

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE INGENIERÍA

MAESTRÍA EN BIOLOGÍA COMPUTACIONAL

Ensamblaje y anotación del genoma de *Actinomyces* 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

EDISON GEOVANNY RODRÍGUEZ CERVANTES

Quito, 2023

DERECHOS DE AUTOR

Expreso que soy autor 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.

Geovanny Rodríguez

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN

Certifico que el trabajo de Edison Geovanny Rodríguez Cervantes 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

A la fuente de donde todo emana

ÍNDICE GENERAL

1.	INTRODUCCIÓN.....	11
1.1.	FORMULACIÓN DEL PROBLEMA.....	11
1.	JUSTIFICACIÓN.....	12
1.1.	OBJETIVOS	13
1.1.1.	OBJETIVO GENERAL	13
1.1.2.	OBJETIVOS ESPECIFICOS	13
2.	REVISION DE LA LITERATURA.....	14
2.1.	MOSCA SOLDADO	14
2.2.	<i>Actinomyces</i> sp.....	15
2.3.	ENSAMBLAJE Y ANOTACIÓN DE GENOMAS.....	15
3.	METODOLOGÍA.....	17
3.1.	ELIMINACIÓN DE LECTURAS DEL HOSPEDERO	17
3.2.	ENSAMBLAJE Y ANOTACIÓN DEL GENOMA	17
3.3.	BÚSQUEDA DE GENES Y RUTAS METABÓLICAS DE INTERÉS	18
4.	RESULTADOS	19
4.1.	ELIMINACIÓN DE LAS LECTURAS DEL HOSPEDERO	19
4.1.1.	CONTROL DE CALIDAD	19
4.1.2.	ELIMINACIÓN DE DUPLICADOS.....	20
4.1.3.	ELIMINACIÓN DE LA CONTAMINACIÓN.....	21
4.2.	ENSAMBLAJE Y ANOTACIÓN DEL METAGENOMA	23
4.2.1.	CLASIFICACIÓN TAXONOMICA	23
4.2.2.	ENSAMBLAJE DEL GENOMA.....	26
4.2.3.	AGRUPACIÓN Y OPTIMIZACIÓN DE BINS.....	27
4.2.4.	CALIDAD DE BINS.....	28
4.2.5.	ANOTACIÓN DEL GENOMA	30

4.3.	BUSQUEDA DE GENES Y RUTAS METABOLICAS	30
4.3.1.	RUTAS METABOLICAS.....	30
4.3.2.	BUSQUEDA DE GENES DE HIDROLISIS DE PET	33
5.	CONCLUSIONES Y RECOMENDACIONES	35
5.1.	CONCLUSIONES	35
5.2.	RECOMENDACIONES.....	35
6.	BIBLIOGRAFÍA.....	37

ÍNDICE DE FIGURAS

Figura 1: Toneladas de desechos sólidos recolectadas por los GADS municipales en Ecuador del 2017 al 2019	12
Figura 2: <i>Hermetia illucens</i> mosca soldado negra.	14
Figura 3: Estado larvario de <i>Hermetia illucens</i>	15
Figura 4: Proceso de ensamblado y anotación de genomas.....	16
Figura 5: Flujo de trabajo para la eliminación de lecturas del hospedero.	17
Figura 6: Flujo de trabajo para el ensamblado del genoma (Chivian et al., 2020).....	17
Figura 7: Flujo de trabajo para búsqueda de genes y rutas metabólicas de interés.	18
Figura 8: Calidad de la secuencia por base.....	19
Figura 9: Contenido de guanina y citocina por secuencia.	20
Figura 10: Resultados obtenidos en Dedupe de BMap.....	21
Figura 11: Resultados obtenidos en Reformat de BMap	21
Figura 12: Descarga de genoma de <i>Hermetia illucens</i> de NCBI.....	22
Figura 13: Genoma de <i>Hermetia illucens</i> en formato FASTA descargado de NCBI ...	22
Figura 14: Base de datos de <i>Hermetia illucens</i>	22
Figura 15: Lecturas mapeadas y no mapeadas de <i>Hermetia illucens</i>	23
Figura 16: Clasificación taxonómica por Kaiju.....	24
Figura 17: Clasificación taxonómica por Kaiju a nivel de género.	25
Figura 18: Clasificación taxonómica generado por Krona.....	25
Figura 19: Longitudes de contigs generados por IBDA, metaSPAdes y MEGA HIT ...	26
Figura 20: Calidad del ensamblaje por IBDA, metaSPAdes y MEGA HIT	27
Figura 21: Comparación de métodos de agrupamiento en bins.....	28
Figura 22: Calidad integridad y contaminación de las secuencias por bins.	29
Figura 23: Mapa de calor de módulos y cobertura de los componentes de la cadena de transporte de electrones en DRAM.	31
Figura 24: Funciones metabólicas generadas en DRAM	32

Figura 25: Resultado en Build FeatureSet from Genome con los genes de la hidrolisis de PET.	33
Figura 26: Resultado en Gblocks Trim Multiple Sequence Alignment.	34
Figura 27: Resultado en HMMER.....	34

ÍNDICE DE TABLAS

Tabla 1: Estadísticas del reporte de la calidad de la secuencia en FastQC	19
Tabla 2: Evaluación en Check M	29
Tabla 3: Resultados de la anotación del genoma en RAST.....	30
Tabla 4: Genes que intervienen en la hidrolisis de PET.....	33

RESUMEN

El objetivo de este trabajo de investigación fue ensamblar y anotar el genoma de *Actinomyces* sp, a partir de datos metagenómicos del tracto digestivo de *Hermetia illucens* en estado larvario para buscar e identificar genes y rutas metabólicas de interés en el proceso de biotransformación de residuos sólidos.

Para este trabajo se inició con la eliminación de lecturas del hospedero, se trabajó con los programas FastQC para el control de calidad de las lecturas, Trimmomatic para la limpieza de secuencias, Dedupe-BBMap para eliminación de datos duplicados, Bowtie2 y Samtools para eliminar la contaminación de la secuencia, Bedtools. Para el ensamblaje y la anotación del genoma se empleó la plataforma bioinformática KBase, se inició con el programa Kaiju para clasificación taxonómica, meta SPAdes, MEGA HIT, e IDBA-UD, para el ensamblaje, KB para comparar resultados y determinar el mejor ensamblaje, para la agrupación (binning) de los contigs se empleó MaxBin2, MetaBAT2 y CONCOCT, la optimización de los grupos (bins) se realizó mediante DAS-Tool y la evaluación de la calidad de los grupos en CheckM. Se procedió con la extracción de un grupo de interés mediante BinUntil, finalmente la anotación del genoma se lo realizó en RAST. Para la búsqueda de genes y rutas metabólicas se utilizó DRAM.

En los resultados se encontró que la muestra analizada tiene 41% de guanina y citosina; se muestra una presencia de Actinobacterias de 2% en la secuencia analizada, el mejor ensamblaje del metagenoma lo ejecutó metaSPAdes, y se logró optimizar 15 bins, el bin correspondiente a *Actinomyces* sp. fue el bin 8.

En la búsqueda de rutas metabólicas con DRAM en el complejo de transporte de electrones *Actinomyces* muestra rutas metabólicas como la ruta de Embden-Meyerhoff y la ruta Entner-Doudoroff; así mismo presentó rutas para el metabolismo del nitrógeno, donde se transforma nitrito en nitrato y viceversa; este microorganismo puede ser de interés en la degradación de material con presencia de nitrógeno como pesticidas, y degradación de materia orgánica, así como en la producción de gas metano, no se pudo comprobar la presencia de genes que participen en la hidrólisis de PET dentro del genoma de *Actinomyces* sp.

1. INTRODUCCIÓN

1.1. FORMULACIÓN DEL PROBLEMA

Uno de los mayores problemas de la actualidad es la generación de basura resultado de la actividad humana, este problema ha permanecido por décadas desde que el ser humano se ha asentado en nichos urbanos. El problema se agudiza con el incremento de la densidad demográfica lo que multiplica por miles las toneladas de residuos en el ambiente, consecuencia de una sociedad adaptada al consumo y al descarte (Bustos, 2019).

Esta basura, resultado de la actividad humana, está compuesta de residuos. Este término es empleado para caracterizar a los materiales sobrantes después de la actividad doméstica de los hogares, donde dichos materiales no son separados para ser reciclados, son descartados en combinación con otro tipo de desechos como los orgánicos, procedentes de la comida humana (Campos-Rodríguez, 2014).

Los desechos residuales llegan a botaderos de basura urbanos donde se aglomeran por el volumen, y al no contar con más alternativas, se emplea a los botaderos como sitios de almacenamiento fuera del alcance de los centros urbanos. Esto como una medida sanitaria para evitar la proliferación de enfermedades, intoxicación de fuentes de agua, y daño ecológico en tierras agrícolas (Bustos, 2019).

En los botaderos de basura se ha presenciado el fenómeno de la proliferación de insectos que estarían alimentándose de los desechos y basura orgánica. Es de gran interés conocer la actividad microbiológica en los intestinos de estos insectos como *Hermetia illucens*, también llamada mosca soldado. Los residuos como el plástico podrían estar siendo degradados por microorganismos dentro del intestino de estos insectos (Arroyave, 2019).

1. JUSTIFICACIÓN

Se conoce como desecho sólido, a los materiales descartados de las actividades animales y humanas. Son desechados como inútiles o superfluos; basándonos en esta definición podemos entender que los desechos sólidos son descartados constante y frecuentemente por los seres humanos (Chávez, 2001).

Según el INEC en el 2019 cada ecuatoriano arrojó 0.84kg de desechos por día, y se recolectaron 12 671.18 toneladas diarias de desechos sólidos; de los cuales el 86% son desechos que no fueron separados para un proceso de reutilización o reciclaje (INEC, 2019).

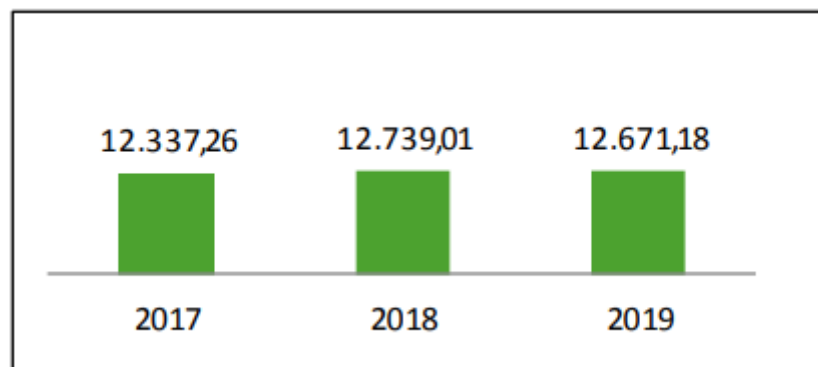


Figura 1: Toneladas de desechos sólidos recolectadas por los GADS municipales en Ecuador del 2017 al 2019

Hermetia illucens o mosca soldado, es un insecto con una gran capacidad para degradar la materia orgánica. Una característica es que se alimenta únicamente en etapa larvaria. Aunque se reconoce que las bacterias en su intestino están relacionadas con la capacidad para degradar la materia orgánica, no existe suficiente información sobre la estructura y capacidad funcional de esta microbiota (Cabrera et al, 2021).

Dentro de la microbiota de la larva de la mosca soldado se ha encontrado *Actinomyces* un bacilo con capacidad de degradación de una diversidad de

materiales, desde desechos agropecuarios hasta residuos de degradación compleja (Tanga et al, 2021).

Esta tesis de maestría propone analizar datos metagenómicos para entender la función de *Actinomyces* sp. dentro del intestino de *Hermetia illucens*, relacionada con la degradación de desechos sólidos, aportando información para un mejor tratamiento de desechos sólidos en el futuro.

1.1. OBJETIVOS

1.1.1. OBJETIVO GENERAL

Ensamblar y anotar el genoma de *Actinomyces* sp. a partir de datos metagenómicos del tracto digestivo de la mosca soldado *Hermetia illucens*.

1.1.2. OBJETIVOS ESPECIFICOS

- Analizar y depurar los datos del hospedero *Hermetia illucens* de las lecturas de los datos metagenómicos.
- Ensamblar el genoma de *Actinomyces* sp. a partir de datos metagenómicos.
- Anotar el genoma ensamblado de *Actinomyces* sp.
- Buscar genes y rutas metabólicas de interés en el genoma anotado de *Actinomyces* sp.

2. REVISION DE LA LITERATURA

2.1. MOSCA SOLDADO

Hermetia illucens es un díptero estratiomido denominado comúnmente como mosca soldado negra. Su origen aún no está claro, pero posiblemente es neotropical. Este insecto está presente en zonas cálidas de todo el mundo, ya sea porque tuvo una introducción accidental por diferentes medios de transporte o fue introducida de manera intencional para ser utilizada en distintos ambientes (Gobbi, 2012).

La especie *Hermetia illucens* puede ser criada de manera masiva ya que el estado larvario se alimenta de una gran cantidad de sustratos de diverso origen. El interés por este insecto se debe a su capacidad para la transformación de residuos orgánicos en biomasa, incluso la obtención de biomoléculas. Así mismo puede ser un bioindicador forense en investigaciones criminalísticas. Esta mosca no pica ni se alimenta en estadio adulto por lo que es completamente inofensiva para los humanos (Arroyave, 2019).



Figura 2: *Hermetia illucens* mosca soldado negra.

Debido a la falta de alimentación en el estadio adulto, la forma larvaria es la encargada de recolectar la energía para la adultez. La actividad enzimática y metabólica de los microorganismos de su intestino es de interés ya que podrían tener la capacidad de degradar desechos como el plástico (Gobbi, 2012).



Figura 3: Estado larvario de *Hermetia illucens*

2.2. *Actinomyces* sp.

Actinomyces es un microorganismo anaerobio Gram positivo, y con forma celular basilada. Su crecimiento óptimo se da entre los 25°C a 30°C, tolera un pH de 5 a 9, es heterótrofo y saprófito. Por estas características este microorganismo se encuentra como parte de la microbiota de varios organismos (Zoya, 2021).

Actinomyces tienen la capacidad de producir varios tipos de enzimas extracelulares con capacidad de degradación a nivel molecular. La capacidad enzimática de *Actinomyces* le permite degradar biopolímeros como almidón, celulosa, hemicelulosa, lignina, lignocelulosa, pectina, queratina, humus y quitina. Este microorganismo cumple un papel esencial en el reciclaje macromolecular de la naturaleza (Alsberi, 2020).

2.3. ENSAMBLAJE Y ANOTACIÓN DE GENOMAS

El ensamblaje del genoma consiste en un proceso bioinformático de descifrar y reconstruir la secuencia genómica a partir de fragmentos de ADN. Se lo realiza utilizando un software especializado para armar la cadena de ácidos nucleicos a partir de los datos obtenidos de la secuenciación (Aguilar-Bultet, 2015).

Luego del ensamblaje el proceso de la anotación del genoma consiste en la identificación estructural de genes, así como las regiones codificantes y la identificación funcional mediante programas bioinformáticos (NIH, 2022).

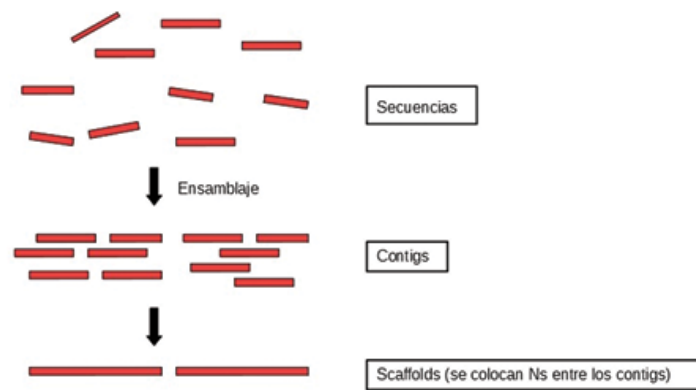


Figura 4: Proceso de ensamblado y anotación de genomas.

3. METODOLOGÍA

Para la ejecución de esta investigación se partió de archivos FASTQ con un output de 12 Gb, producto de una investigación previa de larvas de *Hermetia illucens* (Pazmiño, 2021).

3.1. ELIMINACIÓN DE LECTURAS DEL HOSPEDERO

Para la eliminación de lecturas del hospedero se trabajó en linux, con el flujo de trabajo que se muestra en la figura 5, con los programas FastQC para el control de calidad de las lecturas, Trimmomatic para la limpieza de secuencias, Dedupe-BBMap eliminación de datos duplicados, Bowtie2 y Samtools para eliminar la contaminación de la secuencia y Bedtools para transformar a formato Fastq.

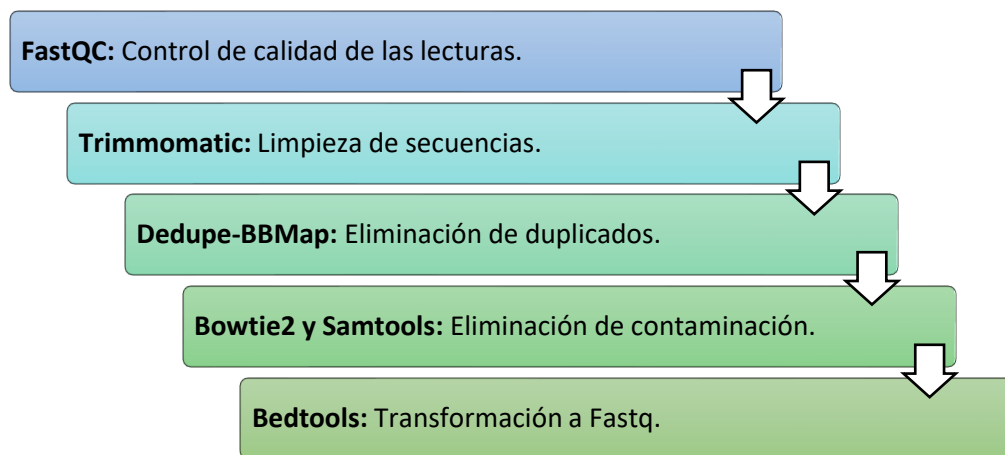


Figura 5: Flujo de trabajo para la eliminación de lecturas del hospedero.

3.2. ENSAMBLAJE Y ANOTACIÓN DEL GENOMA

El ensamblaje y la anotación del genoma se ejecutó en KBase empleando el flujo de trabajo que se muestra en la Figura 6.

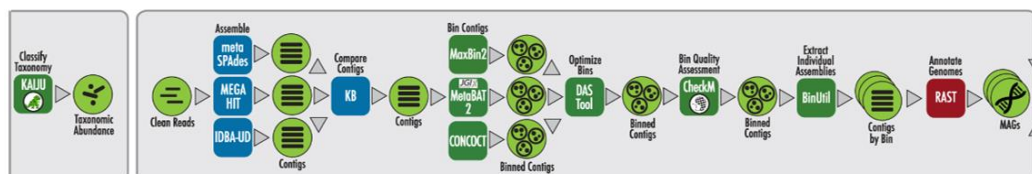


Figura 6: Flujo de trabajo para el ensamblado del genoma (Chivian et al., 2020)

Se inició con el programa Kaiju para determinar la clasificación taxonómica de microorganismos de las lecturas que fueron previamente limpiadas.

Las lecturas fueron introducidas en los programas meta SPAdes, MEGA HIT, e IDBA-UD, para el ensamblaje, cada programa realizó un ensamblaje independiente, para determinar cuál fue el mejor ensamblado se usó KB para comparar resultados y determinar cuál software realizó el mejor proceso. La combinación de contigs se hizo empleando MaxBin2, MetaBAT2 y CONCOCT, la optimización de los mismos se lo realizó mediante DAS-Tool. La evaluación de la calidad de los outputs se lo ejecutó mediante CheckM; y se procedió a la extracción de un bin de interés mediante BinUntil, finalmente la anotación del genoma se lo efectuó mediante RAST.

3.3. BÚSQUEDA DE GENES Y RUTAS METABÓLICAS DE INTERÉS

Para la búsqueda de genes y rutas metabólicas se siguió el flujo de trabajo que se muestra en la figura 7, se trabajó en KBase, para la búsqueda de genes y rutas metabólicas se utilizó DRAM.

Para la búsqueda de genes relacionados con la hidrolisis de PET se realizó la construcción de un set de genes PET con Build FeatureSet from Genome, se procedió a realizar una alineación de secuencias en MUSCLE, la eliminación de regiones variables se ejecutó a través de GBLOCKS, y en las regiones conservadas se buscó las coincidencias para hidrolisis de PET.

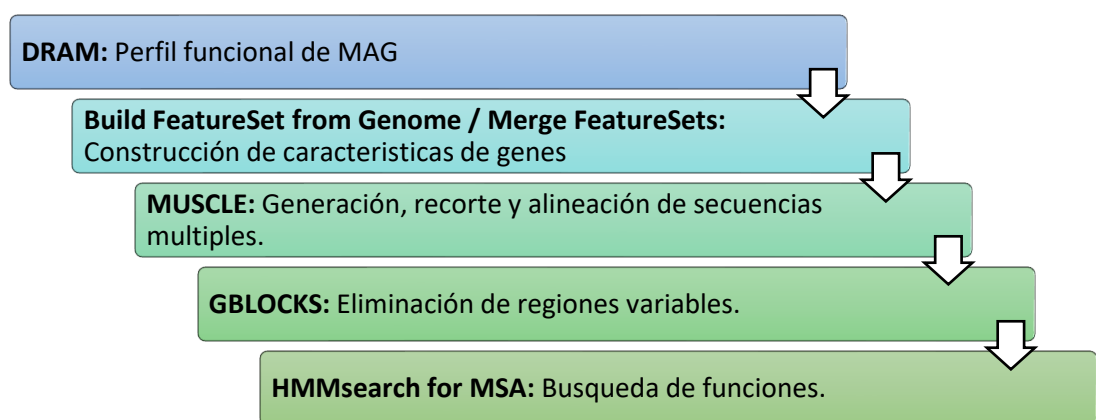


Figura 7: Flujo de trabajo para búsqueda de genes y rutas metabólicas de interés.

4. RESULTADOS Y DISCUSIÓN

4.1. ELIMINACIÓN DE LAS LECTURAS DEL HOSPEDERO

4.1.1. CONTROL DE CALIDAD

En la tabla 1 se muestran las estadísticas resultantes del programa FastQC, en el mismo se aprecia que la codificación para la muestra es Sanger / Illumina 1.9, el total de bases analizadas equivale a 2.9 Gpb; no existieron secuencias marcadas como de baja calidad.

Tabla 1: Estadísticas del reporte de la calidad de la secuencia en FastQC

Measure	Value
Filename	C_1y2_importado_148271_6_1.rev.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	19972191
Total Bases	2.9 Gbp
Sequences flagged as poor quality	0
Sequence length	150
%GC	41

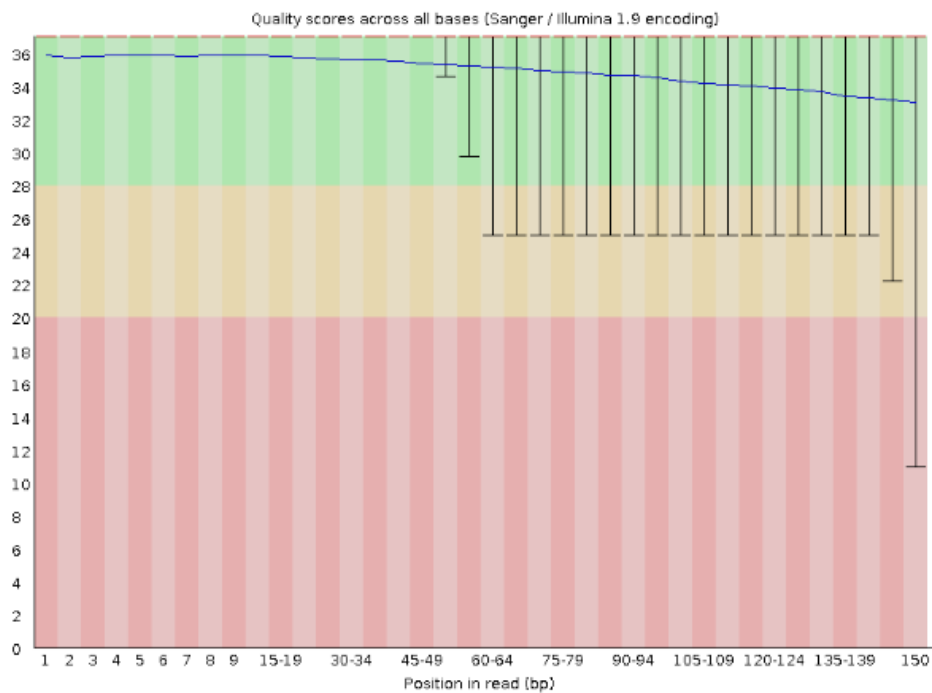


Figura 8: Calidad de la secuencia por base.

En la figura 8 se aprecia la calidad de la secuencia por base, se observa que se mantiene en la zona verde considerada buena calidad, y en la zona naranja de calidad razonable, se observa el final de la secuencia en la posición 150bp que entra en la zona roja por degradación (Babraham, 2020).

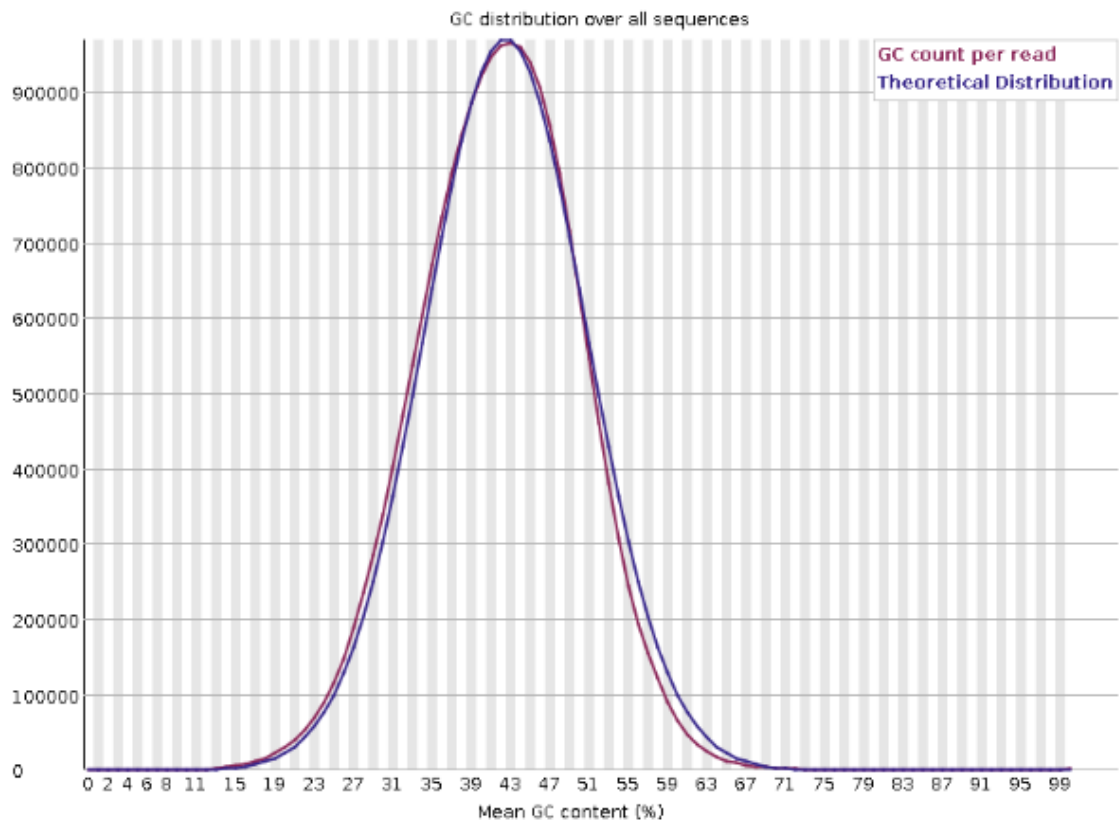


Figura 9: Contenido de guanina y citocina por secuencia.

En la figura 9 se aprecia el contenido de guanina y citocina por secuencia, la línea de color rojo indica el porcentaje de G y C de la muestra. Esta es parecida a la línea en azul que es la distribución normal, la línea teórica para determinar la calidad del contenido de G y C. Por lo que podemos afirmar que el porcentaje de G y C en esta muestra es adecuado debido a la configuración normal que sigue, se contrasta con la tabla 1 que indica que el porcentaje de guanina y citocina es del 41%.

4.1.2. ELIMINACIÓN DE DUPLICADOS

La eliminación de lecturas duplicadas se lo realizó utilizando BBMap, aplicando la herramienta Dedupe. En la figura 10 se puede ver que 10.60% de las lecturas estuvieron duplicadas, dejando un 89.50% libre de duplicados de los nucleótidos originales.

```

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: Resultados obtenidos en Dedupe de BBMap

Después de la aplicación de Dedupe, se empleó la herramienta Reformat, también parte del software BBMap, para reformatear las lecturas. En la figura 11 se aprecia que se obtiene un archivo con 100% de lecturas sin duplicar.

```

geo@geo-Lenovo-Flex-2-15:~/bbmap$ reformat.sh in=/home/geo/Descargas/Shotgun/C_adup.fq out=/home/geo/Descargas/Shotgun/C_reformat.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_reformat.fq
Executing jgi.ReformatReads [in=/home/geo/Descargas/Shotgun/C_adup.fq, out1=/home/geo/Descargas/Shotgun/C_reformat.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: Resultados obtenidos en Reformat de BBMap

4.1.3. ELIMINACIÓN DE LA CONTAMINACIÓN

Para la eliminación de lecturas del hospedero se obtuvo de NCBI el genoma de *Hermetia illucens*, como se ve en la figura 12. Se empleó el genoma con el número de acceso GCF_905115235.1 en formato FASTA como se aprecia en la figura 13 (NCBI, 2020). Con la herramienta Bowtie 2 y el genoma de *Hermetia illucens* se

creó una base de datos, el resultado se aprecia en la figura 14, que consta de 6 archivos, Hi_DB.1; Hi_DB.2; Hi_DB.3; Hi_DB.4; Hi_DB.rev.1 Hi_DB.rev.2.

Figura 12: Descarga de genoma de *Hermetia illucens* de NCBI

```
>LR899009.1 Hermetia illucens genome assembly, chromosome: 1
atctcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcagg
tcctttcattattttgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgattttcagcttttcctca
ttttcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcagg
tcctttcaatattttgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgattttcagcttttcctca
ttttcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcagg
tcctttcattattttgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgattttcagcttttcctca
ttttcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcagg
tactttcattatttccaagtacttaacgagaaattaacatttaaatcctggaaaagtgattttcagcttttcctca
ttttcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcagg
tcctttcattattttgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgattttcagcttttcctca
ttttcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcagg
atctcgatcatcaaatcacaggctatacacttgctctcagacgctgataccctcagatctcaagatcacaatcagg
tcctttcaatattttgcaagtacttaacgagaaattaacatttaaatcctggaaaagtgattttcagcttttcctca
```

Figura 13: Genoma de *Hermetia illucens* en formato FASTA descargado de NCBI

Figura 14: Base de datos de *Hermetia illucens*

Con la misma herramienta Bowtie2 se mapearon y se alinearon 333829 lecturas como se ve en la figura 15, de las cuales 55.20% alineadas concordantemente; 33.29% alineados concordantemente exacto 1 tiempo; 11.51% alineados concordantemente >1 tiempo. De las 184271 lecturas concordantes 2.72% fueron alineados discordantemente.

Así mismo se obtuvieron 179262 pares alineados, de 358524 el 85% esta emparejado; 8.39% emparejado exactamente y el 5.68% emparejado >1. La tasa de alineación general fue del 53.86%.

```
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
geo@geo-Lenovo-Flex-2-15:~$
```

Figura 15: Lecturas mapeadas y no mapeadas de *Hermetia illucens*

4.2. ENSAMBLAJE Y ANOTACIÓN DEL METAGENOMA

4.2.1. CLASIFICACIÓN TAXONOMICA

Los resultados obtenidos por Kaiju se muestran a nivel de filo, clase, orden, familia, género y especie en la figura 16, se puede apreciar que los microorganismos más abundantes a nivel de filo en el tracto intestinal de las larvas de *Hermetia illucens* son Proteobacteria, Firmicutes, Bacteroidetes Actinobacteria y Ascomycota, las clases de microorganismos más relevantes son Gamaproteobacteria que muestra una presencia superior al 80%, mientras que Actinobacteria tiene una presencia inferior al 20%. A nivel de orden las Enterobacterales están presentes en porcentaje superior al 80% y Actinomycetales inferior al 20%. A nivel de familia destaca Orbaceae con un porcentaje superior

al 60% y Actinomycetaceae inferior al 20%; los microorganismos mencionados coinciden con el manual de la mosca soldado negra de Caruso 2013.

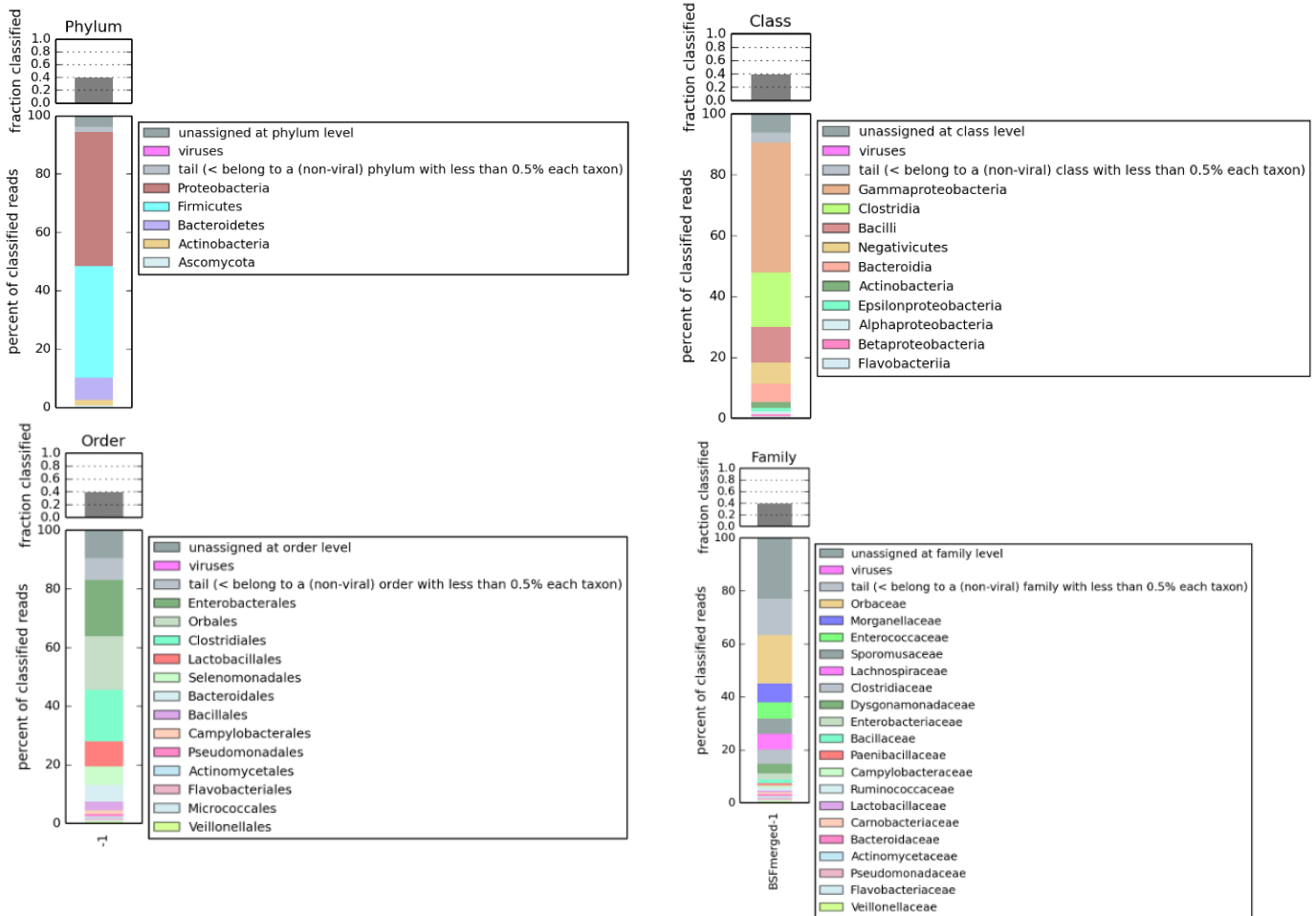


Figura 16: Clasificación taxonómica por Kaiju.

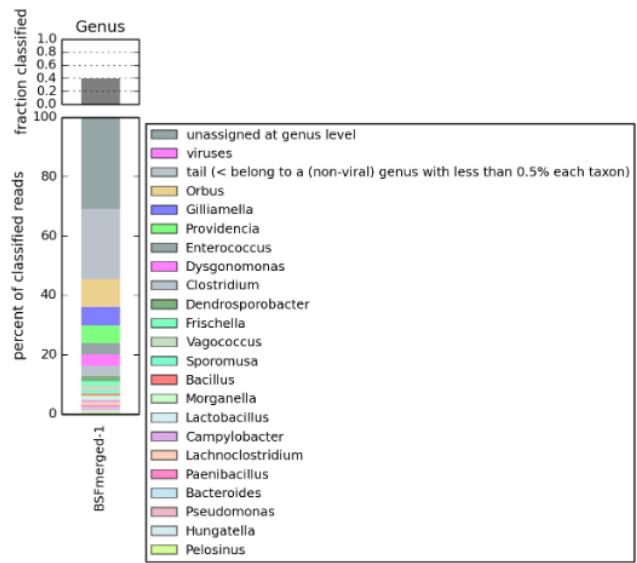


Figura 17: Clasificación taxonómica por Kaiju a nivel de género.

En la figura 17 se muestra los resultados por genero predominando *Orbus* con una presencia superior al 40%. En la figura 18 la clasificación taxonómica generada por Krona con los datos de Kaiju se muestra 2% de presencia de Actinobacterias en el intestino de *Hermetia illucens*.

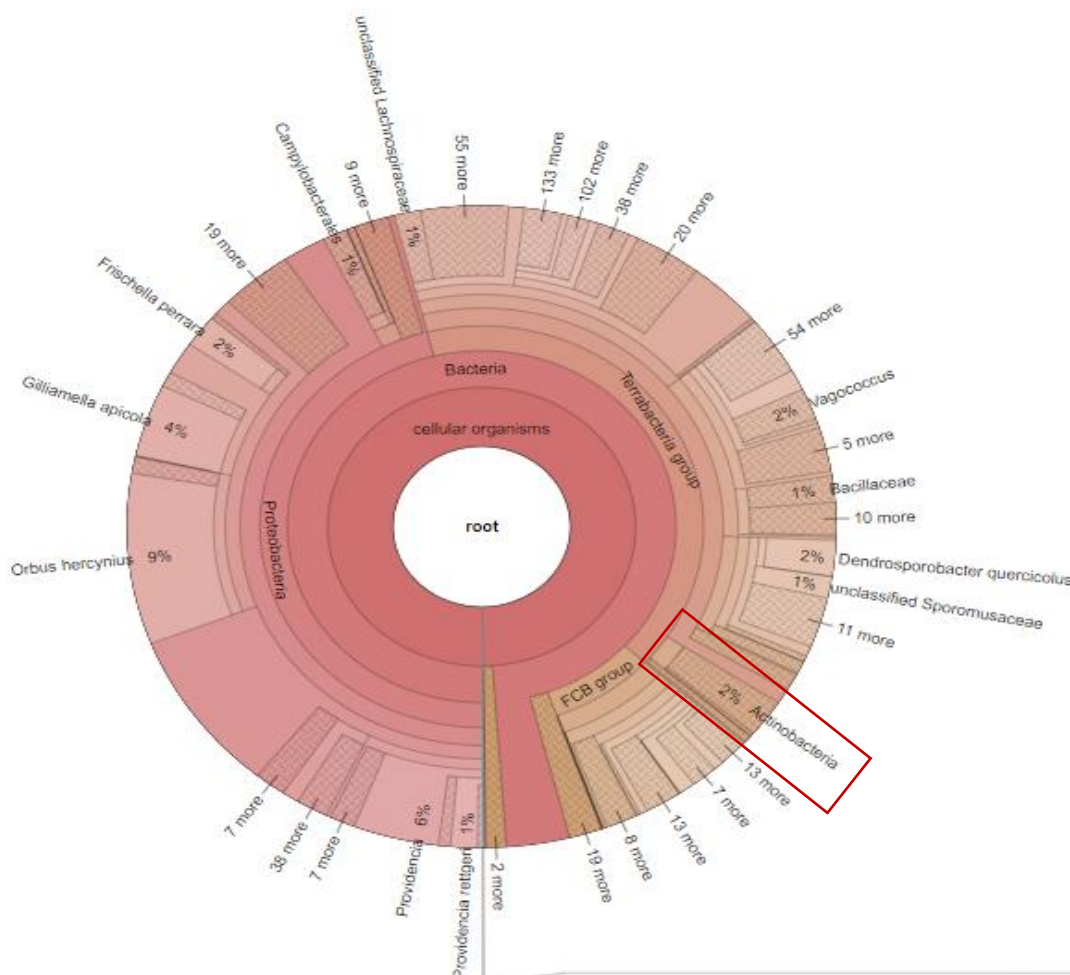


Figura 18: Clasificación taxonómica generado por Krona.

La composición de la microbiota de *Hermetia illucens* es distinta en porcentajes de presencia a la reportada por otros investigadores como Osimani et al., 2021; Sánchez, 2021, o Casariego 2020, cuyas moscas en experimentación fueron alimentadas con diferentes sustratos. La diferencia puede deberse a la composición del alimento que recibieron las larvas como lo resalta Bruno, D. en su investigación publicada en 2019 donde afirma que la microbiota vario no solo por la alimentación sino también por la región en la que la mosca soldado habita.

4.2.2. ENSAMBLAJE DEL GENOMA

En el ensamblaje del genoma se muestra en las figuras 19 y 20 la comparación de la calidad de metaSPAdes, IDBA, y MEGA HIT, evaluado por el programa Compare Assembled Contig Distributions (KB) en KBase.

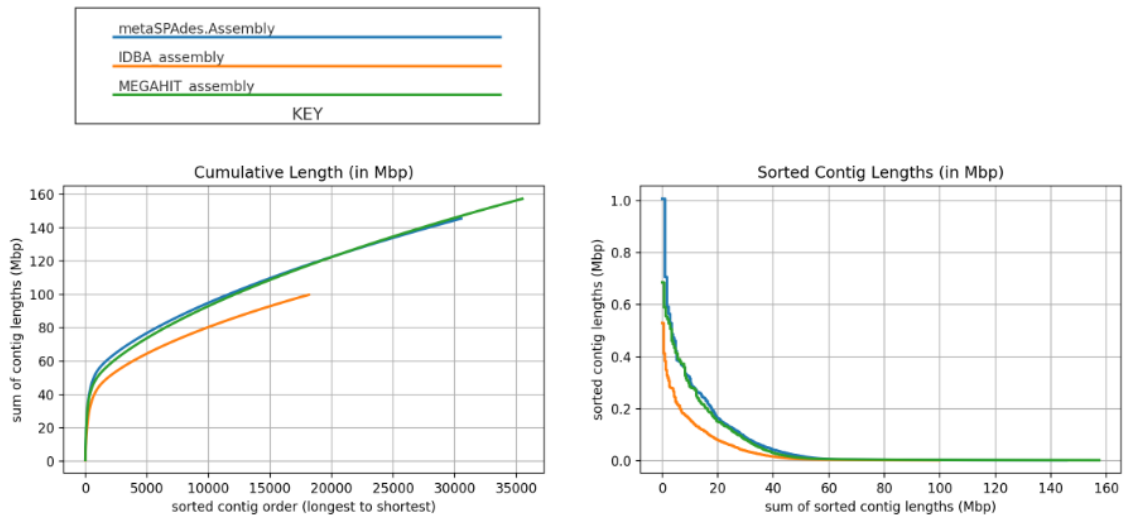


Figura 19: Longitudes de contigs generados por IBDA, metaSPAdes y MEGA HIT

En la figura 13 se aprecia las longitudes de los contigs donde IBDA muestra valores por debajo de los 2000 Mbp en longitud acumulativa frente a sus homologos mestaSPAdes y MEGA HIT que tienen valores superiores a los 3000 Mbp, así también se observa que metaSPAdes muestra una gran diferencia frente a los otros dos programas llegando hasta 1.0 Mbp en la longitud del contig.

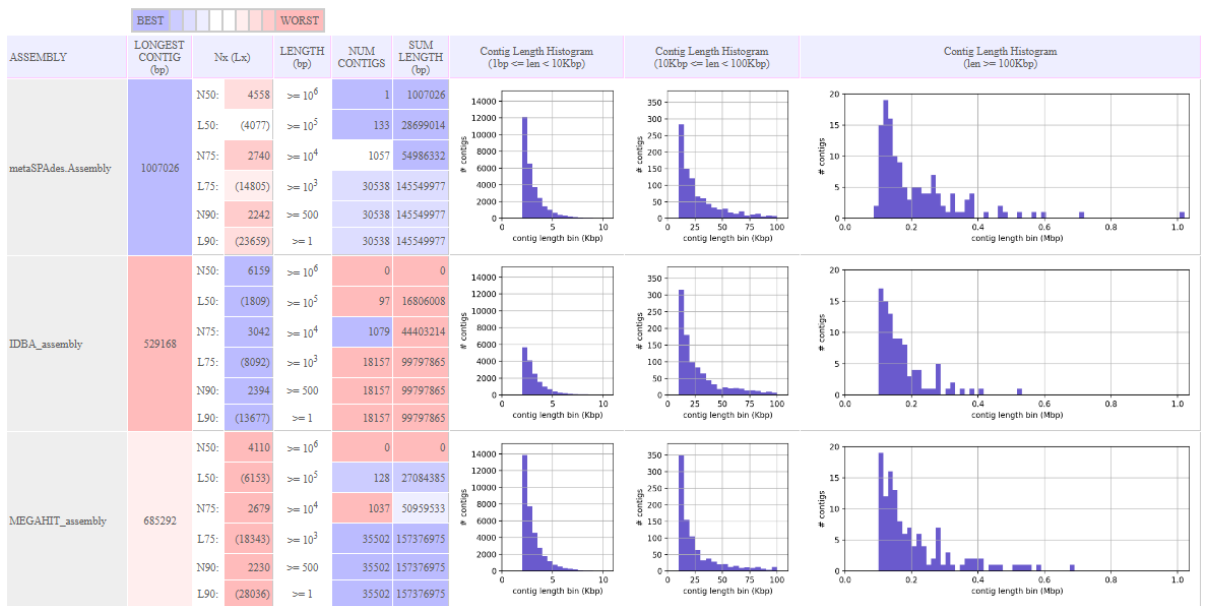


Figura 20: Calidad del ensamblaje por IBDA, metaSPAdes y MEGA HIT

En la figura 20 se aprecia las diferencias de calidad en KB en el siguiente orden, metaSPAdes, MEGA HIT, e IDBA, donde el número de contigs de metaSPADES llega hasta 30538. Los histogramas de longitud y el software Compare Assembled Contig Distributions (KB) indican que el mejor ensamblaje es el que se logró con metaSpades, por lo que se usó éste ensamblaje para los siguientes pasos del flujo de trabajo.

4.2.3. AGRUPACIÓN Y OPTIMIZACIÓN DE BINS

En los resultados de agrupación con el software MaxBin2 se obtuvo 20 bins, con el programa CONCOCT se obtuvo 78 bins, en MetaBAT2 29 bins. La optimización realizada por DAS-Tool obtuvo una cantidad total de contigs en el ensamblaje de 30538, la cantidad de contigs agrupados 1550, y la cantidad de bins optimizados fue de 15.

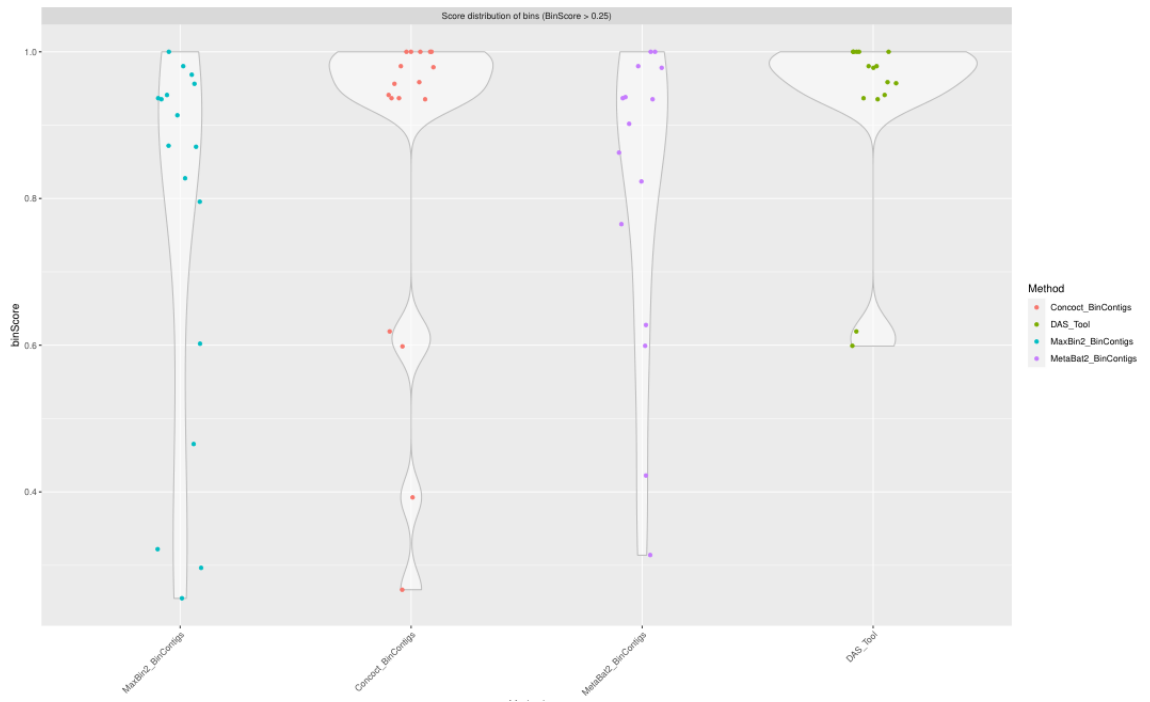


Figura 21: Comparación de métodos de agrupamiento en bins

4.2.4. CALIDAD DE BINS

La calidad de los bins se muestra en la figura 22. El resultado de CheckM, muestra que el bin.008 corresponde al linaje Actinomycetaceae; esto se corrobora al contrastar con la tabla 2 que muestra los marcadores de linaje de cada uno de los bins.

En el bin.008 de la figura 22 se observa presencia de contaminación, esto se aprecia en barras de color amarillo a rojo, así mismo la tabla 2 en la columna 2 tiene un valor de 7, y en la columna contamination lo muestran con un valor de 2.7 muy superior al de las otras muestras; la misma figura 22 muestra datos perdidos en color gris, ya que la herramienta muestra la presencia de una y exactamente una copia de un marcador con una línea verde en la figura 22 y la cuenta en la columna 1 de la tabla 2.

Esto puede deberse a la escasa presencia de las actinobacterias en el intestino de las larvas de *Hermetia illucens* cercana al 2%, lo que indica que existió limitadas moléculas de ADN para la extracción de datos (Chivian et al, 2021).

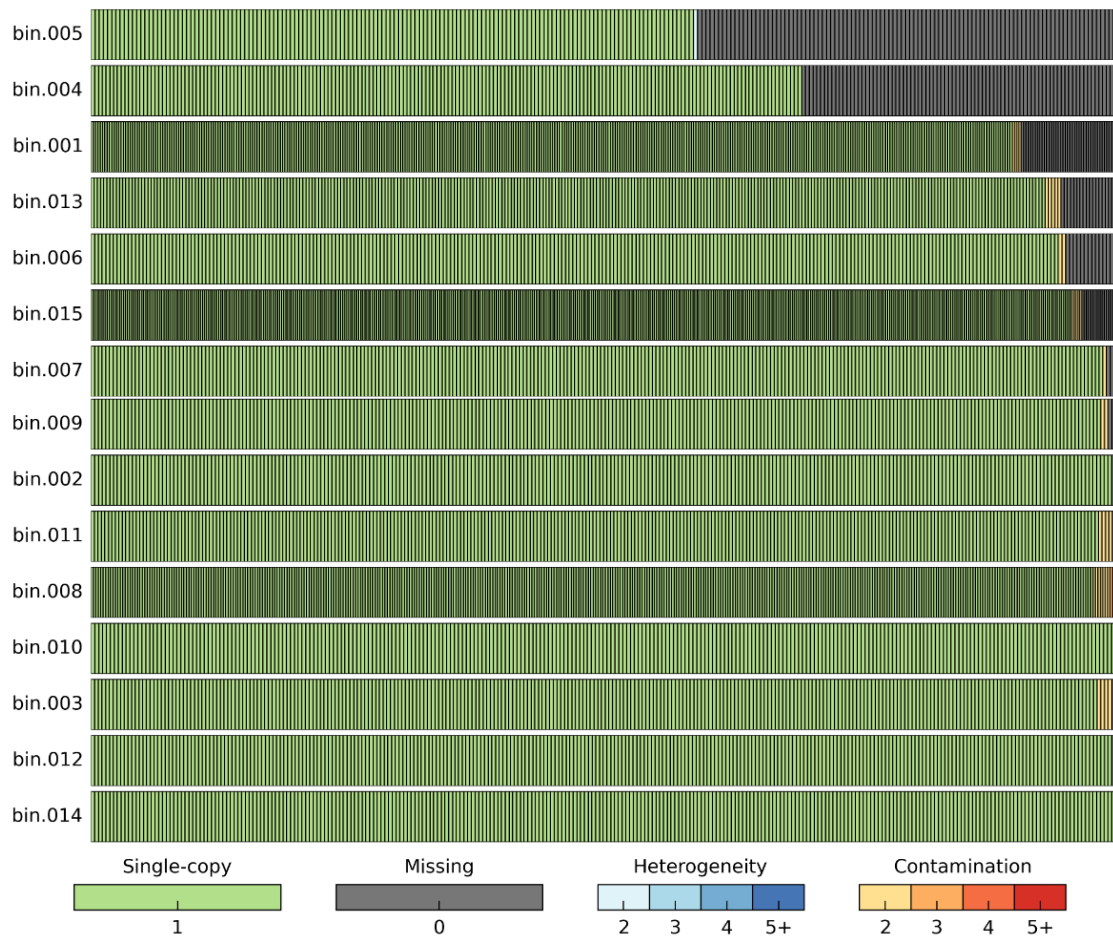


Figura 22: Calidad integridad y contaminación de las secuencias por bins.

Tabla 2: Evaluación en Check M

Bin Name	Marker Lineage	# Genomes	# Markers	# Marker Sets	0	1	2	3	4	5+	Completeness	Contamination
bin_001	c__Epsilonproteobacteria	111	445	271	42	399	4	0	0	0	90.15	0.49
bin_002	k__Bacteria	433	273	183	2	271	0	0	0	0	98.91	0.0
bin_003	o__Clostridiales	155	278	158	2	272	4	0	0	0	99.35	2.22
bin_004	k__Bacteria	433	273	183	84	189	0	0	0	0	70.87	0.0
bin_005	o__Clostridiales	304	247	141	101	145	1	0	0	0	63.03	0.71
bin_006	c__Gammaproteobacteria	965	277	177	14	261	2	0	0	0	93.93	1.13
bin_007	o__Clostridiales	155	278	158	3	274	1	0	0	0	98.42	0.63
bin_008	f__Actinomycetaceae	42	420	211	3	409	7	1	0	0	99.33	2.7
bin_009	o__Clostridiales	155	278	158	3	273	2	0	0	0	98.42	1.27
bin_010	o__Clostridiales	172	263	149	1	262	0	0	0	0	99.33	0.0
bin_011	p__Firmicutes	100	295	158	2	289	4	0	0	0	99.28	1.9
bin_012	c__Gammaproteobacteria	965	277	177	2	275	0	0	0	0	99.44	0.0
bin_013	c__Bacilli	586	325	181	18	302	5	0	0	0	92.4	1.52
bin_014	k__Bacteria	433	273	183	0	272	1	0	0	0	100.0	0.27
bin_015	c__Gammaproteobacteria	263	507	232	18	484	5	0	0	0	94.64	0.94

4.2.5. ANOTACIÓN DEL GENOMA

El resultado de la anotación del genoma en el programa RAST fue de Dominio Bacteria; Nombre científico “taxón desconocido”, código genético 11; el código genético predeterminado para los genomas de bacterias y arqueas es el código genético 11, como se aprecia en la tabla 3 (Land, 2023).

Tabla 3: Resultados de la anotación del genoma en RAST.

Overview		Browse Features	Browse Contigs
KBase Object Name	bin.008.fastaBinUntil_assembly.RAST	Taxonomy	
Scientific Name	unknown taxon	Unconfirmed Organism	
Domain	B		
Genetic Code	11		
Source	KBase		
Source ID	bin.008.fastaBinUntil_assembly.RAST		
Size	3,404,153		

4.3. BUSQUEDA DE GENES Y RUTAS METABOLICAS

4.3.1. RUTAS METABOLICAS

Los resultados de la búsqueda de rutas metabólicas con el programa DRAM, se aprecia en la figura 23, se puede ver que *Actinomyces* presenta rutas metabólicas como la ruta de Embden-Meyerhoff que se conoce como glucólisis, la misma que permite extraer energía de carbohidratos como la glucosa transformándola en piruvato, además presente la ruta de Entner-Doudoroff, esta es una ruta secundaria que permite generar adenosín trifosfato o ATP a partir de carbohidratos. También presenta la ruta TCA el ciclo del ácido tricarboxílico conocido como ciclo de Krebs, que es la vía catabólica para la oxidación de sacáridos. Esto indica que *Actinomyces* está facultada para adaptarse a sustratos ricos en glucosa como banano (Pazmiño, 2021).

En la figura 24 se puede apreciar las rutas para el metabolismo del nitrógeno, donde se muestra que *Actinomyces* puede manejar las rutas de nitrito a nitrato y viceversa (Ataikiru, 2020), además de transformar el nitrito en óxido nítrico, esta capacidad para manejar la transformación de nitrógeno es de interés en la degradación de pesticidas y herbicidas (Ling, 2019). Muchos compuestos químicos para cultivos dejan un excedente de nitratos en el suelo y en la maleza descartada, los nitratos podrían ser procesados por *Actinomyces* que cuenta con las rutas metabólicas para su transformación (Silva, 2006).

En otras investigaciones se ha visto a *Actinomyces* un bacilo anaerobio participar en la digestión anaeróbica (Soriano, 2010). Se ve en la figura 24 que muestra que *Actinomyces* puede tener actividad en las rutas de metanogénesis transformando el acetato a metano. La digestión anaeróbica junto con el manejo del nitrógeno es de interés en la aplicación de biodigestores para la eliminación de compuestos nitrogenados y producción de gas de combustión (Constanza, 2015).

También se aprecian otras rutas metabólicas relacionadas con el ciclo de Krebs como lo son la producción de acetil colina A por varias rutas alternativas (Zoya, 2021), y la ruta del lactato, también se aprecia la ausencia de genes de fotosíntesis, *Actinomyces* destaca por no ser fotosintética ya que puede realizar su actividad en ausencia de luz ya que es un organismo heterótrofo (Constanza, 2015).

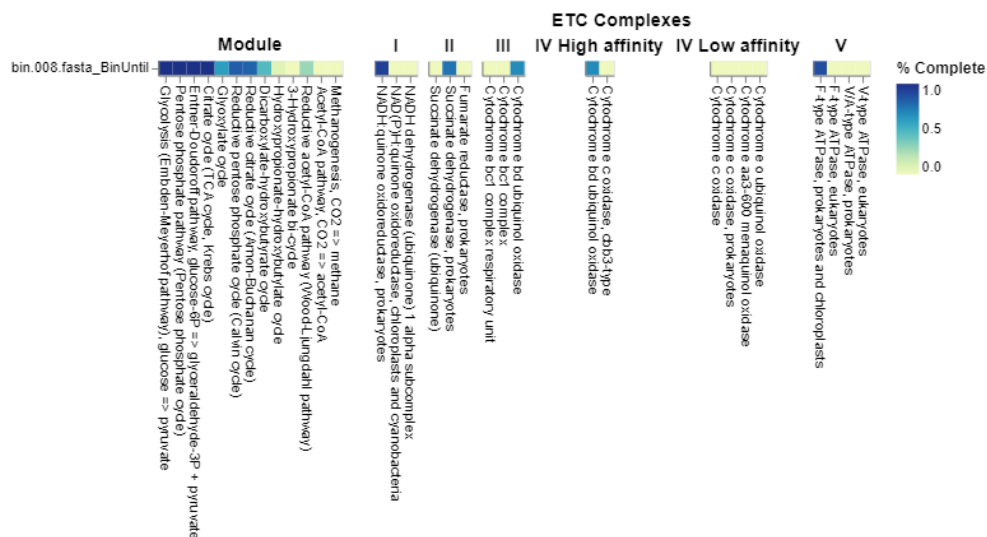


Figura 23: Mapa de calor de módulos y cobertura de los componentes de la cadena de transporte de electrones en DRAM.

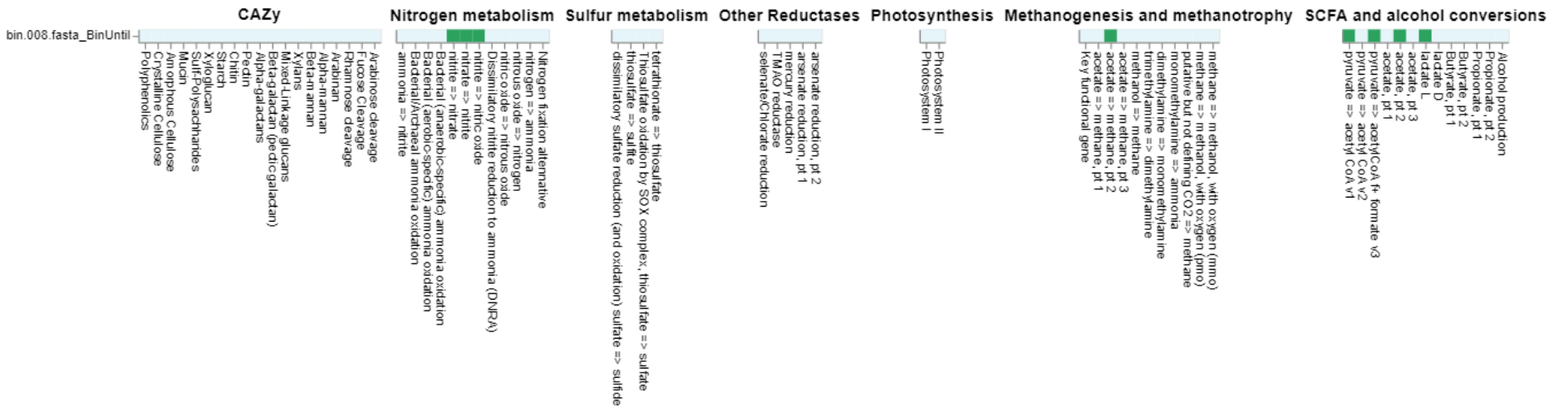


Figura 24: Funciones metabólicas generadas en DRAM

4.3.2. BUSQUEDA DE GENES DE HIDROLISIS DE PET

Para la búsqueda de genes que participan en la hidrólisis de PET se construyó un set en el programa KB Build FeatureSet from Genome, con la información de la tabla 4 de la publicación de Danso et al, 2018. con los genes Cut190 (W0TJ64), cut1 (E9LVI0), cut-2 (E5BBQ3), Tcur_1278 (D1A9G5), cut1 (E9LVH7), cut (H6WX58), cut2 (E9LVH9), y ISF6_4831 (A0A0K8P6T7) obteniendo el resultado que se indica en la figura 25.

Tabla 4: Genes que intervienen en la hidrólisis de PET.

Sequence no.	PDB entry no. ^a	Gene name ^b	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

Fuente: Danso et al, 2018.

Created Object Name	Type	Description
Bin.008_2_BuildFeature	FeatureSet	Feature Set

Showing 1 to 1 of 1 entries

Summary

A new feature set containing 115 features was created.

Figura 25: Resultado en Build FeatureSet from Genome con los genes de la hidrólisis de PET.

El output de KB Build Feature Set from Genome, fue ingresado a Gblocks Trim Multiple Sequence Alignment obteniendo el resultado que se muestra en la figura 26

Created Object Name	Type	Description
Ptase_MSA_trimmed	MSA	GBLOCKS MSA

Showing 1 to 1 of 1 entries

Figura 26: Resultado en Gblocks Trim Multiple Sequence Alignment.

El output de Gblocks fue ejecutado en el programa HMMER Search from MSA (prot-prot), el resultado se muestra en la figura 27 que muestra un error, el protocolo fue ejecutado varias veces para corroborar que no se trata de un error de software o falla en la programación. El resultado fue el mismo en todas las repeticiones, el error muestra un archivo de HMMER vacío. El programa convierte una alineación de secuencias múltiples de proteínas en un modelo oculto de Markov que se utiliza para buscar en una base de datos de proteínas; el programa busca en este archivo frente al modelo de genoma de interés en el cual se desea verificar la existencia de estos genes. Una de las posibles razones es la abundante cantidad de datos ausente en bin.008 correspondiente a *Actinomyces* como se aprecia en la figura 22, como ya se explicó con anterioridad la baja presencia de *Actinomyces* en el intestino de *Hermetia illucens* produce una baja presencia de moléculas de ADN que dificultan la extracción de datos para la correcta ejecución de los programas bioinformáticos (Land, 2023).

Server error: HMMER_SEARCH created empty MSA file

'/kb/module/work/tmp/output.1683050609851/hmmer_run/msaout.txt'

If the app fails consistently, please contact us at <https://www.kbase.us/support>.

▼ Error stacktrace

```
Traceback (most recent call last):
  File "/kb/module/bin/./lib/kb_hmmer/kb_hmmerServer.py", line 95, in _call_method
    result = method(ctx, *params)
  File "/kb/module/lib/kb_hmmer/kb_hmmerImpl.py", line 824, in HMMER_MSA_Search
    raise ValueError("HMMER_SEARCH created empty MSA file '" + output_hit_MSA_file_path +
    "'")
ValueError: HMMER_SEARCH created empty MSA file
'/kb/module/work/tmp/output.1683050609851/hmmer_run/msaout.txt'
```

Figura 27: Resultado en HMMER.

5. CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

Es posible el ensamblado y la anotación del genoma de *Actinomyces* sp. a partir de datos metagenómicos del tracto digestivo de *Hermetia illucens*.

El mejor ensamblaje del genoma de *Actinomyces* sp. a partir de datos metagenómicos, se obtuvo a través de la herramienta metaSPAdes (30538 contigs), la cual fue evaluada frente a MEGA HIT, y IDBA.

RAST no logró identificar al microorganismo del bin008 como *Actinomyces* sp. Sin embargo, el software Krona detecta la presencia de *Actinomyces* en un 2% dentro de los datos crudos obtenidos por secuenciación shotgun del ADN del tracto digestivo de *Hermetia illucens*.

Actinomyces presenta las rutas metabólicas para la transformación de nitrito a nitrato y viceversa, además de la ruta para transformar el nitrito en óxido nítrico, una ruta de interés que puede ser aprovechada en la degradación del excedente de nitratos en el suelo por medio de la biodigestión; así como la ruta de metanogénesis a partir del acetato, puede ser aprovechada en la producción de metano con consorcios microbiológicos.

5.2. RECOMENDACIONES

Se recomienda realizar la identificación taxonómica en Kaiju, para *Actinomyces* con datos extraídos de larvas de *Hermetia illucens* cultivados en medios controlados con distintos sustratos, para adquirir información sobre los sustratos que permiten el aumento y disminución de *Actinomyces* sp, dentro de la mosca soldado.

Realizar la búsqueda de genes para la hidrólisis de PET con datos extraídos de cultivos microbiológicas de *Actinomyces* para corroborar la presencia de estos genes en Actinobacterias.

Ejecutar la búsqueda de genes para la hidrolisis de PET con datos metagenómicos de *Hermetia illucens* con una presencia superior al 10% de *Actinomyces* para evitar la escases de datos por ausencia de moléculas de ADN.

6. BIBLIOGRAFÍA

Aguilar-BultetI, 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

Alsberi, H; Hamad, AA; Hassan, MM. (2020). Biodegradation of petroleum hydrocarbons using indigenous bacterial and actinomycetes cultures. Pakistan Journal of Biological Sciences ISSN: 1028-8880. EISSN: 1812-5735. 726-734. DOI[10.3923/pjbs.2020.726.734](https://doi.org/10.3923/pjbs.2020.726.734)

Ataikiru, T. L. and Okerentugba, P. O. and Okpokwasili, G. C. (2020). Identification of Carbofuran and Paraquat Degrading Microorganisms from Soil. South Asian Journal of Research in Microbiology, 7 (1). pp. 40-52. ISSN 2582-1989.

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

Bruno D, Bonelli M, De Filippis F, Di Lelio I, Tettamanti G, Casartelli M, Ercolini D, Caccia S. The Intestinal Microbiota of *Hermetia illucens* Larvae Is Affected by Diet and Shows a Diverse Composition in the Different Midgut Regions. Appl Environ Microbiol. 2019 Jan 9;85(2):e01864-18. doi: 10.1128/AEM.01864-18. PMID: 30504212; PMCID: PMC6328772.

Cabrera Gutiérrez, D. y López Gutiérrez, A. L. (2021) Evaluación de la larva de mosca soldado negra (*Hermetia Illucens*) como alternativa para la degradación de residuos sólidos urbanos. Institucional Lumieres. <https://hdl.handle.net/20.500.11839/8329>

Campos-Rodríguez, R. Soto-Cordoba, S. (2014). Estudio de generación y composición de residuos sólidos en el cantón de Guácimo, Costa Rica, Tecnología en Marcha. Vol. 27, N° 3, Julio-Setiembre 2014. Pág 122-135.

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>

Constanza Lucia, Antolinez Diana , Bohórquez Johanna, Aura Corredor (2015). Anaerobic bacteria: processes they perform and their contribution to life sustainability on the planet. NOVA. 2015; 13 (23): 55-81

Danso Dominik, Schmeisser Christel, Chow Jennifer, Zimmermann Wolfgang, Ren Wei,b Christian Leggewie,c Xiangzhen Li,d Terry Hazen,e Wolfgang R. Streita (2018) New Insights into the Function and Global Distribution of Polyethylene Terephthalate (PET)-Degrading Bacteria and Enzymes in Marine and Terrestrial Metagenomes. April 2018 Volume 84 Issue 8 e02773-17

Gobbi, F. (2012). Biología reproductiva y caracterización morfológica de los estadios larvarios de *Hermetia illucens*. Bases para su producción masiva en Europa Universidad de Alicante. Centro Iberoamericano de la Biodiversidad | Universidad de Alicante. Departamento de Ciencias Ambientales y Recursos Naturales.

INEC. (2019). Gestión de residuos sólidos, GADS municipales 2019 Unidad de Estadísticas Ambientales basadas en Registros Administrativos (ESARA).

Land Miriam. (2023). Annotate Multiple Microbial Assemblies with RASTtk. Disponible en: <https://narrative.kbase.us/narrative/rast>

Ling Guo, Wen-Wan Fang, Lei-Lei Guo, Chuan-Fei Yao, Yun-Xiu Zhao, Feng Ge, Yi-Jun Dai. (2019). Biodegradation of the Neonicotinoid Insecticide Acetamiprid by Actinomycetes *Streptomyces canus* CGMCC 13662 and Characterization of the Novel Nitrile Hydratase Involved. <https://doi.org/10.1021/acs.jafc.8b06513>

NIH. (2022). Diccionarios del NCI. Instituto nacional del cancer. Disponible: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/genoma>

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.

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. [tps://doi.org/10.46480/esj.5.3.164](https://doi.org/10.46480/esj.5.3.164)

Silva P, A. (2006). Efecto de la fertilización nitrogenada sobre la actividad microbial y rendimiento de avena forrajera en un suelo andisol del departamento de Nariño, Colombia. *Acta Agronómica*, 55(1), 55–63.

Soriano, B. (2010). Degradación de pesticidas por actinomicetos. *UCV - Scientia* Vol. 2 N° 1.

Tanga Chrysantus M, Jacqueline Wahura Waweru¹, Yosef Hamba, Abel Anyega Onyoni, Fathiya M. Khamis¹, Sunday Ekesi and Juan C. Paredes. (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. DOI: <https://doi.org/10.3389/fmicb.2021.635881>

Zoya Javed, Gyan Datta Tripathi, Mansi Mishra, Kavya Dashora. (2021). Actinomycetes – The microbial machinery for the organic-cycling, plant growth, and sustainable soil health, *Biocatalysis and Agricultural Biotechnology*. Volume 31. <https://doi.org/10.1016/j.bcab.2020.101893>.