

Capítulo 14

Estudio de metagenomas en plantas

Abraham Loera Muro
María Goretty Caamal Chan
Aarón Barraza

*Programa de Agricultura en Zonas Áridas
CONACYT-Centro de Investigaciones Biológicas del Noroeste, S.C.*

Resumen

Las plantas, como todos los seres vivos de este planeta, se encuentran en íntima relación con una gran cantidad de microorganismos que habitan en ella y alrededor de ella. Estas comunidades microbianas poseen una función primordial sobre la planta. Estas funciones pueden ser desde promover su crecimiento, ayudar en la absorción de nutrientes y en la protección contra patógenos. Los estados de salud de la planta dependen en gran medida de la cantidad de interacciones que establezca con los diferentes microorganismos que forman el microbioma. Sin embargo, debido a su gran complejidad, estas comunidades y las interacciones que ocurren dentro de ellas y con la planta apenas están siendo comprendidas en estas últimas décadas. Por lo que, además de la importancia ecológica que significa el

estudio del microbioma de las plantas, este también radica en que puede ser aprovechado y explotado para desarrollar nuevos enfoques para la promoción del crecimiento de cultivos y el desarrollo de nuevos mecanismos de biocontrol y biofertilizantes que permitan un mejor aprovechamiento de una manera amigable con el ambiente. Para su estudio, actualmente existen, además de las tecnologías tradicionales, las llamadas tecnologías de secuenciación de nueva generación, con las que se pueden generar una gran cantidad de datos biológicos que permiten una mejor comprensión del microbioma y su relación con la planta huésped.

Introducción

El estudio de microorganismos asociados a plantas tiene una larga historia. Los microorganismos asociados a las plantas deben ser considerados como una pieza clave para aumentar la sanidad vegetal, la productividad, la manutención de las comunidades vegetales y el funcionamiento del ecosistema. El microbioma de las plantas (o fitobioma, al referirnos específicamente al microbioma de plantas) posee un impacto directo sobre las características funcionales de la planta, como, por ejemplo: la longevidad, el crecimiento, la absorción de nutrientes, la protección contra patógenos, entre otros. Todo lo anterior mediante la modificación de las rutas bioquímicas gracias a las interacciones que ocurren entre la planta con su microbioma, y entre las especies que forman al mismo.

El crecimiento y el rendimiento de las plantas en entornos naturales dependen de una gran cantidad de interacciones con diferentes microorganismos. Un ejemplo es la comunidad microbiana asociada con las raíces, mejor conocida como el rizobioma, la cual se postula ensambla de la siguiente manera: primero, la rizosfera (que es la zona localizada entre 1-3 mm alrededor de las raíces y es afectada por los exudados de la misma) es colonizado por un subconjunto de microorganismos de la comunidad del suelo; en segundo lugar, el rizoplano (superficie de la raíz donde se incluyen todos los microorganismos adherida a ella) y el espacio endofítico (compuesto por microorganismos que habitan dentro del tejido de la planta) son colonizados por un subconjunto de la comunidad de la rizosfera (Fig.14.1). De lo anterior, ha surgido un término que ayuda a englobar a un conjunto de microorganismos

que se encuentran recurrentemente asociados a algún compartimiento de las plantas, denominado “core microbiome”, el cual se puede traducir como el microbioma que es localizado en una especie en particular. También podemos encontrar a las comunidades microbianas que habitan la filosfera de las plantas (superficie de la parte aérea de las mismas), que también posee funciones de gran importancia para su huésped.

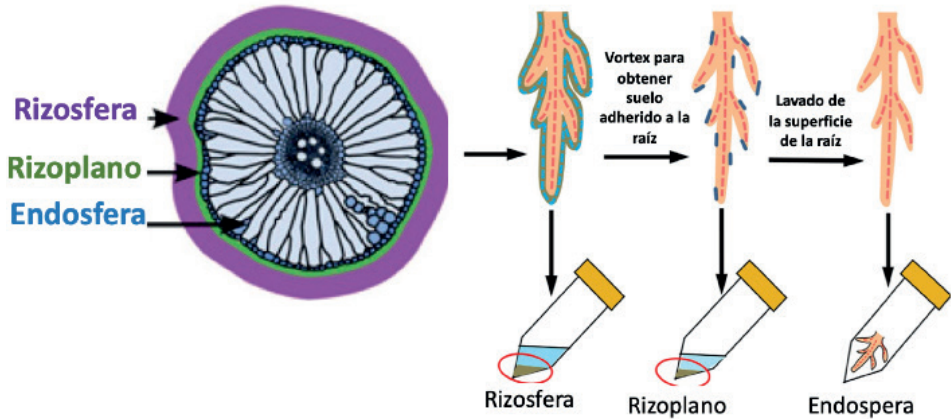


Figura 14.1 Protocolo para colectar muestras de rizosfera, rizoplano y endosfera. Tomada y modificada de Edwards *et al.* (2015) y Luo *et al.* (2017).

Estas comunidades del fitobioma pueden ser estudiadas y divididas según el compartimiento de la planta que habitan: a) los asociados a su raíz se les denomina rizobioma; b) las comunidades asociadas a la filosfera forman el filobioma; y c) las asociadas a sus componentes interiores o endosfera (comunidades endofíticas).

Al igual que el microbioma humano, el fitobioma constituye un aspecto biológico menos apreciado (fisiología, genoma, metaboloma, etc.) de las plantas. Las plantas y su microbioma asociado se afectan unos a otros en variadas y sutiles formas. Por ende, una planta es considerada como un metaorganismo, teniendo una persistente y regulada relación con su fitobioma. La composición del fitobioma está regulado por numerosos factores bióticos y abióticos, donde podemos encontrar una compleja matriz de interacciones microorganismo-planta y microorganismo-microorganismo.

Finalmente, la importancia del estudio del fitobioma, además de la ecológica, es debido a que puede ser aprovechado y explotado de varias maneras: a) desarrollar nuevos enfoques para la promoción del crecimiento de cultivos, b) optimizar procesos de fermentación y formulación relacionados, y c) desarrollar nuevos mecanismos de biocontrol y biofertilizantes que permitan un mejor aprovechamiento de una manera amigable con el medio.

Funciones de las comunidades microbianas asociadas a las plantas

Los microorganismos que conforman al microbioma de las plantas poseen varias funciones. Estos microorganismos son necesarios para promover el crecimiento vegetal. Los microorganismos más estudiados y caracterizados son los que conforman el rizobioma. Estos microorganismos que forman parte del microbioma, pueden estar en varios compartimientos de la planta y estimular el crecimiento de las mismas gracias a varios mecanismos: 1) solubilización de nutrientes (particularmente fósforo), 2) producción de sideróforos que ayudan a asimilar metales, 3) fijación de nitrógeno, 4) producción de fitohormonas, 5) producción de compuestos orgánicos volátiles, 6) inducción de resistencia sistémica (ya sea a través de la resistencia sistémica inducida (ISR), o de la resistencia sistémica adquirida (SAR), y 7) supresión de la enfermedad a través de la producción de compuestos o por competencia directa.

Todos los organismos requerimos nitrógeno (N) para sintetizar biomoléculas tales como proteínas y ácidos nucleicos. Sin embargo, la fuente principal de N en la naturaleza, el nitrógeno atmosférico (N_2), no se encuentra accesible para la mayoría de los organismos vivos, incluidos los eucariotas. La fijación biológica de nitrógeno, proceso por el cual el N_2 se reduce a amoníaco (NH_3) lo llevan a cabo algunas especies de microorganismos, particularmente bacterias y arqueas. Este fenómeno se realiza gracias a la nitrogenasa (genes *nif*, fijación de N_2), una enzima altamente conservada que comprende dos metaloproteínas, la FeMo-proteína y la Fe-proteína.

Por otra parte, el hierro (Fe) es un micronutriente esencial para las plantas y microorganismos, ya que está involucrado en varios procesos biológicos importantes, tales como fotosíntesis, respiración, biosíntesis de clorofila y en la fijación biológica del nitrógeno. En condiciones aeróbicas, la solubilidad del Fe es baja, lo que refleja el predominio de Fe_3+ , limitando así el suministro

de Fe para los seres vivos. Sin embargo, ciertos microorganismos han desarrollado ciertas estrategias para la absorción de Fe. Por ejemplo, las bacterias pueden superar la limitación nutricional de Fe al producir agentes quelantes llamados sideróforos. Los sideróforos son moléculas de bajo peso molecular (<1000 Da) con alta especificidad y afinidad por quelar o unir Fe_3^+ , seguido del transporte y deposición de Fe dentro de las células bacterianas.

Otro mecanismo por el cual las bacterias pertenecientes al microbioma de las plantas promueven beneficios a las mismas, es a través de la solubilización de compuestos como el fósforo. A pesar de que el fósforo se encuentra en grandes cantidades en la mayoría de los suelos, la mayor parte de este es insoluble y por lo tanto no permanece disponible para ser aprovechado por las plantas. El fósforo insoluble está presente como material inorgánico en compuestos como la apatita o como una de varias formas orgánicas, como el fosfato de inositol, fosfomonoésteres y fosfotriésteres. La solubilización inorgánica del fósforo se produce como resultado de la acción de ácidos orgánicos de bajo peso molecular como el ácido glicólico y el ácido cítrico, y por la producción de ciertas enzimas llamadas fosfatasas, que son sintetizadas por diferentes bacterias del suelo.

Así mismo, las hormonas vegetales o fitohormonas juegan un papel crucial en el crecimiento y desarrollo de las plantas. Es bien conocido que los microorganismos del rizobioma también pueden producir o modular los niveles de ciertas fitohormonas en las plantas hospedadas. Por lo tanto, pueden afectar el equilibrio hormonal de la planta huésped y su respuesta al estrés al alterar los niveles de fitohormonas endógenas en éstas plantas. El ejemplo mejor conocido es la producción de la fitohormona auxina (ácido indol-3-acético/ácido indol-acético/IAA). Se estima que cerca del 80% de los microorganismos que se han aislado de la rizosfera de varios cultivos poseen la capacidad de producirlo o modular a esta fitohormona.

Finalmente, una de las formas más comunes por las que las bacterias que habitan en el microbioma de las plantas promueven el crecimiento de las mismas de una forma indirecta, es debido a que pueden prevenir la proliferación de patógenos mediante la síntesis de diferentes tipos de compuestos antimicrobianos, como antibióticos, entre otros, así como por la competencia directa por el espacio contra diferentes patógenos.

Interacciones entre las plantas y su microbioma

Las interacciones entre plantas y su microbioma ocurren a través de procesos simbióticos, endofíticos o asociativos con distintos grados de proximidad con las raíces y el suelo circundante, principalmente. Estas interacciones están influenciadas por factores ambientales, como la composición del suelo, la temperatura, la humedad, la irradiación y el viento. Para establecer estas interacciones entre plantas y microorganismos, algunas barreras preformadas de las plantas y las defensas del sistema inmunitario deben pasarse por alto mediante una efectiva comunicación química entre la planta y el microbioma (Figura 14.2). La percepción extracelular en las plantas está mediada típicamente por receptores de reconocimiento de patrones (Pattern Recognition Receptors o PRR) unidos a la membrana plasmática que detectan epítomos microbianos conservados, denominados patrones moleculares asociados a microbios (Microbe-Associated Molecular Patterns o MAMP). La activación de un PRR por un MAMP asociado induce respuestas de defensa de la planta, denominadas colectivamente inmunidad activada por MAMP (MAMP-Triggered Immunity o MTI), que limitan el crecimiento de los patógenos. La MTI incluye respuestas tales como la alcalinización extracelular, el aumento de la concentración de Ca_2^+ y la activación de la cascada de proteínas quinasas activadas por mitógeno (Mitogen Activated Protein Kinase o MAPK). La MTI funciona en el reconocimiento de muchos componentes del fitobioma, incluidas las bacterias, hongos, oomicetos, nematodos e insectos. Por otra parte, los patrones moleculares asociados a daños (Damage-Associated Molecular Patterns o DAMP) son, por definición, moléculas propias, pero que su presencia señala la actividad dañina en las plantas por otros organismos. Las señales DAMP pueden incluir oligogalacturonidas, que son compuestos formados por la degradación fúngica de las paredes celulares de las plantas durante la infección. La percepción de DAMP también se logra mediante los PRR. Para evitar estos mecanismos de reconocimiento, los microorganismos secretan moléculas llamadas efectores, que son moléculas expresadas que cambian la fisiología del huésped para beneficiar al invasor. La percepción de la planta de los efectores está frecuentemente mediado por proteínas LRR (Leucine-Rich Repeat) de unión a nucleótidos (Nucleotide-binding LRR proteins o NLR) que son proteínas principalmente intracelulares. Las proteínas NLR han evolucionado para detectar la presencia o actividad de proteínas efectoras, que conducen a la

activación de fuertes reacciones inmunes, un proceso conocido como inmunidad activada por efectores (Effector-Triggered Immunity o ETI).

Un aspecto muy interesante y poco entendido es, de hecho, que la mayoría de los microorganismos pertenecientes al microbioma con los que las plantas interactúan, no son patógenos. Sin embargo, muchos de ellos expresan moléculas que son potencialmente reconocidas por el sistema inmune de la planta. Por lo tanto, una importante pregunta sobre el microbioma de las plantas es, ¿cómo el sistema inmune de la planta distingue a los comensales de los patógenos y cómo los hace durante el ensamblaje del mismo? Una de las formas que se ha visto hasta el momento es, que al igual que los patógenos, varios microorganismos no patógenos poseen la capacidad de suprimir la MTI a través de tres posibles mecanismos: i) divergencia de los MAMP, ii) degradación/secuestro de los MAMP, y iii) modificación de los MAMP.

Por otra parte, para establecer una relación simbiótica con las plantas, algunas bacterias del fitobioma secretan o emiten moléculas beneficiosas para la planta. Estas moléculas, que regularmente se producen en la rizosfera, pueden desencadenar cambios o ajustes específicos en el transcriptoma de la planta. Si bien, las fitohormonas son reguladores de crecimiento y defensa producidos por las plantas, algunas bacterias, como las llamadas promotoras del crecimiento vegetal (Plant growth-promoting bacteria o PGPB) también pueden producir estos compuestos que incluyen auxinas, citoquininas, giberelinas, ácido absísico (ABA), ácido salicílico (SA) y ácido jasmónico (JA), entre otros. Así mismo, algunas de estas PGPB del microbioma también producen compuestos volátiles orgánicos (VOC, compuestos lipofílicos de bajo peso molecular) que están involucrados en el mantenimiento de la salud del suelo, la modulación del crecimiento de las plantas y la inducción de resistencia. Algunos ejemplos de éstos VOC conocidos producidos por PGPR son el 2-heptanol, 2-undecanona y el pentadecano.

Así mismo, las plantas sirven como fuentes ricas de nutrientes para las bacterias y otros microorganismos que viven dentro de sus tejidos y alrededor de ellos. A través del proceso de rizodeposición (dado por los exudados de la raíz), las plantas liberan compuestos orgánicos en el suelo. Estos compuestos pueden contribuir a la acumulación del carbono derivado de plantas a la biomasa microbiana circundante. Un ejemplo de estos exudados son los flavonoides, que son metabolitos secundarios ricos en carbono. La adaptación bacteriana a los exudados de las raíces de las plantas es claramente visible en

todos los niveles, desde el genoma hasta el metaboloma de los microorganismos que conforman el fitobioma. Una adaptación evidente es la observada en el enriquecimiento del metabolismo de los carbohidratos y los genes de transporte, junto con sus reguladores transcripcionales en genomas de bacterias asociadas a plantas de diferentes especies.

Finalmente, para que exista un correcto ensamblaje de las comunidades microbianas que forman el fitobioma de las plantas, además de la comunicación que debe existir entre los microorganismos que la habitan con su huésped, también deben existir interacciones microorganismo-microorganismo. Uno de los principales mecanismos a través de los cuales ocurren éstas interacciones es a través del llamado “quorum sensing” (QS). El proceso de QS involucra a la célula bacteriana que produce señales químicas conocidas como autoinductores (AI), que se secretan en el espacio extracelular. Hasta la fecha, tres tipos de autoinductores se han caracterizado en bacterias: AI-1, AI-2 y AI-3. El AI-2 (regularmente en Gram-negativas una acil-homoserin lactona o AHL) está involucrado en la comunicación entre diferentes especies bacterianas.

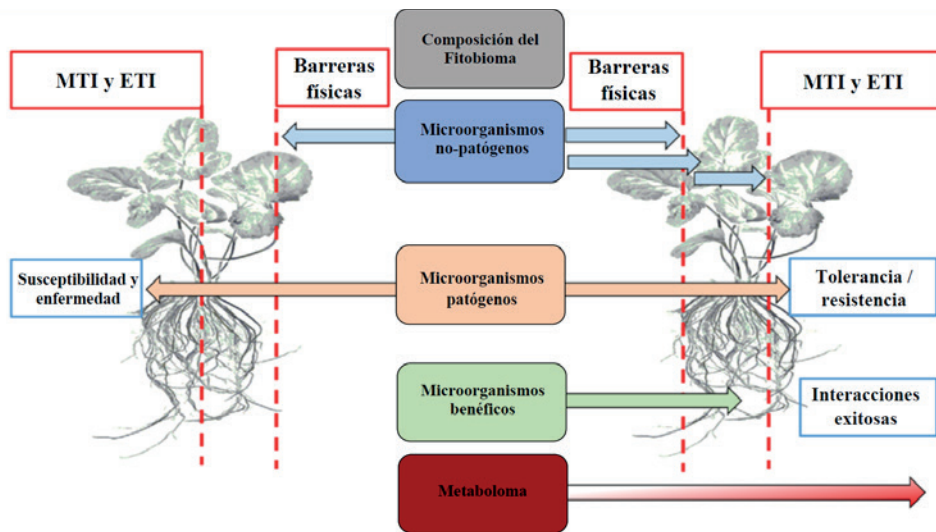


Figura 14.2 Interacciones entre plantas y su fitobioma (MTI: MAMP-Triggered Immunity; ETI: Effector-Triggered Immunity). Tomada y modificada de Mhlongo *et al.* (2018).

La metagenómica y la Secuenciación de Nueva Generación (NGS)

Las secuencias de los genes del ARNr y especialmente el ARNr 16S representan los objetivos actuales de estudio más importantes en la evolución y ecología bacteriana, incluida la determinación de las relaciones filogenéticas entre los taxones, la exploración de la diversidad bacteriana en el medio ambiente y la cuantificación de la abundancia relativa de taxones de varios rangos. El 16S ARNr es adecuado para esto por varias razones: a) el gen está distribuido universalmente, permitiendo el análisis de las relaciones filogenéticas entre taxones distantes; y b) como una parte funcionalmente indispensable del conjunto de genes centrales, se espera que el gen 16S ARNr solo se vea débilmente afectado por transferencia horizontal de genes, que respalda aún más su uso para estudios filogenéticos. A pesar de lo anterior, el 16S ARNr está sujeto a la variación, especialmente en ciertas regiones variables de los genomas. Esta presencia de regiones variables permite una diversificación suficiente para proporcionar una herramienta para la clasificación. Por otra parte, la presencia de regiones conservadas permite el diseño de cebadores para PCR adecuados o sondas de hibridación para varios taxones a diferentes niveles taxonómicos que van desde cepas individuales a phyla.

A pesar del amplio uso de 16S ARNr, hay varios aspectos a considerar durante la interpretación de los resultados derivados de 16S ARNr. Uno de los principales es el hecho de que sus números de copias por genoma varían de 1 hasta 15 o más copias. Los números de copias parecen ser específicos de taxones hasta cierto punto, pero la variación entre cepas de la misma especie también ha sido registrada. El número de copias de ARNr ha sido puesto en contexto con la estrategia de vida de las bacterias porque el número de copias de ARNr de algunos taxones está correlacionado con su capacidad de responder a condiciones de crecimiento desfavorables.

En 1977, Frederick Sanger desarrolló la tecnología de secuenciación de ADN que se basaba en el método de terminación de cadena (conocida como secuenciación de Sanger). Al mismo tiempo, Walter Gilbert desarrolló otra tecnología de secuenciación basada en la modificación química del ADN y la posterior escisión en bases específicas. Sin embargo, debido a su alta eficiencia y baja radioactividad, la secuenciación de Sanger fue adoptada como la tecnología principal. Después de años de mejora, Applied Biosystems introdujo la primera máquina de secuenciación automática (el AB370) en 1987, adop-

tando la electroforesis capilar que hizo que la secuenciación fuera más rápida y precisa. La Secuenciación de Nueva Generación (Next-Generation Sequencing o NGS) es diferente del método Sanger en aspectos tales como el análisis masivo paralelo de datos, el alto rendimiento y los reducidos costos. Aunque la NGS hace que obtener las secuencias de genomas sea más fácil, rápido y con menores errores, el análisis de los datos obtenidos y las explicaciones biológicas que se les den a éstos siguen siendo el cuello de botella en el estudio de los genomas. Muchas plataformas de secuenciación están disponibles hoy en día. Todas ellas usan fragmentos cortos, llamados lecturas o “reads”, para investigar secuencias del genoma. Las tecnologías de NGS analiza millones de fragmentos cortos de ADN durante una secuenciación. Las longitudes de lectura de esos fragmentos dependen del tipo de plataforma de NGS y pueden estar en el rango de 25–450 pares de bases. Si bien las lecturas son mucho más cortas que las creadas por la secuenciación tipo Sanger, NGS tiene un mayor rendimiento y crea conjuntos de datos con hasta 50 gigabases por ejecución. Esto exige algoritmos mejorados que sean capaces de procesar esas enormes cantidades de material de datos sin procesar.

En la metagenómica, la información de la secuencia genética que se captura representa casi la totalidad de las especies de un microbioma, donde se incluyen las que no pueden representarse mediante cultivos microbiológicos tradicionales. Actualmente, se utiliza con mayor frecuencia en las investigaciones de los microbiomas y las comunidades microbianas para el estudio de la abundancia relativa de taxones individuales, el análisis basado en la construcción de grupos de similitud (o unidades taxonómicas operativas, operational taxonomic units o OTU) de amplicones de PCR del gen 16S ARNr. Una alternativa al 16S ARNr, es la secuenciación de un metagenoma del microbioma de la planta (“shotgun metagenomics”). Esta técnica se basa en la secuenciación del ADN total de toda la comunidad del microbiana. El mayor desafío en la metagenómica tipo “shotgun metagenomics” es ensamblar las lecturas de secuenciación en genomas correctamente ensamblados con alta calidad donde se capturan todos los genes en un genoma y las lecturas (en este paso se les denomina contigs) ensambladas se asignen a los organismos correctos. Esto puede ser particularmente desafiante para microorganismos raros o de nueva descripción. Otros obstáculos incluyen la asignación taxonómica adecuada de los genomas ensamblados y la diferenciación entre cepas relacionadas en muestras que contienen un alto grado de heterogeneidad.

Análisis bioinformático

Plataformas de secuenciación de siguiente generación (NGS)

Actualmente las plataformas de secuenciación de siguiente generación o NGS (“Next-generation sequencing”) son las siguientes: IonPGM (Thermo Fisher Scientific, <https://www.thermofisher.com/mx/es/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence.html>), PacBio (Pacific Biosciences, <https://www.pacb.com/>), MinIon (Oxford Nanopore, <https://nanoporetech.com/>) e Illumina (<https://www.illumina.com/>). Estas plataformas tienen la capacidad de secuenciar productos de amplificación por PCR, ya sean únicos, de varios genes o genes marcadores de distintos organismos al mismo tiempo, así como, genomas y transcriptomas únicos, y varios genomas y transcriptomas al mismo tiempo. Dependiendo de las necesidades de trabajo, se puede elegir la plataforma de secuenciación NGS adecuada. Las plataformas de secuenciación PacBio y MinIon se caracterizan por generar lecturas largas (“long reads”) desde 10 Kb hasta 100 Kb, pero con una precisión por base de ~85%. La plataforma de secuenciación IonPGM generan fragmentos de tamaño heterogéneo de 100 pb hasta 400 pb con una precisión por base de ~98%. La plataforma de secuenciación Illumina ofrece distintos niveles de rendimiento (nivel de mesa de trabajo, desde 4 millones a 400 millones de “reads”; y nivel de producción escalada, desde 400 millones hasta 20 mil millones) y una longitud de lecturas uniforme (150 pb, 250 pb, 300 pb, dependiendo el equipo de secuenciación) y con una precisión por base >99%; además, es la única plataforma en que las secuencias generadas son pareadas (“paired-end”), es decir, se producen lecturas en sentido “forward” (a las que se les denomina R1) y lecturas en sentido “reverse” (R2) a partir de la misma secuencia. Por el hecho de producir secuencias pareadas las denominaciones de los productos de secuenciación son: 2 x 150, 2 x 250 y 2 x 300. Para los análisis de diversidad bacteriana mediante la secuenciación de las regiones hipervariables (V1-V9) V3 o V4 del gen 16S rDNA, las plataformas de secuenciación recomendadas son las IonPGM e Illumina (MiniSeq y MiSeq); aunque, por la homogeneidad de los productos de secuenciación (2 x 150 del MiniSeq o 2 x 300 pb del MiSeq), la precisión por base y la generación de secuencias para-

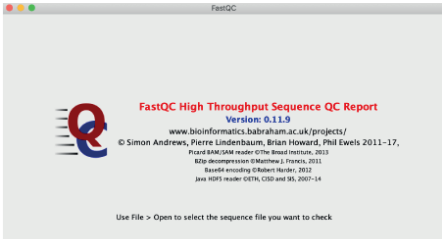
das (R1-R2). Estas características aumentan la precisión por base, en este tipo de estudios y por ende esta plataforma es la de uso más extendido.

Productos de la secuenciación NGS

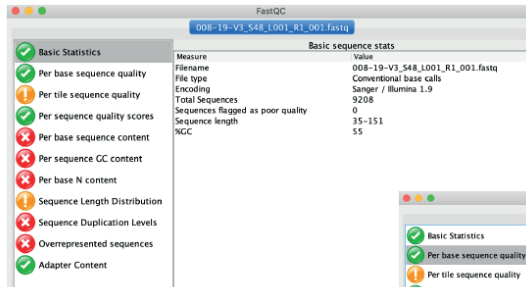
Los productos directos de la secuenciación NGS son las lecturas o “reads”. Las lecturas o “reads” generados poseen información importante para su procesamiento, análisis y manipulación. Esta información es: la secuencia en sí y la calidad por base o nucleótido secuenciado. La calidad por base indica que tan precisa es la información, es decir, que tan fiable es la secuencia en cada base o nucleótido secuenciado. Los rangos de calidad varían de acuerdo a la plataforma de secuenciación. Sin embargo, el rango de calidad más usado es el de las plataformas Sanger (secuenciación automatizada de primera generación, que sólo secuencia una lectura a la vez) e Illumina. El rango de calidad de ambas plataformas de secuenciación es de 0 a 40, lo cual equivale a la tasa de error o calidad por base y se le denomina valor de Phred o calidad (y se representa con la letra “Q”). Entre más bajo sea el valor Q de la base secuenciada, menor es el valor de calidad y entre mayor sea este valor, mayor será la calidad. Además, el valor Q corresponde a la siguiente escala de calidades: 90% de precisión por base $Q = 10$, 99% de precisión por base $Q = 20$, 99.9% de precisión por base $Q = 30$, y 99.99% de precisión por base $Q = 40$. El formato de los archivos en los cuales se encuentra la información de la secuencia de las lecturas y la calidad por base secuenciada es el formato .fastq. Los archivos .fastq contienen todas las secuencias de las lecturas correspondientes a una muestra, que puede ser desde unos cientos hasta cientos de millones, o incluso más, de lecturas o “reads”. La información correspondiente a la calidad por base se encuentra en este archivo y está codificada en formato hexadecimal (Fig. 14.3).

“Terminal” puedes ver los archivos de secuenciación, contar cuantos “reads” se encuentran en el archivo e incluso eliminar los adaptadores que se agregaron en el proceso de síntesis de las “librerías” para el proceso de secuenciación; en el caso de que las secuencias de los adaptadores no se hayan eliminado previamente porque se prepararon estas en nuestro laboratorio acorde a nuestras especificaciones. De haberse preparado las librerías en nuestro laboratorio es necesario tener a la mano la secuencia correspondiente a los adaptadores empleados en el caso de que fuera diseños por nuestro equipo de trabajo, o cuales de los presentes de los “kits” de secuenciación comerciales, se emplearon para quitarlos de la secuencia de interés. Con el uso de la aplicación “Terminal” se pueden identificar los adaptadores usando en conjunto los comandos “grep”, “sed” en conjunto con las secuencias específicas de los adaptadores o con expresiones regulares. Con la aplicación FastQC se puede acceder al archivo de secuenciación NGS usando una interface gráfica de usuario o GUI (“Graphic user interface”, es decir, se puede “abrir” un archivo de manera habitual como accedamos a los archivos con los programas de ofimática de Microsoft Office [Word, Excel]). Al acceder a los archivos de secuenciación NGS, estos son procesados por la aplicación y al término de este procesamiento se despliega toda la información correspondiente a la calidad por base de todas las secuencias o “reads” presentes en el archivo, presencia de adaptadores (si es que los hay), abundancia de ambigüedades en las secuencias (“N”, o nucleótidos no definidos en la secuencia) (Fig. 14.4).

Las secuencias de los archivos de secuenciación NGS se pueden procesar con distintas aplicaciones o programas para eliminar las pares de bases, o incluso secuencias completas, de baja calidad (Fig. 14.4, zona roja). Las aplicaciones o programas más utilizados para este fin son: Trimmomatic y BBDuk. Para ambas aplicaciones es necesario especificar el valor de la calidad mínima por base, para eliminar las pares de bases que no cumplan con este criterio. Este proceso también se le conoce como “Trimming” que es parte del proceso “QC” (“Quality Control”), ya que al eliminar las pares de bases o secuencias de baja calidad se está aplicando un criterio cuantitativo sobre las secuencias presentes en el archivo de secuenciación NGS y así los análisis realizados a estas secuencias serán más confiables (Figura 14.5).



1



2



3

Figura 14.4 Aplicación FastQC. 1) GUI de la aplicación, con la indicación para el uso o apertura de un archivo de secuenciación NGS para su procesamiento. 2) Despliegue de todas las características del archivo de secuenciación NGS procesado por la aplicación. 3) Gráfica de cajas (“boxplot”) de la distribución de las calidades por base de cada secuencia presente en el archivo de secuenciación NGS, con tres zonas de calidad: roja (baja calidad, no recomendable), amarilla (calidad moderada, justificar uso) y verde (calidad alta, recomendable sin justificación de uso).

Control de Calidad y “Trimming”

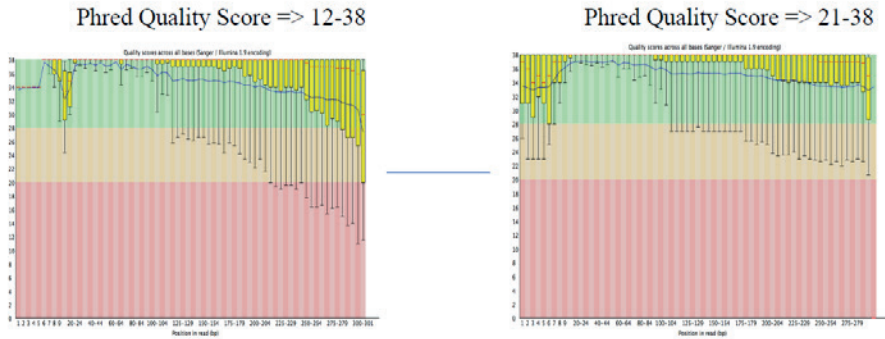


Figura 14.5 Proceso QC mediante el “trimming” de los pares de bases de las secuencias en los archivos de secuenciación NGS. Al inicio (izquierda) están presentes pares de bases con un rango de calidad de 12-38 (es decir que se encuentran en la zona roja a la zona verde) con valores medios de calidad de 28-38 (línea azul). Después del proceso de “trimming” (derecha) el rango de calidad aumentó de 21-38 con valores medios de calidad de 34-37 (línea azul).

Procesamiento de “reads” y generación de tablas de OTUs

Las secuencias procesadas ahora se pueden agrupar por pares, ya que como se generaron a partir de la plataforma de secuenciación Illumina, se cuentan con lecturas en sentido “forward” denominadas R1 y lecturas en sentido “reverse” denominadas R2. Al tener las secuencias en pares se procede al empalme o “merging” de los pares de secuencias R1 y R2 con la aplicación BBMerge. Este proceso elimina las secuencias quiméricas que se generaron en el proceso de creación de las librerías y en el mismo proceso de secuenciación; ya que, las secuencias quiméricas no tienen una secuencia complementaria a la cual complementarse y así generar un empalme de secuencias R1-R2. Previo al proceso de empalme o “merging” de las secuencias R1 y R2 se pueden eliminar las quimeras con el uso de la aplicación UCHIME. Las secuencias resultantes del proceso de empalme o “merging” se denominan ahora con la letra R y se puede proceder al proceso de agrupamiento, mapeo y generación de tablas de unidades taxonómicas operacionales (“Operational taxonomical units or OTU”) (Fig. 14.6).

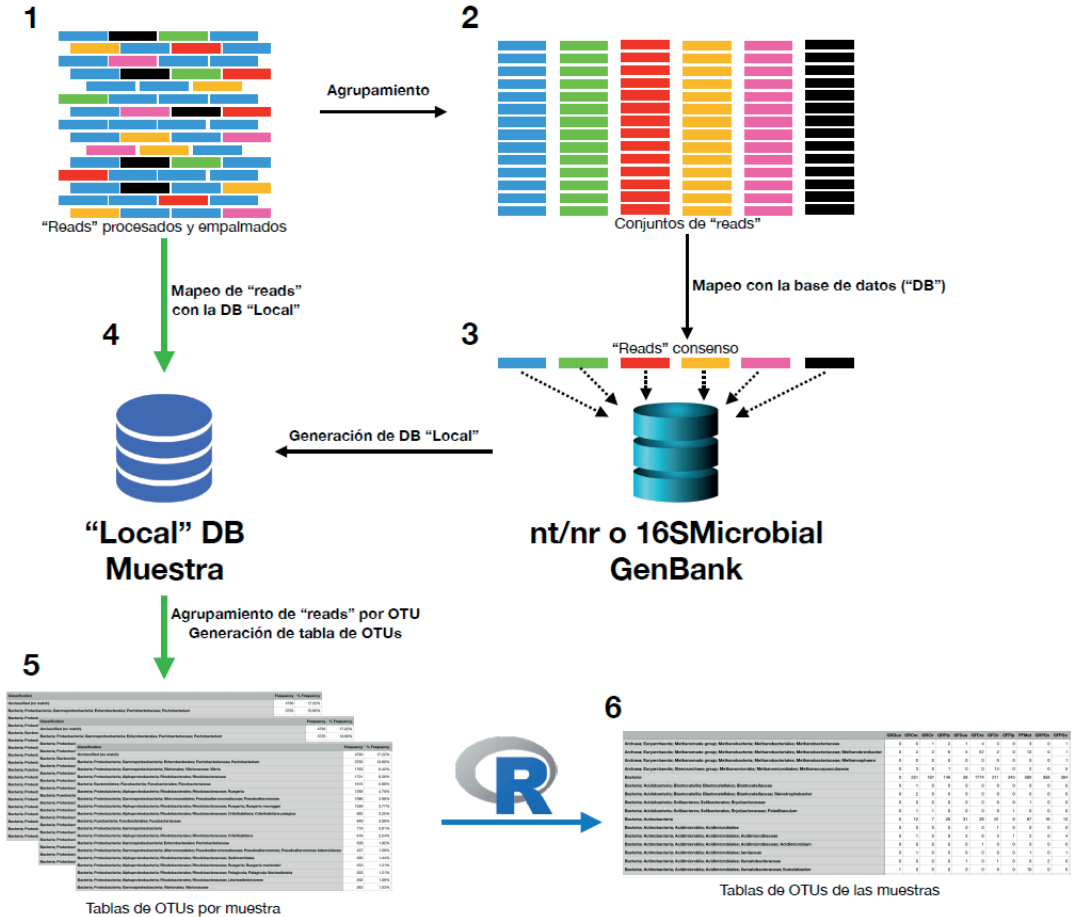


Figura 14.6 Proceso de generación de la tabla de OTUs. 1) Conjunto de "reads" procesados, pareados y empalmados. 2) "Reads" agrupados con base al criterio de identidad/similitud del 98%. 3) "Reads" consenso de cada agrupación que serán "mapeados" con la base de datos ("database" o "DB") del GenBank, para la generación de una DB "Local" o por muestra. 4) Mapeo del conjunto de "reads" procesados, pareados y empalmados de la muestra con la DB "Local" de la muestra. 5) Agrupamiento de los "reads" por cada OTU y generación de la tabla de frecuencias de cada OTU en la muestra. 6) Integración de las tablas de OTUs por muestra a una tabla de OTUs de las muestras.

El proceso de agrupamiento de los “reads” consiste en formar grupos basados en el criterio de “identidad/similitud” de las secuencias de los “reads”, es decir, agrupar las secuencias que se parecen a nivel de secuencia con al menos un porcentaje del 98% de “identidad/similitud” como requisito mínimo de agrupamiento de los “reads” (Fig. 14.6.1 y 14.6.2). Este proceso se puede realizar con las aplicaciones Velvet, SPAdes, CAP3, SOAPdenovo, así como cualquier ensamblador *de novo*. Los ensambladores *de novo*, en principio, se encargan de agrupar los “reads” presentes en un archivo de secuenciación NGS, tomando como criterio principal la “identidad/similitud” de las secuencias o “reads” y así poder formar agrupaciones de secuencias. Los ensambladores *de novo* se utilizan para la generación de “contigs”. Los “contigs” son el producto del empalme de secuencias contiguas que se extienden a longitudes de pares de bases mayores que las lecturas originales. La generación de “contigs” ayuda en la construcción de secuencias completas de ARNm’s, en el caso de la transcriptómica, sin necesidad de contar con un genoma de referencia y determinar que genes están transcripcionalmente activos; así como para la construcción completa de cromosomas y, por ende, del genoma completo de un organismo, o genómica, para la generación de un genoma de referencia.

Los grupos de “reads” generados en el proceso de agrupamiento con el uso del ensamblador *de novo* se designan con un identificador único por grupo (Fig. 14.6.2). Cada grupo identificado genera una secuencia consenso propia, la cual se utiliza para llevar a cabo el proceso de “mapeo” con una base de datos de secuencias de referencia (Fig. 14.6.3). Las bases de datos de las secuencias de referencia pueden ser de genomas completos o de genes marcadores, tales como el 16S para procariotes, 18S y región ITS (“Internal transcribed spacer”) para eucariotes. Los genes 16S ADNr y 18S ADNr ambos poseen 9 regiones hipervariables (V1-V9). Las regiones hipervariables tienen la característica de que son regiones que presentan alta variabilidad en su secuencia para poder diferenciar entre organismos de distintas especies, pero entre organismos de la misma especie, estas regiones se mantienen estables o conservadas. Para los estudios de diversidad procariótica, la o las regiones hipervariables que se recomienda(n) utilizar es la región V3 o V3-V4 (V3 junto con la región V4). El proceso de “mapeo” consiste en el alineamiento de cada secuencia consenso de cada grupo generado en el proceso de “agrupamiento” con una base de datos. El algoritmo o aplicación para llevar a cabo este proceso es “Megablast” y la base de datos puede ser la de nucleótidos no redundantes (“nt/nr”) o la

“16SMicrobial” del GenBank, así como la SSU de Silva y Greengenes, <https://www.arb-silva.de/>, <https://greengenes.secondgenome.com/>). Los resultados de los alineamientos obtenidos servirán para establecer una base de datos “local” por muestra, es decir, este proceso de mapeo se realiza por cada muestra (Fig. 14.6.3, 14.6.4). Una vez obtenida la base de datos “local” por muestra, se procede al proceso de “generación de tablas de unidades taxonómicas operacionales” o de “tablas de OTUs (Operational taxonomic units)” con el uso de todas las secuencias empalmadas y un criterio de asignación taxonómica a nivel de especie del 97% al 99% (Fig. 14.6.5). Este criterio de asignación indica que para que un “read” pueda ser asignado al nivel taxonómico de especie requiere un 97% o 99% de identidad con la secuencia correspondiente en la base de datos “local”, si el porcentaje de identidad es menor al 97% o 99% se asignará al nivel de genero; menor al 95% se asignará a nivel de familia y así sucesivamente hasta llegar al nivel de filo o hasta dominio. Este proceso se lleva a cabo con cada “read”, lo que genera una tabla de frecuencia del total de “reads” asignados a cada OTU en la muestra analizada.

Las tablas de OTUs generadas por muestra se integran en una tabla de OTUs de todas las muestras y de sus replicas basadas en las frecuencias de “reads” de cada OTU (Fig. 14.6.6). El siguiente paso consiste en la normalización de los datos. La normalización de los datos de cada muestra se realiza obteniendo la proporción o porcentaje de la frecuencia de “reads” en cada OTU con respecto al total de “reads” de cada muestra. Esto nos permite comparaciones más claras y evita los sesgos provocados por la cantidad de “reads” generados en cada replica de cada muestra en el proceo de secuenciación NGS; ya que, la obtención de “reads” varia con cada replica de cada muestra. Al tener las tablas de OTUs de las muestras por replicas, estas se pueden integrar sólo por muestra para comparar distintas muestras más fácil y claramente. Para los análisis de diversidad ecológica (alfa y beta, principalmente), es necesario tener las tablas de OTUs por muestra y de sus replicas ya normalizadas.

Análisis de diversidad ecológica microbiana

Los análisis de diversidad alfa son los que se llevan a cabo para conocer que tan homogénea o uniforme es la muestra de trabajo que se eligió para ser estudiada. La muestra de trabajo es cualquier zona de muestreo bien delimitada y con características bien definidas. La muestra de trabajo puede ser la hoja

de una planta, suelo asociado a raíces de plantas (rizoplano), suelo cerca de las raíces de una planta (rizosfera), suelo de un campo de cultivo, suelo de bosque, muestras fecales de pacientes sanos, muestras fecales de pacientes con alguna enfermedad diagnosticada, etc. Las replicas de cada muestra corresponden a tomar varias muestras de la muestra de trabajo, es decir, si son muestras de hojas de una planta tomar varias hojas de esa planta, varias muestras de suelo o varias muestras fecales del mismo paciente. Los índices alfa que se estiman comúnmente son los que nos indican que tan diversa, rica y homogénea es la muestra *per se*. Los índices alfa de diversidad corresponden directamente al número de OTUs observados o identificados en la muestra y al índice de Chao1. Este último toma en cuenta además de la cantidad de OTUs observados en la muestra los OTUs poco representados en la misma, para así darle un peso ecológico relevante a estos organismos en la comunidad microbiana que conforma la muestra. El índice de riqueza o índice de Shannon evalúa que la cantidad de OTUs sea tan variada como la proporción que representa cada OTU en la comunidad microbiana de la muestra analizada. El índice de homogeneidad o índice de Simpson evalúa que la comunidad microbiana este conformada principalmente por un dominio o nivel taxonómico en particular: Bacterias, Archaeas, Eukaryas, Fungi, Metazoa, etc. Los índices alfa se representan gráficamente mediante “boxplots” y se realiza una comparativa estadística para determinar diferencias netas o significativas aplicando las pruebas de Mann-Whitney o Kruskal-Wallis (Fig. 14.7.1). Además, los índices alfa se pueden estimar mediante curvas de muestreo, es decir, realizar una simulación de los OTUs presentes en cada muestra contra el total de “reads” presentes en la muestra; también se les denomina curvas de rarefacción y ayudan para determinar si la cantidad de “reads” presentes en la muestra está representando la totalidad, al menos a nivel teórico, de la diversidad microbiana en la muestra (Fig. 14.7.2). Este análisis de estimación de curvas de muestreo se puede aplicar a los índices de Chao1, Shannon y Simpson con los OTUs presentes en cada muestra (interpolación) e incluso extenderlos, mediante la simulación computacional, al doble de “reads” de la muestra o extrapolación (Fig. 14.7.3). Además, con estos análisis de interpolación/extrapolación se puede estimar de manera más precisa si la cantidad de “reads” en cada muestra está representando la totalidad teórica de la diversidad microbiana (Fig. 14.7.4). Para el manejo de datos, es decir, de las tablas de OTUs, la normalización de los datos, manejo de la integración de las tablas de

OTUs por réplicas y por muestra, cálculo de los índices alfa para diversidad, riqueza y homogeneidad de las muestras, se requiere del uso del lenguaje de programación R y del ambiente de desarrollo integrado o IDE (“Integrative development environment”) de RStudio (<https://www-r-project.org>, <https://rstudio.com>), así como de los paquetes de visualización de gráficos ggplot2 y de análisis ecológicos vegan y iNEXT.

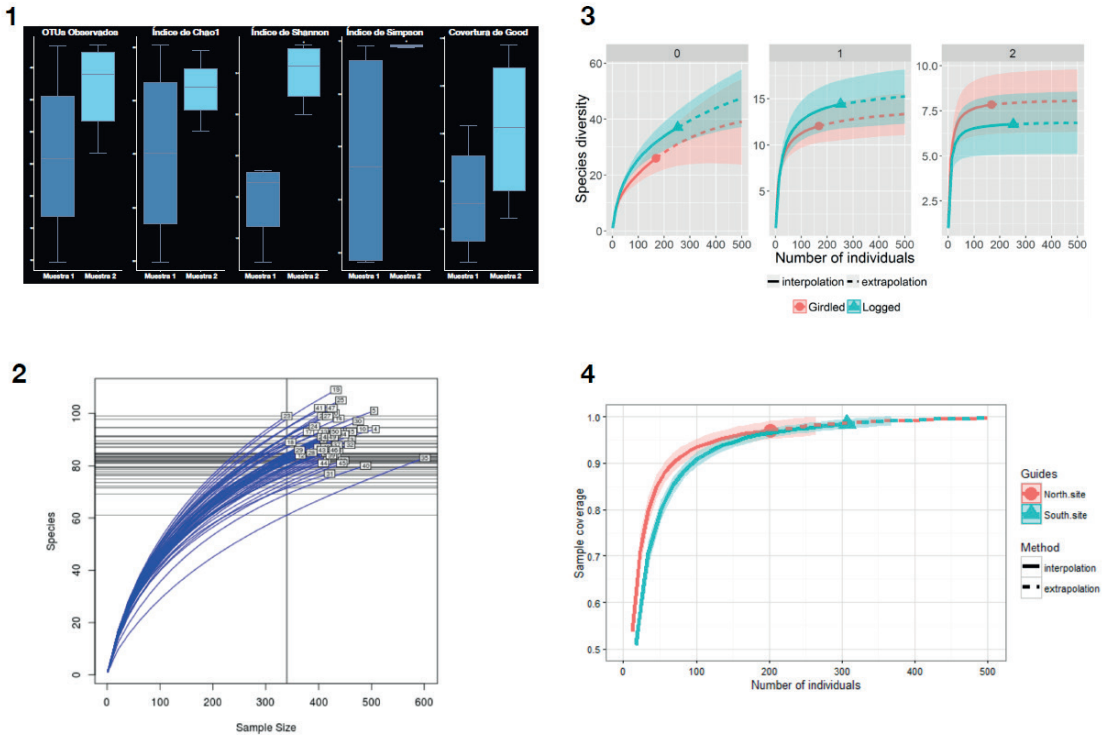


Figura 14.7 Representación de los índices de diversidad alfa. 1) Análisis de distribución de los índices alfa de diversidad (OTUs observados, índice de Chao1), riqueza (índice de Shannon) y homogeneidad (índice de Simpson) de cada muestra en “boxplots”. 2) Estimación de las curvas de rarefacción de cada muestra. 3) Estimación de las curvas de muestra mediante el proceso de inter- y extrapolación de la diversidad ((0) índice de Chao1, riqueza (1) índice de Shannon) y homogeneidad (2) índice de Simpson) de cada muestra. 4) Estimación de la curva de cobertura de diversidad acorde a la cantidad de “reads” mediante el proceso de inter- y extrapolación.

La comparación entre muestras se lleva a cabo por medio de los índices beta. Los índices beta consisten en análisis de agrupamientos de las muestras, los cuales pueden tomar como base la estimación de la matriz de distancias de disimilitud de Bray-Curtis, Jaccard, UniFrac, entre otras. La estimación de la matriz de distancias de disimilitud estima de manera cuantitativa que tan diferentes o similares son las muestras entre sí. Entre más alto el valor entre dos muestras, indica que es mayor la disimilitud o que son muy diferentes. Entre más bajo el valor entre dos muestras, mayor es la similitud entre ellas. Una de las maneras en que se puede visualizar la estimación de la matriz de distancias de disimilitud es mediante un agrupamiento jerarquizado, el cual desplegará como resultado un dendrograma en el cual se agruparán las muestras en ramas o clados acorde a la similitud de las muestras comparadas. Este análisis de agrupamiento se puede extender, tomando como base de los cálculos a la matriz de distancias de similitud estimada previamente, a análisis de coordenadas principales (PCoA), escalamiento multidimensional (NMDS), de correspondencia restringida (CCA) y redundancia basada en distancias (dbRDA), para la representación gráfica para la estimación de los agrupamientos de las muestras en dos o en tres dimensiones. También, se pueden llevar a cabo análisis de agrupamientos independientes de la estimación de la matriz de distancias de disimilitud, los cuales son: análisis de componentes principales (PCA) y el análisis de agrupamiento bidimensional no supervisado. El soporte estadístico de los agrupamientos de las muestras se lleva a cabo mediante los análisis de varianza permutacional (PERMANOVA) y/o de los análisis de similitud (ANOSIM), tomando en cuenta el o los factores que están influyendo en el agrupamiento. Estos factores pueden ser los tratamientos aplicados, ambientales (temperatura, humedad, estación del año), estado de salud, etc. Para la estimación de la matriz de distancia de disimilitud, los análisis de agrupamientos dependientes de esta matriz estimada e independientes, así como para la estimación del PERMANOVA y/o ANOSIM es necesario la utilización de los paquetes `vegan`, `ape`, `ggplot2`, `dplyr` del lenguaje de programación R (Fig. 14.8).

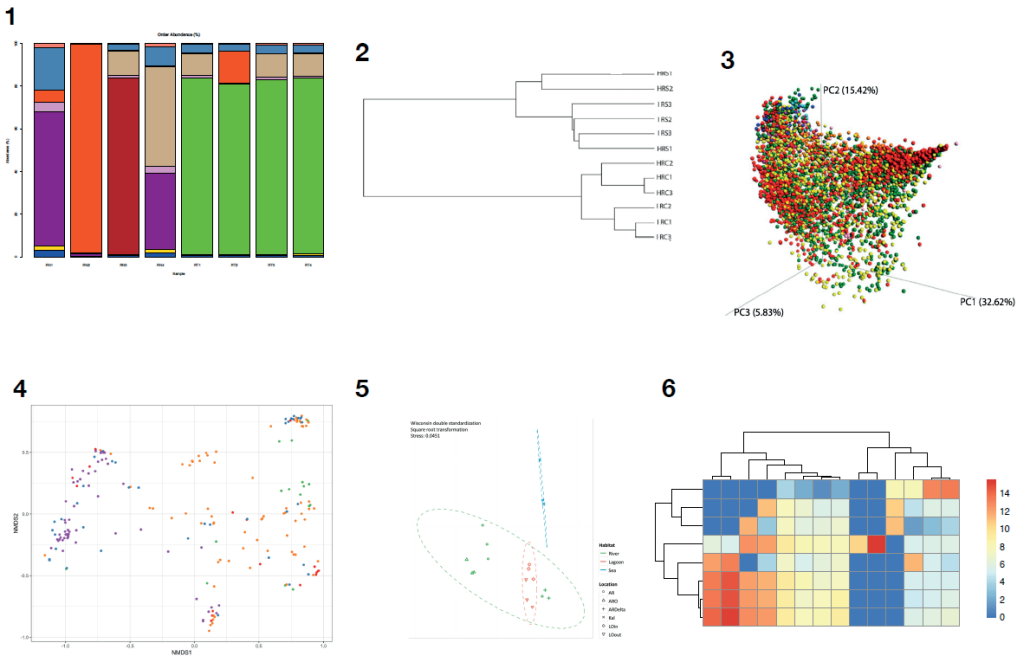


Figura 14.8 Análisis de similitud/disimilitud entre muestras o índices de diversidad beta. 1) Visualización de la estructura de las comunidades bacterianas en las muestras. 2) Dendrograma del análisis de agrupación jerarquizada de muestras. 3) Análisis de componentes principales (PCA). 4.) Análisis de escalamiento multidimensional (NMDS). 5) Análisis de correspondencia restringida (CCA). 6) Análisis de agrupamiento bidimensional no supervisado.

Protocolos

Toma de muestras

La microbiota ha sido analizada por muchos años usando cultivo tradicional y por métodos moleculares (reacción en cadena de la polimerasa, PCR), los cuales excluyen a un gran número de microorganismos que no pueden ser cultivados fuera de su hospedero. La secuenciación de nueva generación

(NGS) es una herramienta que permite considerables avances en el estudio de la ecología microbiana. En esta tecnología, mantener la integridad de las muestras recolectadas, es un desafío importante en la investigación del microbioma. Desde la recolección de la muestra, la composición del microbioma puede comenzar a cambiar, por esta razón, la congelación a menudo se considera un método estándar de preservación; sin embargo, si la colecta se realiza en ubicaciones remotas, en donde transportar y almacenar en frío resulta costoso, la preservación con compuestos químicos son una opción si se desea evitar el sesgo en el momento de analizar los datos.

Las ventajas de un buen método de preservación de muestras para análisis de la microbiota son: estabilizar la muestra desde el punto de recolección evitando sesgos introducidos por el crecimiento microbiano, homogeneizar las muestras creando una buena reproducibilidad en comparación con muestras frescas o congeladas, un agente conservante que lisa toda la muestra, inactivará los virus y hará segura a la muestra permitiendo que todos los tipos de ADN (bacteriano, fúngico, viral, hospedero) se unirán durante el aislamiento del ADN.

En plantas, el análisis de los perfiles taxonómicos bacterianos ha incrementado drásticamente, la microbiota asociada a raíz y las bacterias endófitas de la planta son las de mayor análisis. Para preservar las muestras de suelo sin que estos presenten efectos en sus perfiles taxonómicos bacterianos; las temperaturas recomendadas son de $4 + 2^{\circ}\text{C}$ para un máximo de tiempo de 3 meses, y de $-20 + 2^{\circ}\text{C}$ para un periodo de tiempo de un año.

Muestras de suelo que no pueden ser preservadas por temperatura pueden ser guardadas con amortiguadores ofrecidos por las diferentes casas comerciales, sin embargo, diversos autores recomiendan analizar la adecuada para cada tipo de muestra por procesar.

Se ha observado un efecto en las comunidades bacterianas al preservar muestras de suelo con LifeGuard de MOBIO comparado con muestras recién colectadas o preservadas en una solución con DMSO-EDTA. Los autores demuestran que el método de conservación por refrigeración presenta perfiles taxonómicos bacterianos similares a las muestras recién colectadas, comprobando ser un método de preservación eficiente.

Protocolo para la toma de muestra para análisis de la microbiota asociada a la rizosfera, rizoplano y endófitas de plantas

Importante: Seleccionar 3 plantas las cuales representan una muestra al realizar la combinación del material colectado de ellas (proceso mencionado como pool). Para los análisis de perfiles taxonómicos se requiere un mínimo de dos muestras por punto de análisis para la validación estadística.

- 1) Para muestras de rizosfera: Tomar la muestra de tierra adherida a la superficie de la raíz que no se sacude fácilmente (aproximadamente 1mm desde la superficie de la raíz hacia dirección del suelo), recolectar con ayuda de un amortiguador de lavado (amortiguador de sales de fosfatos PBS es recomendado) en un tubo falcón de 50 ml y preservar (Fig. 14.1).
- 2) Para muestras de rizoplano: Tomar la raíz y eliminar el excedente de sustrato tierra dejando solo aquella que se encuentra estrictamente pegada a las raíces. Esta tierra se recupera por medio de lavados de la raíz con ayuda de vortex (Fig. 14.1).
- 3) Para muestras de bacterias endófitas: Lavar perfectamente la raíz para eliminar el sustrato de tierra pegado a ella por vortex, preservar la raíz en tubo falcón (ver Barraza *et al.* 2020).
- 4) Preservación de las muestras: Guardar los tubos falcón con las muestras previo al procesamiento de extracción a -20 ° C (Fig. 14.1).

Extracción de ADN

Las muestras de suelo a menudo se consideran como uno de los entornos más desafiante, debido a la diversidad de las especies presentes y la variedad de inhibidores enzimáticos (ácidos húmicos o metales pesados) que pueden ser co-extraídos con el ADN.

El método de extracción de ADN es de importancia para los análisis de taxonomía bacteriana, ya que varios grupos de investigación han observado el impacto que puede tener sobre las comunidades bacterianas. La lisis celular es un paso inicial y crucial en los procesos de extracción de ADN, en los que se incluye disrupción física, química y enzimática. La disrupción física puede aumentar la cantidad de ADN, sin embargo, puede potenciar el corte del ADN genómico en fragmentos pequeños que pueden conducir a productos químé-

ricos en el resultado final de la secuenciación. Los métodos de lisis química y enzimática tienen menos probabilidad de dañar el ADN, pero pueden introducir sesgos al no ser general a todos los organismos blancos en una muestra con una amplia gama de contenido bacteriano.

El método tradicional de extracción de ADN de suelos involucra procesos tales como incubación con enzimas líticas y extracción con fenol/cloroformo. Muchos paquetes comerciales para extracción de ADN han sido desarrollados, los cuales se basan en su mayoría en la extracción directa, siendo los componentes secreto comercial. La purificación de ADN puede causar diferencias en la cantidad y pureza de la muestra. La cantidad y pureza del ADN extraído están relacionadas con la diversidad microbiana, en consecuencia como la diversidad de las secuencias obtenidas refleja la presencia de una especie determinada, debido a que cantidad de ADN no equivale a la diversidad representada. Para evaluar el método de extracción para muestras destinadas a los análisis metagenómicos, se debe considerar la cantidad y calidad del ADN recuperado, la diversidad de las secuencias obtenidas, los índices de diversidad de los datos y la reproducibilidad de los resultados entre replicas.

Teng y colaboradores (2018), analizaron seis métodos de extracción. Los autores observaron que entre los paquetes comerciales no se tienen efectos en la cantidad y calidad del ADN. Interesantemente se observó que fue leve el efecto en la calidad del protocolo de extracción de ADN del fenol-cloroformo. El impacto en la microbiota se vio afectado por el método de extracción evaluado, sin embargo, el método de fenol-cloroformo es reproducible entre replicas y las taxas más representativas son identificadas. Fue a nivel de abundancia relativa en donde los métodos de extracción difieren.

Protocolo para la extracción de ADN genómico para análisis de la microbiota asociada a la rizosfera, rizoplano y endófitas de plantas

- 1) Colectar en un tubo eppendorf de 1.5 mL 500 mg de suelo (rizoplano y rizosfera) o de tejido de raíz (previamente pulverizado con ayuda de nitrógeno líquido). Posteriormente adicionar 1 mL de amortiguador de extracción (SDS 2%, NaCl 1.4 M, Tris 100 mM, y EDTA 20 mM).
- 2) Incubar las muestras por 1 hora a 55 °C. Posteriormente dar un segundo período de incubación a temperatura del cuarto por 72 horas.

- 3) Transcurrido el segundo período de incubación, tomar 1 mL del sobrenadante y transferir a un tubo eppendorf de 1.5 mL, adicionar 1 volumen de fenol:cloroformo:alcohol isoamílico (25:24:1 v/v/v) pH 8.0, mezclar y centrifugar a 12,000 x g's por 15 minutos.
- 4) Terminada la centrifugación tomar la fase acuosa con ayuda de una micropipeta y transferir a un tubo eppendorf de 1.5 mL, adicionar 500 µL de isopropanol por 24 horas a 4 °C.
- 5) Transcurrido el tiempo de incubación, realizar una centrifugación a 12,000 x g's por 10 minutos. Eliminar el sobrenadante. Adicionar 500 µL de alcohol al 70 % (v/v), centrifugar a 12,000 x g's por 10 minutos, descartar el sobrenadante y repetir la operación.
- 6) Concluido el último lavado con etanol, descartar sobrenadante y secar la pastilla para eliminar todo residuo de etanol. Como último paso, re-suspender con amortiguador TE (10:1) o H₂O estéril (grado biología molecular) y guardar a -20 °C.
- 7) Las muestras de ADN total son cuantificadas para determinar la calidad y la cantidad. Para ello, en una nanodrop se cuantifica 1 µL de muestra. Se toma importancia a la lectura con relación de 260/230 y la de 260/280 para determinar la calidad. Para determinar la concentración se toma la lectura de 260.

Secuenciación de Nueva Generación

La secuencia del gen 16S ARNr se utilizó por primera vez en 1985 para análisis filogenéticos, debido a que contiene regiones altamente conservadas para el diseño de oligonucleótidos. Las regiones hiper-variables 16S son para identificar las características filogenéticas de los microorganismos, convirtiéndose en el gen marcador más utilizado para perfilar la comunidad bacteriana. La secuencia completa del gen 16S ARNr consiste de nueve regiones hiper-variables que son separadas por nueve regiones altamente conservadas las cuales fueron nombradas de V1 hasta V9. El análisis de cada región del 16S RNAr dará como resultado una composición distinta de una comunidad microbiana dada. La elección de la región hiper-variable del 16S ARNr para secuenciación es un paso importante. En décadas pasadas, la región V1-V3 fue frecuente en la plataforma 454 (ofrece lecturas de una longitud de un solo extremo hasta de 750 pb). En fechas actuales, la región V3-V4 y la V4-V5 han permitido

una caracterización precisa y rentable de muestras de microbioma, empleado por ello por diversos grupos de investigación, siendo las plataformas MiSeq, NextSeq500 y HiSeq 2500 las más empleadas en la última década.

Se ha observado que la estimación de las comunidades bacterianas empleando los fragmentos de la V3 del 16S ARNr tiene una alta resolución para taxones de rango inferior (especie y genero), siendo una buena región para los análisis de taxonomía bacteriana.

Protocolo para la amplificación y secuenciación de la región V3 del 16S ADNr

- 1) Amplificar la región V3-V4 del 16S ADNr por PCR empleando los oligonucleótidos v3-338f y los v4-806r, como templado se emplea ADN total (Figura 14.9).
- 2) Amplificar la región V3 usando los oligonucleótidos V3-338F y V3-533r con adaptadores de illumina siguiendo las recomendaciones del manual, se emplea como templado el producto de PCR amplificado en la sección 1.
- 3) Cuantificar los productos de PCR resultantes de esta segunda amplificación con el instrumento qubit.
- 4) Generar las lecturas de secuenciación usando la plataforma Illumina MiSeq (2x 150 pb) o MiSeq (2 x 300 pb).

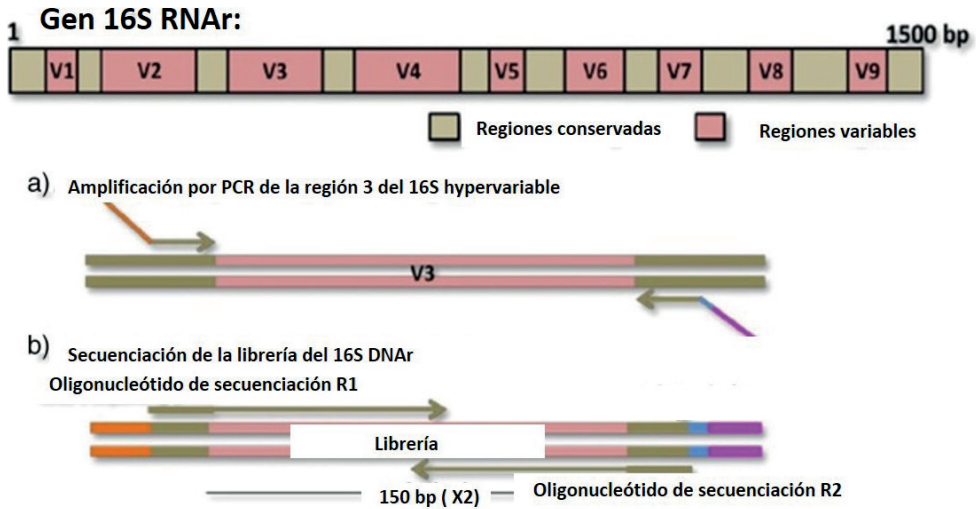


Figura 14.9 Flujo de trabajo del sistema de secuenciación de una librería de 16S ADNr. Tomada y modificada de Ranjith *et al.* (2016).

Análisis bioinformático en MG-Rast

Los primeros pasos en el análisis de cualquier metagenoma implican comparar esas secuencias con secuencias conocidas en bases de datos. Esta tarea intensiva proporciona los tipos de datos básicos para muchos análisis posteriores. Existe un sistema abierto totalmente automatizado para el procesamiento de datos de metagenoma empleando bases de datos ya disponibles. El servidor MG-RAST es un sistema de códigos para la comparativa genómica. Los usuarios pueden cargar datos de secuencias en formato fastq. Las secuencias serán normalizadas y procesadas de forma automática.

Protocolo para el uso de la plataforma del MG-RAST para análisis de secuencias 16S ADNr

- 1) Registrar un usuario en la plataforma: Emplear la página web <http://metagenomics.anl.gov/> para el registro y gestión de usuario, este primer paso tiene dos funciones, limita el acceso a cada conjunto de datos para el usuario y colaboradores y asegura una dirección de correo electrónico válida en caso de que la correspondencia sea requerida. Una vez que inician sesión, los

usuarios pueden ver sus propios metagenomas y los disponibles públicamente. En la plataforma se permite tener los datos confidenciales y pueden ser liberados en el momento que se autorice (Fig. 14.10).

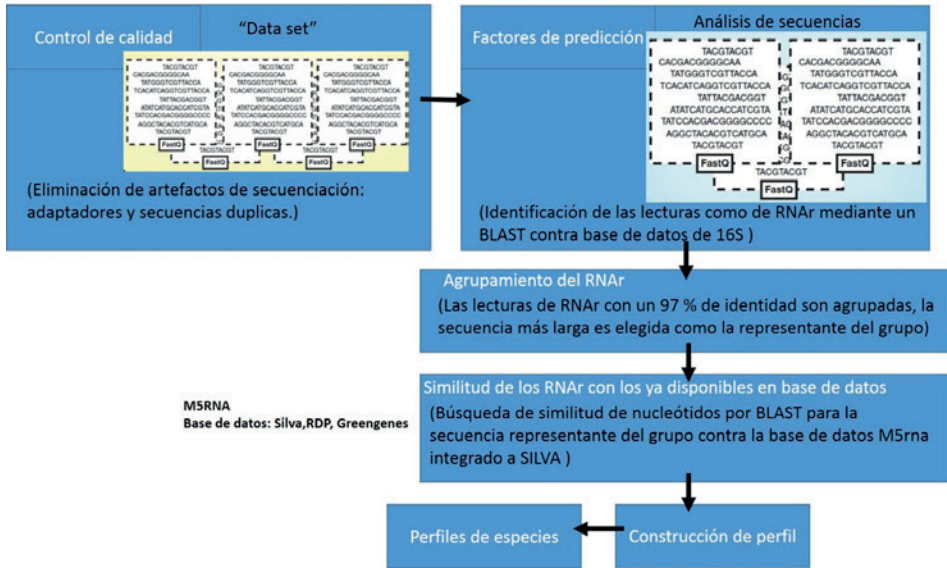


Figura 14.10 Flujo de trabajo desarrollado en la MG-RAST, tomado de Wilke *et al.* (2014).

- 2) Cargar los datos a la plataforma: Los usuarios pueden cargar las lecturas sin ensamblar, sin procesar o como contigs ensamblados. Para ello, acceder a la plataforma del MG-RAST con usuario y contraseña, en la parte superior de la barra se observa un icono de la nube con flecha apuntando arriba, dar click, se despliega una ventana en donde se dan indicaciones de cómo subir los datos en formato fastq en la plataforma. Los datos tienen un tiempo de subida según la demanda de la plataforma. La plataforma permite visualizar el estado del proceso, al concluir se notifica al usuario.

En la plataforma se realiza el “pipeline” o flujo de trabajo de forma automática, el cual implica una curación de la lectura, ensamble, predicción de genes y anotación tanto funcional como taxonómico de los genes resultantes.

Después de cargar los datos, un paso de normalización es ejecutado, generando secuencias únicas y removiendo secuencias duplicadas. En un segundo paso, las secuencias son analizadas en busca de genes codificadores de proteínas a través de BLASTX contra la base de datos “SEED” en el cual se encuentra integrados datos genómicos de una variedad de fuentes como la base de datos de SILVA, RDP entre otros.

- 3) Analizar los datos en la plataforma: Para los análisis en la plataforma se dirige al icono con símbolos de barras, al dar click se despliega una ventana con el título de crear nuevo análisis, se selecciona la base de datos. En las opciones se encuentran RDP, Silva SSU, Silava LSU (para taxonomía bacteriana). Seleccionar Silva SSU y dar click en adicionar. En la parte inferior se selecciona el tipo de secuencia y metagenoma que se analizará mediante el uso del nombre o ID otorgado al momento de haber subido los datos en formato fastq. En el recuadro izquierdo de la parte inferior aparecerán los metagenomas con ID similar, seleccionar el que será sometido al análisis y dar click a la flecha del recuadro céntrico. Posterior a esto se observará que el metagenoma aparecerá en el recuadro derecho. La operación se repite si se quiere analizar varios metagenomas. Concluida la selección dar click a la palomita, para cargar los genomas completos.

En la sección superior derecha aparecerá una ventana con parámetros default, seleccionar el nivel del análisis (Dominio, fila, clase, orden, familia, género u especie), para dar una idea gráfica de la abundancia seleccionar “stacked bar”, seleccionar la normalización de los datos. Los datos pueden ser extraídos de la plataforma, dar click en exportar, seleccionar formato TSV, para la obtención de ellos.

Abreviaturas

ADN: ácido desoxirribonucleico.
ADNr: ácido desoxirribonucleico ribosomal.
ARN: ácido ribonucleico.
ARNr: ácido ribonucleico ribosomal.
ARNm: ácido ribonucleico mensajero.
SAR: resistencia sistémica adquirida.
ISR: resistencia sistémica inducida.
N: nitrógeno.
Fe: fierro.
IAA: ácido indol-acético.
PRR: receptores de reconocimiento de patrones.
MAMP: patrones moleculares asociados a microbios.
MTI: inmunidad activada por MAMP.
MAPK: proteínas quinasas activadas por mitógeno.
DAMP: patrones moleculares asociados a daños.
ETI: inmunidad activada por efectores.
ABA: ácido absicico.
SA: ácido salicilico.
JA: ácido jasmónico.
PGPB: bacterias promotoras del crecimiento vegetal.
VOC: compuestos volátiles orgánicos.
QS: quorum sensing.
AI: autoinductores.
AHL: acil-homoserin lactona.
NGS: secuenciación de nueva generación.
OTUS: unidades taxonómicas operativas.
PCR: reacción en cadena de la polimerasa.
QC: control de calidad.
ITS: espaciador transcrito interno.
PcoA: análisis de coordenadas principales.
NMDS: escalamiento multidimensional.
CCA: análisis de correspondencia restringida.
PCA: análisis de componentes principales.
PERMANOVA: análisis de varianza permutacional.

ANOSIM: análisis de similitud.

SDS: dodecilsulfato sódico.

NaCl: cloruro de sodio.

Tris: tris(hidroximetil)aminometano.

EDTA: ácido etilendiaminotetraacético.

Pb: pares de bases.

SSU: subunidades cortas (16S/18S).

LSU: subunidades largas (23S/28S).

Glosario

Microbioma: comunidad de microorganismos que habitan un lugar determinado, como un ser vivo.

Fitobioma: microbioma específico de plantas.

Rizobioma: microbioma que abarca la raíz y la rizosfera de las plantas.

Rizosfera: zona localizada entre 1-3 mm alrededor de las raíces.

Rizoplano: superficie de la raíz donde se incluyen todos los microorganismos adherida a ella.

Core microbiome: microbioma que es localizado en una especie en particular.

Filosfera: superficie de la parte aérea de las plantas.

Filobioma: microbioma de la filosfera.

Endofítico: parte interna de la planta donde podemos encontrar diferentes microorganismos habitando en ella.

Endosfera: microbioma de la parte endofítica de la planta.

Agradecimientos

Las investigaciones actuales del grupo son apoyadas por CONACYT/México y CIBNOR/México.

Bibliografía

- Barraza, A., Caamal-Chan, M.G., Castellano, T. & Loera-Muro, A. (2020). Bacterial community characterization of plants belonging to *Solanaceae* family cultivated in desert soils. *Annals of Microbiology*, 70, 34.
- Berg, G., Grube, M., Schloter, M. & Smalla, K. (2014). The plant microbiome and its importance for plant and human health. *Frontiers in Microbiology*, 5, 491.
- Bukin, Y.S., Galachyants, Y.P., Morozov, I.V., Bukin, S.V., Zakharenko, A.S. & Zenskaya, T.I. (2019). Data Descriptor: The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Scientific DATA*, 6, 190007.
- Bushnell, B., Rood, J. & Singer, E. (2017). BBmerge – Accurate paired shotgun read merging via overlap. *PLoS ONE*, 12(10), e0185056.
- Caamal-Chan, M.G., Loera-Muro, A., Castellanos, T., Aguilar, C.J., Marfil, M.D. & Barraza, A. (2019). Analysis of the bacterial communities and endosymbionts of natural populations of *Bemisia tabaci* in several crop field from Mexico semi-arid zone. *Annals of Microbiology*, 69, 909-922.
- Davis, A., Kohler, C., Alsallaq, R., Hayden, R., Maron, G. & Margolis, E. (2019). Improved yield and accuracy for DNA extraction in microbiome studies with variation in microbial biomass. *BioTechniques*, 66, 285-289.
- de Souza, R., Ambrosini, A. & Passaglia, L.M.P. (2015). Plant growth-promoting bacteria as inoculants in agricultural soils. *Genetics and Molecular Biology*, 38, 401-419.
- Edwards, J., Johnson, C., Santos, C., Lurie, E., Kumar, N., Bhatnagar, S., Eisen, J.A. & Sundaresan, V. (2014). Structure, variation, and assembly of the root-associated microbiomes of rice. *The Proceedings of the National Academy of Sciences*, 20, E911-E920.
- Ewing, B., Hillier, L., Wendl, M.C. & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, 8, 175-185.
- Ewing, B. & Green, P. (1998). Base-calling of automated sequences traces using phred. II. *Genome Research*, 8, 186-194.
- Fahad, S., Hussain, S., Bano, A., Saud, S., Hassan, S. & Shan, D. (2015). Potential role of phytohormones and plant growth-promoting rhizobacteria in

- abiotic stresses: consequences for changing environment. *Environmental Science and Pollution Research*, 22, 4907-4921.
- Fouhy, F., Clooney, A.G., Stanton, C., Claesson, M. & Cotter, P.D. (2016). 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiology*, 16, 123.
- Garrido-Oter, R., Nakano, R.T., Dombrowski, N., Ma, K.W., AgBiome Team, McHardy, A.C. & Schulze-Lefert, P. (2018). Modular traits of the rhizobiales root microbiota and their evolutionary relationship with symbiotic rhizobia. *Cell Host Microbe*, 24, 155-167.e5.
- Hermosa, R., Viterbo, A., Chet, I. & Monte, E. (2012). Plant-beneficial effects of *Trichoderma* and of its genes. *Microbiology*, 158, 17-25.
- Hsieh, T.C., Ma, K.H. & Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, 7, 1451-1456.
- Jones, J.D.G., Vance, R.E. & Dangl, J.L. (2016). Intracellular innate immune surveillance devices in plants and animals. *Science*, 354, 1117-1125.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Keegan, K.P., Glass, E.M. & Meyer, F. (2016). MG-RAST, a metagenomic service for analysis of microbial community structure and function. In Martin, F. & Uroz, S. (eds) *Microbial Environmental Genomics (MEG)*. *Methods in Molecular Biology*, Vol 1399. Humana press, New York, NY.
- Klammer, S., Mondini, C. & Insam, H. (2005). Microbial community fingerprints of composts stored under different conditions. *Annals of Microbiology*, 55, 299-305.
- Lauber, C.L., Zhou, N., Gordon, J.I., Knight, R., & Firer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiology Letters*, 307, 80-86.
- Leach, J.E., Triplett, L.R., Argueso, C.T. & Trivedi, P. (2017). Communication in the phytobiome. *Cell*, 169, 587-596.

- Levy, S.E. & Myers, R.M. (2016). Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics*, 17, 95-115.
- Levy, A., Conway, J.M., Dangel, J.L. & Woyke, T. (2018). Elucidating bacterial gene functions in the plant microbiome. *Cell Host Microbe*, 24, 475-485.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. (2012). Comparison of Next-Generation Sequencing Systems. *Journal of Biotechnology and Biomedicine*, 2012, 251364.
- Luo, J., Tao, Q., Wu, K., Li, J., Qian, J., Liang, Y., Yang, X. & Li, T. (2017). Structural and functional variability in root-associated bacterial microbiomes of Cd/Zn hyperaccumulator *Sedum alfredii*. *Applied Microbiology and Biotechnology*, 101, 7961-7976.
- Mabood, F., Zhou, X. & Smith, D.L. (2014). Microbial signaling and plant growth promotion. *Canadian Journal of Plant Science*, 94, 1051-1063.
- Mahanty, T., Bhattacharjee, S., Goswami, M., Bhattacharyya, P., Das, B., Ghosh, A. & Tribedi, P. (2017). Biofertilizers: A potential approach for sustainable agriculture development. *Environmental Science and Pollution Research*, 24, 3315-3335.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilkening, J. & Edwards, R.A. (2008). The metagenomics RAST-serve a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 19, 386.
- Mhlongo, M.I., Piater, L.A., Madala, N.E., Labuschagne, N. & Dubery, I.A. (2018). The chemistry of plant-microbe interactions in the rhizosphere and the potential for metabolomics to reveal signaling related to defense priming and induced systemic resistance. *Frontiers in Plant Science*, 9, 112.
- Mutz, K.O., Heilkenbrinker, A., Lönne, M., Walter, J.G. & Stah, F. (2013). Transcriptome Analysis Using Next-Generation Sequencing. *Current Opinion in Biotechnology*, 24, 22-30.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2014). *Vegan: community ecology package*. R package version 2.2-0. <http://CRAN.Rproject.org/package=vegan>.
- Pausch, J. & Kuzyakov, Y. (2018). Carbon input by roots into the soil: quantification of rhizodeposition from root to ecosystem scale. *Global Change Biology*, 24, 1-12.

- Randle-Boggis, R.J., Helgason, T., Sapp, M. & Ashton, P.D. (2016). Evaluating techniques for metagenome annotation using simulated sequence data. *FEMS Microbiology Ecology*, 92(7), fiw095.
- Ranjith, M.T., Chellappan, M., Harish, E.R., Girija, D. & Nazeem, P.A. (2016). Bacterial communities associated with the gut of tomato fruit borer, *Helicoverpa armigera* (Hübner)(*Lepidoptera:Noctuidae*) base on illumina Next-Generation Sequencing. *Journal of Asia-Pacific Entomology*, 19, 333-340.
- Rao, R.M., Pasha, S.N. & Sowdhamini, R. (2016). Genome-wide survey and phylogeny of S-ribosylhomocysteinase (LuxS) enzyme in bacterial genomes. *BMC Genomics*, 17, 742.
- Sasse, J., Martinoia, E. & Northen, T. (2018). Feed your friends: Do plant exudates shape the root microbiome?. *Trends in Plant Science*, 23, 25-41.
- Smith, D.L., Subramanian, S., Lamont, J.R. & Bywater-Ekegård, M. (2015). Signaling in the phytomicrobiome: breadth and potential. *Frontiers in Plant Science*, 6, 709.
- Spence, C., Alff, E., Johnson, C., Ramos, C., Donofrio, N., Sundarsan, V., *et al.* (2014). Natural rice rhizospheric microbes suppress rice blast infections. *BMC Plant Biology*, 14, 130.
- Tatangelo, V., Franzetti, A., Gandolfi, I., Bestetti, G. & Ambrosini, R. (2014). Effect of preservation method on the assessment of bacterial community structure in soil and water samples. *FEMS Microbiology Letters*, 356, 32-38.
- Teixeira, P.J.P., Colaianni, N.R., Fitzpatrick, C.R. & Dangl, J.L. (2019). Beyond pathogens: Microbiota interactions with the plant immune system. *Current Opinion in Microbiology*, 49, 7-17.
- Teng, F., Darveekaran, S.S., Zhu, P., Li, S., Huang, S., Li, X., Xu, J. & Yang, F. (2018). Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Scientific Reports*, 8, 16321.
- Vetrovsky, T. & Baldrian, P. (2013) The Variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS ONE*, 8, e57923.
- Villena, J., Kitazawa, H., Van Wees, S.C.M., Pieterse, C.M.J. & Takahashi, H. (2018). Receptors and signaling pathways for recognition of bacteria in

- livestock and crops: prospects for beneficial microbes in healthy growth strategies. *Frontiers in Immunology*, 9, 2223.
- Whiteley, M., Diggle, S.P. & Greenberg, E.P. (2018). Progress in and promise of bacterial quorum sensing research. *Nature*, 555, 126.
- Whon, T., Chung, W.H., Lim, M., Song, E.J., Kim, P.S., Hyun, D.W., Shin, N.R., Bae, J.W. & Nam, Y.D. (2018). The effects of sequencing platforms on phylogenetic resolution in 16S rRNA gene profiling of human feces. *Scientific Data*, 5, 180068.
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
- Wilke, A., Glass, E.M., Bischof, J., Braithwaite, D., Souza, M.D., Gerlach, W., Harrison, T., Keegan, K.P., Matthews, H., Paczian, T., Tang, W., Trimble, W.L., Wilkening, J., Desai, N. & Meyer, F. (2014). MG-RAST Manual for version 3.3.6, revision 7.
- Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K.P., Paczian, T., Trimble, W.L., Bagchi, S., Grama, A., Chaterji, S. & Meyer, F. (2016). The MG-RAST metagenomics database and portal in 2015. *Nucleic Acids Research*, 44, 590-594.
- Yang, B., Wang, Y. & Qian, P.Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*, 17, 135.
- Zielinska, S., Radkowski, P., Blendowska, A., Ludwig, A.G., Los, M.J. & Los, M. (2017). The choice of the DNA extraction method may influence the outcome of the soil microbial community structure analysis. *Microbiology Open*, 6(4), e00453.