

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE INGENIERÍA

MAESTRÍA EN BIOLOGÍA COMPUTACIONAL

Ensamblaje y anotación del genoma de una bacteria perteneciente a la familia  
*Campylobacteraceae* a partir de datos metagenómicos del tracto digestivo de la mosca

soldado *Hermetia illucens*

Trabajo previo a la obtención del título de

Magíster en Biología Computacional

Juan David Báez Sotomayor

Quito, 2024

## Resumen

La presente investigación aborda la reconstrucción y anotación del genoma de una bacteria de la familia *Campylobacteraceae*, obtenida a partir de información metagenómica del tracto digestivo de la mosca soldado negra (*Hermetia illucens*). Para ello, se empleó secuenciación shotgun con tecnología Illumina. El protocolo incluyó un exhaustivo control de calidad (mediante FastQC y Trimmomatic), seguido de la eliminación de lecturas del hospedador mediante mapeo con Bowtie2 (utilizando como referencia el genoma de *H. illucens*) y la generación final de archivos FASTQ libres de secuencias no deseadas.

Posteriormente, en la plataforma KBase se llevaron a cabo múltiples ensamblajes (empleando metaSPAdes, MEGAHIT e IDBA-UD) y se determinó la calidad de los contigs generados. Para agruparlos en bins de alta calidad, se aplicaron diversas herramientas de binning y optimización (MaxBin2, MetaBAT2, CONCOCT y DAS Tool). Este proceso permitió identificar el bin 003, asignado taxonómicamente a *Campylobacter* mediante GTDB-Tk. El análisis funcional reveló un genoma cercano a 1.7 Mbp, con un nivel elevado de integridad y una contaminación mínima. En dicho genoma se identificaron rutas metabólicas fundamentales para la producción de energía (incluidos los complejos de la cadena de transporte de electrones) y elementos asociados con la biosíntesis de siroheme. Aunque se evidenciaron posibles mecanismos de resistencia a antibióticos, investigaciones complementarias no confirmaron la presencia de genes de resistencia relevantes, lo cual podría deberse a limitaciones en la calidad del ADN o a errores en el proceso de filtrado de datos.

## Tabla de contenido

Introducción .....	10
Justificación .....	10
Objetivos.....	11
Objetivo General .....	11
Objetivos Específicos.....	11
1. Revisión de Literatura .....	12
2. Genómica.....	12
3. Tecnologías de secuenciación.....	12
4. Ensamblaje y anotación de genomas .....	15
5. Metagenómica .....	16
6. Mosca Soldado Negro .....	17
7. Familia <i>Campylobacteraceae</i> .....	18
7.1. Enfermedades .....	18
7.2. Fuentes de transmisión .....	19
8. Metodología.....	25
8.1. Obtención de material genético y secuenciación.....	25
8.2. Ensamblaje y anotación del metagenoma.....	25
8.2.1.1. Evaluación de la calidad de las secuencias.....	26
8.2.1.2. Limpieza de secuencias .....	27

8.2.1.3.	Eliminación de secuencia del hospedero .....	27
8.3.2.	Clasificación taxonómica .....	29
8.3.3.	Ensamblaje de Metagenoma, comparación y agrupamiento de contigs 29	
8.3.4.	Optimización de Bins .....	30
8.3.5.	Evaluación de los bins .....	30
8.3.6.	Anotación del genoma y clasificación taxonómica .....	30
8.3.7.	Obtención de perfil metabólico y genes de resistencia .....	31
9.	Resultados y discusión .....	31
9.1.	Control de calidad.....	31
9.2.	Limpieza de secuencias .....	34
9.3.	Eliminación de secuencia hospedero .....	36
9.4.	Ensamblaje y anotación del metagenoma en KBase .....	39
9.4.1.	Clasificación taxonómica .....	39
9.4.2.	Ensamblaje del metagenoma, comparación de contigs y optimización de bins	45
9.4.3.	Evaluación de la calidad e integridad de los bins obtenidos .....	48
9.4.4.	Anotación del genoma .....	50
9.4.5.	Clasificación taxonómica .....	52
9.5.	Genes de resistencia y rutas metabólicas.....	56
10.	Conclusiones y recomendaciones .....	66
	Bibliografía .....	68

Anexos .....	88
--------------	----

## Indicie de figuras

Figura 1. Cambio de precios a lo largo del tiempo .....	13
Figura 2. Flujo de trabajo para la tecnología de NGS. Tomado de (Rodríguez-Santiago & Armengol, 2012).....	14
Figura 3. Fuentes de contaminación de Campylobacter en humanos.....	20
Figura 4. Flujo de trabajo en supercomputadora CEDIA .....	26
Figura 5. Referencia de <i>H. illucens</i> en NCBI, tomado de (NCBI, 2024) .....	27
Figura 6. Flujograma para metagenómica en KBase. Tomado de (Chivian et al., 2023b).....	28
Figura 7. Evaluación de la calidad de las secuencias forward y reverse antes del proceso de limpieza .....	32
Figura 8. Concentración de GC de las secuencias forward y reverse.....	34
Figura 9. Evaluación de la calidad de las secuencias forward y reverse después del proceso de limpieza .....	36
Figura 10. Genoma de referencia de <i>H. ilucens</i> descargado del NCBI .....	37
Figura 11. Base creada a partir de la secuencia Hospedero y Bowtie2 v2.4.2 ...	37
Figura 12. Resultado de las lecturas mapeadas y no pareadas del metagenoma con relación a <i>H. illucens</i> .....	38
Figura 13. Secuencia forward .....	39
Figura 14. Secuencia reverse .....	39
Figura 15. Carga de las secuencias a KBase.....	39
Figura 16. Distribución de la clasificación taxonómica a nivel de filo en la muestra analizada .....	40
Figura 17. Distribución de la clasificación taxonómica a nivel de clase en la muestra analizada .....	41

Figura 18. Distribución de la clasificación taxonómica a nivel de orden en la muestra analizada .....	41
Figura 19. Distribución de la clasificación taxonómica a nivel de familia en la muestra analizada .....	42
Figura 20. Distribución de la clasificación taxonómica a nivel de género en la muestra analizada .....	42
Figura 21. Distribución de la clasificación taxonómica a nivel de especie en la muestra analizada .....	43
Figura 22. Presencia de microorganismos encontrados en <i>H. illucens</i> , tomado de (Khamis et al., 2020) .....	44
Figura 23. Diagrama de KRONA a partir de KAIJU.....	45
Figura 24. Clasificación de contigs del más largo a más corto.....	45
Figura 25. Suma de contigs con relación a su longitud en Mbp.....	45
Figura 26. Comparación de los contigs generados mediante las diferentes herramientas de ensamblaje.....	46
Figura 27. Comparación del número de bins y la calidad general de cada ensamblado .....	47
Figura 28. Comparación de la eficacia de diferentes métodos de agrupamiento de bins .....	48
Figura 29. Contaminación e integridad de los bins obtenidos.....	49
Figura 30. Taxonomía del bin.003 .....	52
Figura 31. Árbol taxonómico obtenido a partir de los 11 bins obtenidos.....	54
Figura 32. Árbol de especies obtenidos a partir del análisis de los bins .....	55
Figura 33. Mapa de calor de los complementos cadena de transporte de electrones (ETC).....	56

Figura 34. Funciones metabólicas .....	60
Figura 35. Mapa generado a partir del resultado obtenido en Annotate and Distill Assemblies with DRAM.....	65

## Índice da Tablas

Tabla 1. Estadísticas de calidad de las secuencias forward y reverse antes del proceso de limpieza .....	31
Tabla 2. Datos post limpieza de las secuencias forward y reverse .....	35
Tabla 3. Número de bins obtenidos con cada herramienta .....	47
Tabla 4. Caracterización taxonómica y fuente de datos del Bin.003 .....	50
Tabla 5. Evaluación de linaje obtenido a partir de los bins .....	51
Tabla 6. Clasificación taxonómica obtenida .....	53
Tabla 7. Identificación de los genes, nombre, unidad y numero de genes dentro de <i>Campylobacter</i> spp .....	63

## **Introducción**

### **Justificación**

En Ecuador, se exploran vías para lograr una agricultura más sostenible, y la mosca soldado negro ha demostrado un elevado potencial en la biodegradación de residuos orgánicos (Arquivo, 2016; Bortolini et al., 2020; Tanga et al., 2021). No obstante, su utilidad no se limita solo a este ámbito, pues diversos estudios señalan que también constituye una importante fuente de proteínas para la alimentación del ganado (Reátegui et al., 2020; Tanga et al., 2021).

El crecimiento acelerado de la población humana en los últimos años ha impulsado a investigadores, productores, científicos e industrias a buscar estrategias innovadoras para la generación de alimentos y la provisión de fuentes proteicas destinadas a animales de consumo masivo. En este escenario, la mosca soldado negro podría desempeñar un papel fundamental; sin embargo, se debe contemplar la presencia de bacterias asociadas a esta especie, como *Campylobacter*, capaz de provocar zoonosis en los seres humanos (Tanga et al., 2021).

En Ecuador, se ha descrito una variedad de aspectos acerca de *Campylobacter*, destacándose su resistencia antimicrobiana, además de su epidemiología y diversidad genética. La identificación de múltiples secuencias tipo (ST) y complejos clonales pone en evidencia tasas elevadas de resistencia: la ciprofloxacina alcanza el 70,6%, seguida de la tetraciclina (67,6%) y la eritromicina (35,3%). Asimismo, la coincidencia de determinados ST en aislados de granjas, carcasas minoristas y muestras humanas sugiere una posible propagación del patógeno a través de la cadena alimentaria. Sin embargo, la información disponible sobre infecciones por *Campylobacter* en Ecuador es limitada. En la actualidad, la “AGENCIA DE REGULACIÓN Y CONTROL FITO Y ZOOSANITARIO” (centrada en la salud animal) y el “INSTITUTO NACIONAL DE

INVESTIGACIÓN EN SALUD PÚBLICA – INSPI – DR. LEOPOLDO IZQUIETA PÉREZ” (enfocado en el área humana) llevan a cabo el control de estas infecciones. Ambas instituciones realizan la detección de *Campylobacter* mediante la siembra para evaluar su resistencia a antibióticos y emplean métodos de identificación, entre ellos la reacción en cadena de la polimerasa (PCR) y la PCR en tiempo real (qPCR) (Givanoudi et al., 2021; Lv et al., 2020a; Nesla Olivas, 2007; Simaluiza et al., 2018).

Por las razones antes expuesta anteriormente es importante conocer y controlar la propagación de esta bacteria, que causa diarrea crónica en las personas. De igual manera es importante conocer los diversos antibióticos a los cuales son presenta resistencia para poder tratar la infección con esta bacteria (el Sitio Avicola, 2012; ELIKA, 2022; Fernández, 2011; Nesla Olivas, 2007). Esto permitiría conocer si en el tracto digestivo de *Hermetia illucens*, se encuentra la presencia de *Campylobacter*, la cual puede transmitirse a los alimentos, frutas, animales y carnes de consumo humano; lo cual puede causar diarrea crónica al infectarse una persona con la bacteria.

## **Objetivos**

### **Objetivo General**

Ensamblar y anotar el genoma de una proteobacteria perteneciente a la familia *Campylobacteraceae* a partir de datos metagenómicos del tracto digestivo de larvas de mosca *Hermetia Illucens*.

### **Objetivos Específicos**

- Eliminar lecturas del hospedero (*Hermetia illucens*) de los datos metagenómicos
- Ensamblar el genoma de la bacteria de la familia *Campylobacteraceae* a partir de datos metagenómicos libres de lecturas del hospedero
- Seleccionar el genoma ensamblado de la familia

## **1. Revisión de Literatura**

### **2. Genómica**

La genómica puede describirse como la disciplina que estudia el conjunto del material genético presente en los diversos organismos (Mainetti, 2003; National Human Genome Research Institute, 2024a). Este material genético alberga una gran cantidad de información, responsable de transmitir las características necesarias para el funcionamiento y desarrollo de cada organismo. Dentro del estudio del genoma se incluye la identificación, comprensión y análisis de cada gen y de todos los componentes que lo constituyen.

### **3. Tecnologías de secuenciación**

La secuenciación del material genético surgió en la década de los 70, iniciando con el desarrollo de la técnica de Sanger en 1977, lo que representó una revolución en la secuenciación de fragmentos de ADN. Sin embargo, este método presentaba diversas limitaciones en el laboratorio al intentar secuenciar múltiples fragmentos simultáneamente (Metzker, 2010).

La adopción de la secuenciación de nueva generación (Next-Generation Sequencing, NGS) produjo un cambio radical en el campo de la genómica, al posibilitar la secuenciación masiva y paralela de millones de fragmentos de material genético de forma simultánea (Metzker, 2010). Este avance ha elevado de manera significativa la rapidez, eficiencia y exactitud de los procesos de secuenciación, favoreciendo el análisis genómico en diversas áreas, entre ellas la investigación biomédica, la metagenómica y la medicina personalizada (Hernández et al., 2020a; Rubio et al., 2020)

La tecnología de secuenciación NGS tiene múltiples aplicaciones en distintos campos. Entre ellas, se encuentran el estudio de variaciones genéticas, el análisis de expresión génica y la caracterización de proteínas, entre otros enfoques relevantes

(Mardis, 2013). Una de sus aplicaciones más destacadas es la identificación del microbiota para determinar los taxones que coexisten en una muestra, así como la detección de genes de resistencia y rutas metabólicas, lo que permite un análisis detallado de la composición genética en distintos entornos (Hernández et al., 2020a).

El desarrollo de las tecnologías NGS un gran cambio en el ámbito de la investigación dando como resultado que el costo de las tecnologías de NGS fuera disminuyendo, permitiendo que sea más accesible para la investigación, el ámbito clínico, ambiental, etc., generando nuevas aplicaciones para el uso de la información (Rodríguez-Santiago & Armengol, 2012).

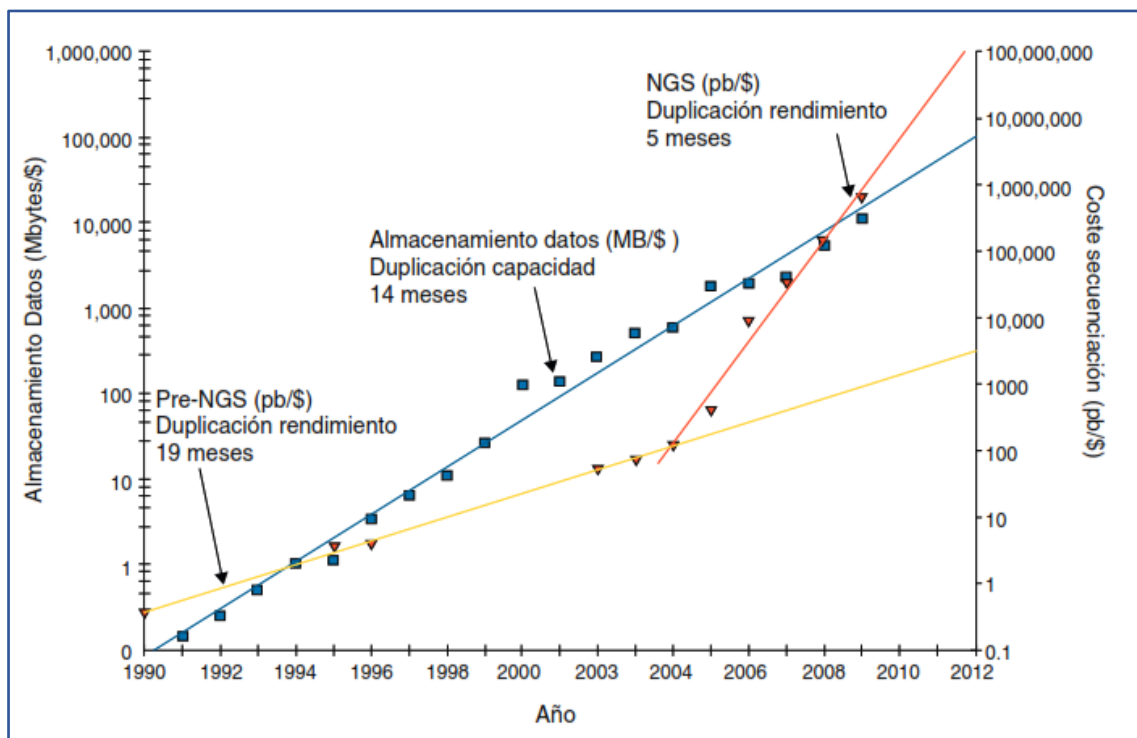


Figura 1. Cambio de precios a lo largo del tiempo

Los triángulos representan el costo de la secuenciación tomando en cuenta la capacidad de secuenciación, la línea de color azul, el incremento del almacenamiento con relación al precio de este. La línea roja muestra cuanto ha variado el costo de secuenciación por cada base con relación al costo, en comparación al inicio de la

secuenciación. Con relación al escalado de la tecnología dando un mayor rendimiento y mejorando la velocidad de secuenciación, esto se observa a partir de la brecha que se observa entre la línea y amarilla (Rodríguez-Santiago & Armengol, 2012).

A continuación, en la figura 2, se presenta un cuadro de flujo para el análisis de datos de secuenciación por NGS.

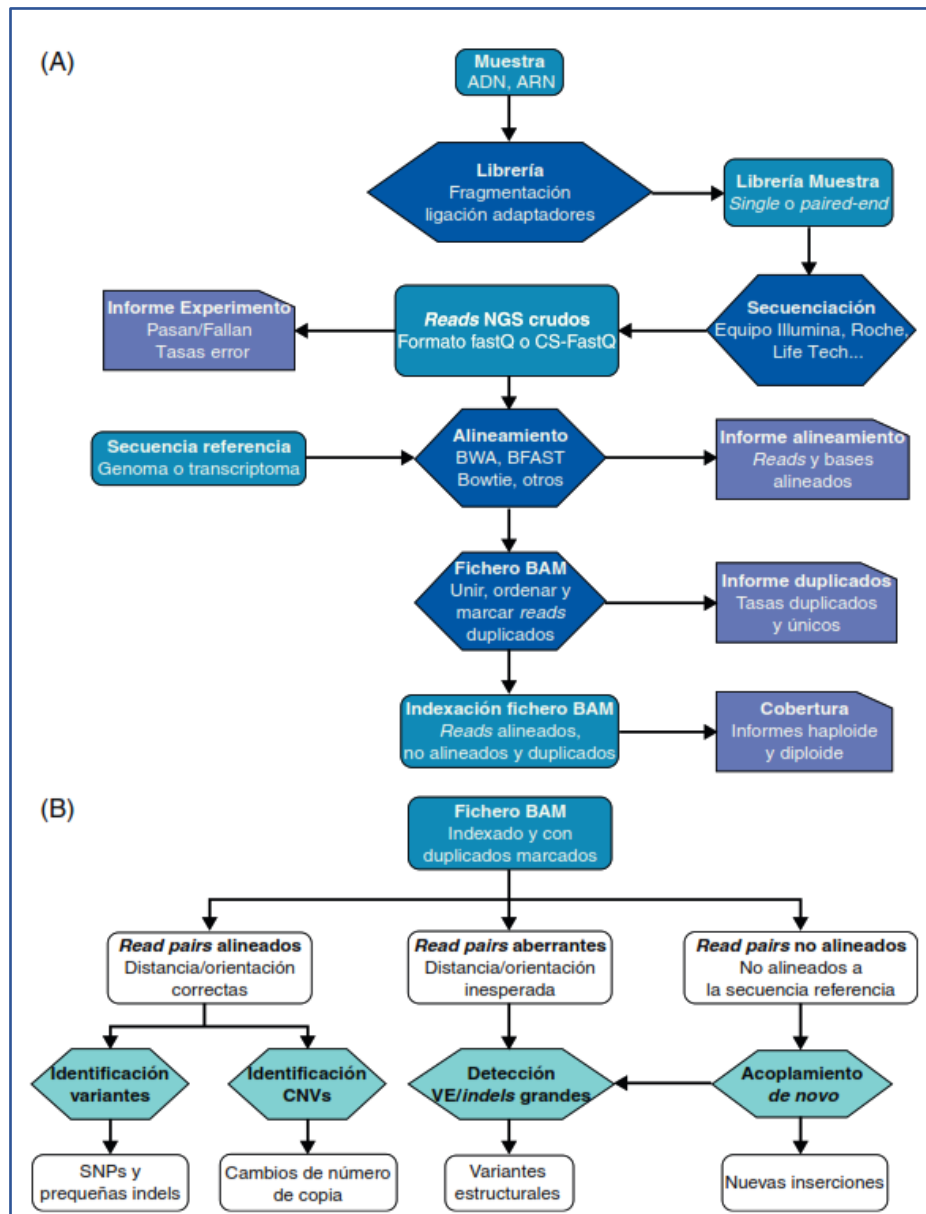


Figura 2. Flujo de trabajo para la tecnología de NGS. Tomado de (Rodríguez-Santiago & Armengol, 2012)

#### 4. Ensamblaje y anotación de genomas

Desde la aparición de la NGS se ha presentado una gran cantidad de datos que contienen información importante y de utilidad para diversas ramas de trabajo (Aguilar-Bultet Lisandra & FalquetLaurent, 2015; Rodríguez-Santiago & Armengol, 2012). Una vez que el material genético se procesa en el secuenciador, se obtiene una gran cantidad de datos, donde se encuentra información muy valiosa, como son los genes que pueden estar asociados a diversas funciones, resistencias entre otros (Aguilar-Bultet Lisandra & FalquetLaurent, 2015). Al trabajar datos metagenómicos, no solo se trabaja con un solo organismo, sino que se tiene una gran diversidad de organismos.

Una vez que se obtienen los datos de secuenciación, en este caso de metagenómica, se procederá a realizar un ensamblaje de estos, lo que consiste en la unión de las lecturas para formar secuencias más largas denominadas *contigs* (Lis Ricardi, 2022). Se pueden realizar distintos tipos de ensamblaje, entre los principales se encuentran el ensamblaje basado en un genoma de referencia y el ensamblaje de *novo*.

El ensamblaje basado en un genoma de referencia permite la alineación de las lecturas con una secuencia genómica previamente caracterizada, lo que facilita la identificación de variantes y regiones conservadas. Entre sus ventajas, destaca su menor demanda de recursos computacionales y su capacidad para generar ensamblajes más precisos en organismos bien estudiados. Sin embargo, su principal desventaja es la dependencia de un genoma de referencia adecuado, lo que puede limitar su aplicación en estudios de comunidades microbianas complejas o en organismos sin información genómica disponible (Blake et al., 2010; Peabody et al., 2015).

Por otro lado, el ensamblaje de *novo* no requiere una secuencia de referencia, lo que permite reconstruir genomas completos de especies previamente no caracterizadas. Entre sus ventajas, se encuentra la capacidad de descubrir nuevas variantes genéticas y

especies desconocidas, lo que resulta esencial en estudios metagenómicos. No obstante, este tipo de ensamblaje presenta desafíos, como la mayor dificultad en la resolución de secuencias repetitivas y la mayor posibilidad de ensamblajes erróneos. Además, demanda un poder computacional significativamente mayor debido a la necesidad de algoritmos avanzados para ensamblar fragmentos cortos sin una referencia previa (Mathai et al., 2020; Vollmers et al., 2017).

Después de contar con el genoma ensamblado, se debe realizar la anotación del genoma obtenido (Aguilar-Bultet Lisandra & FalquetLaurent, 2015; Bikel et al., 2015; Lis Ricardi, 2022). La anotación es importante para conocer la información que contienen los *contigs*.

Existen diversos tipos de anotación, lo cual proporcionan diferentes tipos de información. Un tipo de anotación estructural, la permite realizar una predicción de las regiones codificantes de los *contigs* obtenidos. Otro tipo de anotación es la funcional la cual nos proporciona información importante y relativa de los genes en los *contigs* (Lis Ricardi, 2022).

## **5. Metagenómica**

La metagenómica es una disciplina dentro de las ciencias ómicas que estudia la estructura y función del material genético microbiano presente en diversos entornos, como suelos, océanos, microbiotas humanas y comunidades microbianas no cultivables. Su principal objetivo es identificar los organismos que conforman estos ecosistemas y analizar sus funciones dentro de ellos (National Human Genome Research Institute, 2024b; Rahbari et al., 2017).

La técnica más empleada para este propósito es la secuenciación *shotgun*, que fragmenta y secuencia de forma aleatoria el material genético, brindando una visión

global de la diversidad y el potencial funcional de la comunidad microbiana(Kunin et al., 2008; Rahbari et al., 2017).

## 6. Mosca Soldado Negro

La mosca soldado-negro, tconocida en inglés como Black Soldier Fly – BSF, pertenece al orden de los dípteros en los insectos. Su nombre científico es *Hermetia illucens*. Actualmente ha tomado relevancia el estudio de esta especie, debido a que las larvas poseen la capacidad de degradación de sustratos orgánicos, llegando a transformar alrededor del 50% de la materia orgánica consumida en biomasa (G Del Hierro et al., 2021; INABIO, 2016). La biomasa que se produce de la actividad de las larvas de *H. illucens* se emplea en la alimentación de animales de granja.

*H. illucens* se encuentra distribuida en varios países y regiones como son:

- América Latina
  - Ecuador
- Asia
  - Indonesia
  - Japón
  - Filipinas
- Europa
- África
- Oceanía
  - Nueva Zelanda
  - Australia

Se encuentran dividida en estas regiones debido a que crece en zonas tropicales, subtropicales y templadas (INABIO, 2016).

El ciclo de vida de *H. illucens* incluye un total de seis etapas con un tiempo de duración de cuatro semanas, dependiendo el ambiente donde se desarrolle el individuo, pero puede variar según el clima de desarrollo (G Del Hierro et al., 2021).

## **7. Familia *Campylobacteraceae***

Para poder describir y comprender a la bacteria *Campylobacter*, debemos primero conocer su taxonomía. Pertenece al dominio *Bacteria*, al filo *Pseudomonadota*, a la clase *Epsilonproteobacteria*, a la orden *Campylobacterales*, a la familia *Campylobacteraceae* y al género *Campylobacter* (Alonso et al., 2023; El-Naenaey et al., 2020; Terzolo et al., 2007). El nombre de *Campylobacter* significa “bacilo curvo” (Terzolo et al., 2007).

Dentro de las características que contiene el género *Campylobacter*, son su forma ondulada en espiral con una forma de “S”, son gram negativas con unas medidas aproximadas de 0.5 a 8  $\mu\text{m}$  a lo largo y a lo ancho con unas medidas de 0.2 a 0.5  $\mu\text{m}$  y poseen un solo flagelo (Terzolo et al., 2007). Poseen una reacción positiva a la oxidasa, pero algo a destacar es que la reacción a la catalasa es variable según la especie de *Campylobacter*. Otra característica destacada de este género es la temperatura de incubación que va entre los 25 a 42 °C.

### **7.1. Enfermedades**

*Campylobacter* es una de las causantes de enfermedades intestinales más comunes en las personas, es la causante de la gastroenteritis (Bush & Charles, 2022; ELIKA, 2023; OMS, 2020), de los cuales se conocen que existen más de una docena especies dentro de este género. Esta familia se encuentra presente en diversos lugares, desde el ambiente

hasta el intestino de animales sanos (Bush & Charles, 2022; ELIKA, 2023; OMS, 2020), donde parte de su flora intestinal.

Las personas más propensas a esta enfermedad son niños menores de 2 años, adultos mayores y quienes presentan un sistema inmune comprometido (ELIKA, 2023; OMS, 2020). De acuerdo con la OMS (2020), en ese mismo año la diarrea causada por *Campylobacter* afectó a una de cada diez personas. Asimismo, se estima que, a escala global, 220 millones de niños menores de 5 años se infectaron con este patógeno, según datos de la Organización Mundial de la Salud..

Los primeros síntomas de infección por *Campylobacter* suelen aparecer entre los 2 a 5 primeros días, cuando la bacteria ingresa al organismo (Bush & Charles, 2022; ELIKA, 2023; OMS, 2020). Una vez que la bacteria paso su tiempo de infección presenta los siguientes síntomas, con una duración de 3 a 6 días y que suelen ser los más comunes como son:

- Diarrea, en algunas ocasiones se pueden presentar sangre en las heces
- Calambres
- Dolor abdominal con cólicos
- Fiebre
- Fatiga
- Nauseas
- Vomito

## 7.2. Fuentes de transmisión

Esta bacteria se encuentra presente en alimentos, animales y el ambiente, por lo cual existe la posibilidad de contraer esta bacteria (ELIKA, 2023; OMS, 2020). Con relación a los animales se ha encontrado presencia de esta bacteria en ganado vacuno,

aves de corral, ganado porcino, ganado ovino y mariscos (ELIKA, 2023; OMS, 2020). Pero también en estudios realizados se ha determinado que es posible encontrar esta bacteria en agua dulce y no potable, de igual manera en agua mal potabilizada y hielos comerciales realizados con defectuosos controles de calidad o de manera artesanal para su venta (ELIKA, 2023). No obstante, mediante el uso de buenas prácticas para la elaboración, producción y conservación de alimentos se puede evitar la infección con esta bacteria; con relación al agua, siempre y cuando esta sea potable y cuente con estándares de calidad, la presencia de esta bacteria será nula.

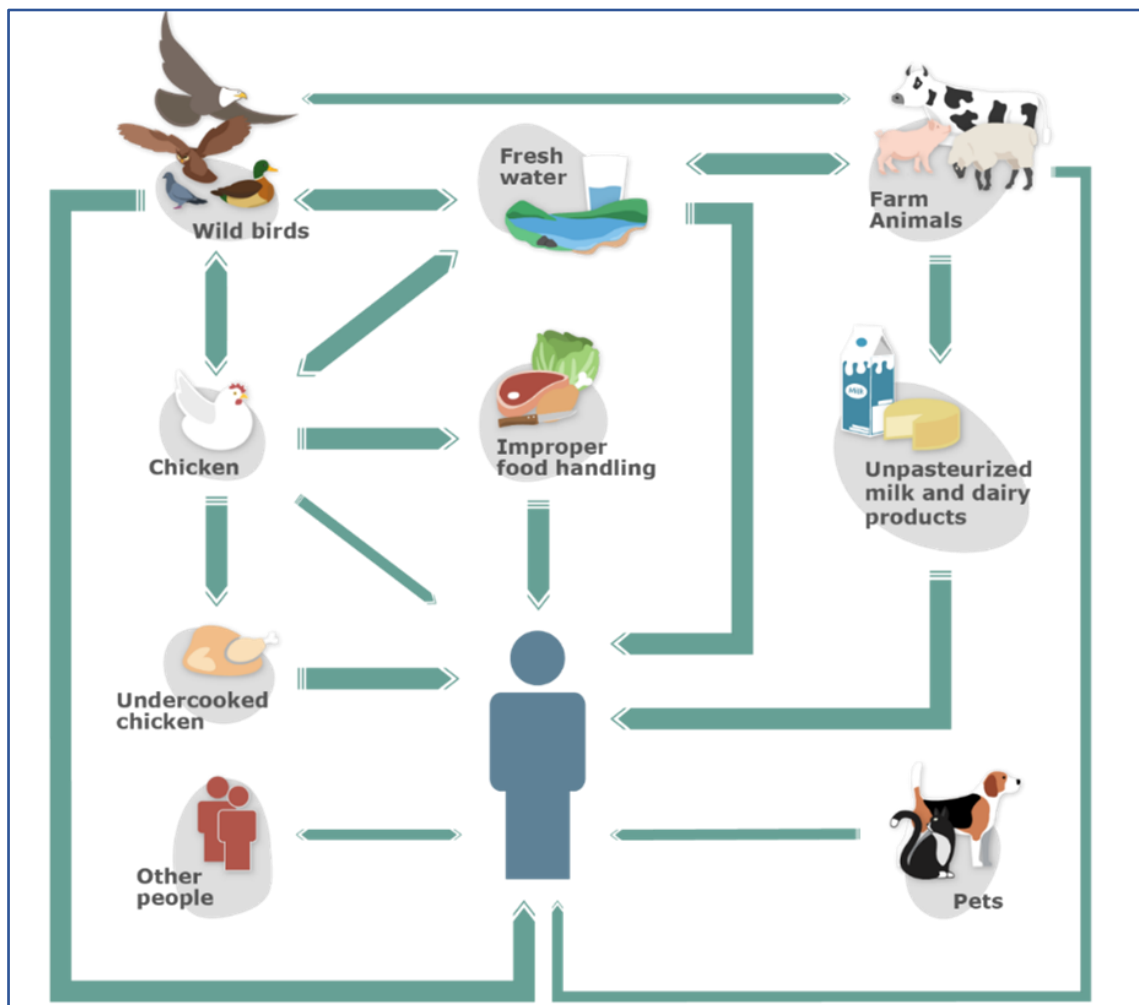


Figura 3. Fuentes de contaminación de *Campylobacter* en humanos

Nota: Mientras más gruesa sea la flecha, se presenta una mayor fuente de contagio para las personas

## Genoma, plásmidos y genes de resistencia en *Campylobacter*

La secuenciación completa del genoma de *Campylobacter* se efectuó por primera vez en el año 2000 (Cuadrado et al., 2018), a través de la colaboración entre The London School of Hygiene & Tropical Medicine (LSHTM) y el Instituto Sanger. El resultado fue un genoma circular de 1,461,481 pb correspondiente a una cepa de *C. jejuni* (Cuadrado et al., 2018). En dicho genoma se determinó un contenido de guanina-citosina de entre 30 y 40% (Cuadrado et al., 2018; De Ciências et al., 2019), así como la presencia de alrededor de 1,654 proteínas.

Un elemento esencial dentro del genoma de *Campylobacter* para la supervivencia del microorganismo lo conforman los plásmidos y los genes de resistencia. Concretamente, se han identificado tres tipos de plásmidos: pVIR, pTET y los denominados plásmidos crípticos (Cuadrado et al., 2018). El plásmido pVIR participa en la exportación de ADN, la secreción de proteínas y la conjugación bacteriana; mientras que pTET cumple una función primordial en la resistencia antimicrobiana. Por otro lado, los plásmidos crípticos contienen diversos genes relacionados con los mecanismos de replicación (Cuadrado et al., 2018).

Con relación a los genes de resistencia que se encuentran en el genoma de *Campylobacter*, se han estudiado 18 genes. De igual manera se han detectado mutaciones dentro de dos genes housekeeping que son *gyrA* and 23S rRNA (Hull et al., 2023; Joshi et al., 2022; Zhao et al., 2016). Se debe destacar la importancia de los genes housekeeping, ya que estos poseen una expresión muy estable en todo el organismo o sus células y las cuales son parte del mantenimiento celular del organismo (Joshi et al., 2022).

Listado de los 18 genes de resistencia que posee *Campylobacter* (Zhao et al., 2016).

- tet(O): El gen **tet(O)** confiere resistencia a la tetraciclina en *Campylobacter*, al codificar una proteína que funciona de forma similar a un factor de elongación. Esta proteína interfiere en la acción de la tetraciclina al inhibir la síntesis proteica por medio de su interacción directa con el ribosoma, lo que libera al antibiótico de su objetivo diana dentro de la célula. Dentro de este grupo de protección ribosomal, se incluyen otras proteínas como Tet(M), Tet(Q), Tet(S), Tet(T), Tet(W) y OtrA (Connell et al., 2003).
- bla OXA-61: Este gen otorga resistencia a los antibióticos  $\beta$ -lactámicos, como la penicilina, al producir una enzima capaz de romper el anillo  $\beta$ -lactámico característico de estos fármacos. Al descomponer dicho anillo, se evita que el antibiótico ejerza su efecto en las bacterias, protegiéndolas y permitiéndoles sobrevivir dentro del organismo (Alfredson & Korolik, 2005; Audu et al., 2022; Casagrande Proietti et al., 2020; Quino et al., 2022a).
- catA: Este gen permite la codificación de la enzima llamada cloranfenicol acetiltransferasa, abreviada como CAT, la cual permite la resistencia al cloranfenicol; el cual es un antibiótico. Este gen se ha encontrado en plásmidos y elementos genéticos móviles en bacterias (Bunduruş et al., 2023; Liao et al., 2022)
- lnu(C): Este gen otorga la resistencia a los antibióticos lincosamidas, esto se logra mediante la codificación de la enzima lincosamida nucleotidiltransferasa (Li et al., 2023; Quino et al., 2022b).
- aph(2'')-Ib: Este gen codifica para la enzima aminoglicósido fosfotransferasa, la cual confiere la resistencia a los antibióticos aminoglucósidos, dentro de este grupo y uno de los más conocidos se encuentra la gentamicina (Fabre et al., 2018a; Habib et al., 2023; Zhao et al., 2014).

- aph(2'')-Ic: Este gen codifica una enzima del tipo aminoglucósido fosfotransferasa conocida como APH, la cual permite la inhibición de los antibióticos del tipo aminoglucósidos; un antibiótico que representa a este tipo de antibióticos es la gentamicina (Fabre et al., 2018b; Ocejo et al., 2021; Yao et al., 2017a, 2017b; X. Zhang et al., 2021a).
- aph(2')-If: Este gen tiene la capacidad de codificar la enzima fosfotransferasa, la cual otorga la resistencia a los antibióticos de tipo aminoglucósidos, entre los más conocidos se encuentran gentamicina y la kanamicina (Yao et al., 2017c).
- aph(2'')-Ig: Este gen codifica para la enzima fosfotransferasa, que otorga resistencia para los antibióticos aminoglucósidos, en específico contra la gentamicina. Algo que preocupa en el área clínica, es la alta transferencia de este gen, ya que se encuentra en elementos móviles, plásmidos y transposones (X. Zhang et al., 2021b; Zhao et al., 2015).
- aph(2'')-Ih: Este gen de igual manera codifica la enzima fosfotransferasas y confiera la resistencia para gentamicina y kanamicina, es similar a los genes el aph(2'')-Ig y aph(2'')-Ih en *Campylobacter* (Kang et al., 2024).
- aac(6')-Ie-aph(2'')-Ia: Este gen codifica para la enzima bifuncional, la cual le confiere resistencia a los antibióticos del tipo aminoglucósidos (Chen et al., 2013).
- aac(6')-Ie-aph(2'')-If: Este gen codifica para la enzima bifuncional, la cual también confiere resistencia a los antibióticos del tipo aminoglucósidos. Esto se logra mediante la modificación estructural de los antibióticos tipo aminoglucósidos. Algunos antibióticos dentro del grupo bifuncional son gentamicina y kanamicina (X. Zhang et al., 2021b).

- aac(6′)-Im: Este gen de igual manera confiere la codificación para la enzima acetiltransferasa, que permite la resistencia a los antibióticos del tipo aminoglucósidos (C. A. Smith et al., 2017).
- aadE: Este gen codifica para la enzima adenililtransferasa, la cual otorga resistencia a los antibióticos del tipo aminoglucósidos, como en este caso sería la estreptomicina y espectinomicina (Guirado et al., 2022; Pinto-Alphandary et al., 1990).
- sat4: Este gen codifica para la enzima streptotricina acetiltransferasa, la cual da resistencia a los antibióticos de clase estreptotricinas (Jacob et al., 1994).
- ant(6′): Este gen codifica la enzima aminoglucósido nucleotidiltransferasa, la cual otorga la resistencia a los antibióticos tipo aminoglucósidos. Uno de los principales antibióticos del grupo aminoglucósidos es la estreptomicina, a la cual le otorga resistencia (Hormeño et al., 2018a, 2018b).
- aad9: Este gen codifica la enzima adenililtransferasa, la cual otorga la resistencia a los antibióticos del tipo aminoglucósidos, en especial confiere resistencia a la espectinomicina (Chen et al., 2013).
- aph(3′)-Ic: Este gen codifica para la enzima fosfotransferasa, lo cual otorga resistencia a los antibióticos del tipo aminoglucósidos, en especial como al antibiótico kanamicina (Guirado et al., 2022).
- aph(3′)-IIIa: Este gen codifica para la enzima fosfotransferasa, para dar resistencia a los antibióticos del tipo aminoglucósidos. A los antibióticos que principalmente confiere resistencia son kanamicina y neomicina (Woegerbauer et al., 2014).

Todos los genes involucrados en la resistencia antimicrobiana permiten que la bacteria obtenga una resistencia del 100% a Tetraciclina, ciprofloxacino/ácido nalidixico y eritromicina. Una resistencia entre 95 a 98% para gentamicina, azitromicina,

clindamicina y telitromicina. Sin embargo, no se ha detectado genes y resistencia contra florfenicol (Zhao et al., 2016)

## **8. Metodología**

### **8.1. Obtención de material genético y secuenciación**

La información utilizada en este análisis metagenómico proviene del estudio “Genetic diversity and organic waste degrading capacity of *Hermetia illucens* from the evergreen forest of the Equatorial Choco lowland” (2023, Ecuador) En dicha investigación, se recolectaron muestras de *Hermetia illucens* en dos regiones del país. Las secuencias analizadas proceden de archivos FASTQ generados mediante secuenciación *shotgun* del ADN proveniente del tracto intestinal de moscas soldado negro (*Hermetia illucens*), empleando tecnología Illumina. El procesamiento de los datos se llevó a cabo en formato *paired-end*, con un archivo *forward* (“PLA\_FDME210308107-1r\_H5Y3FDSX2\_L4\_1.fq”) y otro *reverse* (“PLA\_FDME210308107-1r\_H5Y3FDSX2\_L4\_2.fq”), cada uno con un tamaño cercano a los 7 Gb.

### **8.2. Ensamblaje y anotación del metagenoma**

#### **8.2.1. Obtención e instalación de software**

Con el fin de poder desarrollar el proyecto para el análisis metagenómico de las secuencias, es necesario utilizar las secuencias *paired end* previamente obtenidas y varios programas bioinformáticos. El primer paso para el análisis fue realizar la carga de las secuencias en el espacio de trabajo; en este caso se trabajó la primera parte en la supercomputadora de CEDIA. El programa utilizado para sincronización de archivos entre las dos partes fue *WinSCP 6.3.2*. Las secuencias se subieron al directorio “Tesis”.

Una vez que se subieron las secuencias para trabajar, se creó un directorio, en el cual para almacenar e instalar todos los programas necesarios para trabajar; el nombre del

directorio fue “PROGRAMS”. Cada uno de los programas fue descargado, descomprimido de ser el caso e instalado en este directorio.

- *FastQC* v0.11.9
- *Trimmomatic* v0.39
- *SAMtools* v1.10-3
- *Bowtie 2* v2.4.2

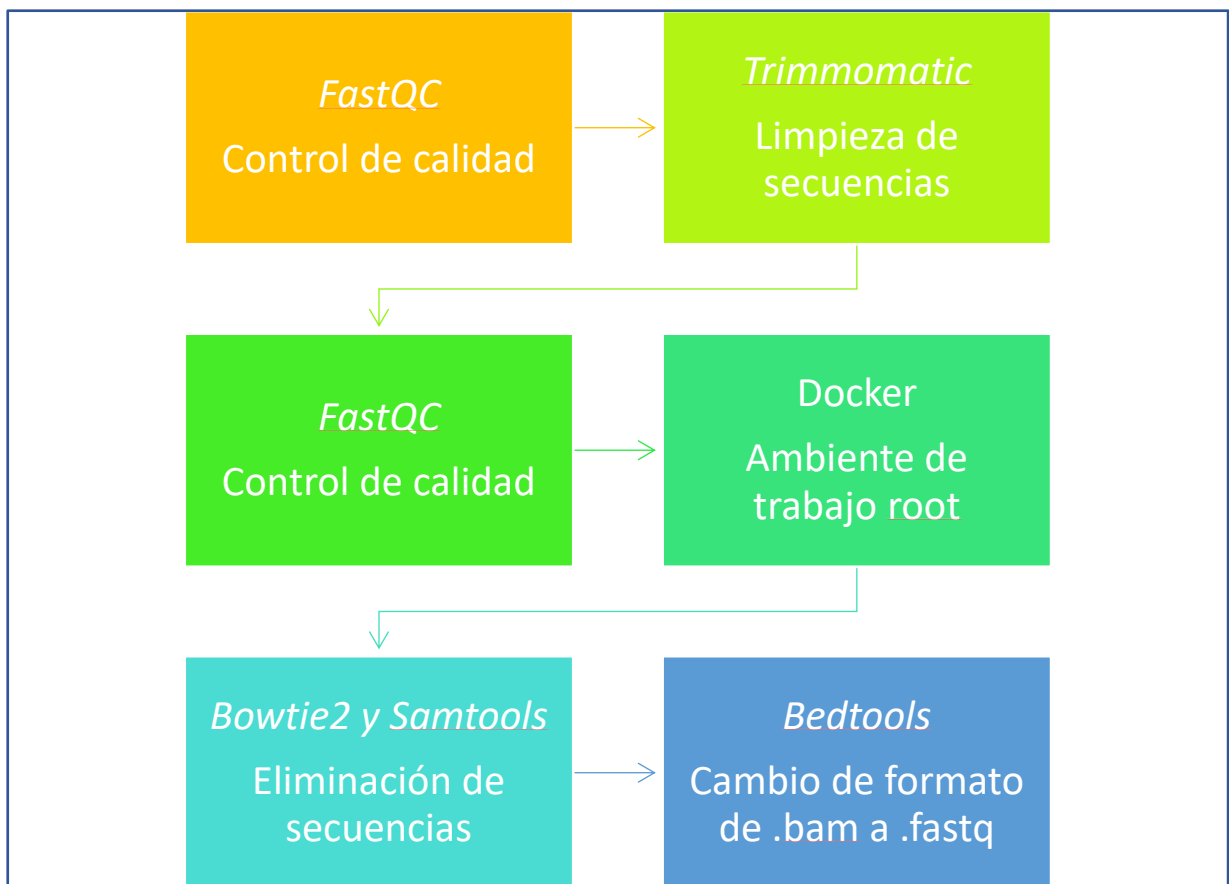


Figura 4. Flujo de trabajo en supercomputadora CEDIA

#### 8.2.1.1. Evaluación de la calidad de las secuencias

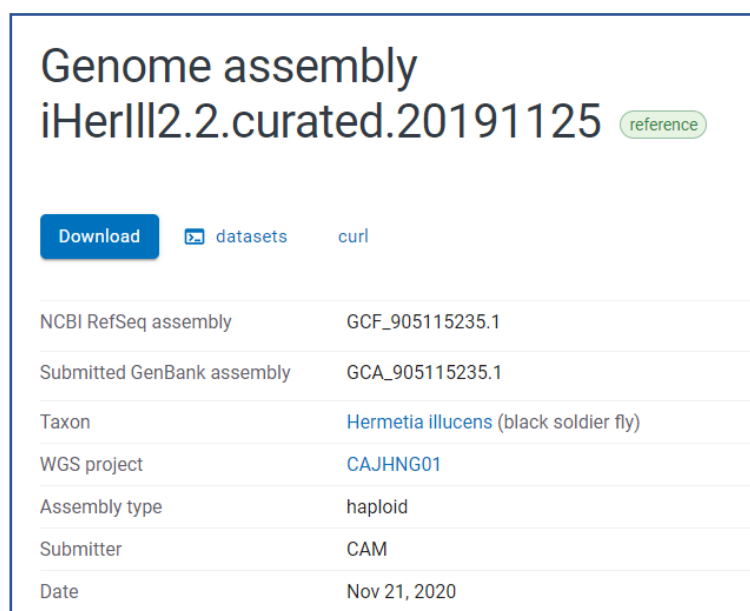
La evaluación de la calidad de las secuencias se llevó a cabo utilizando la herramienta *FastQC* v.0.11.9, la cual permite inspeccionar el nivel de limpieza de las secuencias antes de su procesamiento (Brown et al., 2017; Chivian et al., 2023a).

### 8.2.1.2. Limpieza de secuencias

Tras la evaluación de la calidad de las secuencias, se empleó la herramienta *Trimmomatic v0.39* para su depuración y procesamiento (A. M. Bolger et al., 2014; Chivian et al., 2023a). Este programa permite la filtración de datos, la eliminación de adaptadores y la remoción de secuencias de baja calidad, de acuerdo con los criterios definidos por el analista (A. Bolger et al., 2024; A. M. Bolger et al., 2014).

### 8.2.1.3. Eliminación de secuencia del hospedero

Para la eliminación de las secuencias correspondientes al hospedero, *H. illucens*, se empleó el genoma de referencia obtenido del Centro Nacional para la Información Biotecnológica (NCBI), con el número de accesoión GCF\_905115235.1 (National Center for Biotechnology Information, 2024). Luego de descargar la secuencia a través de línea de comandos, se asignó el nombre “Hospedero.fn” y se generó una base de datos con la información del hospedero utilizando la herramienta *Bowtie2 v2.4.2*.



Genome assembly iHerIII2.2.curated.20191125 <span>reference</span>	
Download	datasets curl
NCBI RefSeq assembly	GCF_905115235.1
Submitted GenBank assembly	GCA_905115235.1
Taxon	<a href="#">Hermetia illucens</a> (black soldier fly)
WGS project	CAJHNG01
Assembly type	haploid
Submitter	CAM
Date	Nov 21, 2020

Figura 5. Referencia de *H. illucens* en NCBI, tomado de (NCBI, 2024)

Posterior a la creación de la base de datos, se mapeo la secuencia que se trabaja con relación a la base de datos del hospedero, con el fin de eliminar la secuencia hospedero; de *H. illucens*.

### 8.2.2. Secuencias no mapeadas y organización de lecturas

Para poder utilizar la herramienta *SAMtools*, primero se requiere un acceso root, por lo cual se crea un contenedor docker, en el cual contiene todo lo necesario para instalar y correr el programa *SAMtools*, a este contenedor se puede acceder de manera independiente, permitiendo que no interfiera con el resto del espacio de trabajo.

Una vez iniciado el docker, se procede a ejecutar la herramienta *SAMtools*, la cual permite interactuar y ayuda al procesamiento para el alineamiento de lecturas del material genético, con el formato SAM. Una vez que se corrió el alineamiento, se procede a organizar las lecturas en un archivo a que contiene las lecturas paired end.

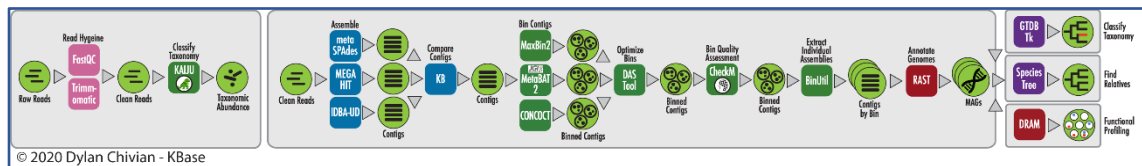


Figura 6. Flujograma para metagenómica en KBase. Tomado de (Chivian et al., 2023b)

### 8.2.3. Generación de lecturas paired end

Para genera los archivos paired end, con todo el análisis que se ha realizado a las secuencias. La generación de estos archivos se realiza mediante la herramienta Bedtools, la cual a partir de archivos en formato .bam genera las lecturas paired end en formato .fastq.

Las secuencias paired end obtenidas al finalizar los análisis, se les coloco el nombre para la secuencia forward “PLA\_F.fastq” y la secuencia reverse “PLA\_R.fastq”.

### **8.3. Ensamblaje y anotación de metagenoma**

#### **8.3.1. Subir secuencias a KBase**

Una vez que se obtuvieron las secuencias paired end “PLA\_F.fastq” y “PLA\_R.fastq”, se cargaron en la narrativa de KBase, en el formato FastQ y unificando las en una sola con el nombre de “PLA”. La cual se utilizó para realizar el ensamblaje y anotación de esta.

#### **8.3.2. Clasificación taxonómica**

Tras obtener la secuencia “PLA”, se procesó en la plataforma KBase. Para el análisis taxonómico, se utilizó la herramienta *Classify Taxonomy of Metagenomic Reads with Kaiju - v1.9.0* (Chivian et al., 2023b), la cual permite determinar la composición microbiana presente en la muestra.

#### **8.3.3. Ensamblaje de Metagenoma, comparación y agrupamiento de contigs**

El ensamblaje del metagenoma se llevó a cabo empleando tres herramientas disponibles en KBase, utilizando la secuencia PLA previamente procesada. Para este procedimiento, se utilizaron *Assemble Reads with metaSPAdes - v3.15.3*, *Assemble Reads with MEGAHIT v1.2.9* y *Assemble Reads with IDBA-UD - v1.1.3* (Chivian et al., 2023b). Estas herramientas permiten la generación de *contigs*, con una longitud mínima de 2000 pb.

Posteriormente, se empleó la herramienta *Compare Assembled Contig Distributions - v1.1.2* (Chivian et al., 2023b), la cual permite visualizar un resumen comparativo de los resultados obtenidos con las tres herramientas utilizadas en el ensamblaje del metagenoma.

Una vez identificado el ensamblaje óptimo, se procedió a su utilización para el agrupamiento de *contigs*, considerando factores como la profundidad de cobertura, la

presencia de genes específicos y los marcadores dentro de cada secuencia. Para ello, se emplearon diversas herramientas disponibles en KBase, incluyendo *Bin Contigs using MaxBin2 - v2.2.4*, *MetaBAT2 Contig Binning - v1.7* y *Bin Contigs using CONCOCT - v1.1*, aplicando la configuración predefinida en cada herramienta (Chivian et al., 2023b).

#### **8.3.4. Optimización de Bins**

Tras la obtención de los *bins* en la etapa previa, se llevó a cabo su optimización siguiendo la metodología descrita en KBase para el ensamblaje de metagenomas del 2023. Para este proceso, se utilizó la herramienta *Optimize Bacterial or Archaeal Binned Contigs using DAS Tool - v1.1.2*. El ensamblaje empleado correspondió al obtenido mediante *metaSPAdes*, junto con los *contigs* generados en pasos anteriores.

#### **8.3.5. Evaluación de los bins**

Una vez obtenidos los *bins*, se realizó una evaluación de su calidad con el fin de determinar el nivel de contaminación presente en cada uno. Este análisis resulta esencial para definir la clasificación taxonómica del ensamblaje. Para ello, se utilizó la herramienta *Assess Genome Quality with CheckM - v1.0.18* (Chivian et al., 2023b).

#### **8.3.6. Anotación del genoma y clasificación taxonómica**

Para llevar a cabo la anotación del genoma, siguiendo el procedimiento descrito en KBase para el ensamblaje de metagenomas, se empleó la herramienta *Annotate Multiple Microbial Assemblies with RASTtk - v1.073*.

Posteriormente, la clasificación taxonómica se realizó utilizando *Classify Microbes with GTDB-Tk - v2.3.2* (Chivian et al., 2023), lo que permitió la organización de los *bins* obtenidos en las etapas previas.

### 8.3.7. Obtención de perfil metabólico y genes de resistencia

Para caracterizar el perfil metabólico del ensamblaje, se empleó la aplicación *Annotate and Distill Assemblies with DRAM - kb\_DRAM v0.1.2* en KBase (Shaffer et al., 2023), utilizando el ensamblaje obtenido en las etapas previas.

Con relación a los genes de resistencia se obtuvo por medio de la plataforma ResFinder (Center for Genomic Epidemiology, 2024; Florensa et al., 2022), perteneciente al Centro de Epidemiología Genómica – CGE (Center for Genomic Epidemiology), perteneciente a la Universidad Técnica de Dinamarca (DTU).

## 9. Resultados y discusión

### 9.1. Control de calidad

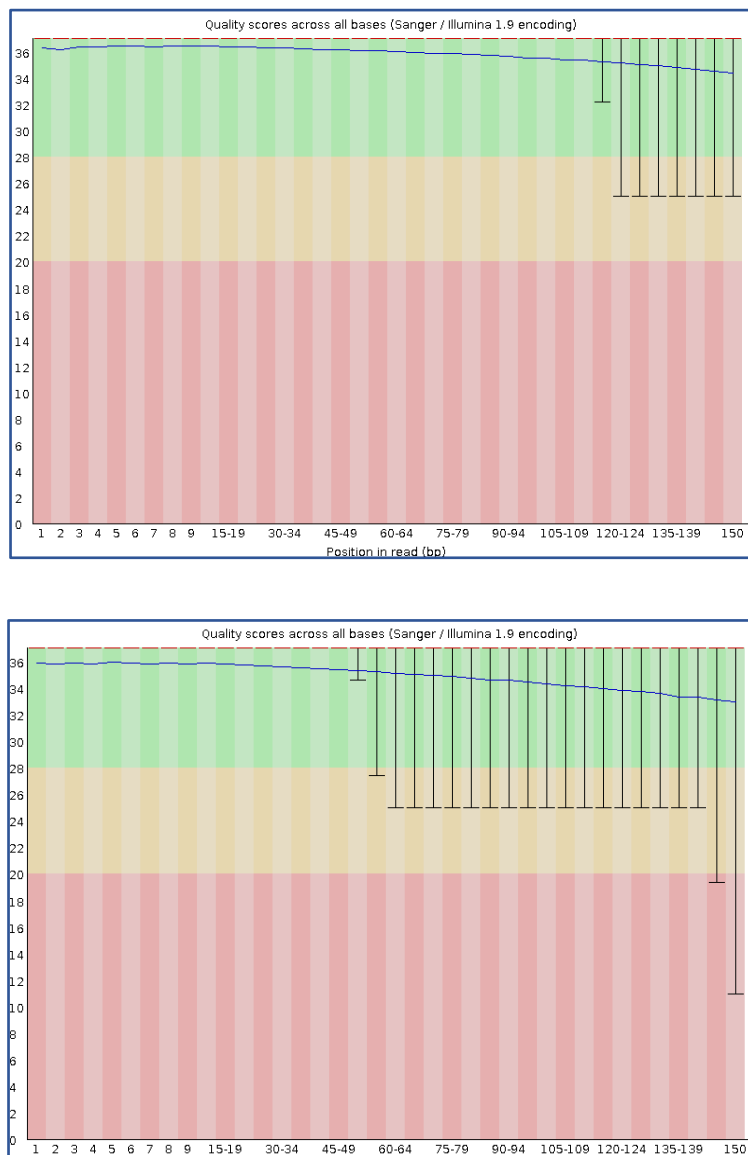
Para iniciar el análisis de las secuencias *forward* y *reverse* obtenidas mediante secuenciación *shotgun* con tecnología Illumina, se llevó a cabo un control de calidad previo.

*Tabla 1. Estadísticas de calidad de las secuencias forward y reverse antes del proceso de limpieza*

Medida	Valor (Forward)	Valor (Reverse)
Nombre del archivo	PLA_FDME210308107-1r_H5Y3FDSX2_L4_1.fq.gz	PLA_FDME210308107-1r_H5Y3FDSX2_L4_2.fq.gz
Tipo de archivo	Llamadas de base convencionales	Llamadas de base convencionales
Codificación	Sanger / Illumina 1.9	Sanger / Illumina 1.9
Secuencias totales	20,327,027	20,327,027
Bases totales	3 Gbp	3 Gbp
Secuencias marcadas como baja calidad	0	0
Longitud de la secuencia	150 pb	150 pb
%GC	41%	41%

En la Tabla 1 se presentan los valores de calidad de las secuencias obtenidas mediante secuenciación con Illumina. Los archivos generados llevan los nombres

*PLA\_FDME210308107-1r\_H5Y3FDSX2\_L4\_1.fq.gz* y *PLA\_FDME210308107-1r\_H5Y3FDSX2\_L4\_2.fq.gz*. No se detectaron errores de calidad durante el proceso de secuenciación; sin embargo, en la gráfica de evaluación de calidad se observa una disminución de la señal al final de la secuenciación, lo que sugiere una caída de la señal en las últimas bases leídas (Figura 6).

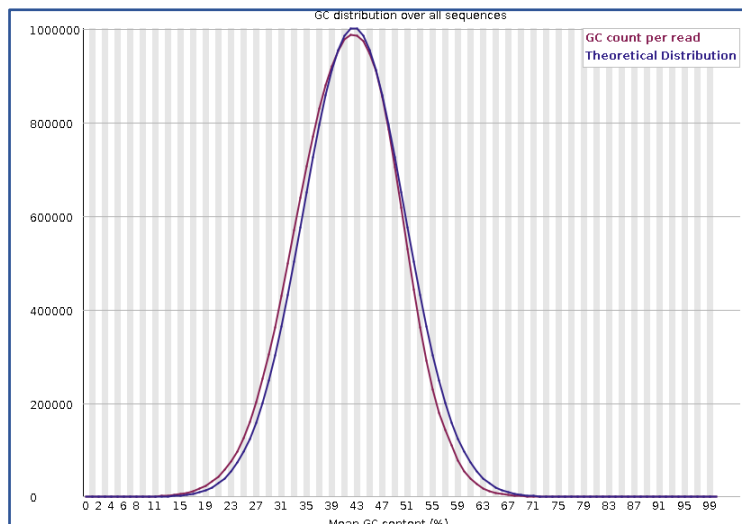


*Figura 7. Evaluación de la calidad de las secuencias forward y reverse antes del proceso de limpieza*  
*Nota: A) secuencia forward. B) secuencia reverse*

En la Figura 6, se observa que en la secuencia representada en la subfigura 6-A, la calidad de las bases a partir de la posición 120 cae por debajo del umbral de aceptación definido por los criterios de calidad (Babraham Institute, 2024). Por otro lado, en la subfigura 6-B, se identifica que las bases comprendidas entre las posiciones 60 y 139 presentan errores dentro de un rango aceptable; sin embargo, a partir de la base 140 hasta la 150, la calidad disminuye significativamente, alcanzando niveles considerados deficientes (Babraham Institute, 2024).

Estos errores pueden estar relacionados con una posible degradación del ADN durante el proceso de secuenciación, lo cual se refleja en ambas subfiguras (Babraham Institute, 2024; Hernández et al., 2020b).

Otro aspecto relevante al evaluar la calidad de las secuencias obtenidas es la concentración de bases GC, ya que puede inducir sesgos durante el proceso de secuenciación. Los resultados de la distribución de GC se presentan en la Figura 7A y 7B.



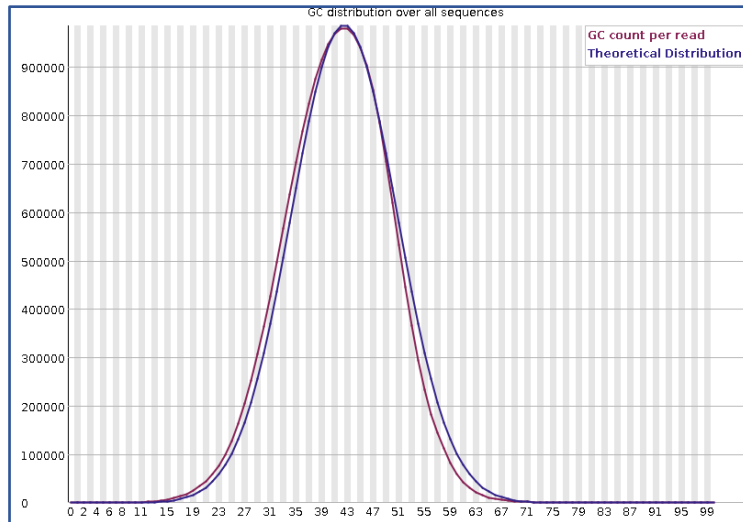


Figura 8. Concentración de GC de las secuencias *forward* y *reverse*  
 Nota: A) Secuencia *forward*. B) Secuencia *reverse*

## 9.2. Limpieza de secuencias

La limpieza de las secuencias *forward* y *reverse* se llevó a cabo utilizando la herramienta *Trimmomatic v0.39*, un software de código abierto diseñado para la depuración de secuencias generadas con tecnología Illumina. Entre sus principales funciones, se encuentra la eliminación de adaptadores y secuencias de baja calidad generadas durante el proceso de secuenciación (A. Bolger, 2024).

Posteriormente, tras la limpieza, se realizó un nuevo análisis de calidad de las secuencias empleando la herramienta *FastQC v0.11.9*.

La cual dio como resultados los datos mostrados en la tabla 2. Los nombres de los archivos cambiaron luego de realizar la limpieza con *Trimmomatic v0.39*, la secuencia *forward* paso a llamarse “*PLA\_1P.fq.gz*” y la secuencia *reverse* “*PLA\_2P.fq.gz*”

Tabla 2. Datos post limpieza de las secuencias forward y reverse

Medida	Valor (Forward)	Valor (Reverse)
Nombre del archivo	PLA_1P.fq.gz	PLA_2P.fq.gz
Tipo de archivo	Llamadas de base convencionales	Llamadas de base convencionales
Codificación	Sanger / Illumina 1.9	Sanger / Illumina 1.9
Secuencias totales	19,279,737	19,279,737
Bases totales	2.8 Gbp	3 Gbp
Secuencias marcadas como de baja calidad	0	0
Longitud de la secuencia (pb)	55 - 150	55 - 150
%GC	41%	40%

Se observa que, en la secuencia forward, antes de aplicar los filtros de *Trimmomatic v0.39* se registraron 20,327,027 lecturas; tras su aplicación, el número de lecturas se redujo a 19,279,737, lo que implica la eliminación de 1,047,290 lecturas durante el proceso de limpieza. Además, el tamaño total de las lecturas disminuyó a 2.8 Gbp, manteniéndose constante la concentración de GC en un 41%.

En la secuencia reverse se observó que, antes de aplicar los filtros, se contaban con 20,327,027 lecturas; tras la limpieza, el número se redujo a 19,279,737, lo que implica la eliminación de 1,047,290 lecturas. El tamaño total de la secuencia fue de 3 Gbp y la concentración de GC disminuyó en 1%, situándose en 40% en comparación con el 41% registrado previamente al uso de *Trimmomatic v0.39*.

En cuanto a la calidad de las secuencias que se observa en la figura

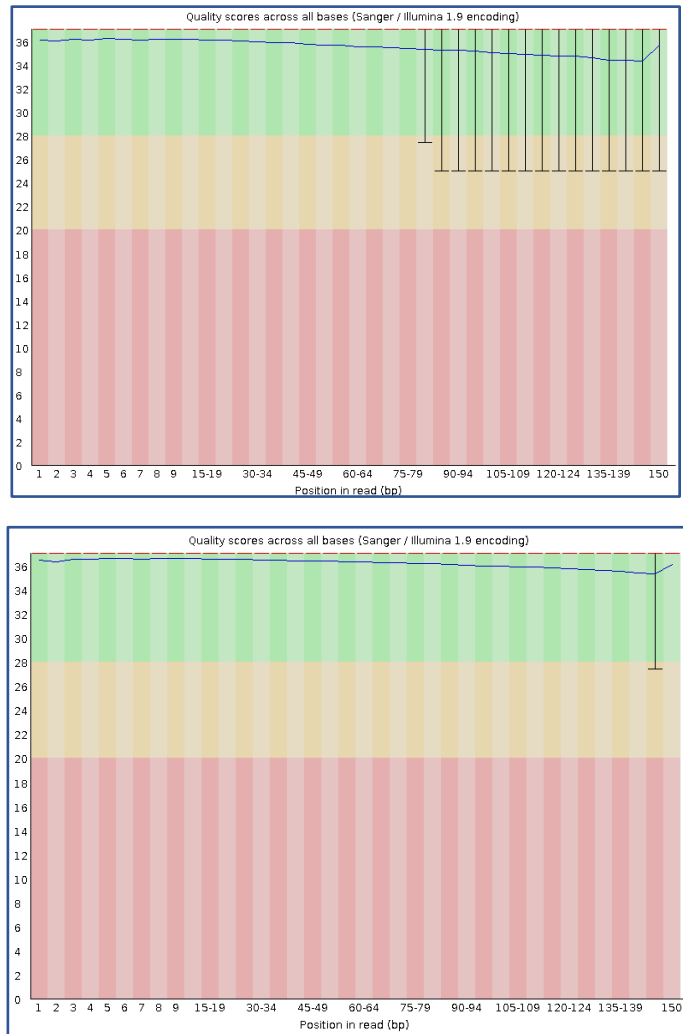


Figura 9. Evaluación de la calidad de las secuencias forward y reverse después del proceso de limpieza  
Nota: A) Secuencia forward. B) Secuencia reverse

Con relación a la secuencia reverse figura 8 – B, se puede observar que luego de la limpieza, se observa que la calidad mejoro; desaparecieron las lecturas que se encontraban en la zona de error. Por lo que se ve una mejor calidad en la secuencia reverse.

### 9.3. Eliminación de secuencia hospedero

Como se mencionó previamente y se observa en la figura 4, el genoma de referencia para *H. ilucens*, con el número de accesoión GCF\_905115235.1, el cual fue descrito en varias publicaciones (Generalovic et al., 2021; Walt et al., 2023). El archivo

fue descargado en formato “.fasta”, para poder trabajar. Mediante el comando “head” se observan las primeras líneas de la secuencia del Hospedero.

```

juan.baez@dgx-node-0-2:~/Secuencias$ head Hospedero.fna
>NC_051849.1 Hermetia illucens chromosome 1, iHerIll2.2.curated.20191125, whole
genome shotgun sequence
atctcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcaggt
tccttttcattatgtgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgattttcagcttttcctcca
ttttcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcaagt
tccttttcattatgtgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgattttcagcttttcctcca
ttttcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcaggt
tccttttcattatgtgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgattttcagcttttcctcca
ttttcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcaggt
tccttttcattatgtgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgattttcagcttttcctcca
ttttcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcaggt

```

Figura 10. Genoma de referencia de *H. illucens* descargado del NCBI

A partir de la secuencia del hospedero (Figuras 4 y 9), la descarga del genoma de referencia desde NCBI y el uso de la herramienta *Bowtie2 v2.4.2*, se generó una base de datos, cuyo resultado se presenta en la Figura 10.

```

bothEndsUnmapped.bam          PLA_1U.fq.gz
bothEndsUnmapped sorted.bam  PLA_2P_fastqc.html
H_DB.1.bt2                    PLA_2P_fastqc.zip
H_DB.2.bt2                    PLA_2P.fq.gz
H_DB.3.bt2                    PLA_2U.fq.gz
H_DB.4.bt2                    PLA_FDME210308107-1r_H5Y3FDSX2_L4_1_fastqc.html
H_DB.rev.1.bt2                PLA_FDME210308107-1r_H5Y3FDSX2_L4_1_fastqc.zip
H_DB.rev.2.bt2                PLA_FDME210308107-1r_H5Y3FDSX2_L4_1.fq.gz
Hospedero.fna                 PLA_FDME210308107-1r_H5Y3FDSX2_L4_2_fastqc.html
mau.sam                        PLA_FDME210308107-1r_H5Y3FDSX2_L4_2_fastqc.zip
PLA_1P_fastqc.html            PLA_FDME210308107-1r_H5Y3FDSX2_L4_2.fq.gz
PLA_1P_fastqc.zip             PLA_F.fastq
PLA_1P.fq.gz                  PLA_R.fastq

```

Figura 11. Base creada a partir de la secuencia Hospedero y *Bowtie2 v2.4.2*

Con la base de datos generada se mapeo las secuencias obtenidas a trabajar se procedió a mapear, donde nos dio como resultado un 56.34% de alineamiento general (figura 11). Resultado en un archivo en formato “.sam” que contiene las secuencias mapeadas.

Una vez que se obtuvo el archivo mapeado, era necesario realizar el cambio de formato mediante la herramienta *SamTools v.1.10-3*, para obtener un archivo en formato “.bam”. Para correr la *SamTools v.1.10-3* en la super computadora de CEDIA; pero al

necesitar acceso root, se utilizó un Docker, el cual permite crear, implementar y probar aplicaciones rápidamente en un espacio separado que cuenta con todo lo necesario para que se ejecute software, bibliotecas, herramientas y código (Amazon Web Services Inc, 2023).

```
19279737 reads; of these:
 19279737 (100.00%) were paired; of these:
   10210286 (52.96%) aligned concordantly 0 times
   5520391 (28.63%) aligned concordantly exactly 1 time
   3549060 (18.41%) aligned concordantly >1 times
----
  10210286 pairs aligned concordantly 0 times; of these:
   281163 (2.75%) aligned discordantly 1 time
----
 9929123 pairs aligned 0 times concordantly or discordantly; of these:
 19858246 mates make up the pairs; of these:
  16836004 (84.78%) aligned 0 times
  1749282 (8.81%) aligned exactly 1 time
  1272960 (6.41%) aligned >1 times
56.34% overall alignment rate
```

Al iniciar el contenedor Docker se pudo correr la aplicación *SamTools v.1.10-3*, para cambiar el archivo “.sam” a un formato “.bam”, y simultáneamente la herramienta permite filtrar las secuencias no alineadas, con este filtro se obtienen únicamente las lecturas que no se encuentran mapeadas dentro del genoma de *H. illucens*, dando un total de 16836004 lecturas.

El archivo “.bam” es el cual se trabaja a partir de ahora, ya que hemos eliminado las lecturas del hospedero, con lo cual ahora contamos únicamente con los organismos que no tienen relación con el hospedero. Para continuar el trabajo de análisis se procedió a acceder a un docker, ya que la configuración de la super computadora de CEDÍA, no permite el acceso a root, por lo cual iniciamos la herramienta

*Figura 12. Resultado de las lecturas mapeadas y no pareadas del metagenoma con relación a H. illucens*

Tras obtener el archivo en formato .bam, que contenía las lecturas no pareadas, se llevó a cabo la conversión a formato .fastq para continuar con el análisis en la

supercomputadora de CEDIA. Para ello, se empleó la herramienta *bedtools* v.2.25.0, generando dos archivos: “PLA\_F.fastq” para la secuencia *forward* y “PLA\_R.fastq” para la secuencia *reverse*, cuyos resultados se presentan en las Figuras 12 y 13.

```
@A00920:768:H5Y3FDSX2:4:1101:1036:16845/1
TGGGTAGAGCACGGGAATACACTCCCTGTTCACTACTCGAAAGCATTCTCCAATCGATGAAGAGCAGGGGAAGCGGTATCTAAACTTCGCA
CACTGTTCCAAAACGATCTATTGGTCAACGCTGGAGGATTTGGAGCGGAAACCCAC
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00920:768:H5Y3FDSX2:4:1101:1036:29371/1
CATTGGAACAGTATAAAAATAGTCCAAGCGCTGTTTTGTACAGATAAACACCAATAAAATGTAGTAAGTAACATTGTGATAAGAGTACTATTTT
ATTTATTTAAGTACTCTATTTCTATTATTTCAAAATCACACTGTCAGATTACTT
+
F,FFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
F, :FFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00920:768:H5Y3FDSX2:4:1101:1036:31125/1
AAAACCACTATTTAGTCAATGGGCTGCGATTATGCAGTCCGAAAAAGGCAACTAATCGAGCTAGATAGTGTGTTGGCGATGGTATTAGGTT
TGACTATGAGTGATGGGTTTACTGCGGCTAATTA
```

Figura 13. Secuencia forward

```
@A00920:768:H5Y3FDSX2:4:1101:1036:16845/2
TGAAGTCTCCTCGCAGATTTATTTATCGCTGCACCTGCAGTTACTGCAGATCTGCTACTTCTCGTAGGAAATCTTGGGAATCCGAAACTTTC
CCAGAGAGTAGAAAAAATGGATGATGCAAAGAGGGAATTCGCTTTAAGTATGA
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
FFFFFFFF:FF,FFFF: :FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00920:768:H5Y3FDSX2:4:1101:1036:29371/2
CTTCCTGGGCACGGTTTTTTCATCACCACCTCCTCGAGTTTCTTAGGAGATGTAAGTTCTGCTTCTTAAGGGAAGAAAATAACTGCCTAAAAAA
ACAACAACGGGAGTCCCTCTGTTTACAAGATGTCAAAGACAGTCATCAATA
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00920:768:H5Y3FDSX2:4:1101:1036:31125/2
TCACCAAATTCAGTGCTATCAATTGATGGACGATCTTTAAACGTTTTTCTACTTGCATCAGGCCCGAAGCCATCAAGGTTACCATTGTTGATGG
AACAGCATTATCATTGTTTACCAG
```

Figura 14. Secuencia reverse

## 9.4. Ensamblaje y anotación del metagenoma en KBase

### 9.4.1. Clasificación taxonómica

Para continuar con el análisis se subió las secuencias a “PLA\_F.fastq” y “PLA\_R.fastq” a KBase (figura 14) con el nombre “PLA”

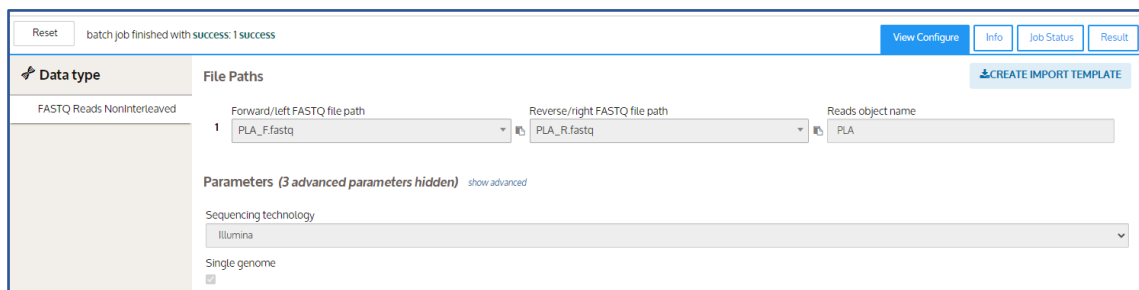


Figura 15. Carga de las secuencias a KBase

Para la predicción de la composición microbiana en las secuencias analizadas, se empleó la herramienta *Classify Taxonomy of Metagenomic Reads with Kaiju - v1.9.0* en KBase, lo que permitió obtener una estimación aproximada del contenido microbiano presente (Lv et al., 2020b; Menzel et al., 2016). Dentro de la configuración de esta herramienta, se utilizó como referencia la base de datos “RsfSeq Genomes (no Euks)”, y se solicitó la generación de todos los niveles taxonómicos. Los resultados obtenidos se presentan en las Figuras 15 a 20, tomadas directamente de KBase(US. Department of ENERGY, 2024).

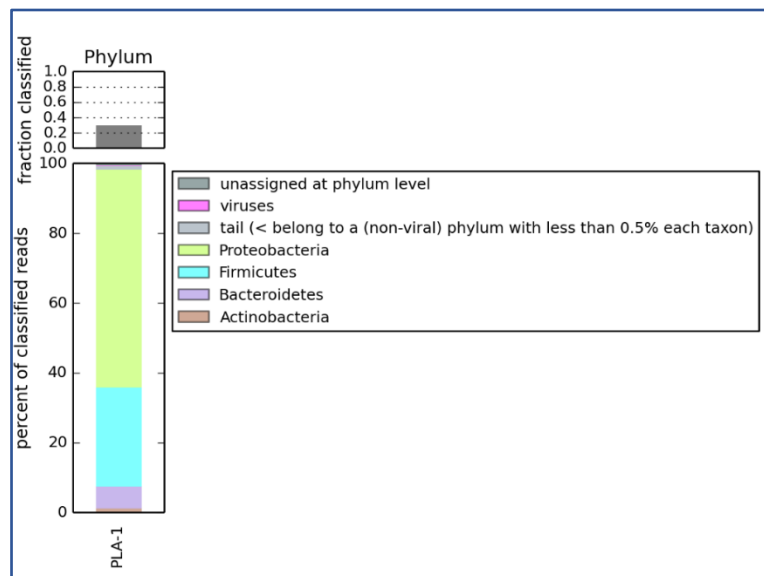


Figura 16. Distribución de la clasificación taxonómica a nivel de filo en la muestra analizada

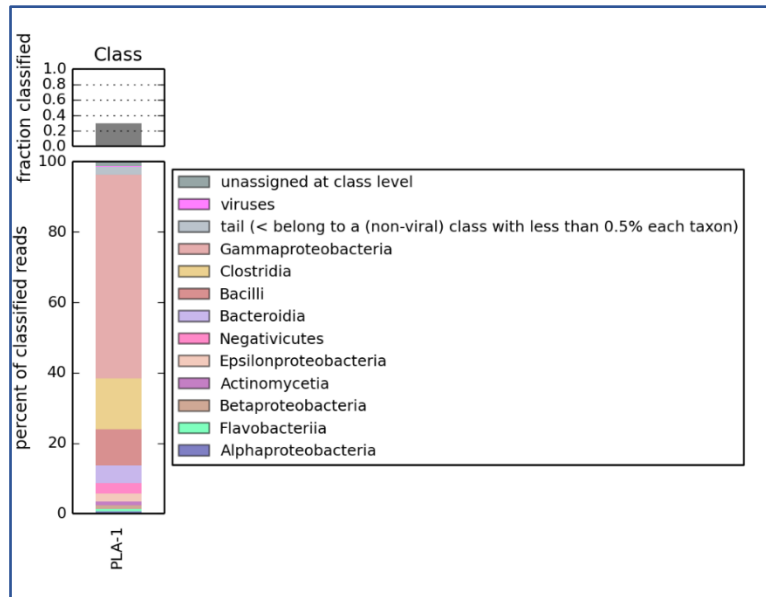


Figura 17. Distribución de la clasificación taxonómica a nivel de clase en la muestra analizada

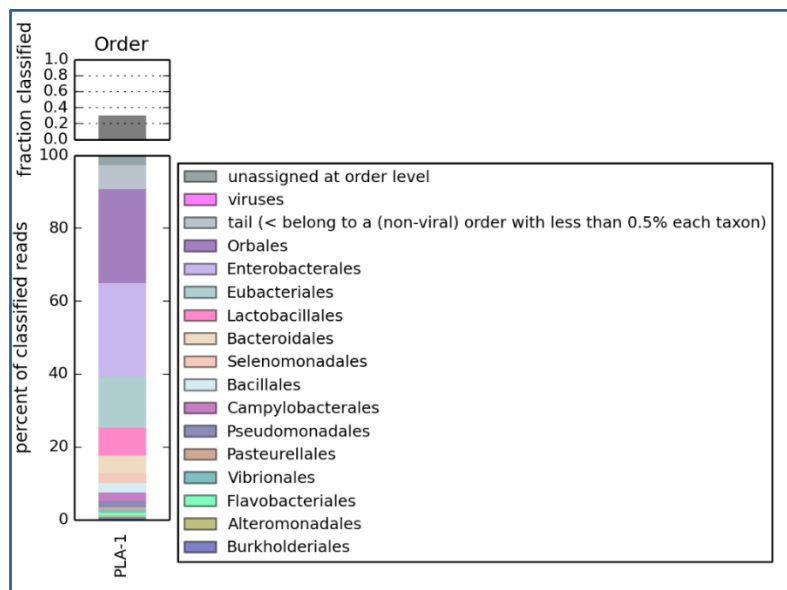


Figura 18. Distribución de la clasificación taxonómica a nivel de orden en la muestra analizada

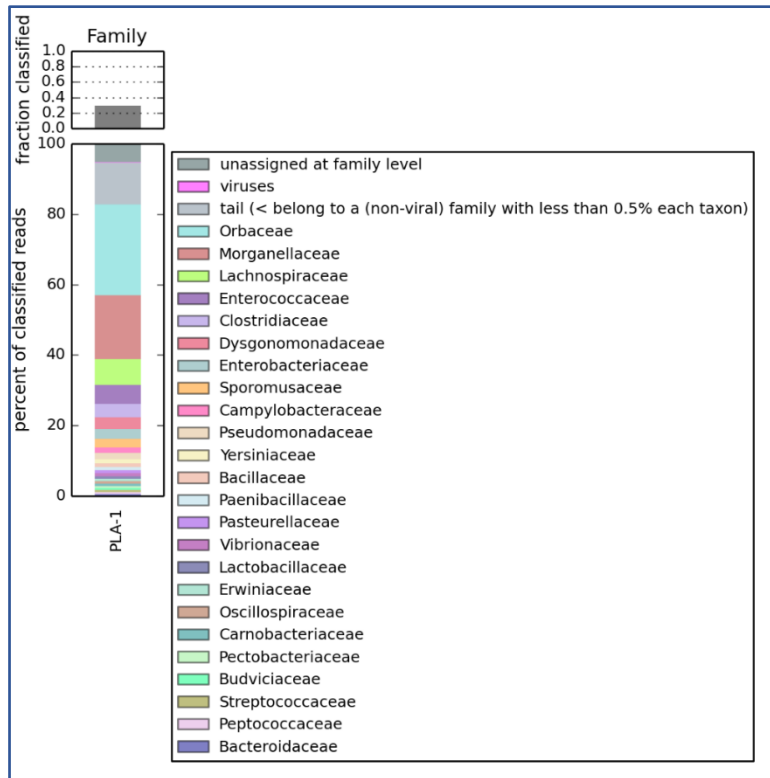


Figura 19. Distribución de la clasificación taxonómica a nivel de familia en la muestra analizada

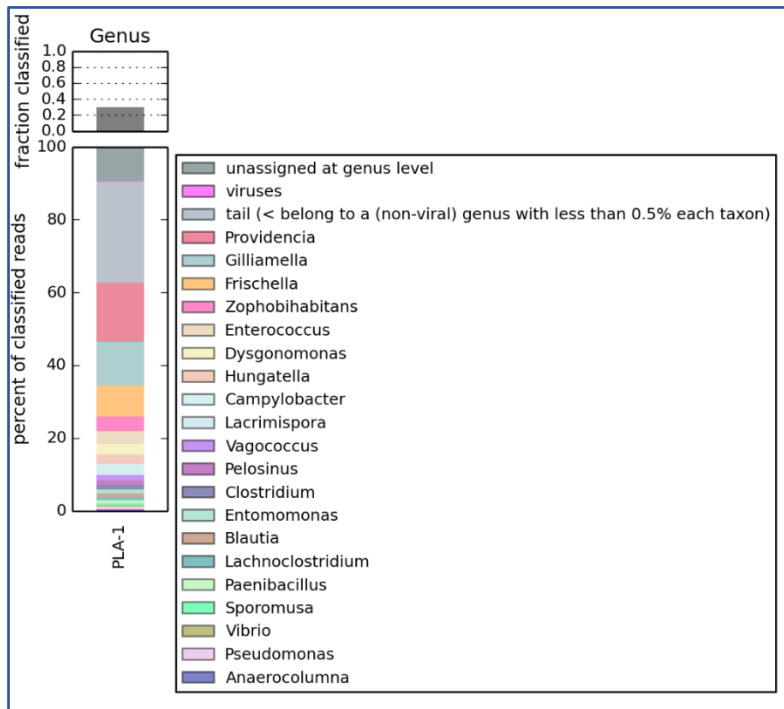


Figura 20. Distribución de la clasificación taxonómica a nivel de género en la muestra analizada

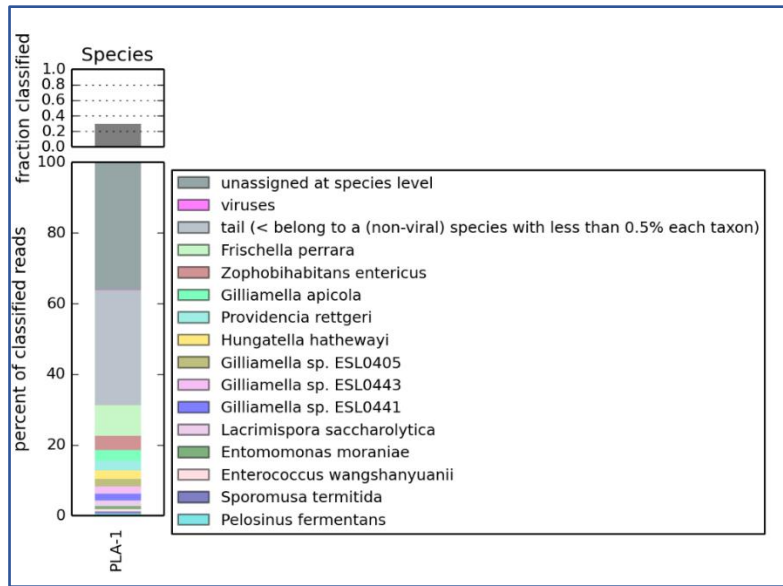


Figura 21. Distribución de la clasificación taxonómica a nivel de especie en la muestra analizada

En diversos estudios se ha encontrado que los microorganismos con mayor predominancia en el tracto digestivo de *H. illucens*, como se puede observar en la figura 21, las siguientes familias *Bacteroidetes*, *Actinobacteria*, *Firmicutes*, *Proteobacteria* y las especies *Acinetobacter*, *Bacillus*, *Klebsiella*, *Providencia*, *Morganella*, etc (Casariego, 2020; Khamis et al., 2020).

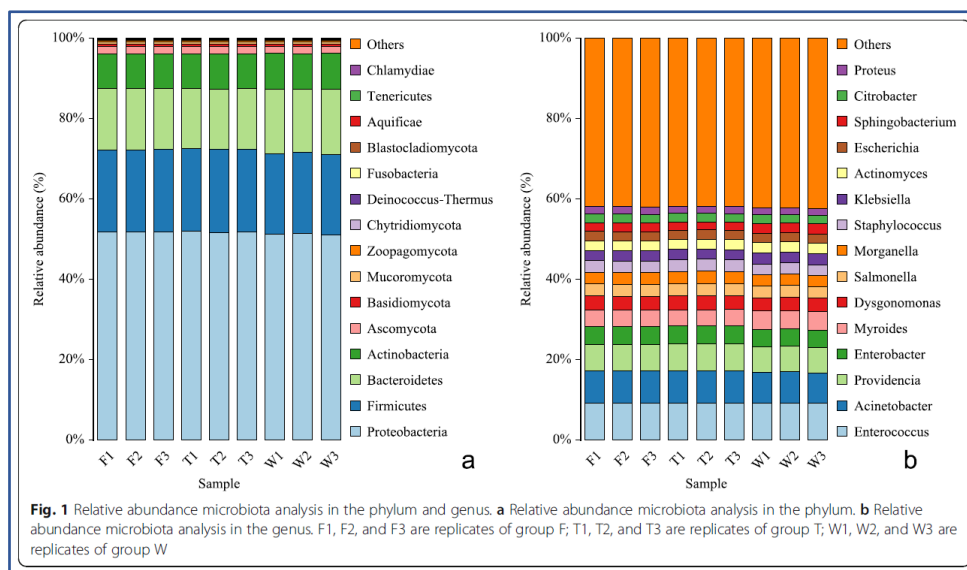


Fig. 1 Relative abundance microbiota analysis in the phylum and genus. **a** Relative abundance microbiota analysis in the phylum. **b** Relative abundance microbiota analysis in the genus. F1, F2, and F3 are replicates of group F; T1, T2, and T3 are replicates of group T; W1, W2, and W3 are replicates of group W

Figura 22. Presencia de microorganismos encontrados en *H. illucens*, tomado de (Khamis et al., 2020)

Dentro del resultado obtenido, se puede observar que dentro de la muestra existe con mayor presencia de *Protobacterias* con una presencia  $\geq 40\%$ , como se puede observar en la figura 15. Pero también se puede observar la presencia de genoma de virus que equivale al 0,1% dentro de la muestra y un 0,3% de otros organismos no identificados. Continuando con la descripción de los datos obtenidos, con relación a la clase se puede observar en la figura 16 la mayor presencia la tiene *Gammaproteobacteria* con una presencia mayor al  $\geq 50\%$ . A nivel de orden se puede observar que la mayor presencia la comparten *Orbales* y *Enterobacterales*; que se puede observar en la figura 17. A nivel de género quien predomina es *Gilliamella*, como se observa en la figura 18.

En relación con *Campylobacter*, se observa que su detección solo alcanza hasta el nivel de género. No se logró identificar a nivel de especie en la predicción obtenida con *Classify Taxonomy of Metagenomic Reads with Kaiju - v1.9.0*.

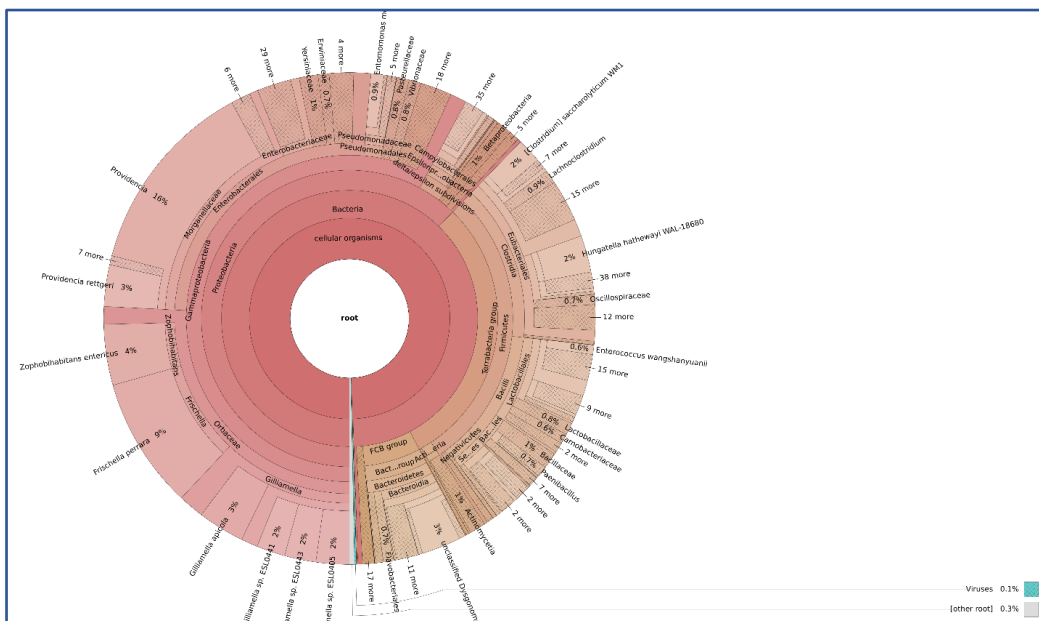


Figura 23. Diagrama de KRONA a partir de KAIJU

#### 9.4.2. Ensamblaje del metagenoma, comparación de contigs y optimización de bins

Para realizar el ensamblaje de la secuencia de trabajo, se emplearon las herramientas *Assemble Reads with metaSPAdes - v3.15.3*, *Assemble Reads with MEGAHIT v1.2.9* y *Assemble Reads with IDBA-UD - v1.1.3*. Posteriormente, se utilizó la herramienta *Compare Assembled Contig Distributions - v1.1.2*, la cual permite comparar los ensamblajes generados en términos de longitud, distribución y calidad de los contigs.

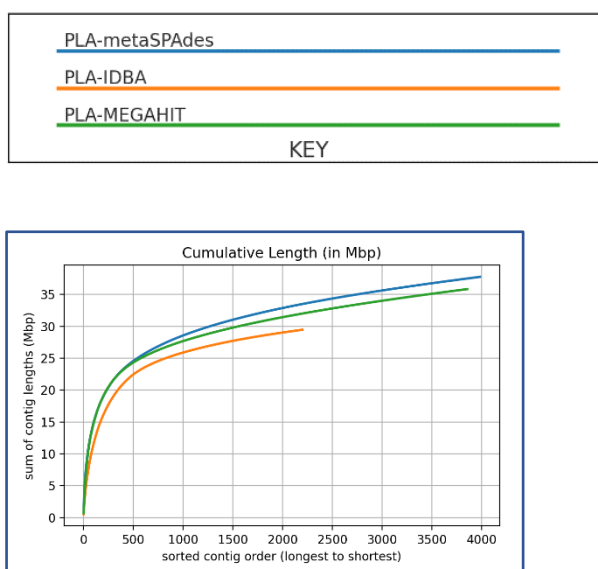


Figura 24. Clasificación de contigs del más largo a más corto

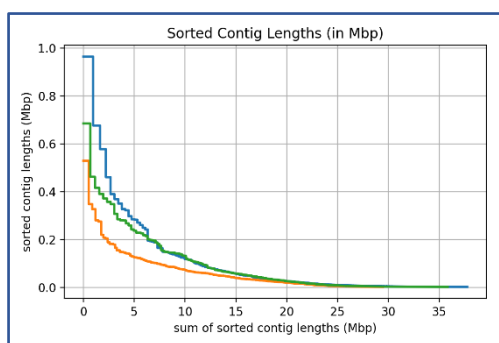


Figura 25. Suma de contigs con relación a su longitud en Mbp

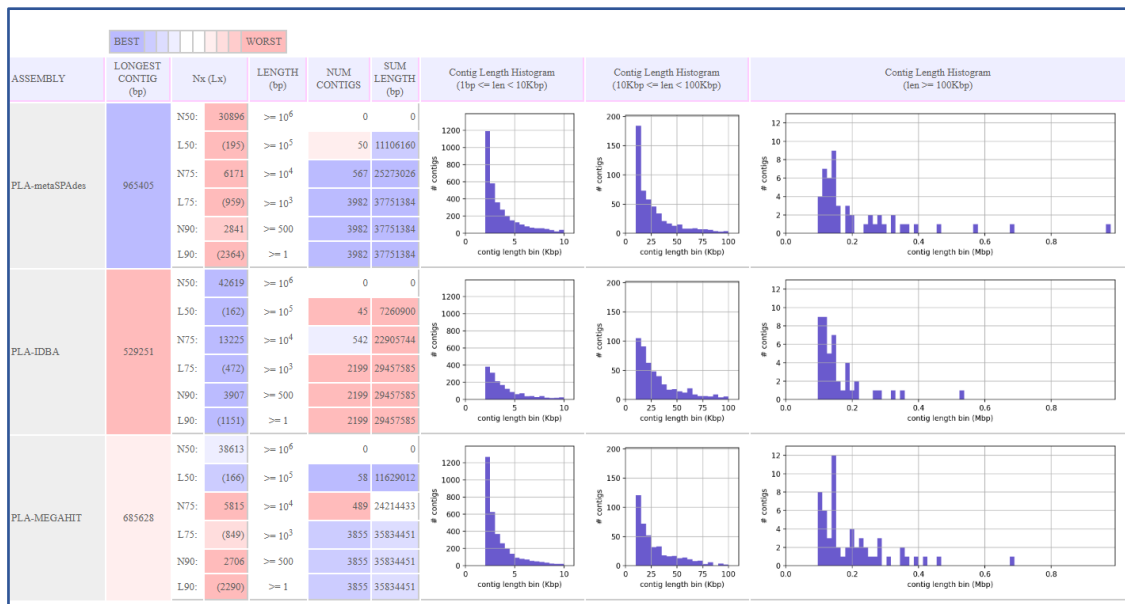


Figura 26. Comparación de los contigs generados mediante las diferentes herramientas de ensamblaje

Se tomó la decisión de utilizar el ensamblaje realizado a partir de *Assemble Reads with metaSPAdes - v3.15.3*, donde se obtuvo el contig de mayor tamaño que fue de 965405 pb, el N50 fue de 30896 pb y el contig de mayor tamaño en pb resultante fue de 11106160.

Tras seleccionar *metaSPAdes* como ensamblaje de referencia, se procedió a realizar el agrupamiento de los contigs utilizando diversas herramientas para generar los bins. La obtención de los bins se llevó a cabo mediante las herramientas *Bin Contigs using MaxBin2 - v2.2.4*, *MetaBAT2 Contig Binning - v1.7* y *Bin Contigs using CONCOCT - v1.1*.

El número de bins generados por cada herramienta se puede visualizar en la tabla

3.

Tabla 3. Número de bins obtenidos con cada herramienta

Nombre de la herramienta	Número de bins obtenidos
<i>Contigs using MaxBin2 - v2.2.4</i>	11
<i>MetaBAT2 Contig Binning - v1.7</i>	13
<i>Bin Contigs using CONCOCT - v1.1</i>	46

Para mejorar la calidad de los bins obtenidos, se realizó un proceso de optimización utilizando la herramienta *Optimize Bacterial or Archaeal Binned Contigs using DAS Tool - v1.1.2*. Como resultado, se obtuvo un total de 11 bins optimizados, los cuales presentaron la mejor calidad dentro del conjunto analizado (Figura 26 y 27).

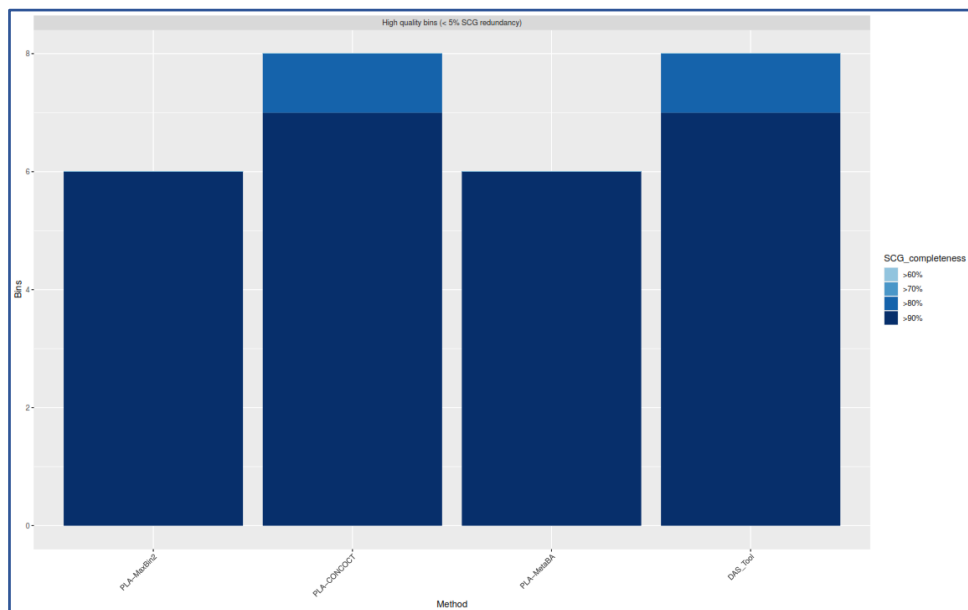


Figura 27. Comparación del número de bins y la calidad general de cada ensamblado

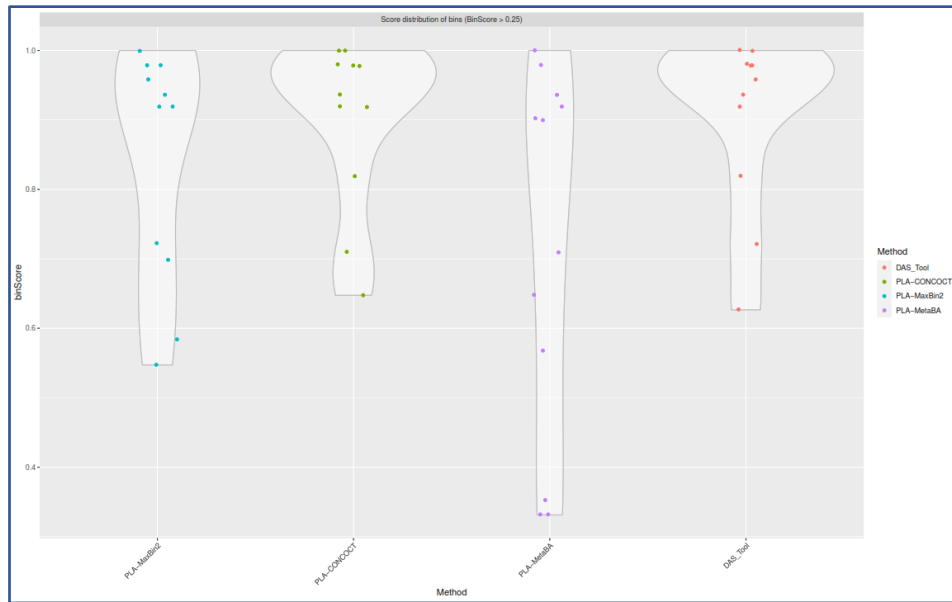


Figura 28. Comparación de la eficacia de diferentes métodos de agrupamiento de bins

#### 9.4.3. Evaluación de la calidad e integridad de los bins obtenidos

on el fin de evaluar la **integridad y contaminación** dentro del ensamblaje metagenómico y de los bins obtenidos, se empleó la herramienta *Assess Genome Quality with CheckM - v1.0.18*. Esta evaluación permite determinar la calidad de los bins mediante el análisis de su completitud y la presencia de contaminación en las secuencias clasificadas.

Los resultados obtenidos a partir de este análisis pueden observarse en la figura 28, donde se representa la distribución de los valores de calidad para cada bin generado.

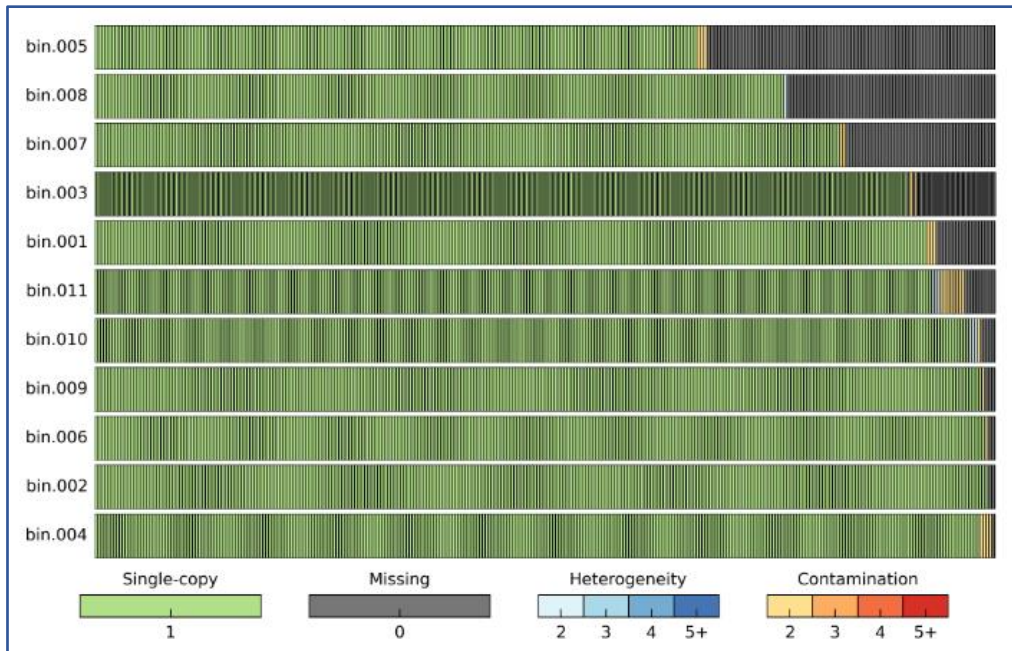


Figura 29. Contaminación e integridad de los bins obtenidos

Dentro del resultado obtenido, se observa que dentro de algunos bins obtenidos, como son el bin.005, bin.007, bin.003, bin.001 y bin.011, existe una contaminación mínima, tomando en consideración que la obtención de los bins, se realizaron mediante el ensamblaje de mejor calidad.

Otra de las funciones que ofrece la herramienta *Assess Genome Quality with CheckM - v1.0.18* es la obtención previa de un linaje asociado a cada bin, cuyos resultados se presentan en la tabla 5.

En esta tabla se observa que el bin.011 es el que presenta el mayor nivel de contaminación, con un valor del 4.01%. Además, en la columna 1 se muestra el número total de genes que actúan como marcadores dentro del genoma, lo que permite evaluar la calidad y composición de cada bin obtenido.

Se puede observar que el resultado obtenido en la tabla 4 presenta una relación con la figura 16; la cual fue obtenida a partir de una predicción realizada con la

herramienta *Kaiju*. Esto denota que se observa una gran presencia de la clase *Gammaproteobacteria*, siendo los bins 002 y 010. También podemos notar la presencia de la clase *Epsilonproteobacteria* en el bin.003, a la cual pertenece *Campylobacter*. El bin.003 posee un total de 445 genes marcadores, una contaminación 0.86% y un total de 111 genomas hasta este punto del análisis.

#### 9.4.4. Anotación del genoma

A partir de los bins obtenidos, se procedió a realizar su anotación taxonómica utilizando la herramienta *Classify Microbes with GTDB-Tk - v2.3.2*. Esta herramienta proporciona información clave, como el dominio y el nombre científico asignado a cada uno de los bins, permitiendo una mejor clasificación de los organismos presentes en la muestra.

Como previamente se observó que el bin.003, representa a la clase *Epsilonproteobacteria*, se realizó observo los resultados de la anotación de este genoma, dando como resultado los siguientes datos presentes en la tabla 4.

*Tabla 4. Caracterización taxonómica y fuente de datos del Bin.003*

Nombre del objeto	PLA-Bin3
Nombre científico	Taxón desconocido
Dominio	B
Código genético	11
Fuente	Kbase
ID de la fuente	bin.003.fasta_assembly.RAST
Tamaño	1709401

Tabla 5. Evaluación de linaje obtenido a partir de los bins

Nombre	Marcador de linaje	Número de genomas	Número de marcadores	Número de sets	0	1	2	3	4	5+	Completo (%)	Contaminación (%)
bin.001	c_Gammaproteobacteria	965	277	177	18	256	3	0	0	0	93.34	1.69
bin.002	c_Gammaproteobacteria	965	277	177	2	275	0	0	0	0	99.44	0
bin.003	c_Epsilonproteobacteria	111	445	271	39	401	5	0	0	0	90.52	0.86
bin.004	p_Firmicutes	100	295	158	1	290	4	0	0	0	99.91	1.9
bin.005	o_Clostridiales	172	263	149	84	176	3	0	0	0	68.58	0.71
bin.006	k_Bacteria	433	273	183	2	270	1	0	0	0	98.91	0.27
bin.007	o_Clostridiales	155	278	158	46	230	2	0	0	0	82.89	0.69
bin.008	k_Bacteria	433	273	183	63	209	1	0	0	0	74.1	0.09
bin.009	o_Clostridiales	155	278	158	3	274	1	0	0	0	98.42	0.63
bin.010	c_Gammaproteobacteria	899	312	185	5	303	4	0	0	0	97.3	1.8
bin.011	c_Bacilli	586	325	181	11	302	12	0	0	0	94.48	4.01

Taxonomy
cellular organisms
Bacteria
Proteobacteria
delta/epsilon subdivisions
Epsilonproteobacteria
Campylobacterales
Campylobacteraceae
Campylobacter

Figura 30. Taxonomía del bin.003

Como se muestra en la figura 29, se puede observar la taxonomía de *Campylobacter* obtenida a partir del análisis metagenómico. Los resultados obtenidos coinciden con lo reportado en la bibliografía, ubicando a *Campylobacter* dentro del dominio *Bacteria*, filo *Pseudomonadota*, clase *Epsilonproteobacteria*, orden *Campylobacterales*, familia *Campylobacteraceae* y género *Campylobacter* (Alonso et al., 2023; El-Naenaeey et al., 2020; Terzolo et al., 2007).

De igual manera los datos resultantes obtenidos muestran que el tamaño del gen era de 1709401 pb, de las cuales 285 eran no codificantes, 1815 pertenecen a genes codificantes, 200 genes no repetidos sin codificación y 37 ARN no codificantes.

#### 9.4.5. Clasificación taxonómica

Para obtener la clasificación taxonómica de los 11 bins previamente obtenidos, se utilizó la herramienta *Classify Microbes with GTDB-Tk - v2.3.2*. Los resultados obtenidos se presentan de manera resumida en la tabla 6.

Según lo encontrado en bibliografía *Campylobacter* se encuentra al momento de secuenciar los microorganismos dentro de *H. illucens* (Yu et al., 2023; Y. Zhang et al., 2022). Se debe destacar que en varios artículos se ha detectado *Campylobacter* hasta nivel de género (Yu et al., 2023; Y. Zhang et al., 2022).

Tabla 6. Clasificación taxonómica obtenida

Bin	Filo	Clase	Orden	Familia	Género	Especie
Bin.001	<i>Pseudomonadota</i>	<i>Gammaproteobacteria</i>	<i>Enterobacterales</i>	<i>Enterobacteriaceae</i>	-	-
Bin.002	<i>Pseudomonadota</i>	<i>Gammaproteobacteria</i>	<i>Enterobacterales</i>	<i>Enterobacteriaceae</i>	<i>Orbus</i>	-
Bin.003	<i>Campylobacterota</i>	<i>Campylobacteria</i>	<i>Campylobacterales</i>	<i>Campylobacteraceae</i>	<i>Campylobacter_B</i>	-
Bin.004	<i>Bacillota_C</i>	<i>Negativicutes</i>	<i>DSM-1736</i>	<i>Dendrosporobacteraceae</i>	-	-
Bin.005	<i>Bacillota_A</i>	<i>Clostridia</i>	<i>Lachnospirales</i>	<i>UBA5962</i>	<i>UBA5962</i>	-
Bin.006	<i>Bacteroidota</i>	<i>Bacteroidia</i>	<i>Bacteroidales</i>	<i>Dysgonomonadaceae</i>	<i>Dysgonomonas</i>	-
Bin.007	<i>Bacillota_A</i>	<i>Clostridia</i>	<i>Lachnospirales</i>	<i>Lachnospiraceae</i>	-	-
Bin.008	<i>Bacteroidota</i>	<i>Bacteroidia</i>	<i>Bacteroidales</i>	<i>Dysgonomonadaceae</i>	<i>Dysgonomonas</i>	-
Bin.009	<i>Bacillota_A</i>	<i>Clostridia</i>	<i>Lachnospirales</i>	<i>Lachnospiraceae</i>	<i>CHH4-2</i>	-
Bin.010	<i>Pseudomonadota</i>	<i>Gammaproteobacteria</i>	<i>Enterobacterales</i>	<i>Enterobacteriaceae</i>	<i>Providencia</i>	<i>providencia rettgeri_D</i>
Bin.011	<i>Bacillota</i>	<i>Bacilli</i>	<i>Lactobacillales</i>	<i>Vagococcaceae</i>	-	-

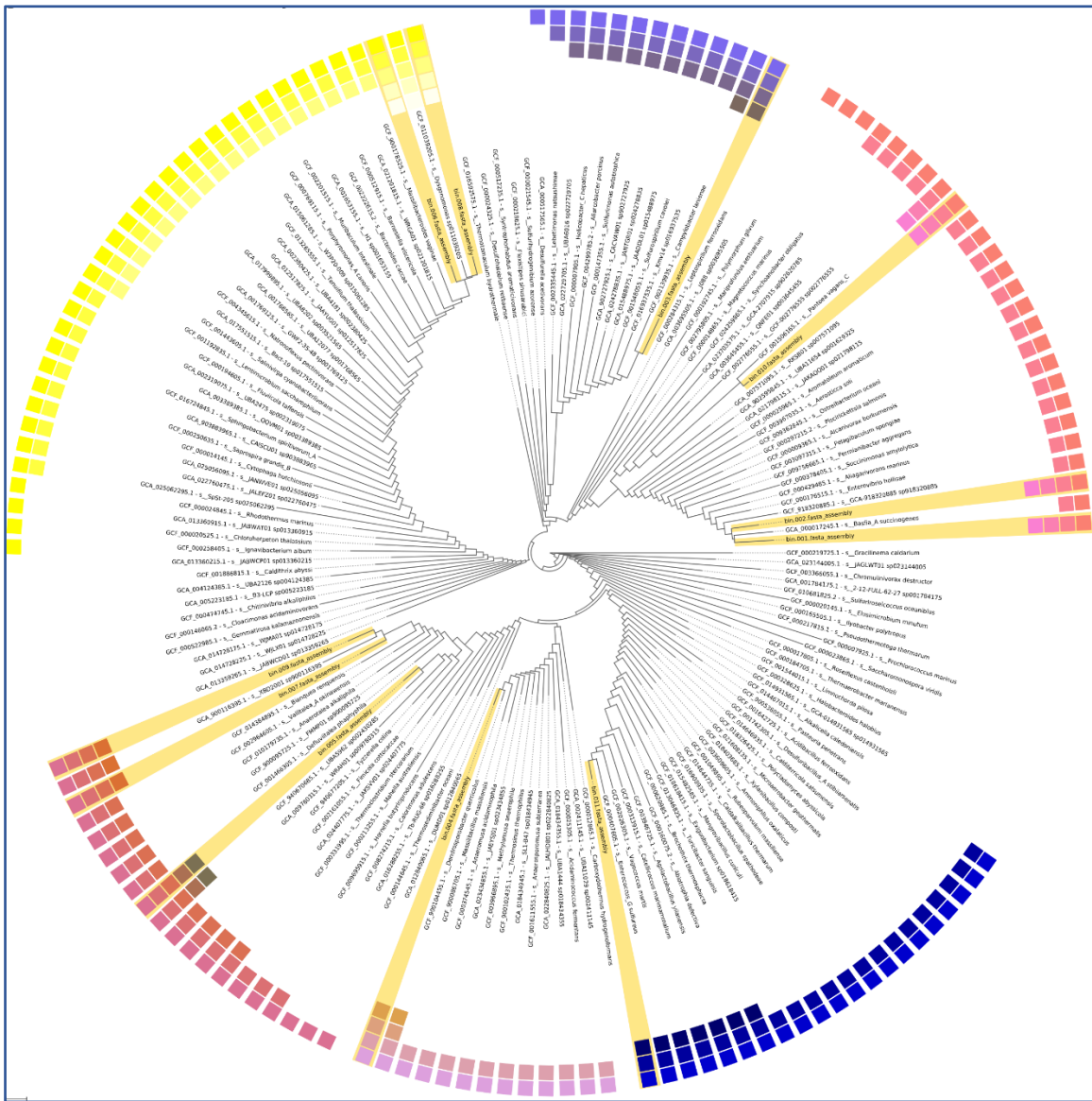
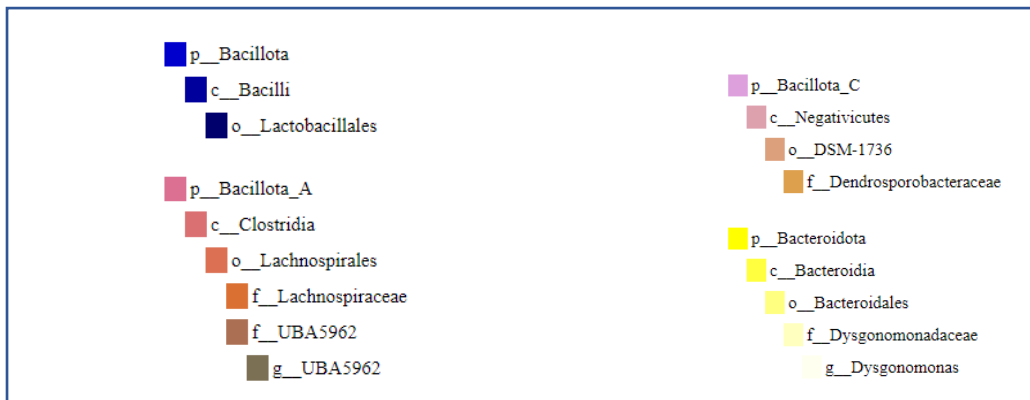


Figura 31. Árbol taxonómico obtenido a partir de los 11 bins obtenidos

Nota: Leyenda de la figura 30, donde se muestran los microorganismos encontrados en la muestra



En la figura 30 se puede observar la relación entre los diferentes organismos presentes en la muestra, así como su interacción con cada una de las bacterias identificadas. En esta representación, se evidencia la presencia de *Campylobacter*.

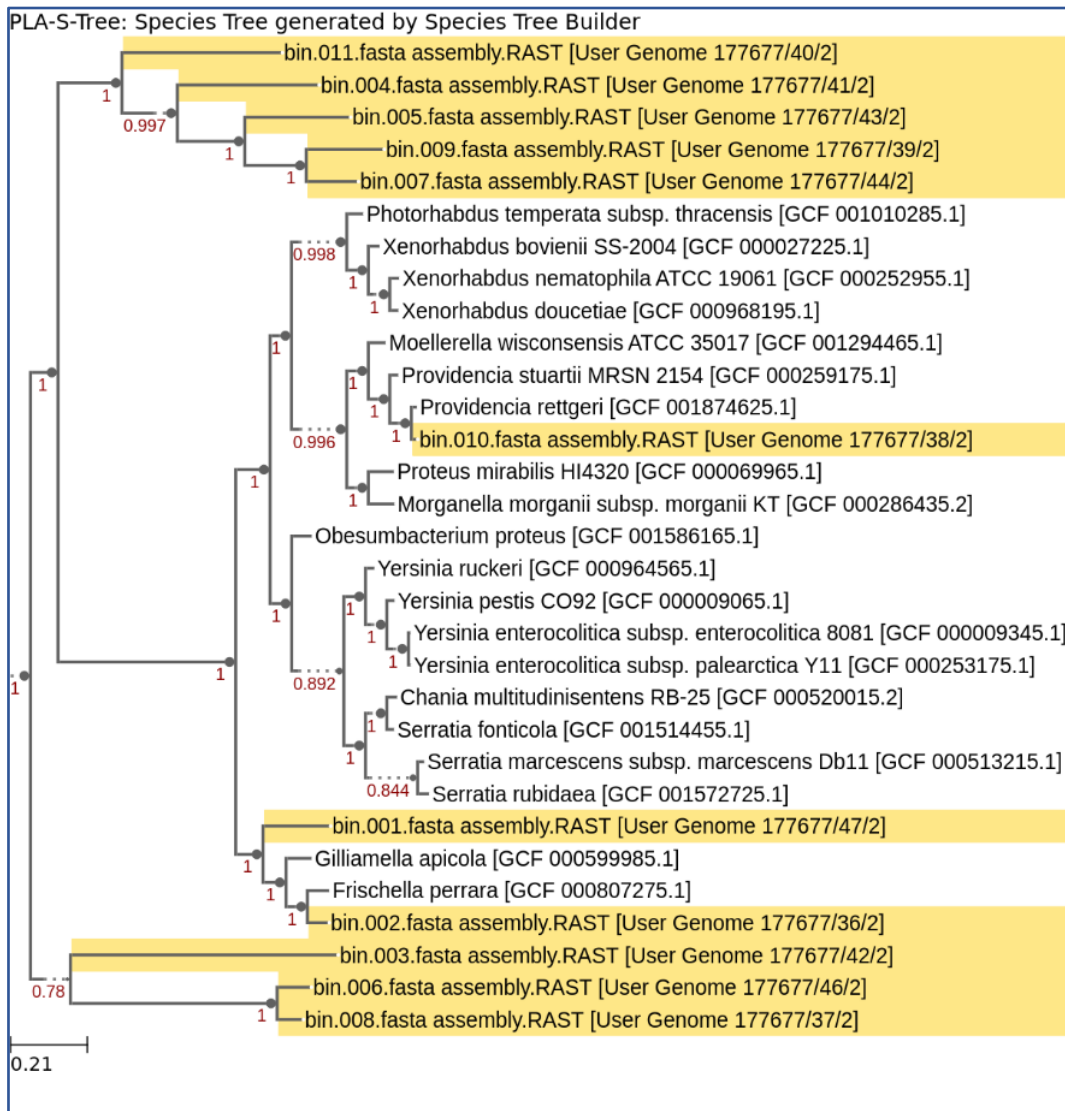


Figura 32. Árbol de especies obtenidos a partir del análisis de los bins

Una vez obtenido el árbol filogenético mediante la herramienta *Insert Set of Genomes Into SpeciesTree - v2.2.0*, se generó una representación que muestra la relación entre los bins obtenidos (figura 31). Se observa que los bins 011, 004, 005, 009 y 007 se agrupan en el mismo clado, lo que sugiere que están relacionados filogenéticamente. Por

otro lado, los bins 003, 006 y 008 comparten un ancestro en común. En contraste, el bin 010 no muestra relación con ningún otro bin dentro del árbol filogenético.

### 9.5. Genes de resistencia y rutas metabólicas

El perfil metabólico del metagenoma se obtuvo mediante el uso de la herramienta *Annotate and Distill Assemblies with DRAM* (Shaffer et al., 2023) en Kbase, este programa nos brinda una gran cantidad de información crucial dentro del análisis metagenómico. Se realizó el análisis metabólico de todos los bin obtenidos previamente, sin embargo, al ser *Campylobacter* el organismo de interés en el presente trabajo se presentará todo lo relacionado con el bin.003; el cual contiene toda la información relacionada con el organismo de interés. Dentro de la información que se presentó se obtuvo la que el bin.003 cuenta con 60 *scraffolds* y 37 genes de tRNA, lo cual fue obtenido posterior al análisis. Se obtuvieron las siguientes figuras, las cuales figura 33 y 34.

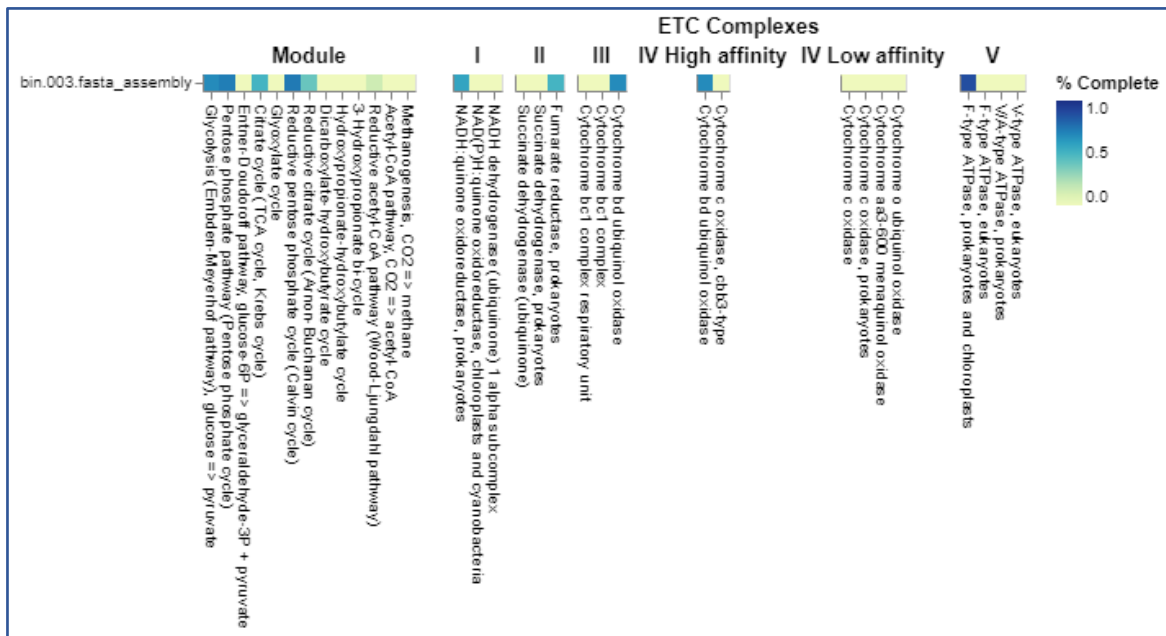


Figura 33. Mapa de calor de los complementos cadena de transporte de electrones (ETC)

Con relación a la figura 33, se observan las rutas metabólicas y la cadena de transporte de electrones, también conocida como *Electron Transport Chain* (ETC, por sus

siglas en inglés). La figura 33 presenta un diagrama de colores donde las rutas representadas en azul oscuro indican que están completas al 100%, mientras que aquellas en verde claro representan rutas incompletas dentro del genoma analizado.

Los complementos de la cadena de electrones – ETC, se compone de distintos módulos numerados del I a l V. El complejo I, corresponde al NADH deshidrogenasa; juega un rol importante dentro del metabolismo energético de *Campylobacter spp*, el cual permite la transferencia desde donantes hacia moléculas que acepten los electrones, lo cual permite la generación de ATP (Klotz et al., 2019; Weerakoon & Olson, 2008; Yagi et al., 2006). Una característica especial de la especie *Campylobacter* que dona electrones a partir de flavodoxina reducida, ya que en algunas ocasiones la disponibilidad de NADH suele ser reducido para la producción de ATP (Weerakoon & Olson, 2008). Otra función dentro de este complejo es que interviene en la patogenicidad y como interactúa *Campylobacter* con el huésped (Klotz et al., 2019; Weerakoon & Olson, 2008; Yagi et al., 2006).

El complejo II o conocido también como succinato deshidrogenasa – SDH, desempeña un rol fundamental del metabolismo energético en *Campylobacter spp*, la enzima que es generada dentro de este complejo juega un papel fundamental dentro del ciclo de Krebs, mediante el transporte de electrones; donde se oxida el succinato para formar sintetizar fumarato y lo cual permite la transferencia de electrones a la ubiquinona que se encuentra en la membrana celular. Otra característica de este complejo dentro de *Campylobacter*, es que permite la sobrevivencia para generar nuevas colonias en el tracto digestivo y gastrointestinal del huésped, en condiciones muy bajas de oxígeno, conocido como condiciones microaerófilas (Huang & Millar, 2013; Weingarten et al., 2009).

Respecto al complejo III o también llamado citocromo bc<sub>1</sub>, juega el papel para la transferencia de electrones en la membrana celular de *Campylobacter*, esta transferencia ayuda a regular la energía para el crecimiento y mantener la vida del organismo. Al igual que el complejo II ayuda a la supervivencia en condiciones bajas de oxígeno o microaerófilas. Otra ruta metabólica el complejo bc<sub>1</sub> realiza el paso de electrones desde ubiquinol hasta el citocromo c, con lo cual se genera la síntesis de ATP; al generarse un gradiente de protones (Myers & Kelly, 2005; Sellars et al., 2002a; M. A. Smith et al., 2000).

Existen dos tipos de complejo IV, los cuales son alto y bajo. Este complejo también lleva el nombre de citocromo oxidasa alto y bajo, el cual cumple una función indispensable en *Campylobacter*; la cual permite presentar una afinidad alto o baja por el oxígeno presente en el medio donde se desarrolla el organismo, lo cual le da una de sus principales características para sobrevivir en condiciones microaerófilas (Kelly, 2014; Sellars et al., 2002a). Cuando se toma en cuenta la alta afinidad del citocromo oxidasa; permite el desarrollo de *Campylobacter* en ambientes con poca presencia de oxígeno como pueden ser el tracto gastrointestinal de aves, mamíferos, insectos, entre otros. En consecuencia, la baja afinidad al citocromo oxidasa permite la supervivencia en ambiente con altas concentraciones de oxígeno para *Campylobacter* (Hoffmant & Goodman, 1982; Kelly, 2014; Weingarten et al., 2008).

El último complejo V o también llamado ATP sintasa, es muy importante para el metabolismo energético en *Campylobacter*, la cual es una de las más importantes en la producción de energía dentro del organismo (Hoffmant & Goodman, 1982; Rupinder Kaur Kanwar et al., 1994; Weingarten et al., 2008). Este complejo funciona sintetizando ATP, tomando como base ADP y fosfato inorgánico, el cual se produce a partir de gradiente de protones que se genera al momento de la respiración celular en la membrana.

Afectar este complejo puede ser perjudicial para *Campylobacter*, debido a que, sin este complejo, se reduce la supervivencia del organismo (Belli & Fryklund, 1995; Dedieu et al., 2020; Jackson et al., 2007; Rupinder Kaur Kanwar et al., 1994; Wassenaar, 1997).

La figura 34 obtenida a partir del bin.003, que previamente fue confirmado como *Campylobacter spp*, donde se puede observar diversas rutas metabólicas con funciones distintas. Las rutas metabólicas obtenidas son CAZy, metabolismo del nitrógeno, metabolismo del azufre, reductasas, fotosíntesis, metanogénesis y metanotrofia y conversión de SCFA. Las rutas metabólicas presentadas permiten que *Campylobacter* adaptarse a diversos tipos de ambientes con condiciones no tan favorables para este (Hofreuter, 2014; Sellars et al., 2002b). Se debe destacar las rutas metabólicas de nitrógeno y azufre son parte de las características de *Campylobacter* que le permite colonizar el tracto gastrointestinal de animales, insectos y humanos (Hofreuter, 2014; Sellars et al., 2002b; van der Stel & Wösten, 2019), lo cual le da la características de ser un patógeno extraordinario, además de las diversas resistencias a los diferentes antibióticos.

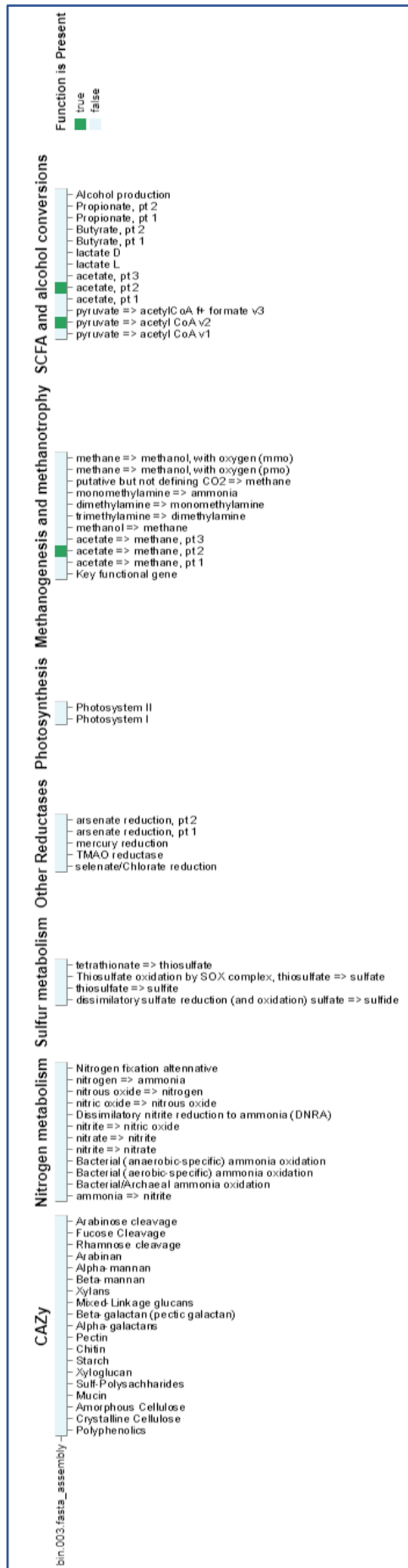


Figura 34. Funciones metabólicas

La herramienta *Annotate and Distill Assemblies with DRAM* proporcionó información clave para el análisis del bin.003, identificado como perteneciente a *Campylobacter*. Esta herramienta genera archivos en formato *.faa*, *.tsv*, *.xlsx* y *.genbank*, permitiendo obtener información detallada sobre los genes identificados, una lista completa de los genes presentes y un mapa genético. Esto facilita la exploración de los genes de resistencia presentes en la muestra (tabla 7).

En la tabla 7, se presenta los diversos genes de resistencia, que presenta dentro del genoma ensamblado, posterior a todos los análisis y lo cual nos permite ver los diversos genes de resistencia dentro de *Campylobacter spp.* Sin embargo, para poder confirmar los genes de resistencia se utilizó la plataforma del “Center for Genomic Epidemiology” dando Los resultados presentados en la tabla 8 indican que no se detectó la **presencia de genes de resistencia a diversos antibióticos**, los cuales se esperaba encontrar en *Campylobacter spp.*, según los datos obtenidos en la plataforma *Center for Genomic Epidemiology* (CGE). Esta ausencia de detección puede deberse a múltiples factores, entre ellos: la calidad del ADN extraído, la eficacia del proceso de limpieza del material genético, la fragmentación del ADN o la posible presencia de material genético del huésped u otros organismos en la muestra metagenómica. Estos factores pueden enmascarar los genes de resistencia o dificultar su detección durante el análisis (Ahrenfeldt et al., 2017; Jurtz et al., 2016).

Una estrategia para mejorar la identificación de genes de resistencia es la utilización de bases de datos especializadas, como *The Comprehensive Antibiotic Resistance Database* (CARD) (Marotta et al., 2020; McArthur et al., 2013). La selección de una base de datos más amplia podría aumentar la capacidad de detección de genes de resistencia (Alcock et al., 2020).

Adicionalmente, una alternativa para optimizar la búsqueda de genes de resistencia es la implementación de algoritmos de *machine learning* o aprendizaje automático. Uno de los modelos más utilizados es Random Forest, también conocido como *bosque aleatorio*, ampliamente empleado en el ámbito de la salud debido a su flexibilidad y facilidad de implementación (Arango-Argoty et al., 2018; IBM, 2024). Otro modelo frecuentemente utilizado es **Support Vector Machine (SVM)**, el cual se basa en algoritmos de clasificación y regresión para mejorar la predicción de resultados, siendo ampliamente aplicado en el análisis biomédico (Arango-Argoty et al., 2018; *Support Vector Machine (SVM) - MATLAB & Simulink*, 2024).

Tabla 7. Identificación de los genes, nombre, unidad y número de genes dentro de *Campylobacter spp*

Gen ID	Descripción del Gen	Módulo	Submódulo	Ensamblajes
K00589	Uroporphyrin-III C-methyltransferase	Siroheme biosynthesis	Antibiotic Resistance	0
K01698	Porphobilinogen synthase	Siroheme biosynthesis	Antibiotic Resistance	1
K01719	Uroporphyrinogen-III synthase	Siroheme biosynthesis	Antibiotic Resistance	1
K01749	Hydroxymethylbilane synthase	Siroheme biosynthesis	Antibiotic Resistance	1
K01845	Glutamate-1-semialdehyde 2,1-aminomutase	Siroheme biosynthesis	Antibiotic Resistance	1
K01885	Glutamyl-tRNA synthetase	Siroheme biosynthesis	Antibiotic Resistance	2
K02302	Uroporphyrin-III C-methyltransferase	Siroheme biosynthesis	Antibiotic Resistance	0
K02303	Uroporphyrin-III C-methyltransferase	Siroheme biosynthesis	Antibiotic Resistance	0
K02304	Precorrin-2 dehydrogenase	Siroheme biosynthesis	Antibiotic Resistance	0
K02492	Glutamyl-tRNA reductase	Siroheme biosynthesis	Antibiotic Resistance	1
K02496	Uroporphyrin-III C-methyltransferase	Siroheme biosynthesis	Antibiotic Resistance	0
K03794	Sirohydrochlorin ferrochelatase	Siroheme biosynthesis	Antibiotic Resistance	0
K13542	Uroporphyrinogen-III synthase	Siroheme biosynthesis	Antibiotic Resistance	0
K13543	Uroporphyrinogen-III synthase	Siroheme biosynthesis	Antibiotic Resistance	0
K14163	Glutamyl-tRNA synthetase	Siroheme biosynthesis	Antibiotic Resistance	0
K11629	Bacitracin resistance sensor kinase BceS	Bacitracin resistance	BceAB transporter	0
K11630	Bacitracin resistance response regulator BceR	Bacitracin resistance	BceAB transporter	0
K11631	Bacitracin transport ATP-binding protein BceA	Bacitracin resistance	BceAB transporter	0
K11632	Bacitracin transport permease protein BceB	Bacitracin resistance	BceAB transporter	0
K11631	Bacitracin transport ATP-binding protein BceA	Bacitracin resistance	VraDE transporter	0

<b>K11632</b>	Bacitracin transport permease protein BceB	Bacitracin resistance	VraDE transporter	0
<b>K19081</b>	BraS; Bacitracin sensor histidine kinase	Tetracycline resistance	TetA transporter	0
<b>K19082</b>	BraR; Bacitracin response regulator	Tetracycline resistance	TetA transporter	0
<b>K19083</b>	BraD; Bacitracin ATP-binding protein	Tetracycline resistance	TetAB(46) transporter	0
<b>K19084</b>	BraE; Bacitracin permease protein	Tetracycline resistance	TetAB(46) transporter	0

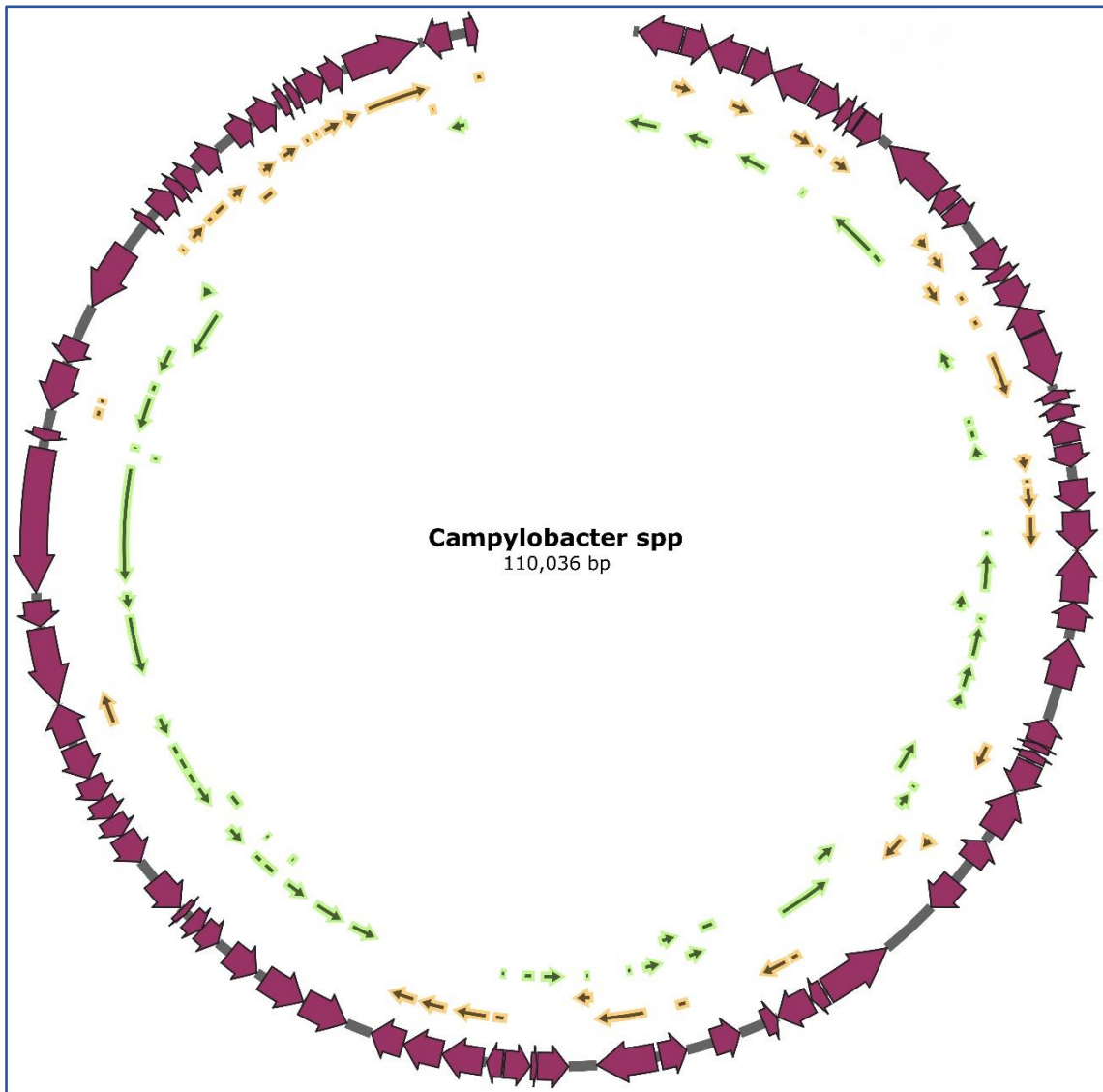


Figura 35. Mapa generado a partir del resultado obtenido en Annotate and Distill Assemblies with DRAM

En la figura 35 se presentan de manera detallada los genes obtenidos tras ensamblar el genoma completo de *Campylobacter spp.*, luego de llevar a cabo la depuración del metagenoma, el análisis correspondiente y la generación de rutas metabólicas y genes de resistencia. Las flechas en color morado señalan cada gen e indican su orientación de expresión. Por otro lado, las flechas amarillas y verdes evidencian el sentido en que se expresan dichos genes, mostrando si la transcripción se realiza en dirección positiva o negativa.

## **10. Conclusiones y recomendaciones**

### **10.1. Evaluación de la calidad y filtrado de datos**

Durante la fase de control de calidad se comprobó que la mayoría de las secuencias mantenía un nivel satisfactorio, aunque se evidenció un leve descenso en los tramos finales, posiblemente atribuido a la pérdida de intensidad de la señal al final de la secuenciación. En total, se evaluaron 20,327,027 pares de bases. Posteriormente, se eliminaron las secuencias con baja calidad y los adaptadores, conservándose 19,279,737 pares; en consecuencia, se descartaron alrededor de 1,047,290 pares, lo que equivale aproximadamente al 5% del total.

### **10.2. Eliminación del genoma del hospedero**

Con las secuencias depuradas se buscó y eliminó el genoma del hospedero usando *Bowtie2* y el genoma de referencia de *H. illucens* (GCF\_905115235.1), identificándose que el 56.34% de las lecturas correspondían al hospedero. Posteriormente, mediante *SamTools* y *bedtools* se filtraron las secuencias no alineadas, obteniéndose 16,836,004 pares de bases y generándose dos archivos FASTQ (PLA\_F.fastq para la secuencia forward y PLA\_R.fastq para la reverse) limpios de adaptadores y secuencias del hospedero.

### **10.3. Clasificación taxonómica y ensamblaje**

Utilizando *Kaiju* en KBase se realizó la clasificación taxonómica, detectándose la presencia de Proteobacterias, especialmente *Gammaproteobacteria*, y de *Campylobacter* a nivel de género (sin asignación a especie). Ante esta verificación, se evaluó el ensamblaje óptimo, identificándose a *metaSPAdes* como el mejor, con resultados que incluyen contigs de hasta 965,405 pb, un N50 de 30,896 pb y el contig más largo de 11,106,106 pb. Los contigs se agruparon en 11 *bins* de alta calidad; en particular, el

bin.003, asignado a *Campylobacter*, mostró una integridad del 90.52%, contaminación del 0.86% y 445 genes, lo que confirma la calidad del ensamblaje.

#### 10.4. **Confirmación taxonómica y perfil funcional**

Con el uso de GTDB-TK en la plataforma KBase, se determinó que el bin.003 se clasifica dentro del dominio Bacteria, filo Campylobacterota, clase Campylobacteria, orden Campylobacterales y familia Campylobacteraceae. Se obtuvo un genoma con una extensión de 1,709,401 pb, lo que confirma su presencia en la muestra.

El perfil funcional del genoma ensamblado arrojó 60 *scaffolds* y 37 genes de tRNA, lo que sugiere una estructura robusta para análisis posteriores. El mapa de calor indicó la presencia de los complejos ETC del I al V, con algunos módulos completos (destacados en azul), esenciales para la generación de ATP y adaptación a condiciones microaerófilas, característica de *Campylobacter spp.* Además, se identificaron rutas metabólicas adicionales (CAZy, y metabolismo del nitrógeno y del azufre) que facilitan la colonización y supervivencia en ambientes hostiles. En cuanto a resistencia, se detectaron genes asociados a la biosíntesis de siroheme y mecanismos de resistencia a bacitracina y tetraciclina; sin embargo, el análisis en el “Center for Genomic Epidemiology” no evidenció la presencia de genes de resistencia.

## Bibliografía

- Aguilar-Bultet Lisandra, & Falquet Laurent. (2015). *Secuenciación y ensamblaje de novo de genomas bacterianos: una alternativa para el estudio de nuevos patógenos*. Departamento de Microbiología, Dirección de Salud Animal, Centro Nacional de Sanidad Agropecuaria (CENSA).  
[http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0253-570X2015000200008](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0253-570X2015000200008)
- Ahrenfeldt, J., Skaarup, C., Hasman, H., Pedersen, A. G., Aarestrup, F. M., & Lund, O. (2017). Bacterial whole genome-based phylogeny: construction of a new benchmarking dataset and assessment of some existing methods. *BMC genomics*, *18*(1). <https://doi.org/10.1186/S12864-016-3407-6>
- Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A. L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H. K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., ... McArthur, A. G. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic acids research*, *48*(D1), D517–D525. <https://doi.org/10.1093/NAR/GKZ935>
- Alfredson, D. A., & Korolik, V. (2005). Isolation and expression of a novel molecular class D  $\beta$ -lactamase, OXA-61, from *Campylobacter jejuni*. *Antimicrobial Agents and Chemotherapy*, *49*(6), 2515–2518. <https://doi.org/10.1128/AAC.49.6.2515-2518.2005>
- Alonso, M., Rojas, B., Andres, S., & Cortes, Q. (2023). *Campylobacter jejuni*, a current problem in Public Health. • *Med*, *45*(3), 472–483. <https://doi.org/10.56050/01205498.2274>

Amazon Web Services Inc. (2023). *Contenedores de Docker | ¿Qué es Docker? | AWS*.  
<https://aws.amazon.com/es/docker/>

Arango-Argoty, G., Garner, E., Pruden, A., Heath, L. S., Vikesland, P., & Zhang, L. (2018). DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome*, 6(1). <https://doi.org/10.1186/S40168-018-0401-Z>

Arquivo. (2016). *Bioconversion of Food Waste: Black Soldier Fly*.  
<https://arquivo.pt/wayback/20160516163727/http://www.esrint.com/pages/bioconversion.html>

Audu, B. J., Norval, S., Bruno, L., Meenakshi, R., Marion, M., & Forbes, K. J. (2022). Genomic diversity and antimicrobial resistance of *Campylobacter* spp. from humans and livestock in Nigeria. *Journal of Biomedical Science*, 29(1), 1–15.  
<https://doi.org/10.1186/S12929-022-00786-2/TABLES/4>

Babraham Institute. (2024). *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*.  
<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Belli, W. A., & Fryklund, J. (1995). Partial Characterization and Effect of Omeprazole on ATPase Activity in *Helicobacter pylori* by Using Permeabilized Cells. En *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY* (Vol. 39, Número 8).  
<https://journals.asm.org/journal/aac>

Bikel, S., Valdez-Lara, A., Cornejo-Granados, F., Rico, K., Canizales-Quinteros, S., Soberón, X., Del Pozo-Yauner, L., & Ochoa-Leyva, A. (2015). Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: Towards a systems-level understanding of human microbiome. En

*Computational and Structural Biotechnology Journal* (Vol. 13, pp. 390–401).

Elsevier B.V. <https://doi.org/10.1016/j.csbj.2015.06.001>

Blake, R. E., Chang, S. J., & Lepland, A. (2010). Phosphate oxygen isotopic evidence for a temperate and biologically active Archaean ocean. *Nature*, *464*(7291), 1029–1032. <https://doi.org/10.1038/nature08952>

Bolger, A. (2024). *Trimmomatic: A flexible read trimming tool for Illumina NGS data*. <http://www.usadellab.org/cms/?page=trimmomatic>

Bolger, A., Lohse, M., & Usadel, B. (2024). *USADELLAB.org - Trimmomatic: A flexible read trimming tool for Illumina NGS data*. <http://www.usadellab.org/cms/?page=trimmomatic>

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>

Bortolini, S., Macavei, L. I., Hadj Saadoun, J., Foca, G., Ulrici, A., Bernini, F., Malferrari, D., Setti, L., Ronga, D., & Maistrello, L. (2020). *Hermetia illucens* (L.) larvae as chicken manure management tool for circular economy. *Journal of Cleaner Production*, *262*. <https://doi.org/10.1016/J.JCLEPRO.2020.121289>

Brown, J., Pirrung, M., & Mccue, L. A. (2017). FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*, *33*(19), 3137–3139. <https://doi.org/10.1093/BIOINFORMATICS/BTX373>

Bunduruş, I. A., Balta, I., Ştef, L., Ahmadi, M., Peţ, I., McCleery, D., & Corcionivoschi, N. (2023). Overview of Virulence and Antibiotic Resistance in *Campylobacter* spp.

<https://doi.org/10.3390/ANTIBIOTICS12020402>

Bush, L., & Charles, E. (2022, septiembre). *Infecciones por Campylobacter - Infecciones - Manual MSD versión para público general*. College of Medicine, Florida Atlantic University. <https://www.msmanuals.com/es-ec/hogar/infecciones/infecciones-bacterianas-bacterias-gramnegativas/infecciones-por-campylobacter>

Casagrande Proietti, P., Guelfi, G., Bellucci, S., De Luca, S., Di Gregorio, S., Pieramati, C., & Franciosini, M. P. (2020). Beta-lactam resistance in *Campylobacter coli* and *Campylobacter jejuni* chicken isolates and the association between blaOXA-61 gene expression and the action of  $\beta$ -lactamase inhibitors. *Veterinary microbiology*, 241. <https://doi.org/10.1016/J.VETMIC.2019.108553>

Casariago, J. (2020). *Identificación molecular de la diversidad bacteriana intestinal de Hermetia illucens (Stratiomyidae-Diptera) mediante técnicas independientes, dependientes y espectrometría de masas Maldi tof/tof*.

Center for Genomic Epidemiology. (2024). *Center for Genomic Epidemiology*. <https://www.genomicepidemiology.org/services/>

Chen, Y., Mukherjee, S., Hoffmann, M., Kotewicz, M. L., Young, S., Abbott, J., Luo, Y., Davidson, M. K., Allard, M., McDermott, P., & Zhao, S. (2013). Whole-genome sequencing of gentamicin-resistant *Campylobacter coli* isolated from U.S. retail meats reveals novel plasmid-mediated aminoglycoside resistance genes. *Antimicrobial Agents and Chemotherapy*, 57(11), 5398–5405. <https://doi.org/10.1128/AAC.00669-13/ASSET/CBEE96C6-934A-4CF7-8A0D-8F1C25C9D0C7/ASSETS/GRAPHIC/ZAC0111322790003.JPEG>

Chivian, D., Jungbluth, S. P., Dehal, P. S., Wood-Charlson, E. M., Canon, R. S., Allen, B. H., Clark, M. M., Gu, T., Land, M. L., Price, G. A., Riehl, W. J., Sneddon, M. W., Sutormin, R., Zhang, Q., Cottingham, R. W., Henry, C. S., & Arkin, A. P. (2023a). Metagenome-assembled genome extraction and analysis from microbiomes using KBase. *Nature Protocols*, *18*(1), 208–238. <https://doi.org/10.1038/s41596-022-00747-x>

Chivian, D., Jungbluth, S. P., Dehal, P. S., Wood-Charlson, E. M., Canon, R. S., Allen, B. H., Clark, M. M., Gu, T., Land, M. L., Price, G. A., Riehl, W. J., Sneddon, M. W., Sutormin, R., Zhang, Q., Cottingham, R. W., Henry, C. S., & Arkin, A. P. (2023b). Metagenome-assembled genome extraction and analysis from microbiomes using KBase. *Nature Protocols*, *18*(1), 208–238. <https://doi.org/10.1038/s41596-022-00747-x>

Connell, S. R., Trieber, C. A., Dinos, G. P., Einfeldt, E., Taylor, D. E., & Nierhaus, K. H. (2003). Mechanism of Tet(O)-mediated tetracycline resistance. *The EMBO Journal*, *22*(4), 945. <https://doi.org/10.1093/EMBOJ/CDG093>

Cuadrado, D. F., Domínguez Rodríguez, L., Concepción, M., Calonge, P., & Molina, A. Q. (2018). *UNIVERSIDAD COMPLUTENSE DE MADRID FACULTAD DE VETERINARIA DEPARTAMENTO DE SANIDAD ANIMAL TESIS DOCTORAL Caracterización genómica y resistencia a antimicrobianos de Campylobacter MEMORIA PARA OPTAR AL GRADO DE DOCTOR PRESENTADA POR.*

De Ciências, F., De, F., & Preto, R. (2019). *UNIVERSIDADE DE SÃO PAULO.*

Dedieu, L., Brunel, J. M., Lorenzi, V., Muselli, A., Berti, L., & Bolla, J. M. (2020). Antibacterial mode of action of the daucus carota essential oil active compounds

- against campylobacter jejuni and efflux-mediated drug resistance in gram-negative bacteria. *Molecules*, 25(22). <https://doi.org/10.3390/MOLECULES25225448>
- el Sitio Avicola. (2012). *Técnicas de detección de Campylobacter*. <https://www.elsitioavicola.com/articles/2269/tacnicas-de-deteccion-de-campylobacter/>
- ELIKA. (2022). *Campylobacter*. <https://seguridadalimentaria.elika.eus/fichas-de-peligros/campylobacter/>
- ELIKA. (2023). *Campylobacter - ELIKA Seguridad Alimentaria*. ELIKA Seguridad Alimentaria. <https://seguridadalimentaria.elika.eus/fichas-de-peligros/campylobacter/>
- El-Naenaeey, E., Abd El-Hamid, M., & Khalifa, E. (2020). Foodborne Campylobacter species: Taxonomy, Isolation, Virulence Attributes and Antimicrobial Resistance. *Zagazig Veterinary Journal*, 48(4), 414–432. <https://doi.org/10.21608/zvzj.2020.40250.1118>
- Fabre, A., Oleastro, M., Nunes, A., Santos, A., Sifre, E., Benejat, L., Buissonnière, A., Floch, P., Mégraud, F., & Dubois, V. (2018a). Whole-Genome Sequence Analysis of Multidrug-Resistant Campylobacter Isolates: a Focus on Aminoglycoside Resistance Determinants.. Whole-Genome Sequence Analysis of Multidrug-Resistant Campylobacter Isolates: a Focus on Amino-glycoside Resistance Determinants. *Journal of Clinical Microbiology*, 56(9), 390–408. <https://doi.org/10.1128/JCM.00390-18i>
- Fabre, A., Oleastro, M., Nunes, A., Santos, A., Sifre, E., Benejat, L., Buissonnière, A., Floch, P., Mégraud, F., & Dubois, V. (2018b). Whole-Genome Sequence Analysis of Multidrug-Resistant Campylobacter Isolates: a Focus on Aminoglycoside

Resistance Determinants.. Whole-Genome Sequence Analysis of Multidrug-Resistant *Campylobacter* Isolates: a Focus on Amino-glycoside Resistance Determinants. *Journal of Clinical Microbiology*, 56(9), 390–408. <https://doi.org/10.1128/JCM.00390-18>

Fernández, H. (2011). *Campylobacter* Y CAMPYLOBACTERIOSIS: UNA MIRADA DESDE AMÉRICA DEL SUR *Campylobacter* AND CAMPYLOBACTERIOSIS: A VIEW FROM SOUTH AMERICA. *Rev Peru Med Exp Salud Publica*, 28(1), 121–148.

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>

G Del Hierro, A., Anrango, M. J., Ortiz, D., & Sánchez, L. (2021). Captura y cría de la mosca soldado negra (*Hermetia Illucens*) para la biodegradación de desechos orgánicos en Puerto Quito, Ecuador. *Ecuadorian Science Journal*, 5(3), 341–354. <https://doi.org/10.46480/esj.5.3.164>

Generalovic, T. N., McCarthy, S. A., Warren, I. A., Wood, J. M. D., Torrance, J., Sims, Y., Quail, M., Howe, K., Papan, M., Durbin, R., & Jiggins, C. D. (2021). A high-quality, chromosome-level genome assembly of the Black Soldier Fly (*Hermetia illucens* L.). *G3: Genes|Genomes|Genetics*, 11(5). <https://doi.org/10.1093/G3JOURNAL/JKAB085>

Givanoudi, S., Cornelis, P., Rasschaert, G., Wackers, G., Iken, H., Rolka, D., Yongabi, D., Robbens, J., Schöning, M. J., Heyndrickx, M., & Wagner, P. (2021). *Selective Campylobacter detection and quantification in poultry: A sensor tool for detecting*

*the cause of a common zoonosis at its source.*

<https://doi.org/10.1016/j.snb.2021.129484>

Guirado, P., Miró, E., Iglesias-torrens, Y., Navarro, F., Campoy, S., Alioto, T. S., Gómez-garrido, J., Madrid, C., & Balsalobre, C. (2022). A New Variant of the aadE-sat4-aphA-3 Gene Cluster Found in a Conjugative Plasmid from a MDR *Campylobacter jejuni* Isolate. *Antibiotics (Basel, Switzerland)*, *11*(4).  
<https://doi.org/10.3390/ANTIBIOTICS11040466>

Habib, I., Ibrahim Mohamed, M. Y., Ghazawi, A., Lakshmi, G. B., Khan, M., Li, D., & Sahibzada, S. (2023). Genomic characterization of molecular markers associated with antimicrobial resistance and virulence of the prevalent *Campylobacter coli* isolated from retail chicken meat in the United Arab Emirates. *Current Research in Food Science*, *6*. <https://doi.org/10.1016/j.crf.2023.100434>

Hernández, M., Quijada, N. M., Rodríguez-Lázaro, D., & Eiros, J. M. (2020a). Aplicación de la secuenciación masiva y la bioinformática al diagnóstico microbiológico clínico. *Revista Argentina de Microbiología*, *52*(2), 150–161.  
<https://doi.org/10.1016/J.RAM.2019.06.003>

Hernández, M., Quijada, N. M., Rodríguez-Lázaro, D., & Eiros, J. M. (2020b). Aplicación de la secuenciación masiva y la bioinformática al diagnóstico microbiológico clínico. *Revista Argentina de Microbiología*, *52*(2), 150–161.  
<https://doi.org/10.1016/J.RAM.2019.06.003>

Hoffmant, P. S., & Goodman, T. G. (1982). Respiratory Physiology and Energy Conservation Efficiency of *Campylobacter jejuni*. En *JOURNAL OF BACTERIOLOGY* (Vol. 150, Número 1). <https://journals.asm.org/journal/jb>

- Hofreuter, D. (2014). Defining the metabolic requirements for the growth and colonization capacity of *Campylobacter jejuni*. *Frontiers in cellular and infection microbiology*, 4(SEP). <https://doi.org/10.3389/FCIMB.2014.00137>
- Hormeño, L., Ugarte-Ruiz, M., Palomo, G., Borge, C., Florez-Cuadrado, D., Vadillo, S., Píriz, S., Domínguez, L., Campos, M. J., & Quesada, A. (2018a). Ant(6)-I genes encoding aminoglycoside O-nucleotidyltransferases are widely spread Among streptomycin resistant strains of *Campylobacter jejuni* and *Campylobacter coli*. *Frontiers in Microbiology*, 9(OCT), 401706. <https://doi.org/10.3389/FMICB.2018.02515/BIBTEX>
- Hormeño, L., Ugarte-Ruiz, M., Palomo, G., Borge, C., Florez-Cuadrado, D., Vadillo, S., Píriz, S., Domínguez, L., Campos, M. J., & Quesada, A. (2018b). ant(6)-I Genes Encoding Aminoglycoside O-Nucleotidyltransferases Are Widely Spread Among Streptomycin Resistant Strains of *Campylobacter jejuni* and *Campylobacter coli*. *Frontiers in microbiology*, 9(OCT). <https://doi.org/10.3389/FMICB.2018.02515>
- Huang, S., & Millar, A. H. (2013). Succinate dehydrogenase: the complex roles of a simple enzyme. *Current Opinion in Plant Biology*, 16(3), 344–349. <https://doi.org/10.1016/J.PBI.2013.02.007>
- Hull, D. M., Harrel, E., Harden, L., & Thakur, S. (2023). Detection of resistance and virulence plasmids in *Campylobacter coli* and *Campylobacter jejuni* isolated from North Carolina food animal production, 2018–2019. *Food Microbiology*, 116, 104348. <https://doi.org/10.1016/J.FM.2023.104348>
- IBM. (2024). *¿Qué es Random Forest?* | IBM. <https://www.ibm.com/mx-es/topics/random-forest>

- INABIO. (2016). *Mosca soldado negra (Hermetia illucens)*. iNaturalist Ecuador.  
<https://ecuador.inaturalist.org/taxa/82177-Hermetia-illucens>
- Jackson, R. J., Elvers, K. T., Lee, L. J., Gidley, M. D., Wainwright, L. M., Lightfoot, J., Park, S. F., & Poole, R. K. (2007). Oxygen reactivity of both respiratory oxidases in *Campylobacter jejuni*: The *cydAB* genes encode a cyanide-resistant, low-affinity oxidase that is not of the cytochrome *bd* type. *Journal of Bacteriology*, *189*(5), 1604–1615. <https://doi.org/10.1128/JB.00897-06>
- Jacob, J., Evers, S., Bischoff, K., Carlier, C., & Courvalin, P. (1994). Characterization of the *sat4* gene encoding a streptothricin acetyltransferase in *Campylobacter coli* BE/G4. *FEMS Microbiology Letters*, *120*(1–2), 13–17.  
<https://doi.org/10.1111/j.1574-6968.1994.tb07000.x>
- Joshi, C. J., Ke, W., Drangowska-Way, A., O'Rourke, E. J., & Lewis, N. E. (2022). What are housekeeping genes? *PLoS Computational Biology*, *18*(7).  
<https://doi.org/10.1371/JOURNAL.PCBI.1010295>
- Jurtz, V. I., Villarroel, J., Lund, O., Voldby Larsen, M., & Nielsen, M. (2016). MetaPhinder-Identifying Bacteriophage Sequences in Metagenomic Data Sets. *PloS one*, *11*(9). <https://doi.org/10.1371/JOURNAL.PONE.0163111>
- Kang, J., Zhang, C., Wan, S., Li, W., Zhao, W., Li, L., Shang, Y., Du, X. D., Liu, D., & Yao, H. (2024). Prevalence and characterization of aminoglycoside resistance gene *aph(2'')*-If-carrying *Campylobacter jejuni*. *International Journal of Food Microbiology*, *419*, 110747.  
<https://doi.org/10.1016/J.IJFOODMICRO.2024.110747>

- Kelly, D. J. (2014). Complexity and Versatility in the Physiology and Metabolism of *Campylobacter jejuni*. *Campylobacter*, 41–61. <https://doi.org/10.1128/9781555815554.CH3>
- Khamis, F. M., Ombura, F. L. O., Akutse, K. S., Subramanian, S., Mohamed, S. A., Fiaboe, K. K. M., Saijuntha, W., Van Loon, J. J. A., Dicke, M., Dubois, T., Ekesi, S., & Tanga, C. M. (2020). Insights in the Global Genetics and Gut Microbiome of Black Soldier Fly, *Hermetia illucens*: Implications for Animal Feed Safety Control. *Frontiers in Microbiology*, 11, 548711. <https://doi.org/10.3389/FMICB.2020.01538/BIBTEX>
- Klotz, M. G., Hemp, J., Hanson, T. E., Wösten, M. M. S. M., Ni, W., & Van Der Stel, A.-X. (2019). Regulation of Respiratory Pathways in Campylobacterota: A Review. *Frontiers in Microbiology*, 10, 424488. <https://doi.org/10.3389/FMICB.2019.01719>
- Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., & Hugenholtz, P. (2008). A Bioinformatician's Guide to Metagenomics. *Microbiology and Molecular Biology Reviews*, 72(4), 557–578. <https://doi.org/10.1128/membr.00009-08>
- 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. <https://doi.org/10.1016/J.IJFOODMICRO.2023.110098>
- Liao, Y. S., Chen, B. H., Teng, R. H., Wang, Y. W., Chang, J. H., Liang, S. Y., Tsao, C. Sen, Hong, Y. P., Sung, H. Y., & Chiou, C. S. (2022). Antimicrobial Resistance in *Campylobacter coli* and *Campylobacter jejuni* from Human Campylobacteriosis in Taiwan, 2016 to 2019. *Antimicrobial Agents and Chemotherapy*, 66(1). <https://doi.org/10.1128/AAC.01736-21>

- Lis Ricardi, L. (2022). *Ensamblado y análisis comparativo de metagenomas de rumen vacuno*.
- Lv, R., Wang, K., Feng, J., Heeney, D. D., Liu, D., & Lu, X. (2020a). Detection and Quantification of Viable but Non-culturable *Campylobacter jejuni*. *Frontiers in Microbiology*, *10*, 2920. <https://doi.org/10.3389/FMICB.2019.02920/BIBTEX>
- Lv, R., Wang, K., Feng, J., Heeney, D. D., Liu, D., & Lu, X. (2020b). Detection and Quantification of Viable but Non-culturable *Campylobacter jejuni*. *Frontiers in Microbiology*, *10*, 2920. <https://doi.org/10.3389/FMICB.2019.02920/BIBTEX>
- Mainetti, J. A. (2003). BIOÉTICA Y GENÓMICA. En *Acta Bioethica* (Número 1).
- Mardis, E. R. (2013). Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry*, *6*(1), 287–303. <https://doi.org/10.1146/annurev-anchem-062012-092628>
- Marotta, F., Di Marcantonio, L., Janowicz, A., Pedonese, F., Di Donato, G., Ardelean, A., Nuvoloni, R., Di Giannatale, E., & Garofolo, G. (2020). Genotyping and Antibiotic Resistance Traits in *Campylobacter jejuni* and *coli* From Pigs and Wild Boars in Italy. *Frontiers in Cellular and Infection Microbiology*, *10*, 592512. <https://doi.org/10.3389/FCIMB.2020.592512/BIBTEX>
- Mathai, N., Chen, Y., & Kirchmair, J. (2020). Validation strategies for target prediction methods. *Briefings in Bioinformatics*, *21*(3), 791–802. <https://doi.org/10.1093/bib/bbz026>
- McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., Bhullar, K., Canova, M. J., De Pascale, G., Ejim, L., Kalan, L., King, A. M., Koteva, K., Morar, M., Mulvey, M. R., O'Brien, J. S., Pawlowski, A. C., Piddock, L. J. V.,

- Spanogiannopoulos, P., ... Wright, G. D. (2013). The Comprehensive Antibiotic Resistance Database. *Antimicrobial Agents and Chemotherapy*, 57(7), 3348. <https://doi.org/10.1128/AAC.00419-13>
- Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, 7. <https://doi.org/10.1038/ncomms11257>
- Metzker, M. L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11(1), 31–46. <https://doi.org/10.1038/nrg2626>
- Myers, J. D., & Kelly, D. J. (2005). A sulphite respiration system in the chemoheterotrophic human pathogen *Campylobacter jejuni*. *Microbiology*, 151(1), 233–242. <https://doi.org/10.1099/mic.0.27573-0>
- National Center for Biotechnology Information. (2024). *Hermetia illucens genome assembly iHerIII2.2.curated.20191125* - NCBI - NLM. [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_905115235.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_905115235.1/)
- National Human Genome Research Institute. (2024a). *Genómica*. <https://www.genome.gov/es/genetics-glossary/Genomica>
- National Human Genome Research Institute. (2024b). *Metagenómica*. <https://www.genome.gov/es/genetics-glossary/Metagenomica>
- NCBI. (2024). *Hermetia illucens genome assembly iHerIII2.2.curated.20191125* - NCBI - NLM. [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_905115235.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_905115235.1/)
- Nesla Olivas. (2007). *Presencia de Campylobacter en el norte de la provincia de Esmeraldas*. <https://repositorio.usfq.edu.ec/handle/23000/489>

- Ocejo, M., Oporto, B., Lavín, J. L., & Hurtado, A. (2021). Whole genome-based characterisation of antimicrobial resistance and genetic diversity in *Campylobacter jejuni* and *Campylobacter coli* from ruminants. *Scientific reports*, *11*(1), 8998. <https://doi.org/10.1038/s41598-021-88318-0>
- OMS. (2020). *Campylobacter*. <https://www.who.int/es/news-room/fact-sheets/detail/campylobacter>
- Pazmiño, M. F., Del Hierro, A. G., & Flores, F. J. (2023). Genetic diversity and organic waste degrading capacity of *Hermetia illucens* from the evergreen forest of the Equatorial Choco lowland. *PeerJ*, *11*. <https://doi.org/10.7717/peerj.14798>
- Peabody, M. A., Van Rossum, T., Lo, R., & Brinkman, F. S. L. (2015). Evaluation of shotgun metagenomics sequence classification methods using in silico and in vitro simulated communities. *BMC Bioinformatics*, *16*(1). <https://doi.org/10.1186/S12859-015-0788-5>
- Pinto-Alphandary, H., Mabilat, C., & Courvalin, P. (1990). Emergence of aminoglycoside resistance genes *aadA* and *aadE* in the genus *Campylobacter*. *Antimicrobial Agents and Chemotherapy*, *34*(6), 1294–1296. <https://doi.org/10.1128/AAC.34.6.1294>
- Quino, W., Caro-Castro, J., Hurtado, V., Flores-León, D., Gonzalez-Escalona, N., & Gavilan, R. G. (2022a). Genomic Analysis and Antimicrobial Resistance of *Campylobacter jejuni* and *Campylobacter coli* in Peru. *Frontiers in Microbiology*, *12*, 802404. <https://doi.org/10.3389/FMICB.2021.802404/BIBTEX>
- Quino, W., Caro-Castro, J., Hurtado, V., Flores-León, D., Gonzalez-Escalona, N., & Gavilan, R. G. (2022b). Genomic Analysis and Antimicrobial Resistance of *Campylobacter jejuni* and *Campylobacter coli* in Peru. *Frontiers in Microbiology*, *12*. <https://doi.org/10.3389/fmicb.2021.802404>

- Rahbari, M., Rahlfs, S., Jortzik, E., Bogeski, I., & Becker, K. (2017). H<sub>2</sub>O<sub>2</sub> dynamics in the malaria parasite *Plasmodium falciparum*. *PLoS ONE*, *12*(4).  
<https://doi.org/10.1371/journal>
- Reátegui, J., Barriga, X., Obando, A., Moscoso, G., Manrique, P., Salazar, I., Reátegui, J., Barriga, X., Obando, A., Moscoso, G., Manrique, P., & Salazar, I. (2020). Harina de larva de *Hermetia illucens* (Diptera: Stratiomyidae) como ingrediente proteico de reemplazo parcial de harina de soja en la alimentación de *Cavia porcellus* (Cuy): efecto en el consumo, ganancia de peso y conversión alimenticia. *Scientia Agropecuaria*, *11*(4), 513–519.  
<https://doi.org/10.17268/SCI.AGROPECU.2020.04.06>
- Rodríguez-Santiago, B., & Armengol, L. (2012). Tecnologías de secuenciación de nueva generación en diagnóstico genético pre- y postnatal. *Diagnostico Prenatal*, *23*(2), 56–66. <https://doi.org/10.1016/j.diapre.2012.02.001>
- Rubio, S., Pacheco-Orozco, R. A., Gómez, A. M., Perdomo, S., & García-Robles, R. (2020). Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica. *Universitas Médica*, *61*(2).  
<https://doi.org/10.11144/javeriana.umed61-2.sngs>
- Rupinder Kaur Kanwar, Ganguly, N. K., Kanwar, J. R., Lata Kumar, & Walia, B. N. S. (1994). Impairment of Na<sup>+</sup>,K<sup>+</sup>-ATPase activity following enterotoxigenic *Campylobacter jejuni* infection: changes in Na<sup>+</sup>, Cl<sup>-</sup> and 3-O-methyl-D-glucose transport in vitro, in rat ileum. *FEMS Microbiology Letters*, *124*(3), 381–385.  
<https://doi.org/10.1111/j.1574-6968.1994.tb07312.x>
- Sellars, M. J., Hall, S. J., & Kelly, D. J. (2002a). Growth of *Campylobacter jejuni* supported by respiration of fumarate, nitrate, nitrite, trimethylamine-N-oxide, or

dimethyl sulfoxide requires oxygen. *Journal of Bacteriology*, 184(15), 4187–4196.

<https://doi.org/10.1128/JB.184.15.4187-4196.2002>

Sellars, M. J., Hall, S. J., & Kelly, D. J. (2002b). Growth of *Campylobacter jejuni* supported by respiration of fumarate, nitrate, nitrite, trimethylamine-N-oxide, or dimethyl sulfoxide requires oxygen. *Journal of bacteriology*, 184(15), 4187–4196.

<https://doi.org/10.1128/JB.184.15.4187-4196.2002>

Shaffer, M., Borton, M. A., Bolduc, B., Faria, J. P., Flynn, R. M., Ghadermazi, P., Edirisinghe, J. N., Wood-Charlson, E. M., Miller, C. S., Chan, S. H. J., Sullivan, M. B., Henry, C. S., & Wrighton, K. C. (2023). kb\_DRAM: annotation and metabolic profiling of genomes with DRAM in KBase. *Bioinformatics*, 39(4).

<https://doi.org/10.1093/BIOINFORMATICS/BTAD110>

Simaluiza, R., Toledo, Z., Fernández, H., Simaluiza, R., Toledo, Z., & Fernández, H. (2018). Prevalencia y caracterización del perfil de susceptibilidad antimicrobiana de *Campylobacter jejuni* y *Campylobacter coli* en niños con diarrea de la ciudad de Loja, Ecuador. *Revista chilena de infectología*, 35(2), 213–215.

<https://doi.org/10.4067/S0716-10182018000200213>

Smith, C. A., Bhattacharya, M., Toth, M., Stewart, N. K., & Vakulenko, S. B. (2017). Aminoglycoside resistance profile and structural architecture of the aminoglycoside acetyltransferase AAC(6′)-Im. *Microbial Cell*, 4(12), 402.

<https://doi.org/10.15698/MIC2017.12.602>

Smith, M. A., Finel, M., Korolik, V., & Mendz, G. L. (2000). Characteristics of the aerobic respiratory chains of the microaerophiles *Campylobacter jejuni* and *Helicobacter pylori*. En *Archives of Microbiology* (Vol. 174, Números 1–2, pp. 1–10).

<https://doi.org/10.1007/s002030000174>

*Support Vector Machine (SVM) - MATLAB & Simulink.* (2024).

<https://la.mathworks.com/discovery/support-vector-machine.html>

Tanga, C. M., Waweru, J. W., Tola, Y. H., Onyoni, A. A., Khamis, F. M., Ekesi, S., &

Paredes, J. C. (2021). Organic Waste Substrates Induce Important Shifts in Gut

Microbiota of Black Soldier Fly (*Hermetia illucens* L.): Coexistence of Conserved,

Variable, and Potential Pathogenic Microbes. *Frontiers in Microbiology*, *12*, 24.

<https://doi.org/10.3389/FMICB.2021.635881/BIBTEX>

Terzolo, H. R., Del, M., & Catena, C. (2007). *Microbiología Veterinaria. Capítulo 34:*

*Campylobacter.* <https://www.researchgate.net/publication/281441926>

US. Department of ENERGY. (2024). *KBase.* <https://narrative.kbase.us/narratives>

van der Stel, A.-X., & Wösten, M. M. S. M. (2019). Regulation of Respiratory Pathways

in *Campylobacterota*: A Review. *Frontiers in microbiology*, *10*.

<https://doi.org/10.3389/FMICB.2019.01719>

Vollmers, J., Wiegand, S., & Kaster, A.-K. (2017). Comparing and Evaluating

Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size

Matters! *PLOS ONE*, *12*(1), e0169662.

<https://doi.org/10.1371/journal.pone.0169662>

Walt, H. K., Kooienga, E., Cammack, J. A., Tomberlin, J. K., Jordan, H. R., Meyer, F., &

Hoffmann, F. G. (2023). Bioinformatic Surveillance Leads to Discovery of Two

Novel Putative Bunyaviruses Associated with Black Soldier Fly. *Viruses*, *15*(8).

<https://doi.org/10.3390/V15081654/S1>

Wassenaar, T. M. (1997). *Toxin Production by Campylobacter spp* (Vol. 10, Número 3).

- Weerakoon, D. R., & Olson, J. W. (2008). The *Campylobacter jejuni* NADH:Ubiquinone Oxidoreductase (Complex I) Utilizes Flavodoxin Rather than NADH. *Journal of Bacteriology*, *190*(3), 915. <https://doi.org/10.1128/JB.01647-07>
- Weingarten, R. A., Grimes, J. L., & Olson, J. W. (2008). Role of *Campylobacter jejuni* respiratory oxidases and reductases in host colonization. *Applied and Environmental Microbiology*, *74*(5), 1367–1375. <https://doi.org/10.1128/AEM.02261-07>
- Weingarten, R. A., Taveirne, M. E., & Olson, J. W. (2009). The Dual-Functioning Fumarate Reductase Is the Sole Succinate:Quinone Reductase in *Campylobacter jejuni* and Is Required for Full Host Colonization. *Journal of Bacteriology*, *191*(16), 5293. <https://doi.org/10.1128/JB.00166-09>
- Woegebauer, M., Zeinzinger, J., Springer, B., Hufnagl, P., Indra, A., Korschineck, I., Hofrichter, J., Kopacka, I., Fuchs, R., Steinwider, J., Fuchs, K., Nielsen, K. M., & Allerberger, F. (2014). Prevalence of the aminoglycoside phosphotransferase genes *aph*(3′)-IIIa and *aph*(3′)-IIa in *Escherichia coli*, *Enterococcus faecalis*, *Enterococcus faecium*, *Pseudomonas aeruginosa*, *Salmonella enterica* subsp. *enterica* and *Staphylococcus aureus* isolates in Austria. *Journal of medical microbiology*, *63*(Pt 2), 210–217. <https://doi.org/10.1099/JMM.0.065789-0>
- Yagi, T., Seo, B. B., Nakamaru-Ogiso, E., Marella, M., Barber-Singh, J., Yamashita, T., & Matsuno-Yagi, A. (2006). Possibility of transkingdom gene therapy for Complex I diseases. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1757*(5–6), 708–714. <https://doi.org/10.1016/J.BBABIO.2006.01.011>
- Yao, H., Liu, D., Wang, Y., Zhang, Q., & Shen, Z. (2017a). High Prevalence and Predominance of the *aph*(2′′)-If Gene Conferring Aminoglycoside Resistance in

- Campylobacter. *Antimicrobial Agents and Chemotherapy*, 61(5).  
<https://doi.org/10.1128/AAC.00112-17>
- Yao, H., Liu, D., Wang, Y., Zhang, Q., & Shen, Z. (2017b). High prevalence and predominance of the aph(2')-if gene conferring aminoglycoside resistance in campylobacter. *Antimicrobial Agents and Chemotherapy*, 61(5).  
<https://doi.org/10.1128/AAC.00112-17>
- Yao, H., Liu, D., Wang, Y., Zhang, Q., & Shen, Z. (2017c). High prevalence and predominance of the aph(2')-if gene conferring aminoglycoside resistance in campylobacter. *Antimicrobial Agents and Chemotherapy*, 61(5).  
<https://doi.org/10.1128/AAC.00112-17>
- Yu, Y., Zhang, J., Zhu, F., Fan, M., Zheng, J., Cai, M., Zheng, L., Huang, F., Yu, Z., & Zhang, J. (2023). Enhanced protein degradation by black soldier fly larvae (*Hermetia illucens* L.) and its gut microbes. *Frontiers in microbiology*, 13.  
<https://doi.org/10.3389/FMICB.2022.1095025>
- Zhang, X., Zhou, Q., Tang, M., Pu, J., Zhang, J., Lu, J., Zhang, Y., & Gao, Y. (2021a). Aminoglycoside Resistance and Possible Mechanisms in *Campylobacter* Spp. Isolated From Chicken and Swine in Jiangsu, China. *Frontiers in Microbiology*, 12.  
<https://doi.org/10.3389/fmicb.2021.716185>
- Zhang, X., Zhou, Q., Tang, M., Pu, J., Zhang, J., Lu, J., Zhang, Y., & Gao, Y. (2021b). Aminoglycoside Resistance and Possible Mechanisms in *Campylobacter* Spp. Isolated From Chicken and Swine in Jiangsu, China. *Frontiers in Microbiology*, 12.  
<https://doi.org/10.3389/fmicb.2021.716185>
- Zhang, Y., Xiao, X., Elhag, O., Cai, M., Zheng, L., Huang, F., Jordan, H. R., Tomberlin, J. K., Sze, S. H., Yu, Z., & Zhang, J. (2022). *Hermetia illucens* L. larvae-associated

intestinal microbes reduce the transmission risk of zoonotic pathogens in pig manure. *Microbial Biotechnology*, 15(10), 2631. <https://doi.org/10.1111/1751-7915.14113>

Zhao, S., Mukherjee, S., Chen, Y., Li, C., Young, S., Warren, M., Abbott, J., Friedman, S., Kabera, C., Karlsson, M., & McDermott, P. F. (2014). Novel gentamicin resistance genes in *Campylobacter* isolated from humans and retail meats in the USA. *Journal of Antimicrobial Chemotherapy*, 70(5), 1314–1321. <https://doi.org/10.1093/jac/dkv001>

Zhao, S., Mukherjee, S., Chen, Y., Li, C., Young, S., Warren, M., Abbott, J., Friedman, S., Kabera, C., Karlsson, M., & McDermott, P. F. (2015). Novel gentamicin resistance genes in *Campylobacter* isolated from humans and retail meats in the USA. *The Journal of antimicrobial chemotherapy*, 70(5), 1314–1321. <https://doi.org/10.1093/JAC/DKV001>

Zhao, S., Tyson, G. H., Chen, Y., Li, C., Mukherjee, S., Young, S., Lam, C., Folster, J. P., Whichard, J. M., & McDermott, P. F. (2016). Whole-genome sequencing analysis accurately predicts antimicrobial resistance phenotypes in *Campylobacter* spp. *Applied and Environmental Microbiology*, 82(2), 459–466. [https://doi.org/10.1128/AEM.02873-15/SUPPL\\_FILE/ZAM999116841SD1.XLS](https://doi.org/10.1128/AEM.02873-15/SUPPL_FILE/ZAM999116841SD1.XLS)

## Anexos

- **Instalación de programas**

Se creó la carpeta PROGRAMS, donde se instalarán todos los programas para correr

- **Inicio del Nodo de CEDIA**

```
salloc -p gpu -n 1 -c 64 --mem=120G --gres=gpu:a100_3g.20gb:2 --  
time=15:00:00
```

```
ssh dgx-node-0-0
```

- **FASTqc**

```
wget
```

```
https://www.bioinformatics.babraham.ac.uk/projects/fastqc/fastqc\_v0.12.1.zip
```

```
unzip fastqc_v0.12.1.zip
```

Ruta del programa

```
export PATH=$PATH:/home/juan.baez/PROGRAMS/FastQC/
```

- **Trimmomatic**

```
wget
```

```
http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic/Trimmomatic-0.39.zip
```

```
unzip Trimmomatic-0.39.zip
```

Ruta del programa

```
/home/juan.baez/PROGRAMS/Trimmomatic-0.39/trimmomatic-0.39.jar
```

- **Bowtie2**

```
wget https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.5.2/bowtie2-
```

```
2.5.2-linux-aarch64.zip
```

```
unzip bowtie2-2.5.2-linux-aarch64.zip
```

Ruta del programa

```
/home/juan.baez/PROGRAMS/bowtie2-2.5.2-linux-aarch64
```

```
export PATH=$PATH:/home/juan.baez/PROGRAMS/bowtie2-2.5.2
```

- **Smatools**

```
enroot import docker://nvcr.io#nvidia/cuda
```

```
enroot create --name micudapersonal nvidia+cuda.sqsh
```

```
enroot list
```

```
enroot start --mount $HOME --root --rw PC sh -c '/bin/bash'
```

```
apt update
```

```
apt install samtools
```

```
cd home/juan.baez/
```

- **Bedtools**

```
wget https://github.com/arq5x/bedtools2/releases/download/v2.25.0/bedtools-
```

```
2.25.0.tar.gz
```

```
tar -zxvf bedtools-2.25.0.tar.gz
```

```
cd bedtools2
```

```
export PATH=$PATH:/home/juan.baez/PROGRAMS/bedtools2/bin
```

- **Análisis de calidad de la secuencia**

```
export PATH=$PATH:/home/juan.baez/PROGRAMS/FastQC/
```

```
fastqc /home/juan.baez/Tesis/Secuencias/PLA_FDME210308107-  
1r_H5Y3FDSX2_L4_2.fq.gz
```

```
fastqc /home/juan.baez/Tesis/Secuencias/PLA_FDME210308107-  
1r_H5Y3FDSX2_L4_1.fq.gz
```

- **Limpieza de secuencias**

```
java -jar /home/juan.baez/PROGRAMS/Trimmomatic-0.39/trimmomatic-0.39.jar  
PE -phred33 /home/juan.baez/Tesis/PLA_FDME210308107-  
1r_H5Y3FDSX2_L4_2.fq.gz /home/juan.baez/Tesis/PLA_FDME210308107-  
1r_H5Y3FDSX2_L4_1.fq.gz PLA_1P.fq.gz PLA_1U.fq.gz PLA_2P.fq.gz  
PLA_2U.fq.gz ILLUMINACLIP:/home/juan.baez/PROGRAMS/Trimmomatic-  
0.39/adapters/TruSeq3-PE.fa:2:30:10:2:True SLIDINGWINDOW:5:20 LEADING:7  
TRAILING:7 MINLEN:55
```

- **Verificación de calidad luego de limpieza**

```
fastqc /home/juan.baez/Prueba5/PLA_1P.fq.gz
```

```
fastqc /home/juan.baez/Prueba5/PLA_2P.fq.gz
```

- **Eliminación de secuencias de hospedero**

```
bowtie2-build /home/juan.baez/Tesis/Hospedero.fna H_DB
```

```
bowtie2 -x /home/juan.baez/Tesis/H_DB -1 /home/juan.baez/Tesis/PLA_1P.fq.gz  
-2 /home/juan.baez/Tesis/PLA_2P.fq.gz -S mau.sam
```

- **Inicio del Docker**

```
enroot list
```

```
enroot start --mount $HOME --root --rw PC sh -c '/bin/bash'
```

- **Seleccionar secuencias no mapeadas**

```
samtools view -b -f 12 -F 256 /home/juan.baez/Tesis/mau.sam >  
/home/juan.baez/Prueba5/bothEndsUnmapped.bam
```

```
samtools sort -n /home/juan.baez/Tesis/bothEndsUnmapped.bam -o  
/home/juan.baez/Prueba5/bothEndsUnmapped_sorted.bam
```

- **Dividir las lecturas en archivos paired end en formato “.fastq”**

```
bedtools bamtofastq -i /home/juan.baez/Tesis/bothEndsUnmapped_sorted.bam -  
fq /home/juan.baez/Tesis/PLA_1F.fastq -fq2 /home/juan.baez/Tesis/PLA_2R.fastq
```