

Diversidad y predicción del potencial funcional de las comunidades bacterianas asociadas a vitroplantas de banano cv. Williams

María Camila Isaza Álvarez

Asesor:
Javier Correa Álvarez
PhD en Genética y Biología

Coasesora:
Gisell García Giraldo
Estudiante MSc. Biotecnología

Universidad EAFIT
Escuela de ciencias
Departamento de ciencias biológicas
Pregrado de biología
Medellín
2019

Agradecimientos

ii

Agradezco a Dios por guiarme siempre en mi camino profesional y personal y por darme fortaleza en cada etapa de mi vida.

A mis Padres por su apoyo incondicional y por sus consejos que siempre me ayudan a aclarar la mente y tomar las mejores decisiones.

A a mi hermana, por su apoyo, motivación y compañía incondicional.

Agradezco a las personas que hicieron posible la realización de este proyecto:
Javier Correa Álvarez, PhD. en Genética y Biología y asesor de este proyecto,
Gisell García Giraldo, Estudiante de MSc. en Biotecnología y coasesora del proyecto por su ayuda y brindar los datos de secuenciamiento,
Profesoras Luisa Posada y Valeska Villegas por brindar los datos de secuenciamiento con los que se realizó el presente proyecto.
Integrantes del semillero de Biología Computacional.

Diferentes estudios se han encargado de caracterizar la microbiota asociada a diferentes especies vegetales. En ellos se han detectado diversos rasgos fenotípicos vegetales producidos por un arreglo específico en la microbiota. La identificación de estas asociaciones tiene grandes aplicaciones en el campo de la agricultura en donde se puede potenciar la producción de los cultivos, estos efectos se incrementan aun más en los cultivos reproducidos por micropropagación en donde se han descubierto ciertos efectos que pueden llevar a que el proceso sea aun más eficiente. En el presente estudio, se analizó la microbiota en la endosfera y filosfera de plantas de banano en estado *in vitro* a partir de la secuenciación de dos regiones del gen ARNr 16S, asimismo se estableció un flujo de trabajo bioinformático para identificar variantes de secuencias de amplicones (ASVs). El perfil taxonómico reveló que la composición y la diversidad de las comunidades bacterianas estaban determinadas principalmente por la variable Lote del cual provinieron las muestras; y que su composición estaba representada en su mayoría por los filos Actinobacteria, Firmicutes y Proteobacteria y por los géneros *Bacillus*, *Stenotrophomonas* y *Pseudomonas*. Finalmente, con las secuencias del 16S se realizó también una predicción del potencial funcional del microbioma de la planta, en la que se encontraron diferentes rutas metabólicas como la biosíntesis de sideróforos, la producción de antibióticos y el metabolismo del nitrógeno que reflejan relaciones simbióticas entre las bacterias y su hospedero.

Introducción	1
Marco teórico	5
El microbioma vegetal	5
Definición de la microbiota y microbioma vegetal.....	5
El microbioma de la rizosfera	6
El microbioma de la filosfera.....	8
El microbioma de la endosfera	9
Ensamble del microbioma vegetal	11
Modulación del microbioma vegetal	13
Microbioma vegetal en plantas de cultivo <i>in vitro</i>	15
Metagenómica para el estudio del microbioma vegetal.....	17
Metataxonomía a partir del ARN 16S	18
- Marcadores moleculares utilizados en el estudio del microbioma	18
-Plataformas para la producción de datos de microbiomas	20
-Aproximaciones y programas bioinformáticos para el procesamiento de secuencias de microbiomas.....	22
-Análisis aplicados en los estudios taxonómicos de microbiomas	30
Estudio de las capacidades metabólicas del microbioma vegetal.....	34
-Secuenciación tipo <i>shotgun</i>	34
-Métodos alternativos para el estudio de las capacidades funcionales del microbioma	36
Musa acuminata AAA como modelo de estudio del microbioma endófito de plantas <i>in vitro</i>	38
-El cultivo del banano	38
Objetivos	44
Objetivo general.....	44
Objetivos específicos	44
Metodología	45
Obtención de secuencias	45
Procesamiento de secuencias	46
Análisis de la composición y diversidad de la microbiota.....	48
Comparación de pipelines.....	49
Predicción funcional	51
Resultados	53
Procesamiento de secuencias	53
Análisis de la composición y diversidad de la microbiota.....	55
Comparación de pipelines.....	62
Predicción funcional	68
Discusión.....	72
Procesamiento de secuencias	72
Análisis de la composición y diversidad de la microbiota.....	74
Comparación de pipelines.....	80
Predicción funcional	83
Conclusiones	88
Perspectivas futuras	90
Lista de referencias	92
Apéndice	117

Tabla 1. Descripción de las librerías de amplicones producidas en la secuenciación con Illumina MiSeq y lecturas procesadas por el método DADA2 (lecturas 300 pb x 2) 54

Figura 1. Flujo de trabajo con los pasos generales para el procesamiento de lecturas pareadas provenientes de las plataformas Illumina y la metodología de construcción de OTUs..... 24

Figura 2. Flujo de trabajo con los pasos generales para el procesamiento de lecturas pareadas provenientes de las plataformas Illumina con el paquete DADA2 de Callahan et al., 2016 29

Figura 3. Abundancia relativa de los filos presentes en la microbiota bacteriana asociada a la endosfera de raíces y pseudotallos y a hojas de plantas de banano en estado in vitro..... 56

Figura 4. Abundancia relativa de los géneros más predominantes (> 2%) en la microbiota bacteriana asociada a la endosfera de raíces y pseudotallos y a las hojas de plantas de banano en estado in vitro..... 58

Figura 5. Abundancia relativa de las especies presentes en la microbiota bacteriana asociada a la endosfera de raíces y pseudotallos y a hojas de plantas de banano en estado in vitro..... 59

..... 59

Figura 6. Curvas de rarefacción que representan el número de ASVs identificadas en 9 muestras de la microbiota de vitroplantas de banano. a) Tamaño de las librerías antes de la normalización y b) Profundidad de las muestras normalizadas por rarefacción a una profundidad de 3,548. 60

Figura 7. Índices de diversidad alfa de la microbiota asociada a la endosfera de raíces, pseudotallos y hojas de plantas de banano in vitro. 61

Figura 8. Análisis de coordenadas principales de la microbiota asociada a la endosfera de raíces, pseudotallos y hojas de vitroplantas de banano basado en la métrica de distancia de Bray-Curtis y normalización por el método GMPR aplicada a la matriz de ASVs..... 61

Figura 9. Diagramas de Venn que muestran la intersección de grupos taxonómicos identificados por cada pipeline (pipeline del NIOO y pipeline DADA2) y por cada versión de la base de datos Silva (DADA2 con Silva versión 128 y DADA2 con1 Silva versión 132). 63

Figura 10. Índices de diversidad alfa de la microbiota identificada por cada pipeline (pipeline del NIOO y pipeline DADA2) y por cada versión de la base de datos Silva (DADA2 Silva versión 128 y DADA2 Silva versión 132). 65

Figura 11. Análisis de coordenadas principales basado en la métrica de distancia de Bray-Curtis de la microbiota asociada a plantas de banano en estado in vitro, identificada por cada pipeline (pipeline del NIOO y pipeline DADA2) y por cada versión de la base de datos Silva (DADA2 Silva versión 128 y DADA2 Silva versión 132). a) DADA2 con Silva versión 128, b) pipeline del NIOO y c) DADA2 con Silva versión 132. 66

Figura 12. Análisis procrustes de los PCoA de las distancias de Bray-Curtis de a) pipeline del NIOO y pipeline DADA2 y b) DADA2 con Silva versión 128 y DADA2 con Silva versión 132 (el valor de significancia de la correlación se indica en la figura). 67

Figura 13. Análisis de componentes principales (PCA) de los ortólogos KEGG identificados en las comunidades bacterianas presentes en la endosfera y filosfera de plantas de banano en estado in vitro, producto de la predicción funcional realizada con PICRUS2 y MicrobiomeAnalyst. .. 69

Figura 14. Mapa de calor que muestra la abundancia relativa de las rutas metabólicas significativas ($P < 0.05$) presentes en los perfiles funcionales predichos con PICRUS2 y MicrobiomeAnalyst filtrados y normalizados con el método CSS para las comunidades bacterianas presentes en la endosfera y filosfera de plantas de banano en estado in vitro. En la parte superior se muestra un dendrograma que muestra el agrupamiento de las muestras a partir de su composición funcional..... 70

Introducción

Las plantas son la columna vertebral de toda la vida en la tierra y un recurso esencial para la supervivencia y el bienestar humano. Estas se enfrentan constantemente a perturbaciones bióticas y abióticas que les pueden causar daños superficiales, sistémicos (enfermedades) y hasta afectar su productividad (Jones & Dangl, 2006). La defensa de la planta se basa en la inmunidad innata de cada célula y en las señales sistémicas que emanan de los sitios de infección (Ausubel, 2005; Jones & Dangl, 2006); sin embargo, recientemente se ha descubierto que la respuesta inmune de las plantas y otros procesos metabólicos involucrados en su desarrollo y crecimiento se ven influenciados por las comunidades microbianas que habitan en ella formando estrechos lazos de relación. Estas relaciones se van construyendo desde las semillas, en donde la planta se ve dotada de un conjunto de microbios provistos por sus progenitores (Berg, Grube, Schloter, & Smalla, 2014) y posteriormente cuando los primeros órganos que salen de la semilla se exponen a otras fuentes de microorganismos de los que puede disponer de acuerdo a las condiciones del ambiente (Kandel, Joubert, & Doty, 2017).

El microbioma de una planta se compone de un microbioma núcleo común a todos los individuos de una variedad vegetal o especie y un microbioma variable dependiente de factores como el tipo de suelo, la ubicación geográfica, la estación climática y las prácticas de manejo de cultivo (Lemanceau, Blouin, Muller, & Moënnelocoz, 2017; Lundberg et al., 2012). Generalmente es el microbioma núcleo el que contiene microorganismos y funciones fundamentales para el *fitness* de la planta y el funcionamiento del holobionte (Berg, Rybakova, Grube, & Köberl, 2016; Hamonts et al., 2018).

El conocimiento del microbioma vegetal está marcando el camino hacia una agricultura de nueva generación, la cual considera a la planta y su microbioma en conjunto para el desarrollo de programas de mejoramiento de cultivos (Sessitsch & Mitter, 2015). Ejemplos de las aplicaciones que tiene este tipo de agricultura son la producción de cultivares con capacidades mejoradas para interactuar con microbios beneficiosos de la microbiota del suelo natural y la aplicación de inoculantes microbianos para brindarle a la planta capacidades de respuesta mejoradas relacionadas con el uso de nutrientes o la inmunidad contra patógenos (Schlaeppi & Bulgarelli, 2015). El éxito de estas estrategias depende de la consideración de los factores que modulan la estructura del microbioma como el suelo nativo, las condiciones climáticas y el contexto geoquímico específico del entorno del suelo en particular (Schlaeppi & Bulgarelli, 2015).

Las relaciones simbióticas entre la planta y su microbiota resultan en la promoción del crecimiento vegetal, la eficiencia en el uso de nutrientes, la tolerancia al estrés abiótico y la resistencia a las enfermedades (Lebeis, 2014); estas cualidades son de gran importancia en la producción agrícola debido a que representan ganancias económicas al optimizar el uso de recursos y reducir los impactos ambientales producidos por agroquímicos conduciendo así hacia una agricultura más sostenible (Busby et al., 2017; Rossmann, Sarango-Flores, Barros Chiaramonte, Pezzo Kmit, & Mendes, 2017).

La microbiota vegetal también forma asociaciones con plantas cultivadas en ambientes *in vitro* que impactan la tasa de supervivencia y el desarrollo de los brotes producidos durante el proceso de la micropropagación (Orlikowska, Nowak, & Reed, 2017). Diferentes géneros de bacterias han sido relacionados con el aumento del número

y la longitud de las raíces, tallos y el área superficial de las hojas; el aumento en el peso seco de las semillas y la biomasa de las microplantas; el aumento en la tasa de germinación de las semillas; el aceleramiento del proceso de enraizamiento; la prevención de pérdida de agua; la inhibición de contaminantes presentes en los medios de cultivo y el aumento en el porcentaje de la plantas aclimatadas (Orlikowska et al., 2017). A pesar de todos los beneficios que experimentan las plantas *in vitro*, se ha encontrado que la utilización de sustratos como medios de cultivo, en lugar de suelo natural reduce la diversidad de la microbiota dejando a la planta desprotegida y sin algunas asociaciones simbióticas que la podrían beneficiar en su etapa adulta (Lucero, Unc, Cooke, Dowd, & Sun, 2011).

El banano es una planta cultivada en los trópicos a lo largo de todo el mundo; caracterizada porque su fruto es altamente energético y hace parte de la seguridad alimentaria de diversos países de África y Latinoamérica (FAO Trade and markets, 2018). El sector bananero es de gran importancia para Colombia debido a que es el producto de exportación que genera mayores ingresos para el país, de hasta 811 millones de dólares y provee más de 200 mil empleos directos e indirectos (Cadena de banano Indicadores e Instrumentos 2018). Los cultivares más comercializados de banano se obtienen bajo el proceso de micropropagación a partir del cormo presente en el tallo verdadero de las plantas adultas (bajo tierra). Al ser un producto cultivado a nivel mundial y que ha recibido mucha atención en estudios de microbiomas, se ha convertido en una planta modelo para estudiar las interacciones bacteria-planta durante el proceso de la micropropagación (Fernandes et al., 2013; Robinson, Fraser, & Eckstein, 1993; Thomas & Sekhar, 2016).

Diversos estudios han demostrado que el conocimiento del ensamble inicial del microbioma de una planta podría tener un impacto significativo en su posterior producción agrícola. Ésto se atribuye al aprovechamiento de los efectos prioritarios del microbioma (ventajas de los primeros colonizadores) (Toju et al., 2018). De esta manera, la modulación de la microbiota en una etapa temprana del desarrollo podría promover la resistencia de la planta a perturbaciones bióticas y abióticas y a mejorar su desarrollo conllevando a mitigar las pérdidas de rendimiento características de esta fase del desarrollo asegurando que un porcentaje mayor de plantas lleguen a estado adulto (Orlikowska et al., 2017). Es por esto por lo que el objetivo de este estudio fue explorar la diversidad taxonómica bacteriana y las posibles rutas metabólicas involucradas en el sano desarrollo de las plantas de banano en estado *in vitro*.

Marco teórico

El microbioma vegetal

Definición de la microbiota y microbioma vegetal

Las plantas, al igual que otros organismos eucariotas, se catalogan como holobiontes, es decir, que se definen como una entidad biológica compuesta de un organismo multicelular complejo que actúa como hospedero de un conjunto de microorganismos denominados microbiota (Berg et al., 2014; Vandenkoornhuyse, Quaiser, Duhamel, Le Van, & Dufresne, 2015). Dentro de la diversidad microbiana asociada a plantas se encuentran microorganismos mutualistas, comensalistas y patógenos (Innerebner, Knief, & Vorholt, 2011; Mansfield et al., 2012); la colección del material genético de cada uno de ellos es lo que se conoce como el microbioma vegetal (Dhariwal et al., 2017; Gill et al., 2006).

La microbiota de las plantas se compone en su mayoría de bacterias, sin embargo, también aloja otros microorganismos como hongos, oomicetos, algas, protozoos y virus, los cuales se distribuyen en los diferentes tejidos de las plantas (Bulgarelli et al. 2013). La distribución de los microorganismos en el cuerpo de la planta permite distinguir tres grandes compartimentos; la rizosfera, la filosfera y la endosfera. La rizosfera fue definida en 1904 como los microorganismos presentes en la región del suelo que se ve influenciada por las raíces por medio de sus exudados y los exudados de los microorganismos presentes en la matriz del suelo (Turner, James, & Poole, 2013), estos microorganismos son aquellos que se adhieren a la superficie de las raíces y habitan a 1 mm de distancia en su suelo adyacente (Hirsch & Mauchline, 2012). La filosfera hace referencia a los microorganismos presentes en la superficie de todos los órganos aéreos

de la planta (Bakker, Berendsen, Doornbos, Wintermans, & Pieterse, 2013) y finalmente, la endosfera abarca todos los microorganismos que habitan, por la mayoría de su ciclo de vida, en los tejidos internos de la planta (Beckers et al., 2016; Kandel et al., 2017).

Diversos estudios han demostrado que diferentes microorganismos de la microbiota, independiente del compartimiento al que pertenezcan, influyen diferentes procesos de la planta como su crecimiento, la resistencia a factores bióticos y abióticos, su salud y productividad, (Lebeis, 2014) Es por esto que muchos autores aseguran que para que las plantas mantengan un buen estado de salud, deben mantener un microbioma diverso (Turner et al., 2013).

El microbioma de la rizosfera

La rizosfera tradicionalmente ha sido el compartimiento más investigado en los estudios del microbioma vegetal, esto se debe a que los microorganismos y el suelo de la rizosfera están directamente relacionados con los nutrientes que toma la planta, la resistencia a enfermedades y la productividad de los cultivos (Edwards et al., 2015; Rossmann, Sarango-Flores, Barros Chiaramonte, et al., 2017). El suelo de la rizosfera es fácilmente distinguible del resto del suelo circundante de la planta, debido a que es una zona abundante en carbono, el cual congrega una comunidad de microorganismos que supera en gran medida a los microorganismos de la matriz suelo (Orlikowska et al., 2017; Rossmann, Sarango-Flores, Chiaramonte, Kmit, & Mendes, 2017). Aunque se ha demostrado que el microbioma de la rizosfera depende en gran medida del tipo de suelo en el que se encuentre (Lundberg et al., 2012), cada planta cuenta con la capacidad de modificar la estructura de la microbiota presente en el suelo de la rizosfera, mediante la modulación del pH, la estructura del suelo, la disponibilidad de oxígeno, la concentración

de antimicrobianos y la fuente energía presente en el suelo (Dennis, Miller, & Hirsch, 2010)

Las asociaciones planta-microorganismos en el compartimiento de la rizosfera trae beneficios para ambas partes: Las plantas producen diferentes fuentes de energía como órganos vegetales muertos y exudados fotosintéticos que son aprovechados por los microorganismos para su alimentación (Mendes et al., 2011; Pandey, 2015). Mientras que diferentes microorganismos proveen a la planta de un pool de nuevas funciones y rasgos genéticos necesarios para la fijación de nitrógeno (Bhattacharjee, Singh, & Mukhopadhyay, 2008; Santi, Bogusz, & Franche, 2013; Schlaeppi & Bulgarelli, 2015) y promueven la adaptación a estrés biótico y abiótico, como: la presencia de nematodos, hongos y oomicetos patógenos (Elbeltagy et al., 2001); condiciones de sequía o inundación; altas concentraciones de sal; bajas temperaturas; pH extremos y altas concentraciones de compuestos tóxicos (Arkhipova et al., 2007).

Se ha encontrado también que diferentes microorganismos como hongos micorrícicos y bacterias productoras de sideróforos proveen a la planta de nutrientes como fósforo, nitrógeno y hierro; mientras que otras rizobacterias promueven el crecimiento de la planta y suprimen enfermedades al competir con los patógenos por nutrientes y espacio o al activar la resistencia sistémica de la planta (Iniguez, Dong, & Triplett, 2004). Todos estos beneficios producidos por los microorganismos permiten evidenciar que el microbioma de la rizosfera contribuye en gran medida al crecimiento y la salud de la planta, su productividad, el secuestro de carbono y la fitoremediación (De Deyn, Cornelissen, & Bardgett, 2008). A pesar de esto, los microorganismos de la

rizosfera no sólo son mutualistas y comensalistas; también se ha registrado la presencia de patógenos furtivos (Hallmann, Quadt-Hallmann, Mahaffee, & Kloepper, 1997).

El microbioma de la filosfera

El microbioma de la filosfera incluye a todos los microorganismos que habitan sobre la superficie de cualquier órgano aéreo de la planta, sin embargo, el microbioma foliar ha sido el más estudiado (Vorholt, 2012). Las bacterias que se organizan en la superficie de las plantas generalmente se encuentran a lo largo de las venas foliares o se organizan en biopelículas sobre la superficie de la hoja teniendo un impacto sobre los posibles epifitos que se integren en la planta a partir del microbioma del viento (Lindow & Brandl, 2003; Lucero et al., 2011).

Diferentes microorganismos presentes en la filosfera intervienen en los procesos de fijación de nitrógeno; protección de la planta contra patógenos; síntesis de fitohormonas y modificación de metabolitos, los cuales son indispensable para el desarrollo de la planta (Rossmann, Sarango-Flores, Barros Chiamonte, et al., 2017). También se ha encontrado que estos microorganismos promueven la captación de sustratos a través de porinas y transportadores ABC; la resistencia a estrés abiótico como las especies reactivas de oxígeno (ROS); y el ciclaje de los nutrientes (Innerebner et al., 2011; Lebeis, 2015). Un género identificado comúnmente en la filosfera es *Methylobacterium*, dentro del cual se encuentran bacterias que, además de participar en las actividades mencionadas anteriormente están involucradas en la metanotrofia (Delmotte et al., 2009).

Las comunidades del microbioma de la filosfera tienden a cambiar más rápidamente que las de los otros compartimentos, debido a que esta zona, especialmente

las hojas están expuestas a constantes perturbaciones físicas y su acceso a nutrientes generalmente es limitado (Bodenhausen, Horton, & Bergelson, 2013).

El microbioma de la endosfera

La endosfera comprende a los microorganismos presentes en la endorizosfera, la endofilosfera, el tejido interno reproductivo y el tejido interno de las semillas (Coombs & Franco, 2003). Estos microorganismos pueden ser sistémicos verdaderos, cuando están influenciados por el metabolismo de la planta y su origen se debe a la transmisión vertical; o transitorios / facultativos cuando su estadía dentro de los tejidos de la planta abarca poco tiempo de su ciclo de vida (Hardoim, Overbeek, & Elsas, 2008). La diversidad, abundancia y acción de los últimos es muy dependiente de factores externos en el ambiente; se han registrado casos de microorganismos que en un principio no muestran signos de patogenicidad en la planta, pero cuando la planta se expone a una condición de recursos limitados, produce una actividad patogénica (Wani, Ashraf, Mohiuddin, & Riyaz-Ul-Hassan, 2015).

Se ha identificado que los endófitos tienen gran influencia en la salud, crecimiento y el desarrollo de la planta y hasta pueden generar alteraciones locales y sistémicas en el metabolismo de ésta (Mathys et al., 2012; Rossmann, Sarango-Flores, Barros Chiaramonte, et al., 2017). Diversas cepas bacterianas tienen la capacidad de suprimir enfermedades y la acción de patógenos (Berg et al., 2014), de reducir los impactos ambientales producidos por diferentes fertilizantes (Kandel et al., 2017), mejorar la captación de nutrientes (Mendes, Pizzirani-Kleiner, Araujo, & Raaijmakers,

2007), promover la tolerancia a estrés biótico y abiótico y el crecimiento vegetal, inclusive en ambientes con escasos nutrientes (Rosenblueth & Martínez-Romero, 2006). La fijación de nitrógeno; la síntesis de fitohormonas como el ácido 3-indol acético, las citoquininas y giberelinas; la síntesis de antibióticos (Schardl et al., 2013); el control de los niveles hormonales de la planta y la adquisición de nutrientes son algunos mecanismos utilizados por las bacterias para promover el crecimiento vegetal en la endosfera (Kandel et al., 2017; Lebeis, 2015). La promoción de la resistencia de la planta a estrés biótico y abiótico lo hacen algunas bacterias al liberar compuestos antimicrobianos, produciendo sideróforos y modulando la respuesta inmune de la planta (Mercado-Blanco & Lugtenberg, 2014; Santoyo, Moreno-Hagelsieb, del Carmen Orozco-Mosqueda, & Glick, 2016). Algunas bacterias que llevan a cabo estos procesos benéficos para la planta son las pertenecientes a los filos Proteobacteria, Actinobacteria, Firmicutes y Bacteroidetes; y a los géneros *Burkholderia*, *Klebsiella* y *Pseudomonas* (Bastian et al., 1998; Elbeltagy et al., 2001; Hardoim et al., 2015; Wemheuer et al., 2017). La planta recibe todos estos beneficios, a cambio de brindar un ambiente en el cual los endófitos se puedan establecer y alimentarse de diferentes fuentes de energía. Muchos endófitos fueron en un principio miembros de la rizosfera y gracias a las interacciones que sostenían con el hospedero y sus atributos de utilización de nicho lograron incorporarse al compartimento de la endosfera (Schulz, Boyle, & Sieber, 2006). Una vez dentro de los tejidos vegetales, se movilizan generalmente hasta ubicarse en algún espacio intercelular, esto debido a que en estas áreas abundan los carbohidratos, aminoácidos y nutrientes inorgánicos (Dong et al., 1994; Kandel et al., 2017).

Ensamble del microbioma vegetal

El microbioma vegetal tiene diferentes orígenes (Berg et al., 2014). La microbiota en un inicio puede ser producto de transferencia vertical, principalmente en el caso de los endófitos, los cuales pueden provenir de las plantas parentales, permanecer en las semillas y luego colonizar los primeros órganos de la planta (Kandel et al., 2017; Lucero et al., 2011; Truyens, Weyens, Cuypers, & Vangronsveld, 2015). Los microorganismos también pueden venir de los microbiomas de fuentes abióticas como el suelo, el viento y el agua o de fuentes animales, cuando tienen contacto con las raíces o los órganos aéreos de la planta (Rastogi et al., 2012) como es el caso del microbioma foliar (Vorholt, 2012). Sin embargo, la fuente principal de microorganismos es el suelo de la rizosfera, en la cual no solo habitan microbios que interactúan con la raíz, sino también aquellos que pueden entrar en la planta y convertirse en endófitos (Bulgarelli et al., 2012; Hirsch & Mauchline, 2012).

Es la planta la encargada de reclutar y seleccionar los microorganismos del suelo que serán parte de su microbiota. La colonización de la rizosfera comienza con la producción de rizodepositos por parte de la planta en forma de exudados, glucosa, aminoácidos, ácidos orgánicos, polisacáridos y proteínas que son dependientes del genotipo de la planta (Kawasaki et al., 2016; Philippot, Raaijmakers, Lemanceau, & Van Der Putten, 2013; Turner et al., 2013). Estos compuestos alteran la composición de la microbiota de la rizosfera y promueven el establecimiento de nuevos microorganismos en la comunidad microbiana, los cuales, se dirigen hacia la raíces hasta adherirse a ellas gracias a su afinidad con los rizodepositos (Kandel et al., 2017). Una vez en la rizosfera, los microorganismos son clasificados y solo permanecerán en ella aquellos que harán parte de este compartimiento o posteriormente de la endosfera, los demás serán repelidos

por acción de los exudados (Rossmann, Sarango-Flores, Chiaramonte, et al., 2017); esta clasificación es altamente dependiente del genotipo de la planta (Bulgarelli et al., 2012; Philippot et al., 2013). Finalmente, los microorganismos seleccionados para entrar a los tejidos internos de la planta exploran posibles sitios de entrada como estomas, heridas y aperturas en la raíz producidas por pelos radiculares o raíces laterales (Hardoim et al., 2015; Kandel et al., 2017; Naveed et al., 2013). Algunas bacterias hasta pueden expresar genes para facilitar la colonización del tejido vegetal y de un nicho específico en la planta (Compant, Clément, & Sessitsch, 2010). Algunos endófitos una vez en el rizoplasma producen enzimas celulolíticas y algunos metabolitos que facilitan su defensa en el proceso competitivo con otros microbios y la entrada y movilización por los tejidos internos de la planta (Brader, Compant, Mitter, Trognitz, & Sessitsch, 2014; Compant et al., 2005). No se ha encontrado evidencia de que la colonización de la raíz y los tejidos internos vegetales se vea afectada por el sistema inmune de la planta, es por esto, que se hipotetiza que las plantas tienen la capacidad de diferenciar microorganismos mutualistas o comensales de posibles patógenos (Kandel et al., 2017).

El primer ensamble del microbioma vegetal es fundamental para determinar la prevalencia de futuros microorganismos en las comunidades microbianas de todos los compartimientos. De esto se trata el efecto de prioridad, el cual, consiste en que el orden de colonización de los primeros microorganismos, definirá si los siguientes colonizan exitosamente o son suprimidos (Fukami, 2015; Werner & Kiers, 2015). Esto se debe a que los primeros colonizadores disponen de más recursos, nichos y microespacios en la planta; tienen la oportunidad de producir barreras o antimicrobianos que pueden retardar la colonización por parte de otros microbios y hasta pueden alterar la respuesta inmune de

las plantas provocando que reaccionen diferente ante la llegada de los nuevos microorganismos (Jacott, Murray, & Ridout, 2017; Pieterse et al., 2014).

Modulación del microbioma vegetal

La estructura del microbioma esta determinada por factores bióticos como la especie, la variedad vegetal, el compartimiento analizado, el estadio de desarrollo y la salud de la planta; y factores abióticos como el tipo de suelo, la ubicación geográfica, la estación climática y las practicas de manejo del cultivo (Berg et al., 2016; Berg & Smalla, 2009; Bulgarelli et al., 2012; Hardoim, Hardoim, van Overbeek, & van Elsas, 2012; Lundberg et al., 2012; Müller, Vogel, Bai, & Vorholt, 2016; Wagner et al., 2016). A pesar de esto, se han identificado taxones comunes en la microbiota de un gran número de especies de plantas. Los filos Proteobacteria, Firmicutes, Actinobacteria y Bacteroidetes son comunmente encontrados en todos los compartimientos de diversas plantas (Bulgarelli et al., 2012; Bulgarelli, Schlaeppi, Spaepen, van Themaat, & Schulze-Lefert, 2013; Lundberg et al., 2012; Rossmann, Sarango-Flores, Barros Chiaramonte, et al., 2017). En la rizosfera además, se registra la vasta presencia de bacterias del filo Acidobacteria; arqueas del filo Crenarchaeota y hongos Ascomycota y Glomeromycota. Este compartimiento también se caracteriza por tener una alta proporción de bacterias no clasificadas (Berg & Smalla, 2009; Mendes, Garbeva, & Raaijmakers, 2013; Philippot et al., 2013). En la filosfera también se registran comúnmente bacterias de las clases Alphaproteobacteria y Gammaproteobacteria (Rastogi et al., 2012; Whipps, Hand, Pink, & Bending, 2008) y géneros como *Pseudomonas*, *Sphingomonas*, *Methylobacterium*, *Bacillus*, *Massilia*, *Arthrobacter* y *Pantoea* (Bulgarelli et al., 2013). Y a nivel de la endosfera predominan junto con las bacterias, hongos Ascomycota y Basidiomycota

(Toju et al., 2013). Por su amplia distribución, los géneros mencionados hacen parte del microbioma núcleo de diversas especies de plantas (Hamonts et al., 2018; Lundberg et al., 2012). El microbioma núcleo se define como los microbios recurrentes en casi todas las comunidades presentes en un hospedero específico (Hamonts et al., 2018), estos microorganismos se caracterizan porque son fundamentales para el *fitness* de la planta y el funcionamiento del holobionte (Berg et al., 2016; Lemanceau et al., 2017).

Se ha encontrado que el microbioma de la rizosfera experimenta cambios dependientes del estadio de desarrollo de la planta; de esta manera el microbioma de algunas plántulas es significativamente diferente del presente en las etapas posteriores de la vida de la planta; esto se debe al tipo de sustratos que libera la planta a la rizosfera. En las primeras etapas de su ciclo de vida los compuestos que más se liberan son alcohol y azúcares, los cuales atraen una gran diversidad de microbios (Chaparro, Badri, & Vivanco, 2014), mientras que en etapas más tardías del desarrollo los compuestos liberados son más específicos por lo que la selección de bacterias que harán parte del microbioma de la rizosfera es más estricta (Bais, Weir, Perry, Gilroy, & Vivanco, 2006). Como se mencionó anteriormente la principal fuente de microorganismos endófitos son aquellos que se incorporan a la rizosfera, es por esto que los cambios en el microbioma endófito también se ven influenciados por los compuestos liberados por la planta y la composición de la rizosfera (Berendsen, Pieterse, & Bakker, 2012). Adicionalmente, las diferentes perturbaciones ambientales provocan la activación de ciertas bacterias endófitas que se encargan de modular la respuesta de la planta (Hirsch & Mauchline, 2012). Los cambios en la composición del microbioma foliar se ven influenciados por la distribución de las fuentes de carbono en las hojas y por factores ambientales como la

temperatura, la humedad y la radiación, los cuales causan que la microbiota se distribuya heterogéneamente por las venas de la hoja (Lindow & Brandl, 2003; Vorholt, 2012).

Microbioma vegetal en plantas de cultivo *in vitro*

Muchas plantas de cultivo importantes se propagan vegetativamente, es decir, que se producen múltiples clones que representan una extensión de la línea celular somática de un individuo (George, Hall, & Klerk, 2008). Esta técnica, denominada micropropagación, es utilizada para realizar la propagación de diversas especies comerciales y se define como el proceso de producción en masa de un genotipo seleccionado a precios competitivos, por medio del cultivo *in vitro* (Debergh & Read, 1991), el cual, permite trabajar en espacios pequeños para la producción de un gran volumen de material vegetal libre de patógenos de campo como gorgojos, nematodos y virus; y asegurar una mayor tasa de multiplicación (Vuylsteke, 1998).

Por mucho tiempo la presencia de microorganismos en plantas de cultivo *in vitro* se considero algo negativo, como un indicativo de falta de desinfección; sin embargo, ahora se sabe que eliminar la totalidad de las bacterias de un explante es casi imposible (Lucero et al., 2011; Orlikowska et al., 2017) y se ha encontrado que varias bacterias que sobreviven los procesos de desinfección tienen efectos positivos que ayudan a la planta durante el proceso de micropropagación (Quambusch, Pirttila, Tejesvi, Winkelmann, & Bartsch, 2014).

La mayoría de los microorganismos detectados en plantas de cultivo *in vitro* se encuentran en el compartimiento de la endosfera, sin embargo, también se han encontrado microbios sobre la superficie de las hojas en explantes que se mantuvieron en condiciones asépticas (Lucero et al., 2011). Los endófitos encontrados hasta el momento

son en su mayoría bacterias, provenientes de diferentes fuentes como el ambiente, la planta, animales, aguas residuales y humanos (de Almeida et al., 2009; Fletcher, Leach, Eversole, & Tauxe, 2013); como las pertenecientes a los géneros *Azorhizobium*, *Azospirillum*, *Azotobacter*, *Bacillus*, *Burkholderia*, *Curtobacterium*, *Enterobacter*, *Halomonas*, *Methylobacterium*, *Microbacterium*, *Methilophylus*, *Paenibacillus*, *Pseudomonas*, *Ralstonia*, *Rhodococcus*, *Rhodopseudomonas* y *Sphingopyxis*; a pesar de esta gran diversidad bacteriana, se ha demostrado que las fuertes desinfecciones de las semillas y el ambiente *in vitro* reducen drásticamente la complejidad de las comunidades (Lucero et al., 2011; Orlikowska et al., 2017).

Dentro de los efectos positivos de la presencia de microorganismos en los explantes de cultivo *in vitro* también se destacan: la promoción del crecimiento vegetal, al influenciar los procesos de desarrollo y morfogénesis *in vitro* (Orlikowska et al., 2017; Vereecke et al., 2000); el mejoramiento del enraizamiento de los brotes de algunas especies, al sintetizar fitohormonas como la rodestrina o el ácido 3-indol acético (Sunayana, Sasikala, & Ramana, 2005); la elongación de brotes axilares; la promoción de la multiplicación de los brotes (Zawadzka, Trzciński, Nowak, & Orlikowska, 2013); el aumento en el número de raíces (P. Thomas, 2004) y la promoción de la embriogénesis somática (Murthy, Vettakkorumakankav, KrishnaRaj, Odumeru, & Saxena, 1999). El conocimiento de las funciones mencionadas y muchas otras presentes en el microbioma endófito de plantas de cultivo *in vitro* podrían ayudar a aumentar la tasa de supervivencia de plantas con genotipos recalcitrantes y a que un porcentaje mayor de plantas supere el proceso de aclimatación, el cuál es un cuello de botella del proceso de micropropagación (Orlikowska et al., 2017)

Un método que pretende asegurar la producción de plantas sanas en el proceso de micropropagación es la eliminación de los brotes visiblemente infectados por bacterias u hongos. Sin embargo, la apariencia asintomática de un brote no significa que este libre de patógenos; es posible que haya desarrollado interacciones comensalistas con los patógenos, o éstos se encuentran en estado de latencia (Lucero et al., 2011). La presencia de algunas bacterias y patógenos no es aparente desde la etapa inicial de la micropropagación; algunos solo son detectados hasta la etapa de multiplicación o inclusive en la etapa de aclimatación cuando las condiciones para su crecimiento sean óptimas (Orlikowska et al., 2017; Poppenberger, Eonhardt, & Edl, 2002), a pesar de esto, la detección de las bacterias no es indispensable para que actúe sobre la planta causando daños y enfermedades (Thomas, 2011; Ye & Doak, 2011).

Metagenómica para el estudio del microbioma vegetal

El estudio del microbioma vegetal ha apuntado principalmente a responder ¿Cuáles microorganismos componen el microbioma de una planta? y ¿Qué funciones están llevando a cabo estos microorganismos?, las respuestas obtenidas en conjunto con diferentes diseños experimentales permiten obtener información sobre microbios y funciones asociados a un estado sano y de enfermedad, que puede ser aprovechada para promover buenas prácticas en los cultivos comerciales.

Anteriormente, la caracterización de las comunidades microbianas asociadas a organismos eucariotas requería una clasificación morfológica por medio de métodos de cultivo (Goodfellow et al., 1984); sin embargo, los resultados obtenidos por estos métodos, no eran representativos de toda la comunidad, pues se estima cerca del 98% de

los microorganismos no son cultivables (Lucero et al., 2011), adicional a esto, la morfología en muchos casos, no permite una detallada clasificación taxonómica (Handelsman & Handelsman, 2004). Con el surgimiento de las tecnologías de secuenciación de ultima generación del ADN se pudo dar una respuesta más concluyente a estos cuestionamientos haciendo uso de diferentes métodos como la secuenciación de genes marcadores, la secuenciación *shotgun*, la metabolómica y la metatranscriptómica.

Metataxonomía a partir del ARN 16S

El primer acercamiento al estudio del microbioma consiste en definir que microorganismos están presentes, es decir, caracterizar taxonómicamente la microbiota presente.

- Marcadores moleculares utilizados en el estudio del microbioma

Actualmente, la mayoría de los estudios que buscan la caracterización taxonómica de las comunidades microbianas asociadas a un eucariota, llevan a cabo una secuenciación masiva en sus muestras de un gen marcador ubicuo y altamente conservado, cuyas regiones hipervariables permitan distinguir polimorfismos que logren la caracterización taxonómica de los microbios (Lundberg, Yourstone, Mieczkowski, Jones, & Dangl, 2013). Los marcadores moleculares más utilizados en estudios de microbiomas son las regiones de la subunidad pequeña del ribosoma 16S en procariontes y 18S en eucariotas y para la identificación de hongos generalmente se secuencian los ITS (espaciadores transcritos internamente - *internal transcribed spacer*) (Balvočiute & Huson, 2017; Lucero et al., 2011)

El ARNr 16S, fundamental para bacterias y arqueas, es un componente de la subunidad pequeña 30S del ribosoma de los procariotas de aproximadamente 1500 pb de longitud (Hao, Pei, & Brown, 2017). Se compone de regiones conservadas, que permiten el diseño de cebadores para la amplificación del gen y regiones variables que permiten la identificación de los microorganismos a partir de la comparación con una base de datos (Lundberg et al., 2013). Las características y propiedades del ARNr 16S permiten caracterizar taxonómicamente bacterias hasta el nivel de género pudiendo llegar en algunas ocasiones hasta el nivel de especie (Hao et al., 2017).

La amplificación de este gen marcador es ampliamente utilizada debido a que selecciona exclusivamente ADN bacteriano, excluyendo el ADN eucariota para la amplificación y brinda información sobre no solamente la taxonomía de los microorganismos, sino que también permite medir la abundancia de cada taxón (Hao et al., 2017). Sin embargo, el ARN 16S presenta algunas limitaciones que pueden sesgar los resultados taxonómicos y de diversidad. Una de ellas es su longitud, la cuál se vuelve insuficiente en muchos casos para la clasificación de los taxones hasta el nivel de especie o cepa (Hao et al., 2017). La capacidad de ser transferido horizontalmente entre taxones cercanos, puede causar sobreestimaciones de diversidad (Acinas, Marcelino, Klepac-Ceraj, & Polz, 2004). También se ha encontrado que en algunas ocasiones este gen no tiene la resolución necesaria para distinguir entre especies cercanas que contienen secuencias idénticas de los genes del 16S (Carlos, Tang, & Pei, 2012). La escogencia de las regiones del ARN 16S a amplificar dependerán de la longitud promedio de lecturas permitido por el equipo de secuenciación y la precisión de clasificación que se desee tener (Nossa et al., 2010).

-Plataformas para la producción de datos de microbiomas

Para identificar los microorganismos que integran la microbiota asociada a una planta, es decir, su perfil taxonómico, tradicionalmente se habían utilizado métodos de cultivo, los cuales, aunque permiten caracterizar bacterias hasta un nivel taxonómico más profundo (cepa) (Mendes et al., 2011) y revelan características más concluyentes acerca de microorganismos específicos dentro de la microbiota (Bakker et al., 2013), limitan los estimados de diversidad al no capturar la totalidad de microorganismos presentes en las muestras (Lebeis, 2014), debido a que la mayor proporción de microorganismos del suelo son no cultivables (Doornbos, Van Loon, & Bakker, 2012; Hirsch, Mauchline, & Clark, 2010). La técnica de cultivo más utilizada consistía brevemente en clonar en un vector los genes del ARNr 16S presentes en una población microbiana, transformar las bacterias, sembrarlas para seleccionar las colonias transformadas, purificar el plásmido y finalmente secuenciar con Sanger los aislados obtenidos (Carlos et al., 2012). A pesar de que el método de Sanger (primera generación de técnicas de secuenciación) produce lecturas de gran longitud y con una baja tasa de error, este tipo de secuenciamiento junto con la técnica de clonación quedo en desuso debido a que son muy laboriosas, tienen un costo de secuenciación alto y limitan la diversidad de las muestras a las pocas secuencias obtenidas por cultivables.

La reciente creación de herramientas para la secuenciación de última generación del ADN (next generation sequencing - NGS) ha permitido ampliar considerablemente el conocimiento acerca de la amplia gama de microorganismos (bacterias, arqueas, hongos y virus) asociados a un hospedero y las funciones que cumplen en éste, partiendo de un volumen más amplio de secuencias, mayor sensibilidad especialmente para la detección

de secuencias poco abundantes, y la posibilidad de secuenciar miles de genes simultáneamente (www.Illumina.com) (Hao et al., 2017; Lebeis, 2014).

Las tecnologías NGS más utilizadas para la secuenciación del ARNr 16S son PacBio (Pacific Biosciences) (Rhoads & Au, 2015) y los sistemas MiSeq y HiSeq de Illumina (<https://www.illumina.com/systems/sequencing-platforms.html>), siendo estos últimos los más utilizados (Beckers et al., 2016; Hao et al., 2017). La tecnología Illumina lleva a cabo una secuenciación por síntesis y comienza con la fragmentación aleatoria del ADN, seguida de la unión de adaptadores a ambos lados de las secuencias y la unión de estos a otros adaptadores complementarios, que posteriormente se unen a una placa sólida; allí cada secuencia se amplifica mediante la “amplificación de puente de PCR” (PCR bridge amplification), la cual genera alrededor de un millón de copias de cada secuencia. El proceso continúa con la adición de un mix que contiene los cebadores, la polimerasa de ADN y nucleótidos modificados (nucleótidos marcados con fluorescencia y un grupo 3'-hidroxilo inactivo que garantiza que solo se una un nucleótido) y comienza el proceso de síntesis de la secuencia complementaria. Una luz específica para cada nucleótido es producida luego de que los grupos de secuencias se exciten con un laser, el proceso finaliza con la eliminación del terminador con la fluorescencia y el inicio de un nuevo ciclo con una nueva incorporación (Reuter, Spacek, & Snyder, 2015). La tasa de error de Illumina es la más baja de las tecnologías comúnmente utilizadas, siendo igual al 1% y consistiendo en su mayoría de sustituciones de nucleótidos (Dohm, Lottaz, Borodina, & Himmelbauer, 2008).

Las plataformas de Illumina permiten la multiplexación de muestras, es decir, la secuenciación simultánea de un gran número de muestras en una única corrida por medio

de la adición de secuencias “códigos de barra” (*barcodes*) durante la preparación de las librerías, que permiten marcar cada fragmento e identificarlo antes del análisis de los resultados finales con su muestra de origen (<https://www.illumina.com>). La secuenciación en los sistemas de Illumina, se obtienen a un bajo costo y con una alta resolución. Adicionalmente, se puede realizar una secuenciación de pares (paired-end sequencing) en la que secuenciando ambos extremos de un fragmento se pueden obtener secuencias alineadas de alta calidad con una longitud de 2 x 150 pb y 2 x 300 pb (HiSeq - MiSeq) capaces de detectar *indels* (variantes de inserción-eliminación) (<https://www.illumina.com>) y llegar a identificar las secuencias hasta niveles taxonómicos como género y especie (Hao et al., 2017). Sin embargo, es importante contar con herramientas bioinformáticas que permitan procesar las secuencias resultantes de esta tecnología, debido a que durante la secuenciación se presentan errores que se suman a posibles errores ocurridos durante la amplificación por PCR (Hao et al., 2017).

-Aproximaciones y programas bioinformáticos para el procesamiento de secuencias de microbiomas

Actualmente existen dos acercamientos ampliamente utilizados para el procesamiento de datos crudos provenientes de secuenciación por plataformas Illumina. Se distinguen a partir del método que utilizan para corregir o eliminar errores en los amplicones producidos durante la amplificación o el secuenciamiento.

El fundamento del método tradicional para el procesamiento de lecturas comenzó a implementarse en el año 1995 para facilitar el manejo informático de la gran cantidad de datos que provienen del microbioma (Bond, Hugenholtz, Keller, & Blackall, 1995). Con el subsecuente desarrollo de diferentes softwares y protocolos bioinformáticos se fue

consolidando hasta formar parte de un flujo de trabajo de seis pasos generales. Para el caso de lecturas pareadas demultiplexadas provenientes de secuenciación en plataformas Illumina, el flujo de trabajo se muestra en la figura 1 y consiste brevemente en un filtrado de secuencias a partir de su información de calidad y un puntaje de calidad Phred deseado (medida de la calidad de la identificación de las bases nucleotídicas generadas por la secuenciación de ADN); continua con el ensamble o unión de las secuencias pareadas forward y reverse correspondientes que comparten una cantidad de bases comunes, identificadas al sobrelapar las secuencias resultando en secuencias consenso por cada pareja de secuencias; la construcción de clusters o grupos de secuencias que comparten un porcentaje de similitud establecido, denominados unidades taxonómicas operativas (OTUs), la identificación y remoción de secuencias quiméricas (amplicones que contienen información de dos secuencias parentales diferentes) (Wani et al., 2015) y la asignación taxonómica que se realiza al comparar las secuencias consenso obtenidas con una base de datos de ribosomas.

Se han desarrollado diferentes paquetes de software como QIIME (Caporaso et al., 2010) y Mothur (Schloss et al., 2009) que integran en una sola plataforma todas las funciones necesarias para llevar a cabo los seis pasos de procesamiento de la figura 1, además incorporan otras funciones para el análisis y la visualización de los resultados como pruebas estadísticas y diferentes representaciones gráficas. Estos programas son ampliamente utilizados en los estudios de la microbiota (Panke-Buisse, Poole, Goodrich, Ley, & Kao-Kniffin, 2015; Peiffer et al., 2013); a pesar de esto, algunos investigadores construyen su propio flujo de trabajo a partir de herramientas especializadas en cada paso

del proceso para tener mayor control de los parámetros o aprovechar funciones especiales de ciertos paquetes.

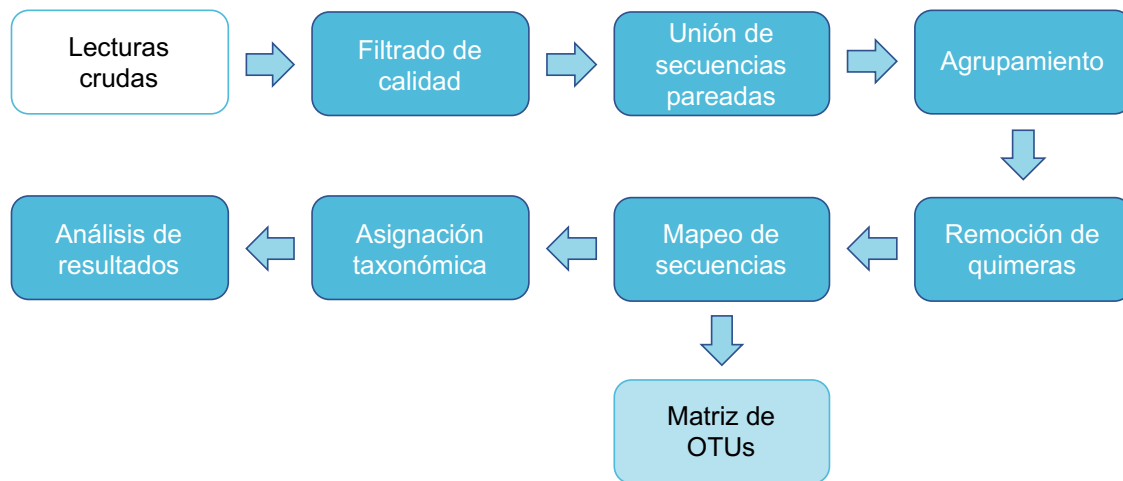


Figura 1. Flujo de trabajo con los pasos generales para el procesamiento de lecturas pareadas provenientes de las plataformas Illumina y la metodología de construcción de OTUs.

Dentro de los paquetes más utilizados para el filtrado de lecturas pareadas se encuentran: BBDuk (<https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbdduk-guide/>) y Trimmomatic (Bolger, Lohse, & Usadel, 2014) que son dos herramientas para la filtración y el trimaje de las lecturas que permiten la remoción de partes no biológicas como cebadores; el corte de las lecturas a partir de una longitud específica y la eliminación de lecturas que no alcancen un puntaje de calidad determinado. Para el filtrado de calidad, BBDuk se basa en el puntaje Phred promedio de una lectura y cuenta con la capacidad de eliminar contaminantes por medio de la comparación de k-mers (todas las posibles subsecuencias de una lectura obtenida a través de la secuenciación de ADN). Por otro lado, Trimmomatic lo desarrolla por medio del parámetro de “*slidingwindow*” el cual, escanea las lecturas cada x bases y las trima cuando el puntaje de las x bases cae por debajo de un puntaje Phred y .

El ensamble de secuencias pareadas se realiza buscando lecturas forward y reverse que tengan un mínimo de bases superpuestas. Un programa ampliamente usado para esto es PANDAseq (<https://github.com/neufeld/pandaseq>), el cuál permite el ensamble de lecturas pareadas provenientes de Illumina con una superposición de 10 bases por defecto y también cuenta con funciones para el corte de cebadores y el filtro de secuencias.

El agrupamiento de lecturas con un porcentaje determinado de similitud, en un principio se realizó para reducir el número de secuencias a analizar (Bond et al., 1995), pero en los años siguientes, se encontró que este agrupamiento en clusters, servía como método de corrección de errores de los amplicones y para caracterizar taxonómicamente las secuencias (Rosen, Callahan, Fisher, & Holmes, 2012). Los clusters de secuencias con un porcentaje de similitud establecido o mayor se denominan unidades taxonómicas operativas (*Operational Taxonomic Units* - OTUs), generalmente se construyen al 97% de similitud (Větrovský & Baldrian, 2013), sin embargo, se ha encontrado que hay especies pertenecientes a un mismo género cuyas secuencias tienen una similitud mayor al 97% (Větrovský & Baldrian, 2013) e inclusive una similitud igual al 99% es insuficiente para diferenciar especies cercanamente emparentadas como diferentes miembros de la familia Enterobacteriaceae, Clostridiaceae y Peptostreptococcaceae (Jovel et al., 2016). El proceso de agrupamiento de secuencias, en inglés llamado *OTU picking*, se puede realizar de 3 formas de acuerdo a la utilización de una base de datos de referencia. La selección de OTUs *de novo* consiste en agrupar las secuencias teniendo en cuenta cual similares son con las otras y sin usar una base de datos de referencia, generalmente se usa cuando se secuencian genes marcadores poco comunes; en la

selección de OTUs con referencia cerrada las secuencias a analizar se agrupan teniendo en cuenta la información de secuencias presente en una base de datos y las lecturas que no tengan una secuencia correspondiente o cercana en la base de datos son descartadas y finalmente, la selección de OTUs con referencia abierta combina los enfoques de las dos anteriores en las que se comparan las secuencias con una base de datos de referencia y las que no tengan su correspondiente se agrupan *de novo* (Caporaso et al., 2010).

La remoción de secuencias quiméricas es un paso importante para evitar resultados sesgados en los cálculos de diversidad alfa (Goodrich et al., 2014). Las quimeras son amplicones mal ensamblados que aparentan ser secuencias recombinantes por lo que muchas veces son difíciles de identificar (Haas et al., 2011; Wylie et al., 2012). Actualmente existen diferentes herramientas que permiten su identificación y eliminación del set de lecturas a analizar. Uno de ellos es UCHIME, un algoritmo con alta sensibilidad que detecta secuencias quiméricas con dos o más segmentos a partir de la comparación con una base de datos de secuencias libres de quimeras o alternativamente detecta quimeras *de novo* mediante la exploración de datos de abundancia (Edgar, Haas, Clemente, Quince, & Knight, 2011).

La asignación taxonómica se realiza a partir del alineamiento de las secuencias consenso obtenidas con secuencias presentes en una base de datos de referencia utilizando herramientas de alineamiento como SINA (Pruesse, Peplies, & Glöckner, 2012) o utilizando técnicas basadas en k-mers (Vinje, Liland, Almøy, & Snipen, 2015). Generalmente para este tipo de estudios se utilizan bases de datos específicas para el 16S como Greengenes (DeSantis et al., 2006), RDP (Maidak et al., 2001) y Silva (Quast et al., 2012). Estas bases de datos son compilados de taxonomías producidas en diferentes

proyectos que difieren en tamaño y resolución y brindan información para clasificar las secuencias en mínimo seis rangos: dominio, filo, clase, orden, familia y género (Balvočiute & Huson, 2017). Greengenes (GG) es una base de datos curada automáticamente que se compone de secuencias pertenecientes al dominio Bacteria y Archaea tomadas de bases de datos públicas que han superado un proceso de filtrado. Su última versión, 13_8 fue publicada en Agosto del 2013 (http://qiime.org/home_static/dataFiles.html). Por otra parte, el proyecto de base de datos ribosomal (*Ribosomal database project* - RDP) es una base de datos de secuencias pertenecientes a bacterias, arqueas y hongos, la última versión (11.5) contiene 3,356,809 ARNr 16S y 125,525 28S ARNr de hongos (<https://rdp.cme.msu.edu/>). Y finalmente, Silva es una base de datos curada manualmente que se compone de secuencias del dominio Bacteria, Archaea y Eukarya principalmente de las subunidades pequeñas de ARNr (Quast et al., 2012); la versión 128 fue publicada el 29 de septiembre del 2016 y cuenta con 1,922,213 secuencias y la última versión, 132 del 13 de diciembre del 2017 contiene 2,090,668 secuencias (<http://www.arb-Silva.de/documentation/release-132/>).

Recientemente se ha demostrado que el agrupamiento de secuencias en clusters produce una inflación en el número de OTUs en las muestras conllevando a sobreestimaciones de diversidad (Huse, Welch, Morrison, & Sogin, 2010). Esto junto con el hecho de que las OTU subutilizan la calidad de la secuenciación moderna al impedir la posibilidad de resolver variaciones de pequeña escala (Callahan, McMurdie, et al., 2016; Huse et al., 2010) han impulsado la creación de nuevos enfoques en los que se busca disminuir la tasa de error sin adicionar los sesgos en la clasificación taxonómica producidos por la construcción de OTUs. Callahan *et al.*, 2016 desarrollaron un paquete

de software, el cual se compone de diferentes funciones que llevan a cabo todos los pasos de procesamiento mencionados anteriormente e incorporan dos nuevos pasos de modelado y corrección de errores; y de inferencia de muestras, que permiten reconstruir con un alto grado de precisión, una comunidad microbiana y disminuir la tasa de falsos positivos (Callahan, Mcmurdie, Rosen, Han, & A, 2016). DADA2 se deriva del “algoritmo divisorio de denoising de amplicón” – DADA (Rosen et al., 2012) y es utilizado para corregir los errores de los amplicones producidos por los sistemas de Illumina, contrario a su versión anterior DADA que procesa lecturas provenientes de secuenciación 454.

El pipeline de DADA2, mostrado en la figura 2, recibe como input las lecturas crudas provenientes del secuenciamiento y comienza su procesamiento con un paso de filtrado de calidad similar al realizado por Trimmomatic y BBDuk, el cual corta las lecturas a una longitud especificada eliminando las lecturas con una longitud inferior y filtra según el número de bases ambiguas, un puntaje de calidad mínimo y los errores esperados en una lectura (Edgar & Flyvbjerg, 2015). Luego, realiza la estimación de la tasa de error, un parámetro fundamental para posteriormente correr el algoritmo núcleo de DADA2.

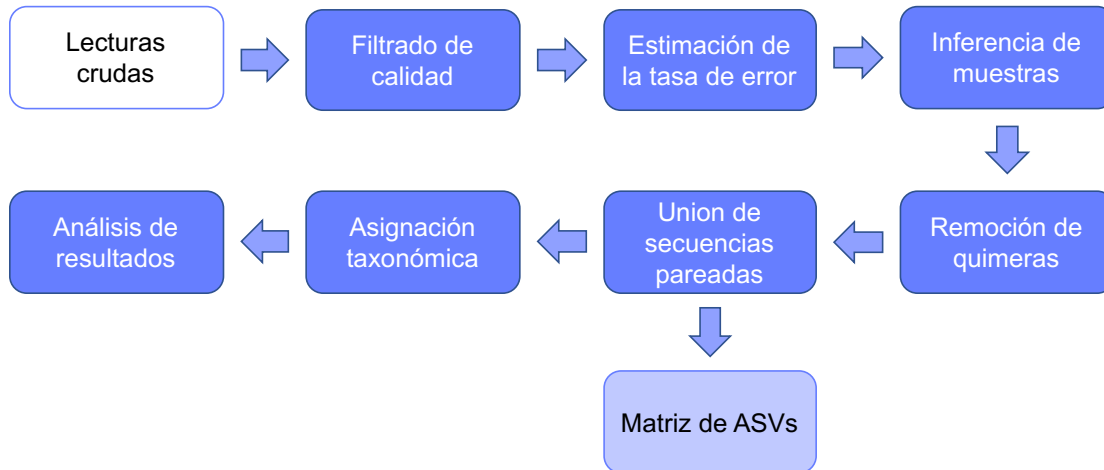


Figura 2. Flujo de trabajo con los pasos generales para el procesamiento de lecturas pareadas provenientes de las plataformas Illumina con el paquete DADA2 de Callahan et al., 2016

El modelo de error calcula la tasa a la cual una lectura de secuencia i proviene de una secuencia muestra j y su p-value de abundancia asociado, cuantifica la noción de que la secuencia i es muy abundante para ser producto de errores ocurridos en la amplificación o secuenciación. Seguidamente se realiza la inferencia de las muestras empleando el algoritmo de partición divisiva. El proceso comienza con la *dereplicación* de las lecturas seguido de la construcción de particiones en las que empleando las tasas de error y los p-values de cada secuencia, se ubica cada lectura única en una única partición claramente diferenciable de las otras secuencias. Este método también se caracteriza por no inferir secuencias *singleton* debido a que les asigna un p-value igual a 1 y no se incorporan a ninguna partición. El pipeline continúa con la remoción de *bimeras* (quimeras) por medio de la identificación de secuencias que contienen información de dos secuencias parentales y concluye con el ensamble de las secuencias forward y reverse con una estricta superposición de 12 bases a partir de un alineamiento

de secuencias (Callahan, McMurdie, et al., 2016). Este método, además de que infiere exactamente las secuencias de estudio y puede resolver diferencias entre secuencias de tan solo un nucleótido; identifica más variantes reales que otros métodos; produce menor número de secuencias equivocadas, lo cual impacta positivamente los análisis posteriores de diversidad, permitiendo que algunas secuencias se puedan caracterizar hasta el nivel de especie; y es altamente escalable, pues al no basarse en comparaciones con bases de datos, puede ser aplicado en cualquier locus (Callahan, McMurdie, et al., 2016).

-Análisis aplicados en los estudios taxonómicos de microbiomas

Una de las propiedades que describe cuantitativamente una comunidad microbiana es la diversidad, la cual según Whittaker 1972 se puede calcular en tres niveles: diversidad alfa para determinar la diversidad dentro de un hábitat o una muestra, diversidad beta para comparar la diversidad entre hábitats o muestras y la diversidad gamma para caracterizar la diversidad en un ecosistema (Whittaker, 1972). La diversidad se cuantifica por medio del cálculo de índices que describen propiedades generales de las comunidades que permiten la comparación de taxones o hábitats (Morris et al., 2014). Actualmente QIIME (Caporaso et al., 2010) tiene disponibles más de 40 métricas para determinar la diversidad alfa y beta, sin embargo, ninguna de ellas resume en totalidad la diversidad de una muestra o hábitat por lo que generalmente en los estudios de comunidades microbianas se reportan varios índices en conjunto (Bodenhausen et al., 2013; Coleman-Derr et al., 2016; Copeland, Yuan, Layeghifard, Wang, & Guttman, 2015).

La diversidad alfa estima la diversidad dentro de una comunidad o una muestra teniendo en cuenta la riqueza (número total de especies observadas) y su abundancia

relativa (uniformidad) (Hao et al., 2017). Dentro de las métricas comúnmente utilizadas para estimarla en la microbiota vegetal, se encuentran el número de OTUs/ASVs, el índice de Shannon, el de Simpson, la uniformidad o equitabilidad y la diversidad filogenética (Edwards et al., 2015; Gopal et al., 2017). El número de OTUs/ASVs mide directamente la riqueza específica dentro de cada muestra, sin embargo, se diferencia del número de especies observadas porque toma como unidades diferentes todas las OTUs/ASVs que fueron identificadas como la misma especie. El índice de Shannon se define como la suma ponderada de la proporción de cada especie en relación con el número total de especies en la comunidad bacteriana, por lo que este índice representa conjuntamente la abundancia y la uniformidad (Shannon, 1948). El componente de uniformidad del índice de Shannon puede ser analizado por separado para cuantificar la distribución de los individuos en los grupos taxonómicos identificados, es decir, que permite identificar si hay dominancia de un grupo o algunos grupos taxonómicos en una muestra o hábitat (Sheldon, 1969). Y el índice de diversidad filogenética, el cuál brinda una idea de los patrones generales de variación a nivel de genes de los taxones identificados a partir de la medición de su distancia filogenética (Faith & Baker, 2006).

La diversidad beta permite comparar la diversidad de diferentes comunidades bacterianas, evidenciando que tan diferentes son entre sí y revelando aspectos ecológicos que no son evidentes al observar la composición de las muestras individuales (Bray & Curtis, 1957). Las métricas de diversidad beta se pueden clasificar en cuantitativas si usan datos de abundancia para los cálculos como lo hace Bray-Curtis o UniFrac ponderado o cualitativas si consideran exclusivamente la presencia-ausencia de secuencias como lo hacen las métricas Jaccard binario o UniFrac sin ponderar (Goodrich

et al., 2014; Kraemer, Ramette, Aebi, Oppliger, & Hilty, 2018). También se pueden clasificar en si basan los cálculos en una filogenia como las métricas de UniFrac o no la toman en cuenta como Bray-Curtis (Bray & Curtis, 1957). Los resultados cuantitativos de las métricas de diversidad beta se grafican haciendo uso de alguna técnica para el análisis multivariado que permiten visualizar tendencias generales de similitud entre muestras y asociarlo con metadatos del estudio (Weiss et al., 2017)

Los datos de abundancia provenientes de estudios metataxonómicos de microbiomas se caracterizan por ser altamente dispersos, contener una gran cantidad de ceros y tener librerías con profundidades de secuenciación diferentes (Xu, Paterson, Turpin, & Xu, 2015). Estas propiedades pueden causar resultados erróneos o sesgos en los cálculos de diversidad (Goodrich et al., 2014; Weiss et al., 2017); es por esto que se han desarrollado diferentes métodos de normalización que ayudan a reducir los sesgos en futuros análisis.

La comparación de la profundidad de secuenciamiento entre las muestras de estudio tradicionalmente se realizaba para preparar los datos para los cálculos de diversidad alfa, sin embargo, esta metodología también es comúnmente utilizada para los cálculos de diversidad beta (Horner-Devine, Lage, Hughes, & Bohannan, 2004). Una diferencia muy grande entre las profundidades requiere la utilización de un método de normalización de datos. El más utilizado es la normalización por rarefacción que consiste en eliminar aleatoriamente lecturas de las muestras hasta que todas las librerías tengan la misma profundidad (Simberloff, 1972). Este método aunque puede llevar a pérdidas de información importante, permite hacer comparaciones de diversidad sin asumir como verdaderas variaciones biológicas las variaciones de profundidad (Dhariwal et al., 2017).

En estudios donde las profundidades de secuenciación son muy diferentes, de hasta 1 orden de magnitud se opta por eliminar completamente las muestras con baja profundidad para conservar más información de otras muestras y no inflar los estimados de diversidad beta (Weiss et al., 2017). Para asegurar que la profundidad de rarefacción escogida en un estudio brinda datos informativos en todas las muestras se construyen curvas de rarefacción, que relacionan los conteos muestreados (profundidad de rarefacción) y valor esperado de la diversidad de especies (Gotelli & Colwell, 2001; Weiss et al., 2017).

La gran dispersión de los datos y la alta abundancia de ceros presentes en la matriz de datos se debe a que las secuencias representativas (OTUs/ASVs) son dependientes de la muestras y generalmente muy pocos taxones, solo aquellos núcleo, se comparten entre muestras (Rosado, Pérez-Losada, Severino, Cable, & Xavier, 2019; Xu et al., 2015), De esta manera, los ceros pueden indicar ausencia física de los taxones en las muestras o en su defecto que la profundidad de secuenciación no fue suficiente para detectarlos (L. Chen et al., 2018). Este último caso puede llevar a obtener una diversidad beta inflada, pues consideraría secuencias compartidas como únicas de una muestra, y también, a obtener un microbioma núcleo con un tamaño inferior al real (L. Chen et al., 2018). Se han propuesto diferentes métodos de normalización para lidiar con este aspecto dentro de los que se destacan la normalización CSS (Paulson, Stine, Bravo-Corrada, & Pop, 2013) y la normalización GMPR (L. Chen et al., 2018).

La metodología CSS perteneciente al paquete Bioconductor metagenomeSeq, se basa en una normalización y un modelo estadístico que explican el submuestreo. Consiste en dividir los conteos de lecturas (*read counts*) por la suma acumulativa de los conteos hasta un percentil determinado que permite capturar la distribución de los recuentos

invariables. Sin embargo, esta metodología enfrenta problemas al seleccionar el percentil cuando los datos tienen una alta variabilidad. La normalización GMPR, implementada como un paquete de R, se basa en el supuesto de que hay una gran parte invariable en los datos de abundancia. Consiste en realizar una comparación por pares de muestras en las que calcula la proporción promedio de abundancias diferentes de cero y luego estos resultados se combinan para calcular el factor de tamaño de cada muestra por la que después se dividen todas las abundancias para obtener los valores normalizados (L. Chen et al., 2018). Esta metodología supera la limitación de la normalización CSS pues no se fundamenta en la variabilidad de los datos.

Aplicar un buen método de normalización es indispensable para reducir o eliminar sesgos y variaciones ocasionados por los procesos de muestreo, extracción de ADN y secuenciamiento y permite revelar la verdadera información biológica presente en los datos (Weiss et al., 2017).

Estudio de las capacidades metabólicas del microbioma vegetal

-Secuenciación tipo *shotgun*

El segundo acercamiento al estudio del microbioma se hace a partir de la secuenciación metagenómica *shotgun*. Este tipo de secuenciación en lugar de estar dirigida a un locus específico para la amplificación, secuencia todo el ADN de los microorganismos presentes en la comunidad al fragmentarlo en pequeñas fracciones y secuenciarlas individualmente. Esto resulta en información taxonómica, de locus genómicos informativos como el 16S, e información funcional, provista por las regiones codificantes (Sharpton, 2014). La capacidad funcional de una comunidad microbiana se identifica por medio de la predicción de genes y la anotación de secuencias, la primera

identifica entre las lecturas genómicas, las secuencias que codifican para proteínas y la segunda compara las predicciones con alguna base de datos de genes, proteínas, familias de proteínas o rutas metabólicas (Sharpton, 2014).

La base de datos más utilizada para la clasificación de genes es KEGG (Kyoto Encyclopedia of Genes and Genomes). Ésta, vincula información genómica con información funcional de orden superior como rutas metabólicas. Se divide a su vez en varias bases de datos especializadas en genes, rutas metabólicas, módulos, enzimas, etc que organizan la información en diferentes formatos (Kanehisa & Goto, 2000). Para realizar la predicción de funciones de genes se utiliza la base de datos KO (KEGG Orthology) que se basa en los ortólogos funcionales para determinar funciones moleculares.

El tipo de secuenciación *shotgun* evidencia varias ventajas sobre el método de caracterización taxonómica por medio del 16S. Con la secuenciación *shotgun* se puede identificar el perfil taxonómico de la comunidad microbiana total (bacterias, hongos, virus, etc), pues no depende de genes marcadores que son específicos para cierto grupo de microorganismos (Hodkinson & Grice, 2015); brinda estimaciones de abundancia más precisas al no realizar amplificación de genes (Dewdney et al., 2000) y brinda una imagen más detallada del microbioma asociado a un organismo, al dar respuesta a las preguntas ¿quiénes están ahí? Y ¿qué pueden estar haciendo? Sin embargo, esta tecnología también enfrenta múltiples retos. Para el ensamble de los genomas de cada microorganismo se debe hacer una clasificación de las lecturas donde se pueden comparar diferentes locus de diferentes genomas, haciendo que este paso sea fundamental en la obtención de resultados verdaderos. Además, la clasificación puede ser poco

confiable debido al limitado número de genomas de referencia disponibles hasta el momento (Hodkinson & Grice, 2015). Caso similar al que se puede presentar durante la predicción de genes, al no encontrar homología entre las secuencias de estudio y las secuencias de referencia (Sharpton, 2014).

-Métodos alternativos para el estudio de las capacidades funcionales del microbioma

Actualmente se han desarrollado métodos alternativos para obtener la información funcional brindada por la secuenciación *shotgun* a partir de las secuencias del 16S.

PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) es uno de los programas desarrollados para este propósito; cuenta con dos versiones disponibles; PICRUSt v1.1.1 el cual hace la predicción de familias de genes por medio de un algoritmo de reconstrucción de estado ancestral extendido, una base de datos de referencia e identificadores de la base de datos de Greengenes (Langille et al., 2013). Y la versión v2.0.3-b (PICRUSt2) (<https://github.com/picrust/picrust2/wiki>), todavía en estado beta, predice funciones de genes usando el algoritmo de predicción de estado oculto que actúa sin necesidad de identificadores propios de una base de datos de referencia, por lo que tiene la capacidad de hacer predicciones para secuencias que no han sido identificadas taxonómicamente. PICRUSt2 recibe una matriz de abundancia con identificadores personalizados y comienza el flujo de trabajo con la colocación de secuencias (*sequence placement*) en el que las secuencias son sometidas a un alineamiento de secuencias múltiples (MSA) con un set de alineamientos provenientes de la base de datos IMG (<https://img.jgi.doe.gov/>) haciendo uso de la herramienta HMMER (Eddy, 1998) que está diseñada para detectar homólogos remotos con la mayor sensibilidad posible, basándose en la fuerza de sus modelos de probabilidad subyacentes

que aplican un método estadístico para estimar la verdadera frecuencia de un residuo en una posición dada en la alineación desde su frecuencia observada. Luego ubica las secuencias en una filogenia de referencia por medio de las herramientas EPA-ng (Barbera et al., 2018) y GAPPa (<https://github.com/lczech/gappa>), el resultado de estos pasos consiste en un árbol filogenético con las secuencias de estudio en formato newick. El siguiente paso se realiza con el paquete castor de R (Louca & Doebeli, 2018), en este se realiza la predicción de estado oculto, el cual es un método filogenético para la predicción de rasgos desconocidos a partir de una colección de rasgos conocidos y una filogenia. Este algoritmo predice el número de copias del 16S, número de comisiones de enzimas y otólogos KEGG de microorganismos modernos a partir de información de número de copias y funciones anotados en las bases de datos y una filogenia (Zaneveld & Thurber, 2014). Los resultados del número de copias del 16S se producen con un valor NSTI asociado (Índice de taxón secuenciado más cercano) que cuantifica que tan diferente es cada secuencia de estudio con respecto a las secuencias de referencia. Un NSTI mayor a 1 indica que la secuencia está muy distante de las secuencias de referencia por lo cual es recomendado eliminar la secuencia de los análisis posteriores para asegurar que las predicciones sean confiables. El paso siguiente de predicción de metagenomas, comienza con la normalización de la matriz de secuencias representativas de acuerdo al número de copias del 16S, luego con esas abundancias, se procede a predecir los perfiles funcionales por muestra. Con la información de las familias de genes presentes en las muestras se puede predecir la abundancia de rutas metabólicas con MinPath al comparar las diferentes enzimas identificadas con la base de datos curada MetaCyc, que contiene

rutas metabólicas experimentalmente dilucidadas de todos los dominios de la vida (Ye & Doak, 2011)

Un método alternativo para la identificación de rutas metabólicas a partir de las familias de genes identificadas por PICRUSt, es utilizar la herramienta web MicrobiomeAnalyst (Dhariwal et al., 2017), que a diferencia de PICRUSt contiene herramientas para el análisis estadístico y gráfico de los metagenomas que identifica. Esta herramienta se compone de 4 módulos que corresponden a diferentes análisis aplicables a datos de microbiomas, siendo el segundo el que contiene las herramientas para analizar datos provenientes de secuenciación *shotgun* o de predicciones de familias de genes (<https://www.microbiomeanalyst.ca/>). Para el análisis de las familias de genes, se cargan individualmente o en una matriz de abundancia con anotaciones de ortólogos KEGG (KO), número de comisiones de enzimas (EC) o grupos de clusters de ortólogos (COG). En el paso siguiente, se muestran los diferentes métodos de filtrado o normalización que se le pueden aplicar a los datos, donde están filtros por baja abundancia, rarefacción y normalización CSS los cuales son aplicados de acuerdo al criterio del usuario. Y finalmente las rutas metabólicas son asignadas al comparar las familias de genes con rutas presentes en la base de datos de rutas metabólicas de KEGG (Dhariwal et al., 2017).

Musa acuminata AAA como modelo de estudio del microbioma endófito de plantas in vitro

-El cultivo del banano

El banano es una hierba perenne de gran tamaño perteneciente a la familia Musaceae y al género *Musa* siendo la *Musa acuminata* la que ha dado origen a las

variedades comercializadas actualmente. Se caracteriza por tener un pseudotallo, en lugar de un tallo aéreo verdadero, construido a partir de vainas foliares, que puede alcanzar los 30 cm de diámetro y los 7 m de altura. Posee rizomas, comúnmente llamados cormos que son estructuras subterráneas de las cuales se desarrollan las hojas y las inflorescencias. Su floración se da aproximadamente después de 10 meses del nacimiento del pseudotallo y consiste en un gran capullo púrpura que, al abrirse, descubre una estructura en forma de espiga, sobre cuyo tallo axial se disponen en espiral hileras dobles de flores, agrupadas en racimos. El fruto, una falsa baya epigínea de 7 a 30 cm de largo se organiza en forma de un racimo compacto que tarda de 80 a 180 días en desarrollarse; la pulpa es rica en almidón y dulce y en plátanos tiene contenido de látex (Generalidades del cultivo de banano en Colombia, 2011).

Actualmente, existen más de 1000 variedades de bananos producidas y consumidas a nivel mundial, sin embargo, la variedad *Cavendish* es la más comercializada y producida globalmente (FAO Trade and markets, 2018); más específicamente, el cultivar Williams, el cuál pertenece al subgrupo *Cavendish*. Se estima que en el año 2017 se produjeron 114 millones de toneladas de banano, provenientes principalmente de cultivos presentes en Asia, Latinoamérica y África. Los principales productores de banano en el mundo son India y China, sin embargo, las frutas producidas allí son utilizadas principalmente en el comercio interno. El principal exportador de esta fruta es Ecuador, el cual representa un tercio del total de exportaciones, le siguen Filipinas, Costa Rica, Guatemala y Colombia, el cuál representa el 10% de las exportaciones totales de banano (FAO Trade and markets, 2018).

El banano es un alimento altamente energético con un alto contenido de vitaminas y minerales que aporta 460 calorías por cada 100 gr; juega un papel fundamental en la seguridad alimentaria de muchos países, debido a que tiene una gran participación en la dieta de los habitantes de países productores e importadores y representa gran parte de la ingesta calórica de las personas en diferentes países de África, América Central y del sur y este de Asia (Caracterización del subsector bananero en Colombia, SENA; FAO Trade and markets, 2018). Adicionalmente, es de gran importancia a nivel económico en algunos países en vía de desarrollo en los cuales, estos cultivos representan hasta el 75% del ingreso mensual de los pequeños agricultores (FAO Trade and markets, 2018).

Los cultivos de banano son afectados por diferentes enfermedades y daños superficiales provocados por nematodos, insectos, hongos, bacterias y virus. Se han identificado 7 nematodos que afectan los cultivos de banano; estos organismos atacan generalmente los órganos subterráneos causando debilidad en la planta y llevándola en ocasiones hasta la muerte. Comúnmente los insectos atacan los órganos aéreos de la planta, como lo es el caso del picudo del banano o picudo negro (*Cosmopolites sordidus*), que afecta la integridad de los pseudotallos; los trips como el trips del banano (*Chaetanaphothrips orchidii*), que afectan la cáscara del banano; la polilla del banano, (*Nacoleia octasema*), que afecta las inflorescencias y el ácaro del banano (*Tetranychus lambi*) que afecta las hojas y el fruto. Por otra parte, las afecciones que padece el banano producidas por hongos son las más drásticas y virulentas; se han registrado diversas micosis producidas por hongos del género *Mycosphaerella*, como la sigatoka amarilla, causada por *Mycosphaerella musicola*, la cuál produce manchas en las hojas que se

extienden por toda la planta y al mismo tiempo reduce el sistema radicular y produce una fruta ácida de menor tamaño. El caso más grave es la enfermedad de la sigatoka negra producida por *Pseudosercospora fijiensis* que se transmite por el viento y comienza con la necrosis de las hojas hasta provocar la caída de los frutos, su control requiere cuatro veces más fungicida que cualquier otro control de sigatoka (Generalidades del cultivo de banano en Colombia, 2011). Otros hongos patógenos son *Fusarium oxysporum* y *Gloeosporium musarum*, que afectan los órganos aéreos de la planta. Otras afecciones del banano, del cual no se conoce cura son las producidas por bacterias como el hereque o moko bacteriano producido por *Ralstonia solanacearum* y el BBTV (*banano bunchy top virus*) el cual en la mayoría de las variedades inhibe por completo la formación de frutos.

El sector bananero es de gran importancia para un país como Colombia porque genera divisas que contribuyen el 1,5% del PIB agrícola y genera alrededor de 296 mil empleos directos e indirectos. Colombia tiene dos zonas bananeras tradicionales ubicadas en el departamento del Magdalena y en la región de Urabá (Antioquia), hasta la fecha, las zonas productoras se han expandido hasta encontrarse también en los departamentos del Valle del Cauca, Guajira, Cundinamarca, Quindío, Tolima, Chocó y Huila. Alrededor del 91% de la producción nacional de banano es destinada a exportaciones principalmente a Bélgica y Estados Unidos. Para el 2018 se estimó que el área de cosecha fue de 93,565 ha con un rendimiento de 18.70 Ton/ha este rendimiento en nuestro país se ve afectado por fenómenos climáticos y por afectaciones de tipo fitosanitarios (Cadena de banano Indicadores e instrumentos. 2018).

Actualmente la mayoría de las variedades comerciales de banano se propagan vegetativamente debido a que la tasa de producción de retoños es muy baja, de máximo

10 retoños al año por planta (Dagnew et al., 2012). La reproducción del banano se realiza a partir de los rizomas (cormos) obtenidos de plantas adultas, colinos de plantas jóvenes o plántulas obtenidas por medio de reproducción *in vitro*. Para asegurar una plantación sana, libre de plagas y enfermedades, se utiliza material vegetativo proveniente de plantaciones de la región o viveros tecnificados y se aplican procesos de desinfección grado alimenticio cuando es necesario (Morena Mena, Cristian Blanco Urina y Ricardo José Mendoza Torres, 2009).

El proceso de micropropagación se compone de 5 etapas. Comienza con la selección de tejidos vegetales (explantes) provenientes de una planta madre *ex vitro* vigorosa y libre de enfermedades (Murashige, 1974). En la siguiente etapa, de iniciación, el explante es desinfectado superficialmente y llevado a un medio de cultivo nutritivo; continua la etapa de multiplicación en la que mediante continuos subcultivos se aumenta el número de propágulos (Saini & Jaiwal, 2002). Posteriormente, en la etapa de enraizamiento se utiliza un medio de cultivo con reguladores de crecimiento para inducir el enraizamiento que puede iniciar desde la etapa de multiplicación (Suman, 2017). Y finalmente en la etapa de aclimatación, las plantas son endurecidas a partir de cambios graduales de humedad e intensidad de luz (similares a condiciones en campo), en esta etapa las plantas son extraídas del medio de cultivo y llevadas a suelos adecuados dentro de un invernadero para luego de 2 o 3 meses ser llevadas a campo (Joy, Anjana, & Jose, 2013).

Las plantas micropropagadas a partir de cormos o retoños demuestran diversas ventajas sobre las plantas cultivadas en campo; como un ciclo de producción más corto y uniforme, rendimientos más altos y crecimiento vigoroso (Robinson et al., 1993). Sin

embargo, estas variables son dependientes de factores como el genotipo, el tipo de explante y el cultivo, la composición de los medios de cultivo, reguladores del crecimiento de las plantas (PGR) y el entorno de cultivo (Vuylsteke, 1998).

Objetivos

Objetivo general

Caracterizar la microbiota bacteriana en diferentes órganos de vitroplantas de banano cv. Williams, asociando rasgos funcionales a las etapas tempranas de desarrollo.

Objetivos específicos

- Determinar la composición y diversidad de la microbiota bacteriana asociada a la endosfera y filosfera de vitroplantas de banano cv. Williams.
- Contrastar dos estrategias bioinformáticas comúnmente utilizadas para el procesamiento de secuencias de ADN provenientes de proyectos de microbiomas.
- Predecir el potencial funcional de la comunidad bacteriana asociada a vitroplantas de banano cv. Williams.

Metodología

Obtención de secuencias

Los datos empleados en el presente proyecto corresponden a los resultados de la secuenciación por síntesis de la región variable V3-V4 del gen ARNr 16S de las bacterias presentes en diferentes tejidos de plantas de banano provenientes de micropropagación en medio MS (Murashigue, T; Skoog, 1962) (Meristemos de Colombia S.A.S. - Rahan Meristems™). Estos datos fueron producidos por García et al., 2017 como parte del proyecto “Efecto de la aplicación de microorganismos en la microbiota, desarrollo y salud vegetal de plantas de banano” (sin publicar). Para su obtención, se seleccionaron 4 lotes, cada uno con 135 plantas *in vitro* de 9 meses de desarrollo aproximadamente. Los investigadores establecieron la endosfera de la raíz, la endosfera del pseudotallo y la filosfera como los compartimientos de estudio. Para capturar los microorganismos presentes en estos compartimientos, fraccionaron cada órgano por separado tomando 20 g de cada material vegetal, sometiendo previamente a una desinfección superficial con etanol al 75% e hipoclorito de sodio al 1.5% las raíces y pseudotallos para eliminar los microorganismos epífitos, proceso que no fue aplicado a las hojas, dado que se buscaba conservar los epífitos y endófitos.

Posteriormente, con cada material vegetal homogenizado se extrajo el ADN total empleando el kit PowerSoil DNA Isolation (QIAGEN) siguiendo las instrucciones del fabricante, y fue enviado a la compañía de secuenciación Macrogen en Korea, donde se realizó la amplificación de la región V3-V4 del gen 16S rRNA empleando los cebadores 341F / 805R. Luego se generaron las librerías de amplicones empleando el kit

Nextera XT Index, para ser finalmente cargados en el secuenciador de nueva generación para ADN Miseq (Illumina, Inc., USA).

En total se generaron 12 librerías que corresponden a secuencias bacterianas presentes en las 12 muestras de los diferentes órganos de las plantas. Las muestras corresponden a extracciones realizadas de cuatro lotes de plantas independientes identificadas por un número de lote. Del lote 1 y 2, se extrajo ADN presente en las raíces, los pseudotallos y las hojas, teniendo el lote 2 réplicas técnicas de cada órgano, del lote 3, solo se extrajo ADN presente en las raíces de las plantas contando con dos réplicas técnicas y del lote 4, se extrajo ADN exclusivamente de las hojas. En la tabla 1 se encuentra la descripción de las librerías, las cuales se componen de lecturas pareadas forward y reverse (*paired-end*) con un promedio de 183,558 lecturas por muestra (equivalente a la suma de lecturas forward y reverse cada una con un tamaño de 300 pb).

Procesamiento de secuencias

Para abordar el primer objetivo de determinar la composición y diversidad de las comunidades bacterianas asociadas a los órganos de plantas de banano *in vitro*, se comenzó con el establecimiento de un flujo de trabajo para el procesamiento de los datos crudos del secuenciamiento. Se escogió el paquete DADA2 (v.1.8.0) de R (versión 3.5.0), el cuál, no solo lleva a cabo los pasos tradicionales para el procesamiento de amplicones como el filtrado de calidad, la dereplicación de secuencias, la identificación de secuencias quiméricas y el ensamblaje de lecturas pareadas sino que también introduce pasos de modelado y corrección de errores de los amplicones para inferir la composición de las muestras y reducir la tasa de error (Callahan, Mcmurdie, et al., 2016).

Para el presente proyecto, el pipeline utilizado se basó en el script “DADA2 Pipeline tutorial (1.8)” (<https://benjjneb.github.io/dada2/tutorial.html>) y los parámetros fueron modificados de acuerdo con las características de los amplicones (Apéndice F); de esta manera, se corrió con las siguientes especificaciones: durante el paso de filtrado, se eliminaron los cebadores de las secuencias haciendo el corte de las primeras 17 bases de las secuencias forward y las primeras 21 bases de las secuencias reverse. Adicionalmente se escogió un puntaje de calidad igual a 25 (Phred Score), el cuál determinó que las lecturas forward fueran cortadas en 290 pb y las reverse en 205 pb. Las lecturas que superaron este paso fueron dereplicadas y procesadas por el algoritmo de DADA2 (estimación de error e inferencia de secuencias por muestra) con el parámetro pool activado en la inferencia de muestras, el cuál permite que la información se comparta entre muestras facilitando la detección de variantes raras presentes solo una o dos veces en una muestra, pero muchas veces en todas las muestras. Luego de eliminar las secuencias quiméricas, se fusionaron las lecturas pareadas que tuvieran una superposición mínima de 20 pb produciendo un total de 444 variantes de secuencias de amplicones (*amplicon sequence variants* - ASVs). Se le asignó a cada ASV su taxonomía utilizando el mismo paquete de DADA2 y el set de entrenamiento de la base de datos Silva versión 132 formateada para este paquete (el script de R se encuentra en el Apéndice A). Las ASVs correspondientes a organelas vegetales como cloroplastos y mitocondrias y aquellas identificadas dentro del dominio Eukarya fueron eliminadas, así, sólo las 405 ASVs identificadas como bacterias continuaron en los análisis posteriores.

Análisis de la composición y diversidad de la microbiota

A partir de la tabla de ASVs, se construyó una tabla de abundancias relativas, la cual se usó para determinar la abundancia por taxón en los diferentes niveles taxonómicos: filo, clase, orden, familia, género y especie por muestra. Los perfiles taxonómicos resultantes se visualizaron en gráficos de barras construidas en Excel, en donde se muestran por cada muestra, todos los filos; las clases y ordenes con una abundancia mayor al 1%, las familias y géneros con abundancia mayor al 2% y la abundancia de todas las especies identificadas.

Para realizar los análisis de diversidad, la tabla de ASVs fue filtrada y normalizada dentro de la plataforma QIIME (Quantitative Insights Into Microbial Ecology) (Caporaso et al., 2010) en donde se eliminaron todas las muestras con una profundidad inferior a 3,548 lecturas, excluyendo así las dos muestras del pseudotallo del lote 2 y la muestra de hojas del lote 4 de los análisis posteriores (tabla 1). Se realizó también la normalización de la tabla por el método de rarefacción llevando todas las muestras a la profundidad mínima, esto es, 3,548 lecturas, para corregir las diferencias que se originan debido a las diferencias en tamaño muestral (Kraker-Castañeda & Cobar-Carranza, 2011). Las métricas escogidas para determinar la diversidad alfa (dentro de cada muestra) fueron las ASVs observadas, el índice de Shannon, Equitabilidad (uniformidad) y la diversidad filogenética que fueron calculadas en QIIME (Caporaso et al., 2010) y graficadas en Excel. Para calcular el índice de diversidad filogenética se construyó un árbol filogenético con las secuencias de las ASVs con los paquetes phangorn (Schliep K.P., 2011) y phyloseq (McMurdie & Holmes, 2013) en R (Apéndice B) a partir del tutorial “Microbiome/Metagenome Analysis Workshop” (Brown

University, 2018). Para comparar la diversidad entre muestras (diversidad beta), se normalizaron de nuevo los datos por abundancia utilizando la normalización GMPR del paquete de R GMPR (L. Chen et al., 2018) y se calculó una matriz de distancias empleando la medida de Bray-Curtis en R usando el paquete vegan (Oksanen et al., 2018). Los resultados fueron graficados usando el método de análisis de coordenadas principales (PCoA) en R.

Comparación de pipelines

El pipeline utilizado para abordar el primer objetivo (DADA2), es una metodología de procesamiento de amplicones desarrollada recientemente que ha mostrado tener una mayor sensibilidad y especificidad que el método tradicional de construcción de OTUs, al evitar el agrupamiento de secuencias a partir de un umbral de similitud arbitrario (Callahan, McMurdie, & Holmes, 2017; Callahan, McMurdie, et al., 2016). Sin embargo, aunque diversos autores recomiendan el reemplazo de las OTUs por ASVs (Callahan et al., 2017), el método de agrupamiento es todavía muy utilizado en el estudio del microbioma vegetal (Agler et al., 2016; Edwards et al., 2015; Hartman, van der Heijden, Roussely-Provent, Walser, & Schlaeppli, 2017; Peiffer et al., 2013; Walters et al., 2018; Yeoh et al., 2017), es por esto, que se realizaron análisis adicionales que permitieran determinar el impacto del método de procesamiento de secuencias en los resultados de composición y diversidad de la microbiota.

Para realizar la comparación, se emplearon los resultados de entrenamiento del proyecto de donde se tomaron las secuencias crudas: “Efecto de la aplicación de microorganismos en la microbiota, desarrollo y salud vegetal de plantas de banano” (sin publicar) (García et al., 2017) en el cual se siguió el flujo de trabajo implementado por el

Instituto Holandés de la Ecología (NIOO-KNAW) de Pérez-Jaramillo et al., 2017 (en adelante llamado pipeline del NIOO) que brevemente consistió en: hacer un filtrado de calidad de las secuencias con un Phred score igual a 25 con el software BBDuk (<https://jgi.doe.gov/data-and-tools/bbtools/>). Luego, el ensamble de las lecturas pareadas (forward y reverse) con PANDASeq (Masella, Bartram, Truszkowski, Brown, & Neufeld, 2012) con una superposición de 10 pb, seguido por el agrupamiento de secuencias con una similitud del 97% en OTUs; la remoción de secuencias quiméricas con el algoritmo UCHIME (Edgar et al., 2011) y la clasificación taxonómica alineando las secuencias a la base de datos Silva versión 128 (Pérez-Jaramillo et al., 2017).

Para realizar la comparación de los dos pipelines en cuanto a grupos taxonómicos identificados, se realizó una nueva corrida del pipeline de DADA2 descrito anteriormente en el procesamiento de secuencias con la única diferencia de la base de datos utilizada, que en lugar de ser Silva versión 132 se corrió la versión 128. Los resultados de las dos versiones de DADA2 también fueron comparadas para determinar el impacto de la integridad de las bases de datos.

De los dos pipelines se comparó el total de unidades taxonómicas identificadas: OTUs/ASVs (riqueza específica), el número de taxones en cada nivel taxonómico (dominio, filo, clase, orden, familia, género, especie) y el impacto que estas tuvieron sobre todos los índices mencionados anteriormente de diversidad alfa y beta. Y finalmente se realizó una correlación de los dos métodos a partir de un análisis gráfico de Procrustes, el cuál permitió evaluar las distancias entre las dos formas de las graficas de diversidad beta construidas con distancias de Bray-Curtis; finalmente, se halló la significancia de la matriz Procrustes con la función `protest` del paquete “vegan” en R.

Predicción funcional

Se realizó una predicción del potencial funcional de la microbiota identificada en el primer objetivo utilizando la herramienta bioinformática PICRUSt2 v2.0.3-b: Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (Langille et al., 2013), la cuál, a partir de secuencias representativas del ARNr 16S (OTUs/ASVs) permite predecir funciones de los metagenomas presentes en las muestras. Se siguió el flujo de trabajo de PICRUSt2 recomendado por su desarrollador (<https://github.com/picrust/picrust2>) que en resumen consistió en: un paso inicial de colocación de secuencias (*Sequence placement*), en el cuál se utilizó HMMER (Eddy, 1998) para alinear las secuencias de la tabla de ASVs, con un alineamiento de secuencias múltiples de referencia, proveniente de la base de datos IMG (Integrated Microbial Genomes). Luego, estas secuencias alineadas se organizaron en una filogenia de referencia con EPA-ng (<https://github.com/Pbdas/epa-ng>) y gappa (<https://github.com/lczech/gappa>). El siguiente paso fue la predicción de estado oculto (*Hidden State Prediction*), en el cuál se predijo las abundancias de las familias de genes presentes con el paquete *castor* de R (Louca, Stilianos and Doebeli, 2017) eliminando las secuencias con un NSTI mayor a 1. Posteriormente, con las abundancias de las familias de genes, las copias del gen 16S y las secuencias representativas, se determinaron los perfiles funcionales predichos para cada muestra. La matriz de abundancia de genes anotados como KOs fue ingresada a la herramienta web MicrobiomeAnalyst (Dhariwal et al., 2017) en donde fue filtrada por abundancia y normalizados por el método CSS. La tabla resultante fue utilizada dentro de la misma plataforma para determinar las posibles rutas metabólicas presentes en las muestras a partir de la presencia y abundancia de los

KOs y la base de datos KEGG (Kanehisa & Goto, 2000). La información de abundancia de las rutas metabólicas encontradas fue llevada al software STAMP (Statistical Analysis of Taxonomic and Functional Profiles) en donde se realizó un PCA para evaluar la relación entre los perfiles funcionales encontrados y posteriormente se evaluó por medio de un test Tukey-Kramer la significancia entre las rutas. Las rutas significativas fueron graficadas en un mapa de calor en el que se mostró su abundancia por muestra y finalmente, se realizó una revisión de literatura para determinar si las rutas identificadas habían sido reportadas en la interacción bacteria-planta.

Resultados

Procesamiento de secuencias

La secuenciación por síntesis de la región V3-V4 del gen ARNr 16S arrojó en promedio $183,558 \pm 10,190$ lecturas crudas por muestra. Posterior al paso de filtrado de calidad se redujo el tamaño de las librerías, descartando en promedio el 29.97% de las lecturas. Una vez ensambladas las lecturas forward con sus lecturas reverse correspondientes, se obtuvieron 1,537,576 secuencias consenso, representadas por 3,071 secuencias únicas con longitudes que oscilan entre 273 y 439 nucleótidos, pero en su mayoría (51.87%) tuvieron una longitud de 427 nucleótidos. La proporción de secuencias únicas parece estar determinada por el lote del cual proviene la muestra, siendo el lote 1 el que posee más cantidad de secuencias únicas con un promedio de 1,362, le sigue el lote 3, con un promedio de 553, luego el lote 2 con un promedio de 243 y por último, el lote 4 con una cantidad de 165 secuencias únicas. En el paso de detección y remoción de secuencias quiméricas se detectaron 2,627 quimeras, lo cual, redujo el número de secuencias únicas de 3,071 a 444. Sin embargo, aunque se mantuvo sólo el 14.45% de las secuencias únicas, éstas equivalen al 93.63% de todas las lecturas consenso del estudio. Las lecturas totales pre-procesadas (tabla 1) fueron aquellas producto de todo el pipeline de DADA2 y cuyas secuencias únicas fueron compiladas en una matriz con su respectiva abundancia y taxonomía para continuar con su procesamiento. La asignación taxonómica dependiente de la base de datos Silva versión 132, clasificó 8 ASVs como pertenecientes al dominio Eukarya. Estas ASVs con una abundancia máxima de 0.04%, no estaban presentes en todas las muestras y como el

enfoque de este estudio son las comunidades bacterianas, fueron excluidas de los análisis posteriores.

En todas las muestras se registró amplificación de organelas vegetales como cloroplastos y mitocondrias, sin embargo, la proporción de cloroplastos fue en todas las muestras mucho mayor que las mitocondrias, teniendo un promedio del 67.61% y alcanzando hasta el 92.73% de las lecturas en la muestra del lote 4. Por otra parte, el

Tabla 1. Descripción de las librerías de amplicones producidas en la secuenciación con Illumina MiSeq y lecturas procesadas por el método DADA2 (lecturas 300 pb x 2)

Característica	Lote 1			Lote 2						Lote 3		Lote 4
	Raíz	Pseudot	Hoja	Raíz-1	Raíz-2	Pseudot-1	Pseudot-2	Hoja-1	Hoja-2	Raíz-1	Raíz-2	Hoja
Lecturas crudas (x2)	170,487	172,871	180,619	182,280	170,217	189,037	199,501	187,866	186,540	176,903	201,857	184,518
Lecturas filtradas (x2)	117,110	120,062	127,962	130,052	124,349	121,972	139,573	138,049	132,608	123,250	137,360	129,601
Lecturas inferidas DADA2 (forward)	116,867	119,906	127,737	129,852	124,324	121,916	139,537	137,832	132,557	123,153	137,269	129,553
Lecturas inferidas DADA2 (reverse)	117,027	120,015	127,879	129,885	124,294	121,884	139,482	137,975	132,571	123,153	137,234	129,565
Lecturas fusionadas	116,462	119,651	127,347	129,452	124,231	121,792	139,368	137,491	132,446	122,900	136,957	129,479
Lecturas quimeras	25,887	21,059	23,570	1,829	534	4,067	3,701	563	2,600	6,411	5,977	1,623
Lecturas totales pre-procesadas	90,575	98,592	103,777	127,623	123,697	117,725	135,667	136,928	129,846	116,489	130,980	127,856
Lecturas totales k_Bacteria	90,573	98,591	103,774	127,590	123,645	117,721	135,661	136,928	129,845	116,485	130,970	127,856
Lecturas totales de cloroplastos	45,439	35,757	56,933	77,189	83,316	88,537	104,833	120,902	118,970	67,835	77,263	118,565
Lecturas totales de mitocondrias	16,771	1,897	5,938	31,566	25,242	28,951	30,611	2,616	5,789	45,102	49,955	9,079
Lecturas totales procesadas	28,363	60,937	40,903	18,835	15,087	233 *	217 *	13,410	5,086	3,548	3,752	212 *
Total de secuencias (ASVs)	23	33	36	188	84	21	22	72	32	53	60	32

porcentaje promedio de lecturas mitocondriales presentes en las muestras fue de 17.39% y siendo de tan solo de 1.91% en la muestra hoja-1 del lote 2. La abundancia de lecturas de organelas vegetales en las muestras no estuvo determinada por lote ni compartimiento; sin embargo, si se observó que la mayoría de las muestras de hojas fueron las que tenían mayor proporción de cloroplastos y que la mayor cantidad de lecturas mitocondriales, se encontraban en muestras de raíz.

La cantidad de lecturas procesadas por muestra es muy variable, la diferencia entre la muestra con menor y mayor número de lecturas fue de 2 ordenes de magnitud.

De esta manera, con el objetivo de hacer una comparación sin sesgos y no perder una gran cantidad de información de las muestras con mayor profundidad, se decidió eliminar las muestras con un recuento de secuencias inferior a la profundidad de la muestra raíz-1 del lote 3 (3,548) señaladas en la tabla 1 con *.

Después de todos los procesos de filtrado presentes en el flujo de trabajo de procesamiento, el total ASVs, es decir, la cantidad de secuencias únicas por muestra sigue estando determinada por el lote, sin embargo, el lote que contiene mayor secuencias únicas es el lote 2, seguido por el 3 y finalmente el 1.

Análisis de la composición y diversidad de la microbiota

En el presente estudio, se identificaron 10 filos, 17 clases, 33 ordenes, 55 familias, 102 géneros y 26 especies bacterianas (figura 3, 4 y 5, Apéndice C) que se distribuyen a lo largo de todas las muestras, siendo en todos los casos, las muestras raíz-1, raíz-2 y hoja-1 del lote 2, las que contienen mayor número de taxones.

El perfil taxonómico reveló que la composición de las comunidades bacterianas estaba mayormente determinada por el lote a excepción de la muestra hoja-2 del lote 2 que difiere mucho de su réplica técnica hoja-1 y parece tener una composición muy similar a las muestras del lote 3; por lo que se sugiere que uno de los factores con gran influencia en la composición de la microbiota de una planta es su procedencia, es decir, el conjunto de factores abióticos y presiones bióticas experimentadas por las plantas, en este estudio corresponde a la planta madre que dio origen a las plántulas analizadas.

Los filos Actinobacteria, Firmicutes y Proteobacteria estuvieron presentes en todas las muestras, sugiriendo que hacen parte de la microbiota núcleo de las vitroplantas de banano analizadas (Hamonts et al., 2018). El filo Proteobacteria fue el de mayor

abundancia en la mayoría de las muestras, alcanzando un 99% de la composición de todas las muestras del lote 1, su abundancia solo fue superada en la muestra hoja-1 del lote 2 por el filo Firmicutes.

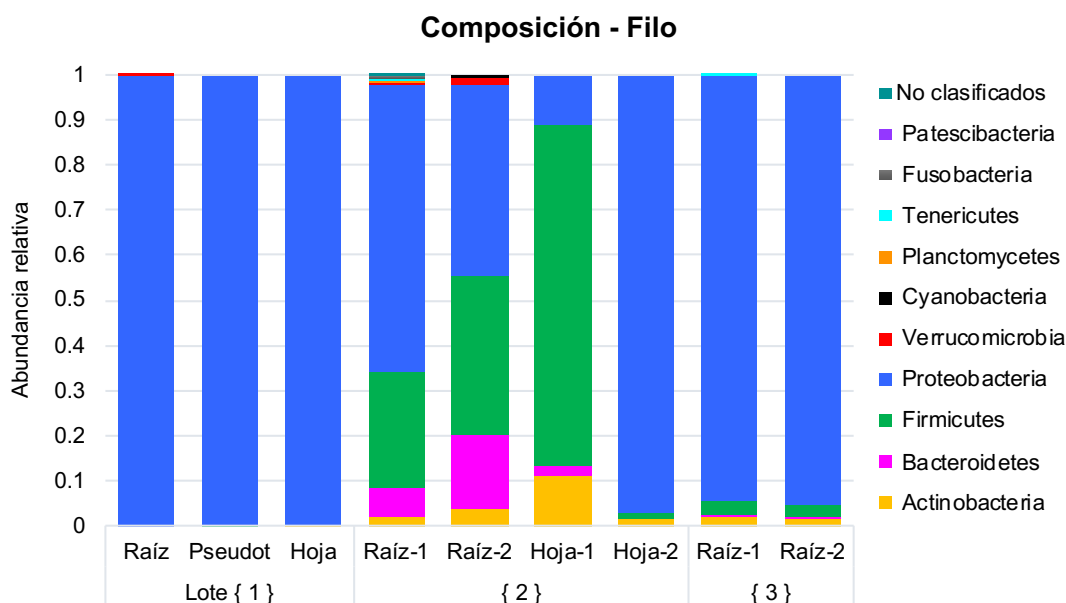


Figura 3. Abundancia relativa de los filos presentes en la microbiota bacteriana asociada a la endosfera de raíces y pseudotallos y a hojas de plantas de banano en estado *in vitro*.

El filo Bacteroidetes fue encontrado exclusivamente en muestras de raíz y hojas, mientras que el filo Verrucomicrobia fue encontrado únicamente en muestras de raíz; sin embargo, estuvo ausente en todas las muestras del lote 3 correspondientes a raíces. Otros filos como Cyanobacteria, Planctomycetes, Tenericutes, Fusobacteria y Patescibacteria solo fueron encontrados en una o dos muestras con una abundancia por muestra máxima del 0.4%.

A nivel de género, se identificó que *Bacillus*, *Stenotrophomonas* y *Pseudomonas* estaban presentes en todas las muestras siendo *Pseudomonas* el género con mayor abundancia en todas las muestras excepto las muestras de hojas del lote 2 (figura 4). La

Figura 4. Abundancia relativa de los géneros más predominantes (> 2%) en la microbiota bacteriana asociada a la endosfera de raíces y pseudotallos y a las hojas de plantas de banano en estado in vitro.

De las 26 de especies identificadas (figura 5) ninguna se encontró en la totalidad de las muestras, sin embargo, *Stenotrophomonas maltophilia* se encontró en todos los compartimientos. El porcentaje de especies identificadas no es representativo de las comunidades bacterianas aisladas, debido a que el porcentaje de especies sin identificar todavía es muy alto, en promedio del 91,23%.

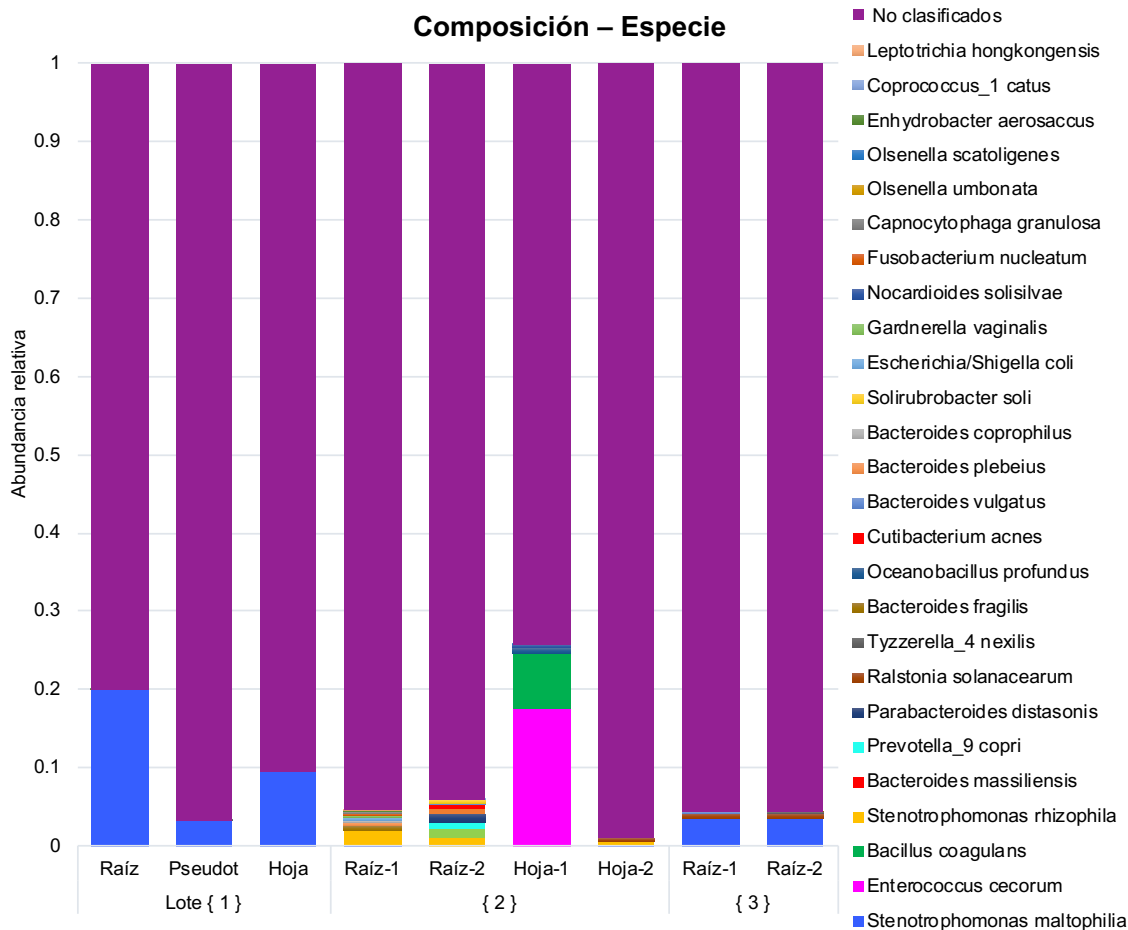


Figura 5. Abundancia relativa de las especies presentes en la microbiota bacteriana asociada a la endosfera de raíces y pseudotallos y a hojas de plantas de banano en estado in vitro.

Las muestras con un mayor porcentaje de especies identificadas fueron la muestra hoja-1 del lote 2 y la muestra de raíz del lote 1, con porcentajes de abundancias del 26% y 20% respectivamente; el primero representado en su mayoría por *Enterococcus cecorum* con una abundancia del 17% y el segundo representado por *Stenotrophomonas maltophilia* en un 19%. De las muestras del lote 1, lote 3 y la hoja-2 del lote 2, se identificaron en promedio 4 especies, en contraste con las demás muestras, raíz-1, raíz-2 y hoja-1 del lote 2 en las que se identificaron 15, 9 y 6 especies respectivamente.

Para calcular la diversidad dentro de cada muestra (diversidad alfa) y compararla entre muestras con diferentes recuentos de secuencias y esfuerzos de muestreo, se normalizaron las lecturas con el método de rarefacción. Las curvas de rarefacción (figura 6a, 6b) muestran que la profundidad de secuenciación escogida (3,548) fue suficiente para capturar la diversidad presente la mayoría de las muestra y para tener un estimado acercado en las muestras submuestreadas.

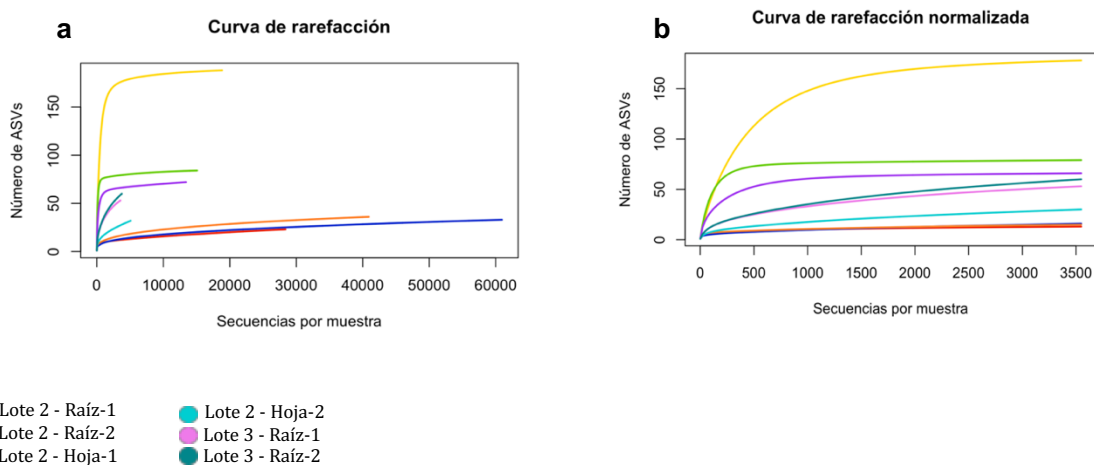


Figura 6. Curvas de rarefacción que representan el número de ASVs identificadas en 9 muestras de la microbiota de vitroplantas de banano. a) Tamaño de las librerías antes de la normalización y b) Profundidad de las muestras normalizadas por rarefacción a una profundidad de 3,548.

Luego de la rarefacción, el número de ASVs pasó de 394 a 375, sin embargo, su distribución en las muestras siguió dependiendo del lote, siendo el lote 2 el de mayor número de ASVs. Los diferentes índices de diversidad alfa calculados (ASV observadas, índice de Shannon, equitabilidad y diversidad filogenética) muestran que la diversidad difiere entre órganos, pero es consistente entre lotes, siendo en general el lote 2 el más diverso (figura 7); similar a los resultados obtenidos en el análisis de coordenadas principales (PCoA) empleando la distancia Bray-Curtis (figura 8) que muestra que el lote es el principal determinante de la diversidad beta al formarse clusters definidos por cada lote. La gran cercanía de los lotes 1 y 3 demuestra que estos lotes tienen una composición similar entre sí, contrario al lote 2 que forma un cluster más disperso con la mayoría de sus muestras y tiene una, la hoja-1 que difiere en alto grado en composición y diversidad.

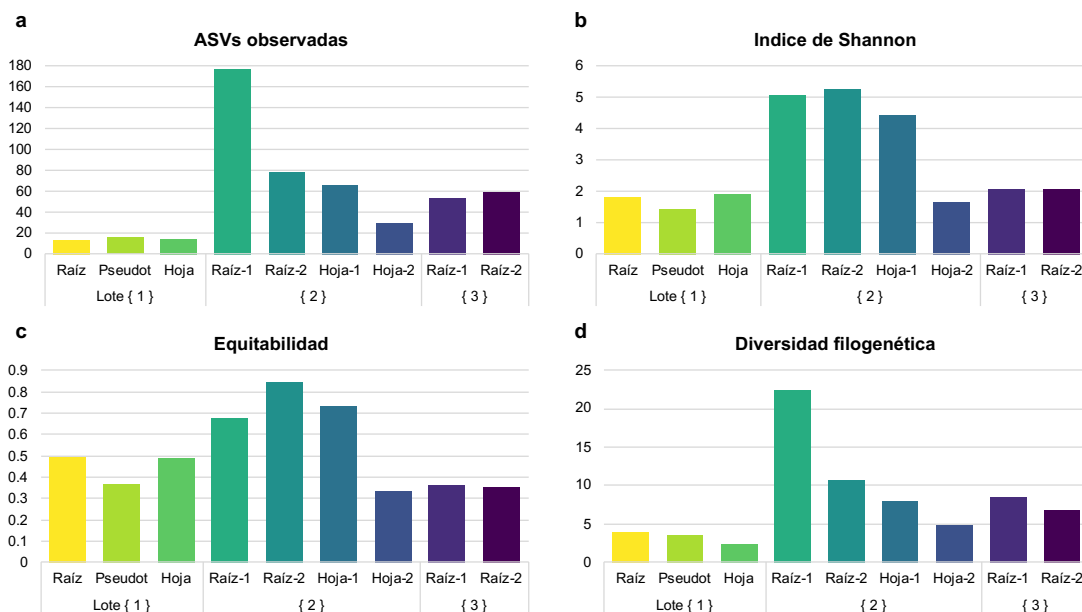


Figura 7. Índices de diversidad alfa de la microbiota asociada a la endosfera de raíces, pseudotallos y hojas de plantas de banano *in vitro*.

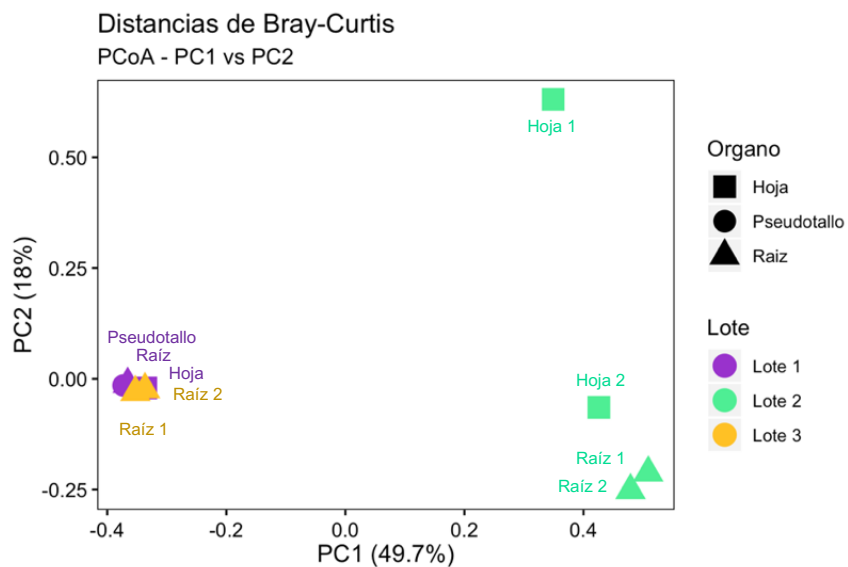


Figura 8. Análisis de coordenadas principales de la microbiota asociada a la endosfera de raíces, pseudotallos y hojas de vitropiantas de banano basado en la métrica de distancia de Bray-Curtis y normalización por el método GMPR aplicada a la matriz de ASVs.

Comparación de pipelines

Al final del procesamiento de secuencias, el flujo de trabajo del NIOO produjo entre 189 y 40,867 lecturas por muestra que se clasificaron en 278 unidades taxonómicas operativas (OTUs), de las cuales 140 no fueron identificadas. Las profundidades obtenidas con este flujo de trabajo fueron menores que las profundidades obtenidas con DADA2, excepto en las dos muestras de pseudotallo del lote 2. Para estas muestras el pipeline del NIOO obtuvo profundidades de 645 y 491 lecturas por muestra (pseudotallo 1 y pseudotallo 2), mientras que DADA2 capturó unas profundidades de 233 y 217 respectivamente. Estas dos muestras, al igual que la muestra del lote 4 fueron eliminados debido a su bajo conteo de lecturas. Para el proceso de rarefacción se escogió una profundidad igual a 2,273 lecturas representadas por 166 OTUs que se tradujeron en 9 filos, 16 clases, 28 ordenes, 40 familias, 66 géneros y sin lograr identificar ninguna especie.

Todos los taxones a nivel de filo, clase, orden y familia identificados por el pipeline del NIOO fueron identificados también con el pipeline DADA2. Éste último pipeline, también encontró en todos los niveles mencionados, mayor número de taxones: 1 filo, 3 clases, 6 ordenes y 15 familias más. El pipeline del NIOO fue el único en dejar secuencias sin identificar, es decir, que hubo secuencias de las cuales no se pudo determinar a que dominio pertenecían. Por otra parte, el pipeline DADA2 logró clasificar todas las secuencias como mínimo hasta el nivel de dominio. Ambos pipelines lograron identificar la misma abundancia de secuencias de cloroplastos y mitocondrias por muestra, sin embargo, el pipeline del NIOO las agrupó en 79 y 42 OTUs respectivamente y DADA2 las condensó en 16 ASVs cada uno. El pipeline DADA2 produjo secuencias

que la base de datos clasificó como eucariotas, por el contrario, el pipeline del NIOO no clasificó ningún OTU fuera del dominio bacteria. El pipeline DADA2 generó 444 ASVs en contraste con el pipeline del NIOO que produjo 977 OTUs. Aunque el número de unidades taxonómicas representativas (OTUs o ASVs) producidas por el pipeline del NIOO fue mayor que el de DADA2, luego de realizar el proceso de filtrado de secuencias vegetales el número de ASVs (404), fue mayor que el de OTUs (278).

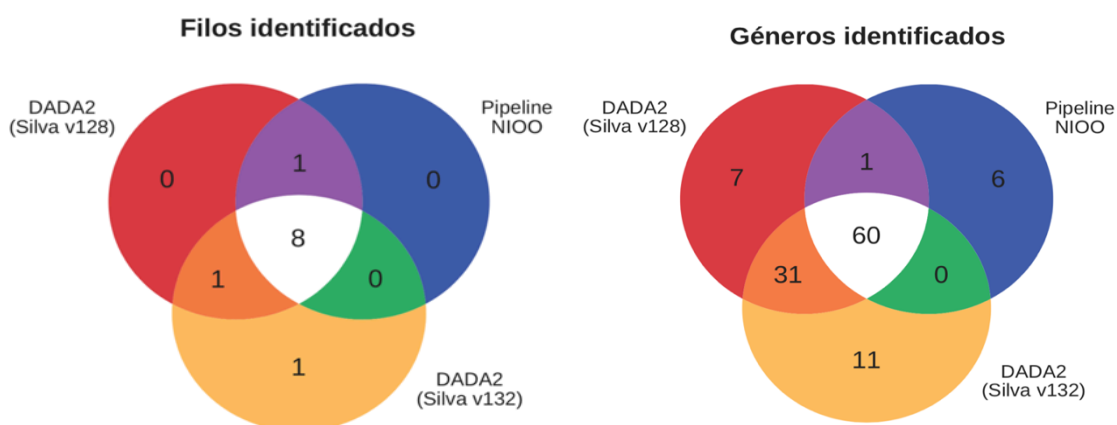


Figura 9. Diagramas de Venn que muestran la intersección de grupos taxonómicos identificados por cada pipeline (pipeline del NIOO y pipeline DADA2) y por cada versión de la base de datos Silva (DADA2 con Silva versión 128 y DADA2 con1 Silva versión 132).

Las diferencias en el pipeline DADA2 usando las dos versiones de la base de datos Silva (versión 128 y versión 132) comienzan a verse a nivel de filo y continúan hasta el nivel de especie (figura 9 y Apéndice D), siendo los niveles de filo y especie los más similares entre versiones. La versión de Silva 128 clasificó una ASV dentro del filo Saccharibacteria, 20 secuencias dentro de las clases Flavobacteriia, OPB35_soil_group,

Sphingobacteriia y Betaproteobacteria y 13 secuencias dentro de los órdenes Rhodocyclales, Thermoanaerobacterales, Neisseriales y Burkholderiales, estas secuencias fueron reclasificadas en la versión 132 de Silva al filo Patescibacteria; las clases Bacteroidia, Verrucomicrobiae y Gammaproteobacteria y el orden Betaproteobacteriales.

A nivel de género, la versión 132 fue la que más grupos identificó logrando detectar 11 taxones adicionales a los identificados por la versión 128. En éstos y en otros géneros comunes, se reclasificaron las 11 secuencias de los géneros identificados únicamente con DADA2 y la versión 128 de Silva. La versión 128 identificó 28 especies dentro de las que se encuentran *Paracoccus yeei* y *Acinetobacter lwoffii*; la última versión de Silva, logró identificar 26 especies, más sin embargo, para las secuencias identificadas como las especies mencionadas, no las llevó hasta el nivel de especie, sino que las clasificó hasta el nivel de género.

A pesar de que se identificaron más ASVs que OTUs, la mayoría de las muestras exhiben una mayor riqueza específica cuando se realizó el agrupamiento de secuencias en OTUs (figura 10). Al comparar la diversidad alfa entre los resultados obtenidos con los dos pipelines y las dos versiones de Silva, se encontró que los índices de OTUs/ASVs observadas, el índice de Shannon, el de equitabilidad y la diversidad filogenética, entre versiones, es muy similar y entre los tres métodos se mantiene la misma tendencia (figura 12).

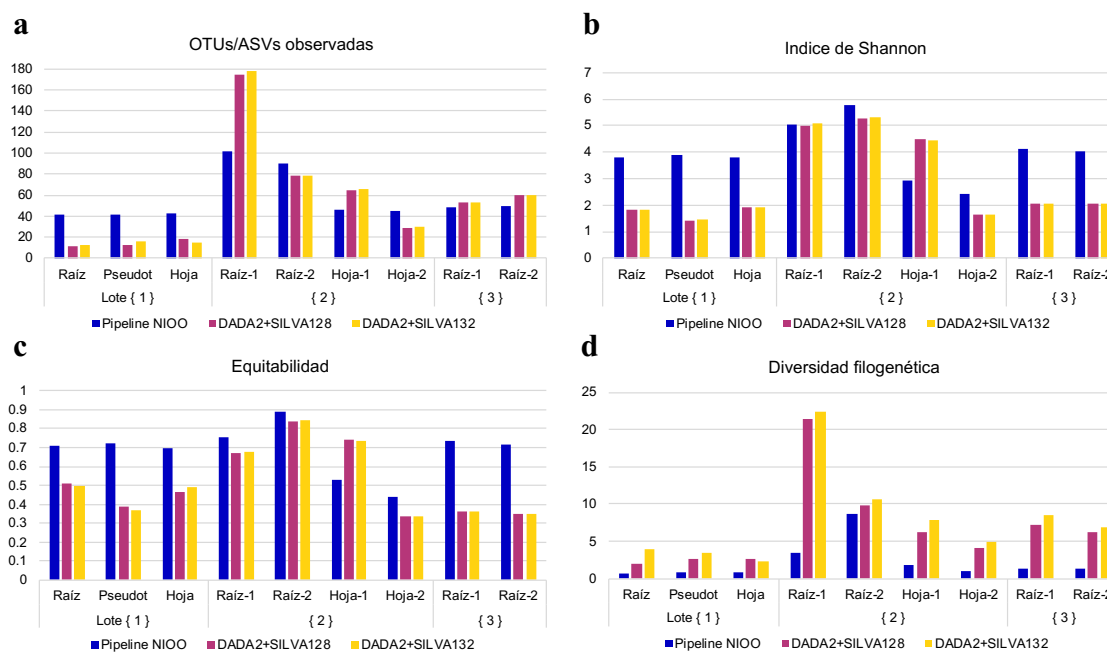


Figura 10. Índices de diversidad alfa de la microbiota identificada por cada pipeline (pipeline del NIOO y pipeline DADA2) y por cada versión de la base de datos Silva (DADA2 Silva versión 128 y DADA2 Silva versión 132).

Las diferencias tan marcadas entre el pipeline del NIOO con DADA2 sugiere que el programa de procesamiento de secuencias si tienen un impacto en los resultados de diversidad. Este resultado es apoyado por los resultados de diversidad beta. Con el pipeline DADA2 los datos se organizan en clusters dependientes del lote y su cercanía está relacionada con similitudes en la composición. Sin embargo, los resultados del pipeline del NIOO no evidencian la formación de clusters entre lotes ni ninguna otra variable contemplada en el estudio (figura 11).

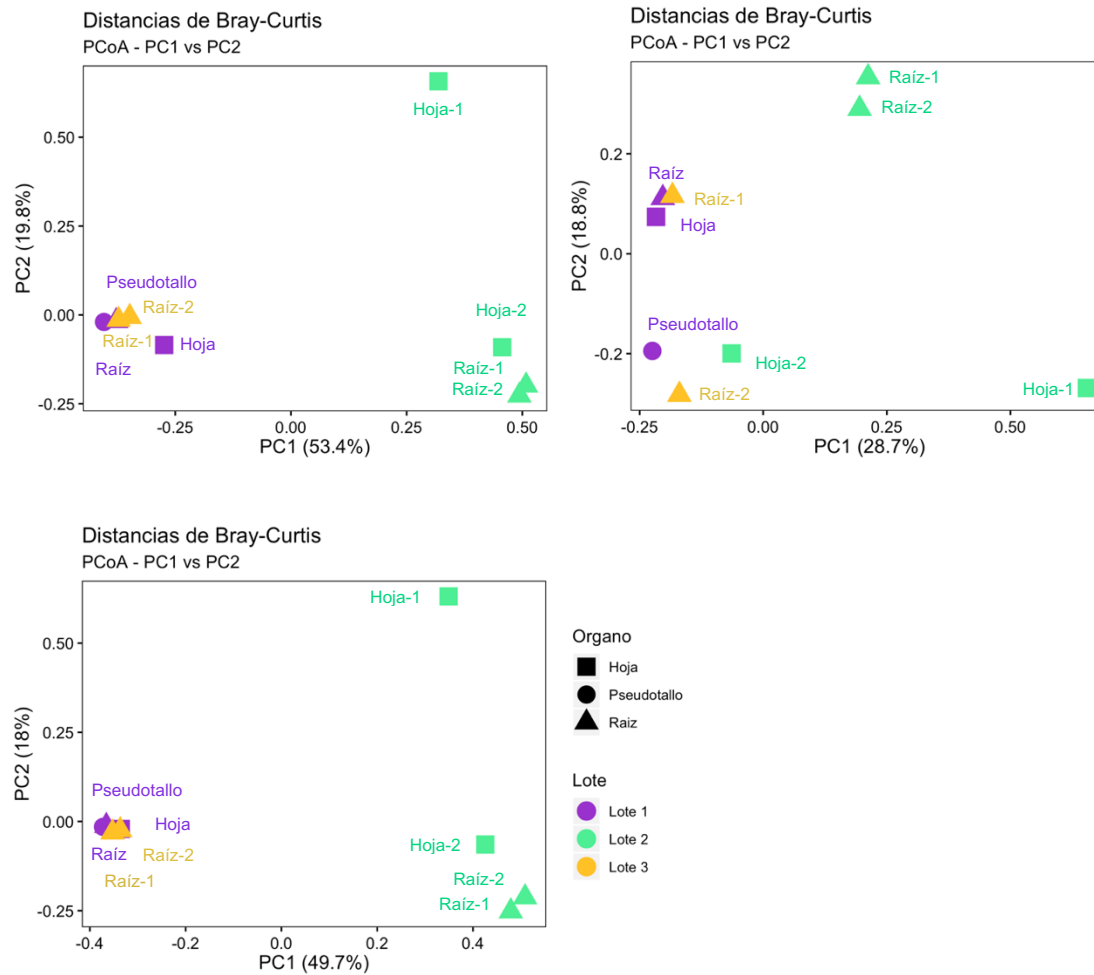


Figura 11. Análisis de coordenadas principales basado en la métrica de distancia de Bray-Curtis de la microbiota asociada a plantas de banano en estado in vitro, identificada por cada pipeline (pipeline del NIOO y pipeline DADA2) y por cada versión de la base de datos Silva (DADA2 Silva versión 128 y DADA2 Silva versión 132). a) DADA2 con Silva versión 128, b) pipeline del NIOO y c) DADA2 con Silva versión 132.

Los análisis de Procrustes (figura 12) mostraron una alta correspondencia entre las versiones de la base de datos las cuales tuvieron con una correlación del 0.99. Por otra parte, al comparar los métodos de procesamiento se evidenció un desplazamiento de las

muestras. Sin embargo, las dos diversidades beta aun mantienen un porcentaje alto de correlación igual al 0.77.

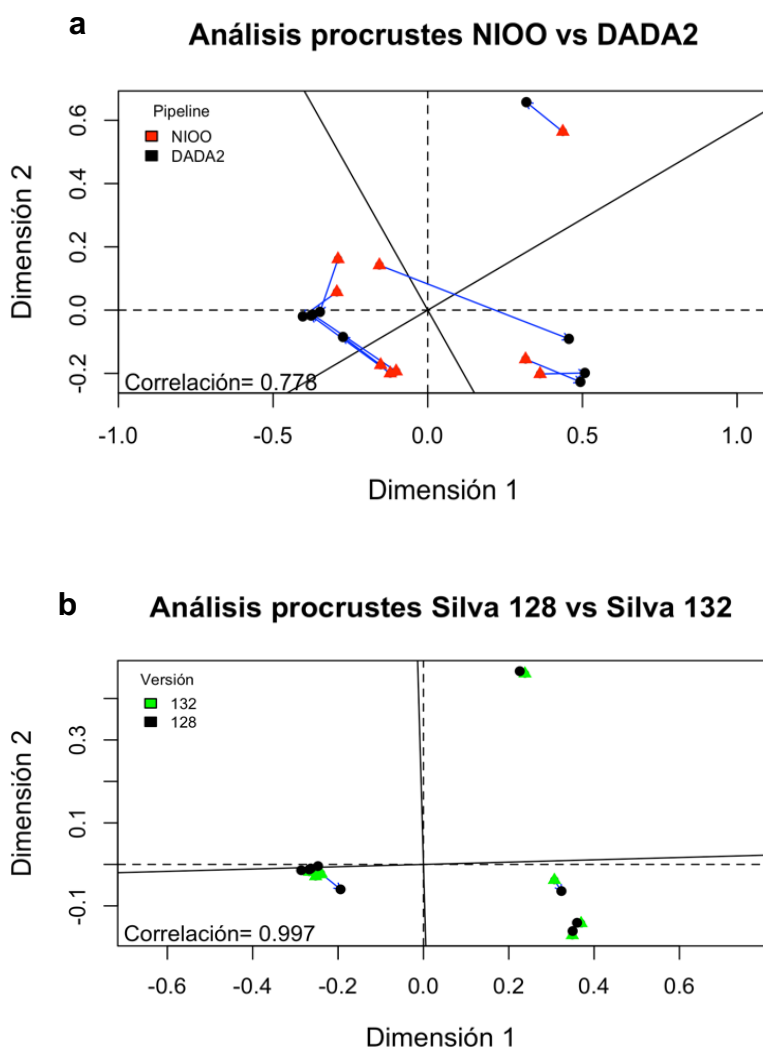


Figura 12. Análisis procrustes de los PCoA de las distancias de Bray-Curtis de a) pipeline del NIOO y pipeline DADA2 y b) DADA2 con Silva versión 128 y DADA2 con Silva versión 132 (el valor de significancia de la correlación se indica en la figura).

Predicción funcional

Se utilizó el software PICRUSt2 para predecir las funciones presentes en los compartimientos de la endosfera y filosfera de plantas de banano en estado *in vitro*. Durante la corrida de PICRUSt2, se identificaron dos secuencias con un NSTI mayor a 1 las cuales fueron eliminadas de los análisis posteriores para evitar ruido en los datos. De las secuencias resultantes la mayoría tenían NSTIs iguales a 0 o 0.1 y muy pocas superaban el 0.6 lo que indica que la información de las secuencias fue buena para realizar una predicción precisa. Se identificaron en total 8,885 funciones bajo las anotaciones de los ortólogos KEGG (KOs) que se distribuyen a lo largo de todas las muestras. Estos KOs junto con sus abundancias fueron filtradas y normalizadas con el método CSS por MicrobiomeAnalyst resultando en 5,477 KOs usados para predecir las rutas metabólicas presentes en el microbioma del banano. Se encontró un total de 131 rutas metabólicas a partir de la comparación con la base de datos de rutas de KEGG.

La predicción funcional también mostró una clara diferencia en la composición de los KOs entre lotes. El primero formó un grupo estrecho a la izquierda, muy cercano al tercero que mostró una superposición entre sus muestras; mientras que el segundo mostró una distribución dispersa a lo largo del eje PC1 de la gráfica PCA (Figura 13).

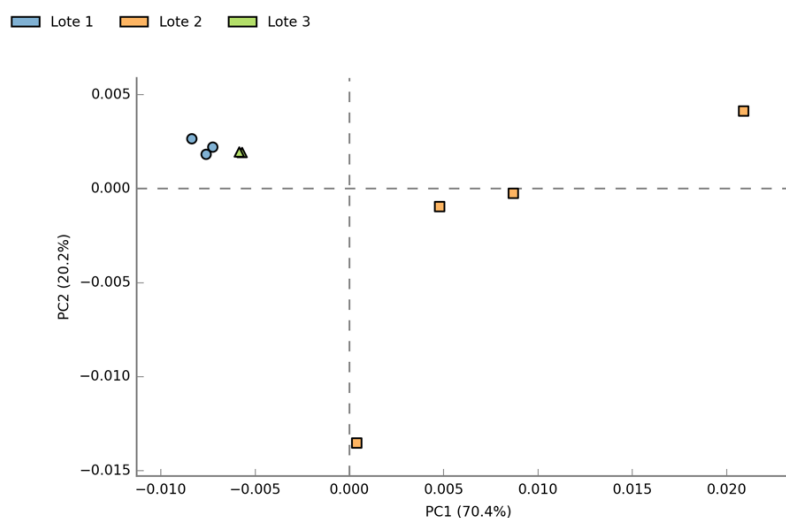


Figura 13. Análisis de componentes principales (PCA) de los ortólogos KEGG identificados en las comunidades bacterianas presentes en la endosfera y filosfera de plantas de banano en estado *in vitro*, producto de la predicción funcional realizada con PICRUSt2 y MicrobiomeAnalyst.

De las 131 rutas identificadas, 44 mostraron una diferencia significativa ($p < 0.05$) entre lotes, las cuales se muestran en la figura 14, y una abundancia relativa entre 0 - 6,4%.

Estas rutas corresponden al metabolismo primario de los microorganismos, es decir, procesos fisiológicos como el metabolismo del carbono, la síntesis de varios aminoácidos ácidos grasos y fuentes de energía; y algunas reportadas dentro de relaciones simbióticas con las plantas como la producción de antibióticos y de lipopolisacáridos para defensa de la planta contra patógenos y la fijación de nitrógeno y la producción de sideróforos para la nutrición de la planta.

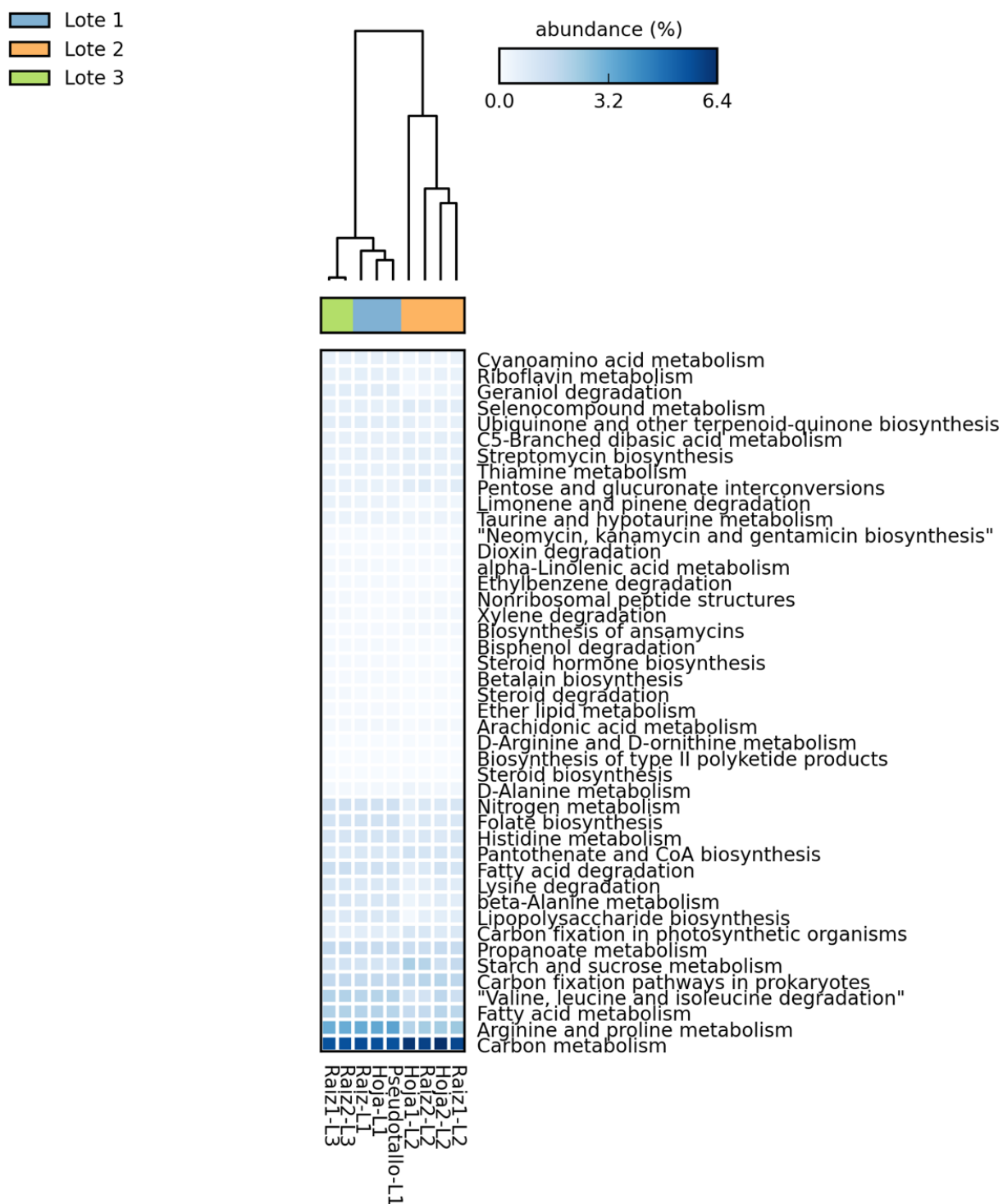


Figura 14. Mapa de calor que muestra la abundancia relativa de las rutas metabólicas significativas ($P < 0.05$) presentes en los perfiles funcionales predichos con PICRUSt2 y MicrobiomeAnalyst filtrados y

normalizados con el método CSS para las comunidades bacterianas presentes en la endosfera y filosfera de plantas de banano en estado in vitro. En la parte superior se muestra un dendrograma que muestra el agrupamiento de las muestras a partir de su composición funcional.

Discusión

Procesamiento de secuencias

Los estudios del microbioma vegetal a partir de genes marcadores como el ARNr 16S se enfrentan a diferentes desafíos que pueden afectar los resultados de diversidad y otros análisis posteriores. El primer inconveniente se fundamenta en los métodos actuales de extracción del ADN y secuenciamiento y consiste en que aunque las lecturas del 16S brindan una imagen detallada del perfil de una comunidad microbiana en términos de composición y diversidad, este método no discrimina entre células vivas y muertas por lo cual puede llevar a sobreestimaciones en la diversidad (Lebeis, 2014).

El segundo se fundamenta en el grado de resolución de las secuencias de estudio y tiene que ver con la gran proporción de secuencias que no son identificadas hasta el nivel de especie. En muchos casos, el gen marcador utilizado no tiene la longitud suficiente para distinguir entre especies o cepas de microorganismos, pues puede ocurrir que varias secuencias del gen marcador sean casi idénticas a otras secuencias bacterianas presentes en la base de datos de referencia (Jovel et al., 2016). La profundidad de secuenciación también es un factor muy influyente en el nivel taxonómico de clasificación al igual que en la detección de miembros poco comunes en las comunidades. Una gran profundidad de secuenciación permite la recolección de secuencias con baja abundancia en la comunidad, sin embargo, con la detección de estas secuencias reales, también se producen secuencias con errores que son cada vez más difíciles de diferenciar de las secuencias reales (Weiss et al., 2017) y tienen un gran impacto en la sobreestimación de OTUs en los métodos tradicionales de procesamiento de secuencias (Rosen et al., 2012).

Un inconveniente muy común en el estudio del microbioma vegetal consiste en la baja proporción de secuencias correspondientes a los microorganismos diana, debido a la masiva presencia de secuencias correspondientes a la planta hospedera como cloroplastos y mitocondrias (Coleman-Derr et al., 2016; Feehery et al., 2013; Sakai, Matsuka, Komura, & Kanazawa, 2004). Estas dos organelas, a pesar de hacer parte de una célula eucariota, cuentan con un linaje procariota (Dyall, Brown, & Johnson, 2004), esto conlleva a que exista una homología entre los genes del 16S bacteriano, los del 16S del cloroplasto y los genes del ARNr 18S de plantas y el mitocondrial (Hanshewa, Masonb, Raffa, & Curriea, 2013). Actualmente existen tres métodos que permiten reducir el impacto de la contaminación por organelas vegetales. El primero actúa a en el paso de extracción del ADN y consiste en aislar el ADN del núcleo de las células para así reducir el número de copias del ADNcp y ADNmt (Lutz, Wang, Zdepski, & Michael, 2011) o alternativamente, metilar el ADN del hospedero y separarlo del ADN microbiano a partir de las diferencias en la densidad de metilación (Feehery et al., 2013). El segundo método consiste en bloquear la amplificación de secuencias pertenecientes al hospedero al momento de la preparación de las librerías al adicionar etiquetas con nucleótidos aleatorios a cada molécula plantilla antes de realizar la PCR y posteriormente revisar la cantidad de amplicones que comparten la secuencia de la etiqueta para identificar las verdaderas secuencias diana (Lundberg et al., 2013). Y por ultimo, la utilización de cebadores *mismatch* que no cuentan con homología para el ADN de las organelas (Sakai et al., 2004). Este ultimo método, es el más utilizado (Beckers et al., 2016) y cuenta ya con un repertorio de cebadores que amplifican la mayoría de regiones del ARN 16S (Thijs et al., 2017).

Análisis de la composición y diversidad de la microbiota

En el presente estudio se esperaba encontrar comunidades bacterianas dependientes de los órganos estudiados, similar a los resultados reportados por Abreu-Tarazai et al., 2010 en plantas de piña *in vitro* (Abreu-Tarazi et al., 2010). Sin embargo, no se observó ninguna relación entre órganos y por el contrario se encontró que las comunidades bacterianas eran muy homogéneas a lo largo del cuerpo de la planta sugiriendo que la comunidad endófito del meristemo inicial, en las primeras etapas del desarrollo de la planta de banano se distribuye por todos los órganos de la planta uniformemente.

Los resultados de composición y diversidad de las comunidades bacterianas mostraron ser muy similares de acuerdo al lote al que pertenecían las muestras; esto podría deberse a las características específicas de la planta madre que dio origen a las plantas procesadas en cada muestra. Esto sugeriría que factores abióticos como como el tipo de suelo, condiciones climáticas, la ubicación geográfica y practicas de manejo de cultivo tienen una mayor influencia que factores bióticos como la especie de la planta o la variedad en la construcción del microbioma o en el microbioma que se hereda a la progenie (Posada U., 2017). Por otra parte las grandes diferencias encontradas entre réplicas técnicas sugiere un impacto debido a los métodos empleados en la extracción del ADN y producto del secuenciamiento. Una distribución heterogénea en el pellet que dio origen a las réplicas técnicas, pudo causar diferencias en su composición al igual que las diferencias en profundidad que evidencian muchas muestras pudo causar que la muestra de hoja-1 del lote 2 tuviera una composición diferente a la de hoja-2 del mismo lote, pues al tener más profundidad la primera muestra, pudo obtener más información de una

proporción más grande de los microorganismos presentes en la comunidad (Rosen et al., 2012).

Los filos Actinobacteria, Firmicutes y Proteobacteria encontrados en todos los órganos se encuentran también presentes en muchas especies (Rossmann, Sarango-Flores, Barros Chiaramonte, et al., 2017) apoyando la hipótesis de Schlaeppi & Bulgarelli, 2015 de que hay un gran microbioma núcleo común a la gran mayoría de especies vegetales (Schlaeppi & Bulgarelli, 2015). Se ha encontrado que estos filos cuando están en el compartimiento endófito ayudan al crecimiento y desarrollo de la planta (Bulgarelli et al., 2012; Hardoim et al., 2015; Wemheuer et al., 2017). En el review de Rossmann et al., 2017 junto con los filos mencionados anteriormente se encuentra Bacteroidetes en el microbioma núcleo, el cual en este estudio solo estuvo presente en muestras de raíz y hojas. Su ausencia en la muestra de pseudotallo no significa necesariamente que este ausente en todos los pseudotallos de las plantas de banano ni tampoco asegura que no este presente en las vitroplantas de banano procesadas en este estudio; tal vez su abundancia fue muy baja en este órgano y con la profundidad de secuenciación obtenida no pudo ser detectada, lo cual es un problema muy común en el estudio de microbiomas (Rosen et al., 2012). Los filos Verrucomicrobia y Planctomycetes encontrados solo en muestras de raíz han, sido identificados en varios estudios como algunos de los filos más dominantes del suelo (Janssen, 2006), y del microbioma de la rizosfera (Turner et al., 2013), sugiriendo que existen unos patrones de distribución de la microbiota que comienzan a ser evidentes dentro de las primeras etapas de desarrollo de las plantas. De los demás filos identificados hay muy pocos reportes en

la literatura debido a que generalmente mantienen una baja abundancia en las comunidades y son difíciles de aislar (Sarhan et al., 2018)

Microorganismos pertenecientes a los géneros *Pseudomonas*, *Bacillus* y *Stenotrophomonas*, los tres géneros del microbioma núcleo identificado en este estudio se han reportado en los tres grandes compartimientos de la planta, sin embargo, han sido catalogados como los géneros más importantes de endófitos vegetales junto con *Burkholderia*, *Micrococcus*, *Pantoea* y *Microbacterium* y más predominantes en la rizosfera de muchas especies (Bulgarelli et al., 2013; Giesler & Yuen, 1998; Nair, Singh, & Sekar, 2002; Romero, Marina, & Pieckenstain, 2014; Santoyo et al., 2016). Algunas cepas de éstos géneros se caracterizan por que promueven el crecimiento vegetal en el compartimiento endófito y en la rizosfera, al adquirir los nutrientes necesarios para el desarrollo de la planta y la modulación del crecimiento (Berg, 2009; Santoyo et al., 2016), y participan en la defensa de la planta a estrés biótico como el ataque de patógenos al producir diferentes compuestos antifúngicos (Giesler & Yuen, 1998; Meyer et al., 2010). Se ha identificado que diferentes cepas de *Pseudomonas* tienen la capacidad de producir antibióticos que junto con los lipopéptidos actúan como reguladores y apoyan la formación de biopelículas, la señalización, la motilidad y la adquisición de micronutrientes (Raaijmakers, de Bruijn, Nybroe, & Ongena, 2010). La presencia del género *Methylobacterium* exclusivamente en hojas está relacionada con la condición de algunos microorganismos de éste género de ser metanótrofos (Delmotte et al., 2009)

Algunos de los géneros mencionados como *Pseudomonas*, *Bacillus* y *Burkholderia* y otros como *Klebsiella*, *Methylobacterium*, *Enterobacter*, *Acinetobacter*, *Paenibacillus* y *Microbacterium* han demostrado beneficiar a diferentes especies de

plantas durante el cultivo *in vitro* (Orlikowska et al., 2017). Endófitos de los géneros *Pseudomonas* y *Burkholderia* por medio de la producción de sideróforos y la solubilización de fosfato ayudan a mantener la salud de la planta (Parray, Kamili, Reshi, Qadri, & Jan, 2015). Se ha encontrado que algunas cepas de *Pseudomonas* y *Bacillus* aumentan la longitud de las raíces y tallos, al mismo tiempo que el área de las hojas (Salomon et al., 2014); otras cepas del género *Klebsiella* pueden aumentar la tasa de germinación y supervivencia de los cormos (Parray et al., 2015); Algunos *Bacillus* por medio de la síntesis de fitohormonas puede aumentar el peso y la proliferación del cormo y promover el enraizamiento como también lo hacen algunos microorganismos del género *Microbacterium* (Quambusch et al., 2014; Salomon et al., 2014). Algunas cepas del género *Acinetobacter* puede por medio de la producción de fitohormonas y la solubilización de fosfato pueden ayudar en el proceso de germinación del cormo (Parray et al., 2015). Algunos microorganismos del género *Methylobacterium* también tiene la capacidad de sintetizar fitohormonas que incrementan el número y la longitud de las raíces y microorganismos de *Paenibacillus* por medio de la producción de fitohormonas, la asimilación de nitrógeno atmosférico y la síntesis de compuestos indol pueden aumentar la longitud de raíces y tallos reduciendo los impactos de la etapa de aclimatación de la micropropagación (Faria, Dias, Melo, & de Carvalho Costa, 2013).

Fernandes et al., 2013 identificaron un microbioma núcleo en plantas de banano del cultivar Prata Anã obtenido por micropropagación en Brasil. Los géneros identificados en esta variedad fueron *Pseudomonas*, *Enterobacter*, *Burkholderia*, *Bacillus* y *Klebsiella* (Fernandes et al., 2013). Todos estos géneros fueron encontrados en el presente estudio, teniendo todos a excepción de *Burkholderia* alta abundancia en las

muestras. Asociados a estos géneros en banano se ha encontrado la capacidad de algunos endófitos de producir fitohormonas como el ácido 3-indol acético y promover el crecimiento de las plántulas, al igual que la elongación de los brotes, el diámetro del pseudotallo y el peso seco de la planta (Pandey, Palni, & Bag, 2000). Del estudio de Posada U, 2017 sobre la microbiota de plantas de banano del cultivar Williams a partir de métodos cultivables, se detectaron 7 de los 11 géneros identificados por este autor, a pesar de esto, se lograron identificar 94 géneros más en la microbiota de este cultivar de banano (Posada U., 2017).

La especie con más abundancia identificada, *S. maltophilia*, ha sido ampliamente estudiada con relación al microbioma vegetal. Se sabe que se transmite por las semillas de las plantas y una vez crece la planta, se convierte en miembro de la rizosfera (Brooke, 2012). Esta especie, cuyo genoma está completamente secuenciado, presenta genes dedicados a la motilidad, adaptación y colonización del tejido vegetal y funciones como el aumento de la biomasa de diferentes cultivos (Hardoim et al., 2012).

La cantidad de taxones identificados en la mayoría de las muestras corresponden a la riqueza total esperada para las muestras, evidenciada en las curvas de rarefacción, donde la mayoría de las muestras alcanzaron el plateau. Algunas muestras que se encuentran próximas a alcanzar el plateau sugieren que la totalidad de la diversidad no pudo ser capturada y que no se registraron especies o taxones “raros” con la profundidad de 3,548 lecturas por muestra (Weiss et al., 2017).

Los índices de diversidad alfa (Figura 7) revelaron que las dos muestras de raíz y la hoja-1 del lote 2 son las muestras más diversas, debido a que contienen la mayor riqueza específica, evidenciada hasta el nivel de género (Figura 4) y aunque se observan

taxones con alta abundancia, no tienen una dominancia fuerte en las muestras. Las muestras del lote 3 tienen una riqueza específica intermedia comparando entre lotes, sin embargo, el índice de equitabilidad sugiere que hay un taxón con alta dominancia (*Pseudomonas sp*) que hace que tenga una diversidad similar a las muestras del lote 1 (índice de Shannon) que tienen el menor número de taxones. El índice de diversidad filogenética reveló, al igual que el índice de Shannon, mayor diversidad del lote 2, seguida del lote 3 y el lote 1, es decir, que este último lote, además de tener un menor número de taxones, estos son filogenéticamente cercanos, perteneciendo la mayoría al filo Proteobacteria.

Los resultados de la diversidad beta evidenciaron una vez más que la diversidad está determinada por los lotes y que los lotes 1 y 3 exhiben casi que exclusivamente el microbioma núcleo identificado, mientras que las muestras del lote 2, que también se componen del microbioma núcleo poseen más diversidad lo que los lleva a ser muy diferentes de los otros lotes.

Los grandes cambios de composición y diversidad de la hoja 2 del lote 2 con respecto a la hoja 1 del mismo lote, al igual que la gran diferencia entre réplicas técnicas, se podrían explicar por errores en el secuenciamiento o fallas en la manipulación de las muestras durante el proceso de extracción de ADN o amplificación (PCR), las cuales impidieron la detección de la totalidad de los taxones esperados.

La identificación de las diferentes comunidades bacterianas en plantas cuyos cormos experimentaron una desinfección superficial y fueron sembradas y mantenidas en ambientes asépticos, evidencia el impacto de la transmisión vertical y el desplazamiento de los microorganismos a través de los compartimientos de la endosfera y filosfera de la

planta. Esta microbiota inicial representa una fuente importante de microorganismos que están involucrados en el desarrollo de la planta en las primeras etapas de su ciclo de vida. Estos microorganismos que no representan la totalidad de la microbiota de la planta madre, fueron heredados debido a que mantenían relaciones estrechas de mutualismo con sus hospederos (Johnston-Monje & Raizada, 2011).

Comparación de pipelines

Para el procesamiento de los datos de la microbiota utilizados en este proyecto se decidió utilizar el paquete de software DADA2 en lugar de un flujo de trabajo fundamentado en el agrupamiento de secuencias en OTUs, debido a que se ha demostrado que por medio de la estimación de error y la inferencia de muestras se producen secuencias con mayor resolución y se reduce la tasa de falsos positivos (Callahan, Mcmurdie, et al., 2016; Kraemer et al., 2018).

Al comparar los resultados de los dos métodos se obtuvo un comportamiento similar al reportado en la literatura con relación a la riqueza específica, en donde el número de OTUs supera en gran medida el número de ASVs una vez se completa el flujo de trabajo de procesamiento (Kraemer et al., 2018). En el presente estudio el número de OTUs superó el doble de la cantidad de ASVs identificadas. Esta proporción se invirtió una vez se filtraron los datos de material vegetal.

Las diferencias encontradas entre metodologías a nivel taxonómico son evidentes en casi todos los niveles jerárquicos (Figura 9, Apéndice D). Todos los taxones identificados por el pipeline del NIOO fueron detectados por el pipeline de DADA2, el cual además de identificarlos, alcanzó a detectar aun más grupos en todos los niveles indicando que el algoritmo de DADA2 permite obtener con más resolución el perfil

taxonómico de las comunidades bacterianas. Las diferencias taxonómicas encontradas entre herramientas, en nuestro caso, a nivel de género, se ha reportado que se deben a diferencias en el volumen de lecturas filtradas: en donde diferentes herramientas que utilizan el aproximamiento de las OTUs tienden a producir un menor número de secuencias filtradas que el filtrado realizado por DADA2 (Dahan, Jude, Lamendella, Keesing, & Perron, 2018), Teniendo en cuenta que la comparación de las herramientas se hace frente a la misma versión de la base de datos, el hecho de que el pipeline del NIOO no identificara taxonómicamente 32 OTUs pertenecientes al dominio bacteria refleja una alta tasa de error al momento de la construcción de los clusters.

Las diferencias entre los acercamientos metodológicos son aun más evidentes en los cálculos de la diversidad alfa (figura 10), en donde hay un contraste evidente y generalmente determinado por el lote. El índice de riqueza específica muestra que en la mayoría de las muestras el número de OTUs es mayor al número de ASVs. Esta tendencia evidencia la gran cantidad de clusters que fueron creados para cuantificar la riqueza, que como se menciona anteriormente contiene errores e introduce en el índice OTUs que no tienen ninguna clasificación taxonómica al igual que se pueden construir clusters a partir de secuencias singleton que añaden ruido a los resultados. Adicional a sobreestimar la riqueza de las muestras generalmente de baja abundancia, reduce el estimado de riqueza de muestras muy diversas, como la muestra raíz-1 del lote 2, al agrupar secuencias que difieren hasta en un nucleótido, pero que pueden ser identificadas taxonómicamente como familia, géneros o especies diferentes (Callahan, McMurdie, et al., 2016). Los índices de equitabilidad y diversidad filogenética calculados por el pipeline del NIOO reflejan muestras con organismos muy similares entre si

(cercanamente emparentados) y con abundancias similares, esto corresponde a la pérdida de riqueza específica como resultado de el agrupamiento de secuencias, las cuales, al no poder ser distinguidas, se clasifican como el mismo taxón que las otras secuencias pertenecientes a su cluster. DADA2 al poder detectar con alta precisión diferencias entre secuencias de hasta 1 nucleótido, permite identificar más grupos taxonómicos y relaciones ecológicas significativas como la dominancia de ciertos taxones. A pesar de estas ventajas del pipeline DADA2, se ha registrado que disminuye los estimados de diversidad alfa, lo cual fue evidente en este estudio en los índices de Shannon y el de equitabilidad.

Las gráficas de diversidad beta (Figura 11) también evidencian un impacto debido al método de procesamiento. En el PCoA producido por DADA2 se observan que las muestras forman clusters de acuerdo a la composición de sus comunidades siendo muy similares la microbiota del lote 1 y la del 3 y estando lejanas al lote 2. Con el pipeline del NIOO no se forman clusters aparentes de acuerdo a una variable determinada en el estudio. Para determinar que tanto variaban las diversidades se construyeron dos gráficos para el análisis Procrustes, uno que comparaba el pipeline DADA2 (versión de Silva 128) con el pipeline del NIOO y el otro, comparó las dos últimas versiones de la base de datos. Los resultados (figura 12) mostraron una diferencia muy pequeña entre las versiones de Silva (Protest; $p = 0.001$; correlación = 0.9968; permutaciones = 999) con una correlación casi perfecta, sin embargo, se muestra que entre versiones si se ven afectados los resultados. Las pequeñas diferencias entre versiones se aprecian como pequeñas variaciones en la taxonomía y en las diversidades indicando que nuevos conocimientos producidos en los últimos años han permitido dilucidar mejor identidades

taxonómicas y relaciones filogenéticas entre la microbiota vegetal. Y en la comparación entre las dos herramientas (Protest; $p = 0.08$; correlación = 0.778; permutaciones = 999) se aprecia un desplazamiento entre las muestras y a pesar, de que se observó un alto grado de similitud, con una correlación igual a 0.77 todavía hay un 23% de diferencia que explica la detección y clasificación de otros taxones por parte del pipeline DADA2. Esto confirma que las diferencias entre el pipeline del NIOO y DADA2 tienen un impacto tanto en composición como en la diversidad alfa y beta de las comunidades identificadas.

La precisión del pipeline del NIOO para identificar taxones, disminuye en los niveles taxonómicos más profundos, es por esto que para los proyectos cuyo objetivo es la caracterización del perfil taxonómico de un microbioma se recomienda la implementación del pipeline de DADA2 (Callahan, Mcmurdie, et al., 2016).

Predicción funcional

El perfil funcional predicho para la microbiota identificada en el presente estudio, se compone de diferentes rutas metabólicas pertenecientes a procesos fisiológicos de los microorganismos de la microbiota y a rutas involucradas en sistemas simbióticos entre éstos y su hospedero.

White Jr et al., 2014 propusieron tres categorías de sistemas endosimbióticos que incluyen (1) Sistemas endosimbióticos defensivos, (2) Sistemas endosimbióticos tolerantes al estrés y (3) Sistemas endosimbióticos nutricionales (White Jr, Torres, Johnson, Irizarry, & Tadych, 2014). En este estudio se modificaron estos sistemas para clasificar las rutas metabólicas simbióticas predichas para el microbioma de la endosfera

y filosfera de vitroplantas de banano en: (1) sistemas defensivos, (2) sistema adaptativo y de desarrollo y (3) sistema nutricional. Dentro del sistema defensivo, se encontraron múltiples rutas destinadas a la biosíntesis de antibióticos como neomicina, estreptomicina, kanamicina, gentamicina y ansamicina. La producción de antibióticos ha estado ligada a los compartimentos de la rizosfera y la endosfera (Berg et al., 2014). Dependiendo de las concentraciones a las que sean producidos, pueden actuar como mecanismos para suprimir patógenos o mediar la comunicación entre microorganismos y de ellos con la planta. La producción de antibióticos a nivel inhibitorio se presenta generalmente en la rizosfera como uno de los mecanismos por medio del cual los microorganismos actúan sobre los patógenos del suelo para suprimir enfermedades. Los antibióticos producidos a concentraciones sub-inhedorias median la comunicación entre microorganismos y entre éstos con la planta para apoyar procesos como la formación de biopelículas, la señalización, motilidad y la adquisición de micronutrientes. Dentro de las bacterias productoras de antibióticos se destacan cepas pertenecientes a la clase Actinomycetes (Turner et al., 2013) y al género *Pseudomonas* (Berg et al., 2014); la alta recurrencia de estos grupos en la microbiota de las plantas de banano analizadas (Figura 4 y apéndice C) se ve evidenciada en que representan las rutas metabólicas más completas predichas para este estudio a partir del número de KOs identificados.

La capacidad de degradar limoneno, pineno y geraniol, encontrado en nuestras comunidades bacterianas, representan componentes importantes de la vía de biosíntesis del monoterpeno y del metabolismo de los terpenoides (H. Chen et al., 2018). Estas rutas se han encontrado en diferentes estudios de microbiomas donde son caracterizadas como vías metabólicas terpenoides comunes en microorganismos y fundamentales para la

defensa de la planta a perturbaciones bióticas (Cheng et al., 2013) al igual que la ruta de biosíntesis de lipopolisacaridos identificada en este estudio (Newman, Dow, Molinaro, & Parrilli, 2007).

Dentro del sistema adaptativo y de desarrollo se identificó la presencia de múltiples rutas metabólicas relacionadas con la biosíntesis de nucleótidos y diferentes azúcares, rasgos que son característicos de los endófitos, los cuales requieren un alto reservorio de energía para llevar a cabo el proceso de colonización de la endosfera (Timm et al., 2015). Las rutas significativas más abundantes en el presente estudio fueron las involucradas con la promoción del crecimiento de las plantas; estas son la fijación de carbono en organismos fotosintéticos y las vías de fijación de carbono en procariontas. Esta última ha sido relacionada con las cianobacterias, un filo poco común en la microbiota de las plantas y detectada en nuestro estudio en una sola muestra correspondiente a raíces del lote 2.

Dentro del sistema nutricional se encontraron rutas involucradas en la promoción de la disponibilidad de elementos como nitrógeno y el hierro. La fijación de nitrógeno es el proceso de reducción del nitrógeno molecular atmosférico a amoníaco, una forma reducida y biológicamente útil que puede ser incorporada en aminoácidos y otros compuestos vitales para las plantas (Lemanceau et al., 2017). Los únicos microorganismos capaces de fijar nitrógeno atmosférico son procariontas llamados diazótrofos que cuentan con el complejo enzimático nitrogenasa representado por los genes *nifD**HK* que codifican para la subunidad de la proteína de hierro de la nitrogenasa. (Bouffaud, Renoud, Moënnelocoz, & Muller, 2016; Lemanceau et al., 2017). Estos genes se encuentran en diferentes bacterias presentes en todos los grandes

compartimentos de la planta y generalmente son producidas por diazótrofos de los géneros *Beijerinckia*, *Rhizobia*, *Sphingomonas*, *Caulobacter*, *Clostridium*, *Pseudomonas*, *Paenibacillus*, *Klebsiella* y *Enterobacter*. En el presente estudio, se registró la presencia de los tres genes dentro de la ruta del metabolismo del nitrógeno (K02586 = *nifD*, K02588 = *nifH* y K02591 = *nifK*). La fijación de nitrógeno en ambientes *in vitro* ha sido relacionada con la promoción del número y tamaño de las raíces de los brotes. De los géneros mencionados anteriormente, sólo *Klebsiella* estaba entre la microbiota identificada; sin embargo, se identificaron otros 29 géneros que podrían contar con este importante complejo enzimático para la fijación de nitrógeno. Esta observación apoya la afirmación de que las funciones metabólicas son redundantes entre microorganismos, es decir, que una función no es única de un microorganismo, sino que puede ser común a un grupo amplio de microorganismos. Al igual que un microorganismo particular puede cumplir múltiples funciones en la ecología de su planta huésped, debido a que las funciones predominantes de un endófito pueden cambiar dependiendo de las circunstancias ecológicas que afecten a su hospedero (White Jr et al., 2014).

La adquisición de hierro por parte de la planta podría estar mediada por algunos microorganismos productores de sideróforos. Los sideróforos microbianos tienen la capacidad de quelar Fe con alta afinidad, causando la solubilización de sus partículas en el suelo y un aumento en su biodisponibilidad para la nutrición de las plantas (Pii et al., 2016). Esta actividad ha sido relacionada comúnmente con bacterias de las familias Pseudomonadaceae, Norcardiaceae y Moraxellaceae y es un mecanismo fundamental para la adquisición de hierro por parte de la planta (Prashar et al., 2014). Dentro del proceso de micropropagación también se ha registrado la producción de sideróforos, los

cuales promueven el crecimiento de raíces y promueven su ramificación al igual que aumentan el número de hojas; uno de los géneros más identificados en este proceso ha sido *Pseudomonas* (Orlikowska et al., 2017). Dentro de nuestro estudio se encontraron las familias Pseudomonadaceae y Moraxellaceae, lo cual apoya la predicción realizada sobre estas rutas metabólicas. Se ha hipotetizado que la asimilación de todos estos nutrientes por parte de la planta, logra aumentar la resistencia al estrés oxidativo (Bravo et al., 2011).

Se identificó la presencia de diversas rutas metabólicas involucradas en procesos de promoción de crecimiento y defensa de la planta, sin embargo, no es posible determinar si estas funciones se están llevando a cabo desde las primeras etapas *in vitro* del desarrollo de las plantas de banano, o corresponden a actividades que por su importancia en la etapa adulta de la planta fueron escogidas para ingresar al compartimento endófito y ser transferidos a las siguientes generaciones.

Conclusiones

El conocimiento del microbioma de una variedad o especie vegetal es de gran importancia para el mantenimiento de cultivos sanos y productivos; más aun en cultivos comerciales donde la necesidad de alimentar a una población mundial en crecimiento bajo los efectos del cambio climático requiere optimizar el uso de los recursos y los impactos ambientales de la producción de alimentos (Busby et al., 2017). Los resultados obtenidos en el presente proyecto permitieron evidenciar una gran diversidad en la microbiota asociada a plantas de banano en estado *in vitro*, que se distribuye de manera uniforme por los diferentes órganos y compartimentos de la planta. La composición y la diversidad de las comunidades bacterianas están determinadas principalmente por la variable lote del cual provinieron las muestras; y su composición estaba representada en su mayoría por los filos Actinobacteria, Firmicutes y Proteobacteria y por los géneros *Bacillus*, *Stenotrophomonas* y *Pseudomonas*.

Las metodologías para el procesamiento de secuencias comparadas en el presente estudio mostraron diferencias en la diversidad alfa, el número de taxones identificados y la capacidad de detección de taxones “raros”. DADA2 fue el método que, aunque produjo una disminución en diversos índices de diversidad alfa, logró identificar más taxones en todos los niveles taxonómicos y detectó individuos con muy baja abundancia; ésto lo convierte en el pipeline recomendado para estudios de metataxonomía.

Asociada a la diversidad identificada, se encontraron unas posibles funcionalidades ampliamente reportadas en literatura como pertenecientes al metabolismo primario de los microorganismos y a relaciones simbióticas presentes en el desarrollo de plantas en estado *in vitro*.

Finalmente, se concluye que el pipeline bioinformático implementado fue lo suficientemente robusto para obtener secuencias con alta resolución que pudieran ser clasificadas taxonómicamente y permitieran una asignación funcional ampliamente soportada por la literatura a partir de información de los genes que codifican para el ARN 16S.

Perspectivas futuras

Este tipo de estudios permiten ampliar el conocimiento de las comunidades microbianas que cohabitan con una planta y generan una mejor comprensión del microbioma vegetal. En el caso específico del banano, esta información podría ser aprovechada a nivel *in vitro* para optimizar la sanidad de las plantas, la nutrición y los rendimientos de la micropropagación en una agricultura sostenible y más adelante, cuando las plantas se encuentren en campo podría ayudar a prevenir el brote de enfermedades y la asociación de patógenos con este cultivo.

El flujo de trabajo establecido en este estudio podría ser implementado con un set de datos más amplio que contenga múltiples replicas biológicas para lograr un entendimiento más amplio de la composición y diversidad de la microbiota asociada a especies vegetales.

La identificación precisa de los procesos metabólicos presentes en las relaciones mutualistas, comensales y patógenas de los microbios con su planta hospedera, requiere de tecnologías más robustas que una predicción funcional para obtener resultados sin sesgos. Una secuenciación metagenómica completa (*shotgun*) es necesaria para determinar qué genes bacterianos son seleccionados en los diferentes órganos y compartimentos de la planta. Esta información podría ser complementada con estudios de metatranscriptómica y metaproteómica que permitan obtener información de cuándo y dónde se expresan los genes identificados por secuenciación *shotgun*. La adición de otras técnicas moleculares específicas para la evaluación de funciones o rutas determinadas podrían complementar los enfoques mencionados anteriormente, como sería el uso de sondas con isótopos estables que permiten la identificación de microorganismos que

metabolizan un sustrato etiquetado como el carbono producido por las plantas. Finalmente, la integración de técnicas dependientes de cultivo, brindarían una comprensión más clara de las interacciones planta-microorganismos a nivel de sistemas (Turner et al., 2013).

Lista de referencias

- Abreu-Tarazi, M. F., Navarrete, A. A., Andreote, F. D., Almeida, C. V., Tsai, S. M., & Almeida, M. (2010). Endophytic bacteria in long-term in vitro cultivated “axenic” pineapple microplants revealed by PCR-DGGE. *World Journal of Microbiology and Biotechnology*, *26*(3), 555–560. <http://doi.org/10.1007/s11274-009-0191-3>
- Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V., & Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *Journal of Bacteriology*, *186*(9), 2629–2635. <http://doi.org/Doi 10.1128/Jb.186.9.2629-2635.2004>
- Agler, M. T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S. T., Weigel, D., & Kemen, E. M. (2016). Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLoS Biology*, *14*(1), 1–31. <http://doi.org/10.1371/journal.pbio.1002352>
- Arkipova, T. N., Prinsen, E., Veselov, S. U., Martinenko, E. V., Melentiev, A. I., & Kudoyarova, G. R. (2007). Cytokinin producing bacteria enhance plant growth in drying soil. *Plant and Soil*, *292*(1–2), 305–315. <http://doi.org/10.1007/s11104-007-9233-5>
- Ausubel, F. M. (2005). Are innate immune signaling pathways in plants and animals conserved? *Nature Immunology*, *6*(10), 973–979. <http://doi.org/10.1038/ni1253>
- Bais, H. P., Weir, T. L., Perry, L. G., Gilroy, S., & Vivanco, J. M. (2006). The Role of Root Exudates in Rhizosphere Interactions with Plants and Other Organisms. *Annu. Rev. Plant Biol*, *57*, 233–266. <http://doi.org/10.1146/annurev.arplant.57.032905.105159>
- Bakker, P. A. H. M., Berendsen, R. L., Doornbos, R. F., Wintermans, P. C. A., & Pieterse, C. M. J. (2013). The rhizosphere revisited: root microbiomics. *Frontiers in Plant Science*, *4*(May), 1–7. <http://doi.org/10.3389/fpls.2013.00165>
- Balvočiute, M., & Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT - how do these taxonomies compare? *BMC Genomics*, *18*(Suppl 2), 1–8. <http://doi.org/10.1186/s12864-017-3501-4>
- Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T., & Stamatakis, A. (2018). EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Systematic Biology*, *0*(0). <http://doi.org/10.1093/sysbio/syy054>

- Bastian, F., Cohen, A., Piccoli, P., Luna, V., Baraldi, R., & Bottini, R. (1998). Production of indole-3-acetic acid and gibberellins A1 and A3 by *Acetobacter diazotrophicus* and *Herbaspirillum seropedicae* in chemically-defined culture media. *Plant Growth Regulation*, 24, 7–11. <http://doi.org/10.1023/A>
- Beckers, B., Op De Beeck, M., Thijs, S., Truyens, S., Weyens, N., Boerjan, W., & Vangronsveld, J. (2016). Performance of 16s rDNA primer pairs in the study of rhizosphere and endosphere bacterial microbiomes in metabarcoding studies. *Frontiers in Microbiology*, 7(MAY). <http://doi.org/10.3389/fmicb.2016.00650>
- Berendsen, R. L., Pieterse, C. M. J., & Bakker, P. A. H. M. (2012). The rhizosphere microbiome and plant health. *Trends in Plant Science*, 17(8), 478–486. <http://doi.org/10.1016/j.tplants.2012.04.001>
- Berg, G. (2009). Plant-microbe interactions promoting plant growth and health: Perspectives for controlled use of microorganisms in agriculture. *Applied Microbiology and Biotechnology*, 84(1), 11–18. <http://doi.org/10.1007/s00253-009-2092-7>
- Berg, G., Grube, M., Schloter, M., & Smalla, K. (2014). Unraveling the plant microbiome: looking back and future perspectives. *Frontiers in Microbiology*, 5, 1–7. <http://doi.org/10.3389/fmicb.2014.00148>
- Berg, G., Rybakova, D., Grube, M., & Köberl, M. (2016). The plant microbiome explored: Implications for experimental botany. *Journal of Experimental Botany*, 67(4), 995–1002. <http://doi.org/10.1093/jxb/erv466>
- Berg, G., & Smalla, K. (2009). Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiology Ecology*, 68(1), 1–13. <http://doi.org/10.1111/j.1574-6941.2009.00654.x>
- Bhattacharjee, R. B., Singh, A., & Mukhopadhyay, S. N. (2008). Use of nitrogen-fixing bacteria as biofertiliser for non-legumes: Prospects and challenges. *Applied Microbiology and Biotechnology*, 80(2), 199–209. <http://doi.org/10.1007/s00253-008-1567-2>
- Bodenhausen, N., Horton, M. W., & Bergelson, J. (2013). Bacterial Communities Associated with the Leaves and the Roots of *Arabidopsis thaliana*. *PLoS ONE*, 8(2).

- <http://doi.org/10.1371/journal.pone.0056329>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.
<http://doi.org/10.1093/bioinformatics/btu170>
- Bond, P. L., Hugenholtz, P., Keller, J., & Blackall, L. L. (1995). Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. *Applied and Environmental Microbiology*, *61*(5), 1910–1916. <http://doi.org/10.1002/anie.200350930>
- Bouffaud, M., Renoud, S., Moënne-locco, Y., & Muller, D. (2016). Is plant evolutionary history impacting recruitment of diazotrophs and nifH expression in the rhizosphere? *Nature Publishing Group*, (February), 1–9.
<http://doi.org/10.1038/srep21690>
- Brader, G., Compant, S., Mitter, B., Trognitz, F., & Sessitsch, A. (2014). Metabolic potential of endophytic bacteria. *Current Opinion in Biotechnology*, *27*, 30–37.
<http://doi.org/10.1016/j.copbio.2013.09.012>
- Bravo, J. A., Forsythe, P., Chew, M. V., Escaravage, E., Savignac, H. M., Dinan, T. G., ... Cryan, J. F. (2011). Ingestion of *Lactobacillus* strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. *Proceedings of the National Academy of Sciences*, *108*(38), 16050–16055.
<http://doi.org/10.1073/pnas.1102999108>
- Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, *27*(4), 325–349.
<http://doi.org/10.2307/1942268>
- Brooke, J. S. (2012). *Stenotrophomonas maltophilia*: An emerging global opportunistic pathogen. *Clinical Microbiology Reviews*, *25*(1), 2–41.
<http://doi.org/10.1128/CMR.00019-11>
- Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., ... Schulze-Lefert, P. (2012). Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature*, *488*(7409), 91–95.
<http://doi.org/10.1038/nature11336>
- Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E. V. L., & Schulze-Lefert, P.

- (2013). Structure and Functions of the Bacterial Microbiota of Plants. *Annual Review of Plant Biology*. <http://doi.org/10.1146/annurev-arplant-050312-120106>
- Busby, P. E., Soman, C., Wagner, M. R., Friesen, M. L., Kremer, J., Bennett, A., ... Dangl, J. L. (2017). Research priorities for harnessing plant microbiomes in sustainable agriculture, (5), 1–14.
- Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal*, 11(12), 2639–2643. <http://doi.org/10.1038/ismej.2017.119>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., & A, A. J. (2016). DADA2: High resolution sample inference from Illumina amplicon data. *Nat Methods*, 13(7), 581–583. <http://doi.org/10.1038/nmeth.3869.DADA2>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A., & Holmes, S. P. (2016). DADA2: High resolution sample inference from Illumina amplicon data. *Nat Methods*, 13(7), 581–583. <http://doi.org/10.1038/nmeth.3869>
- Carlos, N., Tang, Y.-W., & Pei, Z. (2012). Pearls and pitfalls of genomics-based microbiome analysis. <http://doi.org/doi:10.1038/emi.2012.41>
- Chaparro, J. M., Badri, D. V., & Vivanco, J. M. (2014). Rhizosphere microbiome assemblage is affected by plant development. *ISME Journal*, 8(4), 790–803. <http://doi.org/10.1038/ismej.2013.196>
- Chen, H., Wu, H., Yan, B., Zhao, H., Liu, F., Zhang, H., ... Liang, Z. (2018). Core Microbiome of Medicinal Plant *Salvia miltiorrhiza* Seed: A Rich Reservoir of Beneficial Microbes for Secondary Metabolism? *International Journal of Molecular Sciences*, 19(672), 1–15.
- Chen, L., Reeve, J., Zhang, L., Huang, S., Wang, X., & Chen, J. (2018). GMPR: A robust normalization method for zero-inflated count data with application to microbiome sequencing data. *PeerJ*, 6(e4600).
- Cheng, X. Y., Tian, X. L., Wang, Y. S., Lin, R. M., Mao, Z. C., Chen, N., & Xie, B. Y. (2013). Metagenomic analysis of the pinewood nematode microbiome reveals a symbiotic relationship critical for xenobiotics degradation. *Scientific Reports*, 3, 1–10. <http://doi.org/10.1038/srep01869>
- Claude E. Shannon. (1948). A Mathematical Theory of Communication. *Bell System*

- Technical Journal*, 27(July 1928), 623–656. <http://doi.org/10.2307/3611062>
- Coleman-Derr, D., Desgarenes, D., Fonseca-Garcia, C., Gross, S., Clingenpeel, S., Woyke, T., ... Tringe, S. G. (2016). Plant compartment and biogeography affect microbiome composition in cultivated and native *Agave* species. *New Phytologist*, 209(2), 798–811. <http://doi.org/10.1111/nph.13697>
- Compant, S., Clément, C., & Sessitsch, A. (2010). Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. *Soil Biology and Biochemistry*, 42(5), 669–678. <http://doi.org/10.1016/j.soilbio.2009.11.024>
- Compant, S., Reiter, B., Sessitsch, A., Nowak, J., Clément, C., & Barka, E. A. (2005). Endophytic Colonization of *Vitis vinifera* L. by Plant Growth- Promoting Bacterium *Burkholderia* sp. Strain PsJN. *Applied and environmental microbiology*, 71(4), 1685–1693. <http://doi.org/10.1128/AEM.71.4.1685>
- Coombs, J. T., & Franco, C. M. M. (2003). Visualization of an Endophytic *Streptomyces* Species in Wheat Seed Justin, 69(7), 4260–4262. <http://doi.org/10.1128/AEM.69.7.4260>
- Copeland, J. K., Yuan, L., Layeghifard, M., Wang, P. W., & Guttman, D. S. (2015). Seasonal Community Succession of the Phyllosphere Microbiome. *Molecular Plant-Microbe Interactions*, 28(3), 274–285. <http://doi.org/10.1094/MPMI-10-14-0331-FI>
- Dagneu, A., Shibu, S., Debebe, A., Lemma, A., Dessalegn, L., Berhanu, B., ... Sierra, Y. M. (2012). Micropropagation of Banana Varieties (*Musa* spp.) Using Shoot-Tip Culture. *Ethiop. J. Agric. Sci.*, 22, 14–25.
- Dahan, D., Jude, B. A., Lamendella, R., Keesing, F., & Perron, G. G. (2018). Exposure to arsenic alters the microbiome of larval zebrafish. *Frontiers in Microbiology*, 9(JUN), 1–12. <http://doi.org/10.3389/fmicb.2018.01323>
- de Almeida, C. V., Andreote, F. D., Yara, R., Tanaka, F. A. O., Azevedo, J. L., & de Almeida, M. (2009). Bacteriosomes in axenic plants: Endophytes as stable endosymbionts. *World Journal of Microbiology and Biotechnology*, 25(10), 1757–1764. <http://doi.org/10.1007/s11274-009-0073-8>
- De Deyn, G. B., Cornelissen, J. H. C., & Bardgett, R. D. (2008). Plant functional traits and soil carbon sequestration in contrasting biomes. *Ecology Letters*, 11(5), 516–

531. <http://doi.org/10.1111/j.1461-0248.2008.01164.x>
- Debergh, P. C., & Read, P. E. (1991). Micropropagation P.e. In *Micropropagation Technology and Application* (pp. 1–13).
- Delmotte, N., Knief, C., Chaffron, S., Innerebner, G., Roschitzki, B., Schlapbach, R., ... Vorholt, J. A. (2009). Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proceedings of the National Academy of Sciences*, *106*(38), 16428–16433. <http://doi.org/10.1073/pnas.0905240106>
- Dennis, P. G., Miller, A. J., & Hirsch, P. R. (2010). Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiology Ecology*, *72*(3), 313–327. <http://doi.org/10.1111/j.1574-6941.2010.00860.x>
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., ... Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, *72*(7), 5069–5072. <http://doi.org/10.1128/AEM.03006-05>
- Dewdney, J., Lynne Reuber, T., Wildermuth, M. C., Devoto, A., Cui, J., Stutius, L. M., ... Ausubel, F. M. (2000). Three unique mutants of Arabidopsis identify eds loci required for limiting growth of a biotrophic fungal pathogen. *Plant Journal*, *24*(2), 205–218. <http://doi.org/10.1046/j.1365-313X.2000.00870.x>
- Dhariwal, A., Chong, J., Habib, S., King, I. L., Agellon, L. B., & Xia, J. (2017). MicrobiomeAnalyst: A web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Research*, *45*(W1), W180–W188. <http://doi.org/10.1093/nar/gkx295>
- Dohm, J. C., Lottaz, C., Borodina, T., & Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, *36*(16). <http://doi.org/10.1093/nar/gkn425>
- Dong, Z., Canny, M. J., McCully, M. E., Roboredo, M. R., Cabadilla, C. F., Ortega, E., & Rodes, R. (1994). A nitrogen-fixing endophyte of sugarcane stems: A new role for the apoplast. *Plant Physiology (Rockville)*, *105*(4), 1139–1147. <http://doi.org/10.1104/pp.105.4.1139>
- Doornbos, R. F., Van Loon, L. C., & Bakker, P. A. H. M. (2012). Impact of root exudates

- and plant defense signaling on bacterial communities in the rhizosphere. A review. *Agronomy for Sustainable Development*, 32(1), 227–243.
<http://doi.org/10.1007/s13593-011-0028-y>
- Dyall, S. D., Brown, M. T., & Johnson, P. J. (2004). Ancient Invasions: From Endosymbionts to Organelles. *Science*, 304(5668), 253–257.
<http://doi.org/10.1126/science.1094884>
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, 14(9), 755–763.
[http://doi.org/10.1016/S0896-6273\(00\)80510-3](http://doi.org/10.1016/S0896-6273(00)80510-3)
- Edgar, R. C., & Flyvbjerg, H. (2015). Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*, 31(21), 3476–3482.
<http://doi.org/10.1093/bioinformatics/btv401>
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16), 2194–2200. <http://doi.org/10.1093/bioinformatics/btr381>
- Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., ... Sundaresan, V. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proceedings of the National Academy of Sciences*, 112(8), E911–E920. <http://doi.org/10.1073/pnas.1414592112>
- Elbeltagy, A., Nishioka, K., Sato, T., Suzuki, H., Ye, B., Hamada, T., ... Minamisawa, K. (2001). Endophytic Colonization and In Planta Nitrogen Fixation by a *Herbaspirillum* sp. *Applied and Environmental Microbiology*, 67(11), 5285–5293.
<http://doi.org/10.1128/AEM.67.11.5285>
- Faith, D. P., & Baker, A. M. (2006). Phylogenetic Diversity (PD) and Biodiversity Conservation: Some Bioinformatics Challenges. *Evolutionary Bioinformatics*, 2, 117693430600200. <http://doi.org/10.1177/117693430600200007>
- Faria, D. C., Dias, A. C. F., Melo, I. S., & de Carvalho Costa, F. E. (2013). Endophytic bacteria isolated from orchid and their potential to promote plant growth. *World Journal of Microbiology and Biotechnology*, 29(2), 217–221.
<http://doi.org/10.1007/s11274-012-1173-4>
- Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., ... Pradhan, S. (2013). A Method for Selectively Enriching Microbial DNA from

- Contaminating Vertebrate Host DNA. *PLoS ONE*, 8(10).
<http://doi.org/10.1371/journal.pone.0076096>
- Fernandes, T. P., Nietsche, S., Costa, M. R., Xavier, A. A., Pereira, D. F. G. S., & Pereira, M. C. T. (2013). Potential use of endophytic bacteria to promote the plant growth of micropropagated banana cultivar Prata Ana. *African Journal of Biotechnology*, 12(31), 4915–4919. <http://doi.org/10.5897/AJB2012.2958>
- Fletcher, J., Leach, J. E., Eversole, K., & Tauxe, R. (2013). Human Pathogens on Plants: Designing a Multidisciplinary Strategy for Research. *Phytopathology*, 103(4), 306–315. <http://doi.org/10.1094/PHYTO-09-12-0236-IA>
- Fukami, T. (2015). Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 1–23. <http://doi.org/10.1146/annurev-ecolsys-110411-160340>
- García, G., Posada U., L.F., Correa, J. & Villegas, V. (2017). Efecto de la aplicación de microorganismos en la microbiota, desarrollo y salud vegetal de plantas de banano (sin publicar)
- George, E. F., Hall, M. A., & Klerk, G.-J. De. (2008). Micropropagation: Uses and Methods. In *Plant Propagation by Tissue Culture* (pp. 29–64).
<http://doi.org/10.1007/978-1-4020-5005-3>
- Giesler, L. J., & Yuen, G. Y. (1998). Evaluation of *Stenotrophomonas maltophilia* strain C3 for biocontrol of brown patch disease. *Crop Protection*, 17(6), 509–513.
[http://doi.org/10.1016/S0261-2194\(98\)00049-0](http://doi.org/10.1016/S0261-2194(98)00049-0)
- Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., ... Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science*, 312(5778), 1355–1359. <http://doi.org/10.1126/science.1124234>
- Goodfellow, M., Kämpfer, P., Chun, J., Vos, P. De, Rainey, F., Whitman, W. B., ... Tully, J. G. (1984). *Bergey's Manual of Systematic Bacteriology second edition*. Springer-Verlag New York Inc. <http://doi.org/10.1007/978-0-387-68572-4>
- Goodrich, J. K., Rienzi, S. C. Di, Poole, A. C., Koren, O., Walters, W. A., Caporaso, J. G., ... Ley, R. E. (2014). Conducting a Microbiome Study. *Cell*, 158(2), 250–262.
- Gopal, M., Bhute, S. S., Gupta, A., Prabhu, S. R., Thomas, G. V., Whitman, W. B., & Jangid, K. (2017). Changes in structure and function of bacterial communities

- during coconut leaf vermicomposting. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 110(10), 1339–1355.
<http://doi.org/10.1007/s10482-017-0894-7>
- Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, 4, 379–391. <http://doi.org/10.1046/j.1461-0248.2001.00230.x>
- Gregory Caporaso, J., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., ... Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5), 335–336.
<http://doi.org/10.1038/nmeth0510-335>
- Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., ... Birren, B. W. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, 21(March), 494–504. <http://doi.org/10.1101/gr.112730.110>
- Hallmann, J., Quadt-Hallmann, A., Mahaffee, W. F., & Kloepper, J. W. (1997). Bacterial endophytes in agricultural crops. *Can. J. Microbiol.*, 43(10), 895–914.
<http://doi.org/10.1139/m97-131>
- Hamonts, K., Trivedi, P., Garg, A., Janitz, C., Grinyer, J., Holford, P., ... Singh, B. K. (2018). Field study reveals core plant microbiota and relative importance of their drivers. *Environmental Microbiology*, 20(1), 124–140.
- Handelsman, J., & Handelsman, J. (2004). Metagenomics : Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, 68(4), 669–685. <http://doi.org/10.1128/MBR.68.4.669>
- Hanshewa, A. S., Masonb, C. J., Raffa, K. F., & Curriea, C. R. (2013). Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities, 95(2), 149–155.
<http://doi.org/10.1016/j.mimet.2013.08.007.Minimization>
- Hao, Y., Pei, Z., & Brown, S. M. (2017). Bioinformatics in Microbiome Analysis. In *The Human Microbiome* (pp. 1–18).
- Hardoim, P. R., Hardoim, C. C. P., van Overbeek, L. S., & van Elsas, J. D. (2012). Dynamics of seed-borne rice endophytes on early plant growth stages. *PLoS ONE*,

- 7(2). <http://doi.org/10.1371/journal.pone.0030438>
- Hardoim, P. R., Overbeek, L. S. van, & Elsas, J. D. van. (2008). Properties of bacterial endophytes and their proposed role in plant growth. *Cell*, *16*(10), 463–471. <http://doi.org/10.1016/j.tim.2008.07.008>
- Hardoim, P. R., van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., ... Sessitsch, A. (2015). The Hidden World within Plants: Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes. *Microbiology and Molecular Biology Reviews*, *79*(3), 293–320. <http://doi.org/10.1128/MMBR.00050-14>
- Hartman, K., van der Heijden, M. G. A., Roussely-Provent, V., Walser, J. C., & Schlaeppi, K. (2017). Deciphering composition and function of the root microbiome of a legume plant. *Microbiome*, *5*(1), 1–13. <http://doi.org/10.1186/s40168-016-0220-z>
- Hirsch, P. R., & Mauchline, T. H. (2012). Who's who in the plant root microbiome? *Nature Biotechnology*, *30*(10), 961–962.
- Hirsch, P. R., Mauchline, T. H., & Clark, I. M. (2010). Culture-independent molecular techniques for soil microbial ecology. *Soil Biology and Biochemistry*, *42*(6), 878–887. <http://doi.org/10.1016/j.soilbio.2010.02.019>
- Hodkinson, B. P., & Grice, E. A. (2015). Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research. *Advances in Wound Care*, *4*(1), 50–58. <http://doi.org/10.1089/wound.2014.0542>
- Horner-Devine, M. C. C., Lage, M., Hughes, J. B. B., & Bohannan, B. J. M. J. M. (2004). Is there a species-area relationship for bacteria? *Nature*, *432*(7018), 750–753. <http://doi.org/10.1109/KORUS.2000.865987>
- Huse, S. M., Welch, D. M., Morrison, H. G., & Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, *12*(7), 1889–1898. <http://doi.org/10.1111/j.1462-2920.2010.02193.x>
- Iniguez, A. L., Dong, Y., & Triplett, E. W. (2004). Nitrogen Fixation in Wheat Provided by *Klebsiella pneumoniae* 342. *Molecular Plant-Microbe Interactions*, *17*(10), 1078–1085. <http://doi.org/10.1094/MPMI.2004.17.10.1078>
- Innerebner, G., Knief, C., & Vorholt, J. A. (2011). Protection of *Arabidopsis thaliana*

- against leaf-pathogenic *Pseudomonas syringae* by *Sphingomonas* strains in a controlled model system. *Applied and Environmental Microbiology*, 77(10), 3202–3210. <http://doi.org/10.1128/AEM.00133-11>
- Jacott, C., Murray, J., & Ridout, C. (2017). Trade-Offs in Arbuscular Mycorrhizal Symbiosis: Disease Resistance, Growth Responses and Perspectives for Crop Breeding. *Agronomy*, 7(4), 75. <http://doi.org/10.3390/agronomy7040075>
- Janssen, P. H. (2006). Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. *Applied and environmental microbiology*, 72(3), 1719–1728. <http://doi.org/10.1128/AEM.72.3.1719>
- Johnston-Monje, D., & Raizada, M. N. (2011). Conservation and diversity of seed associated endophytes in *Zea* across boundaries of evolution, ethnography and ecology. *PLoS ONE*, 6(6). <http://doi.org/10.1371/journal.pone.0020396>
- Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system, 444(November), 323–329. <http://doi.org/10.1038/nature05286>
- Jovel, J., Patterson, J., Wang, W., Hotte, N., O’Keefe, S., Mitchel, T., ... Wong, G. K. S. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Frontiers in Microbiology*, 7(APR), 1–17. <http://doi.org/10.3389/fmicb.2016.00459>
- Joy, P. P., Anjana, R., & Jose, P. (2013). Protocol for micropropagation of banana. *Pineapple Research Station (Kerala Agricultural University)*, (January 2013), 153–158.
- Kandel, S., Joubert, P., & Doty, S. (2017). Bacterial Endophyte Colonization and Distribution within Plants. *Microorganisms*, 5(4), 77. <http://doi.org/10.3390/microorganisms5040077>
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27–30. <http://doi.org/10.1093/nar/27.1.29>
- Kawasaki, A., Donn, S., Ryan, P. R., Mathesius, U., Devilla, R., Jones, A., & Watt, M. (2016). Microbiome and exudates of the root and rhizosphere of brachypodium distachyon, a model for wheat. *PLoS ONE*, 11(10). <http://doi.org/10.1371/journal.pone.0164533>
- Kraemer, J. G., Ramette, A., Aebi, S., Oppliger, A., & Hilty, M. (2018). Influence of pig farming on the human’s nasal microbiota: The key role of Airbone Microbial

- Communities, *84*(6), 1–13. <http://doi.org/10.1128/AEM.02470-17>
- Kraker-Castañeda, C., & Cóbar-Carranza, A. J. (2011). Uso de rarefacción para comparación de la riqueza de especies: el caso de las aves de sotobosque en la zona de influencia del Parque Nacional Laguna Lachuá, Guatemala. *Naturaleza y Desarrollo*, *9*(1), 60–70.
- Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., ... Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, *31*(9), 814–821. <http://doi.org/10.1038/nbt.2676>
- Lebeis, S. L. (2014). The potential for give and take in plant-microbiome relationships. *Frontiers in Plant Science*, *5*(June). <http://doi.org/10.3389/fpls.2014.00287>
- Lebeis, S. L. (2015). Greater than the sum of their parts: Characterizing plant microbiomes at the community-level. *Current Opinion in Plant Biology*, *24*, 82–86. <http://doi.org/10.1016/j.pbi.2015.02.004>
- Lemanceau, P., Blouin, M., Muller, D., & Moëgne-Loccoz, Y. (2017). Let the Core Microbiota Be Functional. *Trends in Plant Science*, *22*(7), 583–595. <http://doi.org/10.1016/j.tplants.2017.04.008>
- Lindow, S. E., & Brandl, M. T. (2003). Minireview Microbiology of the Phyllosphere. *Applied and Environmental Microbiology*, *69*(4), 1875–1883. <http://doi.org/10.1128/AEM.69.4.1875>
- Louca, S., & Doebeli, M. (2018). Efficient comparative phylogenetics on large trees. *Bioinformatics*, *34*(6), 1053–1055. <http://doi.org/10.1093/bioinformatics/btx701>
- Lucero, M. E., Unc, A., Cooke, P., Dowd, S., & Sun, S. (2011). Endophyte Microbiome Diversity in Micropropagated *Atriplex canescens* and *Atriplex torreyivar griffithsii*. *PLoS ONE*, *6*(3), 1–12. <http://doi.org/10.1371/journal.pone.0017693>
- Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., ... Dangl, J. L. (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature*, *488*(7409), 86–90. <http://doi.org/10.1038/nature11237>
- Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., & Dangl, J. L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, *10*(10), 999–1002. <http://doi.org/10.1038/nmeth.2634>

- Lutz, K. A., Wang, W., Zdepski, A., & Michael, T. P. (2011). Isolation and analysis of high quality nuclear DNA with reduced organellar DNA for plant genome sequencing and resequencing. *BMC Biotechnology*, *11*(1), 54. <http://doi.org/10.1186/1472-6750-11-54>
- M.J.Pieterse, C., Christos Zamioudis, Berendsen, R. L., Weller, D. M., Wees, S. C. M. Van, & Bakker, P. A. H. M. (2014). Induced Systemic Resistance by Beneficial Microbes. *Annu. Rev. Phytopathol*, *52*, 347–375. <http://doi.org/10.1146/annurev-phyto-082712-102340>
- Maidak, B. L., Cole, J. R., Lilburn, T. G., Jr, C. T. P., Saxman, P. R., RyanJ.Farris, ... Tiedje, J. M. (2001). The RDP-II (Ribosomal Database Project) Bonnie. *Hospital Management*, *29*(1), 173–174. <http://doi.org/10.1093/nar/29.1.173>
- Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., ... Foster, G. D. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*, *13*(6), 614–629. <http://doi.org/10.1111/j.1364-3703.2012.00804.x>
- Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., & Neufeld, J. D. (2012). PANDAseq: Paired-end assembler for illumina sequences. *BMC Bioinformatics*, *13*(1), 31. <http://doi.org/10.1186/1471-2105-13-31>
- Mathys, J., De Cremer, K., Timmermans, P., Van Kerckhove, S., Lievens, B., Vanhaecke, M., ... De Coninck, B. (2012). Genome-Wide Characterization of ISR Induced in *Arabidopsis thaliana* by *Trichoderma hamatum* T382 Against *Botrytis cinerea* Infection. *Frontiers in Plant Science*, *3*(May), 1–25. <http://doi.org/10.3389/fpls.2012.00108>
- Mendes, R., Garbeva, P., & Raaijmakers, J. M. (2013). The rhizosphere microbiome: Significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiology Reviews*, *37*(5), 634–663. <http://doi.org/10.1111/1574-6976.12028>
- Mendes, R., Kruijt, M., De Bruijn, I., Dekkers, E., Van Der Voort, M., Schneider, J. H. M., ... Raaijmakers, J. M. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science*, *332*(6033), 1097–1100. <http://doi.org/10.1126/science.1203980>

- Mendes, R., Pizzirani-Kleiner, A. A., Araujo, W. L., & Raaijmakers, J. M. (2007). Diversity of cultivated endophytic bacteria from sugarcane: genetic and biochemical characterization of Burkholderia cepacia complex isolates. *Applied and Environmental Microbiology*, 73(22), 7259–7267. <http://doi.org/10.1128/AEM.01222-07>
- Mercado-Blanco, J., & Lugtenberg, B. J. J. (2014). Biotechnological Applications of Bacterial Endophytes. *Current Biotechnology*, 3(1), 60–75. <http://doi.org/10.3934/bioeng.2015.3.163>
- Meyer, J. B., Lutz, M. P., Frapolli, M., Péchy-Tarr, M., Rochat, L., Keel, C., ... Maurhofer, M. (2010). Interplay between wheat cultivars, biocontrol pseudomonads, and soil. *Applied and Environmental Microbiology*, 76(18), 6196–6204. <http://doi.org/10.1128/AEM.00752-10>
- Morris, E. K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T. S., ... Rillig, M. C. (2014). Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution*, 4(18), 3514–3524. <http://doi.org/10.1002/ece3.1155>
- Müller, D. B., Vogel, C., Bai, Y., & Vorholt, J. A. (2016). The Plant Microbiota: Systems-Level Insights and Perspectives. *Annual Review of Genetics*, 50(1), 211–234. <http://doi.org/10.1146/annurev-genet-120215-034952>
- Murashige, T. (1974). Plant propagation through tissue cultures. *Ann. Rev. Plant Physiol*, 25, 135–166.
- Murashige, T.; Skoog, F. (1962). A revised medium for rapid growth an bioassays with tobacco tissue cultures. *Physioly Plant*, (15), 473–497. <http://doi.org/10.1111/j.1399-3054.1962.tb08052.x>
- Murthy, B. N. S., Vettakkorumakankav, N. ., KrishnaRaj, S., Odumeru, J., & Saxena, P. K. (1999). Characterization of somatic embryogenesis in Pelargonium × hortorum mediated by a bacterium. *Plant Cell Reports (1999)*, 18, 607–613.
- Nair, J. R., Singh, G., & Sekar, V. (2002). Isolation and characterization of a novel Bacillus strain from coffee phyllosphere showing antifungal activity. *Journal of Applied Microbiology*, 93(5), 772–780. <http://doi.org/10.1046/j.1365-2672.2002.01756.x>

- Naveed, M., Mitter, B., Yousaf, S., Pastar, M., Afzal, M., & Sessitsch, A. (2013). The endophyte *Enterobacter* sp. FD17: A maize growth enhancer selected based on rigorous testing of plant beneficial traits and colonization characteristics. *Biology and Fertility of Soils*, *50*(2), 249–262. <http://doi.org/10.1007/s00374-013-0854-y>
- Newman, M.-A., Dow, J. M., Molinaro, A., & Parrilli, M. (2007). Priming, induction and modulation of plant defence responses by bacterial lipopolysaccharides. *Journal of Endotoxin Research*, *13*(2), 71–84.
- Nossa, C. W., Oberdorf, W. E., Yang, L., Aas, J. A., Paster, B. J., Desantis, T. Z., ... National, L. B. (2010). Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome, *16*(33), 4135–4144. <http://doi.org/10.3748/wjg.v16.i33.4135>
- Orlikowska, T., Nowak, K., & Reed, B. (2017). Bacteria in the plant tissue culture environment. *Plant Cell, Tissue and Organ Culture*, *128*(3), 487–508. <http://doi.org/10.1007/s11240-016-1144-9>
- Pandey, A., Palni, L. M. S., & Bag, N. (2000). Biological hardening of tissue culture raised tea plants through rhizosphere bacteria. *Biotechnology Letters*, *22*(13), 1087–1091. <http://doi.org/10.1023/A:1005674803237>
- Pandey, R. (2015). Mineral nutrition of plants. In *Plant Biology and Biotechnology* (Vol. 1, pp. 1–827). <http://doi.org/10.1007/978-81-322-2286-6>
- Panke-Buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E., & Kao-Kniffin, J. (2015). Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME Journal*, *9*(4), 980–989. <http://doi.org/10.1038/ismej.2014.196>
- Parray, J. A., Kamili, A. N., Reshi, Z. A., Qadri, R. A., & Jan, S. (2015). Interaction of rhizobacterial strains for growth improvement of *Crocus sativus* L. under tissue culture conditions. *Plant Cell, Tissue and Organ Culture*, *121*(2), 325–334. <http://doi.org/10.1007/s11240-014-0703-1>
- Paulson, J. N., Stine, O. C., Bravo-Corrada, H., & Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, *10*, 1200–1202. <http://doi.org/10.1038/nmeth.2658>
- Peiffer, J. A., Spor, A., Koren, O., Jin, Z., Tringe, S. G., Dangl, J. L., ... Ley, R. E. (2013). Diversity and heritability of the maize rhizosphere microbiome under field

- conditions. *Proceedings of the National Academy of Sciences*, 110(16), 6548–6553.
<http://doi.org/10.1073/pnas.1302837110>
- Pérez-Jaramillo, J. E., Carrión, V. J., Bosse, M., Ferrão, L. F. V., De Hollander, M., Garcia, A. A. F., ... Raaijmakers, J. M. (2017). Linking rhizosphere microbiome composition of wild and domesticated *Phaseolus vulgaris* to genotypic and root phenotypic traits. *ISME Journal*, 11(10), 2244–2257.
<http://doi.org/10.1038/ismej.2017.85>
- Philippot, L., Raaijmakers, J. M., Lemanceau, P., & Van Der Putten, W. H. (2013). Going back to the roots: The microbial ecology of the rhizosphere. *Nature Reviews Microbiology*, 11(11), 789–799. <http://doi.org/10.1038/nrmicro3109>
- Pii, Y., Borruso, L., Brusetti, L., Crecchio, C., Cesco, S., & Mimmo, T. (2016). The interaction between iron nutrition, plant species and soil type shapes the rhizosphere microbiome. *Plant Physiology and Biochemistry*, 99, 39–48.
<http://doi.org/10.1016/j.plaphy.2015.12.002>
- Poppenberger, B. P., Eonhardt, W. L., & Edl, H. R. (2002). Latent persistence of *Agrobacterium vitis* in micropropagated *Vitis vinifera*. *Vitis*, 41(2), 113–114.
- Posada U., L. F. (2017). *Promoción de crecimiento vegetal de Bacillus subtilis EA-CB0575, colonización rizosférica y potencial genómico y bioquímico*. Retrieved from <http://www.bdigital.unal.edu.co/57419/>
- Pruesse, E., Peplies, J., & Glöckner, F. O. (2012). SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, 28(14), 1823–1829. <http://doi.org/10.1093/bioinformatics/bts252>
- Quambusch, M., Pirttila, A. M., Tejesvi, M. V., Winkelmann, T., & Bartsch, M. (2014). Endophytic bacteria in plant tissue culture: Differences between easy-and difficult-to-propagate *Prunus avium* genotypes. *Journal of Refugee Studies*, 34(5), 524–533.
<http://doi.org/10.1093/treephys/tpu027>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... Glockner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(1), 590–596.
<http://doi.org/10.1093/nar/gks1219>
- Raaijmakers, J. M., de Bruijn, I., Nybroe, O., & Ongena, M. (2010). Natural functions of

- lipopeptides from *Bacillus* and *Pseudomonas*: More than surfactants and antibiotics. *FEMS Microbiology Reviews*, 34(6), 1037–1062. <http://doi.org/10.1111/j.1574-6976.2010.00221.x>
- Rastogi, G., Sbdio, A., Tech, J. J., Suslow, T. V., Coaker, G. L., & Leveau, J. H. J. (2012). Leaf microbiota in an agroecosystem: Spatiotemporal variation in bacterial community composition on field-grown lettuce. *ISME Journal*, 6(10), 1812–1822. <http://doi.org/10.1038/ismej.2012.32>
- Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-Throughput Sequencing Technologies. *Molecular Cell*, 58(4), 586–597. <http://doi.org/10.1016/j.molcel.2015.05.004>
- Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics*, 13(5), 278–289. <http://doi.org/10.1016/j.gpb.2015.08.002>
- Robinson, J. C., Fraser, C., & Eckstein, K. (1993). A field comparison of conventional suckers with tissue culture banana planting material over three crop cycles. *Journal of Horticultural Science*, 68, 831–836. <http://doi.org/10.1080/00221589.1993.11516420>
- Romero, F. M., Marina, M., & Pieckenstain, F. L. (2014). The communities of tomato (*Solanum lycopersicum* L.) leaf endophytic bacteria, analyzed by 16S-ribosomal RNA gene pyrosequencing. *FEMS Microbiology Letters*, 351(2), 187–194. <http://doi.org/10.1111/1574-6968.12377>
- Rosado, D., Pérez-Losada, M., Severino, R., Cable, J., & Xavier, R. (2019). Characterization of the skin and gill microbiomes of the farmed seabass (*Dicentrarchus labrax*) and seabream (*Sparus aurata*). *Aquaculture*, 500(April 2018), 57–64. <http://doi.org/10.1016/j.aquaculture.2018.09.063>
- Rosen, M. J., Callahan, B. J., Fisher, D. S., & Holmes, S. P. (2012). Denoising PCR-amplified metagenome data. *BMC Bioinformatics* 2012, 13(283), 1–16.
- Rosenblueth, M., & Martínez-Romero, E. (2006). Bacterial Endophytes and Their Interactions with Hosts. *Molecular Plant-Microbe Interactions*, 19(8), 827–837. <http://doi.org/10.1094/MPMI-19-0827>
- Rossmann, M., Sarango-Flores, S. W., Barros Chiaramonte, J., Pezzo Kmit, M. C., &

- Mendes, R. (2017). Plant Microbiome: Composition and Functions in Plant Compartments. In *The Brazilian Microbiome: Current Status and Perspectives* (pp. 7–20). <http://doi.org/10.1007/978-3-319-59997-7>
- Rossmann, M., Sarango-Flores, S. W., Chiaramonte, J. B., Kmit, M. C. P., & Mendes, R. (2017). Plant Microbiome: Composition and Functions in plant Compartments. In *The Brazilian Microbiome: Current Status and Perspectives* (pp. 1–123). <http://doi.org/10.1007/978-3-319-59997-7>
- Saini, R., & Jaiwal, P. K. (2002). Age, position in mother seedling, orientation, and polarity of the epicotyl segments of blackgram (*Vigna mungo* L . Hepper) determines its morphogenic response, *163*, 101–109.
- Sakai, M., Matsuka, A., Komura, T., & Kanazawa, S. (2004). Application of a new PCR primer for terminal restriction fragment length polymorphism analysis of the bacterial communities in plant roots. *Journal of Microbiological Methods*, *59*(1), 81–89. <http://doi.org/10.1016/j.mimet.2004.06.005>
- Salomon, M. V., Bottini, R., de Souza Filho, G. A., Cohen, A. C., Moreno, D., Gil, M., & Piccoli, P. (2014). Bacteria isolated from roots and rhizosphere of *Vitis vinifera* retard water losses, induce abscisic acid accumulation and synthesis of defense-related terpenes in in vitro cultured grapevine. *Physiologia Plantarum*, *151*(4), 359–374. <http://doi.org/10.1111/ppl.12117>
- Santi, C., Bogusz, D., & Franche, C. (2013). Biological nitrogen fixation in non-legume plants. *Annals of Botany*, *111*(5), 743–767. <http://doi.org/10.1093/aob/mct048>
- Santoyo, G., Moreno-Hagelsieb, G., del Carmen Orozco-Mosqueda, M., & Glick, B. R. (2016). Plant growth-promoting bacterial endophytes. *Microbiological Research*, *183*, 92–99. <http://doi.org/10.1016/j.micres.2015.11.008>
- Sarhan, M. S., Patz, S., Hamza, M. A., Youssef, H. H., Mourad, E. F., Fayed, M., ... Hegazi, N. A. (2018). G3 PhyloChip Analysis Confirms the Promise of Plant-Based Culture Media for Unlocking the Composition and Diversity of the Maize Root Microbiome and for Recovering Unculturable Candidate Divisions/Phyla. *Microbes and Environments*, *33*(3), 317–325. <http://doi.org/10.1264/jsme2.ME18023>
- Schardl, C. L., Florea, S., Pan, J., Nagabhyru, P., Bec, S., & Calie, P. J. (2013). The epichloae: Alkaloid diversity and roles in symbiosis with grasses. *Current Opinion*

- in Plant Biology*, 16(4), 480–488. <http://doi.org/10.1016/j.pbi.2013.06.012>
- Schlaeppli, K., & Bulgarelli, D. (2015). The Plant Microbiome at Work. *Journal of the Ceramic Society of Japan*, 28(1), 212–217. <http://doi.org/10.1094/MPMI-10-14-0334-FI>
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., ... Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23), 7537–7541. <http://doi.org/10.1128/AEM.01541-09>
- Schulz, B. J. E., Boyle, C. J. C., & Sieber, T. N. (2006). *Microbiell Root Endophytes*. Springer (Vol. 9). <http://doi.org/10.1017/CBO9781107415324.004>
- Sessitsch, A., & Mitter, B. (2015). 21st century agriculture: Integration of plant microbiomes for improved crop production and food security. *Microbial Biotechnology*, 8(1), 32–33. <http://doi.org/10.1111/1751-7915.12180>
- Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Frontiers in Plant Science*, 5, 1–14.
- Sheldon, A. L. (1969). Equitability Indices: Dependence on the Species Count. *Ecology*, 50(3), 466–467. <http://doi.org/10.2307/1933900>
- Simberloff, D. (1972). Properties of the Rarefaction Diversity Measurement. *The American Naturalist*, 106(949), 414–418. <http://doi.org/10.1086/282781>
- Suman, S. (2017). PLANT TISSUE CULTURE : A PROMISING TOOL OF QUALITY MATERIAL PRODUCTION WITH SPECIAL REFERENCE TO MICROPROPAGATION OF BANANA. *Biochem. Cell. Arch.*, 17(1), 1–26.
- Sunayana, M. R., Sasikala, C., & Ramana, C. V. (2005). Rhodestrin: A novel indole terpenoid phytohormone from *Rhodobacter sphaeroides*. *Biotechnology Letters*, 27(23–24), 1897–1900. <http://doi.org/10.1007/s10529-005-3900-5>
- Thijs, S., De Beeck, M. O., Beckers, B., Truyens, S., Stevens, V., Van Hamme, J. D., ... Vangronsveld, J. (2017). Comparative evaluation of four bacteria-specific primer pairs for 16S rRNA gene surveys. *Frontiers in Microbiology*, 8(MAR), 1–15. <http://doi.org/10.3389/fmicb.2017.00494>
- Thomas, P. (2004). Isolation of *Bacillus pumilus* from in vitro grapes as a long-term

- alcohol-surviving and rhizogenesis inducing covert endophyte. *Journal of Applied Microbiology*, 97(1), 114–123. <http://doi.org/10.1111/j.1365-2672.2004.02279.x>
- Thomas, P. (2011). Intense association of non-culturable endophytic bacteria with antibiotic-cleansed in vitro watermelon and their activation in degenerating cultures, 2313–2325. <http://doi.org/10.1007/s00299-011-1158-z>
- Thomas, P., & Sekhar, A. C. (2016). Cultivation Versus Molecular Analysis of Banana (*Musa sp.*) Shoot-Tip Tissue Reveals Enormous Diversity of Normally Uncultivable Endophytic Bacteria. *Microb Ecol.*
- Timm, C. M., Campbell, A. G., Utturkar, S. M., Jun, S. R., Parales, R. E., Tan, W. A., ... Pelletier, D. A. (2015). Metabolic functions of *Pseudomonas fluorescens* strains from *Populus deltoides* depend on rhizosphere or endosphere isolation compartment. *Frontiers in Microbiology*, 6(OCT), 1–13. <http://doi.org/10.3389/fmicb.2015.01118>
- Toju, H., Peay, K. G., Yamamichi, M., Narisawa, K., Hiruma, K., Naito, K., ... Kiers, E. T. (2018). Core microbiomes for sustainable agroecosystems. *Nature Plants*, 4(5), 247–257. <http://doi.org/10.1038/s41477-018-0139-4>
- Toju, H., Yamamoto, S., Sato, H., Tanabe, A. S., Gilbert, G. S., & Kadowaki, K. (2013). Community composition of root-associated fungi in a *Quercus*-dominated temperate forest: “codominance” of mycorrhizal and root-endophytic fungi. *Ecology and Evolution*, 3(5), 1281–1293. <http://doi.org/10.1002/ece3.546>
- Truyens, S., Weyens, N., Cuypers, A., & Vangronsveld, J. (2015). Bacterial seed endophytes: Genera, vertical transmission and interaction with plants. *Environmental Microbiology Reports*, 7(1), 40–50. <http://doi.org/10.1111/1758-2229.12181>
- Turner, T. R., James, E. K., & Poole, P. S. (2013). The plant microbiome, 1–10. <http://doi.org/10.1186/gb-2013-14-6-209>
- Vandenkoornhuysse, P., Quaiser, A., Duhamel, M., Le Van, A., & Dufresne, A. (2015). The importance of the microbiome of the plant holobiont. *New Phytologist*, 206(4), 1196–1206. <http://doi.org/10.1111/nph.13312>
- Vereecke, D., Burssens, S., Simón-Mateo, C., Inzé, D., Van Montagu, M., Goethals, K., & Jaziri, M. (2000). The *Rhodococcus fascians*-plant interaction: Morphological traits and biotechnological applications. *Planta*, 210(2), 241–251.

<http://doi.org/10.1007/PL00008131>

- Větrovský, T., & Baldrian, P. (2013). The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS ONE*, 8(2), 1–10. <http://doi.org/10.1371/journal.pone.0057923>
- Vinje, H., Liland, K. H., Almøy, T., & Snipen, L. (2015). Comparing K-mer based methods for improved classification of 16S sequences. *BMC Bioinformatics*, 16(1), 1–13. <http://doi.org/10.1186/s12859-015-0647-4>
- Vorholt, J. A. (2012). Microbial life in the phyllosphere. *Nature Reviews Microbiology*, 10(12), 828–840. <http://doi.org/10.1038/nrmicro2910>
- Vuylsteke, D. . (1998). *Shoot-tip culture for the propagation, conservation and distribution of Musa germplasm*.
- Wagner, M. R., Lundberg, D. S., Del Rio, T. G., Tringe, S. G., Dangl, J. L., & Mitchell-Olds, T. (2016). Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nature Communications*, 7, 1–15. <http://doi.org/10.1038/ncomms12151>
- Walters, W. A., Jin, Z., Youngblut, N., Wallace, J. G., Sutter, J., Zhang, W., ... Ley, R. E. (2018). Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proceedings of the National Academy of Sciences*, 115(28), 7368–7373. <http://doi.org/10.1073/pnas.1800918115>
- Wani, Z. A., Ashraf, N., Mohiuddin, T., & Riyaz-Ul-Hassan, S. (2015). Plant-endophyte symbiosis, an ecological perspective. *Applied Microbiology and Biotechnology*, 99(7), 2955–2965. <http://doi.org/10.1007/s00253-015-6487-3>
- Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., ... Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5(1), 1–18. <http://doi.org/10.1186/s40168-017-0237-y>
- Wemheuer, F., Kaiser, K., Karlovsky, P., Daniel, R., Vidal, S., & Wemheuer, B. (2017). Bacterial endophyte communities of three agricultural important grass species differ in their response towards management regimes. *Scientific Reports*, 7(January), 1–13. <http://doi.org/10.1038/srep40914>
- Werner, G. D. A., & Kiers, E. T. (2015). Order of arrival structures arbuscular

- mycorrhizal colonization of plants. *New Phytologist*, 205(4), 1515–1524.
<http://doi.org/10.1111/nph.13092>
- Whipps, J. M., Hand, P., Pink, D., & Bending, G. D. (2008). Phyllosphere microbiology with special reference to diversity and plant genotype. *Journal of Applied Microbiology*, 105(6), 1744–1755. <http://doi.org/10.1111/j.1365-2672.2008.03906.x>
- White Jr, J. F., Torres, M. S., Johnson, H., Irizarry, I., & Tadych, M. (2014). A Functional View of Plant Microbiomes: Endosymbiotic systems that Enhance Plant Growth and Survival. *Advances in Endophytic Research*. <http://doi.org/10.1007/978-81-322-1575-2>
- Whittaker, R. H. (1972). Evolution and Measurement of Species Diversity. *Taxon*, 21(2/3), 213–251. <http://doi.org/10.2307/1218190>
- Wylie, K. M., Truty, R. M., Sharpton, T. J., Mihindukulasuriya, K. A., Zhou, Y., Gao, H., ... Pollard, K. S. (2012). Novel bacterial Taxa in the human microbiome. *PLoS ONE*, 7(6), 1–11. <http://doi.org/10.1371/journal.pone.0035294>
- Xu, L., Paterson, A. D., Turpin, W., & Xu, W. (2015). Assessment and selection of competing models for zero-inflated microbiome data. *PLoS ONE*, 10(7), 1–30. <http://doi.org/10.1371/journal.pone.0129606>
- Ye, Y., & Doak, T. G. (2011). A Parsimony Approach to Biological Pathway Reconstruction/Inference for Metagenomes. *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*, 5(8), 453–460. <http://doi.org/10.1002/9781118010518.ch52>
- Yeoh, Y. K., Dennis, P. G., Paungfoo-Lonhienne, C., Weber, L., Brackin, R., Ragan, M. A., ... Hugenholtz, P. (2017). Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. *Nature Communications*, 8(1). <http://doi.org/10.1038/s41467-017-00262-8>
- Zaneveld, J. R. R., & Thurber, R. L. V. (2014). Hidden state prediction: A modification of classic ancestral state reconstruction algorithms helps unravel complex symbioses. *Frontiers in Microbiology*, 5(AUG), 1–8. <http://doi.org/10.3389/fmicb.2014.00431>
- Zawadzka, M., Trzciński, P., Nowak, K., & Orlikowska, T. (2013). The impact of three bacteria isolated from contaminated plant cultures on in vitro multiplication and

rooting of microshoots of four ornamental plants. *Journal of Horticultural Research*, 21(2), 41–51. <http://doi.org/10.2478/johr-2013-0020>

Paquetes de R

Callahan B.J., McMurdie P.J., Rosen M.J., Han A.W., Johnson A.J.A., Holmes S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581-583. doi: 10.1038/nmeth.3869 (URL: <http://doi.org/10.1038/nmeth.3869>).

Schliep K.P. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics*, 27(4) 592-593
McMurdie and Holmes (2013) phyloseq: [An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data](#). PLoS ONE. 8(4):e61217

Louca S, Doebeli M (2017). “Efficient comparative phylogenetics on large trees.” *Bioinformatics* . doi: 10.1093/bioinformatics/btx701 (URL:<http://doi.org/10.1093/bioinformatics/btx701>).

Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlenn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2018). vegan: Community Ecology Package. R package version 2.5-3. <https://CRAN.R-project.org/package=vegan>

Cibergrafía

Brown University. (2018, Abril 13). Microbiome/Metagenome Analysis Workshop: DADA2 [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=wV5_z7rR6yw&t=2429s

ProMusa. Morfología de la planta del banano. Recuperado de <http://www.promusa.org/Morfolog%C3%ADa+de+la+planta+del+banano>

FAO Trade and markets (2018). Banana facts and figures. Recuperado de <http://www.fao.org/economic/est/est-commodities/bananas/bananafacts/en/#.XETiG89KgWo>

Redacción Caribe - El Tiempo. (2018, Agosto 30). El banano sigue abriéndose a nuevas áreas comerciales. Recuperado de <https://www.eltiempo.com/colombia/otras-ciudades/el-banano-sigue-abriendose-a-nuevas-areas-comerciales-261202>

Cadena de banano Indicadores e Instrumentos 2018. Recuperado de <https://sioc.minagricultura.gov.co/Banano/Documentos/002%20-%20Cifras%20Sectoriales/002%20-%20Cifras%20Sectoriales%20-%202018%20Marzo%20Banano.pdf>

Generalidades del cultivo de banano en Colombia. abril 18 de 2011. presentado a: dirección de cadenas productivas- madr. Recuperado de <https://sioc.minagricultura.gov.co/Banano/Documentos/005%20-%20Documentos%20T%C3%A9cnicos/005%20-%20D.T%20-%20Generalidades%20del%20Cultivo%20de%20Banano.doc>

Caracterización del subsector bananero en Colombia. SENA. Recuperado de <https://repositorio.sena.edu.co/bitstream/11404/2151/1/3004.pdf>

Cadena de banano Indicadores e instrumentos. 2018. Recuperado de <https://sioc.minagricultura.gov.co/Banano/Documentos/002%20-%20Cifras%20Sectoriales/002%20-%20Cifras%20Sectoriales%20-%202018%20Marzo%20Banano.pdf>

Apéndice

Apéndice A

Script de R del procesamiento de secuencias con DADA2

```
## Se utilizaron como referencia para construir este pipeline:
## DADA2 Pipeline Tutorial (1.8) (https://benjjneb.github.io/dada2/tutorial.html)
## Microbiome/Metagenome Analysis Workshop: DADA2
(https://www.youtube.com/watch?v=wV5\_z7rR6yw&t=1155s)

## Input files: fastq files obtenidos con Illumina MiSeq, paired-end reads
demultiplexados por muestra y sin barcodes o adaptadores

# Set working direcory
getwd()
setwd("/Users/camilaisaza/Library/Mobile
Documents/com~apple~CloudDocs/Computacional/DADA2/DADA2+SILVA132/")

# 1
# Load Packages
library(dada2)          # Sequence processing
library(DECIPHER)      # Sequence alignment
library(ggplot2)
library(vegan)

# 2
# Specify path where FASTQ files are located
path <- "/Users/camilaisaza/Library/Mobile
Documents/com~apple~CloudDocs/Computacional/DADA2/raw_data"
list.files(path)
```

```
# 3
# Sort files to ensure forward/reverse reads are in the same order
fnFs <- sort(list.files(path, pattern="_1.fastq.gz"))
fnRs <- sort(list.files(path, pattern="_2.fastq.gz"))

# 4
# Extract sample names, assuming filenames have format: SAMPLENAME_XXX.fastq
sample.names <- sapply(strsplit(fnFs, "_"), `[`, 1)

# 5
# Specify the full path to the fnFs and fnRs
fnFs <- file.path(path, fnFs)
fnRs <- file.path(path, fnRs)

# 6
# Plot quality scores and determine trimming cutoffs
plotQualityProfile(fnFs[9:10])
plotQualityProfile(fnRs[9:10])
# Our ideal score is 25
# To be safe, we'll cut the forward reads at around 290 bp
# To be safe, we'll cut the reverse reads at around 205 bp

# 7
# Create a new file path to store filtered and trimmed reads
filt_path <- file.path(path, "filtered-132")

# 8
# Rename filtered files
filtFs <- file.path(filt_path, paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(filt_path, paste0(sample.names, "_R_filt.fastq.gz"))
```

```
# 9
# Quality filtering and trimming
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs,
                    trimLeft=c(17,21), truncLen=c(290, 205),
                    maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                    compress=TRUE, verbose=TRUE)

out

# 10
# Error rate estimation
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)

# 11
# Plot error rates
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

# 12
# Dereplicate reads
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

# 13
# Name the derep-class objects by the sample names
names(derepFs) <- sample.names
names(derepRs) <- sample.names

# 14
# Sample inference
dadaFs <- dada(derepFs, err=errF, multithread=TRUE, pool = TRUE)
```

```
dadaRs <- dada(derepRs, err=errR, multithread=TRUE, pool = TRUE)

# 15
# Merge paired end reads
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)

# 16
# Tabulate denoised and merged reads
seqtab <- makeSequenceTable(mergers)
dim(seqtab)

# 17
# Inspect distribution of sequence lengths
table(nchar(getSequences(seqtab)))

# 18
# Chimera checking and removal (de novo chimera sequence detection)
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus",
multithread=TRUE, verbose=TRUE)
dim(seqtab.nochim)

# 19
# Calculate the proportion of non-chimeric ASVs
sum(seqtab.nochim)/sum(seqtab)

# 20
# Assign taxonomy from phylum to genus using SILVA database version 132
taxa <- assignTaxonomy(seqtab.nochim,
    "~/Library/Mobile
Documents/com~apple~CloudDocs/Computacional/DADA2/SILVAdatabase/Silva_nr_v
ersión 132_train_set.fa.gz",
```

```
        multithread=TRUE, verbose = TRUE)
unname(head(taxa))

# 21
# Assign species
taxa <- addSpecies(taxa, "~/Library/Mobile
Documents/com~apple~CloudDocs/Computacional/DADA2/SILVAdatabase/Silva_speci
es_assignment_versión 132.fa.gz")

# 22
# Replacing the sequences with headers (sq1, sq2...)
asv_seqs <- colnames(seqtab.nochim)
asv_headers <- vector(dim(seqtab.nochim)[2], mode="character")

for (i in 1:dim(seqtab.nochim)[2]) {
  asv_headers[i] <- paste(">sq", i, sep="")
}

# 23
# Export the fasta file
asv_fasta <- c(rbind(asv_headers, asv_seqs))
write(asv_fasta, "ASVs_vitrobanano.fa")

# 24
# Export the sequence table
ASVtab <- t(seqtab.nochim) # Transpose the table
row.names(ASVtab) <- sub(">", "", asv_headers)
ASVtab <- cbind('#OTUID' = rownames(ASVtab), ASVtab) # Add '#OTUID' to the
header (required by biom)
```

```
write.table(ASVtab, "/Users/camilaisaza/Library/Mobile
Documents/com~apple~CloudDocs/Computacional/DADA2/DADA2+SILVA132/seqs_t
able.txt", sep='\t', row.names=FALSE, quote=FALSE)
```

```
# 25
```

```
# Export taxonomy table
```

```
asv_tax <- taxa
```

```
row.names(asv_tax) <- sub(">", "", asv_headers)
```

```
write.table(asv_tax, "taxonomy.txt", sep="\t", quote=F)
```

```
# Save taxonomy with the correct format
```

```
taxonomia <- read.table("/Users/camilaisaza/Library/Mobile
```

```
Documents/com~apple~CloudDocs/Computacional/DADA2/DADA2+SILVA132/taxon
omy.txt", header = TRUE, sep = "\t")
```

```
# Extraer columnas
```

```
kingdomc <- taxonomia$Kingdom
```

```
phylumc <- taxonomia$Phylum
```

```
classc <- taxonomia$Class
```

```
orderc <- taxonomia$Order
```

```
familyc <- taxonomia$Family
```

```
genusc <- taxonomia$Genus
```

```
speciesc <- taxonomia$Species
```

```
i <- 1
```

```
while (i <= length(kingdomc)) {
```

```
  message("k__",kingdomc[i],";" " " "p__",phylumc[i],";" " " "c__",classc[i],";" " " "
    "o__",orderc[i],";" " " "f__",familyc[i],";" " " "g__",genusc[i],";" " " "
    "s__",speciesc[i])
```

```
  i = i+1
```

```
}
```

```
# 26
# Track reads through the pipeline
getN <- function(x) sum(getUniques(x))
track <- cbind(out, sapply(dadaFs, getN), sapply(dadaRs, getN), sapply(mergers, getN),
rowSums(seqtab.nochim))
colnames(track) <- c("Input", "Filtered", "DenoisedF", "DenoisedR", "Merged",
"Nonchim")
rownames(track) <- sample.names
track
```

Apéndice B

Script de R para la construcción de un árbol filogenético con las secuencias de las ASVs bacterianas

```
## Se utilizaron como referencia para construir este pipeline:
## Microbiome/Metagenome Analysis Workshop: DADA2
(https://www.youtube.com/watch?v=wV5\_z7rR6yw&t=1155s)

# Set working direcory
getwd()
setwd("/Users/camilaisaza/Library/Mobile
Documents/com~apple~CloudDocs/Computacional/DADA2/DADA2+SILVA132/")

# 1
# Load packages
library(Biostrings)
library(DECIPHER)
library(phangorn)
library(phyloseq)

# 2
# Import sequences
dna <- readDNAStringSet("/Users/camilaisaza/Library/Mobile
Documents/com~apple~CloudDocs/Computacional/DADA2/DADA2+SILVA132/asv_vi
trobanano_filt.fasta")

# 3
# Align sequences (MSA) using DECIPHER
alignment <- AlignSeqs(DNAStringSet(dna), anchor=NA)
```

```
# 4
# Change sequence alignment output into a phyDat structure
phang.align <- phyDat(as(alignment, "matrix"), type="DNA")

# 5
# Create distance matrix
dm <- dist.ml(phang.align)

# 6
# Perform neighbor joining
treeNJ <- NJ(dm)

# 7
# Internal maximum likelihood
fit = pml(treeNJ, data=phang.align)
fitGTR <- update(fit, k=4, inv=0.2)
fitGTR <- optim.pml(fitGTR, model="GTR", optInv=TRUE, optGamma=TRUE,
  rearrangement = "stochastic", control=pml.control(trace=0))

# 8
# Load asv_table_5_vitrobanano.biom and mapping_2_vitrobanano_R
biomfile5asv <- import_biom("asv_table_5_vitrobanano.biom")
head(otu_table(biomfile5asv))
head(tax_table(biomfile5asv))

# 9
# Import data into phyloseq
ps <- phyloseq(otu_table(biomfile5asv), tax_table(biomfile5asv),
  phy_tree(fitGTR$tree))
```

```
map <- import_qiime_sample_data("/Users/camilaisaza/Library/Mobile
Documents/com~apple~CloudDocs/Computacional/DADA2/DADA2+SILVA132/mappi
ng_2_vitrobanano.txt")

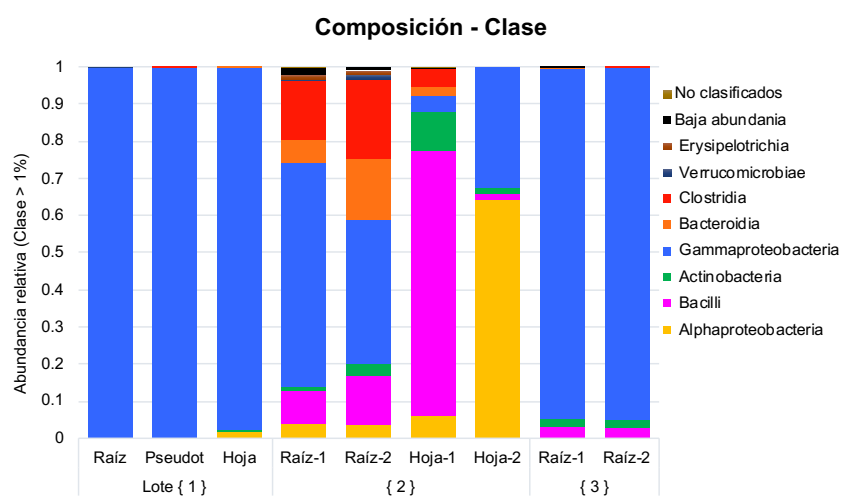
# 10
# Merge ps object with mapping file
ps <- merge_phyloseq(ps, map)
ps

# 11
# Root phylogenetic tree
set.seed(711)
phy_tree(ps) <- root(phy_tree(ps), sample(taxa_names(ps), 1), resolve.root = TRUE)
is.rooted(phy_tree(ps))

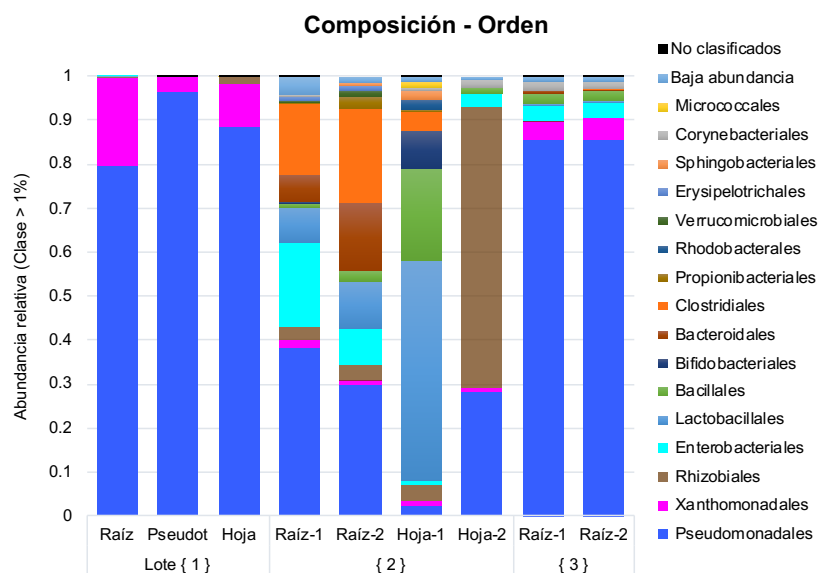
# 12
# Save tree as a file
write.tree(phy_tree(ps), file = "/Users/camilaisaza/Library/Mobile
Documents/com~apple~CloudDocs/Computacional/DADA2/DADA2+SILVA132/tree_fi
ltered_and_rooted.tre")
```

Apéndice C

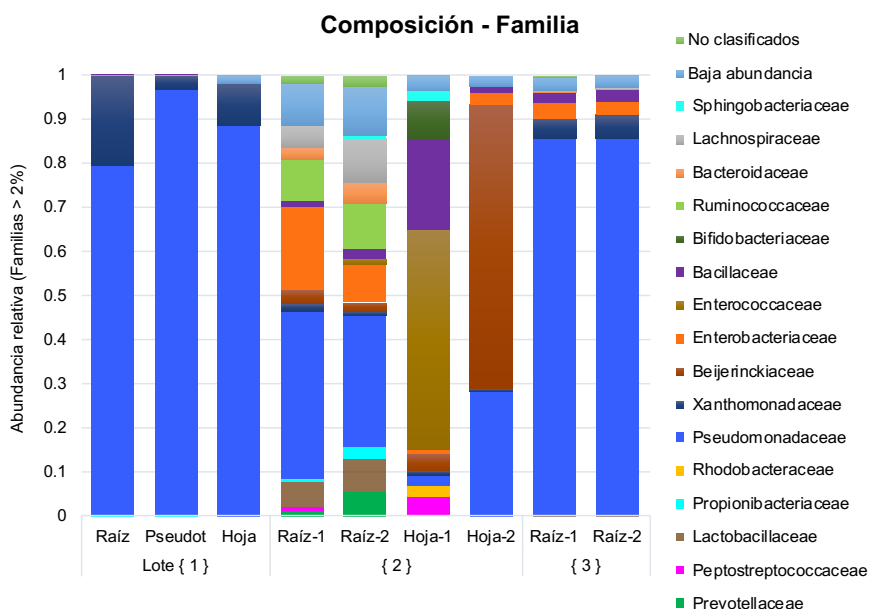
Abundancia relativa de las clases, ordenes y familias por muestra.



Abundancia relativa de las clases más predominantes en las 9 muestras. Dentro de las clases de baja abundancia se encuentran: Melainabacteria, Thermoleophilia, Coriobacteriia, Phycisphaerae, Mollicutes, Fusobacteriia, Deltaproteobacteria y Saccharimonadia.



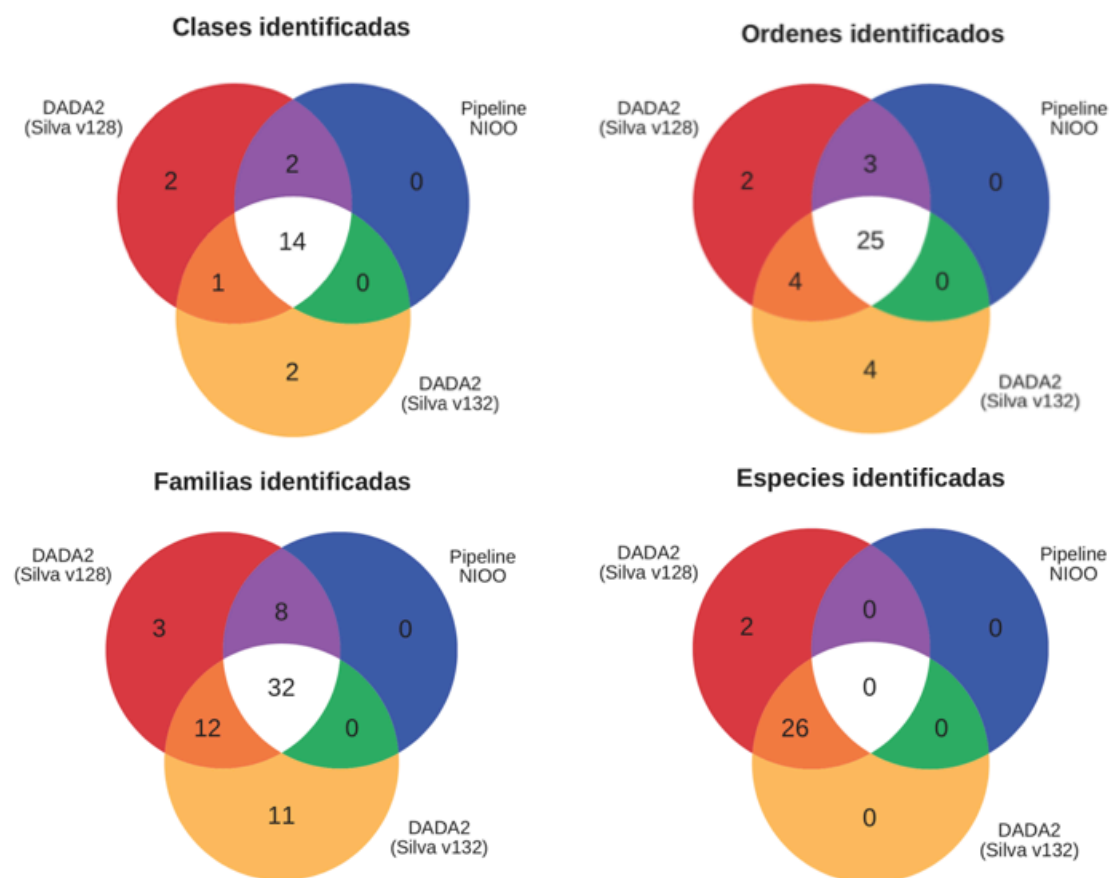
Abundancia relativa de los ordenes más predominantes en las 9 muestras. Dentro de los ordenes de baja abundancia se encuentran: Obscuribacterales, Solirubrobacterales, Coriobacteriales, Tepidisphaerales, Anaeroplasmatales, Actinomycetales, Pasteurellales, Fusobacteriales, Mollicutes_RF39, Flavobacteriales, Caulobacterales, Oligoflexales, Saccharimonadales, Pedosphaerales, Micromonosporales y Betaproteobacteriales.



Abundancia relativa de las familias más predominantes en las 9 muestras. Dentro de los ordenes de baja abundancia se encuentran: Xanthobacteraceae, Devosiaceae, Muribaculaceae, Akkermansiaceae, Family_XIII, Erysipelotrichaceae, Streptococcaceae, Tannerellaceae, Burkholderiaceae, Mycobacteriaceae, Clostridiaceae_1, Corynebacteriaceae, Microbacteriaceae, Solirubrobacteraceae, Atopobiaceae, Tepidisphaeraceae, Rikenellaceae, Anaeroplasmataceae, Actinomycetaceae, Pasteurellaceae, Christensenellaceae, Moraxellaceae, Nocardiodaceae, Fusobacteriaceae, Flavobacteriaceae, Caulobacteraceae, Micrococcaceae, 0319-6G20, Saccharimonadaceae, Pedosphaeraceae, Micromonosporaceae, Weeksellaceae, Rhodocyclaceae, Nocardiaceae, Thermoactinomycetaceae, Neisseriaceae, Staphylococcaceae, Paenibacillaceae y Leptotrichiaceae.

Apéndice D

Diagramas de Venn que muestran la intersección de grupos taxonómicos identificados por cada pipeline (pipeline del NIOO y pipeline DADA2) y por cada versión de la base de datos Silva (DADA2 Silva versión 128 y DADA2 Silva versión 132)



Apéndice E

Taxones identificados a nivel de género con abundancia por muestra inferior al 2% en la microbiota bacteriana asociada a la endosfera de raíces y pseudotallos y a las hojas de plantas de banano en estado *in vitro*.

Géneros identificados con abundancia > 2%	
Bradyrhizobium	Anoxybacillus
Devosia	Nocardioides
Escherichia/Shigella	Fusobacterium
Akkermansia	Capnocytophaga
Erysipelatoclostridium	Caulobacter
Butyrivibrio	Rothia
Prevotella_9	Burkholderia-Caballeronia-Paraburkholderia
Streptococcus	Lachnospiraceae_NK3A20_group
Parabacteroides	Lachnoclostridium
Alloprevotella	Sutterella
Ralstonia	Corynebacterium_1
Tyzerella_4	Alcaligenes
Coprococcus_3	Asanoa
Mycobacterium	Marvinbryantia
Ruminiclostridium_1	Cloacibacterium
Pedobacter	Anaerofilum
Clostridium_sensu_stricto_1	Thauera
Ruminococcaceae_UCG-014	Lactonifactor
Corynebacterium	CAG-873
Oceanobacillus	Lawsonella
Microbacterium	Chryseobacterium
Anaerostipes	Gordonia
Lachnospiraceae_UCG-004	Planifilum
SN8	Blautia
Pseudoclavibacter	Enhydrobacter
Solirubrobacter	Leifsonia
Serratia	Neisseria
Olsenella	Coprococcus_1
Rikenellaceae_RC9_gut_group	Lautropia
Anaeroplasma	Staphylococcus
Actinomyces	Acidipropionibacterium
Ruminiclostridium_9	Propioniciclava
Mucinivorans	Asaccharospora
Erysipelotrichaceae_UCG-003	Breznakia
Subdoligranulum	Pelomonas
Terrisporobacter	Paenibacillus
Gardnerella	Lachnospiraceae_NK4A136_group
Ruminococcaceae_UCG-010	Leptotrichia
Mannheimia	Azorhizophilus
Anaerotruncus	Oligotropha
Christensenellaceae_R-7_group	Falsibacillus
Acinetobacter	Atlantibacter

Apéndice F

Pipeline DADA2 seguido para el procesamiento de las secuencias correspondientes a la región V3-V4 del gen ARNr 16S de las bacterias asociadas a las plantas de banano cv.

Williamns en estado *in vitro*

