

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR**  
**FACULTAD DE INGENIERÍA**  
**MAESTRÍA EN BIOLOGÍA COMPUTACIONAL**

Ensamblaje y anotación del genoma de *Dysgonomonas* sp. a partir de  
datos metagenómicos del tracto digestivo de la mosca soldado *Hermetia*  
*Illucens*

Trabajo previo a la obtención del título de Master en Biología  
Computacional

**MAYRA JAQUELINE SALAZAR ARIAS**

Quito, 2023

## **DERECHOS DE AUTOR**

Expreso que soy autora del presente trabajo de titulación y consulté las referencias indicadas en el mismo. Este trabajo no fue presentado de forma previa para la obtención de ningún grado académico.

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

---

**Mayra Jaqueline Salazar Arias**

## **APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN**

Certifico que el trabajo de Mayra Jaqueline Salazar Arias para la obtención del título de Máster en Biología Computacional se llevó a cabo bajo la normativa y reglamentación institucional y puede ser presentada para su calificación.

---

**Francisco Flores, Ph.D**

## DEDICATORIA

*A José Luis, quien ha vuelto mi  
corazón libre y me ha dado el  
valor para hacerle caso*

## ÍNDICE

DERECHOS DE AUTOR .....	ii
APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN .....	iii
DEDICATORIA .....	iv
ÍNDICE.....	v
ÍNDICE DE TABLAS .....	vii
ÍNDICE DE FIGURAS.....	viii
RESUMEN.....	x
1. FORMULACIÓN DEL PROBLEMA.....	1
2. JUSTIFICACIÓN.....	2
3. OBJETIVOS.....	5
3.1 Objetivo General.....	5
3.2 Objetivos Específicos .....	5
4. REVISIÓN BIBLIOGRÁFICA .....	5
4.1 Mosca Soldado Negra.....	5
4.2 Bioconversión de Residuos .....	7
4.3 <i>Dysgonomonas</i> sp. ....	9
4.4 Ensamblaje y anotación de genomas.....	10
4.4.1 Secuenciación de genomas .....	11
4.4.2 Tecnologías de Secuenciación de Alto Rendimiento (HTS) .....	11
4.4.3 Ensamblaje de genomas .....	12
4.5 Anotación de Genomas .....	13
5. METODOLOGÍA.....	14
5.1 Eliminación de las secuencias del hospedero .....	15
5.2 Ensamblaje y Anotación del Genoma .....	16
5.3 Búsqueda de Genes y Rutas Metabólicas de interés.....	17
6. RESULTADOS Y DISCUSIÓN.....	18
6.1 Eliminación de las secuencias del hospedero .....	18
6.1.1 Control de calidad.....	18
6.1.2 Limpieza de secuencias.....	20
6.1.3 Eliminación de duplicados.....	21
6.1.4 Eliminación de contaminación de la secuencia .....	22
6.1.5 Transformación a Fastq.....	23

6.2	Ensamblaje y Anotación del genoma .....	23
6.2.1	Clasificación Taxonómica .....	23
6.2.2	Ensamblaje del Genoma y Comparación .....	28
6.2.3	Agrupación y Optimización de Bins.....	31
6.2.4	Evaluación de la calidad de los bins.....	32
6.2.5	Anotación del genoma .....	35
6.2.6	Clasificación Taxonómica .....	36
6.3	Búsqueda de Genes y Rutas Metabólicas de Interés .....	39
6.3.1	Perfil Funcional del microorganismo.....	39
6.3.2	Búsqueda de genes relacionados con hidrólisis de PET .....	42
7.	CONCLUSIONES Y RECOMENDACIONES.....	44
7.1	Conclusiones.....	44
7.2	Recomendaciones .....	45
8.	REFERENCIAS BIBLIOGRÁFICAS.....	46

## ÍNDICE DE TABLAS

Tabla 1.	Descripción del procedimiento realizado para la eliminación de las secuencias del hospedero.....	16
Tabla 2.	Descripción del procedimiento realizado para el ensamblado y la anotación del genoma.....	16
Tabla 3.	Descripción del procedimiento realizado para la búsqueda de genes y rutas metabólicas de interés. ....	17
Tabla 4.	Estadística básica para las secuencias en FastQC (forware y reverse) .....	18
Tabla 5.	Estadística básica posterior a la limpieza en trimmomatic utilizando FastQC (forware y reverse).....	20
Tabla 6.	Resultado de la anotación del genoma en Kbase.....	35
Tabla 7.	Resultado de la anotación del genoma y clasificación taxonómica en BV-BRC	36
Tabla 8.	Genes que intervienen en la hidrolisis de PET.....	42
Tabla 9.	Taxonomía de ubicación más cercana obtenida.....	44

## ÍNDICE DE FIGURAS

Figura 1. Mosca Solado Negra <i>Hermetia illucens</i> ( Zamprogna et al., 2017) .....	6
Figura 2. Ciclo de vida de <i>Hermetia illucens</i> (Pazmiño, 2021) .....	7
Figura 3. <i>Dysgonomonas</i> sp. (Kita et al., 2015) .....	10
Figura 4. Representación del flujo de trabajo de un ensamblaje de novo (Aguilar et al., 2015). .....	12
Figura 5. Flujo de trabajo. i) Eliminación de la secuencia del hospedero, ii) Ensamblaje y del Genoma, iii) Anotación del Genoma y iv) Búsqueda de genes y rutas metabólicas de interés .....	15
Figura 6. Flujo de trabajo del ensamblado y la anotación del genoma (Chivian et al., 2020). .....	16
Figura 7. Calidad de la secuencia por par de base para las secuencias forward (derecha) y reverse (izquierda) .....	19
Figura 8. Contenido de GC para las secuencias, forward (izquierda) y reverse (derecha).....	19
Figura 9. Calidad de la secuencia por par de base, posterior a la limpieza en <i>Trimmomatic</i> , forward (derecha) y reverse (izquierda) .....	20
Figura 10. Eliminación de secuencias mediante la herramienta Dedupe de BMap.....	21
Figura 11. Datos obtenidos en Reformat de BMap.....	21
Figura 12. Genoma de <i>Hermetia illucens</i> en formato FASTA .....	22
Figura 13. Base de datos creada con base al genoma de <i>Hermetia illucens</i> .....	22
Figura 14. Lecturas mapeadas y no mapeadas de <i>Hermetia illucens</i> .....	23
Figura 15. Clasificación taxonómica a nivel de Filo .....	24
Figura 16. Clasificación taxonómica a nivel de Clase .....	25
Figura 17. Clasificación taxonómica a nivel de Orden .....	25
Figura 18. Clasificación taxonómica a nivel de Familia .....	26
Figura 19. Clasificación taxonómica a nivel de Género .....	27
Figura 20. Clasificación taxonómica a nivel de Especie .....	27
Figura 21. Evaluación de los ensambladores metaSPAdes, IDBA, MEGA HIT y HipMER..	29
Figura 22. Longitudes de contigs en el eje X y suma acumulada en el eje Y, con escala de porcentaje.....	30
Figura 23. Longitudes de contigs en el eje X y suma acumulada en el eje Y .....	31
Figura 24. Comparación de los métodos de agrupación, de izquierda a derecha se encuentra Max Bin2, MetaBat y Concoct, y la optimización mediante DASTool.....	32

Figura 25. Calidad y contaminación de bins .....	33
Figura 26. Clasificación Taxonómica posible mediante Check M.....	34
Figura 27. Árbol filogenético para los bin 003, 004 y 008 con Species Tree en Kbase .....	38
Figura 28. Árbol filogenético para el bin 004 con BV-BRC en Patric .....	39
Figura 29. Mapa de calor de módulos y cobertura de los componentes de la cadena de transporte de electrones en DRAM.....	41
Figura 30. Funciones metabólicas generadas en DRAM.....	41

## RESUMEN

El aumento de la población está directamente relacionada con el incremento de la contaminación generada por residuos sólidos, debido a sus altas tasas de consumo y generación de desperdicios, los cuales no tienen una gestión adecuada y esto conlleva a problemas ambientales, sociales y de salud pública. Cada año se producen 1300 millones de toneladas de residuos a nivel mundial, de los cuales el 13.5% es reciclado y el 5.5% sirven de abonos, en Ecuador la producción semanal de basura es de 58.829 toneladas, de los cuales el 20% es tratado, el 14% es reciclado, quedando la diferencia en situaciones inapropiadas de descarte, dentro de los mayores contaminantes, se encuentran los residuos plásticos, que son un grave problema medio ambiental.

En busca de soluciones a esta problemática, se ha propuesto la bioconversión de residuos a través de la mosca soldado negra *Hermetia illucens* (BSF), considerando que es un insecto que durante su etapa larvaria tiene la capacidad de degradación de residuos orgánicos y que la bioconversión de desechos guiada por insectos, representa procesos exitosos de biorremediación sostenible, es así que este trabajo de titulación tiene como objetivo ensamblar y anotar el genoma de *Dysgonomonas* sp. a partir de datos metagenómicos del tracto digestivo de la mosca soldado negra *Hermetia Illucens*, con la finalidad de buscar genes de interés y/o rutas metabólicas que estén vinculadas con la degradación de microplásticos de polietileno (PE) y poliestireno (PS).

Se realizó el análisis de calidad mediante la herramienta FastQC y el recorte de las secuencias de baja calidad en *Trimmomatic* para posteriormente eliminar las secuencias del hospedero en bash de Linux. Los datos de salida, en formato FastQ, fueron analizados en Kbase acorde a la narrativa 33233 para el ensamblaje y la anotación del genoma, producto de este proceso, se obtuvo la clasificación taxonómica del organismo para el BIN seleccionado 004, por presentar mayor integridad y ausencia de contaminación en la secuencia.

Mediante RAST y BV-BRC se obtuvo la siguiente clasificación taxonómica Dominio Bacteria, Filo Bacteroidota, Clase Bacteroidia, orden Bacteroidales, Familia Dysgonomonadaceae, Género *Dysgonomonas*, sin poder llegar a determinar especie, empleando Species Tree y BV-BRC se realizó la construcción de árboles filogenéticos, de la cual se observó que *Dysgonomonas capnocytophagoides* guardaba estrecha relación con el bin trabajado.

De la búsqueda de genes y rutas metabólicas de interés, *Dysgonomonas* sp. presenta rutas metabólicas para la fermentación de la glucosa llamada Entner-Doudoroff, rutas metabólicas de sulfatos, carbohidratos y nitrógeno, y el ciclo de Krebs, adicionalmente se encuentra relacionada con la degradación de  $\alpha$ -galactosas lo cual permite la degradación de carbohidratos vegetales. En cuanto a la cadena de transporte de energía, se observa la de citocromo bd ubiquinol oxidasa que participa en la cadena respiratoria de procariotas.

## **1. FORMULACIÓN DEL PROBLEMA**

El incremento de la contaminación generada por residuos sólidos se encuentra relacionado directamente con el aumento de la población, sus altas tasas de consumo y generación de desperdicios, con el desarrollo agroindustrial, comercial y turístico, los cuales no tienen una gestión adecuada y contribuyen a los problemas sociales, ambientales y de salud pública (Aleluya, 2016).

La contaminación ambiental ha tenido una mayor incidencia debido a las malas prácticas de gestión, procesos de tratamiento inadecuado, sobreexplotación de recursos de los ecosistemas, faltas de alternativas para procesos de tratamiento, ausencia de inversión pública y privada y sistemas de eliminación que se encuentran colapsados (Ferronato, 2019).

Los residuos se clasifican de manera general en orgánicos e inorgánicos, los cuales no son separados y no siguen un proceso de reciclaje por parte de la población, sino más bien de producción, consumo y descarte, estos residuos contaminan aguas subterráneas, generan gases de efecto invernadero, erosión de suelos, contaminación por lixiviados, generación de biogás, obstrucción de desagües, entre otras afectaciones ambientales graves.

Cada año se producen 1.300 millones de toneladas de residuos a nivel mundial, de los cuales apenas el 13.5% es reciclado y el 5.5% funge como abonos, quedando la diferencia como basura, que incrementa y agrava los problemas medioambientales (Gustavsson et al., 2011).

En Ecuador, la producción semanal de basura es de 58.829 toneladas, de las cuales el 20% es recolectado y tratado de forma adecuada, el 14% es reciclado, quedando la diferencia en situaciones inapropiadas de descarte como a cielo abierto, botaderos, ríos, lagos y basura que queda sometida a procesos de incineración. Con base a la disposición final de residuos que no son tratados y reciclados, el 19.1% se encuentra

en rellenos sanitarios, el 23.3% en vertederos controlados, el 53.6% en botaderos a cielo abierto y el 4.1% en incineradores (Solís 2020).

De la cantidad total recolectada por parte de las municipalidades, se ha reportado que el 52% es una recolección diferenciada y de este porcentaje, el 24% es sometido a un tratamiento óptimo, quedando el restante en sitios destinados a rellenos sanitarios. (Solís 2012).

De los residuos generados a nivel nacional, el 53.9% son orgánicos, el 24.4% corresponden a papel y cartón, el 10.9% a plástico, el 6.5% a metales y chatarra y el 4.4% a vidrio. Esta problemática empeoró desde 1970 con el incremento de las zonas urbanas, siendo las ciudades con mayor población quienes tienen los índices más altos en generación de residuos (Solís 2012).

La producción de residuos además de estar relacionada con la densidad poblacional, también se centra en los modelos económicos y de administración cantonal y los niveles de pobreza de los sectores (Solís 2020).

## **2. JUSTIFICACIÓN**

La contaminación de los ecosistemas son un grave problema medio ambiental, de los cuales la contaminación por residuos plásticos se ha ido incrementando de forma exponencial, sobrecargando los vertederos, rellenos sanitarios y fuentes de agua, es así que, en busca de soluciones se ha propuesto como una alternativa natural la bioconversión de residuos a través del insecto mosca soldado negra *Hermetia illucens* (BSF), considerando que es un insecto que durante su etapa larvaria tiene la capacidad de degradación de residuos orgánicos, los mismos que los almacena como proteínas, grasas y carbohidratos (Del Hierro et al., 2021).

La bioconversión de desechos guiada por insectos representa procesos exitosos de biorremediación sostenible. La mosca soldado negra es un insecto saprofito, que se

alimenta de materia en descomposición, pero no se considera una plaga, debido a que no tiene piezas bucales funcionales, no muerden o pican y no son conocidas por ser un vector de enfermedades (Oliveira et al., 2015), por el contrario, son ambientalmente beneficiosas, pueden alimentarse de residuos orgánicos en descomposición, restos de comida, vegetación, residuos de la agricultura y estudios actuales arrojan datos de degradación de microplásticos.

Se requieren 40.000 larvas de BSF por metro cuadrado para el tratamiento de 60 Kg de desechos para la conversión de residuos orgánicos, y formar agregados de materia orgánica, en el proceso de degradación, las larvas convierten los residuos orgánicos en proteína (37-63% en peso seco) y grasa (7-39% en peso seco), por lo que tienen amplias aplicaciones en el manejo de residuos, medicina o la elaboración de balanceados (Ao et al., 2021).

Los insectos mantienen dentro de su microbiota intestinal, microorganismos involucrados en el desarrollo de varias funciones vitales, así como capacidades metabólicas de fermentación de carbohidratos complejos, funciones que favorecen al hospedero atribuyéndole beneficios (Zheng et al., 2016), estos microorganismos se encuentran involucrados en varias funciones del hospedero, como mecanismos de defensa ante toxinas de plantas, respuestas a nivel fisiológico, aumento del tiempo de vida, intervención en el desarrollo y potencial reproductivo, entre otros (Zhineng et al., 2021).

Los insectos carecen de enzimas, por lo cual no pueden degradar carbohidratos, por lo que la microbiota que se encuentra en sus intestinos participan con esta función, estos microorganismos a través de sus capacidades metabólicas intervienen en la fermentación de carbohidratos complejos, liberando ácidos grasos de cadena corta que se utilizan como parte del metabolismo energético del huésped.

Yu et al., (2023) analizó mediante metagenómica, bacterias intestinales de las larvas de la mosca soldado negra, la cual arrojó datos de un grupo de bacterias con funciones

benéficas, entre ellas *Dysgonomonas* sp. la misma que participaría en la biosíntesis y el metabolismo de los polisacáridos, la traducción, el transporte de membrana y el metabolismo energético.

Considerando que Ecuador mantiene una producción semanal de basura de 58.829 toneladas, la generación de residuos es una problemática medioambiental que no mantiene políticas y proyectos enfocados en su disminución y tratamiento; y, tomando en consideración la importancia de la mosca soldado negra como una herramienta para la transformación y bioconversión de residuos, este organismo mantiene un rol significativo en el reciclaje de residuos biológicos y otros nutrientes, por lo que es objeto de estudio a nivel mundial, arrojando resultados de actividad enzimática en las larvas que permiten la reducción rápida de materia orgánica. Esta capacidad de degradación se debe a comunidades bacterianas presentes en el sistema digestivo de este organismo, como se ha mencionado anteriormente (Shelomi et al., 2020).

Así también, no se han encontrado en Ecuador estudios con información concerniente a la diversidad poblacional de la mosca soldado negra y su variabilidad genética, estudios que relacionen los microorganismos propios de la microbiota de la BSF asociados con rutas metabólicas y genes vinculados a la degradación de residuos.

El ensamblaje de genomas bacterianos a partir de metagenomas, permite identificar a nivel funcional a los microorganismos, sin necesidad de realizar cultivos, permite buscar relaciones filogenéticas, rutas metabólicas, interacciones tróficas, entre otros (Sangwan, Xia, & Gilbert, 2016).

### **3. OBJETIVOS**

#### **3.1 Objetivo General**

Ensamblar y anotar el genoma de *Dysgonomonas* sp. a partir de datos metagenómicos del tracto digestivo de la mosca soldado negra *Hermetia Illucens*, con la finalidad de buscar genes de interés y/o rutas metabólicas que estén vinculadas con la degradación de microplásticos de polietileno (PE) y poliestireno (PS).

#### **3.2 Objetivos Específicos**

- Limpiar las secuencias de *Dysgonomonas* sp. mediante el uso de un software informático o lenguaje de programación adecuado.
- Ensamblar el genoma de *Dysgonomonas* sp. a partir de los datos metagenómicos disponibles, para generar una secuencia que represente el genoma de este organismo.
- Anotar el genoma ensamblado de *Dysgonomonas* sp. con la finalidad de obtener la identificación estructural genética y regiones codificantes.
- Buscar rutas metabólicas y genes de interés producto del ensamblaje y anotación de genoma de *Dysgonomonas* sp. con incidencia en la degradación de residuos.

### **4. REVISIÓN BIBLIOGRÁFICA**

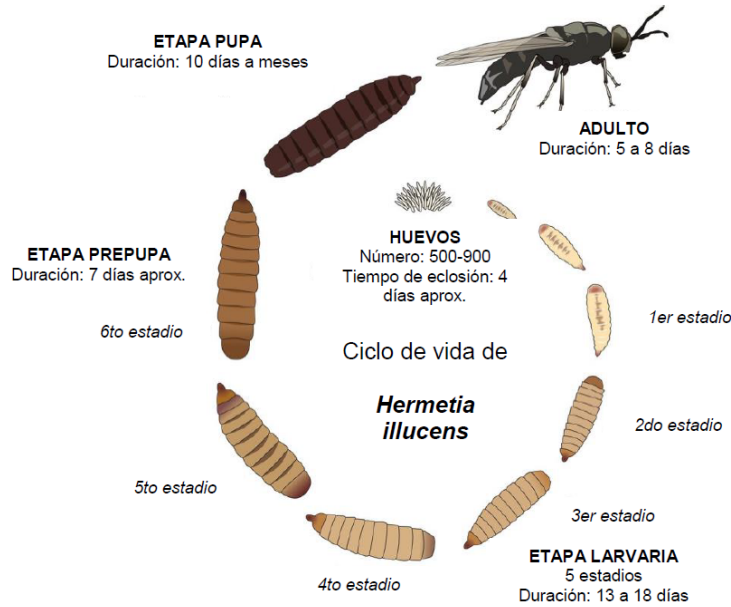
#### **4.1 Mosca Soldado Negra**

La mosca soldado negra es un insecto de orden díptera, de la familia Stratiomyidae, género *Hermetia*, especie *illucens*, se encuentra distribuida en todo el mundo, habitando regiones tropicales y con temperaturas cálidas, su constitución física se caracteriza por una estructura alargada de color negro

(figura 1) diferente a la mosca común, el desarrollo larvario de este insecto está comprendido por cinco etapas: huevecillo, larva, prepupa, pupa y adulta, su ciclo de vida es de 31 días aproximadamente (figura 2), sin embargo puede extenderse por las condiciones ambientales y la cantidad de alimento que dispongan, considerando que a bajas temperaturas y/o falta de alimento, la mosca se puede mantener en etapa larvaria hasta cuatro meses, lo que normalmente la larva se desarrolla en dos semanas, esta etapa es importante pues es en donde se aprovecha con mayor eficiencia los residuos que consume y se generan antibióticos naturales sin afectación de la biomasa, consume cualquier residuo orgánico, en óptimas condiciones como en descomposición, estiércol de animales y animales muertos, estos alimentos no generan patógenos en su producción de biomasa (Boccazzi et al, 2017).



**Figura 1.** Mosca Solado Negra *Hermetia. illucens* ( Zamproгна et al., 2017)



**Figura 2.** Ciclo de vida de *Hermetia illucens* (Pazmiño, 2021)

## 4.2 Bioconversión de Residuos

La BSF se encuentra vinculada a la degradación de materia orgánica, los nutrientes que produce están delimitados por el alimento que consume, si son residuos de alimentos (frutas y verduras) produce hasta el 37.8% de proteína y si su alimentación está basada en estiércol de animales hasta el 44% de proteína; en cuanto a la producción de grasas, estas se han visto definidas también por el tipo de alimento, si consume residuos que tengan aceites, produce de un 42% a un 45% de grasa, si consume estiércol genera un 35% de grasa y gallinaza entre un 15% a un 25% de grasa (Wang y Shelomi, 2017).

Las larvas de BSF mantienen actividad enzimática (amilasa, celulasa, lipasa y proteasa), las cuales participan en la degradación de materia orgánica, las cuales están relacionadas con comunidades bacterianas propias del sistema digestivo de este insecto (Caruso et al., 2013).

Considerando estas características, la BSF se considera un organismo modelo para investigación y se utiliza industrialmente para reciclar ciertos desechos que son específicos, entre ellos desechos orgánicos, con el fin de convertirlos en productos útiles, como proteína para alimentación de animales, compost que se utilice como fertilizante y grasa para usos energéticos (Zhan et al., 2020).

La mosca soldado negra en su estadio de pupa, está compuesta por proteína, grasa y quitina en un 40%, 30% y 9% respectivamente, sin residuos de pesticidas o micotoxinas, lo que la hace importante comercialmente al ser utilizada como un alimento o suplemento animal (Choi et al., 2012). El uso de proteína como alimento animal genera beneficios, es así que el animal al consumir este alimento disminuye la producción de dióxido de carbono, la generación de bacterias nocivas y genera antibióticos naturales (Beskin et al., 2018; De Smet et al., 2018).

Es así que, la empresa ESR International, mantiene una patente con el proceso de bioconversión utilizando larvas de la BSF, estas larvas reducen hasta el 95% de residuos alimenticios en poco tiempo, sin requerimientos de agua, productos químicos o energía y sin producción de contaminantes, únicamente se genera dióxido de carbono en pocas cantidades (ESR International, 2016).

Investigaciones anteriores han trabajado caracterizando las enzimas digestivas que se produce en el intestino de la mosca soldado negra (Ki, et al., 2011), la microbiota del organismo de las larvas (Shelomi et al., 2020) la interrelación entre las comunidades bacterianas y el desarrollo de la BSF y se ha observado que difieren las comunidades bacterianas según la región y el tipo de sustrato que hayan recibido. (De Smet et al., 2018).

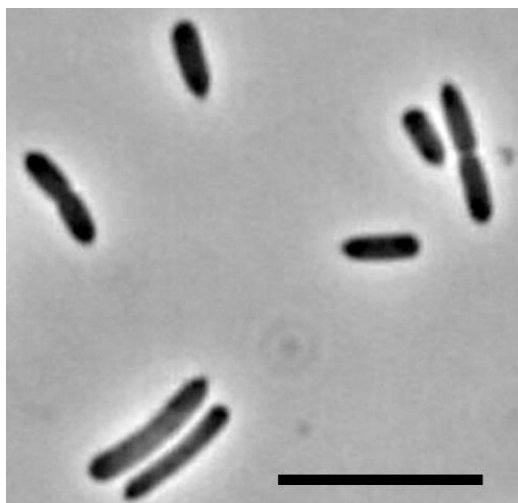
Zhineng et al., (2021) investigaron la microbiota intestinal de las larvas de BSF con base a diferentes estadios y sustratos, de lo cual se observó que el intestino de las larvas contienen alrededor de 11.000 consorcios bacterianos pertenecientes a Proteobacteria, Bacteroidetes, Firmicutes, y Actinobacteria, en

cuanto a especies *Enterococcus*, *Acinetobacter*, *Providencia*, *Enterobacter* y *Myroides*, Firmicutes está vinculada a la degradación de materia orgánica proveniente de estiércol animal y Bacteroides de materia orgánica de alto peso molecular.

Khamis et al., (2020) trabajó en la variabilidad genética del microbioma intestinal de varias poblaciones de BSF, para ello usaron el gen COI mitocondrial y el 16S, producto del estudio se fueron generando secuencias las cuales fueron vinculadas con las accesiones del Gen Bank KM967419.1, FJ794355.1, FJ794361.1, FJ794367.1, KC192965.1 y KY817115.1 para estudios de análisis filogenético, las distancias genéticas fueron de 0,0091 y 0,0407 entre las accesiones del Gen Bank y las secuencias de estudio, se encontraron relaciones en sus haplotipos de los diferentes grupos poblacionales y en cuanto a las bacterias que se encontraron con mayor abundancia formando la microbiota intestinal de la BSF, fueron Enterobactereaceae, Dysgonomonadaceae, Wohlfahrtiimonadaceae y Enterococcaceae, en conclusión, se observó que existe variación filogeográfica entre las diferentes poblaciones de estudio.

#### **4.3 *Dysgonomonas* sp.**

*Dysgonomonas* sp. es un cocobacilo gramnegativo, anaerobio facultativo, perteneciente a la familia Dysgonomonadaceae, forman colonias de 1 a 2 mm de diámetro, color gris, no hemolíticas y de olor ligero (. Al inicio fue aislado de humanos como patógeno oportunista, sin embargo, ahora se ha observado aislados en intestinos de insectos, entre ellos, abejas, moscas y termitas (Auer et al, 2017)



**Figura 3.** *Dysgonomonas* sp. (Kita et al., 2015)

Este microorganismo como parte de la microbiota intestinal de la BSF, se cree que está vinculado con la degradación de polisacáridos complejos para la producción de biocombustibles (Yu et al., 2011) y en conjunto con *Actinomyces* sp. y *Enterococcus* sp. participan en diversas funciones metabólicas de los huéspedes (Khademian et al., 2021), se encuentra relacionada con la digestión y absorción de proteínas, participa en la degradación de celulosa, lignocelulosa e interviene en la absorción de proteínas de los insectos huésped (Auer et al., 2017).

#### **4.4 Ensamblaje y anotación de genomas**

La secuenciación de alto rendimiento (HTS- High throughput Sequencing) permite secuenciar en poco tiempo grandes genomas, reduciendo costos y procesando varias muestras en paralelo, proporciona información a manera de un catálogo completo de genes que conforman un organismo y que se pueden expresar en el transcurso de su vida, estos datos necesitan un amplio análisis computacional, debido al gran volumen de datos biológicos generados (Aguilar et al., 2015).

#### **4.4.1 Secuenciación de genomas**

La secuenciación de genomas es un proceso que permite realizar la identificación de genes de un organismo. La secuenciación sanger permitió desarrollar un amplio espectro de avances científicos, como la secuenciación del primer genoma bacteriano *Haemophilus influenzae* y la secuenciación del genoma humano, este método fue trabajado por más de 20 años, sin embargo tenía varias limitantes como costos y tiempos, de aquí la generación de nuevas alternativas de secuenciación como 454, secuenciación por síntesis, SOLiD, ION Torrent las cuales requieren de un paso previo de amplificación, hasta la actualidad Helicos, SMRT, PACBIO, tecnología ILLUMINA que ya no requieren de pasos previos (Aguilar et al., 2015).

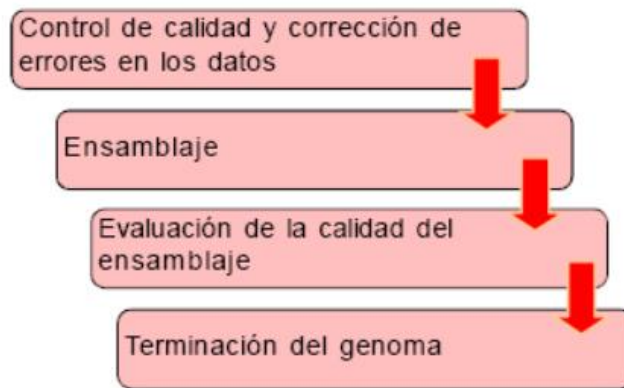
#### **4.4.2 Tecnologías de Secuenciación de Alto Rendimiento (HTS)**

Estas tecnologías se caracterizan por la capacidad de generar un gran volumen de datos a costos relativamente bajos, pueden producir millones y billones de lecturas por cada corrida, tiene una mayor cobertura en comparación con la secuenciación automática sanger, y sus reads son más cortos exceptuando los de PACBIO.

Las tecnologías HTS tiene un amplio campo de aplicación, como biología, genómica, proteómica, transcriptómica, metagenómica, entre otras. Entre sus aplicaciones están la secuenciación de genomas completos, secuenciación de *novo* que no requiere de un genoma de referencia, y buscan la construcción de un genoma completo, la resecuenciación que necesita de un genoma de referencia y buscan variaciones mediante este contraste de comparación, cuantificación de expresión génica, estudio de interacciones ADN-PROTEINA, resistencias a antibióticos, identificación de mecanismos metabólicos, entre otros (Aguilar et al., 2015).

### 4.4.3 Ensamblaje de genomas

El ensamblaje de genomas es un proceso en el cual, al partir de pequeños fragmentos de ADN, se interpreta la secuencia genómica. El ensamblaje puede ser por comparación y de *novo*, el primero parte de un genoma conocido, que sirve de base para la reconstrucción del genoma, mientras que el segundo no parte de un genoma conocido, el proceso de un ensamblaje de *novo* se describe en la figura 4.



**Figura 4.** Representación del flujo de trabajo de un ensamblaje de novo (Aguilar et al., 2015).

La primera parte del ensamblaje de *novo* es el control de calidad, en este paso se hace una revisión de la información para identificar los datos que tienen problemas de calidad y excluirlos, para secuencias largas, se basan en la calidad de las bases, la distribución de los nucleótidos, el contenido de guaninas-citocinas, identificación de secuencias repetidas, entre otros; para secuencias cortas se revisa las lecturas de baja calidad y se elimina a partir del punto de inflexión en donde decae la calidad de secuencias, el porcentaje de error se encuentra entre 0.1 y 1%.

El ensamblaje es basa en unir fragmentos individuales, formando secuencias largas con base a fragmentos iguales de ADN, llamados contigs. Los ensambladores que realizan este trabajo se agrupan en tres tipos:

i) Greedy, que conecta los reads de forma iterativa y se usa principalmente para secuenciación sanger, pero no son eficientes para regiones repetitivas largas; ii) Overlap-Layout-Consensus (OLC) que identifica los reads que se solapan correctamente y a partir de ellos, genera gráficos que constan de un nodo y un conector por cada solapamiento, con estos conectores se generan algoritmos considerando todos los reads, el algoritmo construye un genoma con base a la búsqueda de un camino que atraviese todos los nodos sin repetición, y iii) Gráficos de Brujin que generan modelos basados en la relación que existe entre subcadenas de longitud  $k$  en los reads, nodos en los gráficos que son los  $k$ -mers y los conectores  $k$ -mers que solapan en  $k-1$ , es así que la longitud del  $k$ -mer relaciona la longitud del solapamiento que se detecta.

La evaluación de la calidad utiliza algunos indicadores métricos cuantitativos, calcula la longitud mínima, media y máxima de los contigs y la longitud total del ensamblaje, el mismo que debe coincidir con el tamaño esperado, el indicador N50 es el valor estadístico principal que indica que tan apropiado fue el ensamblaje.

Y finalmente el proceso de terminación que es muy importante, pues es en donde se corrigen los errores del ensamblaje y de la secuenciación, se asocia al cierre de gaps y es el paso que mayor tiempo consume de todo el proceso (Aguilar et al., 2015).

#### **4.5 Anotación de Genomas**

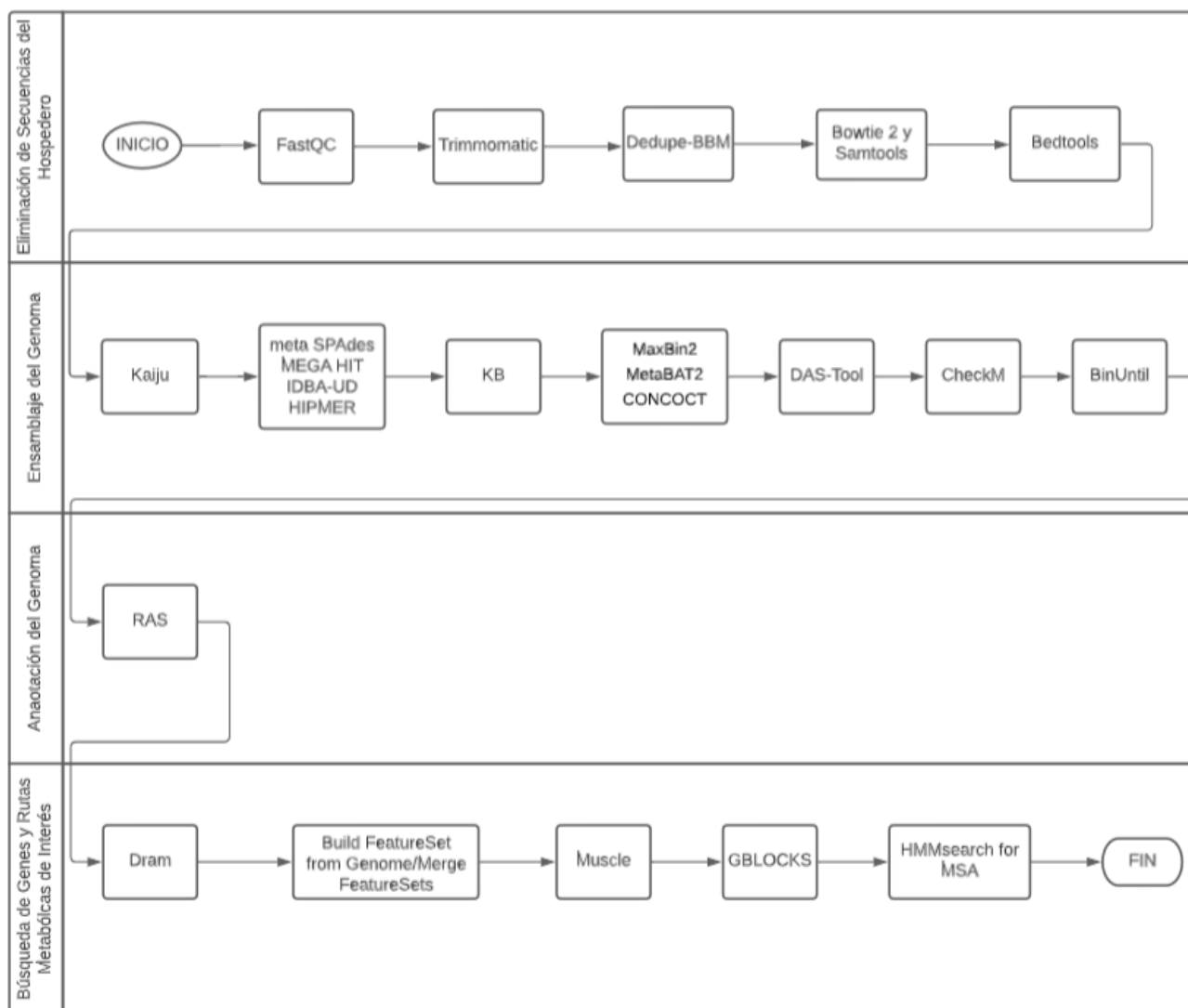
La anotación de genomas consiste en la identificación de las características del genoma que ha sido ensamblado, y consta de dos partes, la anotación estructural y la anotación funcional.

La anotación estructural predice las regiones codificantes, en este tipo de anotación se encuentran dos categorías, el método de *novo* que emplea algoritmos para conocer si la secuencia es codificante mediante motivos específicos; y, el método de comparación que busca regiones de organismos con alta similitud con relación a bases de datos.

La anotación funcional asigna datos biológicos a genes identificados, y se lo realiza mediante algoritmos de alineamiento como BLAST, este encuentra secuencias homologas relacionando bases de datos, también trabaja a través de motivos y dominios funcionales (Aguilar et al., 2015).

## **5. METODOLOGÍA**

Para el desarrollo del proyecto, se utilizó datos de una investigación previa, pertenecientes a muestras de larvas de mosca soldado negra, obtenidos mediante un equipo NovaSeq. Partimos de un dato de salida de 12 Gb, con dos archivos FASTQC (forward y reverse). A continuación, se detalla en la figura 5 el flujo de trabajo realizado con las herramientas de análisis.



**Figura 5.** Flujo de trabajo. i) Eliminación de la secuencia del hospedero, ii) Ensamblaje y del Genoma, iii) Anotación del Genoma y iv) Búsqueda de genes y rutas metabólicas de interés

### 5.1 Eliminación de las secuencias del hospedero

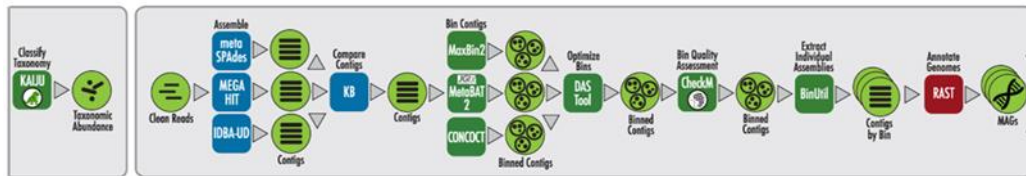
La eliminación de la secuencia del hospedero se realizó en bash de Linux acorde a lo descrito en la tabla 1.

**Tabla 1.** Descripción del procedimiento realizado para la eliminación de las secuencias del hospedero.

Nº	Proceso	Herramienta utilizada
1	Control de calidad de las lecturas	FastQC
2	Limpieza de secuencias	Trimmomatic
3	Eliminación de datos duplicados	Dedupe-BBMap
4	Eliminación de la contaminación de la secuencia	Bowtie2 y Samtools
5	Transformación a Fastq	Bedtools

## 5.2 Ensamblaje y Anotación del Genoma

El ensamblaje y la anotación del Genoma se trabajó desde la plataforma KBase, <https://narrative.kbase.us/narratives>, acorde al flujo de trabajo (figura 6) descrito en la tabla 2.



**Figura 6.** Flujo de trabajo del ensamblado y la anotación del genoma (Chivian et al., 2020).

**Tabla 2.** Descripción del procedimiento realizado para el ensamblado y la anotación del genoma

Nº	Proceso	Herramienta utilizada
1	Clasificación taxonómica de microorganismos con base a los reads obtenidos.	Kaiju
2	Ensamblaje independiente con las cuatro herramientas descritas.	Meta SPAdes MEGA HIT

		IDBA-UD
		HIPMER_ASSEMBLY
<b>3</b>	Comparación de los cuatro ensambladores (punto 2) y elección del mejor ensamblador.	KB
<b>4</b>	Combinación de contigs	MaxBin2 MetaBAT2 CONCOCT
<b>5</b>	Optimización de bins	DAS-Tool
<b>6</b>	Evaluación de la calidad de los bins	CheckM
<b>7</b>	Obtención del bin de interés ( <i>Dysgonomomas</i> sp.)	BinUntil
<b>8</b>	Anotación del genoma	RAST

### 5.3 Búsqueda de Genes y Rutas Metabólicas de interés

La búsqueda de genes y rutas metabólicas de interés se trabajó desde la plataforma KBase, <https://narrative.kbase.us/narratives>, acorde a lo descrito en la tabla 3.

**Tabla 3.** Descripción del procedimiento realizado para la búsqueda de genes y rutas metabólicas de interés.

N°	Proceso	Herramienta utilizada
<b>1</b>	Perfil funcional	Dram
<b>2</b>	Construcción de un set de genes PET	Build FeatureSet from Genome / Merge FeatureSets:
<b>3</b>	Alineamiento múltiple de secuencias	Muscle
<b>4</b>	Eliminación de regiones variables	Gblocks
<b>5</b>	Búsqueda de coincidencias para hidrolisis PET en las regiones conservadas	HMMsearch for MSA

## 6. RESULTADOS Y DISCUSIÓN

### 6.1 Eliminación de las secuencias del hospedero

#### 6.1.1 Control de calidad

El control de calidad hace una revisión de la información para identificar los datos que tienen problemas de calidad y excluirlas, se basa en la calidad de las bases, la distribución de los nucleótidos, el contenido de guaninas-citocinas, entre otros; en la tabla 4. se observa una estadística básica para las secuencias, generada en FastQC en Kbase, la longitud de sus secuencias está entre las 50 y las 150 pares de bases, y su contenido de Guanina Citocina es del 40%.

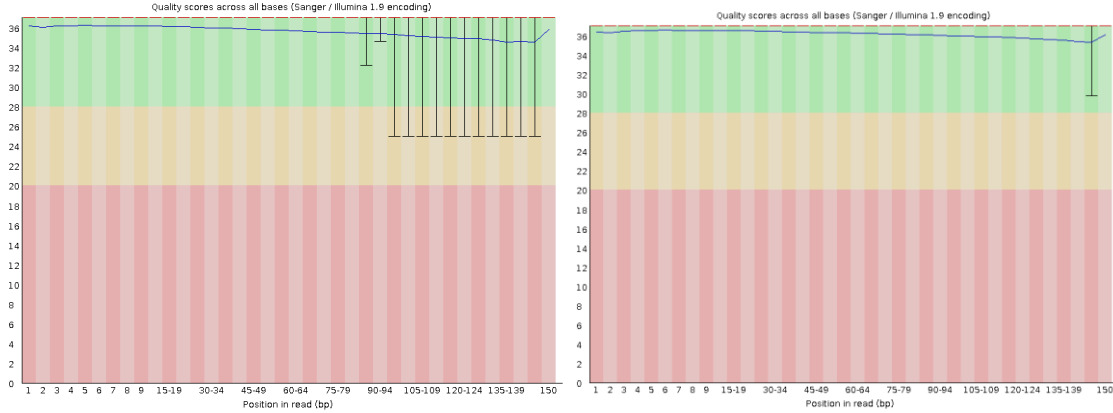
**Tabla 4.** Estadística básica para las secuencias en FastQC (forward y reverse)

Measure	Value	Measure	Value
Filename	PE1_2_151985_2_1.fwd.fastq	Filename	PE1_2_151985_2_1.rev.fastq
File type	Conventional base calls	File type	Conventional base calls
Encoding	Sanger / Illumina 1.9	Encoding	Sanger / Illumina 1.9
Total Sequences	24570279	Total Sequences	24570279
Total Bases	3.6 Gbp	Total Bases	3.5 Gbp
Sequences flagged as poor quality	0	Sequences flagged as poor quality	0
Sequence length	50-150	Sequence length	50-150
%GC	40	%GC	40

La calidad de las secuencias es buena, considerando que se encuentran en la zona verde y naranja para forward, y en la verde para reverse (figura 7), no se observan caídas repentinas o lecturas de baja calidad, por lo que inferimos que no existen errores inesperados.

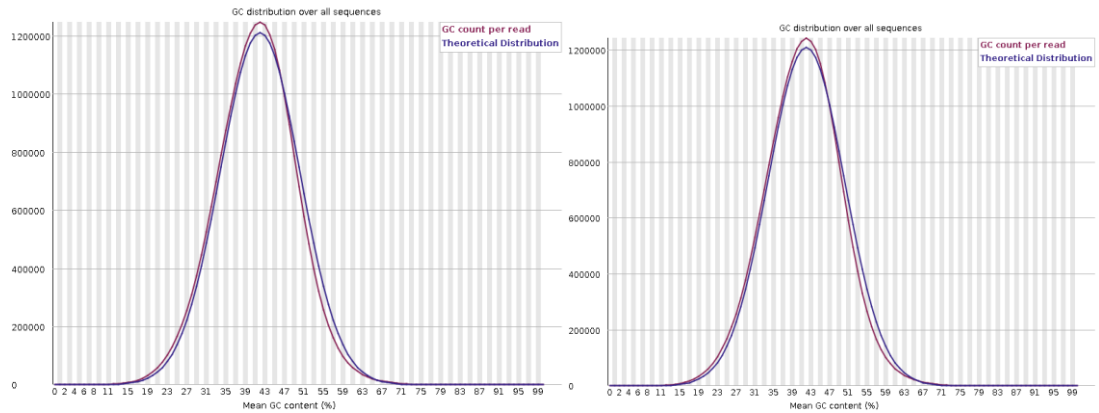
En cuanto a las barras de error, estos alcanzan puntajes de calidad altos, superiores a 30, a partir de las 94 pb para forward y 150 pb para reverse, en términos de PHRED Score, que es una medida de calidad de las bases generadas en la secuenciación que permite evaluar la calidad de la secuencia, se observa que el nivel superior a 30 muestra una precisión de la base el 99.9% y la probabilidad de llamada de base incorrecta de 1 en 1000, en cuanto al puntaje de calidad medio dada por línea azul se

encuentran entre 34 y 36 para forward y reverse respectivamente (Babraham, 2020).



**Figura 7.** Calidad de la secuencia por par de base para las secuencias forward (derecha) y reverse (izquierda)

En cuanto a las gráficas que muestran el contenido de G-C (figura 8) tienen una distribución normal, no presentan picos agudos o anchos que pudieran indicar que existe contaminación por dímeros u otros compuestos, por lo que al observarse picos centrales con distribución normal, y contrarrestándolo con la literatura, inferimos que no existe contaminación (Babraham, 2020).



**Figura 8.** Contenido de GC para las secuencias, forward (izquierda) y reverse (derecha)

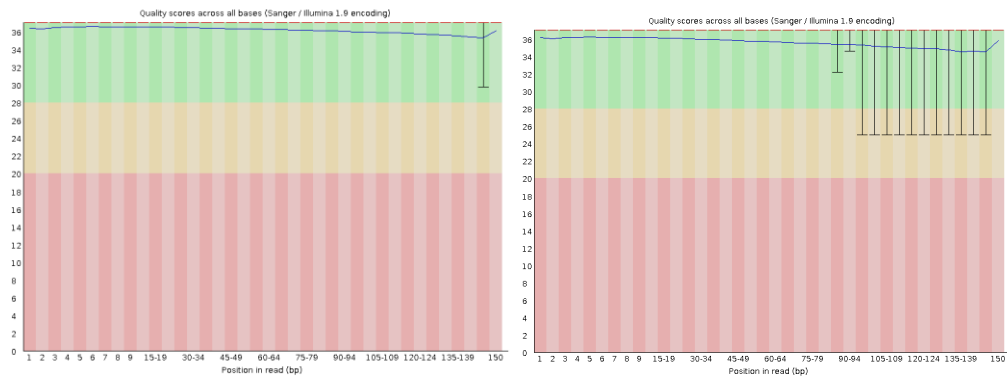
## 6.1.2 Limpieza de secuencias

Para mejorar la calidad de las secuencias, se empleó la herramienta *Trimmomatic*, con la finalidad de eliminar las lecturas de baja calidad y posterior a ello se volvió evaluar la calidad de las secuencias acorde a se indicó en el punto anterior (figura 9).

**Tabla 5.** Estadística básica posterior a la limpieza en trimmomatic utilizando FastQC (forward y reverse)

Measure	Value	Measure	Value
Filename	PE1_2TRIMMOMATIC_paired_151985_5_1.fwd.fastq	Filename	PE1_2TRIMMOMATIC_paired_151985_5_1.rev.fastq
File type	Conventional base calls	File type	Conventional base calls
Encoding	Sanger / Illumina 1.9	Encoding	Sanger / Illumina 1.9
Total Sequences	24570279	Total Sequences	24570279
Total Bases	3.6 Gbp	Total Bases	3.5 Gbp
Sequences flagged as poor quality	0	Sequences flagged as poor quality	0
Sequence length	50-150	Sequence length	50-150
%GC	40	%GC	40

Posterior a la limpieza, se observan mejoras en los puntajes de calidad, sin embargo, no existen cambios en cuanto a las puntuaciones de calidad por lectura.



**Figura 9.** Calidad de la secuencia por de base, posterior a la limpieza en *Trimmomatic*, forward (derecha) y reverse (izquierda)

### 6.1.3 Eliminación de duplicados

Para la eliminación de secuencias duplicadas, se empleó la herramienta Dedupe, de BBDMap, en bash de Linux, con ello se elimina las secuencias que comparten porcentajes de identidad que sean considerados como duplicados, estos se corresponden con el 10.45%.

```
at jgi.Dedupe.main(Dedupe.java:85)
geo@geo-Lenovo-Flex-2-15:~/bbmap$ dedupe.sh -Xmx4g in1=/home/geo/Descargas/Shotgun/C_Pareada_1.fq \ in2=/home/geo/Descargas/Shotgun/C_Pareada_2.fq out=/home/geo/Descargas/Shotgun/C_adup.fq ac=f
java -Djava.library.path=/home/geo/bbmap/jni/ -ea -Xmx4g -Xms4g -cp /home/geo/bbmap/current/ jgi.Dedupe -Xmx4g in1=/home/geo/Descargas/Shotgun/C_Pareada_1.fq out=/home/geo/Descargas/Shotgun/C_adup.fq ac=f
Executing jgi.Dedupe [-Xmx4g, in1=/home/geo/Descargas/Shotgun/C_Pareada_1.fq, in2=/home/geo/Descargas/Shotgun/C_Pareada_2.fq, out=/home/geo/Descargas/Shotgun/C_adup.fq, ac=f]
Version 38.34

Initial:
Memory: max=4294m, total=4294m, free=4288m, used=6m

Exception: java.lang.OutOfMemoryError thrown from the UncaughtExceptionHandler in thread "Thread-4"
Found 667658 duplicates.
Finished exact matches. Time: 529.290 seconds.
Memory: max=4294m, total=4294m, free=5m, used=4289m

Input:          6390117 reads          944602795 bases.
Duplicates:    667658 reads (10.45%) 100127957 bases (10.60%)      229489 collisions.
Result:        5718972 reads (89.50%) 844474838 bases (89.40%)

Exception in thread "main" java.lang.OutOfMemoryError: Java heap space
geo@geo-Lenovo-Flex-2-15:~/bbmap$
```

**Figura 10.** Eliminación de secuencias mediante la herramienta Dedupe de BBDMap

Posterior a la eliminación de secuencias duplicadas, se utilizó la herramienta Reformat de BBDMap, para reformatear las lecturas, en el cual se observa que el total de las secuencias se corresponden a lecturas que no mantienen duplicidad.

```
geo@geo-Lenovo-Flex-2-15:~/bbmap$ reformat.sh in=/home/geo/Descargas/Shotgun/C_adup.fq out=/home/geo/Descargas/Shotgun/C_adup.fq
java -ea -Xmx200m -cp /home/geo/bbmap/current/ jgi.ReformatReads in=/home/geo/Descargas/Shotgun/C_adup.fq out=/home/geo/Descargas/Shotgun/C_adup.fq
Executing jgi.ReformatReads [in=/home/geo/Descargas/Shotgun/C_adup.fq, out1=/home/geo/Descargas/Shotgun/C_adup.fq, out2=/home/geo/Descargas/Shotgun/C_adup.fq]

Set INTERLEAVED to true
Input is being processed as paired
Input:          667658 reads          100127957 bases
Output:         667658 reads (100.00%) 100127957 bases (100.00%)

Time:          1.666 seconds.
Reads Processed: 667k 400.84k reads/sec
Bases Processed: 100m 60.11m bases/sec
geo@geo-Lenovo-Flex-2-15:~/bbmap$
```

**Figura 11.** Datos obtenidos en Reformat de BBDMap

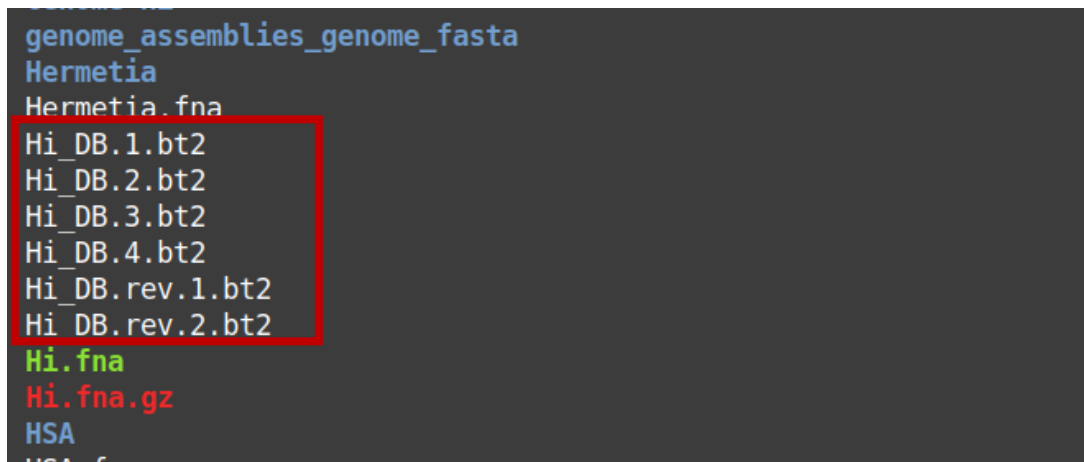
#### 6.1.4 Eliminación de contaminación de la secuencia

Se obtuvo del NCBI el genoma de *Hermetia illucens*, con código de acceso GCF\_905115235.1 en formato FASTA (figura 12).

```
>LR899009.1 Hermetia illucens genome assembly, chromosome: 1
atctcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaaatcaggt
tccttttcattatgtgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgatttttcagcttttcctcca
tttcgcatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaaatcaagt
tccttttcaatattgtgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgatttttcagcttttcctcca
tttcgcatcatcaaatcacaggctatacacttgctctgagacgctgataccctcagatctcaagatcacaaatcaggt
tccttttcattatgtgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgatttttcagcttttcctcca
tttcgcatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaaatcaggt
tacttttcattatgtgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgatttttcagcttttcctcca
tttcgcatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaaatcaggt
tccttttcattatgtgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgatttttcagcttttcctcca
tttcgcatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaaatcaggt
tccttttcattatgtgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgatttttcagcttttcctcca
atcttcgcatcatcaaatcacaggctatacacttgctctcagacgctgacaccctcagatctcaagatcacaaatcaggt
ttccttttcaatattgtgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgatttttcagcttttcctcca
```

**Figura 12.** Genoma de *Hermetia illucens* en formato FASTA

Una vez descargado el genoma de *H. Illucens*, se creó una base de datos de seis archivos (figura 13).



**Figura 13.** Base de datos creada con base al genoma de *Hermetia illucens*

Mediante la herramienta Bowtie2 se mapeó y alineó las lecturas generadas que se corresponden a 333.829 lecturas, de las cuales el 55.20% están

alineadas concordantemente y el 2.72% discordantemente. De las 179.262 pares alineados el 85.93% está emparejado. La tasa de alineación general es del 53.86% (Figura 14).

```
333829 reads; of these:
 333829 (100.00%) were paired; of these:
   184271 (55.20%) aligned concordantly 0 times
   111120 (33.29%) aligned concordantly exactly 1 time
   38438 (11.51%) aligned concordantly >1 times
  ----
   184271 pairs aligned concordantly 0 times; of these:
     5009 (2.72%) aligned discordantly 1 time
  ----
   179262 pairs aligned 0 times concordantly or discordantly; of these:
   358524 mates make up the pairs; of these:
     308062 (85.93%) aligned 0 times
     30093 (8.39%) aligned exactly 1 time
     20369 (5.68%) aligned >1 times
53.86% overall alignment rate
```

**Figura 14.** Lecturas mapeadas y no mapeadas de *Hermetia illucens*

### 6.1.5 Transformación a Fastq

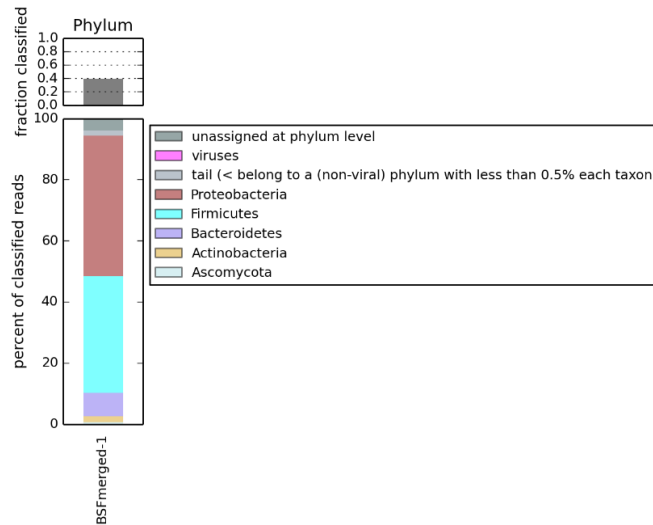
La obtención de archivos en formato FASTQ de las secuencias, se realizó mediante la herramienta bedtools

## 6.2 Ensamblaje y Anotación del genoma

### 6.2.1 Clasificación Taxonómica

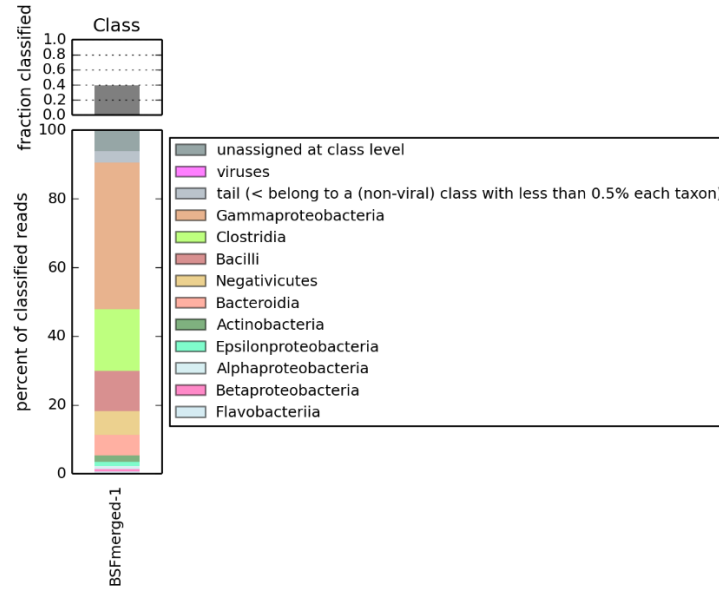
La clasificación taxonómica se realizó con la herramienta *Kaiju* v1.7.3, el cual es un programa que realiza clasificación taxonómica de lecturas de secuenciación de alto rendimiento, a partir de datos metagenómicos o metatranscriptómicos, cada lectura se asigna a un taxón disponible en el NCBI, y compara este taxón con la base de datos de referencia, los resultados son bastante sensibles y robustos, considerando que las comparaciones se realizan con nucleótidos, genomas completos o partes de las bases de datos de proteína (Menzel, Ng, & Krogh, 2016).

Considerando que los datos se corresponden a la microbiota del tracto intestinal de la BSF, los resultados de clasificación taxonómica a nivel de filo, clase, orden, familia, género y especies se observan en las figuras 15, 16, 17, 18 y 19 respectivamente, en cuanto a los microorganismos que tienen mayor abundancia a nivel de filo son Proteobacteria, Firmicutes, Bacteroides y Actinobacteria como se puede observar en la figura 15.



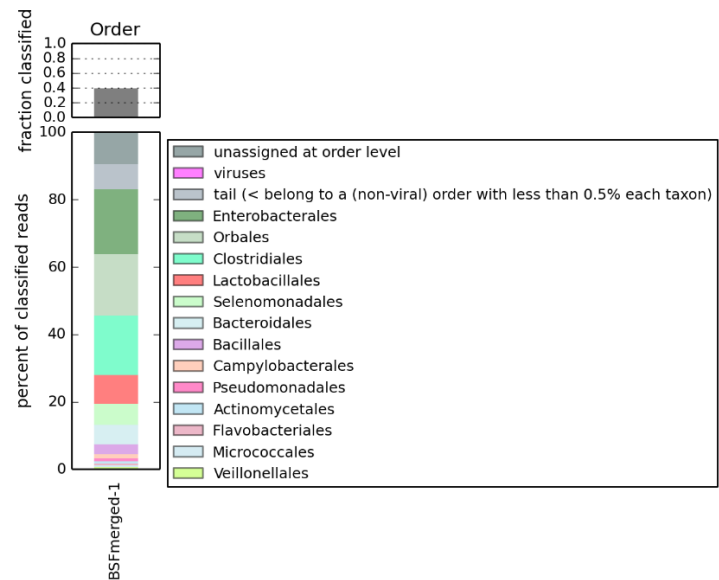
**Figura 15.** Clasificación taxonómica a nivel de Filo

Los microorganismos más abundantes a nivel de clase son Gammaproteobacteria, Clostridia, Betaproteobacteria, Negativicutes, Bacteroidia, Actinobacteria y Alphaproteobacteria, según se indica en la figura 16.



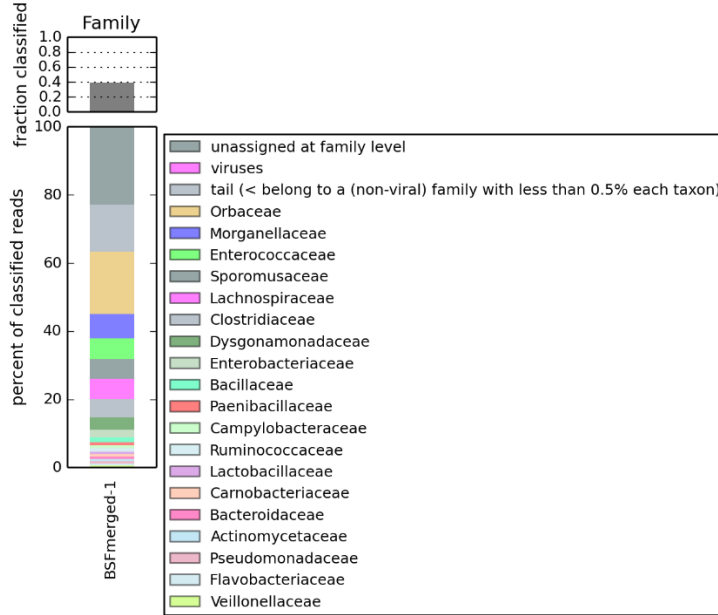
**Figura 16.** Clasificación taxonómica a nivel de Clase

En cuanto a orden, los microorganismos que presentan mayor abundancia son Enterobacteria, Orbitales, Clostridiales, Lactobaciliales y Actinomycetales, acorde se indica en la figura 17.



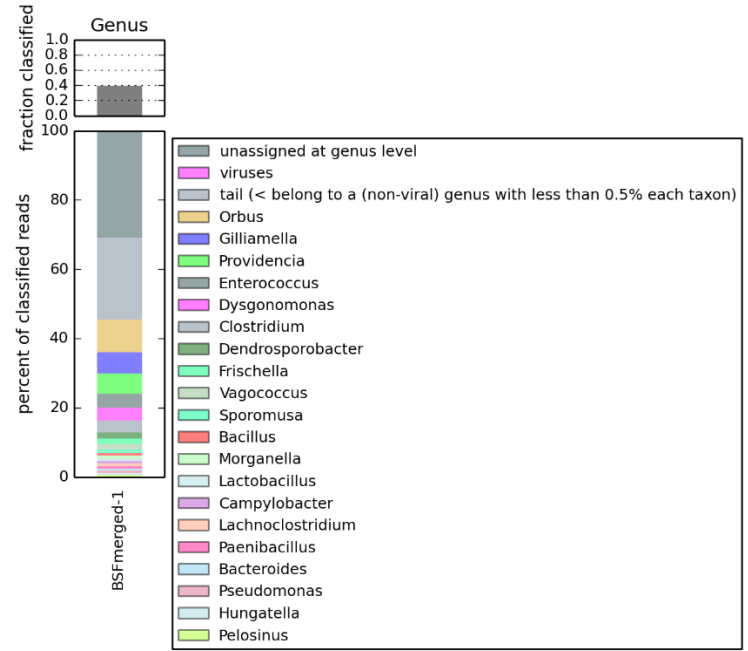
**Figura 17.** Clasificación taxonómica a nivel de Orden

A nivel de familia, los microorganismos que se encuentran son Orbaceae, Morganellaceae, Enterococcaceae, Sporomusaceae, Dysgonomonadaceae, entre otros en menor proporción (figura 18).



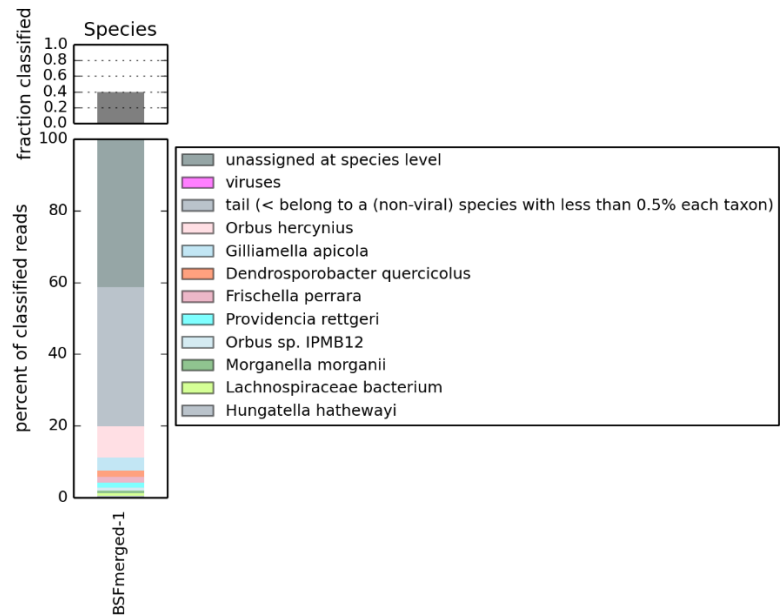
**Figura 18.** Clasificación taxonómica a nivel de Familia

A nivel de género los microorganismos que se encuentran son *Orbus*, *Gilliamella*, *Providencia*, *Enterococcus*, *Dysgonomonas*, *Clostridium*, *Dendosporobacter*, *Frishella*, entre otros en menor proporción, según se indica en la figura 19.



**Figura 19.** Clasificación taxonómica a nivel de Género

Los microorganismos a nivel de especie que se observan en la figura 20 son *Orbus hercynius*, *Gilliamella apicola*, *Dendrosporobacter quercicolus*, *Frichela perrara*, *Providencia rettgeri*, entre otros en menor proporción.



**Figura 20.** Clasificación taxonómica a nivel de Especie

Los microorganismos reportados por Zhineng et al., (2021) se corresponden a los datos presentados en las figuras anteriores, es así que la literatura reporta a los géneros *Providencia*, *Enterococcus*, *Frischella*, *Dysgonomonas*, *Gilliamella* y *Orbus* entre los más abundantes, y *Frischella perrara*, *Gilliamella apicola*, *Enterococcus sp.* y *Providencia rettger* como los más abundantes a nivel de especie.

### **6.2.2 Ensamblaje del Genoma y Comparación**

Se ensambló el genoma con cuatro programas, metaSPAdes, IDBA, MEGAHIT y HipMER, posterior a ello, se evaluaron los cuatro ensamblajes en el programa Compare Assembled Contig Distributions (KB).

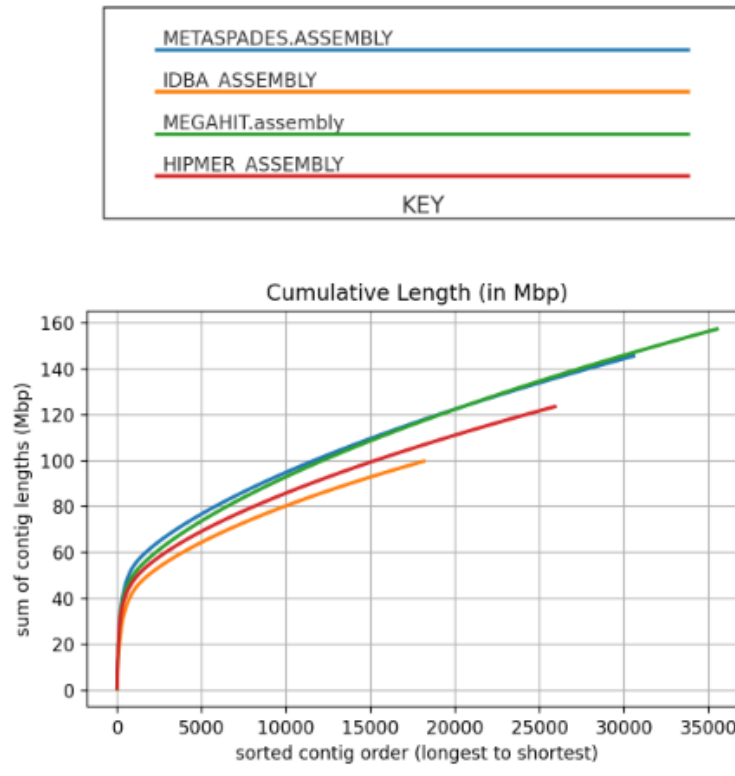
KB permite comparar los ensamblajes con base a la longitud y la distribución de los contigs, considerando que mientras un contig es más largo es más deseable, producto de la valoración se observa un histograma con las mejores longitudes de contig (figura 21) y una tabla vertical con los cuatro ensamblajes (figura 22)

En la figura 20 se observa un diagrama de los cuatro ensamblajes, coloreadas en color azul (mejor ensamblaje) y rojo (peor ensamblaje), siendo el mejor ensamblaje el realizado con metaSPAdes.



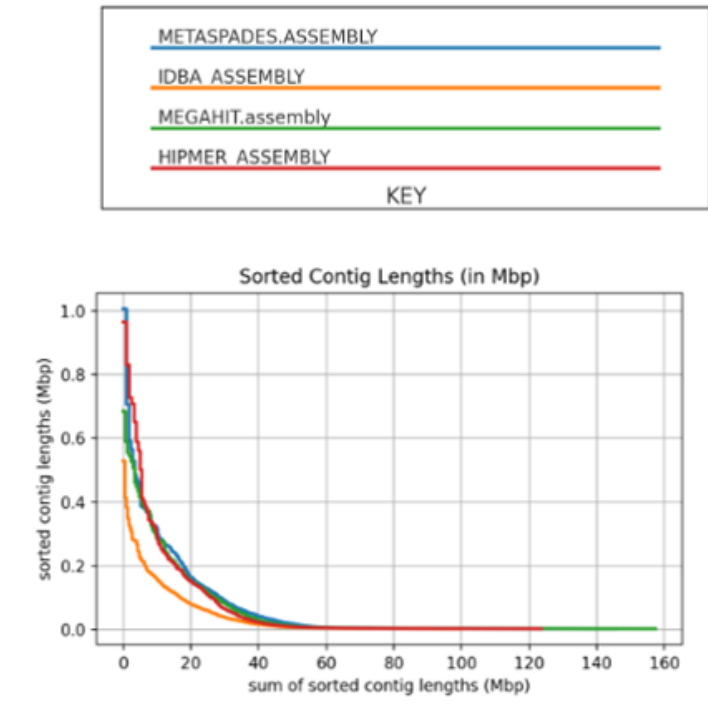
**Figura 21.** Evaluación de los ensambladores metaSPAdes, IDBA, MEGA HIT y HipMER

En la figura 22 se observa en el eje Y la suma acumulada y en el eje X las longitudes de contigs, los contigs se encuentran ordenados del más largo a más corto y la escala es el porcentaje de contigs representados. De la gráfica se observa que IDBA y HIPMER se encuentran por debajo de los 30.000 mbp, y metaSPAdes e IDB superiores a los 30.000 siendo los mejores.



**Figura 22.** Longitudes de contigs en el eje X y suma acumulada en el eje Y, con escala de porcentaje.

En la figura 23 se observa en el eje X la suma acumulada y en el eje Y las longitudes de contigs, a diferencia del gráfico anterior, se usa la suma acumulada en lugar de porcentaje de contigs. De la gráfica se observa en el eje X que metaSPAdes mantiene la longitud más larga, siendo superior el ensamblador metaSPAdes sobre el de IDB.



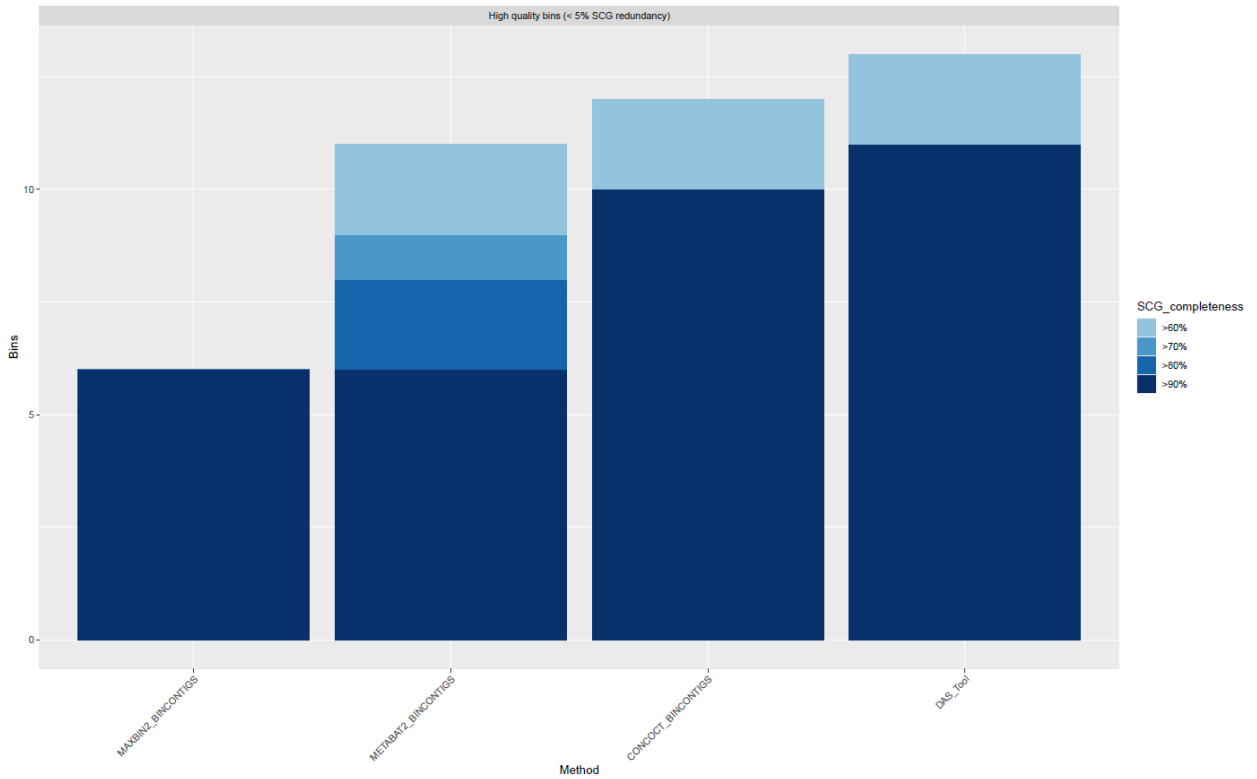
**Figura 23.** Longitudes de contigs en el eje X y suma acumulada en el eje Y

### 6.2.3 Agrupación y Optimización de Bins

La agrupación de contigs se realizó con los softwares MaxBin2, CONCOCT y MetaBat2, en cuanto a la optimización se realizó con DASTool.

MaxBin agrupa secuencias metagenómicas ensambladas en función de un algoritmo de maximización de expectativas (EM), utiliza información de composición de nucleótidos, abundancia de cepas de origen y genes marcadores filogenéticos para realizar las agrupaciones. CONCOCT agrupa secuencias metagenómicas en función de la composición de nucleótidos y la información de profundidad de cobertura y MetaBat2, agrupa secuencias metagenómicas utilizando la composición de nucleótidos y la abundancia, con base a la medida por profundidad de cobertura.

Con MaxBin se obtuvo 20 bins, y 20.448 contigs agrupados, con CONCOCT se obtuvo 76 bins, y 18.383 contigs agrupados mientras que con MetaBat2 se obtuvo 29 bins y 15.289 contigs agrupados, en la figura 24 se observa la comparación de los tres métodos de agrupación utilizados y la optimización con DASTools.



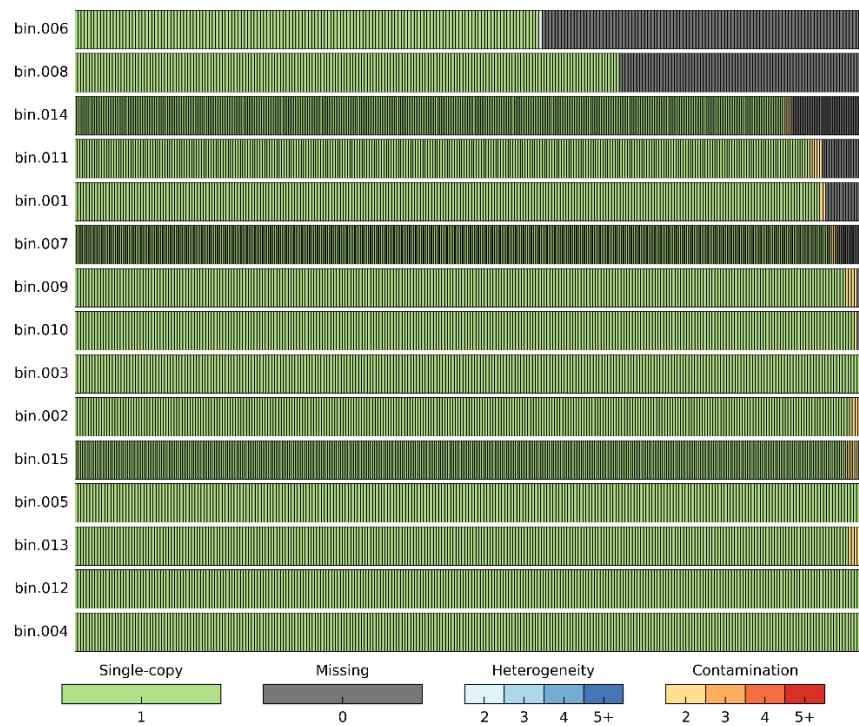
**Figura 24.** Comparación de los métodos de agrupación, de izquierda a derecha se encuentra Max Bin2, MetaBat y Concoct, y la optimización mediante DASTool

#### 6.2.4 Evaluación de la calidad de los bins

Para evaluar la calidad de los bins se utilizó CheckM, este programa evalúa la calidad de los genomas o metagenomas, con base a estimaciones de integridad y contaminación, empleando un conjunto de genes y copias de linajes filogenéticos, el programa no evalúa bins que no superen los umbrales y que mantengan contaminación.

La calidad de los bins evaluados se muestra en la figura 25, los bins 003, 004 y 008 corresponden a *Dysgonomonas* sp. de los tres bins se observa que el bin 004 es el que tiene mejor calidad, lo cual se aprecia en el dominio total de las barras de color verde, no existen secuencias perdidas (ausencia de barras de color plomo) y la contaminación es casi mínima (barras de color amarillo-naranja).

En cuanto al bin 003, su calidad es muy buena, sin embargo existe un poco de pérdida de información lo cual se observa en la presencia de barras en color plomo, el bin 008 por el contrario mantiene una gran cantidad de pérdida de información (barras de color plomo).



**Figura 25.** Calidad y contaminación de bins

Genoma de usuario ▲	Clasificación ▼
bin.001.fasta_BINUNTIL	d_Bacterias; p_Proteobacteria; c_Gammaproteobacteria; o_Enterobacterales; f_Enterobacteriaceae; gramo__; s__
bin.002.fasta_BINUNTIL	d_Bacterias; p_Firmicutes_C; c_negativos; o_Selenomonadales_A; f_Dendrosporobacteraceae; g_Dendrosporobacter; s__
bin.003.fasta_BINUNTIL	d_Bacterias; p_Bacteroidota; c_Bacteroidia; o_Bacteroides; f_Disgonomonadaceae; g_Disgonomonas; s__
bin.004.fasta_BINUNTIL	d_Bacterias; p_Bacteroidota; c_Bacteroidia; o_Bacteroides; f_Disgonomonadaceae; g_Disgonomonas; s__
bin.005.fasta_BINUNTIL	d_Bacterias; p_Firmicutes_A; c_Clostridia; o_Lachnospirales; f_UBA5962; g_UBA5962; s__
bin.006.fasta_BINUNTIL	d_Bacterias; p_Firmicutes_A; c_Clostridia; o_Peptoestreptococos; f_Anaerovoracaceae; gramo__; s__
bin.007.fasta_BINUNTIL	d_Bacterias; p_Proteobacteria; c_Gammaproteobacteria; o_Pseudomonadales; f_Pseudomonadaceae; g_Entomomonas; s__
bin.008.fasta_BINUNTIL	d_Bacterias; p_Bacteroidota; c_Bacteroidia; o_Bacteroides; f_Disgonomonadaceae; g_Disgonomonas; s__
bin.009.fasta_BINUNTIL	d_Bacterias; p_Firmicutes_A; c_Clostridia; o_Lachnospirales; f_Lachnospiraceae; gramo__; s__
bin.010.fasta_BINUNTIL	d_Bacterias; p_Firmicutes_A; c_Clostridia; o_Lachnospirales; f_Lachnospiraceae; g_CHH4-2; s__

**Figura 26.** Clasificación Taxonómica posible mediante Check M

En la figura 26 se observa la taxonomía para todos los bins, de los cuales los bins 003, 004 y 008 fueron identificados como *Dysgonomonas* sp., como se indicó anteriormente, el bin 004 presenta mayor integridad en su secuencia y ausencia de contaminación, por lo que se seleccionó para los análisis posteriores.

### 6.2.5 Anotación del genoma

Se utilizó la herramienta RAST para la anotación del genoma del BIN 004, de la cual el código genético es 11 es el predeterminado para los genomas de bacterias y arqueasa, acorde se indica en la tabla 6, dentro de su clasificación taxonómica se observa Dominio Bacteria, Filo Bacteroidota, Clase Bacteroidia, orden Bacteroidales, Familia Dysgonomonadaceae, Género *Dysgonomonas* Especie desconocida (tabla 6).

**Tabla 6.** Resultado de la anotación del genoma en Kbase

KBase Object Name	bin.004.fasta_BINUNTIL_RAST	Taxonomy
Scientific Name	unknown taxon	d__Bacteria
Domain	B	p__Bacteroidota
Genetic Code	11	c__Bacteroidia
Source	KBase	o__Bacteroidales
Source ID	bin.004.fasta_BINUNTIL_RAST	f__Dysgonomonadaceae
Size	3,071,665	g__Dysgonomonas
		s__

Así también se realizó la anotación del genoma mediante BV-BRC de Patric, en el cual se obtuvo la clasificación taxonómica, Filo Bacteroidota, Clase Bacteroides, Orden Bacteroidales, Familia Dysgonomonadaceae y Género *Dysgonomonas*, acorde a la tabla 7.

**Tabla 7.** Resultado de la anotación del genoma y clasificación taxonómica en BV-BRC

Disgonomonas Bin4 <span style="float: right;">Edit</span>	
Longitud: 3071665bp, Contigs: 21	
Información general	
identificación del genoma	156973.5
Nombre del genoma	Disgonomonas Bin4
Información de taxonomía	
ID de taxón	156973
superreino	bacterias
Filo	Bacteroidota
Clase	Bacteroides
Orden	Bacteroidales
Familia	Dysgonomonadaceae
Género	Dysgonomonas

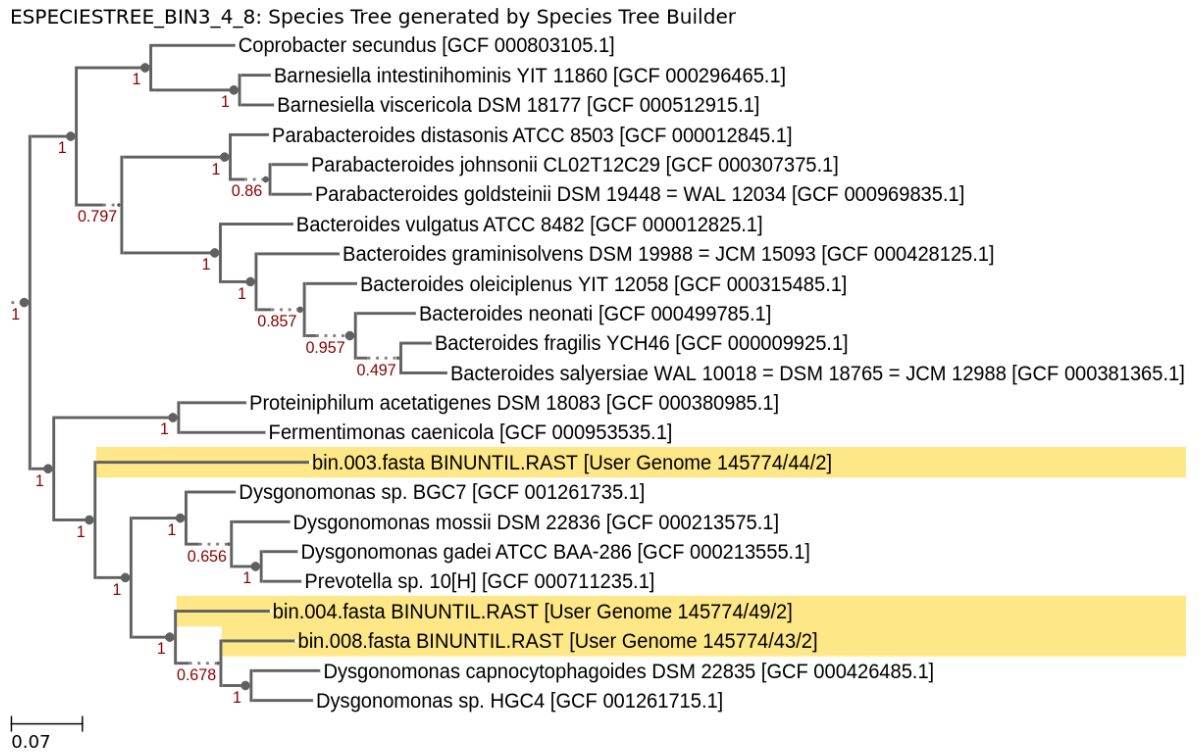
### 6.2.6 Clasificación Taxonómica

A continuación se observa el árbol filogenético generado por *Species tree*, el cual permite construir un árbol de especies utilizando un conjunto de 49 genes universales centrales definidos por familias de genes COG (Clusters of Orthologous Groups), este combina los BIN 003, 004 y 008 con los cuales hemos venido trabajando, con un conjunto de genomas estrechamente relacionados, seleccionados de la importación de genomas KBase públicos de RefSeq, la relación que se generó está determinada por la similitud de alineación con un subconjunto seleccionado de 49 dominios COG, los genomas (bins) se insertan en la alineación de secuencia múltiple (MSA) seleccionada para cada familia COG, cuyas alineaciones se han recortado con GBLOCKS para eliminar las secciones mal alineadas del MSA y luego, los MSA se concatenan y se reconstruye el árbol filogenético usando el método para estimar rápidamente la filogenia de máxima verosimilitud aproximada.

Como producto obtuvimos los microorganismos más cercanos al BIN 003, 004 y 008 que son *Dysgonomonas* sp. *Dysgonomonas mossi*, *Dysgonomonas gadei* y *Dysgonomonas capnocytophagoides*, Kita et al., (2015) reportó siete especies de *Dysgonomonas* sp. entre ellas *D. hofstadii*, *D. oryzaevi*, *D. macrotermidis*, *D. termitidis*, *D. mossi*, *D. Gadei* y *D. capnocytophagoides*, siendo esta última en conjunto con *D. macrotermidis* y *D. termitidis* especies aisladas del intestino posterior de una termita.

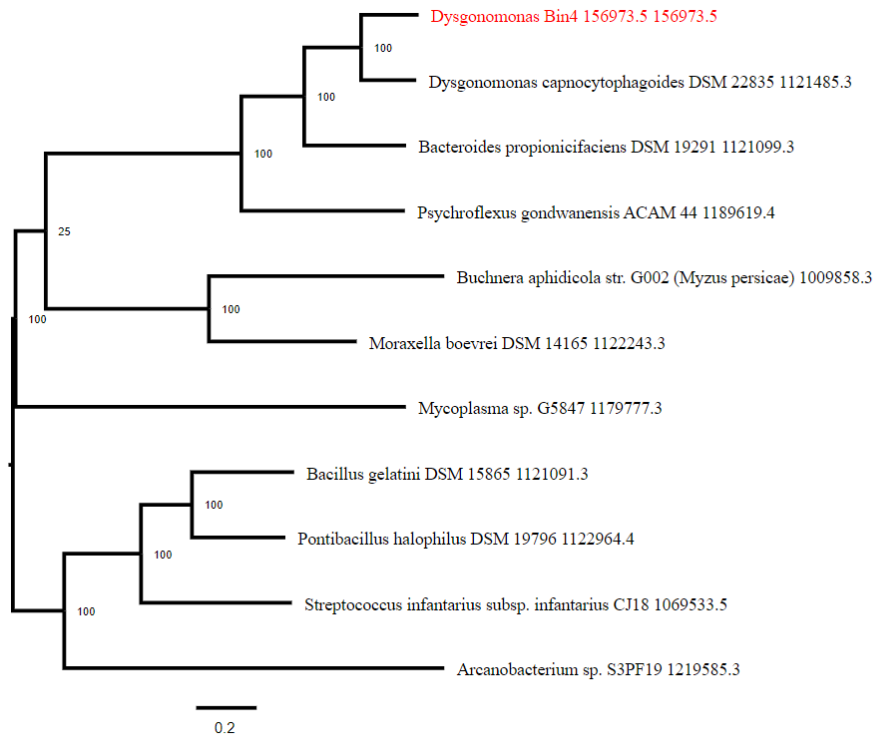
Así también, Kamis et al (2020) reportó a *Dysgonomonas* sp. como parte del microbioma intestinal de la mosca soldado negra, del total de la microbiota intestinal basado en 188.863 secuencias, el 11% se correspondían a *Dysgonomonas* sp. estando entre los géneros más abundantes y procedentes de muestras de América.

Con base a los resultados taxonómicos obtenidos en Kbase y BV-BRC , los datos de Kaiju y RASTtk podemos inferir que la bacteria del bin 004 pertenece al género *Dysgonomonas*. (figura 27).



**Figura 27.** Árbol filogenético para los bin 003, 004 y 008 con Species Tree en Kbase

Así también se elaboró el árbol filogenético del bin 004 en BV-BRC el cual se presenta en la figura 28 y se encuentra relacionado con *Dysgonomonas capnocytophagoides*



**Figura 28.** Árbol filogenético para el bin 004 con BV-BRC en Patric

## 6.3 Búsqueda de Genes y Rutas Metabólicas de Interés

### 6.3.1 Perfil Funcional del microorganismo

Los resultados de las rutas metabólicas se realizaron con la herramienta DRAM, la cual predice secuencias de codificación y proporciona resúmenes metabólicos del genoma, este programa anota un conjunto de genomas y el producto es usado para comparar perfiles metabólicos.

Producto del análisis, obtuvimos, la tabla estadística del genoma trabajado, el resumen metabólico que indica los recuentos de genes funcionales en una amplia variedad de metabolismos y un mapa de calor interactivo (figura 29) muestra la cobertura de los módulos, la cobertura de los componentes de la cadena de transporte de electrones y la presencia de funciones metabólicas.

Con base a las figuras 29 y 30, se aprecia dentro de las rutas metabólicas observadas, la ruta de fermentación de la glucosa llamada Entner-Doudoroff, la cual es una alternativa secundaria a la glucólisis, pues cataboliza la glucosa a piruvato mediante enzimas distintas a la de la glucólisis y a la de pentosa fosfato.

Así también se ha reportado a este microorganismo como parte de la microbiota intestinal de las termitas, en donde participa en las rutas de degradación de lignocelulosa recalcitrante (Yang et al., 2014 & Bruno et al., 2019) y en la de polisacáridos complejos,

Jiang et al. (2019), relaciona a *Dysgonomonas* sp. con rutas metabólicas de sulfatos, carbohidratos y nitrógeno, acorde a la figura 30. se observa que producto del perfil funcional de nuestro microorganismo, se encuentran las rutas metabólicas asociadas a la degradación de sulfatos y de nitrógeno. Por otro lado, permite la oxidación del piruvato para carbohidratos es la del Ciclo del citrato o de Krebs (KEGG, 2023), el cual se encuentra reportado dentro de nuestra información.

Así también, *Dysgonomonas* sp. se encuentra relacionada con la degradación de  $\alpha$ -galactosas, producto del gen  $\alpha$ -galactosidasa, lo cual explica la degradación de carbohidratos vegetales no digeribles (Shelomi et al. 2020). Por lo otro esta bacteria contribuye a la degradación de medicamentos como el ciprofloxacino en conjunto con *Actinomyces* (Lee et al., 2022).

En cuanto a la cadena de transporte de energía, se observa la de citocromo bd ubiquinol oxidasa que participa en la cadena respiratoria de procariotas, en la cual acopla la reducción del oxígeno molecular al agua mediante fuerzas de protones, empleando ATP (Borisov et al., 2021)



### 6.3.2 Búsqueda de genes relacionados con hidrolisis de PET

Danso et al (2018) reporta genes asociados a la degradación de PET, Cut190 (W0TJ64), cut1 (E9LVI0), cut-2 (E5BBQ3), Tcur\_1278 (D1A9G5), cut1 (E9LVH7), cut (H6WX58), cut2 (E9LVH9), y ISF6\_4831 (A0A0K8P6T7), los cuales se encuentran detallados en la tabla 8.

**Tabla 8.** Genes que intervienen en la hidrolisis de PET.

Sequence no.	PDB entry no. <sup>a</sup>	Gene name <sup>b</sup>	Organism	Reference
1	W0TJ64	Cut190	<i>Saccharomonospora viridis</i>	6
2	E9LVI0	cut1	<i>Thermobifida fusca</i> ( <i>Thermomonospora fusca</i> )	7
3	E5BBQ3	cut-2	<i>Thermobifida fusca</i> ( <i>Thermomonospora fusca</i> )	8
4	D1A9G5	Tcur_1278	<i>Thermomonospora curvata</i>	9
5	E9LVH7	cut1	<i>Thermobifida alba</i>	10
6	H6WX58	NA	<i>Thermobifida halotolerans</i>	11
7	E9LVH9	cut2	<i>Thermobifida cellulosityca</i>	12
8	A0A0K8P6T7	ISF6_4831	<i>Ideonella sakaiensis</i>	13
9	G9BY57	NA	Uncultured bacterium	14

Para poder trabajar con estos genes, se utilizó el programa KB (Build FeatureSet from Genome) en donde se construyó un solo grupo de funciones para su asociación con nuestra secuencia.

El grupo de genes reportados en la tabla 8 se sometió a un alineamiento múltiple (MSA) mediante la herramienta MUSCLE, este se almacena como un objeto de datos que contiene las secuencias de la alineación y se utiliza para análisis de búsqueda de genes de interés.

Gblocks recorta las regiones que no se consideran confiables, manteniendo los bloques conservados para la búsqueda de funciones del genoma ensamblado.

HMMER es una aplicación que realiza una búsqueda de modelo oculto de Markov (HMM) en secuencias de proteína utilizando modelos HMMER que fueron creados a partir de una alineación de secuencia múltiple (MSA), esta alineación de proteínas la convierte en un modelo oculto de Markov el cual se usa para buscar una base de datos de proteínas.



Danso et al (2018) indica que entre las enzimas que degradan PET se encuentran  $\alpha$ - y  $\beta$ -hidrolasas, cutinasas, serina hidrolasas, lipasas, carboxilesterasas, entre otras. Identificó 249 hidrolasas PET putativas, y 504 nuevas enzimas que podrían estar relacionadas a las hidrolasas, las cuales se encuentran en Actinobacterias, Proteobacterias y Bacteroidetes, siendo el filo Bacteroidetes quienes se encuentran vinculadas principalmente con genes de hidrolasas de PET.

Es importante indicar que la información que proporciona esta herramienta es una tabla de secuencias de aciertos, en la cual se presentan únicamente los datos que están por encima del umbral y en cuanto a los valores que se encuentran fuera de este rango, no se consideran.

En la tabla 9 se observa el ID de los genes y las funciones asociadas al BIN 004, se reportan dos coberturas de alineamiento, el GEN CDS 524 para hypothetical protein en un 43.8% y el GEN CDS 2290 para putative esterasa/lipasa con un 34%, encontrándose dos coincidencias que muestran la posible presencia de PETasas, considerando que Danso et al (2018) reportó a las enzimas lipasas e hidrolasas PET putativas entre las enzimas degradadoras de PET, así también vinculó estas enzimas al filo Bacteroidota como el más representativo en cuanto a contenido de PETasas, y dentro de cuya clasificación taxonómica está *Dysgonomas* sp.

El GEN CDS 524, se encuentra asociado a actividades enzimáticas de proteínas de la familia 1 de glicosil hidrolasa y deshidrogenasa (NCBI, 2023) y el GEN CDS 2290 a lisina desmetilasa 1<sup>a</sup> (NCBI, 2023).

**Tabla 9.** Taxonomía de ubicación más cercana obtenida

ALIGNMENT COVERAGE (HIT SEQ)	GENE ID	FUNCTION	GENOME	ALN_LEN	E-VALUE	BIT SCORE	H_BEG-H_END
	CDS.524	FIG00653406: hypothetical protein	bin.004.fasta_BINUNTIL.RAST	112 (43.8%)	5.4e-08	29.9	147-258
	CDS.2290	putative esterase/lipase	bin.004.fasta_BINUNTIL.RAST	87 (34.0%)	2.5e-05	21.2	22-108

## 7. CONCLUSIONES Y RECOMENDACIONES

### 7.1 Conclusiones

Mediante datos metagenómicos del tracto digestivo de la Mosca Soldado Negra *Hermetia illucens*, es posible ensamblar y anotar el genoma de *Dysgonomonas* sp., y realizar la búsqueda de rutas metabólicas y genes de interés.

Es posible realizar el control de calidad de las secuencias empleando fastQC, acompañado de Trimmomatic para mejorar la calidad de las secuencias y con ello eliminar secuencias duplicadas con Dedupe usando Bash de Linux

El mejor ensamblaje del genoma para *Dysgonomonas* sp. a partir de datos metagenómicos se realizó con metaSPAdes, en comparación con MEGA HIT, IDBA y HipMER.

Se identificó tres bins para *Dysgonomonas* sp. bin 003, 004 y 008 sin embargo, se observó mayor integridad y ausencia de contaminación para el bin 004.

Mediante RAST y BV-BRC se obtuvo la clasificación taxonómica para el bin 004, que se corresponde a Dominio Bacteria, Filo Bacteroidota, Clase Bacteroidia, orden Bacteroidales, Familia Dysgonomonadaceae, Género *Dysgonomonas* sin poder llegar a determinar especie.

Empleando Species Tree y BV-BRC se realizó la construcción de árboles filogenéticos, donde se observó que *Dysgonomonas capnocytophagoides* se encuentra estrechamente relacionado con el bin 004, y *Dysgonomonas sp.* *Dysgonomonas mossi*, *Dysgonomonas gadei*, con el bin 003 y 008.

*Dysgonomonas sp.* presenta rutas metabólicas para la fermentación de la glucosa llamada Entner-Doudoroff, rutas metabólicas de sulfatos, carbohidratos y nitrógeno, y el ciclo de Krebs, adicionalmente se encuentra relacionada con la degradación de  $\alpha$ -galactosas lo cual permite la degradación de carbohidratos vegetales. En cuanto a la cadena de transporte de energía, se observa la de citocromo bd ubiquinol oxidasa que participa en la cadena respiratoria de procariontas.

Utilizando una búsqueda de modelo oculto de Markov (HMM) en secuencias de proteína con modelos HMMER, se encontraron dos coincidencias con los genes asociados a la degradación de PET, el GEN CDS 524 para hypothetical protein en un 43.8% y el GEN CDS 2290 para putative esterasa/lipasa con un 34%, considerando como base a los genes asociados a degradación de PET, Cut190 (W0TJ64), cut1 (E9LVI0), cut-2 (E5BBQ3), Tcur\_1278 (D1A9G5), cut1 (E9LVH7), cut (H6WX58), cut2 (E9LVH9), y ISF6\_4831 (A0A0K8P6T7).

## 7.2 Recomendaciones

Se recomienda realizar la búsqueda de genes de hidrolisis de PET utilizando otros genes diferentes a los empleados en este trabajo.

Se recomienda realizar cultivos de *Dysgonomonas sp.* y realizar pruebas de actividad funcional considerando que se ha observado que esta bacteria aporta beneficios en la bioconversión de residuos, lignina, lignocelulosa y carbohidratos

de cadena larga, y se ha encontrado dos genes asociados a la actividad enzimática deseada.

Se recomienda realizar análisis de *Dysgonomonas* sp. en asociación con *Actinomyces* sp., considerando que se ha reportado la degradación de medicamentos por parte de estos dos microorganismos cuando actúan en conjunto.

## 8. REFERENCIAS BIBLIOGRÁFICAS

Aguilar-Bultet I, Laurent Falquet. (2015). Sequencing and de novo assembly of bacterial genomes: an approach to study new pathogens. Rev Salud Anim. vol.37 no.2 La Habana. ISSN 2224-4700

Ao, Y., Yang, C., Wang, S., Hu, Q., Yi, L., Zhang, J., et al. (2021). Características y función de los nutrientes de las comunidades bacterianas intestinales en larvas de mosca soldado negra (*Hermetia illucens* L.) en la conversión de estiércol de ganado. Microbio. Biotecnología. 14, 886–896. doi: 10.1111/1751-7915.13595

Aleluia, J.; Ferrão, P. Characterization of urban waste management practices in developing Asian countries: A new analytical framework based on waste characteristics and urban dimension. Waste Manag. 2016, 58, 415–429.

Allen, B. (n.d.). Microbial Genomics in KBase: Gene Feature Analysis. Retrieved September 23, 2022, from <https://narrative.kbase.us/narrative/83681>

Allen, B. (2021). Genomics in KBase: Identifying Features of Interest in Genomes. Retrieved July 15, 2022, from <https://narrative.kbase.us/narrative/83681#>

Auer, L., Lazuka, A., Sillam-Dussès, D., Miambi, E., O'Donohue, M. y Hernandez-Raquet, G. (2017). Descubriendo el potencial del microbioma intestinal de termitas para la

bioconversión de lignocelulosa en biorreactores anaeróbicos por lotes. *Frente. Microbiol.* 8:2623. doi: 10.3389/fmicb.2017.02623

Babraham Institute. (2020). FastQC: Documentation. Retrieved September 4, 2022, from <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Beskin, K. V., Holcomb, C. D., Cammack, J. A., Crippen, T. L., Knap, A. H., Sweet, S. T., & Tomberlin, J. K. (2018). Larval digestion of different manure types by the black soldier fly (Diptera: Stratiomyidae) impacts associated volatile emissions. *Waste Management*, 74, 213–220. <https://doi.org/10.1016/j.wasman.2018.01.019>

Bruno, D., Bonelli, M., De Filippis, F., Di Lelio, I., Tettamanti, G., Casartelli, M., ... Caccia, S. (2019). The intestinal microbiota of *Hermetia illucens* larvae is affected by diet and shows a diverse composition in the different midgut regions. *Applied and Environmental Microbiology*, 85(2), 1–14. <https://doi.org/10.1128/AEM.01864-18>

Bulak, P., Polakowski, C., Nowak, K., Waśko, A., Wiącek, D., & Bieganowski, A. (2018). *Hermetia illucens* as a new and promising species for use in entomoremediation. *Science of the Total Environment*, 633. <https://doi.org/10.1016/j.scitotenv.2018.03.252>

Boccazzi, IV, Ottoboni, M., Martin, E., Comandatore, F., Vallone, L., Spranghers, T., et al. (2017). Un estudio de la microbiota asociada con las larvas de la mosca soldado negra (*Hermetia illucens*) criadas para la producción de alimentos. *PLoS Uno* 12:e0182533. doi: 10.1371/journal.pone.0182533

Borisov, V. B., Siletsky, S. A., Paiardini, A., Hoogewijs, D., Forte, E., Giuffrè, A., & Poole, R. K. (2021). Bacterial Oxidases of the Cytochrome bd Family: Redox Enzymes of Unique Structure, Function, and Utility As Drug Targets. *Antioxidants and Redox Signaling*, 34(16), 1280–1318. <https://doi.org/10.1089/ars.2020.8039>

Caruso, D. Devic, E. Subamia, W. Talamond, P. Baras, E. (2013). Dwi Murti Nastiti, ed. Technical handbook of domestication and production of diptera black soldier fly (BSF) *Hermetia illucens*, Stratiomyidae (en inglés). IRD Editions & IPB.

Chivian, D., Clark, M., & Jungbluth, S. (2020). Genome Extraction from Shotgun Metagenome Sequence Data. Retrieved July 15, 2022, from <https://narrative.kbase.us/narrative/33233>

Choi, WH, Yun, JH, Chu, JP y Chu, KB (2012). Efecto antibacteriano de extractos de larvas de *Hermetia illucens* (diptera: stratiomyidae) contra bacterias gramnegativas. *Entomol. Res.* 42, 219–226. doi: 10.1111/j.1748-5967.2012.00465

Danso, D., Schmeisser, C., Chow, J., Zimmermann, W., Wei, R., Leggewie, C., ... Streit, W. R. (2018). New insights into the function and global distribution of polyethylene terephthalate (PET)-degrading bacteria and enzymes in marine and terrestrial metagenomes. *Applied and Environmental Microbiology*, 84(8). <https://doi.org/10.1128/AEM.02773-17>

De Smet, J., Wynants, E., Cos, P., & Van Campenhout, L. (2018). Microbial community dynamics during rearing of black soldier fly larvae (*Hermetia illucens*) and impact on exploitation potential. *Applied and Environmental Microbiology*, 84(9). <https://doi.org/10.1128/AEM.02722-17>

Del Hierro, A., Anrango, M., 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>

ESR International. (2016). Bio-Conversion of Putrescent Waste. Retrieved July 16, 2022, from

<https://arquivo.pt/wayback/20160516163727/http%3A//www.esrint.com/pages/bioconversion.html>

Ferronato, N., & Torretta, V. (2019). Waste mismanagement in developing countries: A review of global issues. *International journal of environmental research and public health*, 16(6), 1060.

Gustavsson, J., Cederberg, C., Sonesson, U., Van Otterdijk, R., Meybeck A. (2011). *Global Food Losses and Food Waste*. Food and Agriculture Organization of the United Nations. Rom <http://www.fao.org/docrep/014/mb060e/mb060e00.pdf>

Jiang, C. L., Jin, W. Z., Tao, X. H., Zhang, Q., Zhu, J., Feng, S. Y., ... Zhang, Z. J. (2019). Black soldier fly larvae (*Hermetia illucens*) strengthen the metabolic function of food waste biodegradation by gut microbiome. *Microbial Biotechnology*, 12(3), 528–543. <https://doi.org/10.1111/1751-7915.13393>

KBase. (2019). Compare Assembled Contig Distributions - v1.1.2. Retrieved January 6, 2023, from [https://kbase.us/applist/apps/kb\\_assembly\\_compare/run\\_contig\\_distribution\\_compare/release#:~:text=Compare Assembled Contig Distributions allows,contigs are typically more desirable.](https://kbase.us/applist/apps/kb_assembly_compare/run_contig_distribution_compare/release#:~:text=Compare Assembled Contig Distributions allows,contigs are typically more desirable.)

KEGG. (2023). KEGG PATHWAY. Retrieved January 12, 2023, from <https://www.genome.jp/kegg/pathway.html>

Kita A. , Iwasaki Y. , Yano S. , Nakashimada Y. , Hoshino T. , Murakami K. ( 2015 ;). *Dysgonomonas alginatilytica* sp. nov., una bacteria degradadora de alginato aislada de un consorcio microbiano. *Microbiology Society*

Khamis, F. M., Ombura, F. L. O., Akutse, K. S., Subramanian, S., Mohamed, S. A., Fiaboe, K. K. M., ... 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(July), 1–15. <https://doi.org/10.3389/fmicb.2020.01538>

Kruger, M. C., Bertin, P. N., Heipieper, H. J., & Arsène-Ploetze, F. (2013). Bacterial metabolism of environmental arsenic - Mechanisms and biotechnological applications. *Applied Microbiology and Biotechnology*, 97(9), 3827–3841. <https://doi.org/10.1007/s00253-013-4838-5>

Lee, Y.-G., Kim, C., Kuanyshev, N., Kang, NK, Fatma, Z., Wu, Z.-Y., et al. (2022). Ingeniería metabólica basada en Cas9 de *Issatchenkia orientalis* para una mejor utilización de hidrolizados celulósicos. *J. Agric. Química alimentaria* 70, 12085–12094. [doi: 10.1021/acs.jafc.2c04251](https://doi.org/10.1021/acs.jafc.2c04251)

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>

NCBI. (2020). *Hermetia illucens* reference genome iHerIII2.2.curated.20191125. Retrieved December 23, 2022, from [https://www.ncbi.nlm.nih.gov/search/all/?term=GCF\\_905115235.1\\_iHerIII2.2\\_curated.20191125](https://www.ncbi.nlm.nih.gov/search/all/?term=GCF_905115235.1_iHerIII2.2_curated.20191125)

Oliveira, F., Doelle, K., List, R., & Reilly, J. R. O. (2015). Assessment of Diptera: Stratiomyidae , genus *Hermetia illucens* ( L ., 1758 ) using electron microscopy. *Journal of Entomology and Zoology Studies*, 3(5), 147–152.

Osimani, A., Ferrocino, I., Corvaglia, M. R., Roncolini, A., Milanović, V., Garofalo, C. Clementi, F. (2021). Microbial dynamics in rearing trials of *Hermetia illucens* larvae fed coffee silverskin and microalgae. *Food Research International*, 140(July 2020). <https://doi.org/10.1016/j.foodres.2020.110028>

Pazmiño, M. (2021). Caracterización molecular y evaluación de la capacidad de degradación de microplástico de insectos del género *Hermetia* originarios del cantón Puerto Quito. Universidad de las Fuerzas Armadas ESPE.

Sangwan, N., Xia, F., & Gilbert, J. A. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome*, 4, 1–11. <https://doi.org/10.1186/s40168-016-0154-5>

Shelomi, M., Wu, M. K., Chen, S. M., Huang, J. J., & Burke, C. G. (2020). Microbes associated with black soldier fly (Diptera: Stratiomiidae) degradation of food waste. *Environmental Entomology*, 49(2), 405–411. <https://doi.org/10.1093/EE/NVZ164>  
T. Hofstad, I. Olsen, E.R. Eribe, E. Falsen, M.D. Collins, P.A. Lawson. *Dysgonomonas* gen. nov. to accommodate *Dysgonomonas gadei* sp. nov., an organism isolated from a human gall bladder, and *Dysgonomonas capnocytophagoides* (formerly CDC group DF-3). <http://dx.doi.org/10.1099/00207713-50-6-2189> | Medline

Soliz Torres, María Fernanda, ed., Durango Cordero, Juan Sebastián, Solano Peláez, José Luis y Yépez Fuentes, Milena Alía. Cartografía de los residuos sólidos en Ecuador, 2020. Quito, EC: Universidad Andina Simón Bolívar, Sede Ecuador / INEC / VLIR-UOS / GAIA / Alianza Basura Cero Ecuador / Acción Ecológica,

Wang, Y. S., & Shelomi, M. (2017). Review of black soldier fly (*Hermetia illucens*) as animal feed and human food. *Foods*, 6(10). <https://doi.org/10.3390/foods6100091>

Yang Y.J. , Zhang N. , Ji S.Q. , Lan X. , Zhang K.D. , Shen Y. L. , Li F. L. , Ni JF ( 2014 );. *Dysgonomonas macrotermitis* sp. nov., aislado del intestino posterior de una termita productora de hongos. *Int J Syst Evol Microbiol* 64 2956 – 2961

Yu Y, Zhang J, Zhu F, Fan M, Zheng J, Cai M, Zheng L, Huang F, Yu Z y Zhang J (2023) Degradación mejorada de proteínas por larvas de mosca soldado negra (*Hermetia*

illucens L.) y sus microbios intestinales . Frente. microbiol \_ 13:1095025. doi: 10.3389/fmicb.2022.1095025

Yu, G., Cheng, P., Chen, Y., Li, Y., Yang, Z., Chen, Y., et al. (2011). La inoculación de estiércol de aves de corral con bacterias acompañantes influye en el crecimiento y desarrollo de las larvas de la mosca soldado negra ( Diptera: Stratiomyidae ). Reinar. Entomol. 40, 30–35. doi: 10.1603/en10126

Zhan, S., Fang, G., Cai, M., Kou, Z., Xu, J., Cao, Y., ... Huang, Y. (2020). Genomic landscape and genetic manipulation of the black soldier fly *Hermetia illucens*, a natural waste recycler. Cell Research, 30(1), 50–60. <https://doi.org/10.1038/s41422-019-0252-6>

Zheng, H., Nishida, A., Kwong, W. K., Koch, H., Engel, P., Steele, M. I., & Moran, N. A. (2016). Metabolism of toxic sugars by strains of the bee gut symbiont *Gilliamella apicola*. MBio, 7(6). <https://doi.org/10.1128/mBio.01326-16>

Zhineng, Y., Ying, M., Bingjie, T., Rouxian, Z., & Qiang, Z. (2021). Intestinal microbiota and functional characteristics of black soldier fly larvae (*Hermetia illucens*). Annals of Microbiology, 71(1). <https://doi.org/10.1186/s13213-021-01626-8>