



FACULTAD DE INGENIERÍA

MAESTRÍA EN BIOLOGÍA COMPUTACIONAL

**TRABAJO DE TITULACIÓN, PREVIO A LA OBTENCIÓN DEL TÍTULO DE
MAGÍSTER EN BIOLOGÍA COMPUTACIONAL**

TEMA: OBTENCIÓN Y CARACTERIZACIÓN DEL GENOMA DE *Agrobacterium tumefaciens* A PARTIR DEL METAGENOMA DE TUMORES DE ROSA.

AUTOR: CARRILLO FONSECA, DENISE ARIANA

DIRECTOR: FLORES FLOR, FRANCISCO JAVIER, PhD

QUITO

2023

CERTIFICACIÓN

Certifico que el presente trabajo de titulación fue realizado por la señorita Denise Ariana Carrillo Fonseca cumple con los requisitos técnicos y legales establecidos por la Pontificia Universidad Católica del Ecuador, razón por la cual me permito acreditar y autorizar su calificación y posterior sustentación.

.....

Francisco Javier Flores, Ph.D

Director

DEDICATORIA

A mis padres y hermano

AGRADECIMIENTOS

Agradezco a mis padres y hermano, quienes siempre han sido mi mayor fuente de apoyo y aliento. Gracias por ser luz en mi oscuridad.

Agradezco al Dr. Francisco Flores, por su guía y paciencia durante toda la realización del presente trabajo. Gracias por permitirme trabajar bajo su tutela, por su tiempo y consejos que permitieron enriquecer mi conocimiento. También agradezco al laboratorio IDgen, por los datos proporcionados, sin los cuales no hubiera sido posible realizar mi trabajo de titulación.

Finalmente quiero agradecer a mis amigos, por el aguante durante todos estos meses, por escucharme, darme ánimos y sobre todo creer en mí, incluso cuando yo no lo hacía.

ÍNDICE DE CONTENIDOS

ÍNDICE DE TABLAS.....	7
ÍNDICE DE FIGURAS	8
RESUMEN	9
ABSTRACT	10
CAPÍTULO 1: INTRODUCCIÓN.....	11
1.1 FORMULACIÓN DEL PROBLEMA.....	11
1.2 Objetivos.....	12
1.2.1 Objetivo Principal.....	12
1.2.2 Objetivos Específicos	12
1.3 MARCO Teórico	12
1.3.1 <i>Agrobacterium tumefaciens</i>	12
1.3.2 Enfermedad de la agalla de la corona.....	15
1.3.3 Metagenómica	17
CAPÍTULO 2: MATERIALES Y MÉTODOS.....	20
2.1 Evaluación de calidad	20
2.2 Evaluación de la estructura taxonómica	20
2.3 Ensamblaje del metagenoma.....	21
2.4 Anotación del metagenoma.....	21
2.5 Ensamblaje de genomas y mapeo de genes de virulencia.....	21
2.6 Comparacion de los perfiles taxonómicos bacterianos de los metagenomas de tumor de rosa y de tejido sano.	22
CAPÍTULO 3: RESULTADOS Y DISCUSIÓN	23
3.1 Control de calidad de las lecturas de secuenciación	23
3.2 Clasificación taxonómica del metagenoma	24
3.3 Ensamblaje y anotación del metagenoma	31

3.4	Agrupamiento y optimización de bins	32
3.4.1	Asignación taxonómica de los bins	33
3.5	Anotación del Metagenoma	34
3.6	Mapeo del genoma de <i>A. tumefaciens</i> y <i>A. larrymoorei</i>	36
3.7	Mapeo de los genes de virulencia de <i>Pantoea agglomerans</i>	37
3.8	Mapeo de plásmidos <i>ti</i> de <i>A. tumefaciens</i>	38
3.9	Comparación de los perfiles taxonómicos bacterianos de los metagenomas de tumor de rosa y de tejido sano.	40
CAPÍTULO 4: CONCLUSIONES		43
CAPÍTULO 5: RECOMENDACIONES.....		44
BIBLIOGRAFÍA.....		45

ÍNDICE DE TABLAS

Tabla 1. Resultados obtenidos por las herramientas de agrupamiento.	32
Tabla 2. Análisis de los marcadores filogenéticos evaluados por CheckM.	33
Tabla 3. Asignación taxonómica de GTDB-Tk para los bins optimizados.	34
Tabla 4. Análisis de calidad del ensamblaje de <i>Agrobacterium tumefaciens</i>	36

ÍNDICE DE FIGURAS

Figura 1. Regiones del plásmido Ti que contribuyen a la infección (región vir y ADN-T), supervivencia celular en el entorno tumoral (catabolismo de opiones) y transferencia conyugal del plásmido Ti (tra y trb)	13
Figura 2. Modelo del proceso de transformación mediado por <i>Agrobacterium</i>	15
Figura 3. Etapas de desarrollo de la enfermedad de agalla de la corona producida por <i>A. tumefaciens</i>	16
Figura 4. Pasos generales en un flujo de trabajo para el ensamblaje de genomas.	18
Figura 5. Calidad de las secuencias por base para las lecturas forward (izq.) y reverse (der.).	23
Figura 6. Contenido de GC para las lecturas forward (izq.) y reverse (der.).	24
Figura 7. Puntuaciones de calidad por secuencia para las lecturas forward (izq.) y reverse (der.).	24
Figura 8. Clasificación taxonómica a nivel de Filo.	25
Figura 9. Clasificación taxonómica a nivel de Clase.	26
Figura 10. Clasificación taxonómica a nivel de Orden.	26
Figura 11. Clasificación taxonómica a nivel de Familia.	27
Figura 12. Clasificación taxonómica a nivel de Género.	27
Figura 13. Clasificación taxonómica a nivel de Especie.	28
Figura 14. Dendograma generado por GOTTCHA2 - v2.1.7 para bacterias.	29
Figura 15. Clasificación a nivel de género para hongos obtenida a través de Kaiju - v1.7.3.	30
Figura 16. Comparación de las métricas de los distintos ensamblajes.	32
Figura 17. Representación de la integridad y contaminación de los bins optimizados. 33	
Figura 18. Mapa de calor de los módulos y cobertura de los componentes de la cadena de electrones.	35
Figura 19. Funciones metabólicas encontradas.	35
Figura 20. Mapeo del gen PthG contra las lecturas crudas.	38
Figura 21. Alineamiento del ensamblaje con la secuencia completa del plásmido pTiKerr108.	39
Figura 22. Comparación del perfil taxonómico entre los metagenomas de tumor de rosa y de tejido sano.	40

RESUMEN

Agrobacterium tumefaciens es una bacteria Gram negativa patógena que afecta a plantas, causante de la enfermedad de agalla de la corona en rosas. El Ecuador es reconocido mundialmente por su gran variedad de flores de exportación, entre ellas las rosas. Sin embargo, son pocos los estudios a nivel molecular relacionados con patógenos que causan enfermedades a esta planta. Es así como un enfoque de estudio metagenómico de tumores de agalla de corona permitirá ampliar el conocimiento relacionado con su estructura taxonómica y funcional. Por lo que el presente trabajo tuvo por objetivo aplicar herramientas de análisis metagenómico para la obtención y caracterización del genoma de *Agrobacterium tumefaciens* a partir del metagenoma de tumores de rosa. La composición taxonómica del metagenoma reveló que *Pantoea*, *Agrobacterium*, *Pseudomonas* y *Sphingomonas* fueron los géneros más abundantes y presentes de manera concurrente en ambas herramientas de clasificación. Sin embargo, a nivel de especie, se observaron divergencias en la asignación, aunque *Pantoea agglomerans* prevaleció como la especie más común en ambas. A partir del ensamblaje del metagenoma se obtuvo el genoma de *P. agglomerans* con una integridad del 100%, contaminación del 0.52% y un ANI del 98.49%. A nivel funcional el metagenoma presentó un total de 12351 CDSs, 7 tRNAs, 126 tRNAs y 2 tmRNAs. Por medio del mapeo del genoma de referencia de *Agrobacterium tumefaciens* contra las lecturas crudas, se ensambló el 0.861% del genoma, debido al limitado número de lecturas secuenciadas. A pesar de que no se recuperó por completo el genoma de *A. tumefaciens*, el organismo fue seleccionado como el más probable de causar tumores en la muestra. Considerando que, se recuperó la secuencia del plásmido Ti en un 69.23% y no se identificaron genes de virulencia para *P. agglomerans*.

ABSTRACT

Agrobacterium tumefaciens is a Gram-negative bacterial plant pathogen that causes crown gall disease in roses. Ecuador is recognized worldwide for its variety of exported flowers, including roses. However, there are few studies at the molecular level related to pathogens that cause diseases in this plant. Therefore, a metagenomic approach to studying crown gall tumors will allow us to expand the knowledge related to their taxonomic and functional structure. The objective of the work was to apply metagenomic analysis tools to obtain and characterize the *Agrobacterium tumefaciens* genome from the metagenome of rose tumors. The taxonomic composition of the metagenome revealed that *Pantoea*, *Agrobacterium*, *Pseudomonas*, and *Sphingomonas* were the most abundant genera and concurrently present in both classification tools. However, at the species level, divergences in assignment were observed, although *Pantoea agglomerans* prevailed as the most common species in both. From the metagenome assembly, the genome of *P. agglomerans* was obtained with 100% integrity, 0.52% contamination, and 98.49% ANI. At the functional level, the metagenome presented a total of 12,351 CDSs, seven tRNAs, 126 tRNAs, and two tmRNAs. By mapping the *Agrobacterium tumefaciens* reference genome against the raw reads, only 0.861% of its genome was assembled, due to the limited number of sequenced reads. Although the complete *A. tumefaciens* genome was not recovered, the organism was selected as the most likely to cause tumors in the sample. Furthermore, the sequence of the Ti plasmid was recovered in 69.23%, and no virulence genes were identified for *P. agglomerans*.

CAPÍTULO 1: INTRODUCCIÓN

1.1 FORMULACIÓN DEL PROBLEMA

Miembros del género *Agrobacterium* son considerados como organismos ubicuos de la microflora presente en el suelo, en su mayoría son saprófitos y sobreviven fundamentalmente de materia orgánica en descomposición. Sin embargo, especies de este género como a *A. rhizogenes*, *A. rubi*, *A. vitis* y *A. tumefaciens* causan enfermedades neoplásicas en pantas; siendo esta última la causal de la enfermedad de agalla de la corona en rosas. Diversas investigaciones han descrito a la patogénesis de *Agrobacterium* como un proceso único y altamente especializado, ya que involucra la transferencia de genes entre reinos bacterianos y vegetales. La agalla de corona ha sido descrita como una forma de “colonización genética”, donde la transferencia y expresión de un grupo de genes de *Agrobacterium* en una célula vegetal ocasiona la proliferación celular descontrolada y la síntesis de sustancias nutritivas que pueden metabolizarse específicamente por la bacteria. De esta forma se genera un nicho adecuado para la supervivencia de esta (Escobar & Dandekar, 2003). Las plantas de rosas infectadas por *A. tumefaciens* desarrollan agallas (tumores) en las zonas basales de sus tallos y raíces. La enfermedad de la agalla de la corona afecta la absorción, el crecimiento y la producción de nutrientes teniendo como consecuencia una reducción del crecimiento de la planta, lo que conlleva a pérdidas económicas en el sector productivo (Chen et al., 2020).

El Ecuador produce una gran variedad de flores para exportación como: rosas, flores de verano, gypsophila, flores preservadas y claveles. Para febrero de 2022, el sector florícola ocupó el quinto lugar dentro de los productos de exportación no petrolera. Adicionalmente el 75% de exportaciones de flores correspondió a rosas, lo cual colocó al país como el segundo exportador a nivel mundial (ExpoFlores, 2022). Por tanto resulta innegable la importancia a nivel económico del cultivo de rosas en el país. No obstante, esta actividad no se encuentra exenta de desafíos, como el control de plagas y enfermedades, siendo una de ellas la agalla de la corona.

En el futuro, se espera que el manejo de la enfermedad de la agalla de la corona mejore gracias a un mayor conocimiento de los mecanismos de defensa de la planta y las características de los patógenos involucrados en la colonización de las plantas (Torres et

al., 2021). Entre los enfoques que se plantean para la resistencia a la agalla de la corona generada por *A. tumefaciens* se tiene a la ingeniería genética direccionada a proteínas de virulencia y funciones tumorales que contribuyen a la transferencia del material genético y formación de agallas, respectivamente. También se podrían explorar la inhibición de la expresión de oncogenes y la manipulación de genes de plantas para inhibir la síntesis de proteínas que interactúan con las proteínas de virulencia (Otten, Burr, & Szegedi, 2008). Por lo tanto, la obtención y caracterización del genoma de *Agrobacterium tumefaciens* a partir del metagenoma de tumores de rosa de plantas del país facilitará investigaciones futuras sobre la base molecular de la patogénesis y la divergencia evolutiva de las cepas existentes en los cultivos del país.

1.2 OBJETIVOS

1.2.1 Objetivo Principal

Obtener y caracterizar el genoma de *Agrobacterium tumefaciens* a partir del metagenoma de tumores de rosa.

1.2.2 Objetivos Específicos

- Analizar estructuralmente la muestra metagenómica de tumores de rosa.
- Analizar funcionalmente la muestra metagenómica de tumores de rosa.
- Ensamblar y anotar el genoma de *Agrobacterium tumefaciens* a partir del metagenoma de tumores de rosa.
- Caracterizar funcionalmente el genoma de *Agrobacterium tumefaciens*.

1.3 MARCO TEÓRICO

1.3.1 *Agrobacterium tumefaciens*

Agrobacterium tumefaciens es una bacteria Gram negativa de suelo, con forma de bacilo, motilidad y flagelos peritricos, perteneciente a la familia *Rhizobiaceae* (Finer, Fox, & Finer, 2016). Posee la capacidad de infectar plantas por medio de un proceso que implica la entrega de un segmento específico de su genoma a los núcleos de células vegetales susceptibles. El ADN transferido (T-ADN) es una región discreta del genoma bacteriano delimitada por repeticiones directas de 23 pb transportadas por el plásmido inductor de

tumores (*Ti*). El tamaño de los plásmidos *Ti* varía entre los 180 hasta 800 kb (Christie, 2009).

Los plásmidos *Ti* codifican funciones asociadas con (i) replicación y el mantenimiento del plásmido, (ii) transferencia conjugativa, (iii) virulencia, (iv) uso de opina y (v) percepción sensorial de las señales exógenas liberadas por la célula huésped y bacterias vecinas en el sitio de infección. Los genes que codifican cada una de estas funciones comúnmente están agrupadas, con excepción de la región de virulencia *vir* y el T-ADN, y las regiones *tra* y *trb* requeridas para la transferencia conyugal del plásmido. Las plantas no tienen la capacidad de metabolizar opiones. Sin embargo, el plásmido *Ti* porta genes de catabolismo de opines, que son responsables de su transporte y degradación, proporcionando de esta manera una fuente de carbono y nitrógeno para la bacteria. (Figura 1) (Gordon & Christie, 2014).

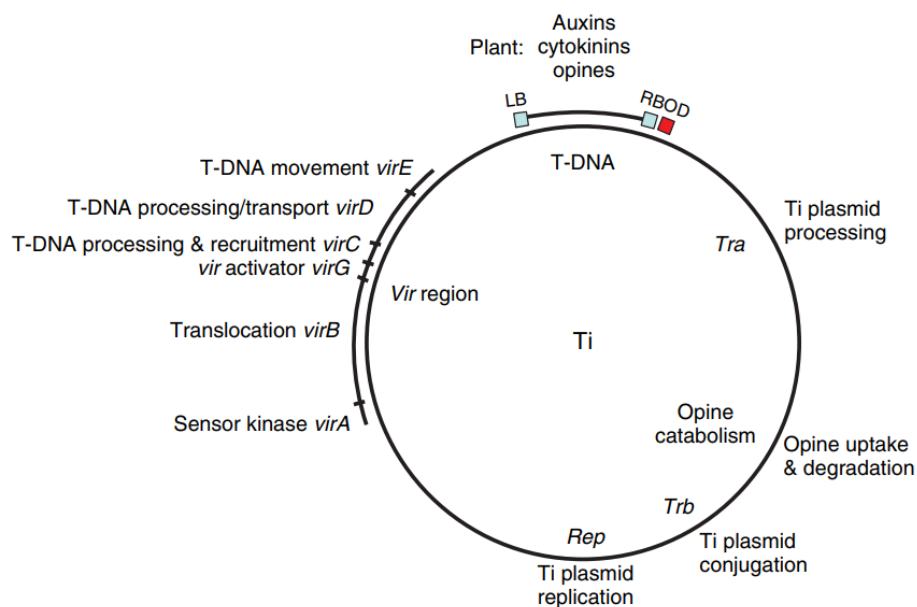


Figura 1. Regiones del plásmido *Ti* que contribuyen a la infección (región *vir* y ADN-T), supervivencia celular en el entorno tumoral (catabolismo de opiones) y transferencia conyugal del plásmido *Ti* (*tra* y *trb*) (Christie, 2009).

1.3.1.1 Proceso de infección

La patogénesis mediada por *Agrobacterium* requiere de dos elementos básicos: (i) entrega del ADN tumorigénico en el genoma de la planta (transformación); y (ii) la alteración resultante del metabolismo de las células vegetales, que en consecuencia resulta en la proliferación celular y síntesis de compuestos nutritivos que dotan de una ventaja selectiva para *Agrobacterium* (tumorigénesis) (Escobar & Dandekar, 2003).

El proceso de transformación inicia con la activación de los genes de virulencia de *Agrobacterium*. La activación de dichos genes se produce ante la presencia de compuestos fenólicos como la acetosiringona o hidroxiaacetosiringona, que son liberados por plantas que han experimentado daño mecánico (Krennek et al., 2015).

Las moléculas liberadas son reconocidas por el sistema de transducción de señales del componente VirA/VirG2. Posteriormente se produce la adhesión de *Agrobacterium* a la pared celular vegetal, lo que conlleva a la activación del operón *Vir* (localizado en el plásmido Ti) y producción de las endonucleasas específicas VirD1/VirD2. Estas proteínas tienen como objetivo procesar el T-ADN del plásmido Ti y liberar un ADN-T monocatenario (cadena ssT) por medio de un mecanismo de reemplazo de cadena. La proteína VirD2 se ensambla de forma covalente al extremo 5' de la cadena T, generando el "Complejo T inmaduro". A continuación, el complejo formado es transferido al citoplasma de la célula por medio de un sistema de secreción tipo IV conformado por 11 proteínas VirB y VirD4, localizado predominantemente en el polo de la bacteria. De forma independiente y por la misma vía, otras proteínas de virulencia (VirE2, VirE3, VirF y VirD5) son transportadas a la célula huésped. La formación del "Complejo T maduro" se da dentro de la célula huésped al ensamblar la cadena T conjugada - VirD2 con VirE2. Las proteínas VirD2 y VirE2 protegen la cadena ssT del ataque exonucleolítico dentro del citoplasma de la planta, además de que contienen señales de ubicación nuclear y guían al "Complejo T maduro" en dirección a el núcleo de la célula. La acción conjunta de factores bacterianos y vegetales dan lugar al paso del complejo a través de la membrana nuclear y posicionamiento en el sitio de integración en el núcleo. Dentro del núcleo, las proteínas acompañantes son liberadas mediante proteólisis dirigida, y el T-ADN monocatenario se convierte en una molécula de doble cadena para su integración en el genoma del huésped mediante una recombinación no homóloga. Una vez concluida la integración, inicia la expresión de genes codificados por T-ADN que conduce a la síntesis de proteínas que promueven la formación de tumores (Figura 2) (Chi et al., 2004; Păcurar et al., 2011).

La tumorigénesis se produce gracias a que los genes presentes en el T-ADN de *Agrobacterium* poseen motivos *cis*, necesarios para la expresión en la célula huésped. Existen dos clases generales de genes en el T-ADN: oncogenes y genes relacionados con la opina. Los oncogenes alteran la síntesis y sensibilidad de fitohormonas en la célula infectada, generando así el fenotipo tumoral. La acumulación masiva de auxinas y

citoquinas generada por la actividad de los oncogenes *iaaM*, *iaaH* e *ipt* conducen a una división celular descontrolada y a la tumorigenicidad. Además, se infiere que los oncogenes *6b* y *5* alteran esencialmente los efectos de las fitohormonas en la célula, aumentando la sensibilidad a las auxinas y disminuyendo la sensibilidad a las citoquinas. En el caso de los genes relacionados con la opina, no se encuentran involucrados directamente en la tumorigénesis. Sin embargo, los opiones proveen un sustrato de crecimiento para *Agrobacterium* y promueven el intercambio conyugal del plásmido Ti y la quimiotaxis (Escobar & Dandekar, 2003).

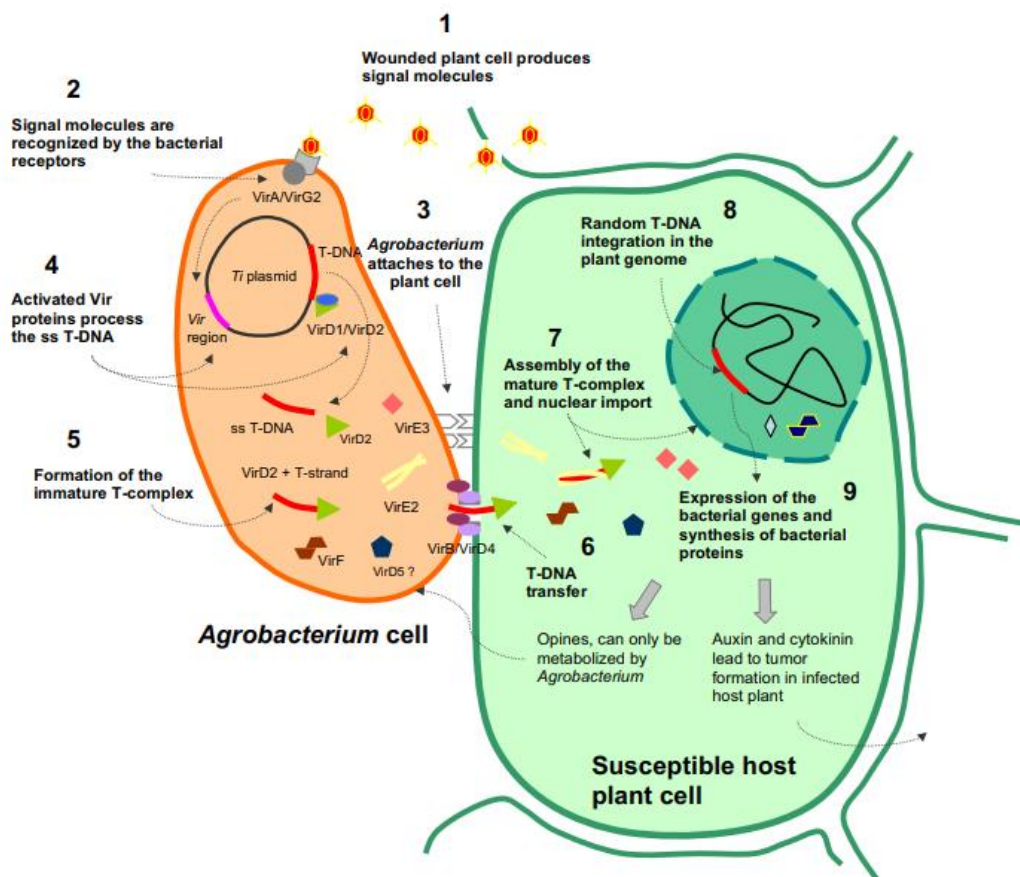


Figura 2. Modelo del proceso de transformación mediado por *Agrobacterium* (Păcurar et al., 2011).

1.3.2 Enfermedad de la agalla de la corona

La agalla de la corona es una enfermedad tumoral vegetal causada por *A. tumefaciens*. El rango de huéspedes de la enfermedad de agalla de la corona está limitada a plantas dicotiledóneas, con ciertas excepciones entre las monocotiledóneas y las gimnospermas (Chandrasekaran et al., 2019). En el caso del género *Rosa* la mayoría de sus especies son susceptibles a la agalla de la corona (Chen et al., 2020).

Las agallas (tumores) son el resultado del crecimiento descontrolado de tejido vegetal no diferenciado en el sitio de infección de la bacteria. La infección en rosas generalmente sucede a nivel de la corona de la planta, es decir en la línea del suelo, pero también puede ocurrir en ramas o raíces. La generación de agallas puede suceder a los 14 días, así como 6 meses después del ingreso de la bacteria en la planta (Duman et al., 2018). En un inicio las agallas aparecen como crecimientos pequeños, redondos, blanquecinos y blandos, pero a medida que aumentan de tamaño comúnmente sus superficies se tornan retorcidas y los tejidos superficiales toman una coloración marrón debido a la muerte y descomposición de las células periféricas, la descomposición de estos tejidos libera bacterias causantes de la enfermedad al suelo donde pueden ser transportadas en el agua e infectar nuevas plantas (Agrios, 2005). En la Figura 3 se esquematiza el ciclo de la enfermedad de la agalla de la corona causada *A. tumefaciens*.

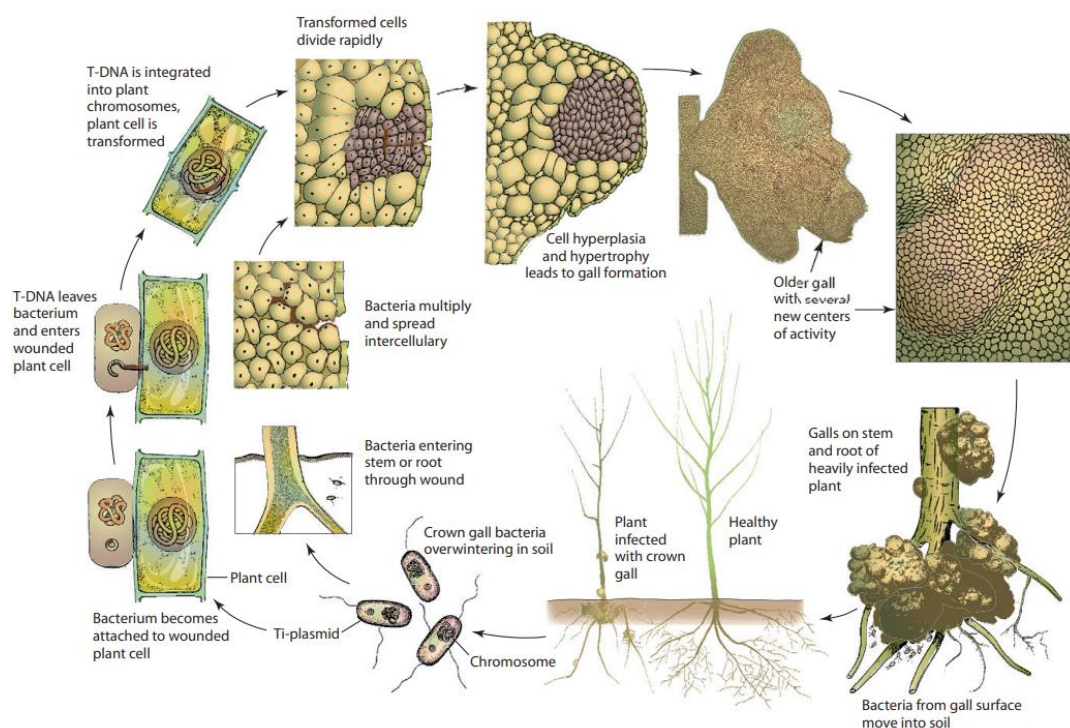


Figura 3. Etapas de desarrollo de la enfermedad de agalla de la corona producida por *A. tumefaciens* (Agrios, 2005).

Debido a que la infección solamente se da a través de una herida, la propagación de la enfermedad comúnmente es el resultado de prácticas hortícolas como la preparación del campo, la poda y el riego. Los insectos, nematodos y materiales de injerto también pueden propagar la bacteria. Además, se debe recalcar que *A. tumefaciens* tiene la capacidad de

sobrevivir en el suelo por hasta tres años, lo que perjudica notablemente las prácticas de manejo de la enfermedad (Duman et al., 2018).

A pesar de que la enfermedad de agalla de corona comúnmente no es fatal a menos que la infección se genere en plantas jóvenes, existe una disminución en el rendimiento y/o vigor del cultivo. La reducción de la productividad de las plantas agalladas puede deberse a factores como la disminución del flujo de agua y nutrientes en consecuencia de la vasculatura dañada o constreñida en el sitio de desarrollo de las agallas, así como a la asignación de agua y nutrientes a las mismas. Adicionalmente las agallas de corona son sitios de infección secundaria para otros fitopatógenos o plagas, que pueden incrementar la susceptibilidad de las plantas al estrés abiótico (Escobar & Dandekar, 2003).

1.3.3 Metagenómica

El término *metagenoma* se refiere a la suma de los genomas de todos los organismos presentes en una muestra ambiental. La *metagenómica* por tanto es el estudio de una colección de material genético de una comunidad mixta de organismos. Las principales áreas de estudio de la metagenómica son la diversidad microbiana, la estructura de la población, las relaciones genéticas y evolutivas, la actividad funcional y las relaciones de cooperación y la relación con el medio ambiente. La investigación metagenómica se ha desarrollado rápidamente en áreas como la medicina, agricultura, protección ambiental, entre otras (Zhang et al., 2021).

En los últimos años, han surgido una serie de nuevas tecnologías de secuenciación que incrementan la viabilidad de los proyectos metagenómicos. Estas tecnologías proporcionan una secuenciación más barata, más rápida y de mayor rendimiento. La diversidad y la abundancia no uniforme de las comunidades microbianas hacen que la secuenciación de alto rendimiento sea fundamental para alcanzar una cobertura adecuada de los miembros de una comunidad. Dado que cada tecnología de secuenciación presenta distintos beneficios y dificultades, la selección de la tecnología dependerá de las características del proyecto y en última instancia puede determinar el éxito relativo del mismo (Bragg & Tyson, 2014).

Las tecnologías de secuenciación de segunda generación (SGS) han dominado el mercado gracias a su capacidad de producir enormes volúmenes de datos a un bajo precio, siendo los secuenciadores Illumina los más utilizados. Sin embargo, cuando existe la presencia

de elementos repetidos en las secuencias el ensamblaje puede verse comprometido. Las tecnologías de tercera generación (TGS) basada en lecturas largas surgieron como una solución prometedora a los problemas que presentaban las SGS. Las TGS son utilizadas para la reconstrucción de regiones altamente contiguas en genomas eucarióticos y genomas microbianos de *novο* con alta precisión, siendo PacBio y Oxford Nanopore las tecnologías más importantes de esta generación. No obstante, las lecturas largas presentan tasas de error por secuencias de hasta 10% a 15%, por lo que es necesario implementar una etapa preliminar de corrección antes o después de un proceso de ensamblaje (Dominguez Del Angel et al., 2018).

1.3.3.1 Ensamblaje y anotación de genomas

Independiente de la tecnología de secuenciación seleccionada el flujo de trabajo para el ensamblaje de genomas será el mismo al mostrado en la Figura 4. La etapa inicial corresponde al control de calidad, las lecturas obtenidas de la secuenciación son examinadas para determinar su calidad, presencia de adaptadores y contaminantes. Durante la etapa de ensamblaje, frecuentemente se prueban distintos ensambladores, para después comparar los resultados en la etapa de validación del ensamblaje, donde a su vez se identifica y corrige los errores de ensamblaje. Un ensamblaje óptimo es aquel que se forma con las secuencias ensambladas más largas posibles con el menor número de errores de ensamblaje (Dominguez Del Angel et al., 2018).

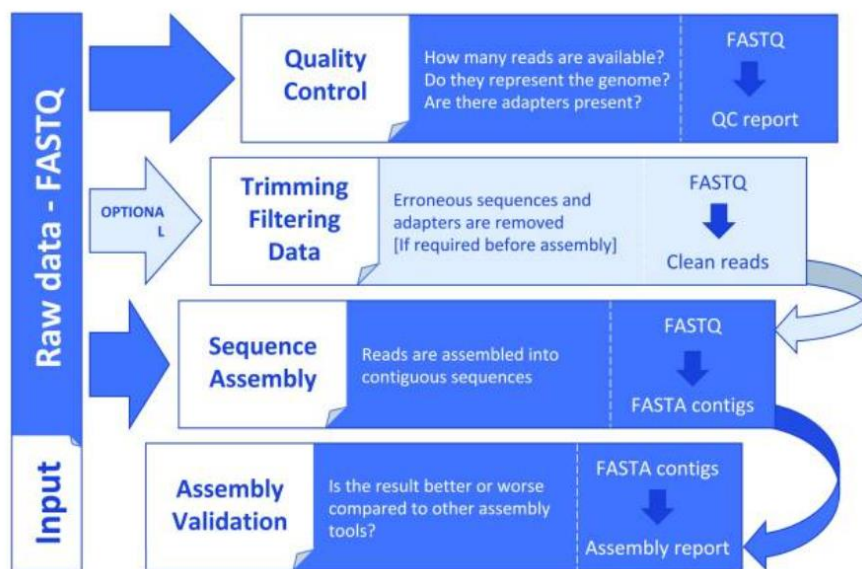


Figura 4. Pasos generales en un flujo de trabajo para el ensamblaje de genomas (Dominguez Del Angel et al., 2018).

La anotación del genoma es el proceso de derivar la información estructural y funcional de una proteína o gen a partir de un conjunto de datos sin procesar por medio de diferentes técnicas de análisis, comparación, estimación, precisión y otras técnicas de minería. La anotación del genoma es fundamental ya que la secuenciación de genoma o ADN produce secuencias sin información acerca de sus características estructurales y roles funcionales (Harbola et al., 2022).

CAPÍTULO 2: MATERIALES Y MÉTODOS

Para el presente proyecto se utilizó una muestra metagenómica obtenida a partir de tumores de rosa perteneciente a una empresa florícola del país, misma que fue secuenciada con tecnología Illumina paired-end. Las lecturas del hospedero fueron mapeadas y filtradas previamente por el laboratorio IDgen.

Los análisis se realizaron en dos ambientes: (i) en el terminal de Linux y (ii) en la plataforma del U.S. Department of Energy Systems Biology Knowledgebase (KBase, <http://www.kbase.us/>). Dentro de KBase se utilizó el protocolo publicado por Chivian et al. (2023) que conjugan las narrativas 33233 y 62384, para la extracción y análisis del genoma ensamblado en metagenomas a partir de microbiomas.

2.1 EVALUACIÓN DE CALIDAD

Para garantizar la calidad de los datos se realizó un control de calidad para eliminar bases de baja calidad, adaptadores y posibles lecturas contaminantes. Inicialmente se importó las lecturas crudas en formato FASTQ dentro de Kbase como un objeto 'Reads Library'. La evaluación de calidad de las lecturas se ejecutó mediante la aplicación FastQC - v0.12.1 (Andrews, 2010).

2.2 EVALUACIÓN DE LA ESTRUCTURA TAXONÓMICA

La evaluación de la estructura taxonómica del metagenoma se realizó en Kbase con las aplicaciones Kaiju - v1.7.3 (Menzel, Kim, & Krogh, 2016) y GOTTCHA2 - v2.1.7 (Freitas et al., 2015). En el caso de Kaiju se valoró la presencia de organismos bacterianos utilizando la base de datos RefSeq Genomes (no Euks), mientras que GOTTCHA2 evaluó la presencia de hongos en la muestra con la base de datos GOTTCHA2 fungal db. La selección del clasificador se llevó a cabo siguiendo las recomendaciones establecidas en el protocolo de Chivian et al. (2023). Donde se destaca que, en caso de que se tenga alguna sospecha de presencia de hongos en la muestra, GOTTCHA2 - v2.1.7 cuenta con una base de datos optimizada especialmente para hongos, lo que permite lograr una clasificación más precisa de las lecturas.

2.3 ENSAMBLAJE DEL METAGENOMA

Se utilizó tres ensambladores distintos: MEGAHIT - v1.2.9 (Li et al., 2015), IDBA-UD - v1.1.3 (Peng et al., 2012) y metaSPAdes - v3.15.3 (Nurk et al., 2017). Posteriormente se examinará y comparará los ensamblajes obtenidos utilizando la herramienta QUASt - v4.4 (Gurevich et al., 2013) por medio de métricas como N50 y L50. Además, se utilizó la variación de los informes de QUASt presente en Kbase denominada Compare Assembled Contig Distributions - v1.1.2 (Arkin et al., 2018) para seleccionar el ensamblaje más apropiado con el que se continúe el análisis.

A continuación se realizó el agrupamiento de bins por medio de las aplicaciones MaxBin2 - v2.2.4 (Wu, Simmons, & Singer, 2016), MetaBAT2 - v1.7 (Kang et al., 2019) y CONCOCT - v1.1 (Alneberg et al., 2014). Los outputs obtenidos de cada una de las aplicaciones sirvieron como input para la herramienta DAS-Tool - v1.1.2 (Sieber et al., 2018), como resultado se obtuvo bins optimizados como resultado del consenso de los resultados de las aplicaciones de agrupamiento individuales. Se evaluó la calidad de los bins por medio de CheckM - v1.0.18 (Parks et al., 2015), entre los resultados se obtuvo una representación gráfica donde se pudo observar los marcadores filogenéticos para cada bin y una tabla donde para cada bin se asignó un clado taxonómico, número de recuentos, integridad y contaminación.

Finalmente se realizó la asignación taxonómica de los bins utilizados por medio de la herramienta GTDB-Tk - v1.7.0 (Chaumeil et al., 2020).

2.4 ANOTACIÓN DEL METAGENOMA

La anotación del metagenoma se realizó en las herramientas Prokka - v1.14.5 (Seemann, 2014) y RAST - v1.073 (Aziz et al., 2008) teniendo como dato de entrada el mejor ensamblaje obtenido previamente. El resultado de RAST - v1.073 se cargó en la herramienta DRAM (Shaffer et al., 2020) para una mejor visualización.

2.5 ENSAMBLAJE DE GENOMAS Y MAPEO DE GENES DE VIRULENCIA

Debido a que ninguno de los bins generados desde el ensamblaje del metagenoma correspondió al genoma de *Agrobacterium tumefaciens*, se procedió a mapear su genoma de referencia contra las lecturas crudas con ayuda del programa Bowtie2-2.5.1

(Langmead & Salzberg, 2012). Adicionalmente se mapeo el genoma de referencia de *Agrobacterium larrymoorei* dado los resultados obtenidos en el análisis taxonómico del metagenoma.

En virtud de que no se logró ensamblar el genoma completo tanto de *A. tumefaciens* como de *A. larrymoorei*. Se utilizó el programa Geneious Prime 2023 (Biomatters Ltd) para mapear las lecturas crudas contra dos genes de virulencia (PthG y HsvG) para *Pantoea agglomerans* y una base de datos de plásmidos *Ti* generada por el laboratorio IDgen, con el objetivo de determinar el organismo causante de tumores de rosa en la muestra analizada.

2.6 COMPARACION DE LOS PERFILES TAXONÓMICOS BACTERIANOS DE LOS METAGENOMAS DE TUMOR DE ROSA Y DE TEJIDO SANO.

La comparación de los perfiles taxonómicos se llevó a cabo utilizando la herramienta Kaiju - v1.7.3, en combinación con la base de datos RefSeq Genomes (no Euks). Se realizó un análisis conjunto de los metagenomas obtenidos de las muestras de tumor de rosa y de tejido sano. La muestra metagenómica de tejido sano (*in vitro*) provino de la misma empresa florícola que la de tumor de rosa, y de igual manera fue secuenciada con tecnología Illumina paired-end. Es relevante mencionar que previamente, el laboratorio IDgen realizó el mapeo y filtrado de las lecturas del hospedero.

CAPÍTULO 3: RESULTADOS Y DISCUSIÓN

3.1 CONTROL DE CALIDAD DE LAS LECTURAS DE SECUENCIACIÓN

Se trabajó con un total de 13 726 514 lecturas paired-end con una longitud de 150 bp. Utilizando el programa FastQC se efectuó el control de calidad de las lecturas - v0.12.1, la Figura 5 muestra la calidad de las secuencias por base. En la gráfica el eje Y contiene las puntuaciones de calidad, mientras que el eje X la posición en la lectura.

Las puntuaciones de calidad (Phred score) indican la probabilidad de que una base haya sido llamada correctamente. Por lo general las puntuaciones varían entre 2 y 40, puntuaciones altas indican una mayor confianza en la llamada (Sathyanarayanan et al., 2019). La gráfica se divide en zonas de color en función a rangos de calidad, el color verde representa una muy buena calidad (>28), el color naranja calidad razonable (20-28) y el color rojo mala calidad (0-20). La calidad de las lecturas tanto forward y reverse se encontraron en la zona verde, por sobre la puntuación de 34; por tanto se las consideró de alta calidad.

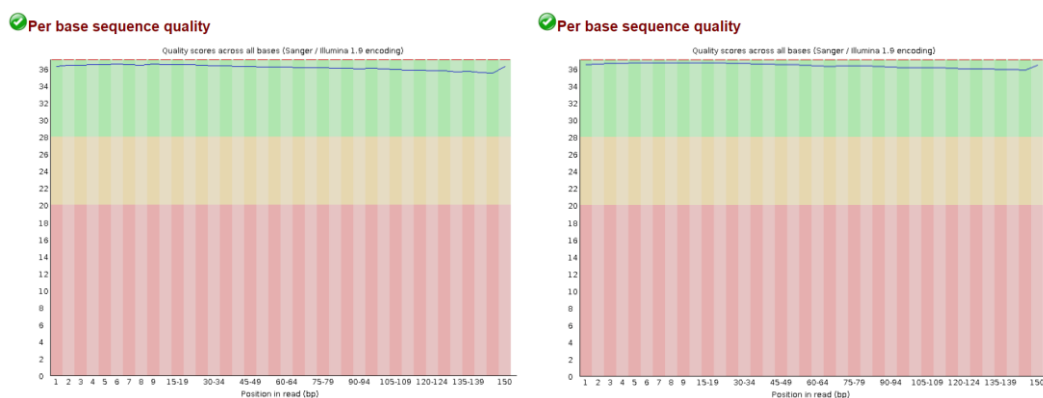


Figura 5. Calidad de las secuencias por base para las lecturas forward (izq.) y reverse (der.).

El contenido de GC de las lecturas se observa en la Figura 6. La curva roja corresponde al contenido de GC de la muestra, en tanto que la curva azul representa una distribución normal de referencia del contenido de GC calculada a partir de los datos de entrada. El pico central indica el contenido general de GC, las lecturas forward y reverse presentaron su pico más alto en 44%. Las curvas de las lecturas presentaron forma similar a la curva de referencia y no se observó picos adicionales o desplazamiento sobre el eje X.

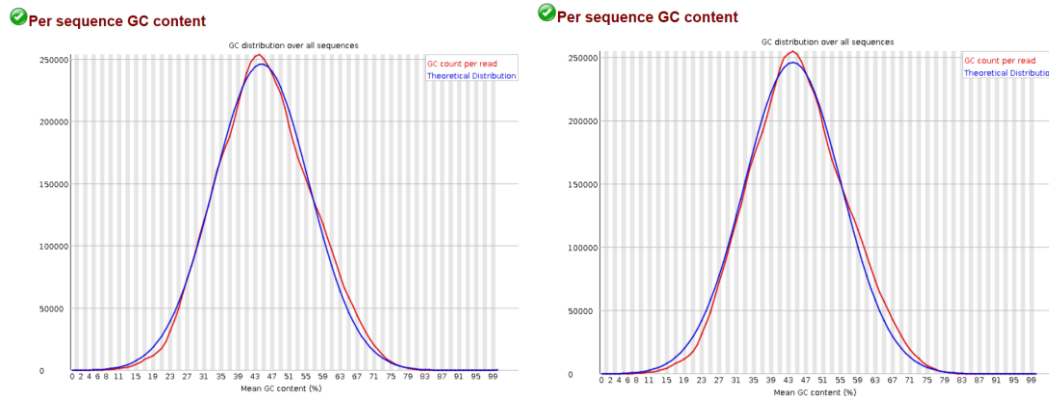


Figura 6. Contenido de GC para las lecturas forward (izq.) y reverse (der.).

Adicionalmente, el reporte generado por FastQC mostró que tanto las lecturas forward y reverse obtuvieron una calidad promedio de 36 (Figura 7) y que no existió la presencia de adaptadores. Por consiguiente, al integrar estos resultados con los expuestos previamente, se infiere que las lecturas son de buena calidad y no requirieron someterse a ningún procedimiento de preprocesamiento.

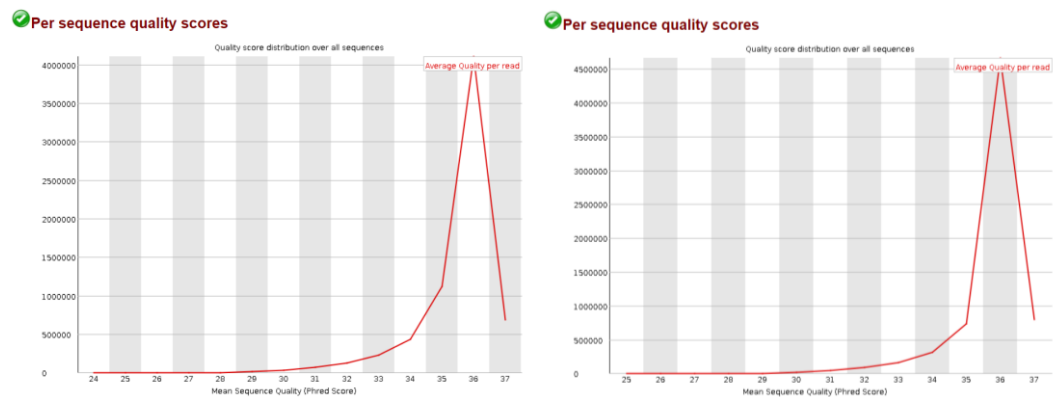


Figura 7. Puntuaciones de calidad por secuencia para las lecturas forward (izq.) y reverse (der.).

3.2 CLASIFICACIÓN TAXONÓMICA DEL METAGENOMA

Posterior al control de calidad se realizó la clasificación taxonómica del metagenoma por medio de la herramienta Kaiju - v1.7.3 utilizando la base de datos RefSeq Genomes (no Euks). La clasificación a nivel de filo se observa en la Figura 8, donde se registra que el porcentaje de lecturas clasificadas fue menor al 20%.

Dentro de las lecturas clasificadas, Proteobacteria fue el filo con mayor presencia seguido de Cyanobacteria, Actinobacteria, Firmicutes y Bacteroidetes. Además, se registró un

pequeño porcentaje de virus, así como filos que tienen una presencia menor al 0.5% y lecturas que no fueron asignadas a ningún filo.

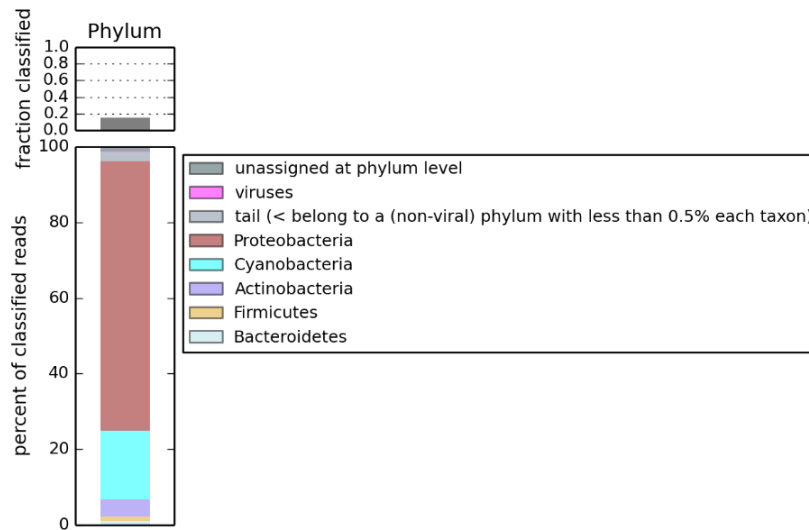


Figura 8. Clasificación taxonómica a nivel de Filo.

El alto porcentaje de lecturas sin clasificar puede deberse principalmente a dos factores que se analizan a continuación. Considerando que Kaiju es un clasificador taxonómico basado en proteínas, su principal desventaja es la incapacidad de clasificar lecturas que se originan en regiones genómicas no codificantes (Menzel, Kim, & Krogh, 2016). Por tanto, los clasificadores basados en proteínas tienden a presentar muchas más lecturas sin clasificar, en comparación con los clasificadores de ADN. No obstante, pueden ser más sensibles a secuencias nuevas y variables debido a que las secuencias de aminoácidos tienen tasas de mutación más bajas que las secuencias de nucleótidos (Ye et al., 2019).

Los metagenomas asociados a un huésped suelen contener una proporción significativa de secuencias contaminantes generadas a partir de genoma del huésped (Mande, Mohammed, & Ghosh, 2012). Dado el alto porcentaje de lecturas no clasificadas se infiere que, lecturas contaminantes del huésped persistieron después del mapeo y remoción de lecturas concordantes con el genoma de *Rosa* sp.

Otros elementos que pueden incidir en el establecimiento de la diversidad taxonómica de una comunidad son: la eficiencia de los pasos experimentales iniciales (recolección de muestras, preparación y extracción de ADN), el proceso de secuenciación puede pasar por alto microorganismos de baja abundancia y que no existe un software “gold standard”

para el procesamiento de secuencias huésped (Mande, Mohammed, & Ghosh, 2012; Zhang et al., 2021).

Las categorías taxonómicas de clase, orden, familia, género y especie se muestran en las Figuras 9, 10, 11, 12 y 13, respectivamente.

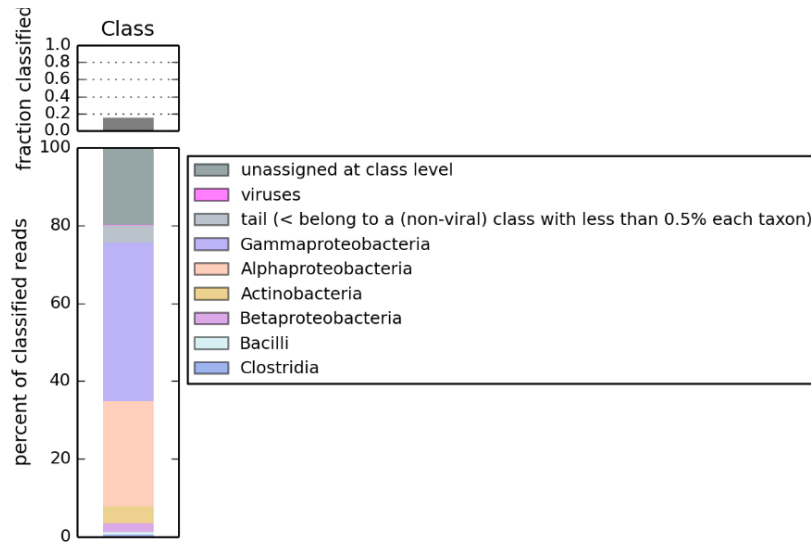


Figura 9. Clasificación taxonómica a nivel de Clase.

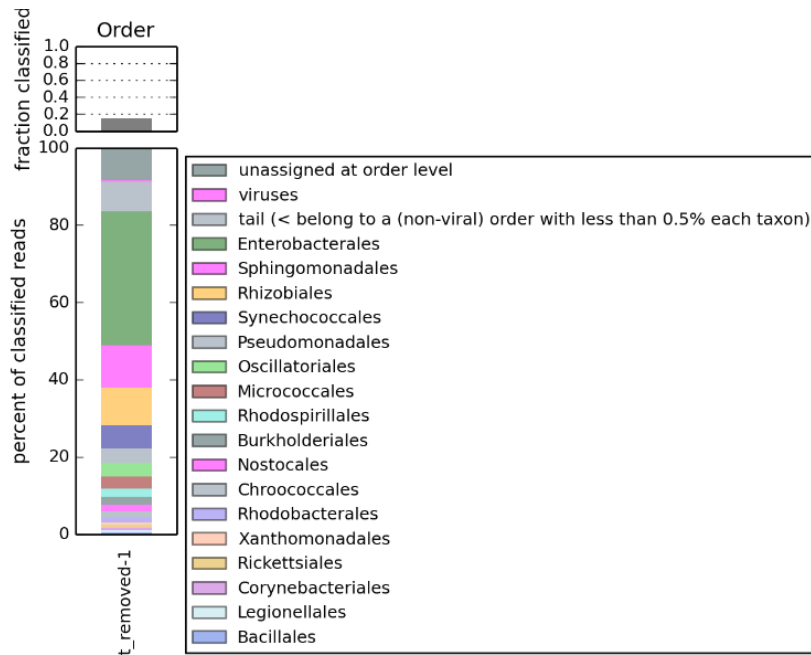


Figura 10. Clasificación taxonómica a nivel de Orden.

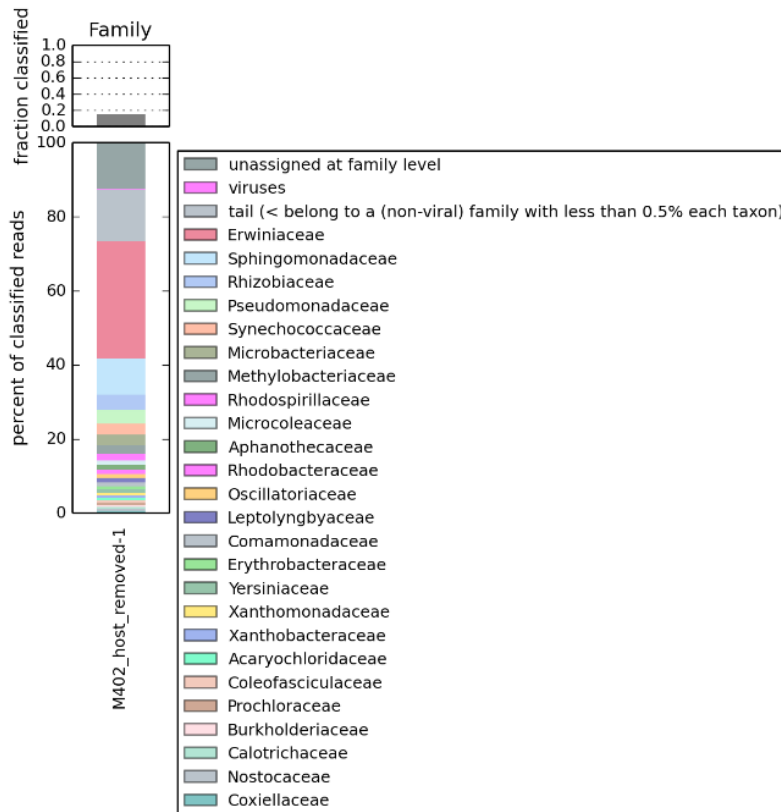


Figura 11. Clasificación taxonómica a nivel de Familia.

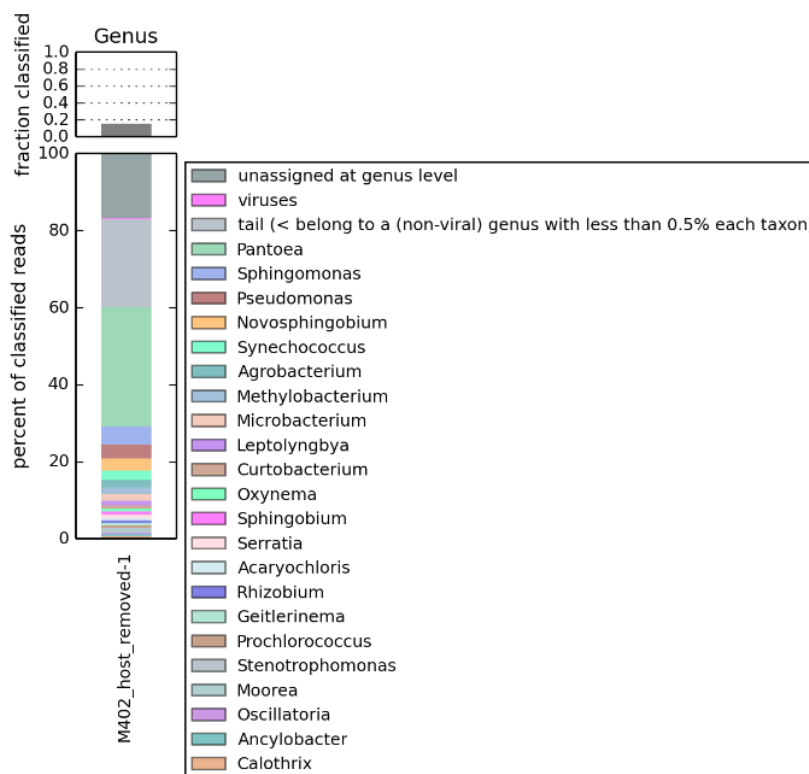


Figura 12. Clasificación taxonómica a nivel de Género.

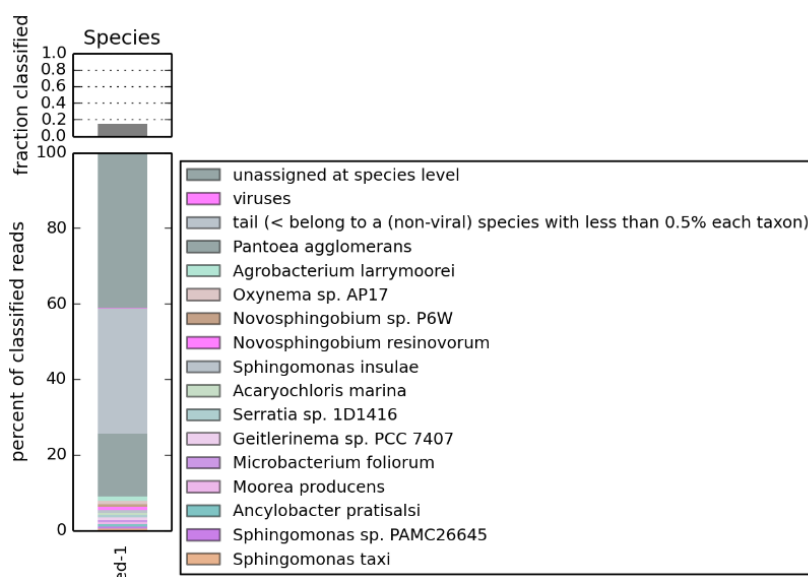


Figura 13. Clasificación taxonómica a nivel de Especie.

Un total de 22 géneros fueron asignados siendo *Pantoea*, *Sphingomonas*, *Pseudomonas*, *Novosphingobium*, *Synechococcus* y *Agrobacterium* los taxones con mayor representación.

A nivel de especie los taxones con mayor presencia fueron *Pantoea agglomerans*, *Agrobacterium larrymoorei*, *Oxynema sp. AP17*, *Novosphingobium sp. P6W*, *Novosphingobium resinovorum* y *Sphingomonas insulae*, de un total de 14.

En relación con las especies de mayor presencia se puede destacar los siguientes aspectos. *Pantoea agglomerans* se encuentra altamente distribuida en la naturaleza y ha sido aislada de plantas, humanos, animales y agua. Comúnmente se relaciona con plantas como una bacteria endófito o epífita. Sin embargo, se han registrado aislados patógenos tumorigénicos que inducen la generación de agallas en plantas de gypsophila, remolacha, abeto de Douglas, glicina y arándano (Weinthal et al., 2007). No se encontró estudios que relacionen a tumores en rosas con *P. agglomerans*.

Agrobacterium larrymoorei provoca tumores aéreos en plantas de *Ficus benjamina* (Bouzar & Jones, 2001). No obstante, no existen registros de otros hospederos para esta especie.

Por otra parte *Novosphingobium* sp. P6W y *Novosphingobium resinovorum* no han sido vinculadas con la generación de tumores en plantas, pero presentan potencial como promotores de crecimiento vegetal.

Novosphingobium sp. P6W se caracteriza por utilizar el ABA (Ácido abscísico) como fuente de carbono y energía. La bioconversión de ABA puede ser relevante para preservar niveles bajos de esta fitohormona en el suelo y contrarrestar los efectos adversos de las concentraciones altas de ABA, en la germinación de semillas y desarrollo de raíces (Gogoleva et al., 2019). Por otra parte, *Novosphingobium resinovorum* ha mostrado la capacidad de solubilizar fosfato inorgánico y producir sideróforos (Rincon-Molina et al., 2021).

En el caso de *Sphingomonas insulae* sp y *Oxynema* sp. AP17 no se encontró investigaciones relacionados con plantas.

Con la herramienta GOTTCHA2 - v2.1.7 se identificó únicamente la presencia *Sarocladium strictum* en la muestra.

Considerando que no existe un motivo específico para seleccionar una herramienta de clasificación particular para cada grupo taxonómico, a continuación en la Figura 14 se muestra dendograma generado por GOTTCHA2 - v2.1.7 para bacterias utilizando la base de datos Viral and bacterial db, y en la Figura 15 la clasificación a nivel de especie para hongos utilizando Kaiju - v1.7.3 con la base de datos Fungi.

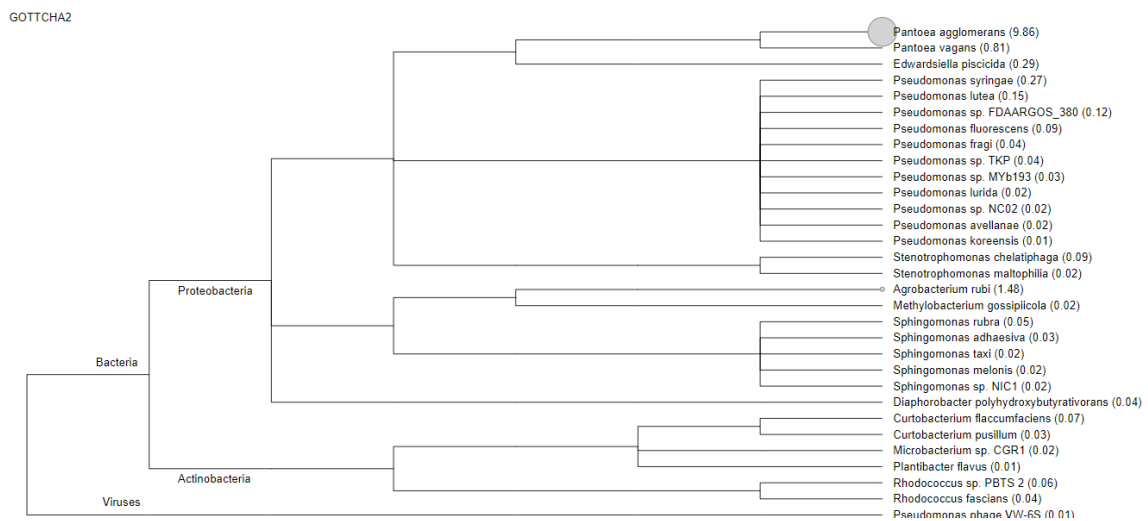


Figura 14. Dendograma generado por GOTTCHA2 - v2.1.7 para bacterias.

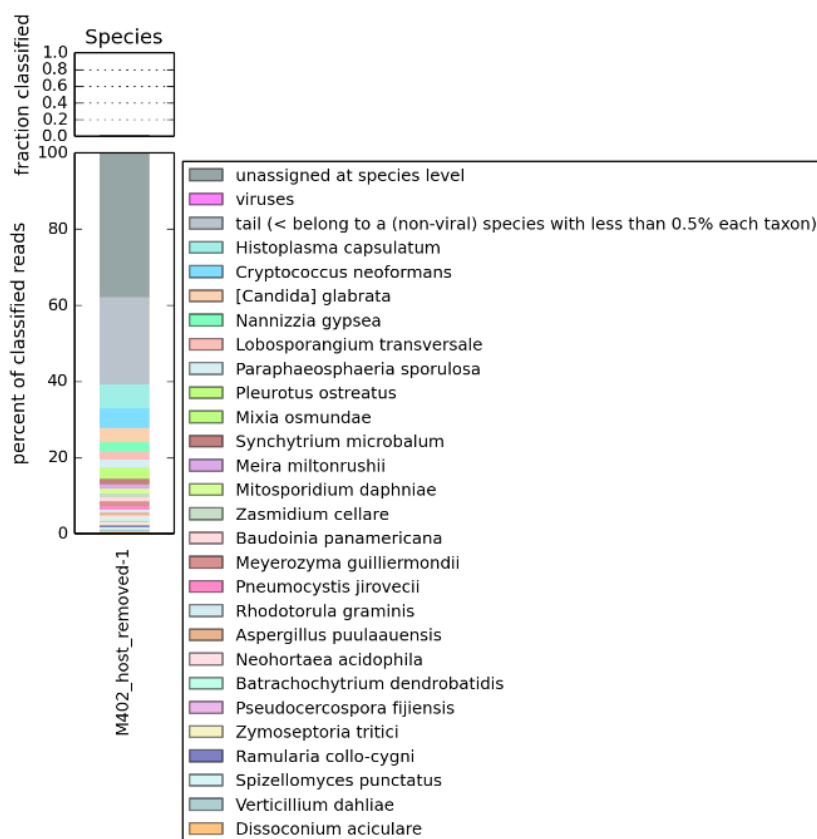


Figura 15. Clasificación a nivel de género para hongos obtenida a través de Kaiju - v1.7.3.

La clasificación obtenida a través de GOTTCHA2 - v2.1.7 reveló que los géneros más predominantes incluyen *Pantoea*, *Agrobacterium*, *Pseudomonas*, *Edwardsiella*, *Sphingomonas* y *Stenotrophomonas*; mientras que a nivel de especie, se destacaron *Pantoea agglomerans*, *Agrobacterium rubi*, *Pantoea vagans* y *Edwardsiella piscicida* como los taxones más abundantes (Figura 14).

En cuanto a la clasificación taxonómica a nivel de especie de hongos mediante Kaiju - v1.7.3, sobresalieron los taxones *Histoplasma capsulatum*, *Cryptococcus neoformans*, *Candida glabrata*, *Nannizzia gypsea* y *Lobosporangium transversale* (Figura 15).

Al comparar los resultados obtenidos con ambas herramientas para la clasificación taxonómica a nivel de bacterias, se observó que a nivel de género, los taxones *Pantoea*, *Agrobacterium*, *Pseudomonas* y *Sphingomonas* fueron los más prevalentes tanto en Kaiju - v1.7.3 como en GOTTCHA - v2.1.7. A nivel de especie, *Pantoea agglomerans* fue la especie más abundante en ambas herramientas. Sin embargo, en el caso de *Agrobacterium*, hubo variaciones en la asignación de especie entre las dos herramientas.

En lo que respecta a los demás taxones a nivel de especie, se identificaron discrepancias entre las herramientas utilizadas.

Con relación a la asignación taxonómica para hongos, se observó una considerable diferencia en los taxones asignados por las herramientas. GOTCHA2 - v2.1.7 solo asignó un taxón a *Sarocladium strictum*, mientras que Kaiju - v1.7.3 asignó un total de 25 taxones a nivel de especie.

La variabilidad en la clasificación taxonómica podría deberse a varios factores que se detallan a continuación. La precisión de la clasificación de lecturas metagenómicas, independientemente de la herramienta de clasificación empleada, depende de la calidad y exhaustividad de la base de datos de referencia, es así como los resultados de la clasificación pueden variar drásticamente en función de la base de datos de referencia empleada (Smith et al., 2022). Asimismo se ha demostrado que el número de genomas utilizados para construir el índice y el sistema taxonómico de clasificación de los genomas puede alterar significativamente las tasas de clasificación (Méric et al., 2019). Además, Nasko et al. (2019) señalan que la precisión de clasificación puede sufrir alteraciones en función de la versión de la base de datos disponible, lo cual comprobaron utilizando la base de datos RefSeq junto con herramientas de clasificación basadas en k-mer. Por tanto, estos factores pueden explicar las diferencias en la clasificación observadas en el estudio.

3.3 ENSAMBLAJE Y ANOTACIÓN DEL METAGENOMA

La herramienta Compare Assembled Contig Distributions - v1.1.2 permitió comparar los resultados de los ensamblajes obtenidos por metaSPAdes - v3.15.3, IDBA-UD - v1.1.3 y MEGAHIT - v1.2.9 (Figura 15). El ensamblador que obtuvo las mejores métricas fue metaSPAdes con un número total de contigs de 2284, un N50 de 8588, un L50 de 159 y su contig más largo fue de 670475 bp.

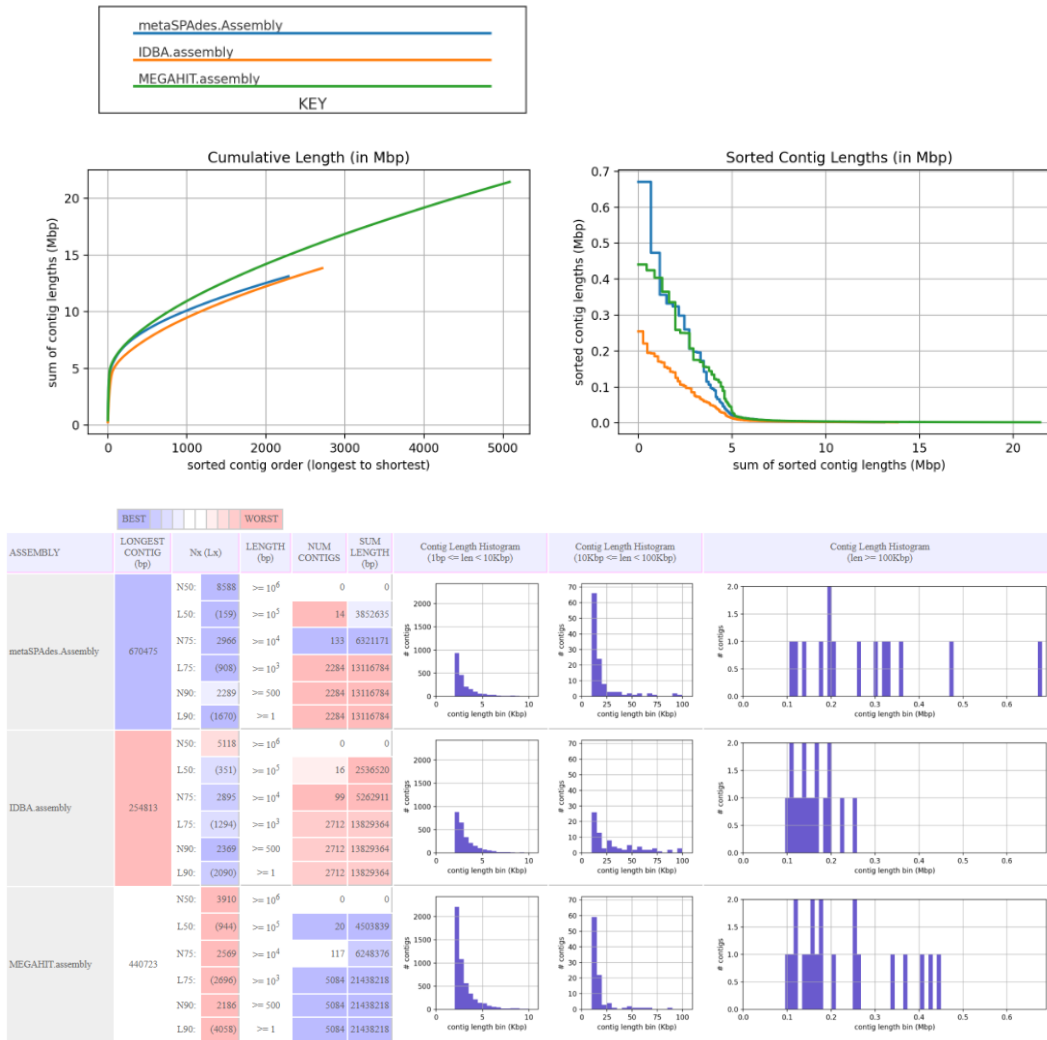


Figura 16. Comparación de las métricas de los distintos ensamblajes.

3.4 AGRUPAMIENTO Y OPTIMIZACIÓN DE BINS

El ensamblaje obtenido de metaSPAdes fue agrupado en *bins* utilizando las herramientas: CONCOCT - v1.1, MaxBin2 - v2.2.4 y MetaBAT2 - v1.7. La Tabla 1 presenta los resultados generados por cada herramienta.

Tabla 1. Resultados obtenidos por las herramientas de agrupamiento.

	Input contigs	Contigs agrupados	Número de bins
CONCOCT		846	22
MaxBin2	2284	2180	3
MetaBAT2		1350	3

Se realizó la optimización de los bins obtenidos a través de la herramienta DAS Tool - v1.1.2 que seleccionó únicamente los bins de mejor calidad. Como resultado se obtuvo 2 bins que correspondían a 2004 contigs agrupados.

Posteriormente se evaluó la calidad de los bins utilizando la herramienta CheckM - v1.0.18. En la Tabla 2 se muestra la asignación taxonómica preliminar realizada por CheckM. El genoma del bin.001 se completó en un 90.81% y presentó una contaminación del 5.24 %, su fila dentro de la tabla fue resaltada puesto que no cumplió con los umbrales de integridad y contaminación establecidos (95% y 2 %).

El bin.002 presenta una integridad de genoma del 100% y una contaminación del 0.52%. Los taxones asignados fueron Sphingomonadales (o) y Enterobacteriaceae (f), para el bin.001 y bin.002, respectivamente.

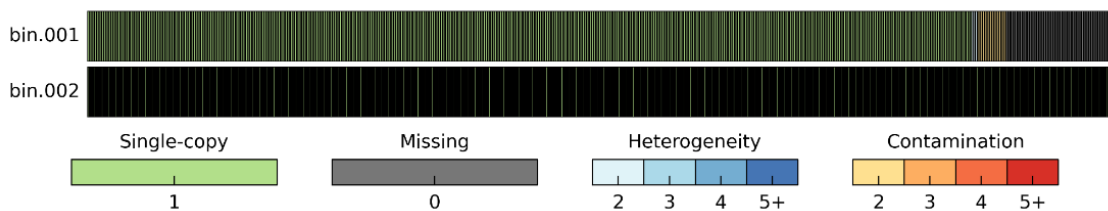


Figura 17. Representación de la integridad y contaminación de los bins optimizados.

Tabla 2. Análisis de los marcadores filogenéticos evaluados por CheckM.

Bin	Linaje	# Genomas	# Marcadores	0	1	2	% I	% C
bin.001	o_Sphingomonadales	26	569	59	491	19	90.81	5.24
bin.002	f_Enterobacteriaceae	223	876	0	872	4	100.0	0.52

*% I: Porcentaje de integridad, %C: Porcentaje de contaminación.

3.4.1 Asignación taxonómica de los bins

Los bins optimizados así como los obtenidos por cada herramienta de agrupamientos fueron asignados a un taxón utilizando la herramienta GTDB-Tk - v1.7.0. De los bins optimizados solo el bin.002 fue asignado a nivel de especie, siendo esta *Pantoea agglomerans*. La Tabla 3 muestra un resumen de los resultados generados.

Tabla 3. Asignación taxonómica de GTDB-Tk para los bins optimizados.

Bin	Clasificación	FastANI			
		Referencia	Taxonomía	ANI	F.A
bin.002	d__Bacteria p__Proteobacteria c__Gammaproteobacteria o__Enterobacterales; f__Enterobacteriaceae; g__Pantoea; s__Pantoea agglomerans	GCF_001598475.1	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Enterobacterales; f__Enterobacteriaceae; g__Pantoea; s__Pantoea agglomerans	98.49	0.94

*F.A: Fracción de alineamiento, ANI: Average Nucleotide Identity.

De los 22 bins obtenidos del agrupamiento por CONCOCT solo dos bins fueron asignados a un taxón, el bin.001 a la especie *Pantoea agglomerans* y el bin.0016 al género *Novosphingobium*. En el caso de MaxBin2 y MetaBAT2 de los 3 bins agrupados con cada herramienta solo dos fueron asignados a nivel de taxón, bin.002 y bin.001 a *Pantoea agglomerans*, y el bin.003 de las dos herramientas a *Novosphingobium*.

3.5 ANOTACIÓN DEL METAGENOMA

La anotación del metagenoma se realizó con el ensamblaje obtenido de la herramienta metaSPAdes. La herramienta Prokka identificó un total de 12351 CDSs, 7 tRNAs, 126 tRNAs y 2 tmRNAs.

En el caso RAST solamente identificó 12365 CDSs, un número similar al identificado por Prokka. Para una mejor visualización el archivo generado por RAST se visualizó en la herramienta DRAM que generó la Figuras 18 donde se presenta un mapa de calor que muestra los módulos y la cobertura de los componentes de la cadena de electrones, y la Figura 19 que contiene las funciones metabólicas presentes.

3.6 MAPEO DEL GENOMA DE *A. tumefaciens* Y *A. larrymoorei*

Puesto que ninguno de los bins generados en el ensamblaje del metagenoma fue clasificado taxonómicamente para *Agrobacterium tumefaciens*, se procedió a mapear el genoma de referencia del microorganismo (ASM366790v1) contra las lecturas crudas. De las 13,726,514 lecturas crudas un total de 3420 lecturas fueron mapeadas en pares, lo que representa aproximadamente un 0.02% del total.

Se filtraron las lecturas mapeadas y se procedió a su ensamblaje. El análisis de calidad de QUAST reportó que el ensamblaje obtenido representó apenas el 0.861% del genoma de *Agrobacterium tumefaciens*. La Tabla 4 contiene las estadísticas destacadas del reporte generado por QUAST para el ensamblaje.

Tabla 4. Análisis de calidad del ensamblaje de *Agrobacterium tumefaciens*.

Estadísticas del Genoma	
Porcentaje del genoma (%)	0.861
Alineamiento más largo	2865
Longitud total alineada	34662
# contigs completamente desalineados	19
Longitud completamente desalineada	12791
# contigs parcialmente desalineados	0
Longitud parcialmente desalineada	0
Estadísticas sin referencia	
# contigs	57
# contigs (≥ 0 bp)	353
# contigs (≥ 1000 bp)	9
Contig más largo	2865
Longitud total	47545
Longitud total (≥ 0 bp)	136418
Longitud total (≥ 1000 bp)	15152
N50	790
L50	19
GC (%)	57.97

Dado que el análisis taxonómico del metagenoma señaló la presencia de *Agrobacterium larrymoorei* en la muestra, se procedió a mapear su genoma de referencia (ASM514442v2). Un total de 4450 lecturas fueron mapeadas. Las lecturas mapeadas fueron filtradas y ensambladas para su posterior análisis de calidad en QFAST. El reporte mostró que el ensamblaje generado correspondía al 1.311% del genoma de *A. larrymoorei*, donde el alineamiento más largo fue de 2865 pb y la longitud total alineada fue de 56885 pb.

3.7 MAPEO DE LOS GENES DE VIRULENCIA DE *Pantoea agglomerans*

Dada la abundante presencia de lecturas de *P. agglomerans* y la escasa presencia de *A. tumefaciens* en la muestra, surgió la sospecha de que *P. agglomerans* podría ser el agente responsable de los tumores en la planta. En consecuencia, se llevó a cabo el mapeo de los genes de virulencia de *P. agglomerans* para validar o refutar esta suposición.

La transformación de *P. agglomerans* en un patógeno vegetal se dio por medio de la adquisición de plásmidos de patogenicidad pertenecientes a la familia pPATH (plásmido de patogenicidad) que contiene una isla de patogenicidad (PAI). *P. agglomerans* pv. *gypsophila* 824-1 (*Pag*) y *P. agglomerans* pv. *betae* 4188 (*Pab*) son las cepas patógenas mejor estudiadas (Gutiérrez-Barranquero et al., 2019).

Los genes seleccionados relacionados con la virulencia fueron PthG (AF071231.3) y HsvG (AF071524.1). HsvG y PthG son efectores de tipo 3 (T3Es) presentes en *Pag* que incitan la formación de agallas en plantas de *gypsophila*. HsvG imita los activadores transcripcionales específicos del huésped y determina la especificidad del patovar, mientras que PthG apoya el progreso de la infección (Nissan et al., 2019).

El mapeo se realizó en el software Geneious Prime 2023. No se encontró lecturas que coincidan con la secuencia de HsvG. En el caso de PthG 167 lecturas fueron mapeadas, sin embargo no corresponden a la región CDS del gen (Figura 20). De esta forma se descartó que *P. agglomerans* fuera la bacteria responsable de la generación de tumores para la muestra analizada.

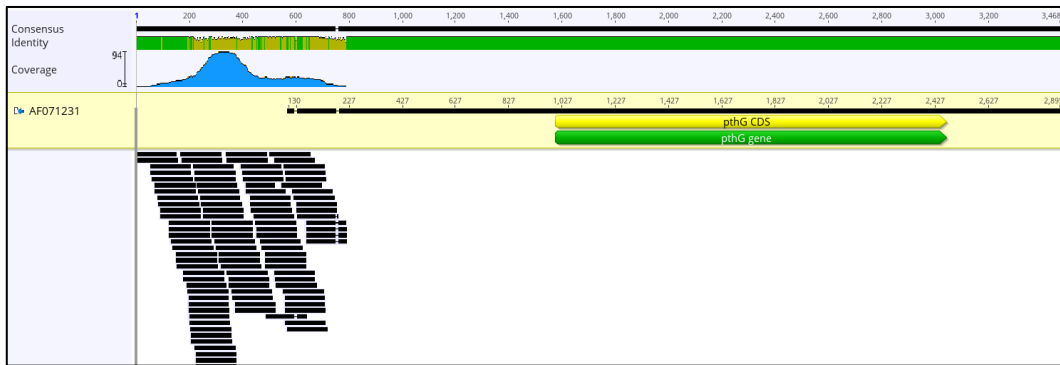


Figura 20. Mapeo del gen PthG contra las lecturas crudas.

Es importante destacar que, a pesar de la presencia de dos especies del género *Novosphingobium* en el análisis taxonómico del metagenoma y la asignación de un bin a nivel de género, no se estimó a *Novosphingobium* para un posterior análisis, ya que no se encontraron reportes previos que relacionen a especies de este género como agentes responsables de la formación de tumores en plantas. No obstante, Gan et al. (2009) han reportado que la cepa Rr 2-17 de *Novosphingobium* sp. aislada de un tumor de agalla de corona de vid, puede influir en la virulencia de *A. tumefaciens* y *A. vitis*, mediante la producción de N-acil homoserina lactonas que activan el receptor de la proteína TraR, involucrada en la regulación del quorum-sensing para la replicación y conjugación del plásmido *Ti* en tumores de plantas.

3.8 MAPEO DE PLÁSMIDOS *Ti* DE *A. tumefaciens*

Al descartar que la formación de tumores se debe a *P. agglomerans*, *A. tumefaciens* se convirtió en la única bacteria presumiblemente responsable de los tumores en la muestra.

Las lecturas crudas fueron mapeadas contra una base de datos de plásmidos *Ti* generada por el laboratorio IDgen. Como resultado un total de 3472 lecturas fueron mapeadas en pares para el plásmido pTiKerr108, se realizó el ensamblaje de estas y el análisis de calidad se ejecutó en QUILT. El reporte reveló que el ensamblaje generado corresponde al 69.23% de la secuencia completa del plásmido pTiKerr108 (MK439384.1) (Figura 21) con una longitud total del alineamiento de 152808 pb.

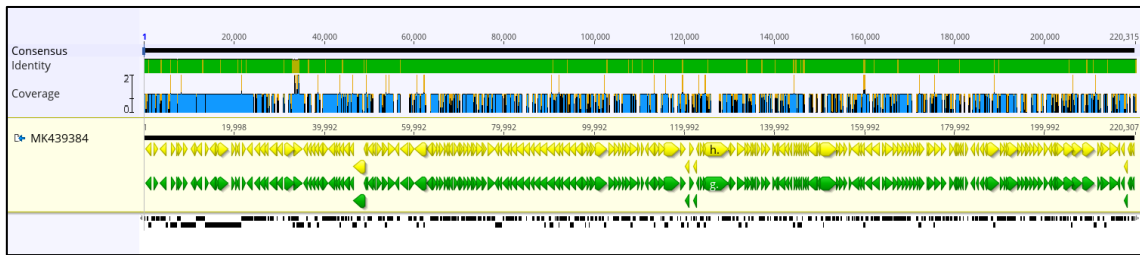


Figura 21. Alineamiento del ensamblaje con la secuencia completa del plásmido pTiKerr108.

pTiKerr108 es un plásmido *Ti* de nopalina (Shao & Hooykaas, 2019). La nopalina es un tipo de opina producida por las células vegetales transformadas que contienen el T-DNA y sirve como fuente de nutrientes para *Agrobacterium*. Adicionalmente se desempeñan como señales que activan el quorum-sensing de *Agrobacterium*, forma de comunicación célula-célula mediante señales que permite la detección de células circundantes y coordinar su comportamiento colectivo (Subramoni et al., 2014).

El análisis de metagenomas de plantas puede ser desafiante puesto que las plantas poseen genomas de gran tamaño, y pueden abrumar la representación genómica de la comunidad microbiana en el metagenoma. En consecuencia, el material genético de la planta será detectado en mayor medida que el de otros microorganismos presentes. Para contrarrestar este efecto se han desarrollado protocolos a nivel de laboratorio que permiten separar físicamente las microbiota del tejido vegetal, lo que permite mejorar la eficacia de las investigaciones metagenómicas en sistemas vegetales (Sharpton, 2014).

Además, los distintos niveles de abundancia de las especies en una muestra metagenómica dan como resultado una cobertura de lectura no muy uniforme en diferentes genomas (Lapidus & Korobeynikov, 2020). Esto sumado a que gran parte de las comunidades microbianas son tan diversas que muchos de los genomas pueden no estar completamente representados por las lecturas (Sharpton, 2014).

El crecimiento bacteriano y tamaño de la población de *A. tumefaciens* en el tejido vegetal está supeditada a los niveles de ciertas fitohormonas de la planta. Niveles altos de auxina generados por los tejidos afectados por tumores, evitan un incremento en la tasa de transformación al inhibir el crecimiento bacteriano y la virulencia. El ácido salicílico (SA) es usado por la planta como mecanismo de defensa, ya que reprime la manifestación de los genes *vir*, el crecimiento bacteriano, la unión bacteriana a las células vegetales y la virulencia (Hwang, Yu, & Lai, 2017).

Los aspectos mencionados pudieron contribuir a la limitada cantidad de lecturas que se obtuvieron para *A. tumefaciens*, lo que condicionó la recuperación de su genoma completo. Sin embargo, se pudo corroborar su presencia en la muestra por medio del mapeo del plásmido pTiKerr108.

3.9 COMPARACION DE LOS PERFILES TAXONÓMICOS BACTERIANOS DE LOS METAGENOMAS DE TUMOR DE ROSA Y DE TEJIDO SANO.

En la Figura 22 se aprecia la comparación de los perfiles taxonómicos de los metagenomas de tumor de rosa (M402) y de tejido sano (M387) a nivel de especie, mediante el empleo de la herramienta Kaiju - v1.7.3 y la base de datos RefSeq Genomes (no Euks). Las especies predominantes en la muestra de tejido sano fueron: *Mycobacterium marseillense*, *Pantoea rwandensis*, *Pantoea vagans* y *Mycobacterium intracellulare*. Ninguno de estos microorganismos, al igual que aquellos con una presencia reducida, han sido previamente asociados con la inducción de tumores.

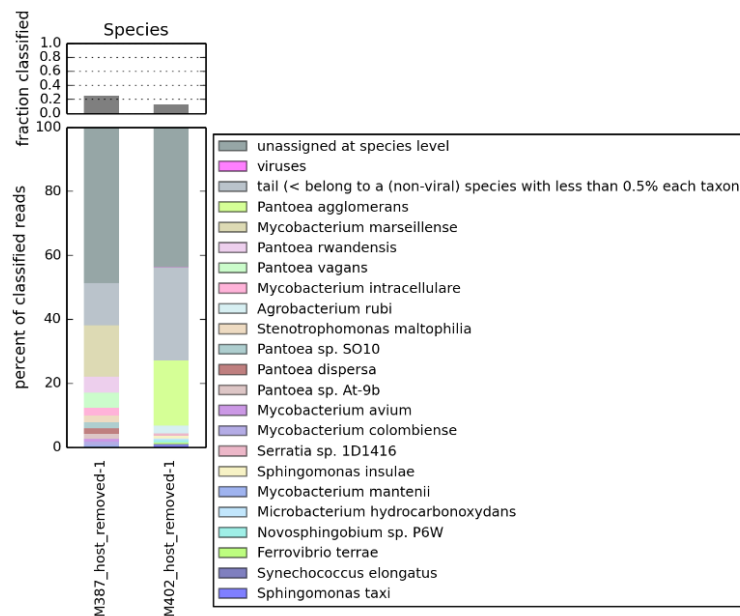


Figura 22. Comparación del perfil taxonómico entre los metagenomas de tumor de rosa y de tejido sano.

Las especies de *M. marseillense* y *M. intracellulare*, miembros del complejo *Mycobacterium avium complex* (MAC), son micobacterias no tuberculosas que pueden causar enfermedades tanto en animales como en seres humanos (Hruska & Kaevska, 2012). La presencia de *M. marseillense* y *M. intracellulare* en el tejido sano puede

explicarse ya que las bacterias pertenecientes al complejo MAC están ampliamente distribuidas en el ambiente, y se detectan con regularidad en fuentes como agua, polvo doméstico y suelo. Además, debido a que sus membranas externas son ricas en lípidos presentan una alta resistencia a la desinfección química, incluyendo agentes como el cloro, las cloraminas y el ozono (Daley, 2017).

En lo que respecta a *P. rwandensis* y *P. vagans* no se han documentado informes que las cataloguen como patógenos de plantas, más bien se ha reportado una relación epífita con plantas de eucalipto (Brady et al., 2009; Brady et al., 2012).

En el caso de la muestra de tumor de rosa, las especies que destacaron fueron: *Pantoea agglomerans*, *Agrobacterium rubi*, *Serratia* sp. 1D1416, *Sphingomonas insulae*, *Novosphingobium* sp. P6W. En comparación con los resultados previos, se observaron discrepancias, principalmente la presencia de *A. rubi* en lugar de *A. larrymoorei* y la exclusión de *Oxynema* sp. AP17. Sin embargo, es relevante destacar que los análisis iniciales y los de la presente sección se realizaron en fechas distintas, por lo que actualizaciones o modificaciones en la base de datos empleada pudo afectar la clasificación.

Al comparar los perfiles taxonómicos de las muestras, se observó que el principal factor distintivo fue la presencia de *Agrobacterium rubi* en la muestra de tumor de rosa, dado que se trata de un patógeno reconocido por inducir agallas en plantas. Esto prevaleció como rasgo distintivo, a pesar de las notables diferencias en la microbiota que presentó cada muestra. No obstante, una característica en común que comparten la muestra de tumor de rosa con la de tejido sano es la presencia de especies del género *Pantoea*. Walterson y Stavrinides (2015) señalan que diversas investigaciones han registrado la presencia de poblaciones epífitas de *Pantoea* en una amplia gama de plantas, que abarcan desde vegetales hasta frutas y cultivos de granos; este hallazgo apunta a la existencia de adaptaciones específicas en estas bacterias que favorecen su capacidad de colonizar la filósfera. Por consiguiente, es plausible inferir que la presencia de especies del género *Pantoea* en ambas muestras es atribuible al potencial versátil de estas bacterias para adaptarse a diferentes entornos vegetales.

Si bien no se encontraron reportes relacionados con el análisis de la microbiota en plantas de rosa afectadas por agalla de corona, el estudio realizado por Faist et al. (2016)

empleando el enfoque de secuenciación de amplicones y la agrupación de OTUs, se enfocó en analizar la variabilidad de taxones bacterianos en las uniones de injertos en plantas de vid con y sin agalla de corona. Los resultados revelaron que las uniones con agalla de corona albergaban un número mayor de especies bacterianas distintas en todas las estaciones en comparación con las uniones sin agalla de corona. En cada estación, *Agrobacterium vitis*, *Pseudomonas* sp. y *Enterobacteriaceae* sp. fueron los tres OTUs más abundantes en las uniones con galla de corona, mientras que en las uniones sin agalla de corona, los tres OTUs más abundantes e incluyeron especies de *Pseudomonas*, *Sphingomonas* sp., *Curtobacterium* sp, *Ralstonia* sp y especies de *Erwinia*. Esto sugiere que las uniones sin agalla de corona no exhiben una microbiota central, llevando a la conclusión de que la enfermedad de agalla de corona contribuye a mantener una composición bacteriana estable a lo largo de las estaciones

En relación con lo expuesto, Mafakheri et al. (2022) mencionan que el tejido de agalla de corona es un ambiente enriquecido en nutrientes y actúa como una reserva de aminoácidos, azúcares y ácidos orgánicos, que a su vez pueden utilizarse como recursos nutricionales para el crecimiento de diversos grupos bacterianos, además de las agrobacterias patógenas. Por lo tanto, el tejido de agalla se configura como un entorno propicio para la posible colonización tanto de poblaciones epífitas como endófitas de diversas especies bacterianas.

CAPÍTULO 4: CONCLUSIONES

- Los géneros *Pantoea*, *Agrobacterium*, *Pseudomonas* y *Sphingomonas* se destacaron por ser los más abundantes y presentaron concordancia en su detección mediante ambas herramientas de clasificación taxonómica.
- A nivel de especie se observaron discrepancias en la asignación taxonómica entre las herramientas de clasificación. A pesar de esto, *Pantoea agglomerans* se destacó como la especie más abundante en ambas herramientas.
- De las taxones detectados, *Pantoea agglomerans* y *Agrobacterium* han sido reportadas como bacterias formadoras de tumores de agalla de corona.
- A nivel funcional en el ensamblaje del metagenoma se identificó un total de 12351 CDSs, 7 tRNAs, 126 tRNAs y 2 tmRNAs en la muestra metagenómica.
- A partir de la optimización de bins se logró ensamblar el genoma de *Pantoea agglomerans* con una integridad del 100%, contaminación del 0.52% y un ANI del 98.49%.
- El ensamblaje de las 3420 lecturas mapeadas del genoma de referencia de *Agrobacterium tumefaciens* representó apenas el 0.861% del genoma. El número de lecturas secuenciadas para *A. tumefaciens* pudo verse afectada por la presencia del genoma del hospedero y los diversos niveles de abundancia de las especies en la muestra metagenómica, y niveles de fitohormonas en la planta al momento del muestreo.
- *Agrobacterium tumefaciens* fue considerado como el organismo más probable de ser responsable de la generación de tumores de rosa en la muestra, debido al mapeo y ensamblaje de la secuencia del plásmido de nopalina pTiKerr108 en un 69.23%. Además de que no se registraron lecturas que correspondieran a genes de virulencia para *Pantoea agglomerans*.

CAPÍTULO 5: RECOMENDACIONES

- Ensayar protocolos para la separación física de la microbiota del tejido vegetal y examinar su efecto sobre los niveles de abundancia.
- Probar métodos de enriquecimiento selectivo de ADN microbiano previo al proceso de secuenciación.
- Probar distintos softwares para la remoción de contaminantes generados a partir del hospedero y evaluar su eficacia.

BILIOGRAFÍA

- Agrios, G. (2005). Plant diseases caused by Prokaryotes: Bacteria and Mollicutes. *Plant Pathology*, 615-703. doi:10.1016/b978-0-08-047378-9.50018-x
- Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., . . . Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, 11(11), 1144-1146. doi:10.1038/nmeth.3103
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Arkin, A., Cottingham, R., Henry, C., Harris, N., Stevens, R., Maslov, S., . . . Yu, D. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology*, 36(7), 566-569. doi:10.1038/nbt.4163
- Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., . . . Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9. doi:10.1186/1471-2164-9-75
- Bouzar, H., & Jones, J. B. (2001). *Agrobacterium larrymoorei* sp. nov., a pathogen isolated from aerial tumours of *Ficus benjamina*. *International journal of systematic and evolutionary microbiology*, 51(Pt 3), 1023–1026. doi:10.1099/00207713-51-3-1023
- Brady, C., Cleenwerck, I., van der Westhuizen, L., Venter, S., Coutinho, T., & De Vos, P. (2012). *antoea rodasii* sp. nov., *Pantoea rwandensis* sp. nov. and *Pantoea wallisii* sp. nov., isolated from Eucalyptus. *International journal of systematic and evolutionary microbiology*, 62(7), 1457–1464. <https://doi.org/10.1099/ijs.0.032615-0>
- Brady, C., Venter, S., Cleenwerck, I., Engelbeen, K., Vancanneyt, M., Swings, J., & Coutinho, T. (2009). *Pantoea vagans* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyi* sp. nov. and *Pantoea anthophila* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 59(9), 2339–2345. <https://doi.org/10.1099/ijs.0.009241-0>

- Bragg, L., & Tyson, G. (2014). Metagenomics Using Next-Generation Sequencing. *Environmental Microbiology*, 183–201. doi:10.1007/978-1-62703-712-9_1
- Chandrasekaran, M., Lee, J., Ye, B.-M., Jung, S., Kim, J., Kim, J.-W., & Chun, S. (2019). Isolation and Characterization of Avirulent and Virulent Strains of *Agrobacterium tumefaciens* from Rose Crown Gall in Selected Regions of South Korea. *Plants*, 8(11). doi:10.3390/plants8110452
- Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, 36(6), 1925-1927. doi:10.1093/bioinformatics/btz848
- Chen, L., Wang, X., Ma, Q., Bian, L., Liu, X., Xu, X., . . . Liu, Y. (2020). *Bacillus velezensis* CLA178-Induced Systemic Resistance of *Rosa multiflora* Against Crown Gall Disease. *Frontiers in Microbiology*, 11. doi:10.3389/fmicb.2020.587667
- Chi, B., Echeverría, S., Kay, A., Acereto, P., & Rodríguez, L. (2004). La bacteria *Agrobacterium tumefaciens* como herramienta biotecnológica. *Ciencia - Academia Mexicana de Ciencias*, 61-68. Obtenido de https://www.revistaciencia.amc.edu.mx/images/revista/55_3/La_Bacteria_Agrobacterium.pdf
- Chivian, D., Jungbluth, S., Dehak, P., Wood-Charlson, E., Canon, R., Allen, B., . . . Arkin, A. (2023). Metagenome-assembled genome extraction and analysis from microbiomes using KBase. *Nature Protocols*, 18(1), 208-238. Obtenido de <https://doi.org/10.1038/s41596-022-00747-x>
- Christie, P. (2009). *Agrobacterium* and Plant Cell Transformation. *Encyclopedia of Microbiology*. doi:10.1016/b978-012373944-5.00115-2
- Daley, C. L. (2017). *Mycobacterium avium* Complex Disease. *ASM Journals - Microbiology Spectrum*, 5(2). Obtenido de <https://doi.org/10.1128/microbiolspec.tnmi7-0045-2017>
- Dominguez Del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Vinnere Pettersson, O., . . . Lantz, H. (2018). Ten steps to get started in Genome Assembly and Annotation. *F1000Research*, 7. doi:10.12688/f1000research.13598.1

- Duman, K., da Silva, S., Iriarte, F., Riddle, B., Knox, G., Orwat, M., . . . Paret, M. (2018). Bacterial Crown Gall of Roses Caused by *Agrobacterium tumefaciens*. *University of Florida - IFAS Extension*(PP343).
- Escobar, M., & Dandekar, A. (2003). *Agrobacterium tumefaciens* as an agent of disease. *Trends in Plant Science*, 8(8). doi:10.1016/s1360-1385(03)00162-6
- ExpoFlores. (Abril de 2022). *Reporte Estadístico Mensual - Abril 2022*. <https://expoflores.com/wp-content/uploads/2022/04/Expoflores-abril-2022.pdf>
- Finer, K., Fox, L., & Finer, J. (2016). Isolation and Characterization of *Agrobacterium* Strains from Soil: A Laboratory Capstone Experience. *Journal of Microbiology & Biology Education*, 444-450. doi:10.1128/jmbe.v17i3.1124
- Freitas, T. A., Li, P. E., Scholz, M. B., & Chain, P. S. (2015). Accurate read-based metagenome characterization using a hierarchical suite of unique signatures. *Nucleic acids research*, 43(10). doi:10.1093/nar/gkv180
- Gan, H. M., Buckley, L., Szegedi, E., Hudson, A., & Savka, M. A. (2009). Identification of an *rsh* gene from a *Novosphingobium* sp. necessary for quorum-sensing signal accumulation. *Journal of bacteriology*, 191(8), 2551–2560. <https://doi.org/10.1128/JB.01692-08>
- Gogoleva, N., Nikolaichik, Y., Ismailov, T., Gorshkov, V., Safronova, V., Belimov, A., & Gogolev, Y. (2019). Complete genome sequence of the abscisic acid-utilizing strain *Novosphingobium* sp. P6W. *3 Biotech*, 9(3). doi:10.1007/s13205-019-1625-8
- Gordon, J., & Christie, P. (2014). The *Agrobacterium* Ti Plasmids. *Microbiology Spectrum*, 2(6). doi:10.1128/microbiolspec.PLAS-0010-2013
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, 29(8), 1072–1075. doi:10.1093/bioinformatics/btt086
- Gutiérrez-Barranquero, J., Carzola, F., Torés, J. A., & de Vicente, A. (2019). *Pantoea* agglomerans as a New Etiological Agent of a Bacterial Necrotic Disease of Mango Trees. *APS Publications*, 109(1), 17-26. doi:10.1094/PHYTO-06-18-0186-R

- Harbola, A., Negi, D., Manchanda, M., & Kesharwani, R. (2022). Bioinformatics and biological data mining. *Bioinformatics*. doi:10.1016/B978-0-323-89775-4.00019-5.
- Hruska, K., & Kaevska, M. (2012). Mycobacteria in water, soil, plants and air: a review. *Veterinarni Medicina*, 623–679. <https://www.agriculturejournals.cz/pdfs/vet/2012/12/02.pdf>
- Hwang, H., Yu, M., & Lai, E. M. (2017). Agrobacterium-mediated plant transformation: Biology and applications. *The Arabidopsis Book*, 15. doi:10.1199/tab.0186
- Kang, D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7. doi:10.7717/peerj.7359
- Krenek, P., Samajova, O., Luptovciak, I., Doskocilova, A., Komis, G., & Samaj, J. (2015). Transient plant transformation mediated by *Agrobacterium tumefaciens*: Principles, methods and applications. *Biotechnology Advances*, 1024 - 1042. doi:10.1016/j.biotechadv.2015.03.012
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359. doi:10.1038/nmeth.1923
- Lapidus, A., & Korobeynikov, A. (2020). Metagenomic Data Assembly – The Way of Decoding Unknown Microorganisms. *Frontiers in Microbiology*, 12. doi:10.3389/fmicb.2021.613791
- Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31(10), 1674-1676. doi:10.1093/bioinformatics/btv033
- Mafakheri, H., Taghavi, S., Zarei, S., Rahimi, T., Hasannezhad, M., Portier, P., . . . Osdaghi, E. (2022). Phenotypic and Molecular-Phylogenetic Analyses Reveal Distinct Features of Crown Gall-Associated *Xanthomonas* Strains. *Microbiology spectrum*, 10(1). <https://doi.org/10.1128/spectrum.00577-21>

- Mande, S., Mohammed, M., & Ghosh, T. (2012). Classification of metagenomic sequences: methods and challenges. *Briefings in bioinformatics*, *13*(6), 669–681. doi:10.1093/bib/bbs054
- Menzel, P., Kim, L. N., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature communications*, *7*(11257). doi:10.1038/ncomms11257
- Méric, G., Wick, R., Watts, S., Holt, K., & Inouye, M. (2019). Correcting index databases improves metagenomic studies. <https://www.biorxiv.org/content/10.1101/712166v1>
- Nasko, D., Koren, S., Phillippy, A., & Treangen, T. (2019). RefSeq database growth influences the accuracy of k-mer-based lowest common ancestor species identification. *Genome biology*, *19*(1). doi:10.1186/s13059-018-1554-6
- Nissan, G., Chalupowicz, L., Sessa, G., & Barash, I. (2019). Two *Pantoea* agglomerans type III effectors can transform nonpathogenic and phytopathogenic bacteria into host-specific gall-forming pathogens. *Molecular Plant Pathology*, *20*(11), 1582-1587. doi:10.1111/mpp.12860
- Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Research*, *27*(5), 824-834. doi:10.1101/gr.213959.116
- Otten, L., Burr, T., & Szegedi, E. (2008). *Agrobacterium*: A disease-causing bacterium. In T. Tzfira, & V. Citovsky, *Agrobacterium: From Biology to Biotechnology*. New York: Springer. doi:10.1007/978-0-387-72290-0_1
- Păcurar, D., Thordal-Christensen, H., Păcurar, M., Pamfil, D., Botez, C., & Bellini, C. (2011). *Agrobacterium tumefaciens*: From crown gall tumors to genetic transformation. *Physiological and Molecular Plant Pathology*, *76*-81. doi:10.1016/j.pmpp.2011.06.004
- Parks, D. H., Imelfort, M., Skennerton, C., Hugenholtz, P., & Tyson, G. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043-1055. doi:10.1101/gr.186072.114

- Peng, Y., Leung, H. C., Yiu, S. M., & Chin, F. Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28(11), 1420-1428. doi:10.1093/bioinformatics/bts174
- Rincon-Molina, C., Rincon-Molina, F., Zenteno-Rojas, A., Ruiz-Valdiviezo, V., Culebro-Ricaldi, J., & Rincon-Rosales, R. (2021). Bacterial diversity with plant growth-promoting potential isolated from *Agave americana* L., rhizosphere. *Agro Productividad*, 3. doi:10.32854/agrop.v14i8.1928
- Sathyanarayanan, A., Manda, S., Poojary, M., & Nagaraj, S. (2019). Exome Sequencing Data Analysis. *Encyclopedia of Bioinformatics and Computational Biology*, 3, 164-175. doi:10.1016/B978-0-12-809633-8.20094-0
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068-2069. doi:10.1093/bioinformatics/btu153
- Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., . . . Wrighton, K. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*, 48(16), 8883-8900. doi:10.1093/nar/gkaa621
- Shao, S., & Hooykaas, P. (2019). Complete Sequence of Succinamopine Ti-Plasmid pTiEU6 Reveals Its Evolutionary Relatedness with Nopaline-Type Ti-Plasmids. *Genome Biology and Evolution*, 11(9), 2480-2491. doi:10.1093/gbe/evz173
- Sharpton, T. (2014). An introduction to the analysis of shotgun metagenomic data. *Frontiers in Plant Science*, 5(209). doi:10.3389/fpls.2014.00209
- Sieber, C. M., Probst, A. J., Sharrar, A., Thomas, B., Hess, M., Tringe, S. G., & Banfield, J. F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, 3(7), 836-843. doi:10.1038/s41564-018-0171-1
- Smith, R., Glendinning, L., Walker, A., & Watson, M. (2022). Investigating the impact of database choice on the accuracy of metagenomic read classification for the rumen microbiome. *Animal Microbiome*, 4(57). doi:10.1186/s42523-022-00207-7

- Subramoni, S., Nathoo, N., Klimov, N., & Yuan, Z. (2014). *Agrobacterium tumefaciens* responses to plant-derived signaling molecules. *Frontiers in Plant Science*, 5(87783). doi:10.3389/fpls.2014.00322
- Torres, M., Jiquel, A., Jeanne, E., Naquin, D., Dessaux, Y., & Faure, D. (2021). *Agrobacterium tumefaciens* fitness genes involved in the colonization of plant tumors and roots. *New Phytologist*, 233(2), 905-918. doi:10.1111/nph.17810
- Walterson, A., & Stavriniades, J. (2015). *Pantoea*: insights into a highly versatile and diverse genus within the Enterobacteriaceae. *FEMS Microbiology Reviews*, 39(6), 968–984,. <https://doi.org/10.1093/femsre/fuv027>
- Weinthal, D. M., Barash, I., Panijel, M., Valinsky, L., Gaba, V., & Manulis-Sasson, S. (2007). Distribution and Replication of the Pathogenicity Plasmid pPATH in Diverse Populations of the Gall-Forming Bacterium *Pantoea agglomerans*. *Applied and Environmental Microbiology*, 73(23), 7552–7561. doi:10.1128/AEM.01511-07
- Wu, Y. W., Simmons, B., & Singer, S. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, 32(4), 605-607. doi:10.1093/bioinformatics/btv638
- Ye, S., Siddle, K., Park, D., & Sabet, P. (2019). Benchmarking Metagenomics Tools for Taxonomic Classification. *Cell*, 178(4), 779-794. doi:10.1016/j.cell.2019.07.010
- Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L., . . . Xie, Y. (2021). Advances in Metagenomics and Its Application in Environmental Microorganisms. *Frontiers in Microbiology*, 21. doi:10.3389/fmicb.2021.766364