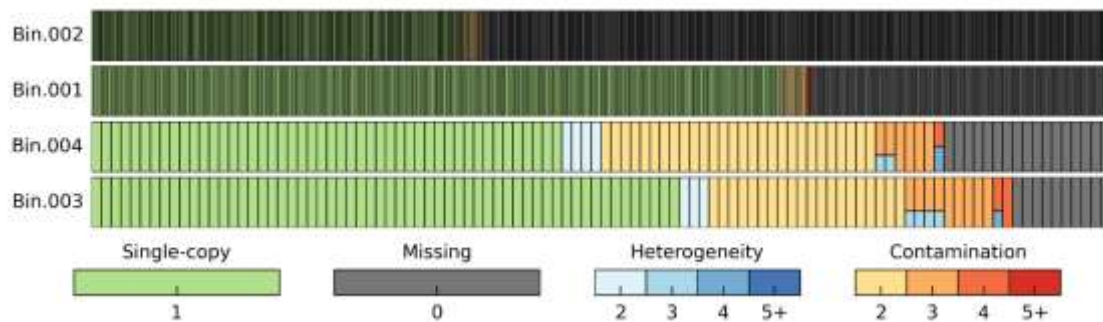


Figura 13. Evaluación de calidad de los bins generados por MaxBin2 mediante CheckM

Por otro lado, el análisis con MetaBAT2 produjo 3 bins, agrupando un total de 502 contigs (Figura 14). Aunque el número de contigs clasificados fue menor al de MaxBin2, la contaminación de los bins obtenidos fue menor. Los resultados de CheckM indicaron una proporción moderada de genes de copia única (color verde uniforme), sin señales visibles de contaminación o duplicación, y con ausencia de genes perdidos en la mayoría de los casos.

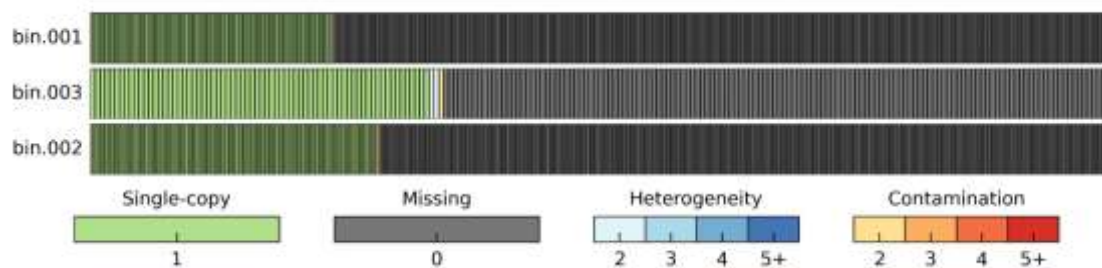


Figura 14. Evaluación de calidad de los bins generados por MetaBAT2 mediante CheckM

Los resultados de CheckM sugieren una mayor precisión en el agrupamiento de contigs en bins por parte de MetaBAT2, en comparación con MaxBin2. Por esta razón, MetaBAT2 fue seleccionado como herramienta estándar para el binning de los 18 metagenomas analizados en este estudio. La realización de esta evaluación de bins es esencial, ya que muchos genomas recuperados a partir de datos metagenómicos pueden encontrarse incompletos o contaminados, y una caracterización adecuada de su calidad es

indispensable para realizar inferencias funcionales confiables en etapas posteriores (Parks et al., 2015).

4.8. Identificación de genes de resistencia a antibióticos

En la Tabla 9 se detallan los resultados de la detección de genes de resistencia a antibióticos en los bins obtenidos a partir del análisis de los 18 metagenomas. Se identificaron genes en la mayoría de las muestras examinadas, entre los genes de resistencia detectados destacan *lnu(A)*, *lnu(B)*, *lnu(C)*, *erm(B)*, *lsaE*, *vat(E)* y *blaACI-1*, con altos porcentajes de identidad y cobertura, destacando valores que van desde 91.77 % hasta 100 % de identidad y una cobertura del 100 % en prácticamente todos los casos.

Los genes de resistencia a antibióticos identificados en los bins de los metagenomas analizados reflejan la presencia de determinantes genéticos funcionales, asociados principalmente a lincosamidas, macrólidos, estreptograminas, beta-lactámicos y otros antimicrobianos de importancia en medicina veterinaria y humana. Muchos de estos genes han sido previamente descritos en bacterias del tracto intestinal de animales de producción, y su detección en la vinaza sugiere que esta matriz actúa como reservorio y posible vehículo de diseminación de resistencia en ambientes agrícolas.

El gen *lnu(A)*, por ejemplo, codifica una nucleotidiltransferasa que confiere resistencia específica a lincomicina mediante inactivación enzimática. Este gen ha sido detectado en *Staphylococcus lugdunensis* en entornos hospitalarios, contenido en un plásmido multirresistencia (*pCGMH_SL138*), junto a genes que confieren resistencia a aminoglucósidos, tetraciclinas y biocidas como la clorhexidina, evidenciando su inserción en clústeres móviles de alta complejidad (Chang et al., 2022). Su presencia en la vinaza indica una posible transferencia desde microbiota intestinal o ambiental expuesta a antibióticos.

De igual manera, el gen *lnu(B)* ha sido reportado en cepas de *Staphylococcus aureus* asociadas a animales, incluidos linajes típicos de origen ganadero (ST398, ST9), y puede estar presente tanto en plásmidos como en el cromosoma (Lozano et al., 2012). La amplia distribución de este gen entre cepas humanas y animales subraya su potencial de transferencia cruzada, especialmente en contextos de uso compartido de antimicrobianos como las lincosamidas en producción animal.

El gen *lnu(C)* representa otra variante funcional con alta actividad frente a lincomicina. En *Brachyspira hyodysenteriae*, patógeno entérico porcino, *lnu(C)* se ha detectado en el transposón MTnSag1, y su presencia se asoció a un aumento significativo en la concentración mínima inhibitoria (MIC) frente a este antibiótico (De Luca et al., 2018). Además, se ha reportado en *Campylobacter coli* de origen avícola, donde se ubica dentro de un transposón MTnSag1-like integrado en el cromosoma, confiriendo una resistencia de al menos 32 veces superior a cepas susceptibles (W. Li et al., 2023). Estas investigaciones demuestran que *lnu(C)* tiene una elevada movilidad genética y puede diseminarse por transformación natural o recombinación, por lo que su detección en vinaza resulta especialmente relevante.

Otro gen ampliamente identificado fue *erm(B)*, codificante de una metiltransferasa ribosomal que otorga resistencia cruzada a macrólidos, lincosamidas y estreptograminas del tipo B (MLS_B). Este gen fue detectado en *Campylobacter coli* aislado de carne de pollo comercializada en Lima, Perú, con una frecuencia superior al 50% en cepas resistentes a eritromicina (Anampa et al., 2020).

El gen *lsa(E)*, perteneciente a la familia ABC-F, fue identificado por primera vez en 2013 en cepas de *Staphylococcus aureus* tanto sensibles como resistentes a meticilina (Wendlandt et al., 2013). Según Shan et al. (2022), este gen confiere resistencia combinada a lincosamidas, pleuromutilinas y estreptograminas tipo A y se ha encontrado

formando parte de clústeres de genes multirresistencia junto a *aadE*, *spw* y *lnu(B)*, entre otros. Hasta la fecha, se han descrito diversas variantes de estos clústeres en géneros como *Streptococcus*, *Enterococcus*, *Staphylococcus* y otras bacterias Gram-positivas (Shan et al., 2022).

Asimismo, el gen *blaACI-1*, también identificado como *ACI-1*, codifica una beta-lactamasa de espectro extendido presente en *Acidaminococcus intestini*, una bacteria comensal de humanos y animales. Este gen ha sido asociado a transposones contenidos en fago-provirus en el microbioma intestinal, y se ha reportado en más del 4% de metagenomas humanos analizados globalmente (Rands et al., 2018). Su localización en elementos móviles complejos sugiere una elevada capacidad de transferencia horizontal, y su presencia en residuos orgánicos como la vinaza refuerza la necesidad de un manejo controlado de este subproducto, especialmente cuando se emplea como fertilizante en sistemas abiertos.

Los resultados obtenidos con las bases de datos ResFinder y CARD fueron en general muy similares. En múltiples ocasiones, ambas bases reportaron genes relacionados al mismo tipo de resistencia dentro del mismo bin, aunque con diferencias en la nomenclatura. Por ejemplo, se observó una coincidencia frecuente entre *lnu(C)_1* en ResFinder y *lnuC* en CARD, así como entre *blaACI-1_1* y *ACI-1*. Estos pares se detectaron con idénticos niveles de identidad y cobertura, reflejando una correspondencia clara entre ambas bases de datos.

Un patrón notable fue el caso de los genes *lnu(B)_2* y *lsaE*, que aparecieron de forma alternada entre ResFinder y CARD, pero casi siempre en los mismos bins. Este comportamiento fue recurrente a lo largo de varias muestras. Este patrón podría generar confusión, pero es importante destacar que, aunque *lnu(B)_2* y *lsa(E)* pueden aparecer en el mismo bin y conferir resistencia a un antibiótico común como la lincomicina, no son

el mismo gen ni actúan por el mismo mecanismo. Como antes se mencionaba, *lnu(B)_2* codifica una nucleotidiltransferasa que inactiva directamente las lincosamidas mediante adenyilación, mientras que *lsa(E)* pertenece a la familia de proteínas ABC-F, las cuales no modifican el antibiótico, sino que protegen el ribosoma bacteriano del efecto del fármaco. Como señalan (Shan et al., 2022), este gen suele localizarse en clústeres multiresistencia, donde puede encontrarse junto a *lnu(B)*, *aadE*, *spw* y otros genes, lo cual explica su co-ocurrencia frecuente, pero no implica equivalencia funcional.

También se puede observar en la Tabla 9 que los genes de resistencia identificados afectan a una diversidad de clases de antibióticos de relevancia veterinaria y médica, entre ellas: lincosamidas, macrólidos, estreptograminas, beta-lactámicos (penem, penam, cefalosporinas), oxazolidinonas, pleuromutilinas, tetraciclinas y fenicoles. Muchas de estas clases son catalogadas por la Organización Mundial de la Salud (OMS) como antibióticos de importancia crítica para la salud humana debido a su uso en infecciones graves, especialmente en casos donde existen pocas alternativas terapéuticas (World Health Organization, 2019).

Dentro de los antibióticos a los que se detectó resistencia en los metagenomas analizados, destacan los beta-lactámicos, una clase fundamental en la medicina moderna que incluye a las penicilinas, cefalosporinas, penem y monobactámicos. Estos antibióticos actúan inhibiendo la síntesis de la pared celular bacteriana al bloquear las proteínas fijadoras de penicilina (PBPs), lo que resulta en la lisis de la célula (Pandey & Cascella, 2023). Su uso generalizado en el tratamiento empírico de infecciones graves ha llevado a un aumento progresivo en los mecanismos de resistencia, entre ellos la producción de beta-lactamasas, enzimas capaces de hidrolizar el anillo beta-lactámico y anular su eficacia (Pandey & Cascella, 2023). Este fenómeno ha sido identificado como una amenaza creciente a la salud pública mundial, pues compromete la eficacia de fármacos de amplio

espectro, incluidos los de última línea como los carbapenémicos y cefalosporinas de tercera generación, utilizados en infecciones multirresistentes en humanos (Gaude & Hattiholli, 2013).

Otro grupo importante de antibióticos a los que se detectó resistencia en los metagenomas analizados fueron las lincosamidas, como la lincomicina y la clindamicina, generalmente usadas en medicina humana y veterinaria. Estos compuestos actúan sobre el ARN ribosomal 23S, inhibiendo la síntesis proteica bacteriana (Yang et al., 2024). La resistencia puede surgir por metilación ribosomal mediada por genes *erm* o por efecto de enzimas modificadoras como las adenililtransferasas codificadas por *lnu* y *lsa* (Sundlov & Gulick, 2009). Esta resistencia es preocupante, ya que estudios como el de (Yang et al., 2024), han mostrado evidencia de transferencia horizontal de genes de resistencia entre bacterias humanas, animales y ambientales, por ejemplo, se ha documentado cómo los ARGs provenientes de humanos pueden ingresar al ambiente a través de aguas residuales y residuos agrícolas, como excretas o subproductos fermentativos aplicados al suelo, permitiendo su posterior captación por bacterias del entorno.

Es importante destacar que, en muchos metagenomas, todos los genes de resistencia detectados se localizaron en un solo bin, como ocurrió con el BIN 4 de las muestras SRX3743992 y SRX3743991, y el BIN 2 de la muestra SRX3743979. En varias muestras se detectaron diversos genes de resistencia, por ejemplo, en SRX3743980, SRX3743981 y SRX3743986, pero, por otro lado, muestras como SRX3743989 y SRX3743978 no presentaron detección de genes con los parámetros aplicados.

La diseminación de estos genes por fertirrigación puede facilitar su incorporación al microbioma del suelo, plantas y animales, incrementando el riesgo de transferencia indirecta a humanos mediante consumo de alimentos, agua o contacto ambiental. Tal como advierten (Forsberg et al., 2012), los suelos pueden tornarse reservorios evolutivos

de genes de resistencia, alimentados continuamente por aportes de origen animal o antrópico, como los residuos orgánicos no tratados.

Tabla 9. Genes de resistencia detectados en los bins de los 18 metagenomas por ResFinder y CARD

Muestra	Accession SRA	Nº de bins obtenidos	Bin asignado	Base de datos	Gen de resistencia	Clase de antibiótico	% Identidad	% Cobertura
Nov. 2013	1.SRX3743992	6	BIN 4	Resfinder	lnu(C)_1	Lincomicina	99,19%	100%
			BIN 4	Resfinder	blaACI-1_1	-	99,77%	100%
			BIN 4	CARD	lnuC	Lincosamida	99,19%	100%
			BIN 4	CARD	ACI-1	Cefalosporina; Penam; Penem	99,77%	100%
	2.SRX3743991	6	BIN 4	Resfinder	lnu(C)_1	Lincomicina	99,19%	100%
			BIN 4	CARD	lnuC	Lincosamida	99,19%	100%
	3.SRX3743990	6	BIN 3	Resfinder	lnu(C)_1	Lincomicina	99,19%	100%
			BIN 3	Resfinder	blaACI-1_1	-	99,77%	100%
			BIN 3	CARD	lnuC	Lincosamida	99,19%	100%
			BIN 3	CARD	ACI-1	Cefalosporina; Penam; Penem	99,77%	100%
	4.SRX3743989	4	-	-	-	-	-	-
	5.SRX3743988	6	BIN 6	Resfinder	lnu(A)_1	Lincomicina	98,97%	100%
BIN 6			CARD	lnuA	Lincosamida	98,97%	100%	
Dic. 2013	6.SRX3743987	6	BIN 6	Resfinder	erm(B)_18	Eritromicina; Lincomicina; Clindamicina; Quinopristina; Pristinamicina IA; Virginiamicina S	99,61%	99,74%
			BIN 6	Resfinder	lnu(A)_1	Lincomicina	98,97%	100%
			BIN 6	CARD	ErmB	Lincosamida; Macrólido; Estreptogramina	99,20%	100%
			BIN 6	CARD	lnuA	Lincosamida	98,97%	100%
7.SRX3743986	5	BIN 2	Resfinder	lnu(C)_1	Lincomicina	99,19%	100%	

Jul. 2014	8.SRX3743985	6	BIN2	CARD	lnuC	Lincosamida	99,19%	100%
			BIN 4	Resfinder	lnu(B)_2	Lincomicina; Clindamicina	100%	100%
			BIN 4	Resfinder	vat(E)_5	Dalfopristina; Pristinamicina IIA; Virginiamicina M	100%	100%
			BIN 4	CARD	lsaE	Lincosamida; Macrólido; Oxazolidinona; Fenicol; Pleuromutilina; Estreptogramina; Tetraciclina	100%	100%
			BIN 4	CARD	vatE	Estreptogramina	100%	100%
			BIN 3	Resfinder	lnu(C)_1	Lincomicina	99,19%	100%
			BIN 3	CARD	lnuC	Lincosamida	99,19%	100%
			BIN 5	Resfinder	lnu(B)_2	Lincomicina; Clindamicina	100%	100%
			BIN 5	CARD	lsaE	Lincosamida; Macrólido; Oxazolidinona; Fenicol; Pleuromutilina; Estreptogramina; Tetraciclina	100%	100%
			9.SRX3743984	3	BIN 2	Resfinder	blaACI-1_1	-
BIN 2	CARD	ACI-1			Cefalosporina; Penam; Penem	99,77%	100%	
Ago. 2014	10.SRX3743983	3	BIN 3	Resfinder	erm(B)_18	Eritromicina; Lincomicina; Clindamicina; Quinopristina; Pristinamicina IA; Virginiamicina S	99,85%	99,74%
			BIN 3	CARD	ErmB	Lincosamida; Macrólido; Estreptogramina	99,73%	100%
11.SRX3743982	5	BIN 2	Resfinder	lnu(A)_1	Lincomicina	98,97%	100%	

			BIN 2	Resfinder	lnu(B)_2	Lincomicina; Clindamicina	100%	100%
			BIN 2	CARD	lnuA	Lincosamida	98,97%	100%
			BIN 2	CARD	lsaE	Lincosamida; Macrólido; Oxazolidinona; Fenicol; Pleuromutilina; Estreptogramina; Tetraciclina	100%	100%
			BIN 1	Resfinder	erm(B)_18	Eritromicina; Lincomicina; Clindamicina; Quinopristina; Pristinamicina IA; Virginiamicina S	99,21%	99,48%
			BIN 2	Resfinder	lnu(B)_2	Lincomicina; Clindamicina	100%	100%
			BIN 2	Resfinder	lnu(C)_1	Lincomicina	99,19%	100%
			BIN 2	Resfinder	blaACI-1_1	-	99,77%	100%
	12.SRX3743981	3	BIN 1	CARD	ErmB	Lincosamida; Macrólido; Estreptogramina	99,60%	99,87%
			BIN 2	CARD	lsaE	Lincosamida; Macrólido; Oxazolidinona; Fenicol; Pleuromutilina; Estreptogramina; Tetraciclina	100%	100%
			BIN 2	CARD	lnuC	Lincosamida	99,19%	100%
			BIN 2	CARD	ACI-1	Cefalosporina; Penam; Penem	99,77%	100%
Oct. 2014	13.SRX3743980	3	BIN 1	Resfinder	vat(E)_5	Dalfopristina; Pristinamicina IIA; Virginiamicina M	100%	100%
			BIN 1	Resfinder	lnu(A)_1	Lincomicina	91,77%	100%

				Eritromicina; Lincomicina; Clindamicina; Quinopristina; Pristinamicina IA; Virginiamicina S	99,72%	96,39%	
		BIN 1	Resfinder	erm(B)_6			
		BIN 2	Resfinder	lnu(B)_2	Lincomicina; Clindamicina	100%	100%
		BIN 3	Resfinder	lnu(C)_1	Lincomicina	99,19%	100%
		BIN 1	CARD	vatE	Estreptogramina	100%	100%
		BIN 1	CARD	lnuA	Lincosamida	91,77%	100%
		BIN 1	CARD	ErmB	Lincosamida; Macrólido; Estreptogramina	99,72%	96,39%
		BIN 2	CARD	lsaE	Lincosamida; Macrólido; Oxazolidinona; Fenicol; Pleuromutilina; Estreptogramina; Tetraciclina	100%	100%
		BIN 3	CARD	lnuC	Lincosamida	99,19%	100%
		BIN 2	Resfinder	blaACI-1_1	-	99,53%	100%
		BIN 2	Resfinder	lnu(B)_2	Lincomicina; Clindamicina	100%	100%
		BIN 2	Resfinder	lnu(C)_1	Lincomicina	99,19%	100%
		BIN 2	CARD	ACI-1	Cefalosporina; Penam; Penem	99,53%	100%
		BIN 2	CARD	lsaE	Lincosamida; Macrólido; Oxazolidinona; Fenicol; Pleuromutilina; Estreptogramina; Tetraciclina	100%	100%
		BIN 2	CARD	lnuC	Lincosamida	99,19%	100%
14.SRX3743979	2						
15.SRX3743978	2	-	-	-	-	-	-

Nov. 2014	16.SRX3743977	2	BIN 2	Resfinder	erm(B)_18	Eritromicina; Lincomicina; Clindamicina; Quinopristina; Pristinamicina IA; Virginiamicina S	99,21%	99,48%
			BIN 2	CARD	ErmB	Lincosamida; Macrólido; Estreptogramina	99,60%	99,87%
	17.SRX3743976	3	BIN 1	Resfinder	lnu(B)_2	Lincomicina; Clindamicina	100%	100%
			BIN 1	Resfinder	erm(B)_18	Eritromicina; Lincomicina; Clindamicina; Quinopristina; Pristinamicina IA; Virginiamicina S	98,95%	99,74%
			BIN 1	CARD	lsaE	Lincosamida; Macrólido; Oxazolidinona; Fenicol; Pleuromutilina; Estreptogramina; Tetraciclina	100%	100%
			BIN 1	CARD	ErmB	Lincosamida; Macrólido; Estreptogramina	99,73%	100%
			BIN 2	Resfinder	lnu(A)_1	Lincomicina	98,97%	100%
			BIN 2	Resfinder	lnu(B)_2	Lincomicina; Clindamicina	100%	100%
	18.SRX3743975	4	BIN 2	Resfinder	erm(B)_18	Eritromicina; Lincomicina; Clindamicina; Quinopristina; Pristinamicina IA; Virginiamicina S	98,95%	100%
			BIN 2	CARD	lnuA	Lincosamida	98,97%	100%

BIN 2	CARD	IsaE	Lincosamida; Macróido; Oxazolidinona; Fenicol; Pleuromutilina; Estreptogramina; Tetraciclina	100%	100%
BIN 2	CARD	ErmB	Lincosamida; Macróido; Estreptogramina	99,60%	99,87%

4.9. Anotación del genoma

Siguiendo con el proceso de la anotación genómica, se integraron los resultados del binning realizado con MetaBAT2 y MaxBin2 sobre el ensamblado del metagenoma SRX3743975. La optimización con DAS Tool permitió consolidar un total de tres bins, a partir de 2372 contigs clasificados sobre un total de 26273 contigs ensamblados. Estos resultados se presentan en la Figura 15, que muestra la distribución de puntuaciones (binScore > 0.25) para cada uno de los métodos de binning utilizados. El gráfico sugiere una agrupación eficiente por parte de DAS Tool, con valores de binScore consistentes y competitivos respecto a los métodos individuales.

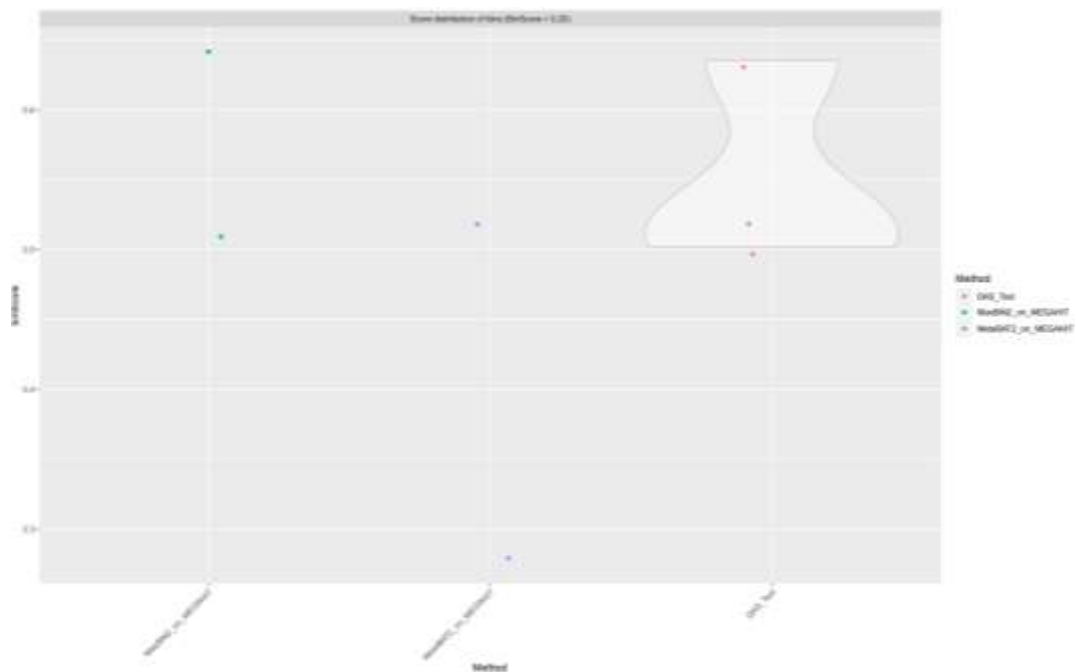


Figura 15. Distribución de puntuaciones de los bins generados con DAS Tool

Posteriormente, los bins obtenidos por DAS Tool fueron evaluados mediante CheckM. La Figura 16 muestra la presencia de genes de copia única, genes faltantes, heterogeneidad y contaminación por cada bin.

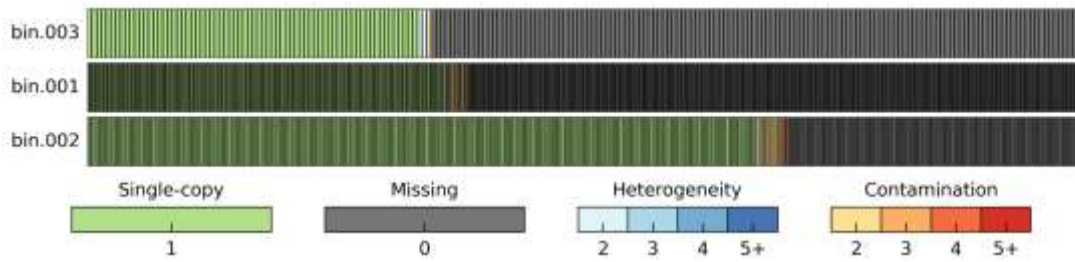


Figura 16. Evaluación de la calidad los bins obtenidos con DAS Tool

Como se puede ver también en la Tabla 10, los resultados evidencian que el bin.002 presentó un nivel de completitud del 66.8 % y una contaminación del 4.01 %, siendo además identificado como perteneciente al género *Lactobacillus*, correspondiente a una bacteria de la familia *Lactobacillaceae*.

Tabla 10. Evaluación de linaje de CheckM

Bin Name	Marker Lineage	# Genomes	# Markers	# Marker Sets	0	1	2	3	4	5+	Completeness	Contamination
bin_001	g_Prevotella	55	567	296	351	203	11	2	0	0	36.22	2.01
bin_002	g_Lactobacillus	91	430	154	128	289	11	1	0	1	66.8	4.01
bin_003	p_Actinobacteria	31	211	124	138	70	3	0	0	0	24.87	1.16

Los resultados obtenidos de la herramienta Annotate Multiple Microbial Assemblies fueron los siguientes:

En el Bin.002, correspondiente a una bacteria del género *Lactobacillus*, se obtuvo un total de 1,181 contigs y una longitud de 2,406,080 pares de bases. Este tamaño se encuentra dentro del rango característico para miembros del género, cuyo genoma puede oscilar entre 1.8 y 3.3 Mb, dependiendo de la especie y su nicho ecológico (Kant et al., 2011). El contenido de GC fue del 37.30 %, valor también concordante con lo reportado para especies adaptadas a ambientes fermentativos o al tracto gastrointestinal, en donde el contenido de guanina y citosina suele situarse entre 33 % y 51 % (Kant et al., 2011; Makarova et al., 2006).

La anotación estructural del bin 002 permitió identificar un total de 4,070 elementos genómicos, de los cuales 3,673 fueron genes codificantes de proteínas y 397 fueron elementos no codificantes. Dentro de estos últimos se incluyeron 84 genes de ARN (rRNA y tRNA), 305 regiones repetidas y elementos asociados a sistemas CRISPR, distribuidos en 1 arreglo CRISPR, 4 repeticiones CRISPR y 3 espaciadores.

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

El control de calidad, realizado con herramientas como FastQC, Trimmomatic y MultiQC, permitió obtener secuencias con altos estándares de calidad, con puntajes Phred superiores a 34 y sin lecturas de baja calidad.

Se logró recuperar, ensamblar y analizar metagenómicamente un total de 18 conjuntos de datos de vinaza provenientes de caña de azúcar, organizados en seis muestras con tres réplicas cada una.

La clasificación taxonómica reveló una riqueza bacteriana en las muestras analizadas, con predominancia de los filos *Firmicutes*, *Actinobacteria* y *Proteobacteria*. La especie, *Megasphaera elsdenii* fue la más predominante en múltiples metagenomas, seguida por representantes del género *Lactobacillus* como *L. amylovorus*, *L. delbrueckii* y *L. fermentum*, lo cual evidencia una comunidad microbiana fermentativa típica de residuos orgánicos agroindustriales

El metagenoma SRX3743975, que presentó la mayor abundancia relativa de bacterias del género *Lactobacillus*, fue utilizado como modelo para la validación metodológica. El ensamblaje de sus secuencias mediante MEGAHIT ofreció mejores métricas que metaSPAdes, destacando en longitud de contigs, ausencia de ambigüedades (N's) y proporción de secuencias largas. La calidad del ensamblaje fue clave para obtener bins representativos y facilitar la identificación precisa de genes de resistencia.

En total, se detectaron genes de resistencia a antibióticos (ARGs) en 16 de los 18 metagenomas analizados, con niveles de identidad y cobertura superiores al 90 %. La búsqueda mediante ABRicate, en conjunto con las bases de datos CARD y ResFinder, permitió identificar genes como *lnu(A)*, *lnu(B)*, *lnu(C)*, *erm(B)*, *lsa(E)*, *vat(E)* y *blaACI-*

1, asociados a la resistencia de antibióticos como lincosamidas, macrólidos, betalactámicos, estreptograminas y otros antibióticos de importancia médica y veterinaria. Los genes detectados fueron encontrados en múltiples bins, y en muchos casos, coexistieron varios ARGs en una misma unidad genómica. Esta coocurrencia sugiere la presencia de clústeres multirresistencia y refleja la circulación activa de mecanismos de resistencia en las comunidades microbianas de la vinaza. Además, se identificaron genes con mecanismos distintos de resistencia a un mismo antibiótico, como *lnu(B)_2* e *lsa(E)*.

A partir del ensamblaje optimizado y del agrupamiento de contigs mediante MetaBAT2 y DAS Tool, se obtuvieron bins de alta calidad evaluados con CheckM. Entre ellos, se destacó el bin.002, correspondiente a *Lactobacillus*, con un genoma de 2,4 Mb, contenido GC del 37,3 %, completitud del 66,8 % y contaminación del 4,01 %, según los parámetros evaluados con CheckM. La anotación estructural identificó 3.673 genes codificantes, 84 ARN y 305 regiones repetidas, incluidos elementos CRISPR. Estas métricas coinciden con lo reportado en estudios de genómica comparativa para bacterias de la familia *Lactobacillaceae*.

5.2. Recomendaciones

Se recomienda incorporar un paso previo de descontaminación de las secuencias, mediante el mapeo y eliminación de lecturas contaminantes antes del ensamblaje con herramientas como Bowtie2, que permiten alinear las lecturas contra genomas de referencia específicos y filtrar aquellas que no corresponden al contenido microbiano de interés.

Se sugiere implementar un enfoque de co-ensamblaje o binning basado en cobertura diferencial entre múltiples muestras, lo cual puede mejorar la calidad y resolución de los

bins generados para obtener reconstrucciones genómicas más completas y menos contaminadas.

Se recomienda replicar este estudio utilizando muestras de vinaza obtenidas directamente en ingenios ecuatorianos, con el fin de generar datos específicos del contexto local y fortalecer la aplicabilidad regional de los resultados.

Se sugiere incorporar técnicas experimentales complementarias, como PCR cuantitativa, cultivo bacteriano selectivo o pruebas de susceptibilidad antimicrobiana, permitiría validar la expresión funcional de los genes de resistencia detectados *in silico* y confirmar su relevancia desde una perspectiva fenotípica.

En futuras investigaciones, se sugiere profundizar en el análisis funcional de los bins obtenidos, explorando la presencia de elementos genéticos móviles, sistemas de secreción, plásmidos y rutas metabólicas asociadas a resistencia, con el fin de caracterizar de manera más completa el potencial adaptativo y evolutivo de las bacterias identificadas.

Se recomienda la integración de enfoques multi-ómicos, como la metatranscriptómica o la metaproteómica, lo cual representaría una línea de trabajo valiosa para identificar genes de resistencia con expresión activa en condiciones ambientales reales, superando la limitación de detectar únicamente su presencia genómica.

Para finalizar, se sugiere mantener una revisión periódica del pipeline bioinformático empleado, evaluando la disponibilidad de herramientas actualizadas que mejoren la sensibilidad, especificidad o eficiencia del análisis metagenómico. En caso de identificar mejoras metodológicas, estas deberían documentarse como recomendaciones para su implementación futura, garantizando la evolución continua del enfoque analítico y la calidad de los resultados generados.

REFERENCIAS

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>
- Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., Edalatmand, A., Petkau, A., Syed, S. A., Tsang, K. K., Baker, S. J. C., Dave, M., McCarthy, M. C., Mukiri, K. M., Nasir, J. A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., ... McArthur, A. G. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*, *51*(D1), D690–D699. <https://doi.org/10.1093/nar/gkac920>
- Ammor, M. S., Belén Flórez, A., & Mayo, B. (2007). Antibiotic resistance in non-enterococcal lactic acid bacteria and bifidobacteria. *Food Microbiology*, *24*(6), 559–570. <https://doi.org/10.1016/j.fm.2006.11.001>
- Anampa, D., Benites, C., Lázaro, C., Espinoza, J., Angulo, P., Díaz, D., Manchego, A., & Rojas, M. (2020). Detección del gen ermB asociado a la resistencia a macrólidos en cepas de *Campylobacter* aisladas de pollos comercializados en Lima, Perú. *Revista Panamericana de Salud Pública*, *44*, 1. <https://doi.org/10.26633/RPSP.2020.60>
- Antony Cristofer, R. R., Diego Santiago, C. C., Ricardo Javier, C. C., & Jissel Catalina, R. R. (2023). Panorama Energético de los Biocombustibles en el Ecuador. *Ciencia*

Latina Revista Científica Multidisciplinar, 7(4), 10254–10275.

https://doi.org/10.37811/cl_rcm.v7i4.7729

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S., Drake, M. M., Brettin, T. S., ... Yu, D. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology*, 36(7), 566–569. <https://doi.org/10.1038/nbt.4163>

Asobanca. (2022). Guía de producción de azúcar en Ecuador. *Asociación de Bancos Privados Del Ecuador*. <https://asobanca.org.ec/wp-content/uploads/2022/12/8.-Guia-Produccion-de-azucar.pdf>

Babraham Institute. (2020). *FastQC: Documentation*. Babraham Institute.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>

Braga, L. P. P., Alves, R. F., Dellias, M. T. F., Navarrete, A. A., Basso, T. O., & Tsai, S. M. (2017). Vinasse fertirrigation alters soil resistome dynamics: an analysis based on metagenomic profiles. *BioData Mining*, 10(1), 17. <https://doi.org/10.1186/s13040-017-0138-4>

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason, J. A., Stevens, R., Vonstein, V., Wattam, A. R., & Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and

annotating batches of genomes. *Scientific Reports*, 5(1), 8365.
<https://doi.org/10.1038/srep08365>

Cabral, L. da S., & Weimer, P. J. (2024). *Megasphaera elsdenii*: Its Role in Ruminant Nutrition and Its Potential Industrial Application for Organic Acid Biosynthesis. *Microorganisms*, 12(1), 219. <https://doi.org/10.3390/microorganisms12010219>

Callejas, C., López, I., Bovio-Winkler, P., Etchebehere, C., & Borzacconi, L. (2022). Temporal analysis of the microbiota involved in the anaerobic degradation of sugarcane vinasse in a full-scale methanogenic UASB reactor. *Biomass Conversion and Biorefinery*, 12(9), 3887–3897. <https://doi.org/10.1007/s13399-021-01281-8>

Casey, E., Mahony, J., Neve, H., Noben, J.-P., Dal Bello, F., & van Sinderen, D. (2015). Novel Phage Group Infecting *Lactobacillus delbrueckii* subsp. *lactis*, as Revealed by Genomic and Proteomic Analysis of Bacteriophage Ldl1. *Applied and Environmental Microbiology*, 81(4), 1319–1326.
<https://doi.org/10.1128/AEM.03413-14>

Cassman, N. A., Lourenço, K. S., do Carmo, J. B., Cantarella, H., & Kuramae, E. E. (2018). Genome-resolved metagenomics of sugarcane vinasse bacteria. *Biotechnology for Biofuels*, 11(1), 48. <https://doi.org/10.1186/s13068-018-1036-9>

Centro de Investigación de la Caña de Azúcar del Ecuador. (2018). *Uso de vinaza como fertilizante en caña de azúcar*. <https://cincae.org/wp-content/uploads/2013/04/A%C3%B1o-20.pdf>

Chang, S.-C., Lin, L.-C., & Lu, J.-J. (2022). Un plásmido multirresistente portador de *lnu* (A) derivado de *Staphylococcus lugdunensis* de secuencia tipo 3 resistente a la meticilina podría contribuir a la resistencia a los antimicrobianos en estafilococos.

Antimicrobial Agents and Chemotherapy, 66(8). <https://doi.org/10.1128/aac.00197-22>

- Christofoletti, C. A., Escher, J. P., Correia, J. E., Marinho, J. F. U., & Fontanetti, C. S. (2013). Sugarcane vinasse: Environmental implications of its use. *Waste Management*, 33(12), 2752–2761. <https://doi.org/10.1016/j.wasman.2013.09.005>
- De Luca, S., Nicholson, P., Magistrali, C. F., García-Martín, A. B., Rychener, L., Zeeh, F., Frey, J., & Perreten, V. (2018). Transposon-associated lincosamide resistance *lnu* (C) gene identified in *Brachyspira hyodysenteriae* ST83. *Veterinary Microbiology*, 214, 51–55. <https://doi.org/10.1016/j.vetmic.2017.12.003>
- Del M. Chaile, A. P., Uboldi, M. E., & Elsa Ferreyra, M. M. (2024). Tratamiento químico de vinaza de caña de azúcar con peróxido de hidrógeno. *Revista de Ciencias Ambientales*, 59(1), 1–19. <https://doi.org/10.15359/rca.59-1.5>
- dos Reis, C. M., Carosia, M. F., Sakamoto, I. K., Amâncio Varesche, M. B., & Silva, E. L. (2015). Evaluation of hydrogen and methane production from sugarcane vinasse in an anaerobic fluidized bed reactor. *International Journal of Hydrogen Energy*, 40(27), 8498–8509. <https://doi.org/10.1016/j.ijhydene.2015.04.136>
- Durso, L. M., & Cook, K. L. (2019). One Health and Antibiotic Resistance in Agroecosystems. *EcoHealth*, 16(3), 414–419. <https://doi.org/10.1007/s10393-018-1324-7>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>

- Farfán-Cano, G., Parra-Vera, H., Ávila-Choez, A., Silva-Rojas, G., & Farfán-Cano, S. (2020). Primera identificación en Ecuador de *Comamonas kerstersii* como agente infeccioso. *Revista Chilena de Infectología*, 37(2), 179–181. <https://doi.org/10.4067/s0716-10182020000200179>
- Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D., & Aarestrup, F. M. (2022). ResFinder – an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microbial Genomics*, 8(1). <https://doi.org/10.1099/mgen.0.000748>
- Forsberg, K. J., Reyes, A., Wang, B., Selleck, E. M., Sommer, M. O. A., & Dantas, G. (2012). The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science*, 337(6098), 1107–1111. <https://doi.org/10.1126/science.1220761>
- Gaude, G., & Hattiholli, J. (2013). Rising bacterial resistance to beta-lactam antibiotics: Can there be solutions? *Journal of Dr. NTR University of Health Sciences*, 2(1), 4. <https://doi.org/10.4103/2277-8632.108504>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- HBC Training. (n.d.). *Assessing quality with FastQC*. Retrieved August 28, 2025, from https://hbctraining.github.io/Intro-to-rnaseq-hpc-salmon/lessons/qc_fastqc_assessment.html
- Hildebrand, F., Meyer, A., & Eyre-Walker, A. (2010). Evidence of Selection upon Genomic GC-Content in Bacteria. *PLoS Genetics*, 6(9), e1001107. <https://doi.org/10.1371/journal.pgen.1001107>

- Jian, Z., Zeng, L., Xu, T., Sun, S., Yan, S., Yang, L., Huang, Y., Jia, J., & Dou, T. (2021). Antibiotic resistance genes in bacteria: Occurrence, spread, and control. *Journal of Basic Microbiology*, *61*(12), 1049–1070. <https://doi.org/10.1002/jobm.202100201>
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, *7*, e7359. <https://doi.org/10.7717/peerj.7359>
- Kant, R., Blom, J., Palva, A., Siezen, R. J., & de Vos, W. M. (2011). Genómica comparativa de *Lactobacillus*. *Microbial Biotechnology*, *4*(3), 323–332. <https://doi.org/10.1111/j.1751-7915.2010.00215.x>
- Leal G., I., Chirinos, E., Leal, M., Morán, H., & Barrera, W. (2003). Caracterización fisicoquímica de la vinaza del Agave cocui y su posible uso agroindustrial. *Multiciencias*, *3*(2). <https://www.redalyc.org/articulo.oa?id=90430202>
- Lemos, L. N., Pedrinho, A., Vasconcelos, A. T. R. de, Tsai, S. M., & Mendes, L. W. (2021). Amazon deforestation enriches antibiotic resistance genes. *Soil Biology and Biochemistry*, *153*, 108110. <https://doi.org/10.1016/j.soilbio.2020.108110>
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics*, *31*(10), 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Li, W., Jiao, D., Kang, J., Yu, R., Zhao, W., Xu, C., Li, R., Du, X.-D., & Yao, H. (2023). Emergence of *lnu(C)* variant conferring lincomycin resistance in *Campylobacter coli* of chicken origin. *International Journal of Food Microbiology*, *388*, 110098. <https://doi.org/10.1016/j.ijfoodmicro.2023.110098>

- Liu, F., Luo, Y., Xu, T., Lin, H., Qiu, Y., & Li, B. (2024). Current examining methods and mathematical models of horizontal transfer of antibiotic resistance genes in the environment. *Frontiers in Microbiology*, 15. <https://doi.org/10.3389/fmicb.2024.1371388>
- Lozano, C., Aspiroz, C., Saenz, Y., Ruiz-Garcia, M., Royo-Garcia, G., Gomez-Sanz, E., Ruiz-Larrea, F., Zarazaga, M., & Torres, C. (2012). Genetic environment and location of the *lnu(A)* and *lnu(B)* genes in methicillin-resistant *Staphylococcus aureus* and other staphylococci of animal and human origin. *Journal of Antimicrobial Chemotherapy*, 67(12), 2804–2808. <https://doi.org/10.1093/jac/dks320>
- Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science*, 3, e104. <https://doi.org/10.7717/peerj-cs.104>
- Luo, Q., Wang, Y., & Xiao, Y. (2020). Prevalence and transmission of mobilized colistin resistance (*mcr*) gene in bacteria common to animals and humans. *Biosafety and Health*, 2(2), 71–78. <https://doi.org/10.1016/j.bsheal.2020.05.001>
- Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., Shakhova, V., Grigoriev, I., Lou, Y., Rohksar, D., Lucas, S., Huang, K., Goodstein, D. M., Hawkins, T., Plengvidhya, V., ... Mills, D. (2006). Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences*, 103(42), 15611–15616. <https://doi.org/10.1073/pnas.0607117103>
- Mayrhofer, S., Domig, K. J., Amtmann, E., Van Hoek, A. H. A. M., Petersson, A., Mair, C., Mayer, H. K., & Kneifel, W. (2007). Antibiotic Susceptibility of *Bifidobacterium*

thermophilum and Bifidobacterium pseudolongum Isolates from Animal Sources. *Journal of Food Protection*, 70(1), 119–124. <https://doi.org/10.4315/0362-028X-70.1.119>

Mondal, A. H., Khare, K., Saxena, P., Debnath, P., Mukhopadhyay, K., & Yadav, D. (2024). A Review on Colistin Resistance: An Antibiotic of Last Resort. *Microorganisms*, 12(4), 772. <https://doi.org/10.3390/microorganisms12040772>

Mota, V. T., Delforno, T. P., Ribeiro, J. C., Zaiat, M., & Oliveira, V. M. de. (2024). Understanding microbiome dynamics and functional responses during acidogenic fermentation of sucrose and sugarcane vinasse through metatranscriptomic analysis. *Environmental Research*, 246, 118150. <https://doi.org/10.1016/j.envres.2024.118150>

Murigneux, V., Rai, S. K., Furtado, A., Bruxner, T. J. C., Tian, W., Harliwong, I., Wei, H., Yang, B., Ye, Q., Anderson, E., Mao, Q., Drmanac, R., Wang, O., Peters, B. A., Xu, M., Wu, P., Topp, B., Coin, L. J. M., & Henry, R. J. (2020). Comparison of long-read methods for sequencing and assembly of a plant genome. *GigaScience*, 9(12). <https://doi.org/10.1093/gigascience/giaa146>

O’Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., ... Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1), D733–D745. <https://doi.org/10.1093/nar/gkv1189>

- Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, *12*(1), 385. <https://doi.org/10.1186/1471-2105-12-385>
- Organización Mundial de la Salud. (2020). *Resistencia a los antibióticos*. <https://www.who.int/es/news-room/fact-sheets/detail/antibiotic-resistance>
- Ospina León, L. J., Manotas-Duque, D., & Ramirez Malule, H. D. (2022). Desafíos y oportunidades de la Vinaza de caña de azúcar. Un análisis bibliométrico. *INGENIERÍA Y COMPETITIVIDAD*, *25*(1). <https://doi.org/10.25100/iyc.v25i1.12144>
- Pandey, N., & Cascella, M. (2023). *Beta-Lactam Antibiotics*.
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Petinaki, E., Guérin-Faublée, V., Pichereau, V., Villers, C., Achard, A., Malbruny, B., & Leclercq, R. (2008). Lincomycin Resistance Gene *lnu(D)* in *Streptococcus uberis*. *Antimicrobial Agents and Chemotherapy*, *52*(2), 626–630. <https://doi.org/10.1128/AAC.01126-07>
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, *35*(9), 833–844. <https://doi.org/10.1038/nbt.3935>
- Rands, C. M., Starikova, E. V., Brüßow, H., Kriventseva, E. V., Govorun, V. M., & Zdobnov, E. M. (2018). ACI-1 beta-lactamase is widespread across human gut

microbiomes in Negativicutes due to transposons harboured by tailed prophages. *Environmental Microbiology*, 20(6), 2288–2300. <https://doi.org/10.1111/1462-2920.14276>

Ríos-Covián, D., Ruas-Madiedo, P., Margolles, A., Gueimonde, M., de los Reyes-Gavilán, C. G., & Salazar, N. (2016). Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Frontiers in Microbiology*, 7. <https://doi.org/10.3389/fmicb.2016.00185>

Rodrigues Reis, C. E., & Hu, B. (2017). Vinasse from Sugarcane Ethanol Production: Better Treatment or Better Utilization? *Frontiers in Energy Research*, 5. <https://doi.org/10.3389/fenrg.2017.00007>

Saremi, B., Kohls, M., Liebig, P., Siebert, U., & Jung, K. (2021). Measuring reproducibility of virus metagenomics analyses using bootstrap samples from FASTQ-files. *Bioinformatics*, 37(8), 1068–1075. <https://doi.org/10.1093/bioinformatics/btaa926>

Seemann, T. (2016). ABRicate: mass screening of contigs for antibiotic resistance genes. *GitHub*. <https://github.com/tseemann/abricate>

Shan, X., Li, X.-S., Schwarz, S., Chen, Y., Xu, C., & Du, X.-D. (2022). Plasmid-Assisted Horizontal Transfer of a Large *lsa(E)*-Carrying Genomic Island in *Enterococcus faecalis*. *Microbiology Spectrum*, 10(4). <https://doi.org/10.1128/spectrum.00154-22>

Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., & Banfield, J. F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, 3(7), 836–843. <https://doi.org/10.1038/s41564-018-0171-1>

- Skandalis, N., Maeusli, M., Papafotis, D., Miller, S., Lee, B., Theologidis, I., & Luna, B. (2021). Environmental Spread of Antibiotic Resistance. *Antibiotics*, *10*(6), 640. <https://doi.org/10.3390/antibiotics10060640>
- Suárez Valles, M. Á., & Garzón Avilés, M. J. (2024). Efecto de la vinaza en el rendimiento azucarero y calidad de caña de azúcar, aplicada con el riego, en el Ingenio Valdez; Milagro. *RECIMUNDO*, *8*(3), 371–391. [https://doi.org/10.26820/recimundo/8.\(3\).julio.2024.371-391](https://doi.org/10.26820/recimundo/8.(3).julio.2024.371-391)
- Sundlov, J. A., & Gulick, A. M. (2009). Insights into Resistance against Lincosamide Antibiotics. *Structure*, *17*(12), 1549–1550. <https://doi.org/10.1016/j.str.2009.11.001>
- Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, *2*(1), 3. <https://doi.org/10.1186/2042-5783-2-3>
- Tien, Y.-C., Li, B., Zhang, T., Scott, A., Murray, R., Sabourin, L., Marti, R., & Topp, E. (2017). Impact of dairy manure pre-application treatment on manure composition, soil dynamics of antibiotic resistance genes, and abundance of antibiotic-resistance genes on vegetables at harvest. *Science of The Total Environment*, *581–582*, 32–39. <https://doi.org/10.1016/j.scitotenv.2016.12.138>
- Topp, E., Chapman, R., Devers-Lamrani, M., Hartmann, A., Marti, R., Martin-Laurent, F., Sabourin, L., Scott, A., & Sumarah, M. (2013). Accelerated Biodegradation of Veterinary Antibiotics in Agricultural Soil following Long-Term Exposure, and Isolation of a Sulfamethazine-degrading *Microbacterium* sp. *Journal of Environmental Quality*, *42*(1), 173–178. <https://doi.org/10.2134/jeq2012.0162>

- van Hoek, A. H. A. M., Mevius, D., Guerra, B., Mullany, P., Roberts, A. P., & Aarts, H. J. M. (2011). Acquired Antibiotic Resistance Genes: An Overview. *Frontiers in Microbiology*, 2. <https://doi.org/10.3389/fmicb.2011.00203>
- von Wintersdorff, C. J. H., Penders, J., van Niekerk, J. M., Mills, N. D., Majumder, S., van Alphen, L. B., Savelkoul, P. H. M., & Wolfs, P. F. G. (2016). Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. *Frontiers in Microbiology*, 7. <https://doi.org/10.3389/fmicb.2016.00173>
- Wendlandt, S., Lozano, C., Kadlec, K., Gómez-Sanz, E., Zarazaga, M., Torres, C., & Schwarz, S. (2013). The enterococcal ABC transporter gene *lsa(E)* confers combined resistance to lincosamides, pleuromutilins and streptogramin A antibiotics in methicillin-susceptible and methicillin-resistant *Staphylococcus aureus*. *Journal of Antimicrobial Chemotherapy*, 68(2), 473–475. <https://doi.org/10.1093/jac/dks398>
- Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20(1), 257. <https://doi.org/10.1186/s13059-019-1891-0>
- Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, 15(3), R46. <https://doi.org/10.1186/gb-2014-15-3-r46>
- World Health Organization. (2019). *Critically Important Antimicrobials for Human Medicine: 6th revision*. <https://www.who.int/publications/i/item/9789241515528>
- Wu, L., Wen, C., Qin, Y., Yin, H., Tu, Q., Van Nostrand, J. D., Yuan, T., Yuan, M., Deng, Y., & Zhou, J. (2015). Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiology*, 15(1), 125. <https://doi.org/10.1186/s12866-015-0450-4>

- Wu, Y., Zaiden, N., & Cao, B. (2018). The Core- and Pan-Genomic Analyses of the Genus *Comamonas*: From Environmental Adaptation to Potential Virulence. *Frontiers in Microbiology*, 9. <https://doi.org/10.3389/fmicb.2018.03096>
- Yang, Y., Xie, S., He, F., Xu, Y., Wang, Z., Ihsan, A., & Wang, X. (2024). Recent development and fighting strategies for lincosamide antibiotic resistance. *Clinical Microbiology Reviews*, 37(2). <https://doi.org/10.1128/cmr.00161-23>
- Zúñiga Cerón, V., & Gandini Ayerbe, M. A. (2013). Caracterización ambiental de las vinazas de residuos de caña de azúcar resultantes de la producción de etanol. *DYNA*, 80(177), 124–131. <https://revistas.unal.edu.co/index.php/dyna/article/view/27518>

ANEXOS

Anexo I. Tabla resumen de las estadísticas generales de MultiQC

<i>Sample</i>	<i>Dups</i>	<i>GC</i>	<i>Median len</i>	<i>Seqs</i>
<i>Trimmomatic on SRX3743975_forward_R1 paired</i>	1,61968282	46	224	0,426478
<i>Trimmomatic on SRX3743975_reverse_R2 paired</i>	2,26613639	45	174	0,426478
<i>Trimmomatic on SRX3743976_forward_R1 paired</i>	1,52365912	46	224	0,357855
<i>Trimmomatic on SRX3743976_reverse_R2 paired</i>	1,9911822	45	174	0,357855
<i>Trimmomatic on SRX3743977_forward_R1 paired</i>	2,94888255	39	224	0,345323
<i>Trimmomatic on SRX3743977_reverse_R2 paired</i>	3,19051049	38	184	0,345323
<i>Trimmomatic on SRX3743978_forward_R1 paired</i>	2,40806705	39	224	0,275443
<i>Trimmomatic on SRX3743978_reverse_R2 paired</i>	2,72367248	39	184	0,275443
<i>Trimmomatic on SRX3743979_forward_R1 paired</i>	1,2431237	44	224	0,315382
<i>Trimmomatic on SRX3743979_reverse_R2 paired</i>	1,36548425	44	174	0,315382
<i>Trimmomatic on SRX3743980_forward_R1 paired</i>	1,05035394	44	224	0,492812
<i>Trimmomatic on SRX3743980_reverse_R2 paired</i>	1,55173356	44	174	0,492812
<i>Trimmomatic on SRX3743981_forward_R1 paired</i>	0,9029244	44	224	0,436696
<i>Trimmomatic on SRX3743981_reverse_R2 paired</i>	1,53333466	44	174	0,436696
<i>Trimmomatic on SRX3743982_forward_R1 paired</i>	2,47838022	45	224	0,463837
<i>Trimmomatic on SRX3743982_reverse_R2 paired</i>	2,99380313	44	174	0,463837
<i>Trimmomatic on SRX3743983_forward_R1 paired</i>	3,63373444	39	224	0,48001
<i>Trimmomatic on SRX3743983_reverse_R2 paired</i>	4,12092263	39	174	0,48001
<i>Trimmomatic on SRX3743984_forward_R1 paired</i>	0,80564721	51	214	0,425293
<i>Trimmomatic on SRX3743984_reverse_R2 paired</i>	1,48329701	51	164	0,425293
<i>Trimmomatic on SRX3743985_forward_R1 paired</i>	0,95430064	51	214	0,492911
<i>Trimmomatic on SRX3743985_reverse_R2 paired</i>	1,8773214	51	164	0,492911
<i>Trimmomatic on SRX3743986_forward_R1 paired</i>	0,83555874	51	214	0,356227
<i>Trimmomatic on SRX3743986_reverse_R2 paired</i>	1,60967497	51	164	0,356227
<i>Trimmomatic on SRX3743987_forward_R1 paired</i>	0,41367948	47	224	0,534057

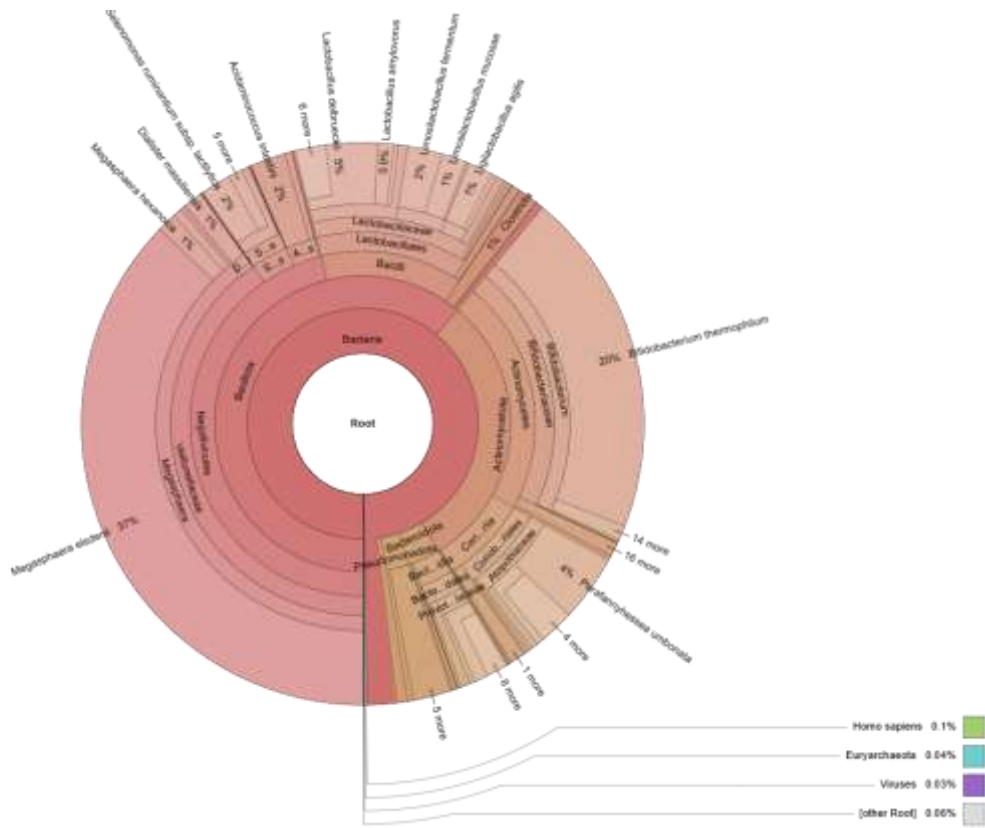
<i>Trimmomatic on SRX3743987_reverse_R2 paired</i>	1,02138334	47	174	0,534057
<i>Trimmomatic on SRX3743988_forward_R1 paired</i>	0,40479324	48	224	0,509714
<i>Trimmomatic on SRX3743988_reverse_R2 paired</i>	0,95021153	48	174	0,509714
<i>Trimmomatic on SRX3743989_forward_R1 paired</i>	0,31369515	49	224	0,40687
<i>Trimmomatic on SRX3743989_reverse_R2 paired</i>	1,06137035	48	174	0,40687
<i>Trimmomatic on SRX3743990_forward_R1 paired</i>	0,70196316	52	214	0,460746
<i>Trimmomatic on SRX3743990_reverse_R2 paired</i>	1,54494581	51	164	0,460746
<i>Trimmomatic on SRX3743991_forward_R1 paired</i>	0,64870977	51	224	0,447786
<i>Trimmomatic on SRX3743991_reverse_R2 paired</i>	1,56163991	51	164	0,447786
<i>Trimmomatic on SRX3743992_forward_R1 paired</i>	0,71335447	51	224	0,454737
<i>Trimmomatic on SRX3743992_reverse_R2 paired</i>	1,64166375	51	164	0,454737

Anexo 2. Parámetros utilizados en el Trimmomatic

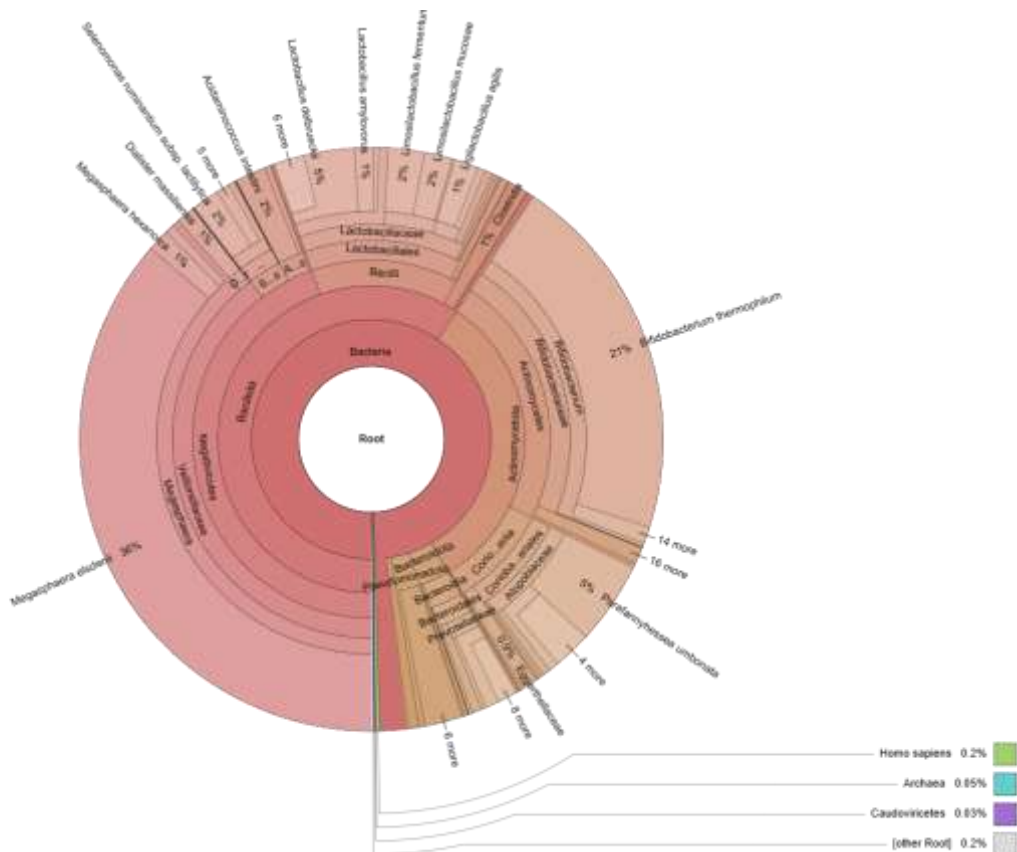
Tool Parameters

Input Parameter	Value
Single-end or paired-end reads?	pair_of_files
Input FASTQ file (R1/first of pair)	3: SRX3743978:forward
Input FASTQ file (R2/second of pair)	4: SRX3743978:reverse
Perform initial ILLUMINACLIP step?	false
Select Trimmomatic operation to perform	SLIDINGWINDOW
Number of bases to average across	4
Average quality required	30
Output trimlog file?	false
Output trimmomatic log messages?	false

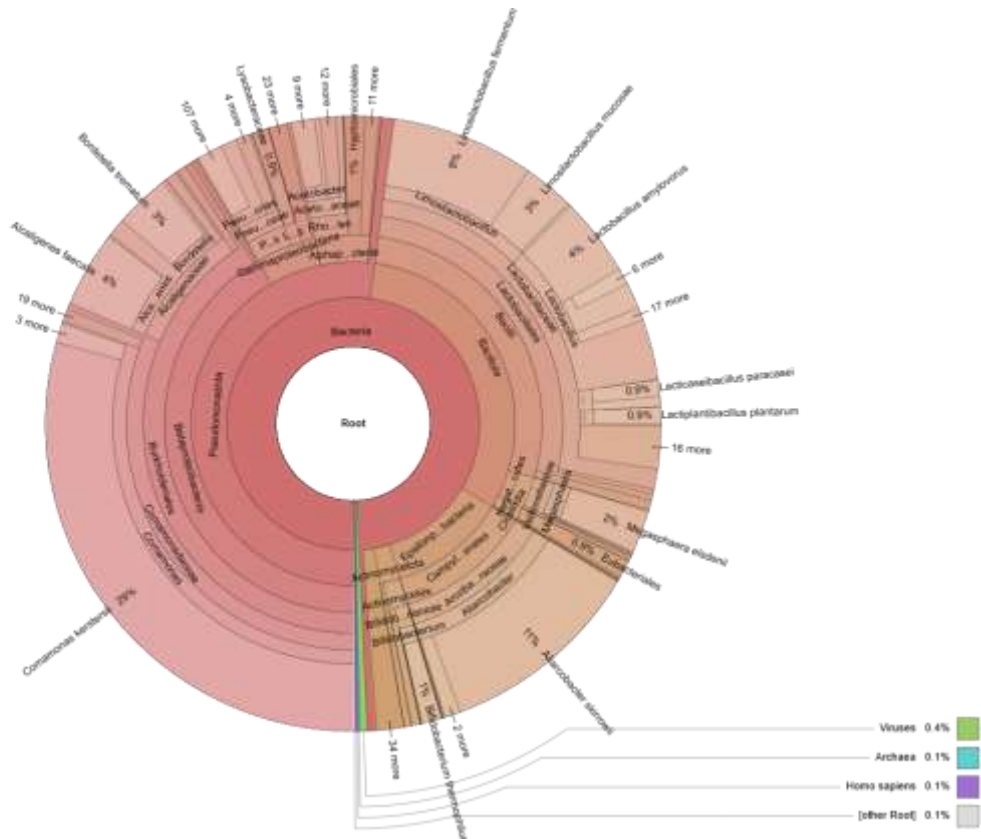
Anexo 3. Diagrama circular de Krona del metagenoma SRX3743991



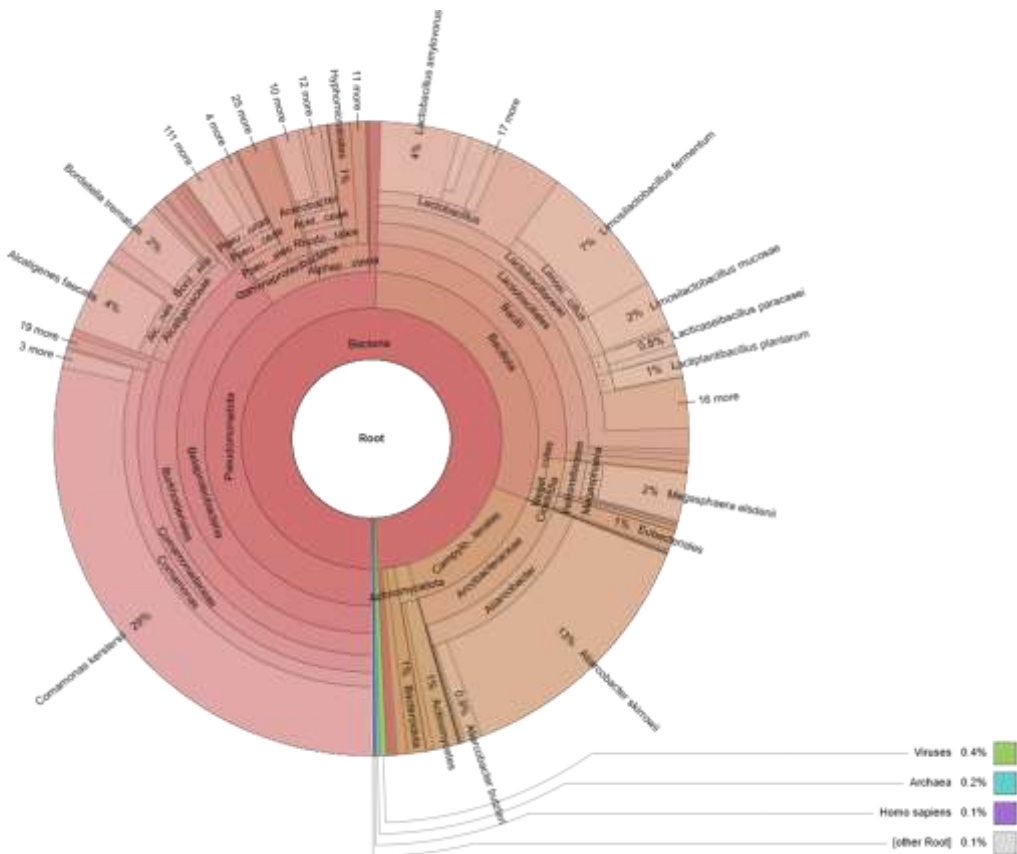
Anexo 4. Diagrama circular de Krona del metagenoma SRX3743990



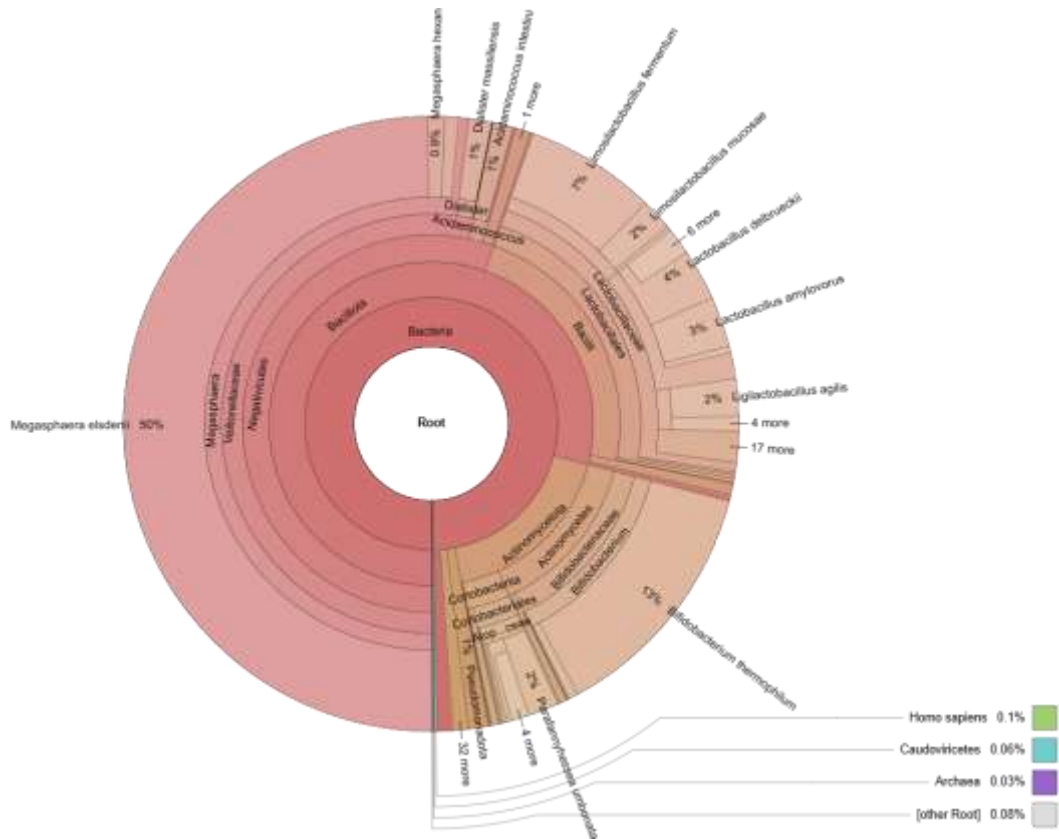
Anexo 5. Diagrama circular de Krona del metagenoma SRX3743989



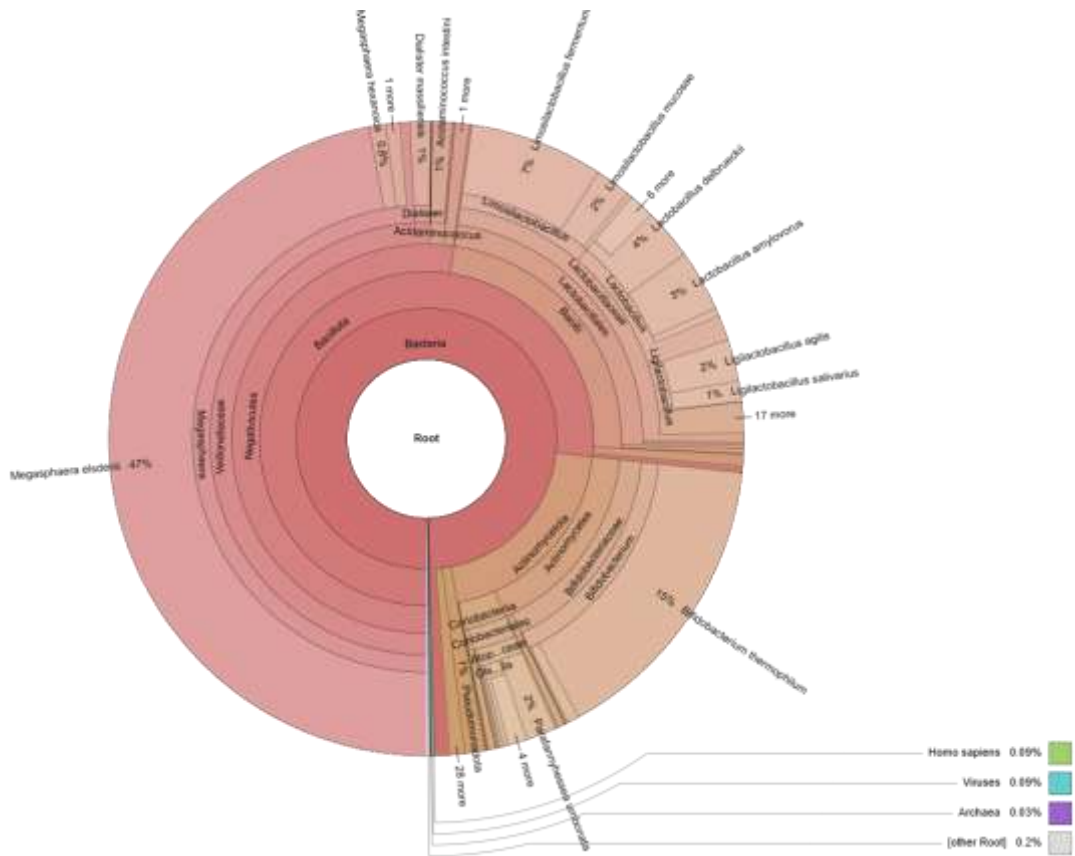
Anexo 6. Diagrama circular de Krona del metagenoma SRX3743988



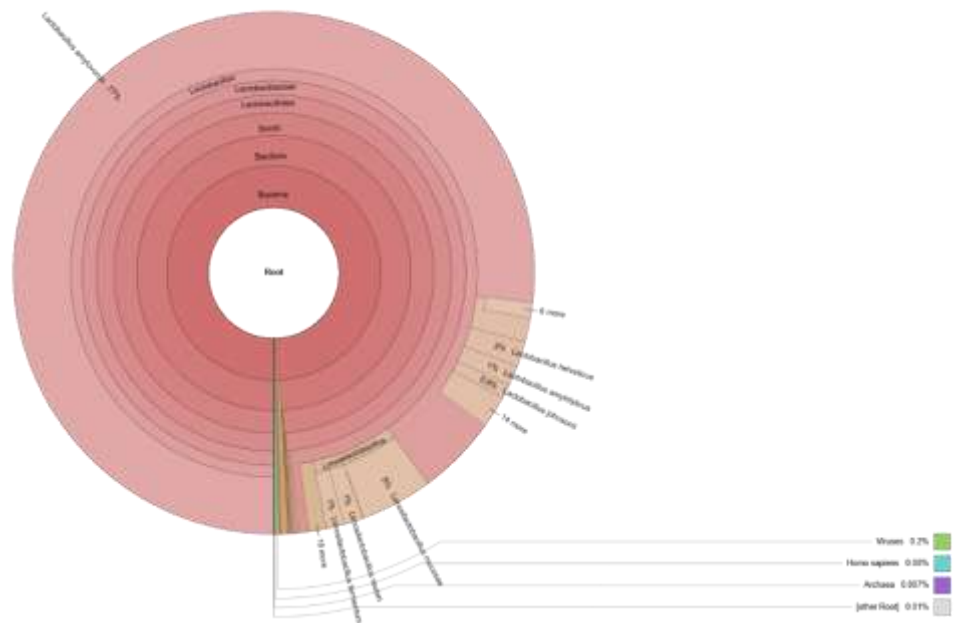
Anexo 7. Diagrama circular de Krona del metagenoma SRX3743985



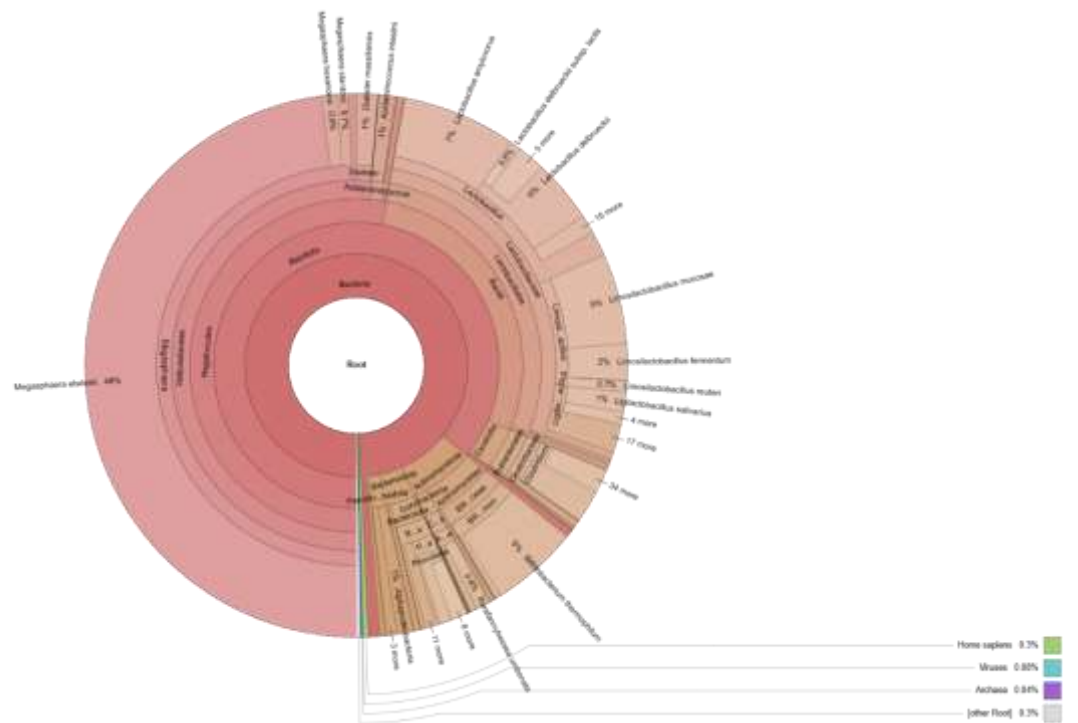
Anexo 8. Diagrama circular de Krona del metagenoma SRX3743984



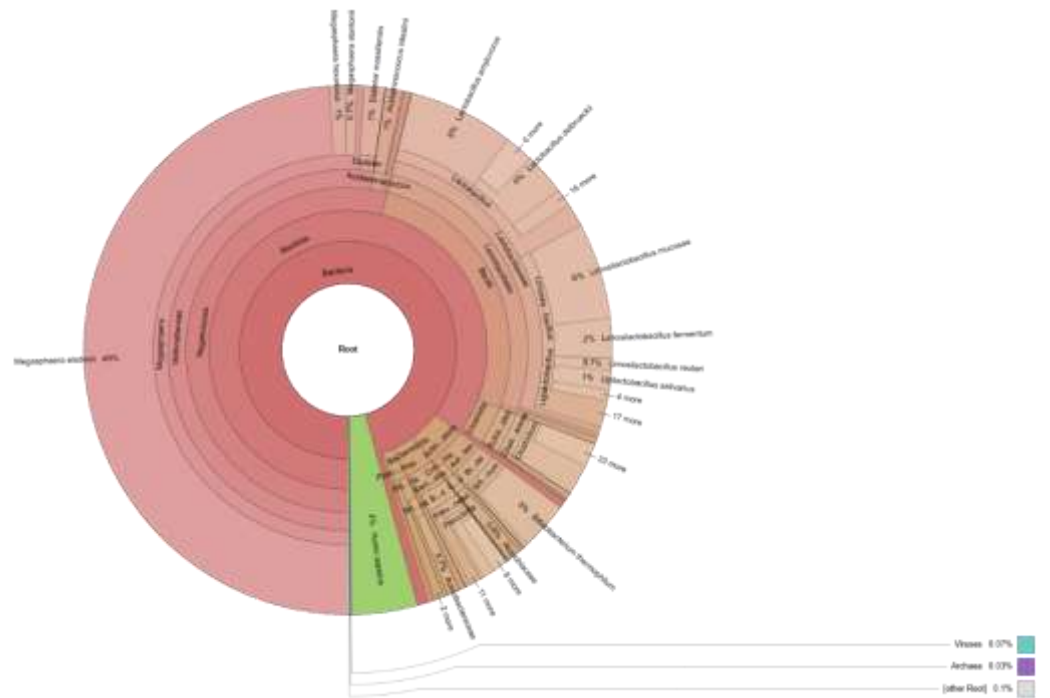
Anexo 9. Diagrama circular de Krona del metagenoma SRX3743983



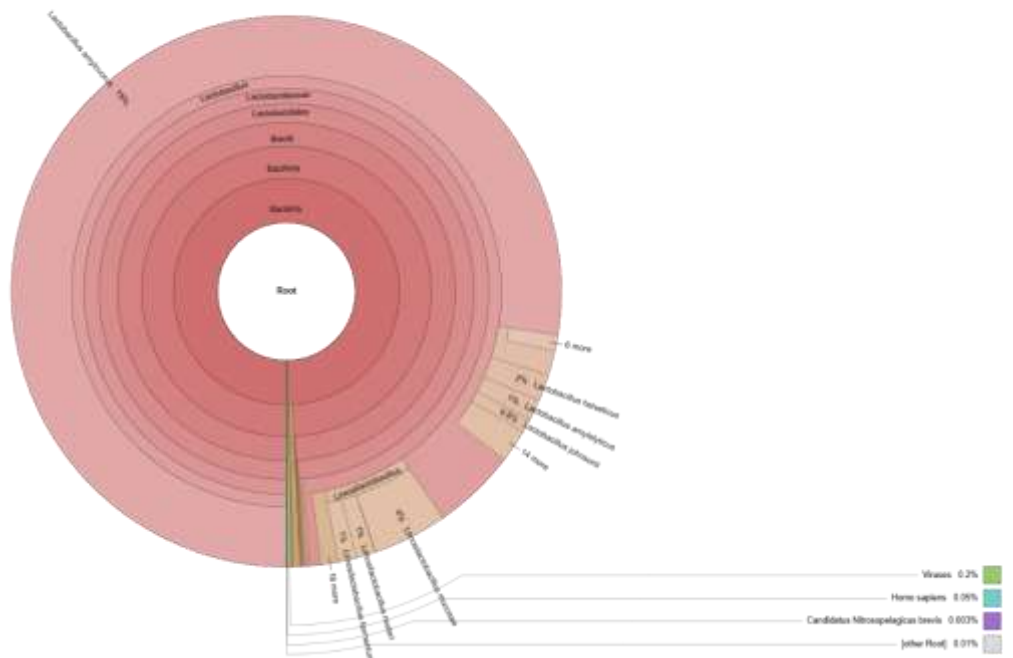
Anexo 10. Diagrama circular de Krona del metagenoma SRX3743981



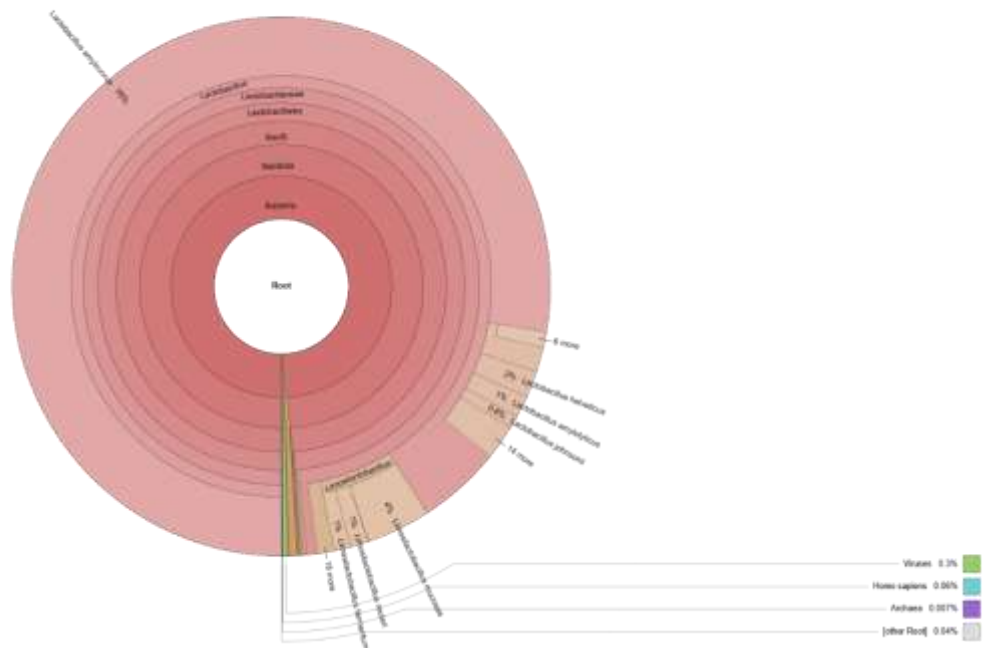
Anexo 11. Diagrama circular de Krona del metagenoma SRX3743979



Anexo 12. Diagrama circular de Krona del metagenoma SRX3743978



Anexo 13. Diagrama circular de Krona del metagenoma SRX3743977



Anexo 14. Diagrama circular de Krona del metagenoma SRX3743976

