

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR



FACULTAD DE INGENIERÍA

MAESTRÍA EN BIOLOGÍA COMPUTACIONAL

TRABAJO DE TITULACIÓN

TEMA:

“Aplicación de métodos bioinformáticos para identificar endófitos de la familia Burkholderiaceae en hojas de especies de la familia Rubiaceae a partir de datos NGS”

Gabriela Inés Méndez Silva

Tutor: Romain A. Guyot

Quito, 2022-2023

Aprobación del director del Trabajo de Titulación



**FACULTAD DE INGENIERÍA
MAESTRÍA EN biología computacional
MODALIDAD VIRTUAL**

Quito, 23 de Marzo de 2023

Sr.
Mgt. Charles Escobar
DECANO
FACULTAD DE INGENIERÍA
PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

De mis consideraciones:

Se envía el Informe correspondiente a la tutoría realizada al Trabajo de Titulación que se detalla a continuación:

TITULO DEL TRABAJO DE INTEGRACIÓN CURRICULAR	Aplicación de métodos bioinformáticos para identificar endófitos de la familia Burkholderiaceae en hojas de especies de la familia Rubiaceae a partir de datos NGS	
DIRECTOR	Nombre	Cédula
	Romain Guyot	20FV06833
ESTUDIANTE(S)	Nombre	Cédula
	Gabriela Ines Mendez Silva	1722305057

Se informa que el trabajo ha cumplido con todos los parámetros establecidos, mediante el cual el Maestrante demuestra el desarrollo de competencias en el campo de conocimiento de su profesión y presenta una propuesta en el área de conocimiento, con un nivel de argumentación coherente.

Dando por concluida esta tutoría de trabajo de titulación, CERTIFICO, para los fines pertinentes, que el Maestrante está apto para continuar con el proceso de LECTURA-EVALUACIÓN.

Atentamente,

DIRECTOR/TUTOR DE TRABAJO DE TITULACIÓN

Romain Guyot

Agradecimientos

Quiero expresar mi sincero agradecimiento al Dr. Romain A. Guyot quien, a lo largo de este tiempo, me ha brindado soporte científico y humano para la ejecución de este trabajo de investigación. A los docentes de la maestría a quien les debo los conocimientos adquiridos.

Agradezco a mis compañeros de maestría por su apoyo y esos buenos momentos que compartimos.

Finalmente, agradezco especialmente a mi familia por su apoyo y comprensión constante durante todo este proceso. De manera especial a mi esposo Ricardo, por su apoyo incondicional, quien ha estado a mi lado impulsándome para que me supere día a día.

Índice General

Contenido

1	Introducción	1
2	Revisión de la literatura	4
2.1	Familia <i>Rubiaceae</i>	4
2.1.1	Género <i>Coffea</i>	4
2.1.1.1	<i>Coffea arabica</i>	6
2.1.1.2	<i>Coffea canephora</i>	7
2.1.1.3	<i>Coffea eugenioides</i>	8
2.1.1.4	<i>Coffea pseudozanguebariae</i>	10
2.1.1.5	<i>Coffea racemosa</i>	11
2.1.1.6	<i>Coffea stenophylla</i>	12
2.1.1.7	<i>Coffea humilis</i>	14
2.1.1.8	<i>Coffea dolichophylla</i> o <i>Coffea perrieri</i>	15
2.1.1.9	<i>Coffea humblotiana</i>	16
2.1.1.10	<i>Coffea ebracteolata</i>	17
2.1.1.11	<i>Coffea ambongensis</i>	19
2.1.1.12	<i>Coffea boinensis</i>	20
2.1.1.13	<i>Coffea bissetiae</i>	21
2.1.2	Género <i>Pavetta</i>	22
2.1.2.1	<i>Pavetta schumanniana</i>	22
2.1.3	Genero <i>Vangueria</i>	24
2.1.3.1	<i>Vangueria infausta</i>	24
2.2	Microorganismos endófitos en plantas de la familia <i>Rubiaceae</i>	25
2.2.1	Microbioma en el género <i>Coffea</i>	26
2.2.1.1	Microbioma de la rizósfera del café.....	27
2.2.1.2	Microbioma de los granos de café.....	31
2.2.1.3	Microbioma en las hojas del café.....	31
2.2.2	Microbioma en el Género <i>Pavetta</i> , <i>Psychotria</i> y <i>Sericanthe</i>	32
2.1	Familia <i>Burkholderiaceae</i> como endófito de plantas de la familia <i>Rubiaceae</i>	33
2.2	Tecnologías de secuenciación de nueva generación (NGS)	35
2.2.1	Ilumina.....	36
2.3	Herramientas bioinformáticas para análisis de datos metagenómicos	37
2.3.1	Control de calidad.....	38

2.3.2	Mapeo con Bowtie2.....	38
2.3.3	Ensamblaje con AByss	38
2.3.4	NGPhylogeny.fr.....	38
3	Metodología.....	40
3.1	Selección de datos de secuenciación masiva para análisis bioinformático	40
3.2	Desarrollo de pipeline con herramientas bioinformáticas, para la identificación de microorganismos endófitos del género <i>Burkholderia</i>	43
3.2.1	Evaluación de la calidad de las secuencias de Illumina	45
3.2.2	Limpieza de las secuencias con Trimmomatic	45
3.2.3	Mapeo de secuencias	46
3.2.4	Ensamblaje con AByss.....	47
3.2.5	Taxonomía y análisis filogenético	48
4	Resultados y discusión	49
4.1	Evaluación de la calidad de las secuencias	49
4.1.1	Calidad de la secuencia por base	50
4.1.2	Puntajes de calidad por secuencia	51
4.1.3	Contenido de bases por secuencia.	52
4.1.4	Contenido de GC por secuencia.	52
4.1.5	Contenido de bases N	53
4.1.6	Distribución de la longitud de las secuencias.....	53
4.1.7	Nivel de duplicación de las secuencias.....	54
4.1.8	Contenido de adaptadores.....	55
4.1.9	Secuencias sobreexpresadas y k-mers sobreexpresados.....	55
4.2	Mapeo de secuencias	55
4.3	Visualización cobertura en IGV	57
4.4	Ensamblaje	60
4.4.1	Gen <i>16s</i> RNA	60
4.4.2	Gen <i>gyrB</i>	61
4.4.3	Gen <i>recA</i>	62
4.5	Taxonomía y análisis filogenético	62
4.5.1	Bacterias de la familia Burkholderiaceae en especies del género <i>Tricalysia</i> 66	
4.5.2	Bacterias de la familia Burkholderiaceae en especies del género <i>Vangeria</i> 66	
4.5.3	Bacterias de la familia Burkholderiaceae en especies del género <i>Paveta</i> 66	
4.5.4	Bacterias de la familia Burkholderiaceae en especies del género <i>Coffea</i> 67	
5	Conclusiones y Recomendaciones	69
6	Referencias	70

7 Anexos	80
-----------------------	-----------

Índice de Figuras

Figura 1 Planta de café arábica.....	5
Figura 2 <i>Coffea arabica</i> L. (A) Bayas (B) Flores (C) Hojas.....	6
Figura 3 <i>Coffea canephora</i> Pierre (A) Hojas (B) Flores (C) Bayas	8
Figura 4 <i>Coffea eugenioides</i> . (A) Bayas (B) Flores (C) Hojas.....	9
Figura 5 <i>Coffea pseudozanguebariae</i> (A) Bayas (B) Flores (C) Hojas.....	11
Figura 6 <i>Coffea racemosa</i> Lour. (A) Bayas (B) Flores (C) Hojas	12
Figura 7 <i>Coffea stenophylla</i> G.Don. (A) Bayas (B) Flores (C) Hojas	13
Figura 8 <i>Coffea humilis</i> A.Chev. (A) Flor (B) Bayas (C) Hojas.....	14
Figura 9 <i>Coffea dolichophylla</i> J.-F.Leroy (A) Bayas (B) Flor (C) Hojas	16
Figura 10 <i>Coffea humblotiana</i> Baill. (A) Bayas (B) Flor (C) Hojas	17
Figura 11 <i>Psilanthus ebracteolata</i> Hiern (A) Bayas (B) Flor (C) Hojas	18
Figura 12 <i>Psilanthus . C. ambongensis</i> (A) Arbusto (B) Fruto (C) Granos	20
Figura 13 <i>C. boinensis</i> (A) Hoja (B) Fruto.....	21
Figura 14 <i>C. bissetiae</i> (A) Hoja (B) Fruto.....	22
Figura 15 <i>Pavetta indica</i> (A)Flores (B) Hojas y Frutos	23
Figura 16 <i>Vangueria infausta</i> Burch (A)Hojas (B) Flores y (C) Frutos	24
Figura 17 Perfiles taxonómicos del microbioma endófito de <i>Coffea canephora</i> en la región de las Tierras Altas Centrales, Vietnam	29
Figura 18 Diagrama de Venn con la distribución diferencial de los géneros <i>C.</i> <i>stenophylla</i> (círculo azul), <i>C. racemosa</i> (círculo verde) y <i>C. arabica</i> (círculo rojo).	30
Figura 19 Clasificación taxonómica de secuencias bacterianas asociadas a hojas en muestras de café.....	32
Figura 20. Asociaciones del género <i>Burkholderia</i>	34
Figura 21 Generación de hebras mediante PCR puente.	37
Figura 22 Flujo de trabajo de NGPhylogeny.fr	38
Figura 23 Flujo del algoritmo de MUSCLE.....	39
Figura 24 Flujo de trabajo para identificación de bacterias endófitas del género <i>Burkholderia</i>	44
Figura 25 Visualización en IGV.....	47
Figura 26 Flujo de trabajo Phylogeny.fr.....	48
Figura 27 Calidad de las secuencias por base.	51
Figura 28 Puntuación de calidad por secuencia.....	51
Figura 29 Buena calidad en el contenido de bases por secuencia.	52
Figura 30 Mala calidad en el contenido de bases por secuencia.	52
Figura 31 Contenido de GC por secuencia.	53
Figura 32 Contenido de Ns por base.	53
Figura 33 Distribución de la longitud de las secuencias.	54
Figura 34 Niveles de duplicados en secuencias.....	54
Figura 35 Niveles de duplicación por secuencia de <i>Tricalysia semidecdua</i>	55
Figura 36 Contenido de adaptadores.	55
Figura 37 Cobertura obtenida luego del mapeo para el gen 16s RNA:.....	58
Figura 38 Cobertura obtenida luego del mapeo para el gen recA.	58
Figura 39 Cobertura obtenida luego del mapeo para el gen gyrB.....	59
Figura 40 Cobertura obtenida luego del mapeo para el gen 16s RNA en <i>Tricalysia</i> <i>semidesidua</i>	59
Figura 41 Cobertura obtenida luego del mapeo para el gen recA en <i>Tricalysia</i> <i>semidesidua</i>	59

Figura 42 Cobertura obtenida luego del mapeo para el gen <i>gyrB</i> en <i>Tricalysia semidesidua</i>	60
Figura 43 Visualización ensamblaje.con BANDAGE de contigs de la especie <i>Tricalysia lasiodelphys</i>	61

Índice de Tablas

Tabla 1 Taxonomía de la especie <i>Coffea arabica</i>	6
Tabla 2 Taxonomía de la especie <i>Coffea canephora</i>	7
Tabla 3 Taxonomía de la especie <i>Coffea eugenoides</i>	8
Tabla 4 Taxonomía de la especie <i>Coffea pseudozanguebariae</i>	10
Tabla 5 Taxonomía de la especie <i>Coffea racemosa</i>	11
Tabla 6 Taxonomía de la especie <i>Coffea stenophylla</i>	12
Tabla 7 Taxonomía de la especie <i>Coffea humilis</i>	14
Tabla 8 Taxonomía de la especie <i>Coffea dolichophylla</i>	15
Tabla 9 Taxonomía de la especie <i>Coffea humblotiana</i>	16
Tabla 10 Taxonomía de la especie <i>Coffea ebracteolata</i>	18
Tabla 11 Taxonomía de la especie <i>Coffea ambongensis</i>	19
Tabla 12 Taxonomía de la especie <i>Coffea boinensis</i>	20
Tabla 13 Taxonomía de la especie <i>Coffea bissetiae</i>	21
Tabla 14 Taxonomía del género <i>Pavetta</i>	22
Tabla 15 Taxonomía de la especie <i>Vangueria infausta</i>	24
Tabla 16 Resumen del microbiota del café	26
Tabla 17 Datos utilizados en estudio.....	40
Tabla 18 Resultado evaluación de la calidad de las secuencias.	49
Tabla 19 Resultado del mapeo realizado con Bowtie 2.....	56
Tabla 20 Resultados cobertura en IGV.....	57
Tabla 21 Estadísticos obtenidos del proceso de ensamblaje con Abyss con el gen 16S RNA.....	60
Tabla 22 Estadísticos obtenidos del proceso de ensamblaje con Abyss con el gen gyrB.	61
Tabla 23 Estadísticos obtenidos del proceso de ensamblaje con Abyss con el gen recA.	62
Tabla 24 Identificación taxonómica de bacterias con BLASTn.....	63

Índice de anexos

Anexo 1 Tipo de técnicas de Illumina.....	80
Anexo 2 Genomas de referencia.....	82

Resumen

Las especies vegetales en entornos naturales y agrícolas están continuamente expuestas a una gran cantidad de microorganismos, formando diferentes tipos de interacciones plantas-bacterias, como con las bacterias endófitas de la familia Burkholderiaceae. En la familia Rubiaceae se ha reportado actualmente el mayor número registrado de especies que se caracterizan por la nodulación en las hojas, sin embargo, son pocos los estudios de las bacterias endófitas no-nodulares de la familia Burkholderiaceae. Dado la dificultad en la identificación de microorganismos endófitos, se aplican técnicas de secuenciación para su estudio; y es por ello que se vuelve importante desarrollar herramientas bioinformáticas para detectar y estudiar endófitos bacterianos a partir de datos obtenidos de análisis metagenómicos por tecnologías de secuenciación masiva. En este estudio se desarrolló un pipeline con diferentes herramientas bioinformáticas para detectar y clasificar endófitos (especies de la familia Burkholderiaceae) de muestras de hojas de diferentes géneros de la Familia Rubiaceae (*Coffea*, *Vangeria* y *Pavetta*), por su importancia económica, principalmente del género *Coffea*, del cual se cultiva comercialmente el café. Se utilizó secuencias obtenidas de tecnologías de nueva generación (NGS) tipo Illumina pareados. Para probar la eficacia del pipeline desarrollado se utilizó como controles positivos muestras del género *Tricalysia* (Rubiaceae). Como resultados se detectó la presencia de *Paraburkholderia phenoliruptrix* en el género *Vangeria*; *Caballeronia ptereochtonis* y *Candidatus Burkholderia schumanniana* en *Pavetta*; sin embargo, no se detectó la presencia de *Burkholderia* en el género *Coffea*. Estos resultados podrían ser la base para posteriormente comprender de mejor manera las interacciones entre la planta y las bacterias de la familia Burkholderiaceae.

Palabras clave: endófitos, Burkholderiaceae, *Coffea*, NGS, *Vangeria*, *Pavetta*.

Abstract

Plant species in natural and agricultural environments are continuously exposed to many microorganisms, forming different plant-bacteria interactions, such as with endophytic bacteria of the Burkholderiaceae family. The family Rubiaceae has been reported to have the most significant recorded number of species characterized by nodulation in the leaves. However, there are few studies of non-nodular endophytic bacteria of the family Burkholderiaceae. By the difficulty in identifying endophytic microorganisms, sequencing techniques are applied for their study. Therefore, it is essential to develop bioinformatics tools to detect and study bacterial endophytes from metagenomic data. In this study, a pipeline was developed with different bioinformatic tools to detect and classify endophytes (species of the Burkholderiaceae family) of sequences obtained from new generation technologies (NGS) type Illumina paired, of leaves of the family Rubiaceae (*Coffea*, *Vangeria*, and *Pavetta*), due to its economic importance, mainly the genus *Coffea*, of which coffee is commercially grown. To test the effectiveness of the pipeline developed, samples of the genus *Tricalysia* (Rubiaceae) were used as positive controls. As a result, *Paraburkholderia phenoliruptrix* was detected in the genus *Vangeria*; *Caballeronia ptereochtonis* and *Candidatus Burkholderia schumanniana* in *Pavetta*; however, the presence of Burkholderia in the genus *Coffea* was not detected. These results could be the basis for a better understanding of the interactions between the plant and bacteria of the Burkholderiaceae family.

Keywords: endophytes, Burkholderiaceae, *Coffea*, NGS, *Vangeria*, *Pavetta*

1 Introducción

Las especies vegetales en entornos naturales y agrícolas están continuamente expuestas a una gran cantidad de microorganismos, que llevan a cabo la colonización microbiana, con una intrincada red de interacciones simultáneas entre microorganismo y planta (Cadena et al., 2016).

La interacción de plantas-bacterias es considerada ubicua y un fenómeno común en angiospermas (Orozco & Santoyo, 2021). Muchos estudios han demostrado el impacto beneficioso de tales interacciones; uno de los ejemplos más conocido es la bacteria *Rhizobia*, fijadora de nitrógeno que se encuentra en la rizosfera de varios miembros de la familia Fabaceae. Por otro lado, un ejemplo de endófitos de plantas en la filosfera, es la simbiosis bacteriana de nódulos de hojas, que se encuentra en varios taxones en eudicotas (como en Rubiaceae). La colonización de estas bacterias endófitas puede tener diferentes efectos sobre la planta, como por ejemplo, si se encuentran en la raíz pueden ayudar a la fijación de nitrógeno atmosférico, promover el crecimiento y/o actuar como agentes de control biológico frente a patógenos, reduciendo la gravedad de la enfermedad; en cuanto a los nódulos de las hojas se asocian a maximizar la eficiencia fotosintética al aumentar la cantidad de cloroplastos y, por lo tanto, la cantidad de luz capturada y la tasa de fotosíntesis, sin embargo, en algunos casos aún es desconocida su función específicamente los efectos de estas interacciones (Shiomi et al., 2006).

La familia Rubiaceae tiene actualmente el mayor número registrado de especies que se caracterizan por la nodulación bacteriana de las hojas. Actualmente, se han observado nódulos foliares en Rubiaceae en los géneros *Pavetta*, *Psychotria* y *Sericanthe*. Cuando los nódulos de las hojas están presentes, generalmente es sencillo reconocer la presencia de esta interacción particular planta-bacteria. Sin embargo, debido a la naturaleza obligada de la simbiosis, fueron necesarios métodos independientes del cultivo para establecer la identidad de las endófitas noduladas como pertenecientes al género *Burkholderia* s.l. y cada especie de planta parece contener su propia especie bacteriana única (M. Pinto et al., 2018). Además, estas estructuras distintivas a veces se pueden utilizar para la caracterización taxonómica de ciertos linajes de plantas en rubiáceas.

La familia Rubiaceae también contiene especies con endófitos foliares que no se alojan en nódulos, sino se caracteriza por endófitos que ocurren en el espacio intercelular entre las células mesófilas de la hoja. Este fenotipo no nodulado se ha observado en el género *Psychotria* y *Vangueria*, entre otros. Estas endófitas no noduladas también han sido identificadas como *Burkholderia* s.l (Verstraete et al., 2013) .

La verdadera extensión de la simbiosis foliar (tanto nodulada como no nodulada) en Rubiaceae sigue siendo desconocida hasta la fecha. Como se mencionó anteriormente, actualmente se conoce que está presente en 8 géneros en 4 tribus. Sin embargo, es probable que la simbiosis de las hojas esté más extendida y también pueda ocurrir en otros géneros. Se ha encontrado que varias especies de *Burkholderia* están asociadas con las raíces de las plantas de café (es decir, *Coffea arabica* y *C. canephora* en Caballero-Mellado et al.(2004) y *C. liberica* en Duong et al. (2021). También se ha encontrado que la bacteria *Burkholderia* está asociada con las hojas de *C. arabica*, aunque solo como epífitas, no como endófitas; por lo tanto, los endófitos putativos de las hojas a menudo permanecen sin ser detectados.

Un especial interés se enfocará en la identificación de bacterias endófitas del género *Burkholderia* en diferentes muestras de especies del café cultivado, dado que está relacionado por un lado con el fin de mejorar la sostenibilidad de la producción de café como alternativas sostenibles a los insumos agroquímicos, esto dado que se conoce que varios microorganismos (incluyendo bacterias y hongos) son capaces de mostrar capacidades promotoras del crecimiento de las plantas y/o capacidades de biocontrol hacia las plagas y enfermedades del café (Duong et al., 2020a). Sin embargo también hay un gran interés en trabajar con especies de cafés silvestres, que muestran una gran resistencia a los cambios globales y a las enfermedades, dado que al conocer la existencia de bacterias asociadas a las hojas de estas plantas, permitirá posteriormente conocer los beneficios asociados a esta interacción, para conservar las especies de café silvestre existentes *in situ*, garantizar la preservación de la diversidad genética, promover su uso en la búsqueda de genes agronómicos de interés, así como permitirá caracterizar los beneficios que proporcionan estas cepas cuando se inoculan en diversos cultivos agrícolas (Orozco & Santoyo, 2021)

Para el estudio de estos microorganismos, actualmente se utilizan las tecnologías de secuenciación de nueva generación, con un enfoque metagenómico. La metagenómica se define como el análisis genético directo de genomas contenidos en una muestra ambiental (Chouvarine et al., 2016). La secuenciación masiva proporciona un perfil excepcionalmente rico de comunidades microbianas, produciendo miles de millones de lecturas cortas de ADN, que permiten clasificar a los microorganismos taxonómicamente hasta el nivel de especie (Liu et al., 2010). De las tecnologías NGS (Next Generation Sequencing), tanto el sistema 454/Roche como el Illumina/Solexa se han aplicado ampliamente a muestras metagenómicas, aunque en los últimos años se ha extendido el uso de tecnologías que producen lecturas largas como Oxford Nanopore (Ciuffreda et al., 2021).

Actualmente con el desarrollo de métodos moleculares modernos y el rápido aumento de diferentes técnicas de secuenciación de alto rendimiento, es más barato secuenciar un organismo o una muestra completa con el ADN de la planta y de todos los microorganismos asociados. Esta estrategia puede evitar el sesgo de identificación debido a las amplificaciones de la PCR, en las cuales los cebadores utilizados pueden hibridarse de forma no específica con secuencias que son similares, pero no completamente idénticas, al ADN objetivo, que puede causar errores posteriores en la identificación. También durante la amplificación la ADN polimerasa puede incorporar nucleótidos incorrectos en la secuencia (Garibyan & Avashia, 2013). Adicionalmente es importante tener en cuenta que los datos generados durante los análisis metagenómicos necesitan un tratamiento tanto estadístico como bioinformático y que los algoritmos utilizados aún deben mejorarse, e identificarse el más adecuado para el procesamiento de datos; que será en lo que se enfocará este estudio.

Las herramientas para el análisis metagenómico a partir de la secuenciación masiva no están bien desarrolladas y surgen muchas preguntas. Se plantean varios problemas bioinformáticos, como i) la estrategia de identificación de microorganismos basada en un gen (*16s*) o en varios tipos de genes, ii) la estrategia de identificación de secuencias, a partir de lecturas crudas o de ensamblajes, iii) la velocidad de análisis de los algoritmos para analizar decenas de millones de lecturas por muestra.

Es por ello, que se vuelve importante desarrollar herramientas para detectar y estudiar endófitos bacterianos de la familia Burkholderiaceae, a partir de datos obtenidos por

tecnologías de secuenciación masiva de hojas de especies de la familia Rubiaceae, a través de análisis metagenómicos, para posteriormente comprender de mejor manera las interacciones entre la planta y los microorganismos; considerando que de los microorganismos patógenos se ha estudiado con mayor detalle sus mecanismos de propagación, colonización, supervivencia y patogenicidad, sin embargo menos comprendidos son los microorganismos no patógenos que habitan en la filosfera de las plantas, que podrían llegar a ser relevantes para la salud y el desarrollo de las plantas.

El objetivo de este trabajo de titulación es desarrollar un método bioinformático para detectar y clasificar endófitos (especies del género *Burkholderia*) en hojas de diferentes géneros de la Familia Rubiaceae (*Coffea*, *Vangeria* y *Pavetta*). especialmente especies silvestres de *Coffea*, a partir de datos obtenidos de tecnologías de nueva generación (NGS) tipo Illumina pareados. Para probar la eficacia del pipeline desarrollado se utilizará como controles positivos muestras del género *Tricalysia* (Rubiaceae); y finalmente se determinará si se detectó *Burkholderias* en el género *Coffea*.

2 Revisión de la literatura

2.1 Familia *Rubiaceae*

Las rubiáceas son miembros de las gencianas y representan una de las cuatro familias más ricas en especies de las angiospermas, con 13143 especies clasificadas en 611 géneros (Naga & García, 2021), que se encuentran en regiones tropicales excepto en el continente antártico (Torres & Torres, 2020).

Están formadas por tres subfamilias Cinchonoideae, Ixoroideae y Rubioideae. Esta familia tienen una importante diversidad de productos naturales como alcaloides, antraquinonas, flavonoides, indoles, iridoides terpenoides y otros derivados fenólicos; a los cuales se les atribuye una amplia gama de usos medicinales y actividades farmacológicas como antioxidantes, antifúngicas, antidiabéticas, antibacterianas, antidiarreicas, antiinflamatorias, anticancerígenas, entre otras (Ekalu, 2021).

Las especies de Rubiaceae son consideradas económicamente importantes, especialmente el género *Coffea*, del cual se cultiva comercialmente el café. El café es uno de los productos más importantes del mundo, solo superado por el petróleo. Otras plantas de importancia económica en esta familia incluyen el género *Cinchona*, un árbol del cual se extrae de su corteza el alcaloide quinina para el tratamiento de malaria (Sultan et al., 2022).

También algunas especies de este género se han estudiado por su actividad biológica, por ejemplo, el género *Mitracarpus* en el que se ha detectado especies con metabolitos secundarios bioactivos con actividad anticancerígena como la psicorubina, el ácido ursólico, el kaempferol-3-O-rutinósido, el kaempferol, el estigmasterol y la quercetina (Ekalu, 2021). Por otro lado, el género *Pavetta* ha sido estudiado por su actividad antiinflamatoria; y el género *Vangueria* en el cual se ha identificado una alta cantidad de saponinas, podría tener diferentes aplicaciones a futuro (Valadez & Mendoza, 2021).

2.1.1 Género *Coffea*

El género *Coffea* pertenece a la familia Rubiaceae, subfamilia Ixoroideae, que abarca un grupo de especies de plantas perennes (Sultan et al., 2022). El género *Coffea* contiene 130 especies, de las cuales solo dos tienen importancia económica mundial: *C. arabica* conocida como café arábica (Figura 1) que representa el 60% del mercado de exportación de café debido a su mayor calidad de bebida; y *C. canephora*, conocida como café robusta o conilon que abarca el 40% restante, dado que tiene un perfil químico menos deseable que afecta a las características sensoriales de la bebida, a pesar de ser más resistente a plagas y enfermedades y tener una mejor adaptación a las altas temperaturas. Estas especies se encuentran distribuidas en África, Madagascar y Asia (Castillo et al., 2020; R. Pinto et al., 2021).



Figura 1 Planta de café arábica

Nota: a) planta, b) Flores c) Frutos

Fuente: (Guyot et al., 2020)

Las especies de *Coffea* son ricas en cafeína, ácidos clorogénicos, ácido cafeico e hidroxihidroquinona, que ayudan a reducir los niveles de especies reactivas de oxígeno. La cafeína es un derivado de la xantina, presenta un sabor amargo característico que se considera importante en la calidad del café. Los ácidos clorogénicos son responsables de parte del sabor amargo, la pigmentación del café, la formación del aroma y la astringencia. La trigonelina es un derivado de la piridina, influye en la puntuación final de la calidad sensorial de *C. arabica* y *C. canephora* (Castillo et al., 2020).

A pesar del abrumador éxito agronómico y económico de Arábica y Robusta, nuevas amenazas son ahora evidentes para el sector mundial del café. Estos incluyen el cambio climático, especialmente la creciente incidencia y duración de la sequía, la propagación y la creciente gravedad de patógenos fúngicos devastadores, especialmente la roya de la hoja del café (CLR) para Arábica en América Central, del Norte y del Sur; y la enfermedad del marchitamiento del café. Las variantes silvestres de Arábica y Robusta serán de importancia primordial, pero es probable que se requieran otras especies de café silvestre (A. Davis et al., 2019).

Hace catorce años, Davis y sus colaboradores revelaron que numerosas especies de café silvestre son vulnerables (23 especies), en peligro de extinción (30 especies) o gravemente amenazadas (19 especies). Una reevaluación reciente confirmó que el 60% de ellos están ahora amenazados de extinción, lo que sugiere una mala perspectiva para las especies de café silvestre en todo el mundo tropical. El análisis de las especies de café silvestre conservadas en colección viva reveló que presentan una diversidad fenotípica muy elevada, una diversidad de la composición bioquímica de las semillas, una alta adaptación al medio ambiente y una resistencia a los patógenos. Por ejemplo *C. stenophylla*, una especie silvestre del Alto Oeste de África revela un sabor superior y un perfil sensorial análogo al café Arábica de alta calidad; también *C. racemosa* Lour. recibió una medalla de oro en una feria de Lisboa en 1906 debido a características únicas de sabor y aroma, considerado el mejor café de todas las colonias portuguesas, estando presente en tierras bajas (por debajo de 200 m) y áreas de baja disponibilidad de agua en Mozambique (Koutouleas et al., 2022).

2.1.1.1 *Coffea arabica*

Tabla 1 Taxonomía de la especie *Coffea arabica*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea</i> L.
Especie	<i>Coffea arabica</i> L.

Fuente: (GBIF, 2022)

Coffea arabica es una especie nativa de Etiopía, Kenia y Sudán. Se encuentra distribuida en la Amazonia, Andes, Guayana, Serranía de La Macarena, Sierra Nevada de Santa Marta, tierras altas de Brasil y América Central (Ferreira et al., 2019).

Es una especie dicotiledónea, autógama con el 10% de polinización cruzada natural. Tienen hojas perennes, pequeñas, con una superficie oscura y brillante que alcanzan de 8 a 10 m de altura. Presenta una ramificación dimórfica, con inflorescencias del tipo glomerular, y se desarrolla a partir de yemas seriadas en cada nudo de las ramas plagiotrópicas (Figura 2) (Melese & Kolech, 2021).

Las bayas inmaduras tienen un color verde apagado; sin embargo, al madurar el color de la piel cambia de amarillo a rosa brillante. Cada baya de café contiene dos semillas con una longitud de 8,5-12,5 mm, son elipsoidales, con una superficie aplanada y profundamente acanalada; la superficie exterior es convexa (Melese & Kolech, 2021) tiempo de maduración de la fruta es de 220 días (Ferreira et al., 2019).

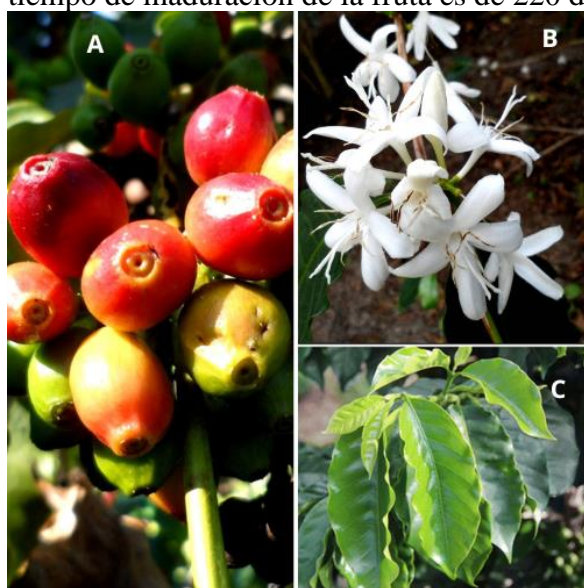


Figura 2 *Coffea arabica* L. (A) Bayas (B) Flores (C) Hojas

Fuente: (Guyot et al., 2020)

En general, las plantas de café tienen efectos positivos en las actividades agrícolas en términos de generación de oxígeno, captura de carbono y protección de la biodiversidad. Según Nab & Maslin, (2020a) estimaron que las plantas de café capturan aproximadamente el 35% del CO₂ que la misma área de bosque. Dado que los sistemas de producción de café pueden variar desde simples plantaciones de monocultivo hasta complejos agrobosques, sus reservas de carbono superficiales y subterráneas varían ampliamente, con estimaciones que van desde 14,1 Mg/ha en un monocultivo de café sin sombra a 46,3 Mg/ha en un agrobosque muy diverso.

Por otro lado, esta especie tiene varias aplicaciones, Perera et al., (2021) utilizó extracto de hoja de *Coffea arabica* como agente reductor verde del óxido de grafito a óxido de grafeno formando un nanocompuesto de *Coffea arabica* en el cual se destacó que el uso de hojas jóvenes de esta especie tiene buenas propiedades antioxidantes e inflamatorias, que permite el diseño de un fármaco eficaz con aplicaciones biomédicas para tratar diversas enfermedades inflamatorias y del envejecimiento.

Las semillas de *Coffea arabica* almacenan entre 10%-16% de aceite graso como sustancia de almacenamiento. El aceite de semillas inmaduras se puede utilizar como fuente alternativa de biodiésel, y la cubierta de las semillas se utiliza para la producción de bioetanol. La cubierta de la semilla y las tortas prensadas de los granos defectuosos son útiles como adsorbentes para la eliminación de colorantes de medios acuosos (Patay et al., 2016).

2.1.1.2 *Coffea canephora*

Tabla 2 Taxonomía de la especie *Coffea canephora*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea</i> L.
Especie	<i>Coffea canephora</i> Pierre ex A.Froehner

Fuente: (GBIF, 2022)

Coffea canephora es una especie que procede de una amplia región cálida, húmeda y de baja altitud, que se extiende desde Guinea hasta el Congo, África central y occidental (Ferrão et al., 2019). Es nativa de los bosques tropicales alrededor del lago Victoria en Uganda. Se cultiva en países tropicales como Indonesia, Vietnam y Brasil (Campuzano et al., 2021).

Coffea canephora tiene arbustos de varios tallos. Las hojas son de color verde más claro, elípticas, lanceoladas menos cerosas, más grandes y ligeramente onduladas. Presenta inflorescencias del tipo glomerular y las flores son autoincompatibles y hermafroditas con

estambres adheridos al tubo de la corola, de color blanco, en gran número por inflorescencia y por axila de hoja. El fruto es pequeño, poco esférico, de 30 a 60 por verticilo foliar (Figura 3), de superficie lisa, con colores rojo, amarillo y naranja cuando está maduro y exocarpio más fino (Ferrão et al., 2019). El tiempo de maduración del fruto es de 300 días (Ferreira et al., 2019). Las semillas son de tamaño variable, con película plateada adherente y marrón, endosperma verde y mayor contenido de cafeína. Esta especie incluye diversas variedades siendo las más representativas Kouillou (Conilon) y Robusta (Ferrão et al., 2019).



Figura 3 *Coffea canephora* Pierre (A) Hojas (B) Flores (C) Bayas

Fuente: (Guyot et al., 2020)

Coffea canephora se utiliza en la medicina tradicional para el tratamiento de diversas enfermedades, como la diabetes, las enfermedades cardiovasculares, el cáncer, hígado, Parkinson y la gota (Mewaba et al., 2022). De acuerdo con Mewaba et al., (2022) destaca la actividad antimicrobiana de un extracto de *C.robusta* frente a bacterias grampositivas y gramnegativas. Así mismo, afirmó que el extracto de acetato de etilo de las hojas de *Coffea canephora* inhiben *Edwardsiella tarda* y *Streptococcus agalactiae*.

El tallo y las hojas de *C.robusta* poseen ácido ursólico que tiene actividad antimicrobiana contra las cepas de *S.aureus* y presenta una concentración inhibitoria mínima de 8-16 µg/mL (Mewaba et al., 2022).

2.1.1.3 *Coffea eugenoides*

Tabla 3 Taxonomía de la especie *Coffea eugenoides*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida

Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea</i> L.
Especie	<i>Coffea eugenioides</i> S.Moore

Fuente: (Guyot et al., 2020)

El área de distribución nativa de esta especie va desde Sudán del Sur hasta el este de África central y tropical. Se desarrolla principalmente en ambientes tropicales, estacionalmente secos (Kewscience, 2022).

Esta especie es un arbusto cónico con numerosas ramas delgadas que generalmente crece hasta 2-3 m de altura, pero puede crecer hasta 5 m de altura. Las ramas tienen una corteza brillante de color marrón pálido con hojas de 7 a 8 cm longitud y hasta 3 cm de ancho, de forma lanceolada-acuminada, ligeramente coriáceas. Tiene inflorescencias axilares, las flores son pequeñas y blancas con 4-6 sépalos florales. La floración en condiciones naturales se produce 7 días después del inicio de las lluvias. Los frutos son de 1 cm de ancho y 1,2 cm de largo (Figura 4). El promedio de meses de desarrollo de las semillas es de 4 a 5 meses (Solano, 2016).

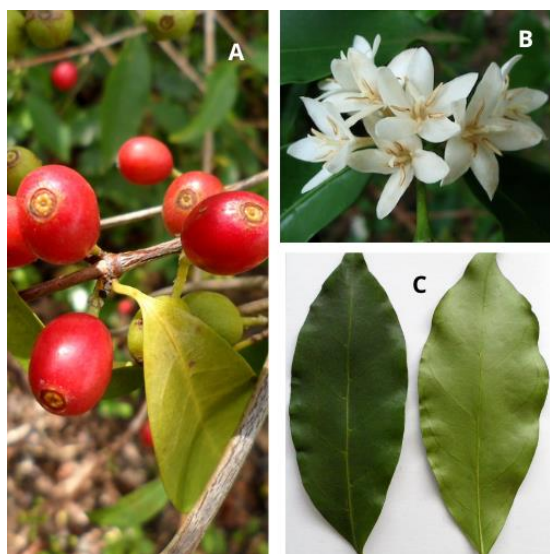


Figura 4 Coffea eugenioides. (A) Bayas (B) Flores (C) Hojas.

Fuente: (Guyot et al., 2020)

Las semillas de *Coffea eugenioides* tienen un bajo contenido de cafeína, dispone de un sistema radicular profundo que le permite tolerar la sequía. Presenta una alta resistencia a la broca del café (*Hypothenemus hamper*) y al hongo (*Mycena citricolor*) (Solano, 2016).

Adicionalmente, se ha identificado en esta especie, las xantomonas como el mangostino y la mangiferina, que tienen un gran potencial para aplicaciones farmacológicas como agentes quimiopreventivos de cáncer, antioxidantes y antiinflamatorios. Las hojas jóvenes de la especie *Coffea eugenioides* presentan un contenido elevado de mangiferina de 53.875 y 76.686 mg /g de materia seca. Estas hojas permiten producir hoja de té de café, principalmente debido a su baja relevancia económica y mayor contenido en polifenoles (mangiferina), lo cual permite el desarrollo de una bebida descafeinada de forma natural para las personas sensibles a la cafeína o que desean limitar su consumo (Monteiro et al., 2019).

2.1.1.4 *Coffea pseudozanguebariae*

Tabla 4 Taxonomía de la especie *Coffea pseudozanguebariae*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea</i> L.
Especie	<i>Coffea pseudozanguebariae</i> Bridson

Fuente: (GBIF, 2022)

Se encuentra al sur de Tanzania, norte de Zimbabue y norte de Mozambique (A. Davis et al., 2021). Es un árbol pequeño de 3 a 6 m de altura, lampiño, las ramas jóvenes están cubiertas de una corteza de color marrón oxidado o cremoso. Las hojas son elípticas o casi redondas, tienen 5,4 cm largo y 2,2 cm de ancho, de textura fina. El color de las hojas por encima es opaca o raramente brillante, tiene venas prominentes en ambas superficies, presenta de 5-6 pares principales de nervios laterales; el margen es plano o ligeramente ondulado. Presenta de seis a ocho flores. La corola tiene una coloración rosa claro que se vuelve blanca. La inflorescencia es una por axila, cada una con dos a cinco flores. Tiene de dos a cinco frutos por axila (4-10 por nudo), son lisos o ligeramente acanalado, de color verde pálido que se vuelve negro violáceo cuando está maduro, tienen un tamaño de 0,9-1,3 cm de longitud (Figura 5) (A. Davis et al., 2021).



Figura 5 *Coffea pseudozanguebariae* (A) Bayas (B) Flores (C) Hojas

Fuente: (Guyot et al., 2020)

Coffea pseudozanguebariae tiene tolerancia extrema al calor, bajas precipitaciones, estacionalidad de precipitaciones elevadas (tolerancia a la estación seca) y desarrollo rápido del fruto (alrededor de cuatro meses de floración a fruto maduro). Posee un rendimiento estimado entre 300-400 g de café limpio por planta (A. Davis et al., 2021).

El ácido 5-feruloilquínico es un ácido clorogénico que se encuentra abundantemente en la especie *C.pseudozanguebariae*. El contenido en base de materia seca es del 2,08 - 12,5% en las hojas de *C.pseudozanguebariae* con potencial actividad antioxidante que se sintetiza en las células vegetales para formar un sistema de defensa y mejorar el estrés oxidativo celular que es causado por factores abióticos o bióticos del medio ambiente (Chen, 2019).

2.1.1.5 *Coffea racemosa*

Tabla 5 Taxonomía de la especie *Coffea racemosa*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea</i> L.
Especie	<i>Coffea racemosa</i> Lour

Fuente: (GBIF, 2022)

Coffea racemosa se encuentra en el centro y sur de Mozambique, este de Zimbabue y norte de Sudáfrica (Kwa-Zulu Natal) (A. Davis et al., 2021). Es un arbusto de 1-3,5 m de altura, con tallos jóvenes puberulentos (ligeramente cubierto de pelos finos). Las hojas son caducifolias o semicaducifolias, estrechamente elípticas, con pecíolo de 1-2,5 mm de longitud y onduladas. Presenta de cuatro a seis venas secundarias. Tiene de cinco a nueve flores, la corola tiene una coloración rosa claro que se vuelve blanco. La inflorescencia es de uno a dos por axila, cada una con una a tres flores. El fruto es moderadamente acanalado en la fase verde y roja, liso y de color negro violáceo cuando está maduro (Figura 6). Tiene de uno a tres frutos por axila (2-6 por nudo) (A. Davis et al., 2021).



Figura 6 *Coffea racemosa* Lour. (A) Bayas (B) Flores (C) Hojas

Fuente: (Guyot et al., 2020)

La especie *Coffea racemosa* tiene tolerancia a temperaturas altas y a las bajas precipitaciones. Presenta resistencia a diversas plagas del café particularmente al minador de la hoja de café (*Perileucoptera coffeella*) y nemátodos (*Meloidogyne*). Además, reduce los tiempos de fructificación para híbridos interespecífico y tiene bajo contenido de cafeína. De la misma forma, la especie *Coffea racemosa* ha sido exitosamente cruzada con *C.arabica* para producir una serie de cultivares derivados de híbridos, en particular *C.arabica* cv., con un alto grado de tolerancia a la sequía. Por otra parte, esta especie se ha utilizado en programas de mejoramiento para reducir el contenido de cafeína en *C.arabica* (A. P. Davis et al., 2021).

De acuerdo con Junior et al., (2022) describe el uso de líneas de *C.arabica* portadoras de genes de *C.racemosa* con resistencia a las temperaturas de congelación enfatizando que la línea H0107-32 presentó daños leves en las hojas por lo tanto no hubo cambio en la coloración de esta ni deterioro celular.

2.1.1.6 *Coffea stenophylla*

Tabla 6 Taxonomía de la especie *Coffea stenophylla*

Reino	Plantae
--------------	---------

Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea</i> L.
Especie	<i>Coffea stenophylla</i> G.Don

Fuente: (GBIF, 2022)

Se encuentra en la Costa de Marfil, Sierra Leona, Senegal, Guinea, Liberia, Ghana, Trinidad y Tobago (Guyot et al., 2020). Es un árbol de tamaño mediano, de 3-6 m de altura con alta productividad. Los brotes de las hojas son de color rosa. Las hojas son oblongas o elípticas, 3 cm de ancho y entre 9-15 cm de largo, y crecen una a una densamente agrupadas. El color de las hojas por encima es brillante, verde oscuro y lustroso y más pálido por debajo; posee de seis a diez pares de nervios, con pequeñas glándulas en las axilas, que son blancas y perforadas en la superficie superior. Las flores tienen cinco pétalos y ocho o más lóbulos de la corola. Son blancas, con forma de estrella y fragantes. Las formas de las flores difieren entre las poblaciones nativas de Costa de Marfil; las poblaciones occidentales tienen flores oblongas y redondas; las flores de las poblaciones orientales son oblongas y ovaladas. Los frutos tienen forma globosa y tienen un tamaño de aproximadamente de 1,25 cm. Tienen una cáscara negra cuando están maduros. La piel es fina y los granos se pueden exprimir fácilmente. Las semillas son hemisféricas, con un estrecho surco ventral. La piel apergaminada es fina y de color blanco grisáceo. La piel plateada tiene una línea engrosada en el dorso de la semilla donde se encuentra un ligero surco en la semilla (Figura 7) (Solano, 2016).

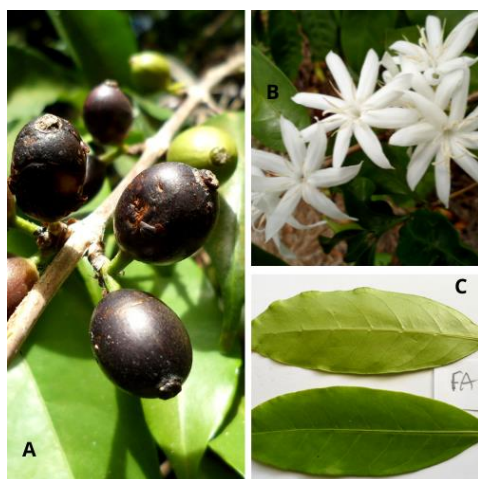


Figura 7 *Coffea stenophylla* G. Don. (A) Bayas (B) Flores (C) Hojas

Fuente: (Guyot et al., 2020)

Las semillas de *Coffea stenophylla* tienen un alto contenido de cafeína. Esta especie posee al menos dos genes recesivos que le brindan una alta resistencia al minador de la hoja (*Perileucoptera coffeella*) (Solano, 2016), también la especie *Coffea stenophylla* tiene

resistencia al tizón del halo y la mancha bacteriana de la hoja, estos resultados son útiles en programas de mejora genética para la resistencia duradera a múltiples agentes bióticos, proporcionando nuevas combinaciones de alelos de resistencia por hibridación, así como estudios fitopatológicos, para identificar la variabilidad intraespecífica de los patógeno (Rodrigues et al., 2019).

2.1.1.7 *Coffea humilis*

Tabla 7 Taxonomía de la especie *Coffea humilis*

Reino	Plantae
Filo	Tracheophyta
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea</i> L.
Especie	<i>Coffea humilis</i> A.Chev.

Fuente:(GBIF, 2022)

Coffea humilis, se encuentra en Liberia, Guinea y Costa de Marfil (Guyot et al., 2020). Es un arbusto leñoso, monocauleso de 1,83 m de altura, raramente ramificado, crece en biomas tropicales húmedos, tiene hojas grandes acuminadas sin peciolo, con criptas glabras extra axilares. Tiene estípulas cortas de 3 a 4 mm de largo. El fruto es liso, pedunculado y en la maduración presentan una coloración roja (Figura 8) (Kewscience, 2022).El período de floración y fructificación es de 8 a 9 meses (Hamon et al., 2015).

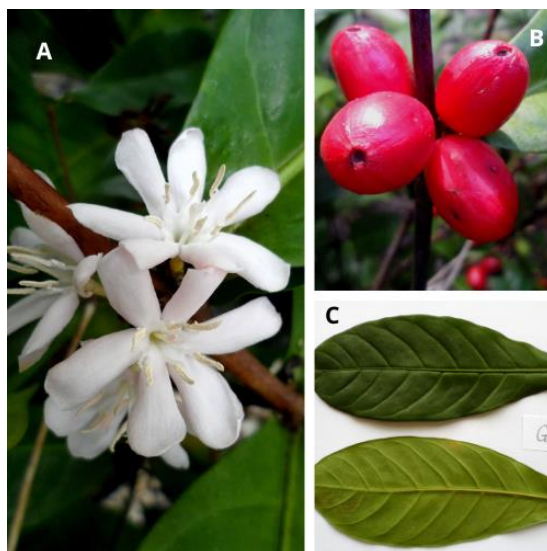


Figura 8 *Coffea humilis* A.Chev. (A) Flor (B) Bayas (C) Hojas

Fuente: (Guyot et al., 2020)

Esta especie, forma raíces adventicias en los nudos cuando están en contacto con el suelo húmedo y tienen estípulas muy grandes que con las bases de las hojas forman copas recolectoras de residuos. Estas adaptaciones permiten una absorción adicional de agua y nutrientes (Stoffelen et al., 1997).

Las raíces son una buena fuente de fitoquímicos debido a su estabilidad genética y biosintética, su potencial para la alta producción de biomasa y la síntesis de metabolitos específicos (Chawla et al., 2020). Adicionalmente, la esta especie tiene 6,16 % de ácidos clorogénicos en materia seca, los cuales están involucrados en el amargor del café debido a su descomposición en compuestos fenólicos durante el tostado (Preedy, 2015).

2.1.1.8 *Coffea dolichophylla* o *Coffea perrieri*

Tabla 8 Taxonomía de la especie *Coffea dolichophylla*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea</i> L.
Especie	<i>Coffea dolichophylla</i>

Fuente: (GBIF, 2022)

Se encuentran en el oeste, sur y el extremo sureste de Madagascar (A. Davis et al., 2020). Es un árbol pequeño y delgado, de 4-8 m de altura, aunque los ejemplares maduros y envejecidos pueden llegar hasta los 15 m. El tronco tiene un diámetro de 5-10 cm, la corteza (y los brotes) son de color gris a gris blanquecino. Las hojas son perennes, opuestas, la forma del limbo es variable (elíptica-oblonga, estrechamente elíptica, elíptica-obovada, o raramente obovada), el ápice es obtuso a agudo, base atenuada, presenta una coloración verde pálido y tiene un tamaño de 9-16 cm de largo y 3-6.5 cm de ancho, posee de ocho a diez pares de venas laterales. La inflorescencia axilar en brotes es laxa, de tres a nueve flores. Las flores son blancas, perfumadas, con una corola de cinco lóbulos. Los frutos tienen forma elipsoidal o estrechamente ovoide a ovoide, son de color rojo a marrón rojizo cuando están maduros. Las semillas presentan un tamaño de 2,11 mm de largo y 6,5-7.5 mm de ancho, son afiladas en ambos extremos (Figura 9) (A. Davis et al., 2020)

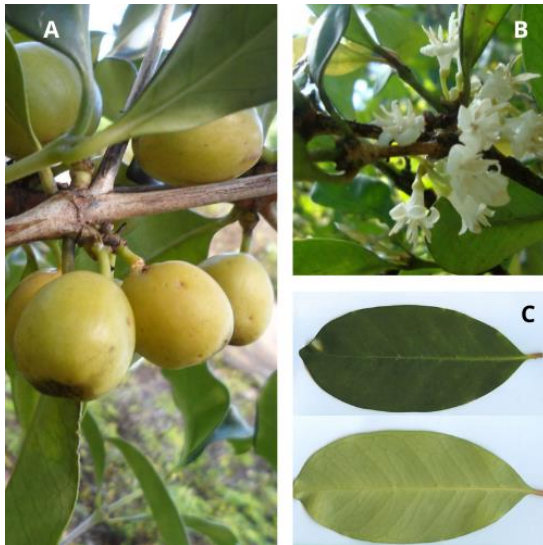


Figura 9 *Coffea dolichophylla* J.-F.Leroy (A) Bayas (B) Flor (C) Hojas

Fuente: (Guyot et al., 2020)

Esta especie se utiliza como patrón de injerto fiable y robusto, debido al fuerte crecimiento que presentan sus raíces. En los híbridos de *Coffea dolichophylla* se pueden injertar esquejes de especies más débiles para mejorar el vigor y aumentar las tasas de crecimiento y supervivencia (A. Davis et al., 2020).

Además, la pulpa del fruto de *Coffea dolichophylla* es muy dulce y deliciosa y en algunos países la pulpa seca se utiliza como ingrediente de alimentos (harina) y bebidas (zumos y té) . Y la decocción de las hojas de *Coffea dolichophylla* se utiliza para aliviar los dolores del embarazo (Bussmann et al., 2021).

2.1.1.9 *Coffea humblotiana*

Tabla 9 Taxonomía de la especie *Coffea humblotiana*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea</i> L.
Especie	<i>Coffea humblotiana</i> Baill.

Fuente: (GBIF, 2022)

Coffea humblotiana Baill, se encuentra en África tropical, Madagascar, Comoras, Mauricio, las Islas Reunión, y se extiende hasta el sur y el sureste de Asia y Australasia (Raharimalala et al., 2021).

Es un árbol que crece en los biomas tropicales estacionalmente secos. Tiene hojas glabras, lanceoladas, membranosas, acuminadas y atenuadas en la base. El pecíolo está completamente desprovisto de la punta del limbo. Los frutos tienen forma ovoide, sin pelos y son de 1 cm de largo. La flor presenta un cáliz glandular y verrugoso. La cáscara es pálida, subregular, con una hendidura en el centro y un acumen pronunciado en la parte inferior, que encierra una semilla plana en su interior de 1 cm de longitud (Figura 10) (Raoul, 1894).

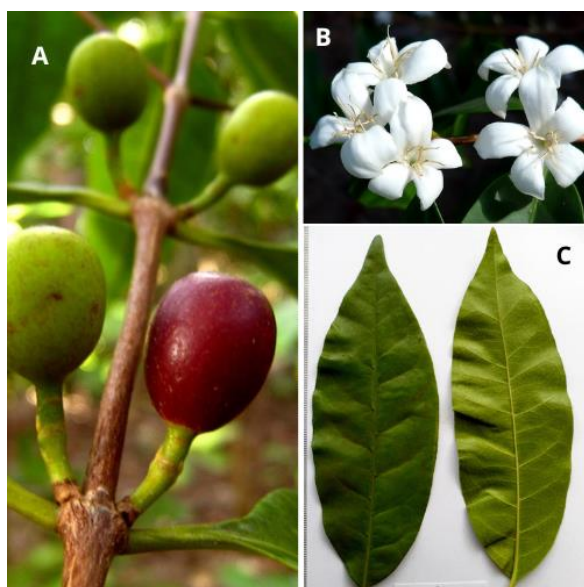


Figura 10 *Coffea humblotiana* Baill. (A) Bayas (B) Flor (C) Hojas

Fuente: (Guyot et al., 2020)

La especie *Coffea humblotiana* tiene ausencia total de cafeína en semillas y hojas debido a que carece del gen *DXMT* en su genoma provocando una reducción de la acumulación de teobromina y también presenta un bajo contenido de ácidos clorogénicos. Esta característica reduce fuertemente el amargor en la bebida de café (Raharimalala et al., 2021).

La especie *Coffea humblotiana* posee cafamarina que es un diterpeno glucosídico, el cual le otorga el amargor a la bebida de café (Raharimalala et al., 2021).

2.1.1.10 *Coffea ebracteolata*

Es una especie diploide con un número cromosómico de x igual a 11 y un tamaño total del genoma que oscila entre 460 y 900 Mbp; con un contenido de compuestos bioquímicos muy bajo o libre de cafeína. Adicionalmente este género presenta un crecimiento simpodial y monopodial (J. C. Charr et al., 2020), y es autógama (Hue, 2017).

Tabla 10 Taxonomía de la especie *Coffea ebracteolata*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Psilanthus</i> Hook.f.
Especie	<i>Coffea ebracteolata</i>

Fuente: (Anthony et al., 2011)

Son endémicas de Costa de Marfil en África Occidental, se encuentran en bosques húmedos y semisecos de baja altitud (Couturon et al., 1998).

Es un arbusto pequeño, delgado, de 1,5 – 2,5 m, presenta un crecimiento monopodial y tiene ramificaciones secundarias a lo largo de toda la longitud de las ramas primarias. Las flores son fragantes y presentan una posición terminal, el estilo es corto, las anteras y el estigma son insertados (Couturon et al., 1998). Las hojas son empapeladas y tiene una domatia extra axilar. Los frutos son ovoides o elipsoides y pedunculados con constricción conspicua a lo largo del tabique (Figura 11) (Guyot et al., 2020).

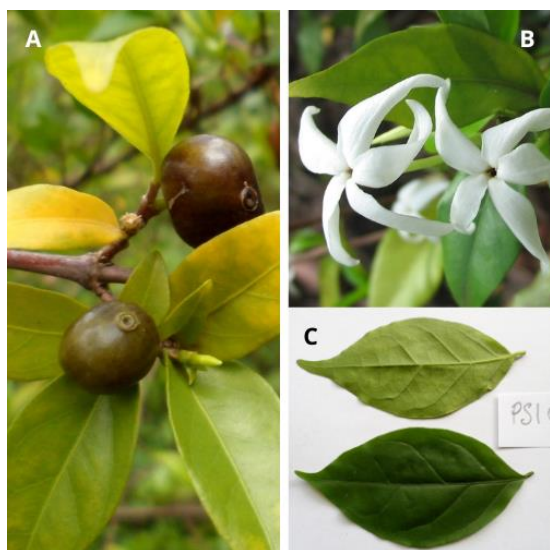


Figura 11 *Psilanthus ebracteolata* Hiern (A) Bayas (B) Flor (C) Hojas

Fuente:(Guyot et al., 2020)

Esta especie se caracteriza por presentar un bajo contenido de cafeína y ácido clorogénico por lo cual es idónea para realizar una hibridación interespecífica con otras especies cercanas como *C.arabica* con la finalidad de mejorar la calidad (Hue, 2017).

Las hojas de *Coffea ebracteolata* son empleadas contra el gusano de Guinea (Quattrocchi, 2016). Las ramas de esta especie son fibrosas y se utilizan como escobas, con la madera se fabrican cepillos de dientes, los frutos y semillas molidas son usados para realizar tatuajes negros (Guyot et al., 2020).

2.1.1.11 *Coffea ambongensis*

Tabla 11 Taxonomía de la especie *Coffea ambongensis*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea</i> L.
Especie	<i>Coffea ambongensis</i> J.F.Leroy

Fuente: (GBIF, 2022)

Se encuentran al oeste de Madagascar, en la provincia de Mahajanga, en el distrito de Mahajanga II (Ambongo) y en el distrito de Soalala (Boina) a 30 msnm (Kewscience, 2022).

Es un arbusto pequeño, con múltiples tallos, que crece hasta 5 m de altura. Los tallos tienen un diámetro de 0,5-4 cm (a la altura del pecho) y la corteza lisa es de color gris claro a blanquecino. Las hojas poseen estípulas que están unidas y forman un tubo. Las inflorescencias presentan una coloración blanca, nacen en pares a lo largo de un brote floral. Los pétalos se unen para formar un tubo de 3,1-3,7 cm de longitud, que se divide en 6 o 7 lóbulos de la corola. Los frutos tienen forma ovoide, sin pelos y son de color amarillo o amarillo rojizo cuando están maduros. Las semillas se encuentran encerradas en una estructura denominada pireno (cáscara dura) y presentan una coloración marrón claro (Kewscience, 2022). Presenta los frutos más grandes que no se parecen a la de ninguna otra especie de *Coffea*, a menudo superan los 2,5 cm de piriformes (Figura 12) (GBIF, 2022)



Figura 12 *Psilanthus . C. ambongensis* (A) Arbusto (B) Fruto (C) Granos

Fuente:(Guyot et al., 2020)

C.ambongensis es una especie endémica de los bosques secos caducifolios occidentales, sobre las arenas blancas no consolidadas, en el cual constituyen un hábitat especializado que exige adaptaciones especiales de las especies vegetales que allí se encuentran debido a los suelos pobres en nutrientes que no tienen la capacidad de retención de humedad, esta extrema falta de nutrientes hace que la regeneración sea casi imposible una vez que la vegetación ha sido despojada (de Block et al., 2020).

2.1.1.12 *Coffea boinensis*

Tabla 12 Taxonomía de la especie *Coffea boinensis*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea L.</i>
Especie	<i>Coffea boinensis</i> A.P.Davis & Rakotonas

Fuente:(Tropical Plants Database, 2022)

Es una planta nativa de Madagascar occidental, provincia de Mahajanga, distrito de Marovoay, restringida al Parque Nacional de Ankarafantsika (A. Davis & Rakotonasolo, 2008). Crece principalmente en el bioma tropical estacionalmente seco (Kewscience, 2022).

Es un árbol pequeño (de un solo tallo) de 3-6 m de altura, ramificado. Posee una corteza lisa, con una coloración gris a marrón. Las ramillas tienen un diámetro de 1-3 mm y son glabras, las hojas presentan un peciolo de 1,5-2 mm de longitud, pubescente, limbo es deltoide. Las inflorescencias son de 12,2-16,5 mm de longitud, posee botones florales y el pedicelo es de 0,6-1 mm de longitud. El fruto es ovoide, piriforme o turbinado, liso, glabro. La pared del fruto es de 3,5 -4 mm de grosor, pedicelo ausente y al madurar posee un color amarillo. Las semillas en su contorno son elípticas, obovadas, en su interior presenta una invaginación ventral ramificada y posee una coloración marrón claro cuando está seca (Figura 13). Y posee los granos de café más grandes del mundo, estos tienen el doble de tamaño de *C. arabica* (A. Davis & Rakotonasolo, 2008).



Figura 13 *C. boinensis* (A) Hoja (B) Fruto.

Fuente:(Guyot et al., 2020)

2.1.1.13 *Coffea bissetiae*

Tabla 13 Taxonomía de la especie *Coffea bissetiae*.

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Coffea</i> L.
Especie	<i>Coffea bissetiae</i> A.P.Davis & Rakotonas

Fuente: (GBIF, 2022)

Se encuentra en Madagascar occidental, provincia de Mahajanga; distrito de Marovoay, Mahajanga II y Boriziny (A. Davis & Rakotonasolo, 2008)

Es un arbusto o árbol pequeño (de uno o varios tallos), bien ramificado, de 2-3 m de altura, con un diámetro de 1-5 cm o tallos de 0,7-1,5 cm en la base y tiene una corteza lisa de color gris. Presenta inflorescencias de 5-6,4 mm de longitud, cálculos pubescentes, lóbulos estipulares y lóbulos foliares a veces glabros y márgenes de los lóbulos estipulares sin coléteres. Las flores tienen un pedicelo de 1-1,5 mm de largo y el cáliz es turbinado o abcnico. Los estamentos son filamentosos de 0,1-0,2 mm de longitud, anteras de 4-4,5 mm de longitud, estilo es de 2,2-2,4 cm de largo y los lóbulos del estigma son aplanados. El fruto es globoso, bilobado, pubescente (de escaso a moderadamente cubierto de pelos blancos apresados de 0,1-0,3 mm de longitud), y el fruto maduro es de color rojo a rojo marrón (Figura 14). Las semillas en su contorno son ampliamente elípticas, presenta una invaginación principal no ramificada endosperma entero (no rumiado), y de color marrón blanquecino cuando está seco (A. Davis & Rakotonasolo, 2008)



Figura 14 *C. bissetiae* (A) Hoja (B) Fruto

Fuente: (Guyot et al., 2020)

C.bissetiae es una especie endémica del bosque caducifolio estacionalmente seco, incluyendo hábitats litorales (en el borde de las salinas y en las dunas); arenas, suelos lateríticos, arenas lateríticas, es decir, que poseen resistencia a las variaciones abruptas de precipitación y temperatura (A. Davis & Rakotonasolo, 2008)

2.1.2 Género *Pavetta*

2.1.2.1 *Pavetta schumanniana*

Tabla 14 Taxonomía del género *Pavetta*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales

Familia	Rubiaceae
Género	<i>Pavetta</i> L.

Fuente:(Hyde et al., 2023)

Pavetta schumanniana F.Hoffm. es un arbusto o árbol pequeño de hasta 5 m de altura, con flores blancas, crema o verdosa. Se desarrolla en bosques abiertos o matorrales, sobre suelo arenoso o pedregoso, a 700–1900 m de altitud. Son cerca de 400 especies en los trópicos del viejo mundo y se encuentran desde África hasta el Sudeste Asiático, Nueva Guinea, Australia, Nueva Caledonia y Vanuatu (Figueiredo, 2008; Reynolds, 1993a).

Las hojas de este árbol son venenosas para rumiantes y causan una enfermedad llamada gousiekte. En esta forma de intoxicación, pequeñas dosis de hojas causan insuficiencia cardíaca. Adicionalmente, se caracterizan por presentar nódulos bacterianos, y las bacterias que se refugian en ellos fijan el nitrógeno del aire, liberándolo en una forma que la planta huésped puede usar (Glen, 2007). Posee inflorescencias terminales en ramificaciones principales o laterales con o sin hojas (ramificaciones florales), generalmente sésiles por encima del último par de hojas; generalmente con muchas flores. El fruto en forma de drupa, globosa, coronada por lóbulos de cáliz persistentes, en su mayoría de color negro seco y brillante (Figura 15) (Reynolds, 1993a).



Figura 15 *Pavetta indica* (A) Flores (B) Hojas y Frutos .

Fuente:(Lemoine, Correia, Lefort, et al., 2019)

Los frutos verdes de la planta se utilizan como vegetales o en encurtidos (Prasad & Bisht, 2011). Mientras que el extracto de metanol de las hojas y ramas de *P.indica* es un potencial terapéutico en el tratamiento del Triple negativo cancer de mama (TNBC), debido a que induce la apoptosis y la detención del ciclo celular en la fase sub G1, inhibe la metástasis y reduce la expresión de MRP1 (proteína 1 asociada a la resistencia a múltiples fármacos). Además, contiene 12 compuestos, entre los cuales DK (5,6 – dehidrokawaína), el componente principal, que ejerce efectos anti-invasión y de migración (Nguyen et al., 2019). De la misma forma esta especie se caracteriza por ser antiinflamatoria, diurética, hepatoprotectora, antibacteriana y antioxidante (Rajesh et al., 2021)

2.1.3 Genero Vangueria

2.1.3.1 *Vangueria infausta*

Tabla 15 Taxonomía de la especie *Vangueria infausta*

Reino	Plantae
Filo	Tracheophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Rubiaceae
Género	<i>Vangueria</i> Juss.
Especie	<i>Vangueria infausta</i> Burch

Fuente: (Maroyi, 2018)

Es una planta nativa de Angola, Botsuana, Kenia, Malawi, Mozambique, Namibia, Ruanda, Sudáfrica, Suazilandia, Tanzania, Uganda, Zambia y Zimbabue (Maroyi, 2018)

Es un arbusto caducifolio que crece de 3-8 m de altura con un tronco corto de varios tallos que sostiene ramificaciones colgantes. Las ramas están cubiertas de pelos cortos, lanosos y la corteza es de color grisáceo a marrón amarillento, liso y con tendencia a desprenderse en varias tiras irregulares. Las hojas son de tamaño medio a grande (5-10 mm de longitud), densamente peludas, opuestas y de color verde claro. Las flores son pequeñas y de color blanco verdoso a amarillento y nacen en pequeños grupos multi-ramificados, densamente agrupados a lo largo de ramas cortas y laterales. Los frutos son redondos, de color verde oscuro brillante cuando son jóvenes y cambian a un color marrón claro cuando maduran. Los frutos suelen nacer solos o en parejas en las ramitas, debajo de las hojas (Figura 16) (Maroyi, 2018)

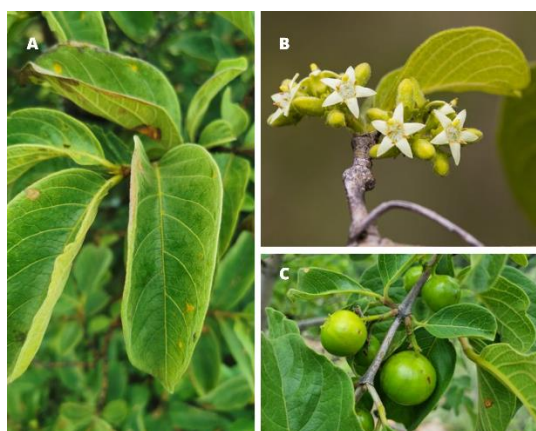


Figura 16 *Vangueria infausta* Burch (A)Hojas (B) Flores y (C) Frutos

Fuente: (Maroyi, 2018)

Son árboles frutales polivalentes tolerantes a la sequía y son adecuadas para la agrosilvicultura de secano por su capacidad de soportar una amplia gama de regímenes de temperatura y precipitaciones, ayudan a las comunidades locales con alimentos, medicinas, forraje y otras necesidades. De la misma forma son de gran importancia durante la hambruna y la pérdida de cosechas, ya que sus frutos proporcionan una variedad de nutrientes, como carbohidratos, vitaminas y proteínas, a las personas que residen en las zonas áridas y semiáridas (Karani et al., 2022).

Esta especie posee una amplia gama de actividades farmacológicas de los extractos y compuestos fitoquímicos entre los que se destaca la actividad antibacteriana, antimicobacterias, antifúngica e inhibidoras de la síntesis de prostaglandinas (Maroyi, 2018)

2.2 Microorganismos endófitos en plantas de la familia Rubiaceae

Casi todas las especies de plantas en la tierra interactúan de una manera u otra con microorganismos endófitos. Los microorganismos endófitos son microorganismos que colonizan los tejidos internos de las plantas sin causar daños visibles a sus huéspedes (Shiomi et al., 2006). Estos microorganismos se pueden transmitir verticalmente mientras que otros se caracterizan por diversos patrones de colonización. Así mismo la capacidad de colonización interna permite que los microorganismos endófitos se encuentren menos expuestos a los cambios en las condiciones del suelo y a la competencia con otros microorganismos (Duong et al., 2020a).

Aunque la gran mayoría de los endófitos son comensales, hay algunos endófitos fúngicos y bacterianos que actúan de manera mutualista. Las bacterias endófitas no siempre muestran signos externos de infección, no tienen un efecto perjudicial en su huésped. (Orozco & Santoyo, 2021). Las asociaciones planta-bacteria se pueden designar en una de las siguientes categorías: (1) una interacción negativa en la que uno de estos tiene un impacto negativo en la otra; (2) una interacción positiva, en la que ambos obtienen beneficios de la asociación; y (3) una interacción neutral, en la que ninguno obtiene un beneficio directo de la interacción y donde ninguno de los dos se ve perjudicado (Verstraete et al., 2013).

Cuando los microorganismos endófitos tienen una interacción positiva con la planta huésped, pueden promover el crecimiento, reducir la gravedad de enfermedades, inducir los mecanismos de defensa de las plantas, etc; por ello el interés de identificar la presencia de este tipo de microorganismos en plantas de interés. Los mecanismos implicados en la promoción del crecimiento de las plantas incluyen la producción de hormonas vegetales o sus análogos, la fijación de nitrógeno y el aumento de la disponibilidad de nutrientes mediante la solubilización del fósforo y la mineralización de la materia orgánica (Silva et al., 2012)

Por otro lado, los mecanismos de control de enfermedades incluyen la competencia por el espacio y los nutrientes en las plantas huésped, la producción de compuestos antimicrobianos y la inducción de la resistencia sistémica (Silva et al., 2012)

Los microorganismos endófitos pueden clasificarse en:

- **Bacterias endofíticas:** son bacterias capaces de penetrar y diseminarse sistémicamente en la planta huésped, colonizando activamente el apoplasto, los

vasos conductores y los espacios intracelulares (Shiomi et al., 2006). Los endófitos bacterianos desempeñan funciones como la producción o modulación de los reguladores de crecimiento vegetal, suministran nutrientes a través de la fijación del nitrógeno, solubilización de fósforo, control biológico contra los patógenos mediante la producción de antibióticos y una mayor tolerancia a la salinidad (Fulthorpe, Martin, & Isaac, 2020)

- **Hongos asociados:** según el sitio de colonización pueden formar: ectomicorrizas que son asociaciones de hifas de hongos que no penetran en el interior de las células de la raíz, sino que se ubican sobre y entre las separaciones de éstas, es decir, penetran en los espacios intercelulares; endomicorrizas, en las cuales el micelio fúngico penetra en las células del córtex de la raíz, es decir, colonizan intracelularmente; y micorrízicos arbusculares que son hongos simbióticos endófitos que colonizan las raíces de las plantas mientras que el micelio externo busca alimento en el suelo para transferir agua y compuestos orgánicos (fósforo, nitrógeno y otros nutrientes esenciales) a su huésped a cambio de una fuente de carbono. Al mismo tiempo son responsables en el incremento de la tolerancia al estrés biótico y abiótico (Duong et al., 2020).

2.2.1 Microbioma en el género *Coffea*

Un gran número de microorganismos asociados al café que tienen potencial de aplicación agrícola y/o industrial han sido identificados, principalmente identificados en el microbioma de la rizosfera del café, granos de café, e incluso los residuos de café (L. P. de Sousa, da Silva, & Costa, 2018). En cuanto a estudios realizados sobre el microbiota del género *Coffea*, han identificado como microbiota general 22 filos, que abarcan 129 órdenes, 607 géneros y 923 especies pertenecientes principalmente a los reinos bacteriano y fúngico. En la Tabla 16 se resume los microorganismos identificados, en relación con su diversidad taxonómica y tipo de microorganismo, en la que se incluyen los microorganismos autóctonos, que se pueden encontrar en diferentes compartimentos vegetales, como, la rizosfera, la epífita (fuera de la planta, incluye raíces, hojas, semillas, yemas) y la endófita (dentro de la planta) (Duong et al., 2020).

Tabla 16 Resumen del microbiota del café

MICROBIOTA	REINO	FILO	ORDEN	GÉNEROS	ESPECIES
AUTÓCTONA	Archaea	2	4	5	0
	Bacteria	9	43	152	174
	Hongo (HMA; levaduras)	4 (1;2)	53 (4;6)	248 (24;10)	380 (126;24)
	Número total	15	100	405	554
POST COSECHA	Bacterias	14	59	227	176
	Hongos (levaduras)	4 (2)	34 (10)	119 (51)	270 (117)
	Total	18	93	346	446

TOTAL	Archaea	2	4	5	0
	Bacteria	15	67	279	265
	Hongos (HMA; levaduras)	5 (1;2)	58 (4;11)	315 (24;53)	610 (126;133)
	Número total	22	129	607	923

Para el reino de hongos, el número de hongos micorrízicos arbusculares (HMA) y levaduras están indicados entre paréntesis.

Fuente: (Duong et al., 2020)

Por otra parte, los microorganismos endófitos de las plantas se separan en partes constituyentes aéreas y subterráneas. Las hojas, el tallo y los órganos reproductivos forman la filosfera o filobioma, mientras que las raíces y el pequeño volumen de suelo asociado representan la rizosfera o rizobioma (Lucaciu et al., 2019).

2.2.1.1 Microbioma de la rizósfera del café

La rizosfera es la principal fuente de organismos endófitos para las plantas capaces de mejorar la adquisición de nutrientes, así como la resistencia al estrés biótico y abiótico (Duong et al., 2020). Los componentes ambientales, como el manejo del suelo, tienen efectos importantes en las estructuras de los microbiomas de *Coffea* (L. P. de Sousa, Guerreiro, & Mondego, 2022), destacando que la edad de las plantas también influye en la formación del microbioma debido a que las plantas más viejas tienen microbiomas con menos diversidad (De et al., 2022).

Las comunidades bacterianas y los nutrientes minerales se correlacionan en la rizosfera, siendo los iones de calcio uno de los elementos más importantes para las plantas debido a que actúan como elementos estabilizados en las membranas, agentes fortalecedores en las paredes celulares y mensajeros secundarios obicuos. Los iones de calcio son utilizados en la transducción de señales durante la nodulación de rizobacterias. Además, aumentan la adherencia a la superficie y la formación de biopelículas de bacterias en plantas, por ejemplo, la precipitación de carbonato de calcio inducida por microbios que ocurren como un subproducto de las actividades metabólicas bacterianas, como la fotosíntesis y la desnitrificación, etc (L. P. de Sousa, da Silva, & Costa, 2018)

Otro elemento esencial para las plantas es el manganeso, ya que varias proteínas y enzimas fotosintética contienen Mn^{2+} en sus estructuras. La absorción de manganeso ocurre en las células de la raíz y se acumulan en los tejidos aéreos, especialmente en las hojas (L. P. de Sousa, da Silva, & Costa, 2018). Las bacterias como la *Pseudomonas fluorescens* y la *Pseudomonas putida* oxidan el Mn^{2+} soluble a óxido de Mn (IV) insoluble que recubre las células con precipitados de color marrón oscuro de nanopartículas de MnO_2 . La capacidad de las *Pseudomonas* para oxidar el manganeso es una ventaja competitiva sobre otras bacterias en la colonización de *Coffea arabica* Obatã que a su vez evita los efectos tóxicos del manganeso, en las hojas como la disminución de la tasa de fotosíntesis (L. P. de Sousa et al., 2018).

Por otro lado, la presencia de cafeína en el suelo también determina el microbiota presente en las raíces, se ha determinado que la cafeína aumenta hasta 1,7 veces el micoparasitismo

de *Trichoderma*. Además, el ácido clorogénico inhibe la germinación de esporas y el crecimiento micelial de *Sclerotinia*, *Fusarium*, *Botrytis*, y *Cercospora*. En efecto, la cafeína y el ácido clorogénico seleccionan microorganismos que toleran mayores concentraciones de estos metabolitos. Así mismo, estos microorganismos pueden incluso utilizar la cafeína y el ácido clorogénico como fuentes de carbono y nitrógeno (L. de Sousa et al., 2021).

Cabe resaltar que el cultivo tradicional de *Coffea arabica* tiene más de 35 veces la concentración de cafeína y 2,5 veces la concentración del ácido clorogénico que las especies silvestre *Coffea racemosa* y *Coffea stenophylla*. Sin embargo, la elevada concentración de estos metabolitos, por un lado, trata de inhibir los posibles patógenos; por otro lado, disminuye la abundancia del microbioma de la hoja, permitiendo que los microorganismos oportunistas como el hongo *Phoma* causante de la mancha foliar, se instalen más rápido (L. de Sousa et al., 2021).

Coffea racemosa y *Coffea stenophylla* producen en menor concentración la cafeína y el ácido clorogénico, albergando un microbioma más complejo que las protege contra infecciones más graves. Este fenómeno se debe a que las plantas domesticadas aprenden a tener una pérdida de diversidad en sus microorganismos asociados en comparación con las plantas silvestres. En el caso de la *C. arabica*, tanto la alta concentración de sus metabolitos como el proceso de domesticación se correlacionaron negativamente con la formación de su microbioma (L. de Sousa et al., 2021).

En cuanto a datos de diversidad microbiana rizosférica realizado por Tran, D en Vietnam, en *Coffea canephora* L, conocida como Robusta café, se ha identificado microbioma endófito representada por diferentes filos, como son las proteobacterias (70,72%) que fueron las más abundantes, seguidas de Actinobacteriota (28,12%), Acidobacteriota (0,58%) y Bacteroidota (0,58%). De los ocho órdenes bacterianos detectados, Burkholderiales (32,24%) demostró ser el más dominante, seguido de Pseudonocardiales (28,12%), Xanthomonadales (24,06%), Sphingomonadales (8,99%), Rhizobiales (4,06%), Pseudomonadales (1,16%), Holophagales (0,58%) y Chitinophagales (0,58%) Figura 17 (Tran et al., 2022). El orden Burkholderiales ha sido identificado como un microorganismo endófito, en otras especies de la familia Rubiaceae, asociándolo aplicaciones en biocontrol, dado que producen algunas enzimas como hitinasas, gelatinasas, lipasas y proteasas así como algunos otros compuestos activos como HCN y sideróforos (Duong et al., 2020).

Los suelos con grandes cantidades de actinobacterias suelen suprimir enfermedades, debido a que este grupo produce compuestos antimicrobianos. De hecho, ambientes eutróficos, como la rizosfera, inducen una mayor interacción negativa de las actinobacterias con otros microorganismos debido a que los altos niveles de nutrientes aumentan su metabolismo y la producción de metabolitos secundarios, haciéndolas colonizadoras más agresivas. Esta prevalencia explica el hecho de que las plantas de café tengan relativamente pocas enfermedades relevantes transmitidas por el suelo en comparación con otros cultivos, especialmente en su ambiente nativo (L. P. de Sousa, Guerreiro, & Mondego, 2022).

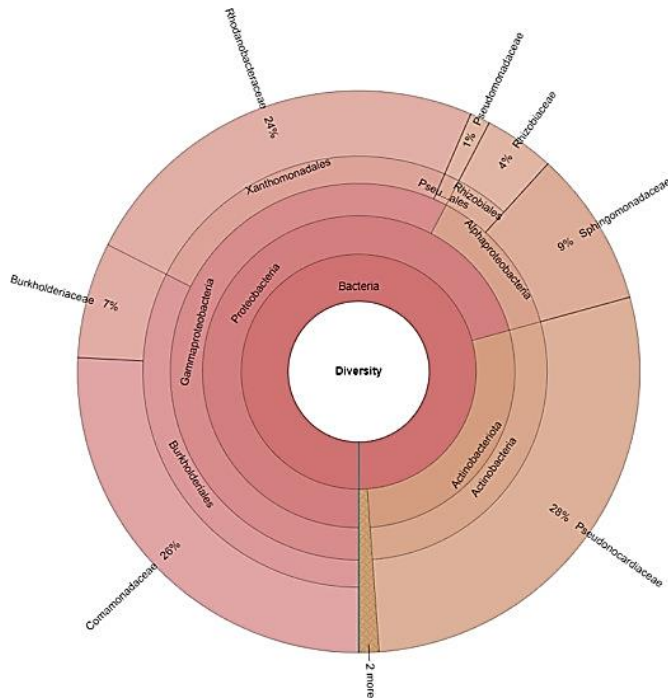


Figura 17 Perfiles taxonómicos del microbioma endófito de *Coffea canephora* en la región de las Tierras Altas Centrales, Vietnam

Fuente: (Tran et al., 2022)

Resultados similares fueron obtenidos en un estudio realizado por Fulthorpe, et al en el cual se estudió el microbioma de la raíz del café (*C.arabica* “Caturra” y *C.arabica* “Pacas”) en América Central (Costa Rica y Nicaragua), donde las principales bacterias incluían los géneros *Pantoea* (17.2%), *Enterobacter* (12.2%) y *Burkholderia* (7.6%), mientras que entre los hongos predominaban los géneros *Cladosporium* (40%), *Penicillium* (12%), *Exidiopsis* (6.6%), *Trechispora* (5.4%) y *Mycena* (3.9%) . Entre los géneros identificados, se conoce que *Burkholderia* tiene especies con una relación cercana con muchas especies de plantas, siendo considerada como un simbiote obligado en algunas hojas de miembros de la familia Rubiaceae en África (Fulthorpe, Martin, & Isaac, 2020).

De acuerdo con L. de Sousa et al., (2021) el micobioma central de los tres cafetos (*C. stenophylla*, *C. racemosa* y *C. arabica*) está formado por *Hannaella*, *Cryptococcus*, *Erythrobasidium*, *Cladosporium*, y *Alternaria* (Figura 18).

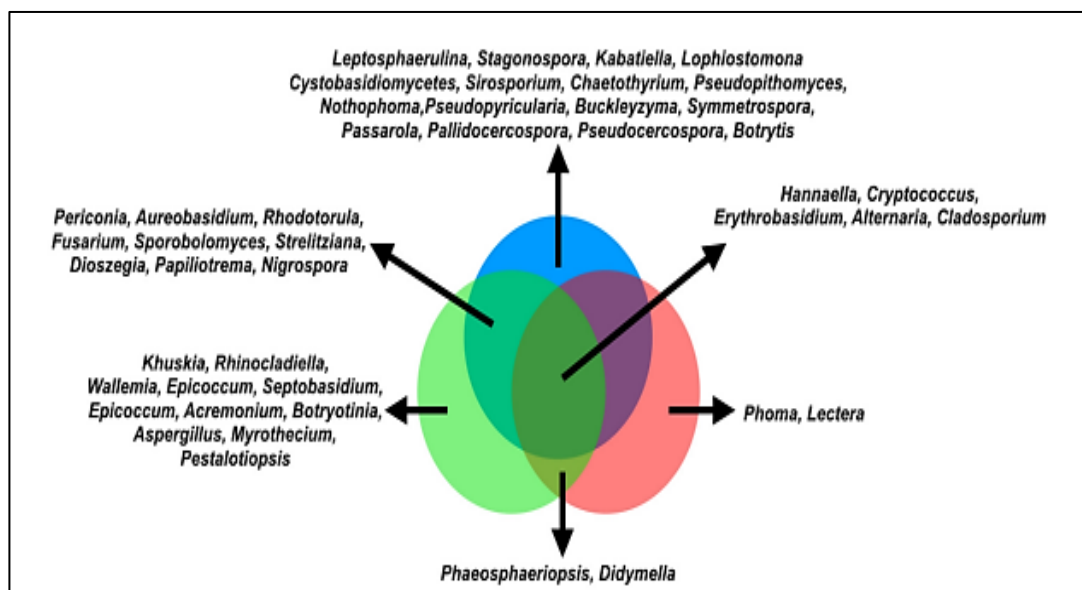


Figura 18 Diagrama de Venn con la distribución diferencial de los géneros *C. stenophylla* (círculo azul), *C. racemosa* (círculo verde) y *C. arabica* (círculo rojo).

Cada flecha representa hongos específicos para cada especie de café.

Fuente: (L. de Sousa et al., 2021)

La interacción entre las plantas y las bacterias endófitas en la formación de nódulos radiculares especializados es bien conocida y estudiada intensamente para la familia Fabaceae. La mayoría de los rizobios fijadores de nitrógeno descritos pertenecen a la subclase α de proteobacterias, pero también se sabe que varias especies β -proteobacterianas del género *Burkholderia* nodulan leguminosas (Verstraete et al., 2013).

Las bacterias promotoras del crecimiento en la rizosfera del café pueden aumentar la producción de los cultivos, actuando como promotoras del crecimiento de plantas o suministrándoles nutrientes, como las bacterias solubilizadoras de fosfato y las bacterias fijadoras de nitrógeno (Caldwell et al., 2015). Una de las primeras bacterias fijadoras de nitrógeno que se asociaron a las plantas de café es la *Acetobacter diazotrophicus* que se encuentran en los tejidos de las plantas, así como en la rizosfera (Caldwell et al., 2015).

Fulthorpe et al., (2020) menciona que las raíces de café cultivadas en Centroamérica (Fulthorpe, Martin, & Marney, 2020) como *Glomus* (Fulthorpe, Martin, & Marney, 2020); también los hongos micorrízicos y *Bradyrhizobium* actúan como antagonistas, especialmente para el control de *Fusarium* y *Colletotrichum*, que son patógenos potencialmente destructivos para las plantas de café (L. P. de Sousa, Guerreiro, & Mondego, 2022).

También se determinó que la abundancia relativa de las bacterias y hongos se encontraba relacionada con la zona climática y las técnicas de manejo; observándose en lugares húmedos una abundancia mayor de las bacterias *Pantoea* y *Pseudomonas*, los hongos *Cladosporium* y *Dothidea*; mientras que en lugares secos se presentaban en mayor cantidad las bacterias *Burkholderia* y *Ktedenobacter*, y los hongos *Exidiopsis* y *Codinaeopsis* (Fulthorpe, Martin, & Isaac, 2020). Así como se determinó que algunas bacterias como *Pseudomonas*, *Methylobacterium*, *Bradyrhizobium*, *Azospirillum* y los hongos micorrízicos favorecen en la minimización de los efectos de sequía en las especies

de *C. racemosa*, *C. stenophylla*, y *C. liberica* (L. P. de Sousa, Guerreiro, & Mondego, 2022).

2.2.1.2 Microbioma de los granos de café

Las semillas de café (*Coffea* spp.) son los principales productos básicos en el mundo, producido principalmente en países tropicales como Brasil, Vietnam, Indonesia, Colombia y Ecuador; y consumido en los países del norte como EEUU y Europa (A. Davis et al., 2019).

La colonización de los tejidos internos por parte de un microorganismo endofítico comienza con la migración a los sitios de las semillas en germinación o las raíces en crecimiento. Por ejemplo, la cepa *Escherichia fergusonii*, produce sideróforos y ácido indol acético. Estas características hacen que esta cepa sea prometedora para su uso en el sistema de raíces y brotes, ya que la disponibilidad de sideróforos mejora la disponibilidad de hierro para las raíces, y el ácido indol acético actúa directamente sobre la parte aérea de la planta. Mientras que la cepa *Acinetobacter calcoaceticus* produce fosfatasa, la cual es de gran utilidad en la rizosfera. Si un endófito coloniza los tejidos internos de una planta viva en la que vive, se multiplica y sobrevive (Silva et al., 2012).

2.2.1.3 Microbioma en las hojas del café

Para las plantas terrestres, la filosfera representa la interfaz entre las partes sobre el suelo de las plantas y el aire. En el estudio realizado por de Sousa, et al en Brazil, se identificó una prevalencia de Enterobacteriales en *Coffea canephora*, Pseudomonadales en *C. arabica* 'Obatã', y un alta de dominancia bacteriana de Bacillales en *C. arabica* 'Catuaí'. Actinomycetales, Enterobacteriales, y Lactobacillales también se encontraron en 'Catuaí', a veces compitiendo con Bacillales por la primacía en la comunidad bacteriana (Figura 19). En este estudio también se evaluó la relación de los nutrientes presentes en las hojas con la presencia de microorganismos, identificándose una correlación entre la cantidad de Mn y la abundancia (L. P. de Sousa et al., 2018).

En estudios realizados sobre el microbiota de las hojas se ha evidenciado que la misma puede variar por su composición química, difiriendo entre plantas cultivables y silvestres. Por ejemplo, en el estudio de Pio de Sousa, 2021, *C. arabica* mostró una población menor de hongos, que *C. stenophylla*. Este fenómeno podría deberse que las plantas domesticadas tienen una pérdida de diversidad en sus microorganismos asociados (L. de Sousa et al., 2021). Sin embargo, no se sabe nada sobre la presencia y la diversidad de los endófitos en estas especies, ni su relación con la adaptación al medio ambiente (climas más cálidos y secos) y la resistencia a los patógenos (Guyot et al., 2020)

Recientemente, se han secuenciado las hojas de 72 especies silvestres de *Coffea* mediante la técnica Illumina con una profundidad de > 40X del tamaño del genoma. Estas especies proceden de dos colecciones de árboles vivos, una en la estación de Bassin Martin en Isla Reunión, que reúne especies de África, las Mascareñas y las Comores, mientras que la otra, en la estación de Kianjavato en Madagascar, incluye especies malgaches. Estos datos están disponibles en una base de datos de especies silvestres del café (WCSdb) para el análisis de la diversidad genética y metagenómica (Guyot et al., 2020).

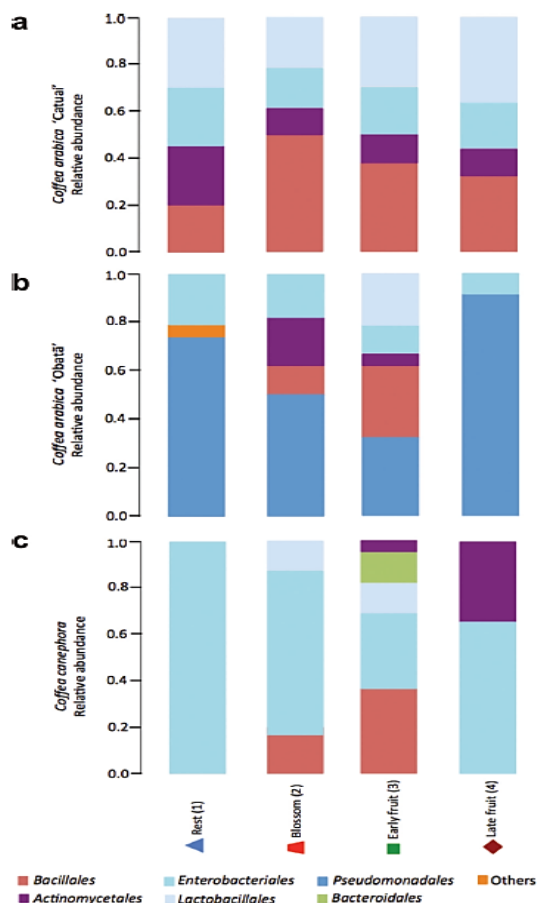


Figura 19 Clasificación taxonómica de secuencias bacterianas asociadas a hojas en muestras de café. a) *C. arabica* 'Catuai', b) *C. arabica* 'Obatã', and c) *C. canephora*.

Fuente: (L. P. de Sousa et al., 2018)

El conocimiento sobre los beneficios exactos conferidos por estos endosimbiontes aún es incompleto. Muchos estudios han propuesto que los endófitos de las especies noduladas están involucrados en la producción de fitohormonas. Desde la perspectiva del endosimbionte, la colonización de los tejidos internos de las plantas puede proporcionar un entorno estable, uniforme y protector (Lemaire, van Oevelen, et al., 2012).

Como se ha mencionado previamente, se ha detectado la presencia de endófitos en algunas muestras de café cultivado (*C. arabica* y *C. canephora*) cuya abundancia parece estar relacionada con las zonas climáticas, cabe preguntarse si también están presentes en las especies de café silvestre.

2.2.2 Microbioma en el Género *Pavetta*, *Psychotria* y *Sericanthe*

Otro tipo de endosimbiosis se encuentra en la familia Rubiaceae donde las agallas de las hojas con endófitos bacterianos se producen en los géneros *Pavetta*, *Psychotria* y *Sericanthe* (M. Pinto et al., 2018). *Psychotria* sensu stricto (aproximadamente 80 de 1400 especies), *Sericanthe* (11-12 de 17 especies) y *Pavetta* con aproximadamente 350 especies noduladas aceptadas (Lemaire, van Oevelen, et al., 2012).

Pavetta contiene el mayor número de especies noduladas, con el 85% de las especies del género que albergan endosimbiontes bacterianos dentro de los nódulos de las hojas en

África tropical continental y Asia (Lemaire, van Oevelen, et al., 2012). Se sabe que las especies de Rubiaceae albergan endosimbiontes bacterianos dentro de las cavidades internas en la lámina de la hoja, conocidas como nódulos bacterianos de la hoja o agallas de las hojas (Lemaire, van Oevelen, et al., 2012). La presencia de bacterias endófitas es visible a simple vista debido a la formación de manchas oscuras o agallas en las hojas de las plantas. Las endófitas de las plantas de Rubiaceae solo se han identificado recientemente y todas ellas pertenecen a *Burkholderia* (M. Pinto et al., 2018).

2.1 Familia *Burkholderiaceae* como endófito de plantas de la familia *Rubiaceae*

La familia Burkholderiaceae contiene los géneros *Burkholderia* (género tipo), *Cupriavidus*, *Lautropia*, *Limnobacter*, *Pandoraea*, *Paucimonas*, *Polynucleobacter*, *Ralstonia* y *Thermothrix* (Garrity et al., 2015). La diversidad filogenética dentro del género *Burkholderia* se ha utilizado para una reclasificación, la gran mayoría de las especies de *Burkholderia* se ubican en *Burkholderia sensu stricto* y los nuevos géneros son *Caballeronia*, *Paraburkholderia* y *Robbsia*. Excepto por este último, cada uno de los géneros consiste en especies ambientales de las cuales se informa que un número creciente causa infecciones en humanos. (Vandamme & Eberl, 2018).

El género *Burkholderia* incluye bacilos gramnegativos no formadores de esporas, que, dependiendo de la especie, utilizan uno o varios flagelos polares para moverse. Los individuos que pertenecen al género *Burkholderia* tienen una relación G+C entre 59 y 69,5%, son aerobios, oxidasa y catalasa positiva, mesófilos, y sintetizan polihidroxitirato como material de reserva. Los miembros del grupo *Burkholderia* son ubicuos ya que se encuentran en el agua y el suelo, así como en tejidos vegetales y animales. La versatilidad nutricional es una característica sobresaliente de este grupo bacteriano porque puede vivir como un biotrofo o como un saprofito. *Burkholderia* es un género bifuncional porque algunas de sus especies establecen relaciones simbiótico-mutualistas con las plantas, mientras que otras establecen asociaciones simbiótico-patógenas con plantas, animales y humanos (Figura 20) (Espinosa et al., 2020; Vandamme & Eberl, 2018).

Gran parte de la literatura reciente sobre especies endófitas o beneficiosas para plantas de *Burkholderia* se refiere a especies ahora clasificadas en el género *Paraburkholderia*. Sin embargo, estas bacterias son colonizadoras de las rizosferas y filosferas de varias especies de plantas, incluido el maíz, arroz, guisante, algodón, etc. (Vandamme & Eberl, 2018).

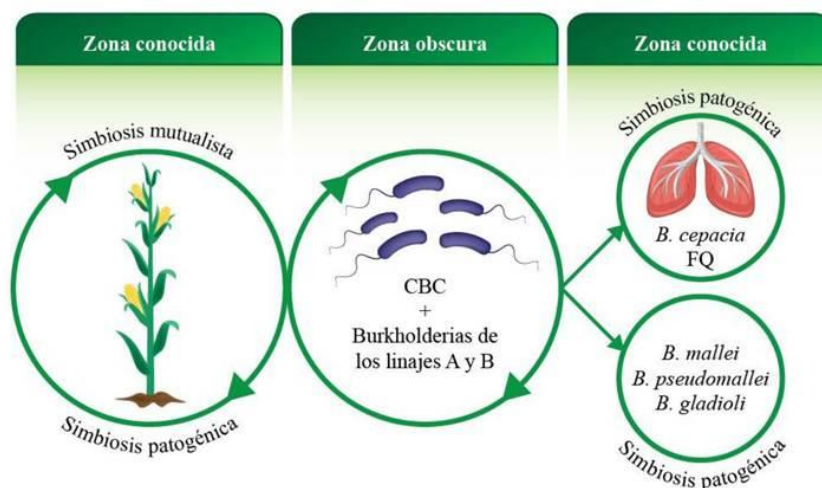


Figura 20. Asociaciones del género *Burkholderia*

Nota: El círculo en el centro muestra el grupo de especies *Burkholderia* (BCC y burkholderias de los linajes A y B). A la izquierda se representa el grupo de *Burkholderias* que establecen relaciones mutualistas y/o patógenas con las plantas; a la derecha se indican las relaciones patogénicas que las especies de *Burkholderia* establecen con los seres humanos y los animales.

Fuente: (Espinosa et al., 2020).

Se ha identificado en diferentes estudios la presencia de este microorganismo asociado a diferentes plantas. Por ejemplo, en las leguminosas, la especie β -proteobacterias del género *Burkholderia* se encargan de la nodulación y de la fijación del nitrógeno (Verstraete et al., 2013). En plantas de tomate la actividad nitrogenasa y la presencia de genes *niFH* en aislamientos de *Burkholderia* de la rizosfera tiene capacidad diazotrófica, es decir son intermediarias en el proceso de fijación biológica del nitrógeno, reduciendo el nitrógeno atmosférico que no puede ser asimilado por las plantas a amoníaco contribuyendo así a suplir las carencias de este elemento en los cultivos (Verstraete et al., 2013).

Burkholderia se ha identificado en el género *Coffea* de la familia Rubiaceae, como microorganismo endófito que se caracterizan por su capacidad de colonizar la parte interna los tejidos vegetales sin causar ningún síntoma negativo al huésped. (Duong et al., 2020). También, en el género *Coffea* se ha realizado el aislamiento bacteriano de *Burkholderia* de raíces y semillas, para la inhibición del crecimiento del hongo *F.oxysporum*, que es el agente causante del marchitamiento, también conocido como traqueomicosis que afecta a las especies de *C.arabica*, *C.canephora* y *C.liberica* logrando su reducción hasta un 40% (Duong et al., 2021).

También en el perfil taxonómico del microbioma endofítico de *Coffea canephora* el género *Burkholderia* se encuentra ocupando el 32,24% siendo el más dominante (Figura 17).

Por otro lado, en el análisis de la microbiota de la raíz de café Robusta, la función principal es la biosíntesis (69,97%), seguida de la generación de metabolitos precursores y energía (13,15%), y la degradación, utilización y asimilación (7,98%). Entre las funciones implicadas en la biosíntesis, la biosíntesis de aminoácidos (19,16%) fue la más abundante, seguida de la biosíntesis de nucleósidos y nucleótidos (17,73%), la biosíntesis de los cofactores, grupos prostéticos, portadores de electrones y vitaminas (12,15%),

biosíntesis de ácidos grasos y lípidos (7,97%), biosíntesis de carbohidratos (7,61%), biosíntesis de estructuras celulares (5,09%) y biosíntesis de metabolitos secundarios (3,62%). Mientras que, la microbiota epífita del café está compuesta por alta diversidad de microorganismos capaces de adherirse y vivir en la superficie de los tejidos vegetales aéreos y subterráneos. Aquí se encuentra el género *Burkholderia* conformado por dos especies (Duong et al., 2020a). La filosfera posee proteínas relacionada en la respuesta al estrés, como la glioxalasa, la ferretina, la peroxidasa, la catalasa y el superóxido dismutasa (Tran et al., 2022).

Con respecto al microbiota de la cereza del café, los microorganismos juegan un papel importante durante el proceso de postcosecha, especialmente al degradar la capa de mucilago durante la fermentación y también influyendo positiva y negativamente en la calidad organoléptica del producto final, así como en la seguridad frente a la presencia de micotoxinas. Entre los microorganismos encontrados se destaca el género *Achromobacter* y *Burkholderia* conformada con una especie (Duong et al., 2020a)

En *C.arabica* se demostró la capacidad de fijar nitrógeno atmosférico y mejorar la nutrición de la planta mediante uso de las especies bacterianas asociadas a la raíz, en la cual se destaca el género *Burkholderia* (L. P. de Sousa, Guerreiro, & Mondego, 2022).

Cabe destacar que el género *Burkholderia* también permite la solubilización de fósforo en raíces y semillas de *C.arabica*, *C.canephora* y *C.liberica*. Otra capacidad es la movilización a través de la producción de sideróforos asociado con raíces, hojas y semillas de *C.arabica*, *C.canephora* y *C.liberica*. Finalmente producen fitohormonas como la auxina o reguladores como la enzima ACC desaminasa capaz de reducir el nivel de etileno y actúa como mecanismo de biocontrol frente a patógenos mediante la producción de enzimas (quitinasas, gelatinasas, lipasas y proteasas), así como algunos compuestos activos (HCN) y sideróforo (Duong et al., 2020a).

También se ha aislado *Paraburkholderia coffeinitolerans* spp. como una bacteria que degrada la cafeína del suelo en plantaciones de té. La bacteria *Burkholderia* spp. puede tolerar un alto nivel de cafeína, hasta 20 g/L, pero solo puede degradar el 10% de cafeína inicial (Win et al., 2019).

Por otro lado, el género *Burkholderia* es el principal candidato para el control biológico de patógenos fúngicos de plantas, ya que posee la capacidad de producir compuestos antifúngicos y permite la degradación enzimática de la pared celular. La inhibición de *F.oxysporum* por un aislado de *Burkholderia seminalis* se atribuyó principalmente a la producción de sideróforos. Además, la actividad de la gelatinasa (proteasa que degrada el colágeno desnaturalizado) esta correlacionada con la actividad antifúngica de este aislado (Duong et al., 2021).

2.2 Tecnologías de secuenciación de nueva generación (NGS)

La secuenciación de nueva generación (NGS) es una tecnología de secuenciación paralela masiva que se utiliza para determinar el orden de los nucleótidos en genomas completos o regiones específicas de ADN o ARN, ofreciendo un rendimiento, escalabilidad y velocidad muy altos (Slatko et al., 2018). La NGS puede ser aplicada en análisis metagenómicos, cuyo objetivo es secuenciar partes aleatorias del material genético de un individuo, e incluso todas las partes de un genoma presentes en un organismo para buscar el origen de estos, es decir su taxonomía, logrando incluso si se tiene suficientes partes

de este, obtener genomas individuales completos, que pueda dar un conjunto de información acerca de las especies en estudio (Slatko et al., 2018).

Las técnicas de nueva generación más conocidas son: Pirosecuenciación-Roche 454, PacBio, Nanopore e Illumina.

2.2.1 Illumina

Esta técnica se ejecuta mediante el método de amplificación de puente, en el cual los fragmentos de ADN se colocan sobre una superficie sólida de vidrio separada por carriles. Cada carril está completamente recubierto por oligonucleótidos complementarios a los adaptadores de cada fragmento que se va a secuenciar, por lo que permiten que cada fragmento se pueda anclar a la celda de flujo. Una vez anclados los segmentos, la polimerasa inicia el proceso de copia en la hebra de ADN y genera una hebra reversa complementaria. La hebra original es retirada; mientras que la hebra reversa, a través de una secuencia terminal, se pliega y se ancla a su respectiva secuencia complementaria de oligonucleótido, y así queda en forma de puente. Después la polimerasa genera una hebra complementaria idéntica a la original, resultando en dos hebras clonadas del segmento inicial. Este proceso se repite masivamente hasta formar millones de copias de cada fragmento. Al finalizar la ampliación clonal, se retiran todas las hebras reversas y quedan únicamente las hebras idénticas a las originales (Figura 21). En este punto, en la placa se introducen nucleótidos modificados con etiquetas fluorescentes específicas para cada tipo (Rubio et al., 2020).

Los nucleótidos utilizados presentan una modificación química (terminadores reversibles) que evitan la unión de más de un nucleótido marcado en cada sitio de reacción, de tal manera que se puede ubicar el que corresponde a cada punto en la secuencia y se disminuye el riesgo de errores en la secuenciación. Cada vez que una base se adhiere, emite una fluorescencia propia que permite su identificación. La etiqueta debe ser removida antes de la colocación del siguiente nucleótido para evitar que dos bases emitan señal al mismo tiempo. Al finalizar la primera lectura, el fragmento resultante es retirado. Este paso se repite simultáneamente con todas las hebras del mismo clúster de forma paralela hasta completar la secuenciación. La exactitud de la secuenciación está determinada por la intensidad de la señal, y la longitud de las lecturas, por el número de ciclos realizados, alcanzando una longitud por lectura de hasta 300 pb (Rubio et al., 2020)

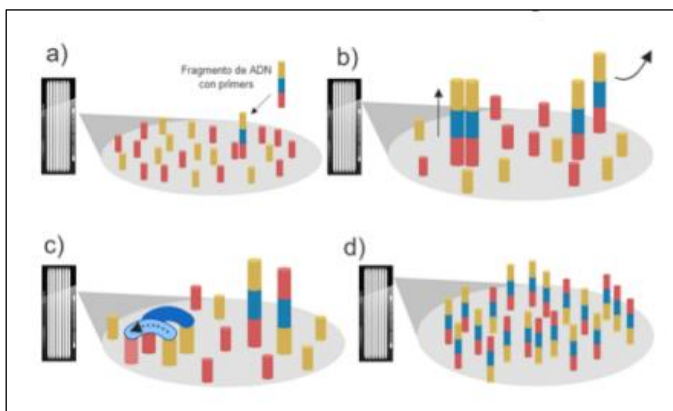


Figura 21 Generación de hebras mediante PCR puente.

Nota: a) El fragmento de ADN se ancla a la celda de flujo por medio de la unión de sus adaptadores a los oligonucleótidos complementarios, b) La polimerasa genera una hebra reversa complementaria y la hebra original es retirada, c) La hebra reversa se pliega y queda en forma de puente, donde la polimerasa genera una hebra complementaria idéntica a la original, d) El proceso se repite masivamente.

Fuente: (Rubio et al., 2020).

En Anexos

Anexo 1 se visualiza los diferentes tipos de Illumina que actualmente se utilizan para la secuenciación, indicando características generales de las mismas, como: longitud de lectura, tiempo de secuenciación, bases totales por secuenciación, etc.

Por otro lado, las técnicas de Illumina han sido ampliamente utilizadas en la metagenómica de diferentes organismos, incluyendo las plantas, por ejemplo, se ha utilizado las diferentes tecnologías de Illumina para acceder a la fracción no cultivable de endófitos bacterianos en tejido vegetales, lo cual permite comprender los microbios que viven sobre o dentro de las plantas en condiciones naturales y estudiar la manera en la que interactúan (Mushtaq et al., 2022). Por ejemplo, en el estudio de Tran et al. se realizó el análisis de datos de microbioma endófito de raíces de pimienta negra (*Piper nigrum* L.) de las Tierras Altas Centrales de Vietnam, obteniendo como resultado los perfiles taxonómicos y funcionales del microbioma endofítico de las raíces de pimienta negra, clasificando un total de 189.883 lecturas de las cuales se analizaron 190.058. Se identificaron dos filos, Proteobacteria (87,5%) y Actinobacteria (12,5%) de la muestra de raíz. Entre los géneros identificados, *Burkholderiales* (43,8%) fue el más dominante, seguido de *Xanthomonadales* (21,9%), *Micrococcales* (7,8%), *Micromonosporales* (4,7 %) y *Rhizobiales* (4,7%) (Tran et al., 2022).

También se ha utilizado algunas tecnologías de Illumina como MiSeq para correlacionar las alteraciones microbianas con la presencia de enfermedades en plantas, por ejemplo, Jongman et al., (2020) estudió la pudrición del extremo del tallo en el mango, llegando a la conclusión que varias condiciones afectan los microbiomas del mango, lo que influye en la aparición de esta enfermedad. A su vez destacaron que esta técnica se puede utilizar para explicar detalladamente las relaciones y las variaciones de composición entre las poblaciones de hongos y bacterias en el mango, así como, en otras frutas, permitiendo el desarrollo de aplicaciones más efectivas contra patógenos específicos y, a su vez, mejorar el control y la gestión de enfermedades.

2.3 Herramientas bioinformáticas para análisis de datos metagenómicos

Cuando se trabaja con datos resultantes de secuenciación de nueva generación, los datos crudos obtenidos de la secuenciación necesitan pasar a través de diferentes herramientas, con el objetivo de generar el resultado final esperado. El conjunto de herramientas a utilizar para un análisis en un orden en específico es comúnmente referido como “workflow” o “pipeline”.

Estos flujos de trabajo en bioinformática se caracterizan porque la salida de una herramienta se puede usar fácilmente como entrada para otra herramienta sin una configuración extensa.

2.3.1 Control de calidad

Uno de los primeros pasos en los análisis bioinformáticos es analizar la calidad de las lecturas contenidas en archivos FASTQ, que son obtenidos luego de la secuenciación, para ello una de las herramientas más utilizadas es FastQC. FastQC lee un conjunto de archivos de secuencia y produce a partir de cada uno un informe de control de calidad que consta de una serie de reportes diferentes, cada uno de los cuales ayudará a identificar un tipo potencial diferente de problema en los datos (Ahmad et al., 2019).

2.3.2 Mapeo con Bowtie2

La coincidencia (también conocida como mapeo) de secuencias biológicas es esencialmente una tarea computacional. Bowtie2 se utiliza a menudo para alinear las lecturas cortas contra un genoma de referencia, dado que requiere menos RAM y es capaz de realizar una buena alineación. El resultado del Bowtie2 es el formato SAM. Posteriormente con la herramienta samtools se convierte los archivos SAM al formato BAM para eliminar variantes duplicadas y falsos positivos de los datos, además de cifrar y reducir el tamaño de los datos de secuencia (Ahmad et al., 2019).

2.3.3 Ensamblaje con AByss

ABySS son las siglas de Assembly By Short Sequencing, y es una herramienta para el ensamblaje del genoma de novo utilizando datos de lectura corta. Esta herramienta utiliza una representación de grafos de Bruijn, que permiten el cálculo paralelo del algoritmo de ensamblaje. El algoritmo ABySS consiste en dos etapas: 1) todas las posibles subcadenas de longitud k (denominadas k-mers) se generan a partir de las lecturas de secuencia. El conjunto de datos k-mer se procesa para eliminar los errores de lectura y se construyen los contigs iniciales; 2) se extienden los contigs resolviendo ambigüedades en superposiciones (Simpson et al., 2009).

2.3.4 NGPhylogeny.fr

NGPhylogeny.fr es un servicio web dedicado al análisis filogenético, que proporciona un conjunto completo de herramientas filogenéticas y flujos de trabajo adaptados a diversos contextos y diversos niveles de experiencia del usuario (NGPhylogeny, 2022). En la Figura 22 se visualiza el flujo de trabajo básico utilizado en NGPhylogeny.fr :

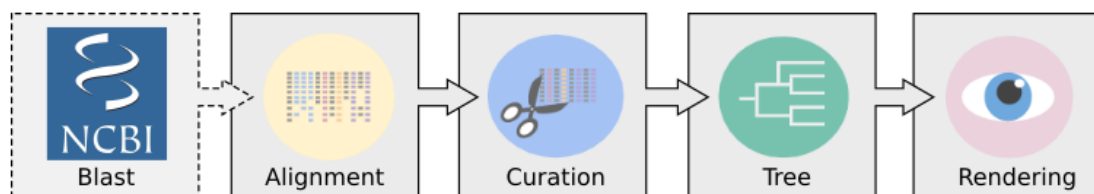


Figura 22 Flujo de trabajo de NGPhylogeny.fr

Fuente: (NGPhylogeny, 2022).

Todos los flujos de trabajo toman un archivo FASTA como entrada, preferiblemente no alineado, y producen múltiples archivos de alineación de secuencia (FASTA o PHYLIP) y archivos de árbol filogenético (formato Newick). Los árboles filogenéticos se visualizan dinámicamente en iTOL (Lemoine et al., 2019).

Para el alineamiento de secuencias múltiples se utiliza MUSCLE, que produce una medida de distancia más precisa para evaluar la relación de dos secuencias. La medida de distancia se actualiza entre las etapas de iteración (Chagoyen, 2013).

La configuración predeterminada de MUSCLE está diseñada para ofrecer la mejor precisión. Se utiliza la agrupación UPGMA. Que es más rápida para gran número de secuencias. En la Figura 23, se visualiza el flujo del algoritmo MUSCLE. Hay tres etapas principales: Etapa 1 (borrador progresivo), Etapa 2 (mejora progresiva) y Etapa 3 (refinamiento). Una alineación múltiple está disponible al finalizar cada etapa, momento en el cual el algoritmo puede terminar (R. C. Edgar, 2004).

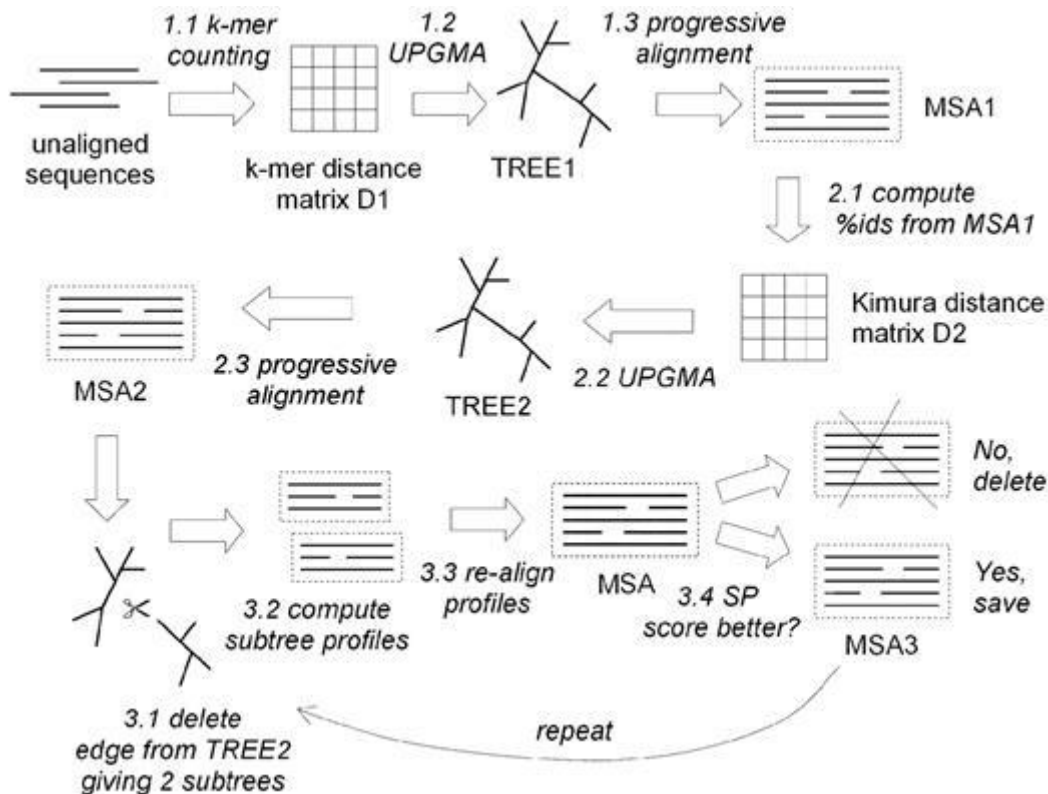


Figura 23 Flujo del algoritmo de MUSCLE.

Fuente: (R. C. Edgar, 2004)

Por otro lado, FastTree es el encargado de inferir árboles filogenéticos de máxima probabilidad aproximada, a partir de alineamientos de secuencias de nucleótidos o proteínas. Utiliza una estrategia de neighbor-joining para construir rápidamente un árbol filogenético aproximado y luego lo mejora mediante el método de máxima verosimilitud. FastTree puede manejar alineaciones con hasta un millón de secuencias en una cantidad razonable de tiempo y memoria (NGPhylogeny, 2022).

3 Metodología

3.1 Selección de datos de secuenciación masiva para análisis bioinformático

En este estudio se utilizó secuencias pareadas de Illumina obtenidas por secuenciación masiva de estudios previos realizados por (J. Charr et al., 2020) y (Ly et al., 2020), en los cuales se analizaron hojas de plantas de diferentes especies de los géneros *Coffea*, *Vangeria* y *Pavetta*.

Como controles positivos para la evaluación del funcionamiento del pipeline desarrollado, se utilizaron secuencias obtenidas de hojas de plantas, de dos especies de la familia Rubiaceae (*Tricalysia lasiodelphys* y *Tricalysia semidecidua*), en las cuales en estudios previos de secuenciación de la planta se identificó en el genoma de la planta como contaminación la presencia de bacterias del género *Burkholderia* (Tabla 17).

Tabla 17 Datos utilizados en estudio.

Nombre de especie	Origen geográfico	Nombre de archivos	Número de secuencias	Longitud de secuencias
<i>Tricalysia semidecidua</i> (Control positivo)	Tanzania y el norte de Mozambique (Darbyshire et al., 2019)	Tri-sem-86_1.fq Tri-sem-86_2.fq	65.171.012	100
<i>Tricalysia lasiodelphys</i> (Control positivo)	Camerún y Gaibón(Onana, 2013)	Tri-las-84_1.fq Tri-las-84_2.fq	62.973.455	100
<i>Coffea arabica</i>	Sur oeste de Etiopía (provincias de Kaffa e Illubador). Sudán del Sur. Norte de Kenia (Mt Marsabit) (Ferreira et al., 2019)	ET39-7_S12_R1 ET39-7_S12_R2	9.286.4788	35-126

<i>Coffea canephora</i>	África Occidental (Guinea y Liberia) a Uganda y Sudán del Sur. De República Centroafricana a Angola (Campuzano et al., 2021)	C399_HD_200_S12_R1 C399_HD_200_S12_R2	12.606.8297	35-126
<i>Coffea eugenoides</i>	Kenya y Uganda (PlantNet, 2022)	DA_NoIndex_L002_R1_001.clean.fastq DA_NoIndex_L002_R2_001.clean.fastq	74.087.113	50-100
<i>Coffea pseudozangebariae</i>	África del Este (PlantNet, 2022)	HP_ACAGTG_L004_R1_001 HP_ACAGTG_L004_R2_001	75.986.686	100
<i>Coffea racemosa</i>	África del Este (PlantNet, 2022)	IB_ACAGTG_L005_R1_001 IB_ACAGTG_L005_R2_001	72.751.510	100
<i>Coffea stenophylla</i>	África del Oeste-Sierra Leona, Costa de Marfil, Guinea (PlantNet, 2022)	FB_GTGAAA_L002_R1_001 FB_GTGAAA_L002_R2_001	75.588.827	100
<i>Coffea humilis</i>	África del Oeste-Costa de marfil (PlantNet, 2022)	GH_GCCAAT_L002_R1_001 GH_GCCAAT_L002_R2_001	59.130.390	100

<i>Coffea dolichophylla</i> o <i>Coffea perrieri</i>	Madagascar (PlantNet, 2022)	DOL_ACAGTG_L001_R1_001 DOL_ACAGTG_L001_R2_001	75.069.239	100
<i>Coffea humblotiana</i>	Madagascar-canal de Mosambique (PlantNet, 2022)	HUMB_GTGAAA_L004_R1_001 HUMB_GTGAAA_L004_R2_001	67.786.173	100
<i>Coffea ebracteolata</i>	Costa Marfil (PlantNet, 2022)	PSI_ACAGTG_L006_R1_001 PSI_ACAGTG_L006_R2_001	85.464.288	100
<i>Coffea ambongensis</i>	Madagascar (Kewscience, 2022).	NG-28980_CoffeaAmb_lib551863_7805_2_1 NG-28980_CoffeaAmb_lib551863_7805_2_2	8.000.000	21-151
<i>Coffea bissetiae</i>	Madagascar occidental (Tropical Plants Database, 2022)	NG-28980_CoffeaBis_lib551862_7805_2_1 NG-28980_CoffeaBis_lib551862_7805_2_2	8.000.000	21-151
<i>Coffea boinensis</i>	Madagascar occidental (A. Davis & Rakotonasolo, 2008)	NG-28980_CoffeaBoi_lib551861_7805_2_1 NG-28980_CoffeaBoi_lib551861_7805_2_2	8.000.000	27-151
<i>Vangeria infausta</i>	Angola, Botsuana, Kenia, Malawi, Mozambique, Namibia, Ruanda,	M19_6_1.fq.gz M19_6_2.fq.gz	814.839.924	100

	Sudáfrica, Suazilandia , Tanzania, Uganda, Zambia y Zimbabue (Maroyi, 2018)			
<i>Pavetta schumanniana</i>	África hasta el Sudeste Asiático, Nueva Guinea, Australia, Nueva Caledonia y Vanuatu (Reynolds, 1993b)	M9129_R1.fastq.gz M9129_R2.fastq.gz	54.089.287	100

3.2 Desarrollo de pipeline con herramientas bioinformáticas, para la identificación de microorganismos endófitos del género *Burkholderia*

Los análisis bioinformáticos fueron realizados por la línea de comandos en Ubuntu 22.04 con el uso de la máquina virtual VirtualBox Versión 6.1.34, en Anaconda con el ambiente llamado genómica, donde se instalaron todos los paquetes necesarios para los análisis respectivos.

El flujo de trabajo (pipeline) desarrollado para la identificación de la presencia de microorganismos endófitos del género *Burkholderia*, en datos de secuenciación masiva, de hojas de diferentes especies de los géneros *Coffea*, *Vangeria* y *Pavetta*, se visualiza en la Figura 24.

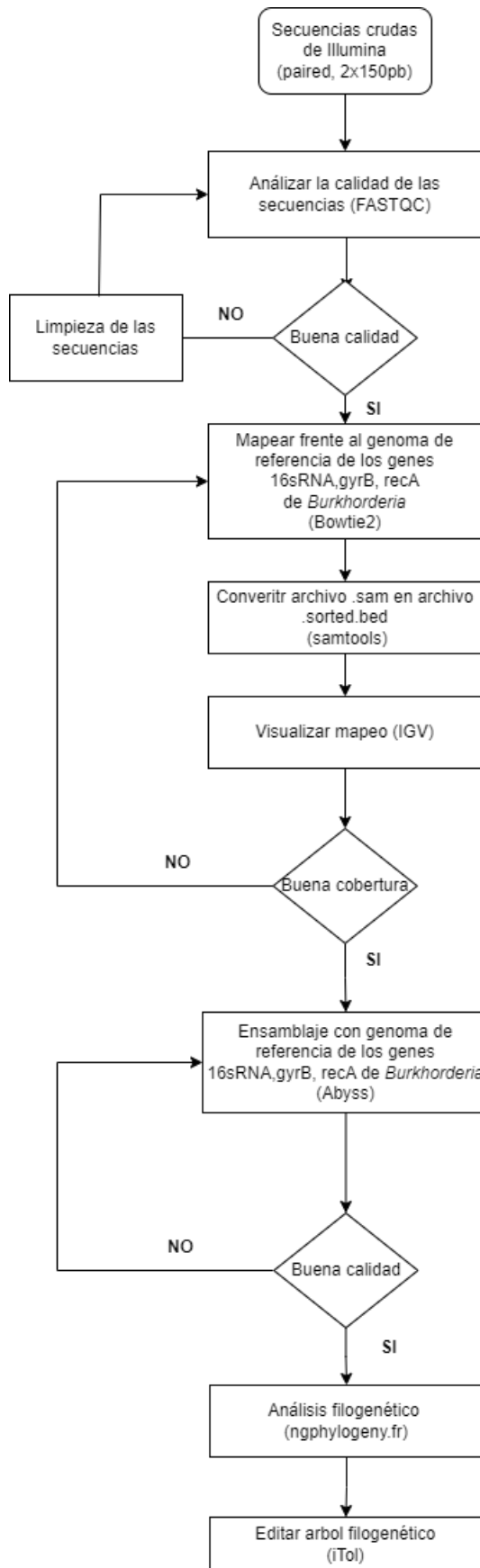


Figura 24 Flujo de trabajo para identificación de bacterias endófitas del género *Burkholderia*.

3.2.1 Evaluación de la calidad de las secuencias de Illumina

Para corroborar la longitud de las secuencias y la calidad de estas, se analizó la calidad por muestra utilizando el software FastQC, con el uso del comando:

```
>fastqc nombre_del_archivo.fq
```

Como resultado se generó un reporte en versión html, con los siguientes estadísticos básicos: calidad de la secuencia por base, puntuación de calidad por secuencia, contenido por base por secuencia, contenido de GC, contenido de GC por secuencia, contenido de base N, distribución de longitud de secuencia, secuencias duplicadas, secuencias sobreexpresadas, k-mers sobreexpresados (Babraham Institute, 2022).

3.2.2 Limpieza de las secuencias con Trimmomatic

Si la calidad de las secuencias no era buena en alguno de los estadísticos evaluados previamente, se utilizó el programa Trimmomatic para mejorar la calidad de estas, utilizando el comando:

```
>java -jar <path to trimmomatic jar> SE [-threads <threads>] [-phred33 | -phred64] [-trimlog <45logfile>] <input> <output> <step 1> ...
```

Donde:

SE: corresponde al modo single end

[-threads <threads>]: indica el número de subprocesos informáticos utilizados para ejecutar.

[-phred33 | -phred64]: nivel de la calidad de las bases en la secuenciación dependiendo del pipeline de Illumina utilizado.

[-trimlog <logFile>]: un archivo trimlog que crea un registro de todas las lecturas leídas y limpias, donde se incluye nombre de la lectura, la longitud de la secuencia resultante, la ubicación de la primera base, la ubicación de la última base en la lectura original y la cantidad recortada desde el final.

<input>: archivos en formato fastq, que pueden encontrarse comprimidos con extensión .gz.

<output>: archivos en formato fastq, que pueden encontrarse comprimidos con extensión .gz.

<step 1> ...: se pueden especificar varios pasos según sea necesario, utilizando argumentos adicionales al final. Aquí se puede incluir: ILLUMINACLIP que se encarga de retirar los adaptadores de Illumina proporcionados en el archivo TruSeq3-PE.fa (incluido); LEADING: corte las bases del comienzo de una lectura, si está por debajo de un umbral de calidad; TRAILING: corte las bases del final de una lectura, si está por debajo de un umbral de calidad.

3.2.3 Mapeo de secuencias

Si las secuencias de las lecturas forward and reverse, son de buena calidad, se procede a realizar el mapeo de estas frente al genoma de referencia para los genes 16s RNA, *gyrB* y *recA*, con la herramienta Bowtie2, para identificar y filtrar solo las secuencias correspondientes a bacterias del género *Burkholderia*, y separarlas del huésped de *Coffea*.

Previo al mapeo se crearon bases de datos con secuencias de referencia descargadas de NCBI, para los genes 16s RNA, *gyrB* y *recA* p (Anexo 2) para incrementar el poder discriminatorio; a través del comando:

```
>bowtie2 -build <fasta> <bt2-idx>
```

Donde:

<fasta> corresponde al archivo que contiene los genomas de referencia con el cual se creará la base de daos.

<bt2-idx> corresponde al índice generado, requerido para el mapeo.

Posteriormente, se realizó el mapeo para lo cual se utilizó el comando:

```
>bowtie --very-fast-local -x <bt2-idx> -U <r> -U <r> --no-unal -S <sam>
```

Donde:

x <bt2-idx>: el nombre base del índice para el genoma de referencia.

--very-fast-local: en este modo, Bowtie 2 no requiere que toda la lectura se alinee de un extremo a otro. Más bien, algunos caracteres pueden omitirse desde los extremos para lograr la mayor puntuación de alineación posible.

-U <r>: lista separada por comas de archivos que contienen lecturas no emparejadas que deben alinearse.

--no-unal: suprime registros SAM para lecturas que no se pudieron alinear

-S <sam>: archivo SAM en el que se escribe el resultado de las alineaciones

Para la interpretación de los resultados, los archivos obtenidos del mapeo en formato .SAM, se convierten en formato BAM con la herramienta *samtools* para posteriormente visualizarlos en el programa Integrative Genomics Viewe (IGV), el mismo que se lo descargo en <https://software.broadinstitute.org/software/igv/AlignmentData>.

En IGV, se visualiza la cobertura del mapeo en relación con los genomas de referencia para los tres genes en estudio. Cuando IGV se acerca al umbral de visibilidad de lectura de alineación (de forma predeterminada, 30 KB), la pista de cobertura muestra la profundidad de las lecturas que se muestran en cada locus como un gráfico de barras grises. Si un nucleótido difiere de la secuencia de referencia en más del 20% de las lecturas ponderadas de calidad, IGV colorea la barra en proporción al recuento de lectura de cada base (A, C, G, T), como se visualiza en la Figura 25 (Integrative Genomics Viewer, 2022).

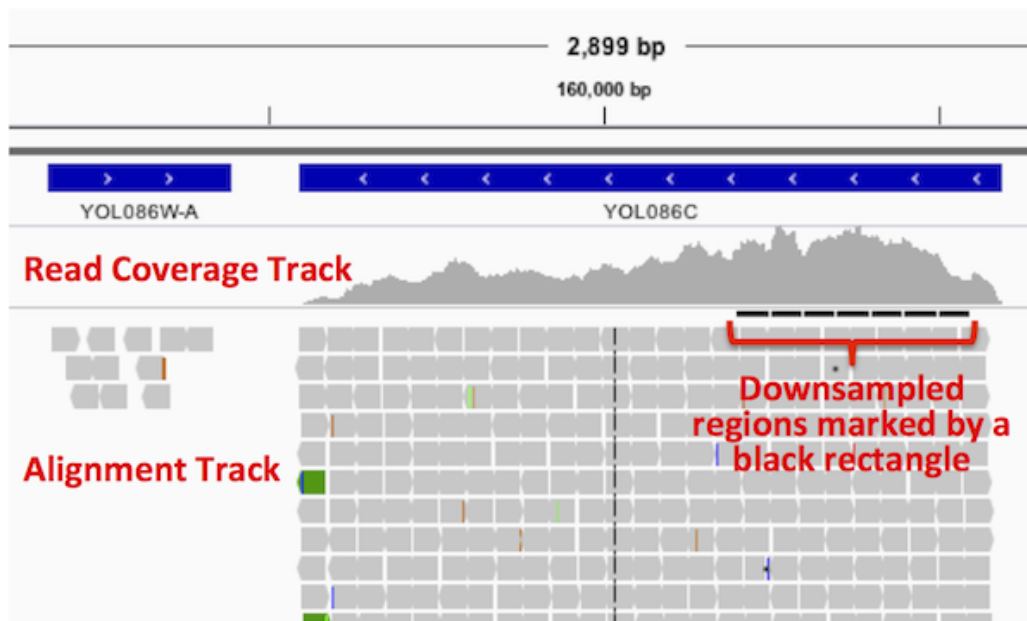


Figura 25 Visualización en IGV.

Fuente: (Integrative Genomics Viewer, 2022)

Si se identifica una buena cobertura, se procede con el proceso de ensamblaje.

3.2.4 Ensamblaje con Abyss

Las lecturas obtenidas del mapeo frente a los tres genes (16s RNA, *gyrB* y *recA*) fueron ensambladas usando Abyss, dado que esta herramienta permite trabajar con secuencias cortas como las obtenidas por Illumina, tiene un consumo de memoria eficiente y el tiempo de corrida es bastante rápido.

Para los ensamblajes, se utilizó el siguiente comando:

```
>abyss-se name=nombre_archivo_output k=71 se='nombre_archivo_input'fa'
```

Donde:

-pe: se refiere a secuencias pareadas.

name: el nombre del archivo de salida donde se almacenarán los contigs resultado del ensamblaje.

k: se refiere al valor de *k*-mers considerado.

se: nombre del archivo.fasta resultado del mapeo a utilizarse para el ensamblado.

Para la evaluación de la calidad del ensamblaje se deben analizar los diferentes estadísticos generados, por ejemplo: N25 que calcula las longitudes de los contigs más grandes hasta alcanzar el 25 % de la longitud total de contig. La longitud mínima de contig en este conjunto es el número que se utiliza normalmente para informar el valor N25 de un conjunto de novo. Lo mismo ocurre con N50 y N75, que son el 50 % y el 75 % de la longitud total del contig, respectivamente.

Adicionalmente se observó de forma gráfica los contigs obtenidos por medio del uso de la herramienta BANDAGE.

3.2.5 Taxonomía y análisis filogenético

Para el análisis filogenético, se utiliza la herramienta NGPhylogeny.fr (<https://ngphylogeny.fr/>) disponible de forma libre, online (Lemoine, Correia, Lefort, et al., 2019)

Phylogeny.fr se basa en la aplicación de flujos de trabajo, con las siguientes etapas: i) búsqueda de secuencias basada en BLAST; ii) alineación de secuencias múltiples; iii) curación de la alineación; iv) inferencia filogenética de árboles; (v) visualización de árbol (Lemoine, Correia, Lefort, et al., 2019), como se visualiza en la Figura 26:

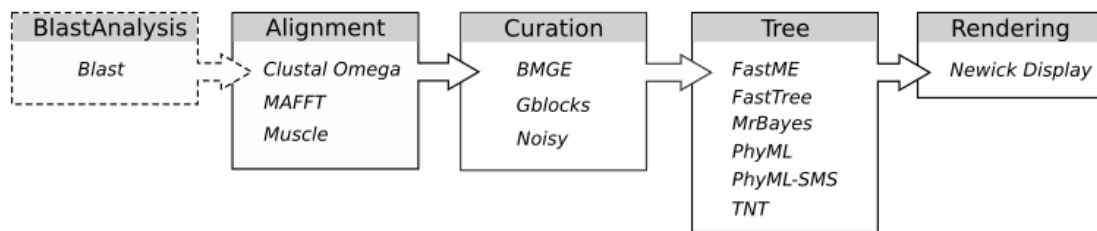


Figura 26 Flujo de trabajo Phylogeny.fr

Fuente: (Lemoine, Correia, Lefort, et al., 2019)

En primer lugar, se realizó la identificación taxonómica de las bacterias detectadas en las muestras, por medio del uso de la herramienta BLAST, que es un programa proporcionado por NCBI para alinear secuencias query con las presentes en una base de datos de destino seleccionada; en este caso inicialmente se trabaja con la base de datos de 16S Microbial rRNA sequences, para lo cual se debe cargar la secuencia en formato FASTA.

Como regla general para la determinación a nivel de especie de muchos grupos de microorganismos, si dos especímenes tienen una similitud de secuencia del 97% o superior se consideran una sola especie, con porcentajes menores no se asegura la identificación a nivel de género. También se debe analizar el valor de cobertura, un alto valor en un rango de >70% indica buenos resultados, si los resultados principales caen por debajo de este rango se debería revisar la secuencia (NCBI, 2023).

Posteriormente, de ser necesario se elabora un árbol filogenético utilizando Muscle, que será el encargado del alineamiento múltiple, para el cual se adjunta un archivo .fasta de input con no más de 50 secuencias, con la agrupación UPGMA (R. Edgar, 2004).

FastTree se aplicará solo si se identifica la presencia de bacterias del género *Burkholderia* en las diferentes especies del género *Coffea*, para analizar relaciones filogenéticas. En este caso se utilizará la herramienta FastTree, que utiliza una estrategia de neighbor-joining para construir rápidamente un árbol filogenético aproximado y luego lo mejora mediante el método de máxima verosimilitud. FastTree puede manejar alineaciones con hasta un millón de secuencias en una cantidad razonable de tiempo y memoria (NGPhylogeny, 2022). El árbol filogenético creado se manipuló con iTol para generar el árbol final.

4 Resultados y discusión

4.1 Evaluación de la calidad de las secuencias

El primer paso en el pipeline fue evaluar la calidad de las secuencias con las que se trabajó, para lo cual se utiliza el programa FASTQC. Cuando se obtuvo una mala calidad en algún parámetro, se procedió a realizar una limpieza con el uso de Trimmomatic. Los resultados finales de calidad obtenidos luego de la limpieza, se visualizan en la Tabla 18.

Tabla 18 Resultado evaluación de la calidad de las secuencias.

Nombre de especie	Calidad de la secuencia por base	Puntuación de calidad por secuencia	Contenido por base por secuencia	Contenido d GC	Contenido de GC por secuencia	Contenido de bases N	Distribución de longitud de secuencia	Secuencias duplicadas	Contenido de adaptadores	Secuencias sobreespresadas	kmers sobreespresados
<i>Tricalysia lasiodelphys</i> (Control positivo)	SI	SI	SI	SI	NO	SI	SI	SI	SI	SI	SI
<i>Tricalysia semidecdua</i> (Control positivo)	SI	SI	SI	SI	NO	SI	SI	Advertencia	SI	SI	SI
<i>Coffea arabica</i>	SI	SI	SI	SI	SI	SI	SI	Advertencia	SI	SI	SI
<i>Coffea canephora</i>	SI	SI	SI	Advertencia	NO	SI	SI	SI	SI	SI	SI
<i>Coffea eugenoides</i>	SI	Advertencia	SI	SI	Advertencia	SI	Advertencia	SI	SI	SI	SI
<i>Coffea pseudozangebariae</i>	SI	Advertencia	SI	SI	SI	SI	SI	SI	SI	SI	SI

<i>Coffea racemosa</i>	SI	Advertencia	SI	SI	Advertencia	SI	SI	SI	SI	SI	SI
<i>Coffea stenophylla</i>	SI	NO	SI	SI	SI	SI	SI	SI	SI	SI	SI
<i>Coffea humilis</i>	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI
<i>Coffea dolichophylla</i> o <i>Coffea perrieri</i>	SI	NO	SI	SI	SI	SI	SI	SI	SI	SI	SI
<i>Coffea humblotiana</i>	SI	Advertencia	SI	SI	Advertencia	SI	SI	SI	SI	SI	SI
<i>Coffea ebracteolata</i>	SI	Advertencia	SI	SI	SI	SI	SI	SI	SI	Advertencia	SI
<i>Coffea amboensis</i>	SI	SI	SI	SI	SI	SI	Advertencia	SI	SI	SI	SI
<i>Coffea bissetia</i>	SI	SI	SI	SI	NO	SI	Advertencia	SI	SI	SI	SI
<i>Coffea bionensis</i>	SI	SI	SI	SI	Advertencia	SI	Advertencia	SI	SI	SI	SI
<i>Vangeria infusta</i>	SI	SI	SI	SI	NO	SI	SI	SI	Advertencia	SI	SI
<i>Pavetia schumaniana</i>	SI	SI	SI	SI	Advertencia	SI	SI	SI	SI	SI	SI

Nota: se visualiza por parámetro como SI=cumple los criterios de calidad establecidos, NO=no cumple los criterios de calidad establecidos; y Advertencia cuando se emite una notificación en el nivel de calidad obtenido.

A continuación, se explica la interpretación de cada uno de los estadísticos generados en el programa FASTQC.

4.1.1 Calidad de la secuencia por base

Este estadístico indica para cada posición en la secuencia, los valores de calidad de las mimas, considerándose que un valor que no sea menor a 32 se considera de alta calidad (Babraham Institute, 2022). La representación de los resultados se visualiza en la **¡Error! No se encuentra el origen de la referencia.**, con una codificación por colores para

identificar puntajes de calidad, siendo buenos (verdes), aceptables (amarillos) y malos (rojos).

En todas las muestras analizadas se obtuvo como resultado una calidad alta en este parámetro.

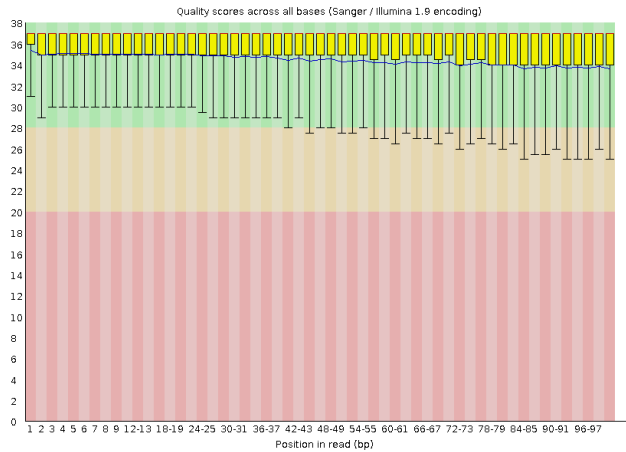


Figura 27 Calidad de las secuencias por base.

4.1.2 Puntajes de calidad por secuencia

La Figura 28; **Error! No se encuentra el origen de la referencia.**, indica la puntuación de calidad por secuencia, lo que permite ver si un subconjunto de las secuencias tiene valores de calidad bajos; si una proporción significativa de las secuencias tienen baja calidad general, esto podría indicar algún tipo de problema sistemático (Babraham Institute, 2022)

En la mayoría de las muestras se obtuvo como resultado una calidad alta en este parámetro. Sin embargo, en las muestras de las especies *Coffea eugenioides*, *Coffea pseudozangebariae*, *Coffea racemosa*, *humboldtiana* y *Coffea ebracteolate*, se obtuvo una advertencia, que se emite si la calidad media observada con mayor frecuencia es inferior a 27, lo que equivale a una tasa de error del 0,2% (Babraham Institute, 2022)

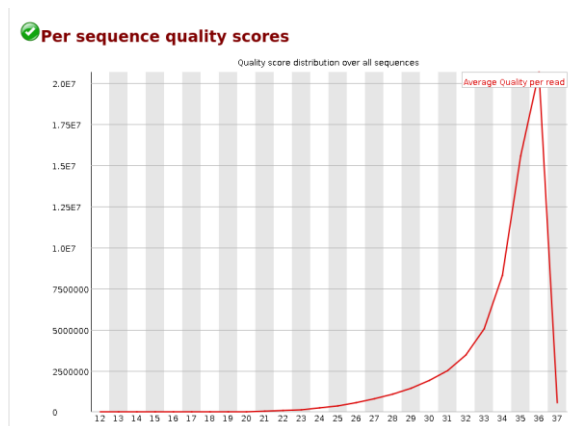
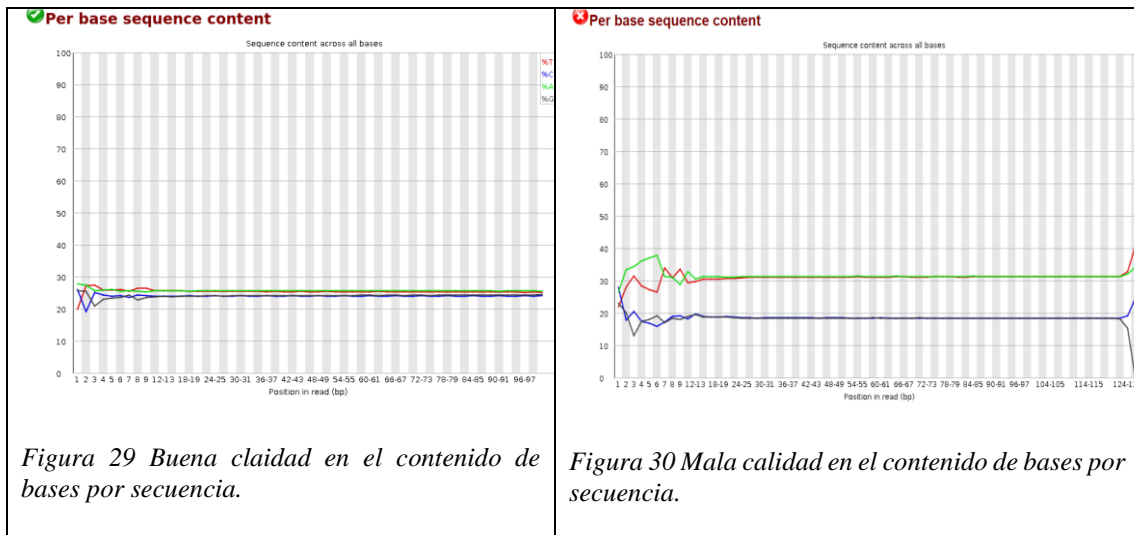


Figura 28 Puntuación de calidad por secuencia.

4.1.3 Contenido de bases por secuencia.

En la Figura 29, se observa los resultados en este parámetro; si existe una buena calidad las líneas de esta gráfica corren paralelas entre sí, lo que significa que hay poca diferencia entre las diferentes bases de la secuencia. A diferencia de la Figura 30, en la que se visualiza una mala calidad en la muestra de *Coffea canephora* antes de la limpieza, que se debe a que la diferencia entre A y T, o G y C es mayor que 20% en cualquier posición. Este error puede deberse a secuencias sobrerrepresentadas, error por fragmentación o error en la composición de las librerías (Babraham Institute, 2022).



En todas las muestras analizadas se obtuvo un buen valor en este parámetro.

4.1.4 Contenido de GC por secuencia.

En una biblioteca normal, se esperaría ver una distribución aproximadamente normal del contenido de GC donde el pico central corresponde al contenido general de GC del genoma. En este caso la distribución de forma inusual con varios picos más amplios se debe a una contaminación de especies diferentes de organismos en la muestra original (Figura 31) (planta del género *Coffea*), lo que se debe a que las secuencias son resultado de análisis de secuenciación masiva. Sin embargo, en este caso hay que considerar, que el objetivo del estudio es justamente separar las secuencias correspondientes al huésped, de microorganismos endófitos, dentro de los cuales podrían encontrarse bacterias del género *Burkholderia*.

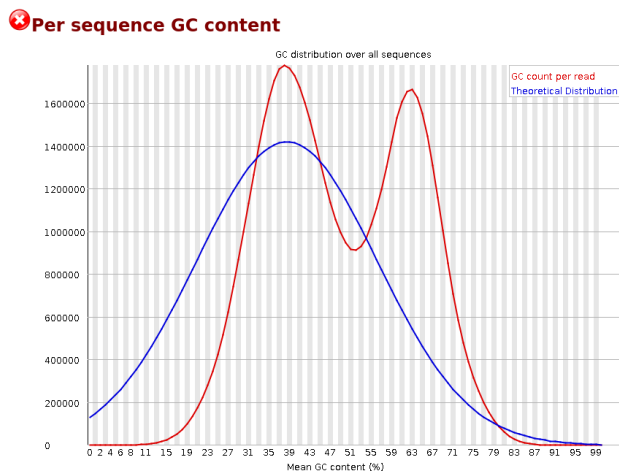


Figura 31 Contenido de GC por secuencia.

4.1.5 Contenido de bases N

El contenido de bases N hace referencia al porcentaje de N que fueron llamadas para sustituir bases, cuando durante el proceso de secuenciación no se pudo llamar una base con suficiente certeza (Babraham Institute, 2022). En este caso en todas las muestras se visualizó que el porcentaje es cero, lo que indica una buena calidad de las secuencias; siendo considerada como mala calidad porcentajes de N >20% (Figura 32; **Error! No se encuentra el origen de la referencia.**).

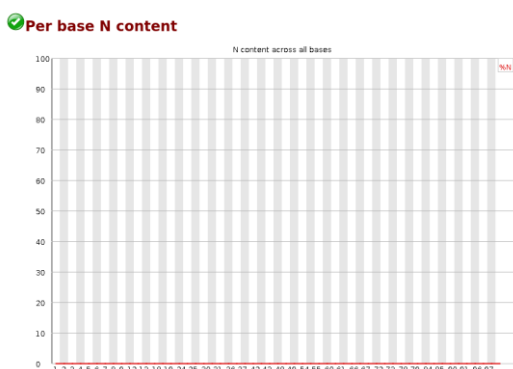


Figura 32 Contenido de Ns por base.

4.1.6 Distribución de la longitud de las secuencias.

En la Figura 33, se visualiza la distribución de los tamaños de fragmento, en el cual se visualiza un gráfico simple que muestra un pico solo en el tamaño de los fragmentos que corresponda, en este caso de 100pb. En todas las muestras se tuvo una distribución uniforme a excepción de la muestra de *Coffea eugenoides*, en la cual se presentó una advertencia que significa que todas las secuencias no tienen la misma longitud, sin embargo en la mayoría de las técnicas de secuenciación esto es completamente normal (Babraham Institute, 2022)

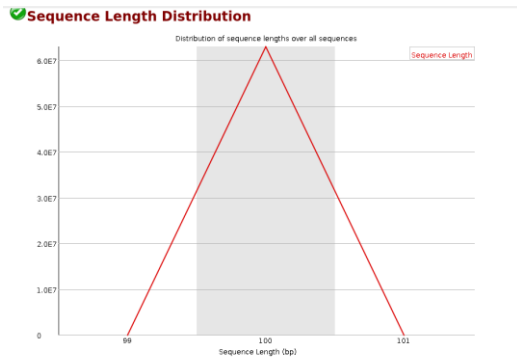


Figura 33 Distribución de la longitud de las secuencias.

4.1.7 Nivel de duplicación de las secuencias

Este estadístico indica el número relativo de secuencias con diferentes grados de duplicación, indicando si las secuencias ocurren solo una vez en el conjunto de datos. En la Figura 34 se visualiza dos líneas; donde la línea azul toma el conjunto de secuencia completo y muestra cómo se distribuyen sus niveles de duplicación. En la línea roja las secuencias se deduplican y las proporciones mostradas son las proporciones del conjunto deduplicado que provienen de diferentes niveles de duplicación en los datos originales (Integrative Genomics Viewer, 2022).

Adicionalmente, se visualiza en la parte superior de la gráfica, la pérdida general esperada de secuencias si la biblioteca se deduplicara; para dar una impresión razonable del posible nivel general de pérdida.

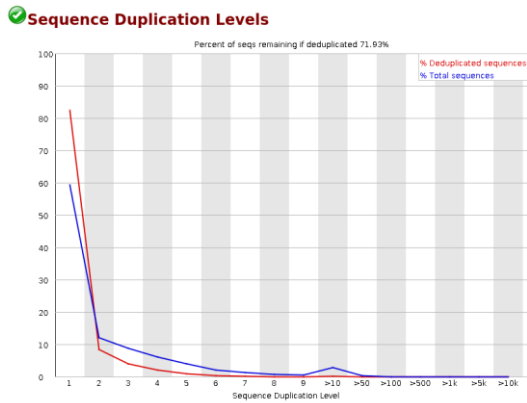


Figura 34 Niveles de duplicados en secuencias.

Por otro lado, en la Figura 35 se visualiza un nivel bajo de duplicación, que es un indicativo de un nivel muy alto de cobertura de la secuencia objetivo, donde la línea azul tiene un porcentaje mayor al 10% de duplicados en las secuencias forward and reverse, que podría deberse a un sesgo de enriquecimiento, por ejemplo, PCR sobre amplificado, que se consideran duplicados técnicos; pero en la gráfica no se puede distinguir de los duplicados biológicos. Se considera como un error cuando las secuencias duplicadas representan más del 50% del total, lo que podría indicar un conjunto de contaminantes o una duplicación técnica muy grave (Babraham Institute, 2022), que no se visualiza en ninguna de las muestras analizadas.

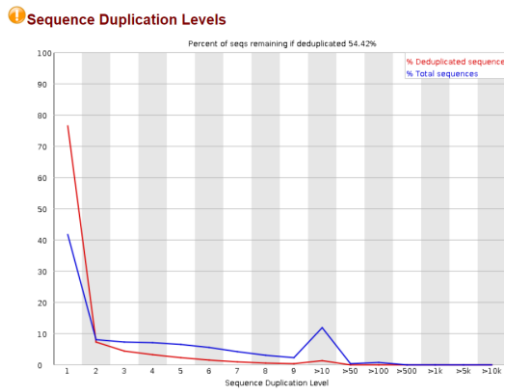


Figura 35 Niveles de duplicación por secuencia de *Tricalisia semidecidua*.

4.1.8 Contenido de adaptadores

El contenido de adaptadores indica si la biblioteca contiene una cantidad significativa de adaptadores, con el cual se puede evaluar si necesita recortar el adaptador o no; en si se visualiza un recuento porcentual acumulativo de la proporción de secuencias del adaptador en cada posición (Figura 36) (Babraham Institute, 2022)

Para las muestras analizadas se obtuvo como resultado una calidad alta en este parámetro.

Adapter Content

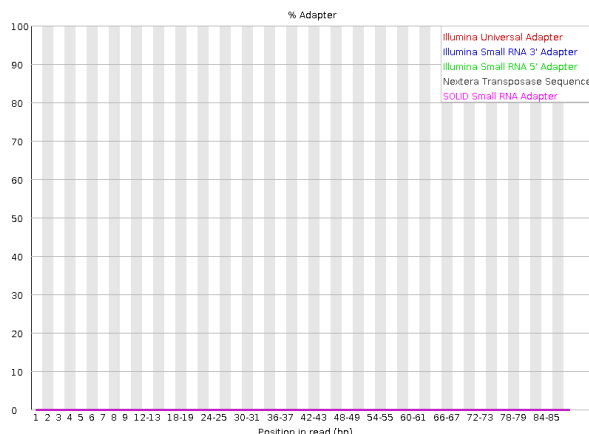


Figura 36 Contenido de adaptadores.

4.1.9 Secuencias sobreexpresadas y k-mers sobreexpresados

No se identificó secuencias sobreexpresadas en las muestras analizadas, excepto en la muestra de *Coffea ebracteolate*, lo que significa que una sola secuencia está muy sobrerrepresentada, en el conjunto significa que es altamente significativa biológicamente, que la biblioteca está contaminada, o no es tan diversa como se espera (Babraham Institute, 2022).

4.2 Mapeo de secuencias

Luego de analizar la calidad de las secuencias y realizar el mapeo con Bowtie 2, se obtuvo los siguientes resultados (Tabla 19).

Tabla 19 Resultado del mapeo realizado con Bowtie 2

Nombre de especie	Tasa de alineamiento global (%) Gen <i>16S RNA</i>	Tasa de alineamiento global (%) Gen <i>gyrB</i>	Tasa de alineamiento global (%) Gen <i>recA</i>
<i>Tricalysia lasiodelphys</i> (Control positivo)	0.04	0.00	0.00
<i>Tricalysia semidecidua</i> (Control positivo)	0.08	0.01	0.01
<i>Coffea arabica</i>	0.01	0.00	0.00
<i>Coffea canephora</i>	0.00	0.00	0.00
<i>Coffea eugenioides</i>	0.01	0.00	0.00
<i>Coffea pseudozangebariae</i>	0.01	0.00	0.00
<i>Coffea racemosa</i>	0.01	0.00	0.00
<i>Coffea stenophylla</i>	0.01	0.00	0.00
<i>Coffea humilis</i>	0.01	0.00	0.00
<i>Coffea dolichophylla</i> o <i>Coffea perrieri</i>	0.01	0.00	0.00
<i>Coffea humblotiana</i>	0.01	0.00	0.00
<i>Coffea ebracteolate</i>	0.00	0.00	0.00
<i>Coffea ambogensis</i>	0,03	0,00	0,00
<i>Coffea bissetia</i>	0.01	0.00	0.00
<i>Coffea bionensis</i>	0.01	0.00	0.00
<i>Vangeria infusta</i>	0.05	0.00	0.00
<i>Pavetta schumanniana</i>	0.01	0.00	0.00

Esto nos indica que en algunas de las muestras se pudo identificar secuencias que mapeaban con las secuencias de las bases de datos de *Burkholderia* para el gen 16s RNA,

gyrB o *recA*, indicando la posible presencia de este microorganismo en las muestras analizadas. Para comprobar la identidad de estos microorganismos se procedió a visualizar la cobertura para un posterior ensamblaje.

4.3 Visualización cobertura en IGV

Luego del mapeo se visualizó la cobertura obtenida utilizando el programa IGV. En la *Tabla 20*, se observa los resultados obtenidos.

Tabla 20 Resultados cobertura en IGV.

Nombre de especie	Gen 16S RNA	Gen <i>gyrB</i>	Gen <i>recA</i>
<i>Tricalysia lasiodelphys</i> (Control positivo)	Buena cobertura	No muy buena cobertura	No muy buena cobertura
<i>Tricalysia semidecidua</i> (Control positivo)	Buena cobertura	No muy buena cobertura	No muy buena cobertura
<i>Coffea arabica</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea canephora</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea eugenioides</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea pseudozangebariae</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea racemosa</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea stenophylla</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea humilis</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea dolichophylla</i> o <i>Coffea perrieri</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea humblotiana</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea ebracteolate</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea ambogensis</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea bissetia</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Coffea bionensis</i>	Mala cobertura	Mala cobertura	Mala cobertura
<i>Vangeria infusta</i>	Buena cobertura	No muy buena cobertura	No muy buena cobertura

<i>Pavetta schumanniana</i>	Buena cobertura	No muy buena cobertura	No muy buena cobertura
-----------------------------	-----------------	------------------------	------------------------

Si visualmente se observó una buena cobertura en la mayoría del gen de interés con las secuencias resultantes del mapeo, se procedió al ensamblaje.

A continuación, se visualizan los resultados con las muestras del género *Tricalysia* spp que fue el control positivo utilizado. En la Figura 37, se visualiza los resultados de cobertura obtenidos con la especie *Tricalysia lasiodelphys* para los tres genes en estudio. En el caso de el gen 16s RNA se visualiza que se tiene una buena cobertura (Figura 37).

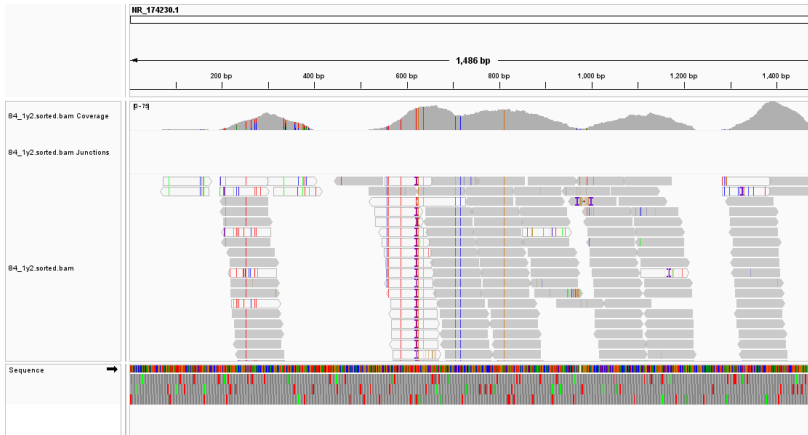


Figura 37 Cobertura obtenida luego del mapeo para el gen 16s RNA:

Para el gen *recA* se visualiza que se tiene una cobertura relativamente buena (Figura 38).

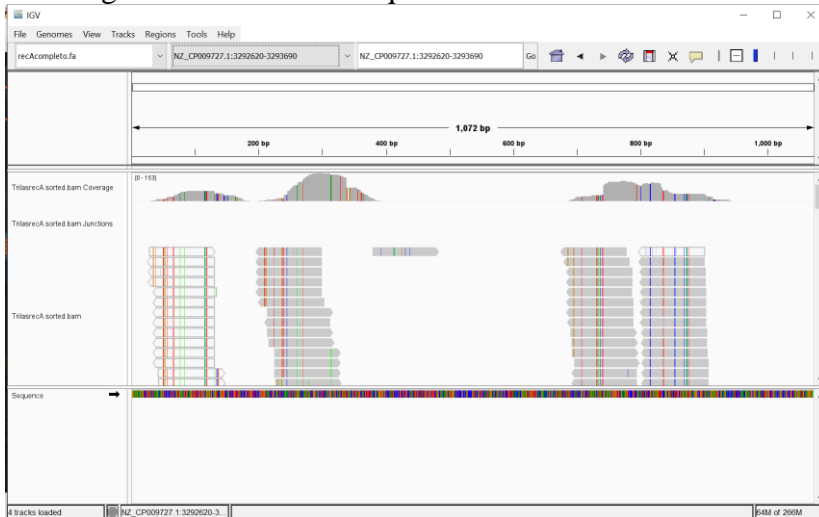


Figura 38 Cobertura obtenida luego del mapeo para el gen *recA*.

Y para el gen *gyrB* la cobertura no es buena, como se visualiza en la Figura 39

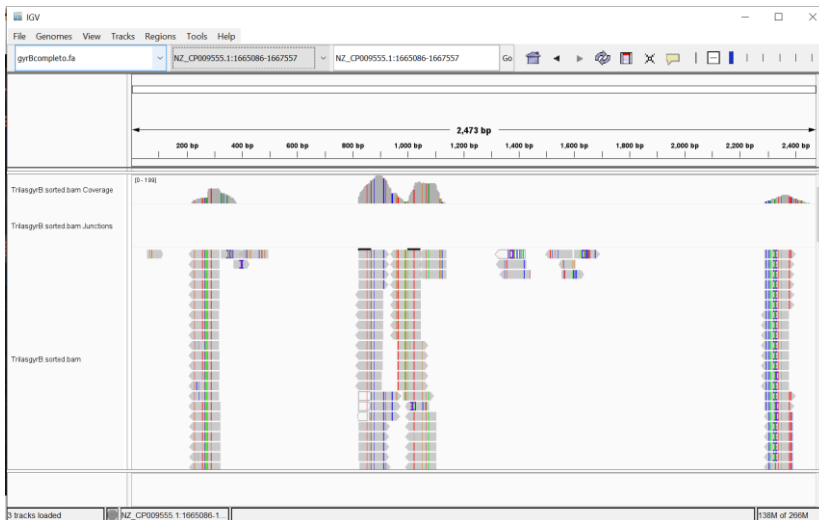


Figura 39 Cobertura obtenida luego del mapeo para el gen *gyrB*

En el caso de *Tricalysia semidesidua*, se visualizan resultados similares al anterior, en el gen 16sRNA se ve una buena cobertura (Figura 40), en el gen *recA* de igual manera (Figura 41) y en el gen *gyrB* una cobertura no tan buena (Figura 42).

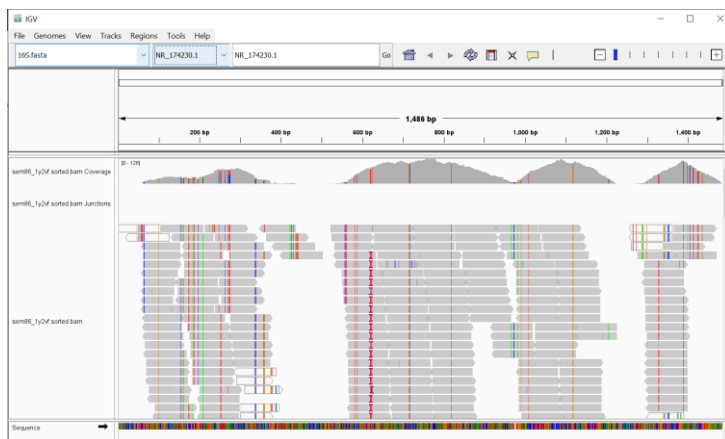


Figura 40 Cobertura obtenida luego del mapeo para el gen 16s RNA en *Tricalysia semidesidua*

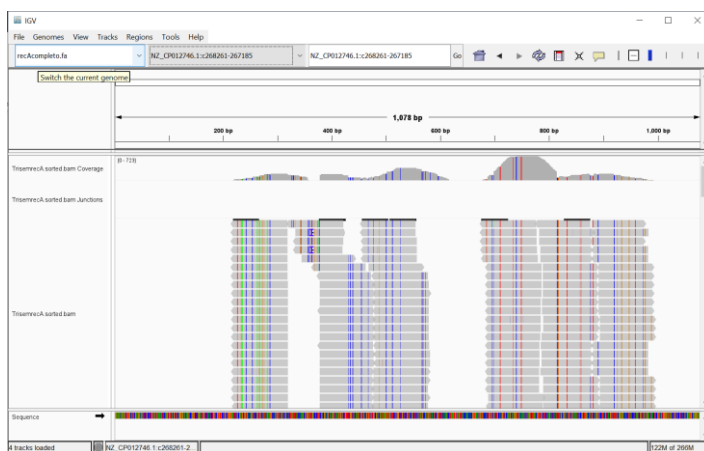


Figura 41 Cobertura obtenida luego del mapeo para el gen *recA* en *Tricalysia semidesidua*

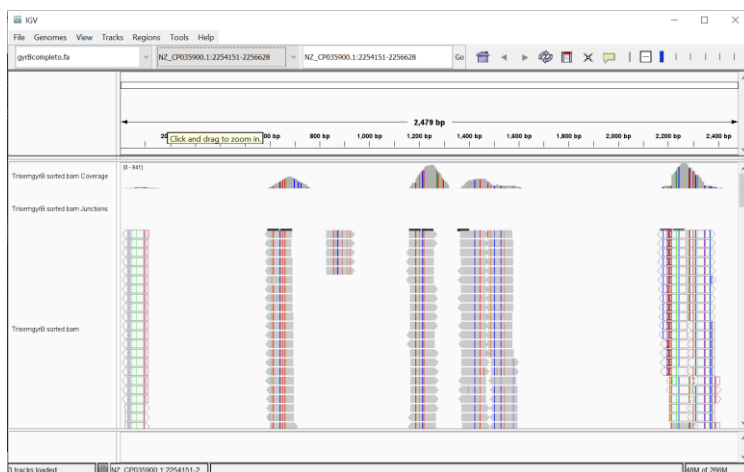


Figura 42 Cobertura obtenida luego del mapeo para el gen *gyrB* en *Tricalysia semidesidua*

Solo con aquellas muestras que se visualizó una buena cobertura, se procedió al proceso de ensamblaje.

4.4 Ensamblaje

Para la evaluación de la calidad del ensamblado, se analizan los valores de los estadísticos; en los cuales se tiene el valor de N50 que define la calidad del ensamblaje en función de la contiguidad, considerando que el valor de pb correspondería a la longitud de secuencia del contig más corto que representa el 50% de la longitud total de la secuencia completa. L50 en cambio define el recuento del menor número de contigs cuya suma de longitud constituye la mitad del tamaño del genoma (Simpson et al., 2009).

A continuación, se presentan los resultados de ensamblaje para cada uno de los genes evaluados:

4.4.1 Gen *16s* RNA

Se realizaron varias pruebas de ensamblaje con valores de k-mers diferentes, en la Tabla 21 se visualizan los estadísticos con los mejores resultados obtenidos. En el caso de las muestras en las cuales no se obtuvo una cobertura adecuada luego del mapeo no se realizó el ensamblaje y no se presentan los resultados en la tabla; como fue en el caso de *Coffea arabica*, *Coffea canephora*, *Coffea eugenioides*, *Coffea pseudozangebariae*, *Coffea racemosa*, *Coffea stenophylla*, *Coffea humilis*, *Coffea dolichophylla* o *Coffea perrieri*, *Coffea humblotiana*, *Coffea ebracteolata*, *Coffea ambogensis*, *Coffea bissetiae* y *Coffea boinensis*.

Tabla 21 Estadísticos obtenidos del proceso de ensamblaje con Abyss con el gen *16S* RNA.

Nombre especie	de	Valor k-mers	N	L50	min	N50	N25	E-size	Max
<i>Tricalysia lasiodelphys</i>		71	2	1	1565	1565	1565	1565	1565

(Control positivo)								
<i>Tricalysia semidecidua</i> (Control positivo)	71	4	1	1662	1662	1662	1662	1662
<i>Vangeria infusta</i>	75	5	1	1616	1616	1616	1616	1616
<i>Pavetta schumanniana</i>	75	8	1	1488	1488	1488	1488	1488

A continuación, se presenta un ejemplo de la visualización de los contigs obtenidos del ensamblaje utilizando BANDAGE (Figura 43) de la especie *Tricalysia lasiodelphys* que corresponde al control positivo, en el cual se observa que uno de los contigs tiene el tamaño de 1500 pb aproximadamente que corresponde al tamaño reportado para el gen *16s* RNA (Faniyan et al., 2023).

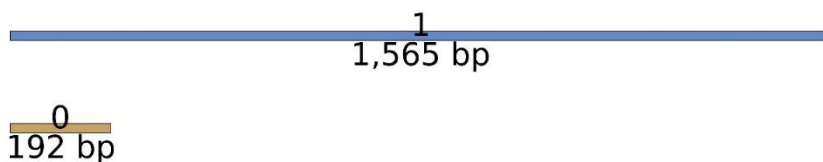


Figura 43 Visualización ensamblaje.con BANDAGE de contigs de la especie *Tricalysia lasiodelphys*.

4.4.2 Gen *gyrB*

Se realizaron varias pruebas de ensamblaje con valores de K-mers diferentes, en la Tabla 22 se visualizan los estadísticos con los mejores resultados obtenidos. En el caso de las muestras en las cuales no se obtuvo una cobertura adecuada luego del mapeo no se realizó el ensamblaje y no se presentan los resultados en la Tabla 22; como fue en el caso de *Coffea arabica*, *Coffea canephora*, *Coffea eugenioides*, *Coffea pseudozangebariae*, *Coffea racemosa*, *Coffea stenophylla*, *Coffea humilis*, *Coffea dolichophylla* o *Coffea perrieri*, *Coffea humblotiana*, *Coffea ebracteolata*, *Coffea ambogensis*, *Coffea bissetiae* y *Coffea boinensis*.

Tabla 22 Estadísticos obtenidos del proceso de ensamblaje con Abyss con el gen *gyrB*.

Nombre de especie	Valor k-mers	N	L50	min	N50	N25	E-size	max
<i>Tricalysia lasiodelphys</i> (Control positivo)	4	1	1	1199	1199	1199	1199	1199

<i>Tricalysia semidecidua</i> (Control positivo)	4	1	1	1169	1169	1169	1169	1169
<i>Vangeria infusta</i>	75	8	0	0	0	0	0	0
<i>Pavetta schumanniana</i>	75	8	0	0	0	0	0	0

4.4.3 Gen *recA*

Se realizaron varias pruebas de ensamblaje con valores de K-mers diferentes, en la se visualizan los estadísticos con los mejores resultados obtenidos. En el caso de las muestras en las cuales no se obtuvo una cobertura adecuada luego del mapeo no se realizó el ensamblaje y no se presentan los resultados en la Tabla 23; como fue en el caso de *Coffea arabica*, *Coffea canephora*, *Coffea eugenioides*, *Coffea pseudozangebariae*, *Coffea racemosa*, *Coffea stenophylla*, *Coffea humilis*, *Coffea dolichophylla* o *Coffea perrieri*, *Coffea humblotiana*, *Coffea ebracteolata*, *Coffea ambogensis*, *Coffea bissetiae* y *Coffea boinensis*.

Tabla 23 Estadísticos obtenidos del proceso de ensamblaje con Abyss con el gen *recA*.

Nombre especie	de	Valor k-mers	N	L50	min	N50	N25	E-size	max
<i>Tricalysia lasiodelphys</i> (Control positivo)		71	1	1	989	989	989	989	989
<i>Tricalysia semidecidua</i> (Control positivo)		71	1	1	1169	1169	1169	1169	1169
<i>Vangeria infusta</i>		75	0	0	0	0	0	0	0
<i>Pavetta schumanniana</i>		75	1	1	531	531	531	531	531

4.5 Taxonomía y análisis filogenético

En este caso se procedió a la identificación taxonómica con las secuencias identificadas, En primer lugar, se comparó las secuencias obtenidas con las bases de datos disponibles a través de BLASTn; para la identificación del microorganismo presente en las secuencias obtenidas de las hojas de las diferentes especies analizadas. No se realiza un árbol

filogenético dado que en el presente estudio entre las especies de *Coffea* no se identificó bacterias de la familia Burkholderiaceae, sin embargo en el artículo científico titulado “Metagenomics of African Empogona and *Tricalysia* (Rubiaceae) reveals the presence of leaf endophytes” de Verstraete1, B, et al, 2020 que se encuentra sometido a peer review en una revista científica se incluye un árbol filogenético de los microorganismos endófitos de los géneros *Empogona* and *Tricalysia*, identificados con la metodología propuesta en esta tesis.

En la Tabla 24, se presentan los resultados obtenidos de la identificación taxonómica de las muestras analizadas:

Tabla 24 Identificación taxonómica de bacterias con BLASTn.

Especie	Gen	Especie	No de acceso	Quer y Cover (%)	E-valor	% de identidad
<i>Tricalysia lasiodelphys</i> (Control positivo)	16s RNA	<i>Caballeronia terrestris</i> strain R-23321	NR_125558.1	96	0.0	97.56
	<i>gyrB</i>	<i>Burkholderia</i> sp. SFA1.	AP022305.1	99	0.0	94.14
	<i>recA</i>	<i>Burkholderia</i> sp. KK1	CP015999.1	99	0.0	94.42
<i>Tricalysia semidecidua</i> (Control positivo)	16s RNA	<i>Caballeronia fortuita</i> strain LMG 29320	NR_145600.1	89	0.0	98.18
	<i>gyrB</i>	<i>Caballeronia zhejiangensis</i> strain A33_M4_a	CP084283.1	99	0.0	93.53
	<i>recA</i>	<i>Burkholderia</i> sp. RPE67	AP014576.1	99	0.0	95.42
<i>Vangeria infusta</i>	16s RNA	<i>Paraburkholderia phenoliruptrix</i> strain AC1100	NR_042901.1	91	0.0	99.66
	<i>gyrB</i>	/	/	/	/	/
	<i>recA</i>	/	/	/	/	/
<i>Pavetta schumanniana</i>	16s RNA	<i>Caballeronia ptereochtonis</i> strain LMG 29326	NR_145596.1	95	0.0	98.25

	<i>gyrB</i>	/	/	/	/	/
	<i>recA</i>	Candidatus <i>Burkholderia</i> <i>schumanniana</i> strain 2001944257	HQ849182.1	93	0.0	100

Con los resultados indicados es posible concluir que el pipeline desarrollado para la identificación de microorganismos endófitos es efectivo; dado que se identificó la presencia de bacterias endófitas de la familia *Burkholderiaceae* con los controles positivos; y también se identificó la presencia de estas, en el género *Vangeria* y *Paveta*. Se identificaron tres géneros principales como son *Burkholderia*, *Caballeronia* y *Paraburkholderia*. Es importante resaltar que las especies de *Caballeronia* forman un clado distintivo entre clados que contienen especies de los géneros *Burkholderia* y *Paraburkholderia*, en árboles filogenéticos basados en secuencias de genes *16S rRNA* (Sawana et al., 2014).

Sin embargo para las muestras de *Vangeria infusta* no se pudo realizar un ensamblaje con los genes *gyrB* y *recA*; y para la muestra de *Pavetta schumanniana*, no fue posible realizar un ensamblaje con el gen *gyrB*; esto puede deberse a que el ADN de la planta es más dominante que el de los microorganismos endófitos, lo que a su vez dificulta el aislamiento y la secuenciación del metagenoma del endófito con una alta cobertura (Fadiji & Babalola, 2020), lo que mejoraría los resultados del ensamblaje. Otra alternativa sería la aplicación de una técnica de secuenciación que produzca fragmentos más largos para facilitar posteriormente el ensamblaje, como la secuenciación Oxford Nanopore.

El género *Burkholderia* se caracteriza por ser polifilético y contener al menos dos grupos filogenéticos, y tres géneros: los primeros corresponden a bacterias patógenas animales y vegetales, el segundo grupo incluye a bacterias ambientales aisladas del suelo y el agua, así como microorganismos asociados a plantas, endófitos y noduladores de leguminosas, denominadas *Paraburkholderia*, y el tercero es el género *Caballeronia* que reagrupa a especies como *Burkholderia glathei*, *Burkholderia grimmiae*, *Burkholderia megalochromosomata*, *Burkholderia jiangsuensis*, *Burkholderia sordicola*, *Burkholderia telluris*, *Burkholderia terrestris*, *Burkhold* (Dobritsa & Samadpour, 2016; Uroz & Oger, 2017)ritsa & Samadpour, 2016; Uroz & Oger, 2017). Sin embargo, actualmente, hay una escasez de características bioquímicas o moleculares que puedan distinguir de manera confiable diferentes grupos de especies de *Burkholderia* (Sawana et al., 2014).

El género *Burkholderia* es conocido como un grupo versátil de bacterias, que son capaces de establecer una asociación estrecha y simbiótica/mutualista con otros organismos, desempeñando un papel biológico crucial, al determinar la supervivencia del huésped; dado que se ha identificado que cuando la planta huésped no se encuentra en asociación con la bacteria, se afecta el crecimiento y desarrollo normal de la planta huésped, lo que sugiere un equilibrio hormonal alterado (Lemaire, Robbrecht, et al., 2011).

Por otro lado, el género *Caballeronia* incluye bacteria gramnegativa, ovoides con forma de varilla, no formadoras de esporas, que se las visualiza microscópicamente solo, en pares o en cadenas; que han sido identificadas en varios estudios de asociación planta-bacteria. Estas bacterias se caracterizan por ser quimio-organotróficas, mesófilas; capaces

de utilizar una amplia gama de compuestos para su crecimiento (Dobritsa & Samadpour, 2016)

La importancia de la identificación de estos géneros (*Burkholderia*, *Paraburkholderia* y *Caballeronia*) en las muestras analizadas, es su relación con el hecho que algunos de estos se han identificado como endófitos, reportándose varios endosimbiontes de nódulos foliares no cultivados que viven en nódulos foliares de un gran número de especies de plantas hospedadoras pertenecientes a los géneros *Ardisia*, *Pavetta*, *Psychotria* y *Sericanthe*, siendo esenciales para un correcto desarrollo y supervivencia del hospedador. Ninguna de estas bacterias, sin embargo, ha sido cultivada in vitro, y por lo tanto algunas se clasifican provisionalmente como especies candidatas en la categoría Candidatus de *Burkholderia*. Actualmente todos estos endosimbiontes de nódulos foliares no cultivados se agrupan filogenéticamente en el género *Caballeronia* (Vandamme y Eberl, 2018).

Estos microorganismos endófitos son capaces de colonizar los tejidos internos de la planta sin inducir síntomas negativos en su huésped. La colonización interna permite que los endófitos se vean menos afectados por las condiciones edáficas y la competencia con otros microorganismos. Los endófitos bacterianos son ahora bien reconocidos como los principales candidatos para mejorar el crecimiento de las plantas a través de la producción de fitohormonas, la reducción del nivel de etileno, la fijación de nitrógeno y el aumento de la disponibilidad de nutrientes. El uso de bacterias endófitas como agentes de biocontrol también es de creciente interés, ya que colonizan el mismo nicho ecológico que varios fitopatógenos, incluidos hongos y nematodos. Los endófitos son capaces de estimular las defensas de las plantas y mostrar propiedades antagónicas hacia los patógenos mediante la producción de diversos compuestos activos como antibióticos, sideróforos, cianuro de hidrógeno y varias (Duong et al., 2021). También, se sabe que varios de estos géneros y especies, se conocen porque mejoran la resistencia a las enfermedades en las plantas, contribuyen a una mejor gestión del agua y mejoran la fijación de nitrógeno y la adaptación general del huésped a las tensiones ambientales; sin embargo el conocimiento sobre los beneficios exactos conferidos por estos endosimbiontes aún es incompleto (Lemaire, Robbrecht, et al., 2011), por ello el interés en este estudio de en primer lugar identificar si existe la presencia de microorganismos endófitos en estas plantas (Compant et al., 2008).

Se han realizado observaciones morfológicas de endosimbiontes bacterianos en todos los géneros rubiáceos. Se sabe que alrededor de 500 especies de *Rubiaceae* albergan endosimbiontes bacterianos dentro de cavidades internas en la lámina de la hoja, conocidos como nódulos bacterianos de las hojas o agallas foliares, la cual es una interacción única e íntima entre las bacterias y las plantas con flores, en la que los endosimbiontes se organizan en estructuras foliares especializadas. Anteriormente, la simbiosis bacteriana de las hojas se ha descrito como una interacción cíclica y obligada en la que los endosimbiontes se transmiten verticalmente entre las generaciones de plantas y carecen de crecimiento autónomo. (Lemaire, Robbrecht, et al., 2011). Este fenómeno ha sido reportado en tres géneros de *Rubiaceae* (*Pavetta*, *Psychotria* y *Sericanthe*) Pero no han sido reportados en otros géneros como *Tricalysia*, sin embargo, en este estudio se identificó su presencia (Yang y Hu, 2018). Adicionalmente no se ha identificado con certeza los microorganismos que estarían en este grupo de endófitos.

4.5.1 Bacterias de la familia Burkholderiaceae en especies del género *Tricalysia*

En este estudio en la especie *Tricalysia lasiodelphys* se identificó que los microorganismos endófitos son bacterias cercanas a *Caballeronia terrestres* cuyo sinónimo homotípico es *Burkholderia terrestris*; y en *Tricalysia semidecidua* se determinó que pueden ser bacterias cercanas a *Caballeronia fortuita*, que tiene como sinónimo homotípico *Caballeronia fortuita*.

Burkholderia terrestris son bacterias gram negativas, que fueron aisladas por primera vez de suelo a granel recolectado de un campo cultivable a largo plazo bajo monocultivo de maíz en 2003 en los Países Bajos (Vandamme et al., 2013)

Caballeronia fortuita cepa LMG 29320, es una bacteria gram negativa, aerobia, mesófila que se ha identificado en suelo de la rizósfera (Peeters et al., 2016).

En otros estudios también se identificó la presencia de microorganismos endófitos como en los resultados reportados por *Guyput, 2020 (estudio no publicado) con Tricalysia congesta*, donde se identificó la presencia del grupo *Burkholderia*, donde obtuvieron que 198/769 scaffolds tenían coincidencia con los genomas de *Burkholderia* con eValue < 10-50.

4.5.2 Bacterias de la familia Burkholderiaceae en especies del género *Vangeria*

Por otro lado, en *Vangeria infusta*, se determinó que probablemente el microorganismo presente incluye bacterias cercanas a *Paraburkholderia phenoliruptrix strain AC1100*.

En cuanto al género *Paraburkholderia* forma parte de *Burkholderia sensu lato*, que incluye patógenos en plantas, animales y el ser humano; sin embargo algunos también se les asocia efectos benéficos mientras se asocian con plantas. Específicamente a *Paraburkholderia phenoliruptrix* se la ha asociado como promotora del crecimiento vegetal y nodulador (Beukes et al., 2021), y se la ha identificado como simbiote de *Mimosa flocculosa* como fijador de nitrógeno, y del grupo *Piptadenia* (*Mimosoideae*) (Zuleta et al., 2014).

4.5.3 Bacterias de la familia Burkholderiaceae en especies del género *Pavetta*

En el género *Pavetta schumanniana* se determinó que pueden ser bacterias cercanas a la bacteria *Caballeronia ptereochtonis* y *Candidatus Paraburkholderia schumanniana*; similar a otros estudios en dos especies de *Pavetta* y una de *Psychotria*, en los cuales con la secuenciación directa del gen 16S rRNA se identificó endosimbiontes nodulantes de hojas, estrechamente relacionados con el género *Burkholderia* y también se descubrió que cada especie huésped está infectada por una sola cepa bacteriana. Con el tiempo, se han hecho varios intentos para aislar, cultivar e identificar los endosimbiontes bacterianos en *Rubiaceae* "bacteriófilas", pero ninguno ha tenido éxito debido a la incultivabilidad de los endosimbiontes; he aquí la importancia de las técnicas de secuenciación; es por ello que en este estudio se ha utilizado los datos de ARNr 16S y marcadores genéticos comunes (*recA* y *gyrB*) para estudiar las relaciones filogenéticas de endosimbiontes bacterianos entre estas especies (Lemaire, Robbrecht, et al., 2011).

Con aproximadamente 350 especies noduladas aceptadas, *Pavetta* contiene el mayor número de especies noduladas, con el 85% de las especies del género que albergan endosimbiontes bacterianos dentro de los nódulos foliares en África continental tropical

y Asia. Se planteó la hipótesis que en *Psychotria* y *Pavetta*, la simbiosis bacteriana de las hojas era obligatoria y cíclica: la presencia del compañero bacteriano es necesaria para la supervivencia del huésped y los endosimbiontes se retienen dentro de la planta huésped durante todas las etapas de su ciclo de vida (Lemaire, van Oevelen, et al., 2011).

Específicamente con referencia a *Caballeronia ptereochthonis* no hay mucha descripción de esta, se conoce que una primera cepa fue aislada de suelo del jardín botánico de Bélgica en 2014 (Peeters et al., 2016). Y en cuanto a Candidatus *Burkholderia schumanniana*, su presencia ha sido reportada en otros estudios como un microorganismo epifito del huésped *Pavetta* (Lemaire, Oevelen, et al., 2012). Mientras que los endosimbiontes nodulantes de hojas en los otros géneros de angiospermas permanecían desconocidos hasta ahora. (Lemaire, van Oevelen, et al., 2011)

4.5.4 Bacterias de la familia *Burkholderiaceae* en especies del género *Coffea*

Luego de la aplicación del pipeline desarrollado, no fue posible demostrar la presencia de *Burkholderia* en las secuencias de hojas del género *Coffea*; por lo que se considera que es posible que las bacterias simbióticas analizadas no estén asociadas a este género a diferencia del género *Tricalysia*; a pesar que si se ha identificado la presencia de *Burkholderia* en otros tejidos del género *Coffea*; es así que en un estudio realizado en Hawai, México y Colombia ha revelado la presencia de una amplia diversidad de endófitos bacterianos en varios tejidos de café, incluida la semilla; donde se encontraron endófitas como *Burkholderia*, *Cedecea*, *Curtobacterium*, *Enterobacter*, *Pantoea* y *Serratia* (Vega et al., 2005). También se han estudiado hongos en algunas especies de café como en *Coffea arabica* en Puerto Rico, donde se han identificado como endófitos a los géneros *Colletotrichum*, *Xylaria*, y *Guignardia* (Santamaría & Bayman, 2005).

Sin embargo, la ausencia de *Burkholderia* puede deberse a otros factores, como el muestreo, dado que no se han secuenciado todas las especies de *Coffea*; y también al enfoque bioinformático utilizado; considerando que se optó por un enfoque directo mediante la búsqueda de genes específicos de *Burckholderia* (16S, *gyrB* y *recA*). En este sentido, siendo un enfoque muy estricto, puede provocar que no se encuentren genes con más variación de nucleótidos utilizando la herramienta de mapeo Bowtie2. En el futuro se recomendaría el uso de una herramienta de búsqueda menos estricta como:

- BMAP es un alineador global, que puede alinear lecturas de todas las plataformas principales: Illumina, 454, Sanger, Ion Torrent, Pac Bio y Nanopore. BMAP es rápido y extremadamente preciso, particularmente con genomas altamente mutados o lecturas con InDeles (inserción, delección, o inserción y delección) largos, incluso delecciones de genes completos de más de 100 kbp de largo. Además, la fase de indexación es muy rápida en comparación con otros alineadores (JGI, 2023).
- BWA: se utiliza para encontrar coincidencias inexactas, a través de un nuevo algoritmo que busca coincidencias entre la 'substring' del genoma de referencia y la secuencia 'query' dentro de una cierta distancia definida (Hatem et al., 2013).

Adicionalmente, se podrían plantear otras herramientas bioinformáticas para el pipeline, como por ejemplo incluir inicialmente un alineamiento con el genoma del café para eliminar secuencias del huésped que serían consideradas como contaminación en este caso; y con las secuencias no alineadas, trabajar en un alineamiento y ensamblaje para la identificación de microorganismos endófitos (Akinsanya et al., 2015).

Por otro lado, un enfoque más abierto de la investigación de microorganismos también sería de interés en el género *Coffea*, con el fin de construir un catálogo de microorganismos asociados a las hojas de café; que también pueda servir de base para futuras aplicaciones de estas, como por ejemplo como biocontroladores, como una alternativa adicional al concepto de manejo integrado, racionalizando el uso de fungicidas (Silva et al., 2012)

Por último, si se desea una mayor precisión en la identificación de los microorganismos de interés, se puede partir de datos obtenidos de una secuenciación de genes marcadores; para lo cual se debe realizar la extracción de ADN genómico de las muestras de tejido de hojas, para una posterior amplificación del gen de interés, por ejemplo el 16s RNA, seguida del proceso de secuenciación. Posteriormente se podría utilizar algoritmos como Kaiju (<https://kaiju.binf.ku.dk>) que interpreta lecturas y hace una comparación entre ellas y una base de datos de proteínas de referencia; luego de haber obtenido una anotación funcional de los microorganismos endófitos de interés (Fadiji & Babalola, 2020).

5 Conclusiones y Recomendaciones

El pipeline generado es útil para identificar la presencia de endófitos de la familia *Burkholderiaceae* en hojas de especies de los géneros *Vangeria* y *Pavetta*, de la familia *Rubiaceae*, a partir de datos NGS.

No se identificó la presencia de bacterias del género *Burkholderia* en el género *Coffea*, lo que se atribuye a un posible no asociación de microorganismos endófitos-hojas de café.

Se recomienda extender el estudio, al ensamblaje del genoma completo de las especies identificadas, para poder realizar una anotación funcional que ayude a identificar la función de estas bacterias endófitas.

Se recomienda un enfoque más abierto de la investigación de microorganismos en el género *Coffea*, con el fin de construir un catálogo de microorganismos asociados a las hojas de café.

6 Referencias

- Ahmad, N., Khalid, S., Hakeem, R., Banaganapalli, B., & Ramu, E. (2019). *Essentials of Bioinformatics, Volume I: Vol. I*. <https://doi.org/https://doi.org/10.1007/978-3-030-02634-9>
- Akaçin, İ., Ersoy, Ş., Doluca, O., & Güngörmüşler, M. (2022). Comparing the significance of the utilization of next generation and third generation sequencing technologies in microbial metagenomics. *Microbiological Research*, *264*, 1–14. <https://doi.org/10.1016/j.micres.2022.127154>
- Akinsanya, M. A., Goh, J. K., Lim, S. P., & Ting, A. S. Y. (2015). Metagenomics study of endophytic bacteria in Aloe vera using next-generation technology. *Genomics Data*, *6*, 159–163. <https://doi.org/10.1016/J.GDATA.2015.09.004>
- Anthony, F., Bertrand, B., Etienne, H., & Lashermes, P. (2011). Coffea and Psilanthus. In *Wild Crop Relatives: Genomic and Breeding Resources* (pp. 41–61). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21201-7_3
- Babraham Institute. (2022). *FastQC A Quality Control tool for High Throughput Sequence Data*. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Balzer, S., Malde, K., Lanzén, A., Sharma, A., & Jonassen, I. (2010). Characteristics of 454 pyrosequencing data-enabling realistic simulation with flowsim. *Bioinformatics*, *27*(13), i420–i422. <https://doi.org/10.1093/bioinformatics/btq365>
- Beukes, C. W., Venter, S. N., & Steenkamp, E. T. (2021). The history and distribution of nodulating Paraburkholderia, a potential inoculum for Fynbos forage species. *Grass and Forage Science*, *76*(1), 10–32. <https://doi.org/10.1111/GFS.12522>
- Bussmann, R. W., Paniagua, N. Y., & Njoroge, G. N. (2021). Coffea arabica L. Rubiaceae. In *Etnobotany of the Mountain Regions of Africa* (pp. 313–318). https://doi.org/10.1007/978-3-030-38386-2_45
- Caballero-Mellado, J., Martínez-Aguilar, L., Paredes-Valdez, G., & Estrada-de los Santos, P. (2004). Burkholderia unamae sp. nov., an N₂-fixing rhizospheric and endophytic species. *International Journal of Systematic and Evolutionary Microbiology*, *54*(4), 1165–1172. <https://doi.org/10.1099/ij.s.0.02951-0>
- Cadena, J., Martínez, M., Guzmán, L., & Arteaga, G. (2016). *Aplicación de secuenciación masiva para el estudio y exploración de diversidad microbiana y su aprovechamiento biotecnológico*. *9*(2), 70–80.
- Caldwell, A. C., Silva, L. C. F., da Silva, C. C., & Ouverney, C. C. (2015). Prokaryotic diversity in the rhizosphere of organic, intensive, and transitional coffee farms in Brazil. *PLoS ONE*, *10*(6). <https://doi.org/10.1371/journal.pone.0106355>
- Campuzano, L., Herrera, J., Ged, C., & Blair, M. (2021). Bases for the establishment of Robusta Coffee (Coffea canephora) as a new crop for Colombia. In *Agronomy* (Vol. 11, Issue 12, pp. 1–11). MDPI. <https://doi.org/10.3390/agronomy11122550>
- Castillo, J., Prado, E., Rincón, V., Torres, T., García de los Santos, G., & Ramos, V. (2020). Quantitative comparison of three main metabolites in leaves of Coffea accessions by UPLC-MS/MS. *European Food Research and Technology*, *247*(2), 375–384. <https://doi.org/10.1007/s00217-020-03631-8>
- Chagoyen, M. (2013). *Introductory course on Multiple Sequence Alignment* (pp. 1–30).
- Charr, J. C., Garavito, A., Guyeux, C., Crouzillat, D., Descombes, P., Fournier, C., Ly, S. N., Raharimalala, E. N., Rakotomalala, J. J., Stoffelen, P., Janssens, S., Hamon, P., & Guyot, R. (2020). Complex evolutionary history of coffees revealed by full plastid genomes and 28,800 nuclear SNP analyses, with particular emphasis on Coffea canephora (Robusta coffee). *Molecular Phylogenetics and Evolution*, *151*, 1–13. <https://doi.org/10.1016/j.ympev.2020.106906>

- Charr, J., Garavito, A., Guyeux, C., Crouzillat, D., Descombes, P., Fournier, C., Ly, S. N., Raharimalala, E. N., Rakotomalala, J. J., Stoffelen, P., Janssens, S., Hamon, P., & Guyot, R. (2020). Complex evolutionary history of coffees revealed by full plastid genomes and 28,800 nuclear SNP analyses, with particular emphasis on *Coffea canephora* (Robusta coffee). *Molecular Phylogenetics and Evolution*, *151*, 1–13. <https://doi.org/10.1016/j.ympev.2020.106906>
- Chawla, A., Kumar, A., Warghat, A., Singh, S., Bhushan, S., Sharma, R. K., Bhattacharya, A., & Kumar, S. (2020). Approaches for conservation and improvement of Himalayan plant genetic resources. In *Advancement in Crop Improvement Techniques* (pp. 297–317). Elsevier. <https://doi.org/10.1016/b978-0-12-818581-0.00018-8>
- Chen, X. (2019). A review on coffee leaves: Phytochemicals, bioactivities and applications. In *Critical Reviews in Food Science and Nutrition* (Vol. 59, Issue 6, pp. 1008–1025). Taylor and Francis Inc. <https://doi.org/10.1080/10408398.2018.1546667>
- Chouvarine, P., Wiehlmann, L., Losada, P. M., DeLuca, D. S., & Tümmeler, B. (2016). Filtration and Normalization of Sequencing Read Data in Whole-Metagenome Shotgun Samples. *PLOS ONE*, *11*(10), 1–16. <https://doi.org/10.1371/JOURNAL.PONE.0165015>
- Ciuffreda, L., Rodríguez-Pérez, H., & Flores, C. (2021). Nanopore sequencing and its application to the study of microbial communities. *Computational and Structural Biotechnology Journal*, *19*, 1497–1511. <https://doi.org/10.1016/J.CSBJ.2021.02.020>
- Compant, S., Nowak, J., Coenye, T., Clément, C., & Ait Barka, E. (2008). Diversity and occurrence of Burkholderia spp. in the natural environment. *FEMS Microbiology Reviews*, *32*(4), 607–626. <https://doi.org/10.1111/J.1574-6976.2008.00113.X>
- Couturon, E., Lashermes, P., & Charrier, A. (1998). *First intergeneric hybrids (Psilanthus ebracteolatus Hiern × Coffea arabica L.) in coffee trees.*
- Darbyshire, I., Timberlake, J., Osborne, J., Rokni, S., Matimele, H., Langa, C., Datizua, C., de Sousa, C., Alves, T., Massingue, A., Hadj-Hammou, J., Dhanda, S., Shah, T., & Wursten, B. (2019). The endemic plants of Mozambique: diversity and conservation status. *PhytoKeys* *136*: 45-96, *136*(2019), 45–96. <https://doi.org/10.3897/PHYTOKEYS.136.39020>
- Davis, A., Chadburn, H., Moat, J., O’Sullivan, R., Hargreaves, S., & Lughadha, E. (2019). High extinction risk for wild coffee species and implications for coffee sector sustainability. *Science Advances*, *5*(1), 1–9. https://doi.org/10.1126/SCIADV.AAV3473/SUPPL_FILE/AAV3473_SM.PDF
- Davis, A., Gargiulo, R., Almeida, I., Caravela, M., Denison, C., & Moat, J. (2021). Hot Coffee: The Identity, Climate Profiles, Agronomy, and Beverage Characteristics of *Coffea racemosa* and *C. zanguebariae*. *Frontiers in Sustainable Food Systems*, *5*, 1–13. <https://doi.org/10.3389/fsufs.2021.740137>
- Davis, A. P., Gargiulo, R., Almeida, I. N. das M., Caravela, M. I., Denison, C., & Moat, J. (2021). Hot Coffee: The Identity, Climate Profiles, Agronomy, and Beverage Characteristics of *Coffea racemosa* and *C. zanguebariae*. *Frontiers in Sustainable Food Systems*, *5*, 1–14. <https://doi.org/10.3389/fsufs.2021.740137>
- Davis, A., & Rakotonasolo, F. (2008). A taxonomic revision of the baracoffea alliance: nine remarkable *Coffea* species from western Madagascar. *Botanical Journal*, *158*(3), 355–390. <https://doi.org/10.1111/J.1095-8339.2008.00936.X>
- Davis, A., Rakotonasolo, F., & Yamanaka, M. (2020). *Coffea Perrieri Rubiaceae.*

- de Block, P., Rakotonasolo, F., Vrijdaghs, A., & Dessein, S. (2020). Two new species of phialiphora (Spermacoaceae, rubiaceae) exemplify drought adaptations in western madagascar. *Plant Ecology and Evolution*, 153(2), 267–269. <https://doi.org/10.5091/plecevo.2020.1675>
- de Sousa, L., Filho, O., & Costa, J. (2021). Differences between the leaf mycobiome of Coffea arabica and wild Coffee species and their modulation by caffeine chlorogenic acid content. *Microorganisms*, 9(11), 1–9. <https://doi.org/10.3390/microorganisms9112296>
- de Sousa, L. P., da Silva, M. J., & Costa, J. M. (2018). Leaf-associated bacterial microbiota of coffee and its correlation with manganese and calcium levels on leaves. *Genetics and Molecular Biology*, 41(2), 455–465. <https://doi.org/10.1590/1678-4685-gmb-2017-0255>
- de Sousa, L. P., Guerreiro, O., & Costa, J. (2022). The Rhizosphere Microbiomes of Five Species of Coffee Trees. *Microbiology Spectrum*, 10(2), 1–12. <https://doi.org/10.1128/spectrum.00444-22>
- de Sousa, L. P., Guerreiro, O., & Mondego, J. M. (2022). The Rhizosphere Microbiomes of Five Species of Coffee Trees. *Microbiology Spectrum*, 10(2). https://doi.org/10.1128/SPECTRUM.00444-22/SUPPL_FILE/REVIEWER-COMMENTS.PDF
- Dobritsa, A. P., & Samadpour, M. (2016). Transfer of eleven species of the genus Burkholderia to the genus Paraburkholderia and proposal of Caballeronia gen. nov. to accommodate twelve species of the genera Burkholderia and Paraburkholderia. *International Journal of Systematic and Evolutionary Microbiology*, 66(8), 2836–2846. <https://doi.org/https://doi.org/10.1099/ijsem.0.001065>
- Duong, B., Marraccini, P., Maeght, J. L., Vaast, P., Lebrun, M., & Duponnois, R. (2020a). Coffee Microbiota and Its Potential Use in Sustainable Crop Management. A Review. *Frontiers in Sustainable Food Systems*, 4, 1–24. <https://doi.org/10.3389/FSUFS.2020.607935/BIBTEX>
- Duong, B., Marraccini, P., Maeght, J. L., Vaast, P., Lebrun, M., & Duponnois, R. (2020b). Coffee Microbiota and Its Potential Use in Sustainable Crop Management. A Review. In *Frontiers in Sustainable Food Systems* (Vol. 4, pp. 1–31). Frontiers Media S.A. <https://doi.org/10.3389/fsufs.2020.607935>
- Duong, B., Nguyen, H. X., Phan, H. V., Colella, S., Trinh, P. Q., Hoang, G. T., Nguyen, T. T., Marraccini, P., Lebrun, M., & Duponnois, R. (2021). Identification and characterization of Vietnamese coffee bacterial endophytes displaying in vitro antifungal and nematicidal activities. *Microbiological Research*, 242, 1–13. <https://doi.org/10.1016/J.MICRES.2020.126613>
- Edgar, R. (2004). *MUSCLE User Guide*. <http://www.drive5.com/muscle/muscle.html>
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/NAR/GKH340>
- Ekalu, A. (2021). Medicinal uses, phytochemistry, and pharmacological activities of Mitracarpus species (Rubiaceae): A review. In *Scientific African* (Vol. 11, pp. 2–11). Elsevier B.V. <https://doi.org/10.1016/j.sciaf.2020.e00692>
- Espinosa, D., López, L., Carcaño, M. G., & Serret, M. (2020). The Burkholderia genus: between mutualism and pathogenicity. *Revista Mexicana de Fitopatología*, 38(3), 337–354. <https://doi.org/10.18781/R.MEX.FIT.2004-5>
- Fadiji, A. E., & Babalola, O. O. (2020). Metagenomics methods for the study of plant-associated microbial communities: A review. *Journal of Microbiological Methods*, 170, 105860. <https://doi.org/10.1016/J.MIMET.2020.105860>

- Faniyan, O., Akpe, V., & Cock, I. E. (2023). Analyzing Bacterial Species from Different Environments Using Direct 16S rRNA Gene Sequencing Methods. *Pharmacognosy Communications*, 13(1), 24–33. <https://doi.org/10.5530/pc.2023.1.5>
- Ferrão, R., Almeida da Fonseca, A., Gava, M., & de Muner, L. (2019). *Conilon Coffee* (INCAPER). <https://biblioteca.incaper.es.gov.br/digital/bitstream/123456789/3540/1/chapter-04-origin-geographical-dispersion-taxonomy-genetic-diversity.pdf>
- Ferreira, T., Shuler, J., Guimarães, R., & Farah, A. (2019). Introduction to Coffee Plant and Genetics. In *Coffee* (pp. 1–25). Royal Society of Chemistry. <https://doi.org/10.1039/9781782622437-00001>
- Figueiredo, E. (2008). The Rubiaceae of Angola. *Botanical Journal of the Linnean Society*, 156(4), 537–638. <https://doi.org/10.1111/J.1095-8339.2007.00750.X>
- Fulthorpe, R., Martin, A. R., & Isaac, M. E. (2020). Root endophytes of coffee (*Coffea arabica*): Variation across climatic gradients and relationships with functional traits. *Phytobiomes Journal*, 4(1), 27–36. <https://doi.org/10.1094/PBIOMES-04-19-0021-R/ASSET/IMAGES/LARGE/PBIOMES-04-19-0021-RF3.JPEG>
- Fulthorpe, R., Martin, A. R., & Marney, Isacc. (2020). Root endophytes of coffee (*Coffea arabica*): Variation across climatic gradients and relationships with functional traits. *Phytobiomes Journal*, 4(1), 27–39. <https://doi.org/10.1094/PBIOMES-04-19-0021-R>
- Garibyan, L., & Avashia, N. (2013). Research Techniques Made Simple: Polymerase Chain Reaction (PCR). *The Journal of Investigative Dermatology*, 133(3), e6. <https://doi.org/10.1038/JID.2013.1>
- Garrity, G. M., Bell, J. A., & Lilburn, T. (2015). Burkholderiaceae fam. nov. . In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1–1). Wiley. <https://doi.org/10.1002/9781118960608.fbm00181>
- GBIF. (2022). <https://www.gbif.org/fr/>
- Glen, H. (2007). *Pavetta schumanniana* | *PlantZAfrica*. <https://pza.sanbi.org/pavetta-schumanniana>
- Guyot, R., Hamon, P., Couturon, E., Raharimalala, N., Rakotomalala, J.-J., Lakkanna, S., Sabatier, S., Affouard, A., & Bonnet, P. (2020). *WCSDb: a database of wild Coffea species*. http://publish.plantnet-project.org/project/wildcofdb_en
- Hamon, P., Rakotomalala, J. J., Akaffou, S., Razafinarivo, N. J., Couturon, E., Guyot, R., Crouzillat, D., Hamon, S., & de Kochko, A. (2015). Caffeine-free Species in the Genus *Coffea*. In *Coffee in Health and Disease Prevention* (pp. 39–44). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-409517-5.00005-X>
- Hatem, A., Bozdağ, D., Toland, A. E., & Çatalyürek, Ü. v. (2013). Benchmarking short sequence mapping tools. *BMC Bioinformatics*, 14(1), 1–25. <https://doi.org/10.1186/1471-2105-14-184/TABLES/4>
- Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, 82(11), 801–811. <https://doi.org/10.1016/j.humimm.2021.02.012>
- Hue, T. (2017). *Genomic studies of biochemical compounds determining arabica coffee (Coffea arabica L.) quality* Centre for Nutrition and Food Sciences.
- Hyde, M. A., Wursten, B. T., Ballings, P., & Coates Palgrave, M. (2023). *Flora of Zimbabwe: Species information: Pavetta schumanniana*. https://www.zimbabweflora.co.zw/speciesdata/species.php?species_id=156360
- Integrative Genomics Viewer. (2022). *Viewing Alignments | Integrative Genomics Viewer*. <https://software.broadinstitute.org/software/igv/AlignmentData>

- JGI. (2023). *BBMap Guide - DOE Joint Genome Institute*. <https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbmap-guide/>
- Jongman, M., Carmichael, P. C., & Bill, M. (2020). Technological Advances in Phytopathogen Detection and Metagenome Profiling Techniques. In *Current Microbiology* (Vol. 77, Issue 4, pp. 675–681). Springer. <https://doi.org/10.1007/s00284-020-01881-z>
- Junior, V. M., Shigueoka, L. H., Pereira, C. T. M., Carducci, F. C., Sera, T., & Sera, G. H. (2022). Resistance to frost in Arabica coffee lines introgressed with *Coffea racemosa* Lour. genes. *Australian Journal of Crop Science*, 16(3), 338–342. <https://doi.org/10.21475/ajcs.22.16.03.p2925>
- Karani, S., Njuguna, J., Runo, S., Muchugi, A., Machua, A., & Mwaniki, P. (2022). Molecular and morphological identification of fungi causing canker and dieback diseases on *Vangueria infausta* (Burch) subsp. *rotundata* (Robyns) and *Berchemia discolor* (Klotzsch) Hemsl in lower Eastern Kenya. *African Journal of Biotechnology*, 21(1), 6–8. <https://doi.org/10.5897/ajb2020.17297>
- Kewscience. (2022). *World Checklist of Selected Plant Families: Royal Botanic Gardens, Kew*. https://wcsp.science.kew.org/namedetail.do?name_id=207877
- Koutouleas, A., Sarzynski, T., Bordeaux, M., Bosselmann, A., Campa, C., Etienne, H., García, N., Rigal, C., Vaast, P., Ramalho, J., Marraccini, P., & Ræbild, A. (2022). Shaded-Coffee: A Nature-Based Strategy for Coffee Production Under Climate Change? A Review. *Frontiers in Sustainable Food Systems*, 6, 1–14. <https://doi.org/10.3389/FSUFS.2022.877476/BIBTEX>
- Lemaire, B., Oevelen, S. van, de Block, P., Verstraete, B., Smets, E., Prinsen, E., & Dessein, S. (2012). Identification of the bacterial endosymbionts in leaf nodules of *Pavetta* (Rubiaceae). *Burkholderia Schumanniana*, 62, 202–209. <https://doi.org/10.1099/ijcs.0.028019-0>
- Lemaire, B., Robbrecht, E., van Wyk, B., van Oevelen, S., Verstraete, B., Prinsen, E., Smets, E., & Dessein, S. (2011). Identification, origin, and evolution of leaf nodulating symbionts of *Sericanthe* (Rubiaceae). *The Journal of Microbiology* 2011 49:6, 49(6), 935–941. <https://doi.org/10.1007/S12275-011-1163-5>
- Lemaire, B., van Oevelen, S., de Block, P., Verstraete, B., Smets, E., Prinsen, E., & Dessein, S. (2011). Identification of the bacterial endosymbionts in leaf nodules of *Pavetta* (Rubiaceae). *International Journal of Systematic and Evolutionary Microbiology*, 62(1), 202–209. <https://doi.org/10.1099/IJS.0.028019-0/CITE/REFWORKS>
- Lemaire, B., van Oevelen, S., de Block, P., Verstraete, B., Smets, E., Prinsen, E., & Dessein, S. (2012). Identification of the bacterial endosymbionts in leaf nodules of *Pavetta* (Rubiaceae). *International Journal of Systematic and Evolutionary Microbiology*, 62(1), 202–209. <https://doi.org/10.1099/IJS.0.028019-0/CITE/REFWORKS>
- Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, Mareuil, F., Cohen-Boulakia, S., & Gascuel, O. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Research*, 47. https://watermark.silverchair.com/gkz303.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAsEwggK9BgkqhkiG9w0BBwagggKuMIICqgIBADCCAqMGCSqGSIB3DQEHATAeBgIghkgBZQMEAS4wEQQM9LCv8njC4H_WYjKOAqEQgIICdAcfDW6yGUswNVo_Hqtz8q9QirOYQLxD-xtRUancrJgexXeAbkXutmCBhlhB_txYJb0_WYS8-PiTlhDtcpHxjGCAt8RXmpvKtIq3i4I1y14iHHG8oL45ciokqGIg0h85xg7ajPJ3oR3x96YhS9s4Ok7c0rTTw-

- rPfa7Ek9RTvilf1cSpTD_5OKiYuDtAXP_1GmZQySdfMwp0UcIT5ZDuHxZtSJ
Ay6JC7Z4o59n-
rQ0pdwwZWmIcxKTjlNqfirYMzrPrknvLdVRqb37880K_0Xi0GGXFFQKVPB9c
_nP-KmYWKVD4F8PhziIrAiLdn2f-
eNCxtDboScfdgwRuUUkqscJFOxCzubNAyME6ezfIzCMnun4ZJJjXb2VR1WLb
_xBFXoGF-
hfu5PeadMNP3wIQPsne8nlZ62v1ZnpTVvxbzyffelPeH4HIVtFSN_N7Dr6y7fU_q
tmVcj9zsO9ULrB2rRJOIdGmQFlfET433rAXQwlp5iMM5eGCmOsLbT9Mi0GeL
HpxAcnGem2XTFYCK4UMWhOYwZLRScGMBTQIxNRsm6203oI17sbz8jiP00
T5mVu6BoC76XCrTEExwElisynw4sG233KrtDCWBAPAhgTB1uBkO3pAFzYZjx
vSoMcMOjDFFRjcf8E4RRLS6YYXLfcHE0hFfilOQ07xAhJmV18ipo1_9wHYG
OFhAKuGBfooM3PYDE_wVt2QAoBsCbDjfeYViDKuH-
wBYuuy_ss_xvtwqB2TffQxMVBtJEFKi27XvszOTBAQ3HE13SZqy9B8PUXm
CE4hCIPrQzvsGsqs5KAXoI0rfH4S9d99TI7qDjgy1m0L5kdsORA
- Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., & Gascuel, O. (2019). NGPhylogeny.fr: New generation phylogenetic services for non-specialists. *Nucleic Acids Research*, 47(W1), W260–W265. <https://doi.org/10.1093/NAR/GKZ303>
- Liu, B., Gibbons, T., Ghodsi, M., Treangen, T., & Pop, M. (2010). Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genomics*, 12(Suppl 2), 1–10. <https://doi.org/10.1186/1471-2164-12-S2-S4>
- Lucaciu, R., Pelikan, C., Gerner, S., Zioutis, C., Kostibacher, S., Marx, H., Herbold, C., Schmidt, H., & Rattei, T. (2019). A Bioinformatics Guide to Plant Microbiome Analysis. *Frontiers in Plant Science*, 1–18.
- Ly, S. N., Garavito, A., de Block, P., Asselman, P., Guyeux, C., Charr, J. C., Janssens, S., Mouly, A., Hamon, P., & Guyot, R. (2020). Chloroplast genomes of Rubiaceae: Comparative genomics and molecular phylogeny in subfamily Ixoroideae. *PLoS ONE*, 15(4). <https://doi.org/10.1371/journal.pone.0232295>
- Maroyi, A. (2018). Nutraceutical and Ethnopharmacological Properties of *Vangueria infausta* subsp. *infausta*. *Molecules: A Journal of Synthetic Chemistry and Natural Product Chemistry*, 23(5). <https://doi.org/10.3390/MOLECULES23051089>
- Melese, Y., & Kolech, S. (2021). Coffee (*Coffea arabica* L.): Methods, objectives, and future strategies of breeding in Ethiopia Review. In *Sustainability* (Vol. 13, Issue 19, pp. 1–17). MDPI. <https://doi.org/10.3390/su131910814>
- Mewaba, S. G., Happi, E. N., Nangmou, B. M. N., Langat, M. K., Siddique, H., Sadgrove, N., Kamdem, A. F. W., Vardamides, J. C., Mas-Claret, E., Isyaka, S. M., Rotich, W., & Azebaze, B. G. A. (2022). Antibacterial compounds of the cultivated Robusta coffee: *Coffea canephora* Pierre ex A. Froehner. *Scientific African*, 16, 1–4. <https://doi.org/10.1016/j.sciaf.2022.e01274>
- Monteiro, Â., Colomban, S., Azinheira, H. G., Guerra-Guimarães, L., Silva, M. D. C., Navarini, L., & Resmini, M. (2019). Dietary antioxidants in coffee leaves: Impact of botanical origin and maturity on chlorogenic acids and xanthenes. *Antioxidants*, 9(1), 1–16. <https://doi.org/10.3390/antiox9010006>
- Mushtaq, S., Shafiq, M., Ashraf, T., Haider, M. S., Atta, S., Almaary, K. S., & Elshikh, M. S. (2022). Enumeration of citrus endophytic bacterial communities based on illumine metagenomics technique. *PLOS ONE*, 17(4 April), 1–10. <https://doi.org/10.1371/journal.pone.0263144>
- Nab, C., & Maslin, M. (2020). Life cycle assessment synthesis of the carbon footprint of Arabica coffee: Case study of Brazil and Vietnam conventional and sustainable

- coffee production and export to the United Kingdom. *Geo: Geography and Environment*, 7(2), 1–16. <https://doi.org/10.1002/geo2.96>
- Naga, M., & García, L. (2021). Innovations in Biotechnology for a Sustainable Future. In *Innovations in Biotechnology for a Sustainable Future*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-80108-3>
- NCBI. (2023). *Quick Start Guide — BlastQuickStartGuide 0.1.1 documentation*. <https://blast.ncbi.nlm.nih.gov/doc/blast-quick-start-guide/>
- NGPhylogeny. (2022). *NGPhylogeny.fr*. <https://ngphylogeny.fr/documentation>
- Nguyen, Y. T. K., Moon, J. Y., Ryu, J. Y., Eum, S., Bach, T. T., & Cho, S. K. (2019). Methanol Extract of Aerial Parts of *Pavetta indica* L. Enhances the cytotoxic effect of doxorubicin and induces radiation sensitization in MDA-MB-231 triple-negative breast cancer cells. *Molecules*, 24(12), 1–2. <https://doi.org/10.3390/molecules24122273>
- Onana, J. (2013). *Sinopsis de especies de plantas vasculares endémicas y raras en Camerún: lista de verificación para la gestión sostenible y la conservación de la biodiversidad, Yaundé, Ministerio de Investigación Científica e Innovación, coll. "Flora de Camerún."* 40, 192.
- Orozco, M., & Santoyo, G. (2021). Plant-microbial endophytes interactions: Scrutinizing their beneficial mechanisms from genomic explorations. *Current Plant Biology*, 25, 1–7. <https://doi.org/10.1016/J.CPB.2020.100189>
- Patay, É. B., Bencsik, T., & Papp, N. (2016). Phytochemical overview and medicinal importance of *Coffea* species from the past until now. *Asian Pacific Journal of Tropical Medicine*, 9(12), 1127–1133. <https://doi.org/10.1016/j.apjtm.2016.11.008>
- Peeters, C., Meier-Kolthoff, J. P., Verheyde, B., de Brandt, E., Cooper, V. S., & Vandamme, P. (2016). Phylogenomic Study of Burkholderia glathei-like Organisms, Proposal of 13 Novel Burkholderia Species and Emended Descriptions of Burkholderia sordidicola, Burkholderia zhejiangensis, and Burkholderia grimmiae. *Frontiers in Microbiology*, 7(JUN). <https://doi.org/10.3389/FMICB.2016.00877>
- Perera, D. S. M., de Silva, R. C. L., Nayanajith, L. D. C., Colombage, H. C. D. P., Suresh, T. S., Abeysekera, W. P. K. M., & Kottegoda, I. R. M. (2021). Anti-Inflammatory and Antioxidant Properties of *Coffea Arabica*/Reduced Graphene Oxide Nanocomposite prepared by green synthesis. *Material Science Research India*, 18(3), 305–317. <https://doi.org/10.13005/msri/180306>
- Pinto, M., Gademann, K., Eberl, L., & Carlier, A. (2018). Leaf nodule symbiosis: function and transmission of obligate bacterial endophytes. *Current Opinion in Plant Biology*, 44, 23–24. <https://doi.org/10.1016/J.PBI.2018.01.001>
- Pinto, R., Cardoso, T., Paiva, L., & Benedito, V. (2021). Genomic and transcriptomic inventory of membrane transporters in coffee: Exploring molecular mechanisms of metabolite accumulation. *Plant Science*, 312, 1–9. <https://doi.org/10.1016/j.plantsci.2021.111018>
- PlantNet. (2022). *Especies de Coffea*. http://publish.plantnet-project.org/project/wildcofdb_en/collection/coffee-species/species/details/C.%20arabica%20L.
- Prasad, K., & Bisht, G. (2011). Evaluation of Nutritive, Antioxidant and Mineral Composition of *Pavetta indica*. *Phytochemistry*, 54–55.
- Preedy, V. R. (2014). *Coffee in Health and Disease Prevention*. Elsevier Science. <https://books.google.com.ec/books?id=cFECBAAQBAJ>

- Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, 26(7), 1641–1650. <https://doi.org/10.1093/MOLBEV/MSP077>
- Quattrocchi, U. (2016). *CRC World Dictionary of Medicinal and Poisonous Plants: Common Names, Scientific Names, Eponyms, Synonyms, and Etymology (5 Volume Set)*. CRC Press. <https://books.google.com.ec/books?id=-37OBQAAQBAJ>
- Raharimalala, N., Rombauts, S., McCarthy, A., Garavito, A., Orozco-Arias, S., Bellanger, L., Morales-Correa, A. Y., Froger, S., Michaux, S., Berry, V., Metairon, S., Fournier, C., Lepelley, M., Mueller, L., Couturon, E., Hamon, P., Rakotomalala, J. J., Descombes, P., Guyot, R., & Cruzillat, D. (2021). The absence of the caffeine synthase gene is involved in the naturally decaffeinated status of *Coffea humblotiana*, a wild species from Comoro archipelago. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-87419-0>
- Rajesh, A., Sathasivampillai, S., & Rajamanoharan, P. (2021). Medicinal Values of *Pavetta indica* L. Extracts. *Sabuncuoglu Serefeddin Health Science (SSHS)*, 3(1), 1–12.
- Raoul, E. (1894). *Culture du caféier, semis, plantation, taille, cueillette, dépulpation, décorticage, expédition, commerce, espèces et races*. <https://books.google.com.ec/books?id=XBYbAAAAYAAJ>
- Reynolds, S. T. (1993a). *The genus Pavetta L. (Rubiaceae) in Australia* (Vol. 4, Issue 1).
- Reynolds, S. T. (1993b). *The genus Pavetta L. (Rubiaceae) in Australia* (Vol. 4, Issue 1).
- Rodrigues, L. M. R., Destéfano, S. A. L., Almeida, I. M. G. de, Beriam, L. O. S., Braghini, M. T., & Guerreiro Filho, O. (2019). Multiple resistance to bacterial halo blight and bacterial leaf spot in *Coffea* spp.*. *Arquivos Do Instituto Biológico*, 86, 1–9. <https://doi.org/10.1590/1808-1657000632018>
- Rubio, S., Pacheco, R. A., Gómez, A. M., Perdomo, S., & García, R. (2020). Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica. *Universitas Médica*, 61(2), 2–10. <https://doi.org/10.11144/javeriana.umed61-2.sngs>
- Santamaría, J., & Bayman, P. (2005). Fungal epiphytes and endophytes of coffee leaves (*Coffea arabica*). *Microbial Ecology*, 50(1), 1–8. <https://doi.org/10.1007/S00248-004-0002-1/METRICS>
- Sawana, A., Adeolu, M., & Gupta, R. S. (2014). Molecular signatures and phylogenomic analysis of the genus burkholderia: Proposal for division of this genus into the emended genus burkholderia containing pathogenic organisms and a new genus paraburkholderia gen. nov. harboring environmental species. *Frontiers in Genetics*, 5(NOV), 429. <https://doi.org/10.3389/FGENE.2014.00429/ABSTRACT>
- Shiomi, H., Silva, H., de Melo, I., Nunes, F., & Bettiol, W. (2006). Bioprospecting endophytic bacteria for biological control of coffee leaf rust. *Scientia Agricola*, 63(1), 32–39. <https://doi.org/10.1590/S0103-90162006000100006>
- Silva, H. S. A., Tozzi, J. P. L., Terrasan, C. R. F., & Bettiol, W. (2012). Endophytic microorganisms from coffee tissues as plant growth promoters and biocontrol agents of coffee leaf rust. *Biological Control*, 63(1), 62–67. <https://doi.org/10.1016/J.BIOCONTROL.2012.06.005>
- Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., & Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research*, 19(6), 1117. <https://doi.org/10.1101/GR.089532.108>
- Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, 122(1), 2–11. <https://doi.org/10.1002/cpmb.59>

- Solano, W. (2016). *Coffea eugenioides* S. Moore and *Coffea stenophylla* G. Don.
- Stoffelen, P., Robbrecht, E., & Smets, E. (1997). Adapted to the Rain Forest Floor: A Remarkable New Dwarf Coffea (Rubiaceae) from Lower Guinea (Tropical Africa). In *Source: Taxon* (Vol. 46, Issue 1).
- Sultan, S., Azwani, K., Ali, M., Dayana, N., Akram, M., Ashraf, K., Ashraf, M., Kaur, G., & Singh, S. (2022). Antimicrobial Activity of Secondary Metabolites Isolated from Endophytic Fungi Associated with Rubiaceae Species. In *International Journal of Pharmaceuticals, Nutraceuticals and Cosmetic Science* (Vol. 5, Issue 1).
- Thangadurai, D., Nollet, L., Islam, S., & Sangeetha, J. (2021). *Sequencing Technologies in Microbial Food Safety and Quality*. <https://www.crcpress.com/Food-Analy>
- Torres, A., & Torres, A. N. (2020). Las Rubiáceas de México: ¿Ya está hecho el trabajo? *Botanical Sciences*, 1(1), 446–463. <https://doi.org/10.17129/botsci.2847>
- Tran, D. M., Nguyen, T. H., Huynh, T. U., Do, T. O., Nguyen, Q. V., & Nguyen, A. D. (2022). Analysis of endophytic microbiome dataset from roots of black pepper (*Piper nigrum* L.) cultivated in the Central Highlands region, Vietnam using 16S rRNA gene metagenomic next-generation sequencing. *Data in Brief*, 42, 1–6. <https://doi.org/10.1016/j.dib.2022.108108>
- Tropical Plants Database. (2022). *Coffea ebracteolata* - Plantas tropicales útiles. <https://tropical.theferns.info/viewtropical.php?id=Coffea%20ebracteolata>
- Uroz, S., & Oger, P. (2017). Caballeronia mineralivorans sp. nov., isolated from oak-Scleroderma citrinum mycorrhizosphere. *Systematic and Applied Microbiology*, 40(6), 345–351. <https://doi.org/10.1016/j.syapm.2017.05.005>
- Valadez, M. del C., & Mendoza, L. (2021). Descripción de las plantas medicinales de la familia Rubiaceae con actividad. *Educación y Salud Boletín Científico Instituto de Ciencias de La Salud Universidad Autónoma Del Estado de Hidalgo*, 9(18), 168–174. <https://doi.org/10.29057/icsa.v9i18.6576>
- Valencia, A., Pervaiz, A., Husami, A., Quian, Y., & Zhang, K. (2013). *Next Generation Sequencing Technologies in Medical Genetics*. <http://www.springer.com/series/8923>
- Vandamme, P., de Brandt, E., Houf, K., Salles, J. F., van Elsas, J. D., Spilker, T., & LiPuma, J. J. (2013). Burkholderia humi sp. nov., Burkholderia choica sp. nov., Burkholderia telluris sp. nov., Burkholderia terrestris sp. nov. and Burkholderia udeis sp. nov.: Burkholderia glathei-like bacteria from soil and rhizosphere soil. *International Journal of Systematic and Evolutionary Microbiology*, 63(PART 12), 4707–4718. <https://doi.org/10.1099/IJS.0.048900-0/CITE/REFWORKS>
- Vandamme, P., & Eberl, L. (2018). Burkholderia. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1–45). Wiley. <https://doi.org/10.1002/9781118960608.gbm00935.pub2>
- Vega, F. E., Pava-Ripoll, M., Posada, F., & Buyer, J. S. (2005). Endophytic bacteria in Coffea arabica L. *Journal of Basic Microbiology*, 45(5), 371–380. <https://doi.org/10.1002/JOBM.200410551>
- Verstraete, B., Janssens, S., Smets, E., & Dessein, S. (2013). Symbiotic β -Proteobacteria beyond Legumes: Burkholderia in Rubiaceae. *PLOS ONE*, 8(1), 1–7. <https://doi.org/10.1371/JOURNAL.PONE.0055260>
- Win, Y. Y., Singh, M., Sadiq, M. B., & Anal, A. K. (2019). Isolation and identification of caffeine-degrading bacteria from coffee plantation area. *Food Biotechnology*, 33(2), 109–124. <https://doi.org/10.1080/08905436.2019.1570854>
- Yang, C.-J., & Hu, J.-M. (2018). Bacterial Leaf Nodule Symbiosis in Flowering Plants. *Symbiosis*, 1–18. <https://doi.org/10.5772/INTECHOPEN.73078>

Zuleta, L. F. G., Cunha, C. D. O., de Carvalho, F. M., Ciapina, L. P., Souza, R. C., Mercante, F. M., de Faria, S. M., Baldani, J. I., Stralio, R., Hungria, M., & de Vasconcelos, A. T. R. (2014). The complete genome of *Burkholderia phenoliruptrix* strain BR3459a, a symbiont of *Mimosa flocculosa*: Highlighting the coexistence of symbiotic and pathogenic genes. *BMC Genomics*, 15(1), 1–19. <https://doi.org/10.1186/1471-2164-15-535/FIGURES/6>

7 Anexos

Anexo 1 Tipo de técnicas de Illumina.

Tipos	Longitud de lectura (pb)	Tiempo de secuenciación	Bases totales por secuenciación (Gb)	Número de lecturas generadas (Millón)
MiSeq v2 Nano	2x250	28 horas	0,5	1
MiSeq v2 Micro	2x150	19 horas	1,2	4
MiSeq v2	2x250	39 horas	7,5-8,5	15
MiSeq v3	2x300	56 horas	13-15	25
Illumina NextSeq 550 Mid-Output	2x75	26 horas	16-19	130
GAIIx	2x100	9,5 días	54-60	320
Illumina NextSeq 550 High-Output	2x75	18 horas	50-60	400
Hiseq Rapid-Run	2x100	27 horas	100-120	600
HiScansSQ	2X100	8,5 días	135-150	750
Hiseq 1000	2X100	8,5 días	270-300	1500
Hiseq 2000	2X100	11 días	540-600	3000
Hiseq 2500 High-Output	2x125	6 días	450-500	4000
Hiseq 4000	2x75	48 horas	650-750	5000
Hiseq X Ten	2x150	72 horas	1600-1800	6000

Nova 6000	Seq	2x100	36 horas	1600-2000	9000

Fuente: (Akaçin et al., 2022; Rubio et al., 2020)

Anexo 2 Genomas de referencia

Gen 16sRNA

Secuencia de referencia de NCBI	Definición
NR_174230.1	<i>Burkholderia aenigmatica</i> strain LMG 13014
NR_165695.1	<i>Burkholderia alba</i> strain AD18
NR_159299.1	<i>Burkholderia puraquae</i> strain CAMPA 1040
NR_153664.1	<i>Burkholderia catarinensis</i> strain 89
NR_152631.1	<i>Burkholderia humptydooensis</i> strain MSMB43
NR_152630.1	<i>Burkholderia humptydooensis</i> strain MSMB43
NR_152629.1	<i>Burkholderia humptydooensis</i> strain MSMB43
NR_152628.1	<i>Burkholderia humptydooensis</i> strain MSMB43
NR_132708.1	' <i>Burkholderia humi</i> ' Srinivasan et al. 2013 strain Rs7
NR_126274.1	<i>Burkholderia australis</i> strain Q208
NR_118872.1	<i>Burkholderia vietnamiensis</i> strain TVV75
NR_118637.1	<i>Burkholderia rinojensis</i> strain A396
NR_118629.1	<i>Burkholderia thailandensis</i> strain BD10-00323
NR_102890.1	<i>Burkholderia lata</i> strain 383
NR_074687.1	<i>Burkholderia ambifaria</i> strain AMMD
NR_043552.1	<i>Burkholderia oklahomensis</i> C6786
NR_042636.1	<i>Burkholderia metallica</i> strain R-16017
NR_042634.1	<i>Burkholderia arboris</i> strain R-24201
NR_042633.1	<i>Burkholderia diffusa</i> strain R-15930
NR_042632.1	<i>Burkholderia latens</i> strain R-5630
NR_024882.1	<i>Burkholderia ambifaria</i> strain AMMD
NR_152632.1	<i>Burkholderia singularis</i> strain LMG 28154
NR_136496.1	<i>Burkholderia territorii</i> strain LMG 28158
NR_136495.1	<i>Burkholderia stagnalis</i> strain LMG 28156
NR_037064.2	<i>Burkholderia plantarii</i> strain NIAES 1723
NR_118082.1	<i>Burkholderia vietnamiensis</i> strain LMG 10929
NR_118078.1	<i>Burkholderia stabilis</i> strain LMG 14294
NR_118075.1	<i>Burkholderia pyrrocinia</i> strain LMG 14191
NR_118074.1	<i>Burkholderia plantarii</i> strain LMG 9035
NR_118070.1	<i>Burkholderia oklahomensis</i> strain LMG 23618
NR_118068.1	<i>Burkholderia multivorans</i> ATCC BAA-247 strain LMG 13010
NR_118063.1	<i>Burkholderia glumae</i> LMG 2196 = ATCC 33617
NR_118061.1	<i>Burkholderia gladioli</i> strain LMG 2216
NR_118058.1	<i>Burkholderia dolosa</i> strain LMG 18943
NR_118057.1	<i>Burkholderia cepacia</i> ATCC 25416 strain LMG 1222
NR_118053.1	<i>Burkholderia anthina</i> strain LMG 20980
NR_118051.1	<i>Burkholderia ambifaria</i> strain LMG 19182
NR_117661.1	<i>Burkholderia pseudomultivorans</i> strain LMG 26883
NR_117553.1	<i>Burkholderia gladioli</i> strain CFBP 2427
NR_116157.1	<i>Burkholderia stabilis</i> strain CIP 106845

NR_116156.1	<i>Burkholderia pyrrocinia</i> strain CIP 105874
NR_116155.1	<i>Burkholderia glumae</i> LMG 2196 = ATCC 33617 strain CIP 106418
NR_116154.1	<i>Burkholderia vietnamiensis</i> strain CIP 105875
NR_116153.1	<i>Burkholderia ubonensis</i> strain CIP 107078
NR_116152.1	<i>Burkholderia multivorans</i> ATCC BAA-247 strain CIP 105495
NR_116151.1	<i>Burkholderia plantarii</i> strain CIP 105769
NR_116150.1	<i>Burkholderia thailandensis</i> strain CIP 106301
NR_116149.1	<i>Burkholderia cepacia</i> ATCC 25416 strain LMG 1222
NR_116051.1	<i>Burkholderia thailandensis</i> E264
NR_114523.1	<i>Burkholderia multivorans</i> ATCC BAA-247 strain LMG 13010
NR_114522.1	<i>Burkholderia stabilis</i> strain LMG 14294
NR_114491.1	<i>Burkholderia cepacia</i> ATCC 25416
NR_113645.1	<i>Burkholderia cepacia</i> strain NBRC 14074
NR_113629.1	<i>Burkholderia gladioli</i> strain NBRC 13700
NR_112060.1	<i>Burkholderia pyrrocinia</i> strain ATCC 15958
NR_104978.1	<i>Burkholderia contaminans</i> strain J2956
NR_104975.1	<i>Burkholderia anthina</i> strain W92B
NR_104973.1	<i>Burkholderia dolosa</i> strain LMG 18943
NR_074312.2	<i>Burkholderia thailandensis</i> E264
NR_074299.2	<i>Burkholderia mallei</i> strain ATCC 23344
NR_044378.1	<i>Burkholderia gladioli</i> strain CIP 105410
NR_043553.1	<i>Burkholderia pseudomallei</i> strain ATCC 23343
NR_042635.1	<i>Burkholderia seminalis</i> strain R-2419
NR_041725.1	<i>Burkholderia mallei</i> strain ATCC 23344
NR_041720.1	<i>Burkholderia vietnamiensis</i> strain LMG 10929
NR_041719.1	<i>Burkholderia stabilis</i> strain LMG 14294
NR_040830.1	<i>Burkholderia ubonensis</i> strain GTC-P3-415
NR_029210.1	<i>Burkholderia pyrrocinia</i> strain 2327
NR_025013.1	<i>Burkholderia cenocepacia</i> strain LMG 16656
NR_029358.1	<i>Burkholderia multivorans</i> strain Struelens
NR_029211.1	<i>Burkholderia glumae</i> strain P 1-22-1
NR_029209.1	<i>Burkholderia cepacia</i> strain 717
NR_027587.1	<i>Burkholderia thailandensis</i> E264

Gen *gyrB*

Secuencia de referencia de NCBI	Definición	Gen ID
NZ_CP035900	<i>Burkholderia glumae</i>	58136272
NC_007510.1	<i>Burkholderia lata</i>	45093090
NZ_CP013401.1	<i>Burkholderia metallica</i>	67902368
NZ_CP020397.1	<i>Burkholderia multivorans</i>	66524923
NZ_CP009555.1	<i>Burkholderia oklahomensis</i> C6786	60547763
NZ_CP008781.1	<i>Burkholderia pseudomallei</i>	56527639
NZ_CP008781.1	<i>Burkholderia pseudomallei</i>	56527639

NZ_CADFDQ010000031.1	<i>Burkholderia pseudomultivorans</i>	60979693
NZ_CP011503.1	<i>Burkholderia pyrrocinia</i>	55504863
NZ_CP013398.1	<i>Burkholderia seminalis</i>	62009471
NZ_LR025742.1	<i>Burkholderia stabilis</i>	71054937
NZ_CP013459.1	<i>Burkholderia stagnalis</i>	46324467
NZ_CP013366.1	<i>Burkholderia territorii</i>	46198695
NZ_CP035900.1	<i>Burkholderia glumae</i>	58136272
NZ_CP020737.1	<i>Burkholderia ubonensis</i> subsp. mesacidophila	69005413
NZ_LOWL01000016.1	<i>Burkholderia vietnamiensis</i>	69524425
NZ_CP066037.1	<i>Burkholderia ambifaria</i>	69545870
NZ_CABVLY010000031.1	<i>Burkholderia anthina</i>	6503988
NZ_CP013450.1	<i>Burkholderia cenocepacia</i>	56556494
NZ_JAIZPY010000010.1	<i>Burkholderia cepacia</i>	69456493]
NZ_AP018358.1	<i>Burkholderia contaminans</i>	61084410
NZ_CP012746.1	<i>Paraburkholderia caribensis</i> MBA4	69967739
NZ_CP009795.1	<i>Burkholderia dolosa</i> AU0158	64462767
NZ_CP068049.1	<i>Burkholderia gladioli</i>	66455886
NZ_CP008781.1	<i>Burkholderia pseudomallei</i>	56527639
NZ_CP009727.1:	<i>Burkholderia mallei</i>	56593631

Gen recA

Secuencia de referencia de NCBI	Definición	Gen ID
NC_007510.1	<i>Burkholderia lata</i>	45095862
NZ_CP009727.1	<i>Burkholderia mallei</i>	56596587
NZ_CP013401.1	<i>Burkholderia metallica</i>	67905054
NZ_CP009555.1	<i>Burkholderia oklahomensis</i>	60547143
NZ_CP008781.1	<i>Burkholderia pseudomallei</i>	56527045
NZ_CADFDQ010000005.1	<i>Burkholderia pseudomultivorans</i>	60975492
NZ_CABVLY010000001.1	<i>Burkholderia anthina</i>	56498142
NZ_CP013398.1	<i>Burkholderia seminalis</i>	62012102
NZ_LR025742.1	<i>Burkholderia stabilis</i>	71055467
NZ_CP013459.1	<i>Burkholderia stagnalis</i>	46327234
NZ_CP013366.1	<i>Burkholderia territorii</i>	46201238
NZ_CP013411.1	<i>Burkholderia thailandensis</i>	66546179
NZ_CP020737.1	<i>Burkholderia ubonensis</i> subsp. mesacidophila	69003682
NZ_CP066037.1	<i>Burkholderia ambifaria</i>	69543103
NZ_CP012746.1	<i>Paraburkholderia caribensis</i> MBA4	69967300
NZ_CP013450.1	<i>Burkholderia cenocepacia</i>	56559233
NZ_JAIZPY010000004.1	<i>Burkholderia cepacia</i>	69453946
NZ_AP018358.1	<i>Burkholderia contaminans</i>	61087190
NZ_CP009795.1	<i>Burkholderia dolosa</i> AU0158	64460232
NZ_CP068049.1	<i>Burkholderia gladioli</i>	66459051

NZ_CP035900.1	<i>Burkholderia glumae</i>	58135802
NZ_CP020397.1	<i>Burkholderia multivorans</i>	66527557