



Herramientas moleculares utilizadas para el análisis metagenómico. Revisión



Nohemí Gabriela Cortés-López ^a

Perla Lucía Ordóñez-Baquera ^{a*}

Joel Domínguez-Viveros ^a

^a Universidad Autónoma de Chihuahua, Facultad de Zootecnia y Ecología. Chihuahua, México.

*Autor de correspondencia: plordonez@uach.mx

Resumen:

La metagenómica utiliza técnicas de biología molecular para analizar la diversidad de los genomas microbianos (metagenomas). La diversidad de los metagenomas se ha analizado mediante marcadores moleculares para clasificar bacterias y arqueas en grupos taxonómicos a nivel de género. Entre los marcadores moleculares más utilizados se encuentran los genes ribosomales, genes que codifican subunidades del citocromo C y algunos genes constitutivos (*gyrB*, *rpoB*, *rpoD*, *recA*, *atpD*, *infB*, *groEL*, *pmoA*, *sodA*). El marcador más utilizado es el gen 16S rRNA para clasificar bacterias y arqueas de muestras metagenómicas, aunque no permite clasificar de forma adecuada algunas secuencias. Sin embargo, con la secuenciación del gen completo 16S rRNA se identifican todas las secuencias de las regiones hipervariables, por lo que se ha logrado clasificar hasta nivel taxonómico de especie con este marcador molecular. La secuenciación de próxima generación, también llamada secuenciación masiva o de alto rendimiento ha ayudado a describir metagenomas complejos como los de muestras ambientales, con importancia ecológica, así como metagenomas que crecen en ambientes extremos. También han ayudado a estudios relacionados con sanidad animal y en humanos, y en el ámbito agroalimentario. Específicamente, tanto el uso del marcador molecular 16S rRNA como la secuenciación de alta eficiencia combinadas con el uso de las herramientas bioinformáticas para el análisis metagenómico se han usado para describir el metagenoma

ruminal, una comunidad microbiana de gran importancia debido a que está involucrada en la producción animal de carne y leche. A pesar de los muchos estudios que se han realizado en este campo, aún faltan microorganismos por descubrir y caracterizar.

Palabras clave: Marcador molecular, Gen 16S rRNA, Metagenómica, Diversidad microbiana, Secuenciación de alto rendimiento.

Recibido: 17/12/2018

Aceptado: 23/10/2019

Introducción

La metagenómica se basa en el uso de técnicas de biología molecular para analizar la diversidad de genomas microbianos, llamados también metagenomas, a partir de muestras ambientales. La diversidad microbiana de los metagenomas se ha analizado mediante el uso del gen 16S rRNA, que codifica para el ARN ribosómico que conforma la subunidad pequeña de los ribosomas. Este gen comprende regiones conservadas y variables en bacterias y arqueas. El gen 16S rRNA se ha utilizado como marcador molecular, ya que permite clasificar a las bacterias y arqueas en grupos taxonómicos de acuerdo con las familias o géneros.

En los primeros estudios de diversidad microbiana de muestras ambientales se utilizaban métodos dependientes de cultivo, donde sólo se lograban estudiar aquellos microorganismos que se pudieran aislar en el laboratorio. Mediante el avance de las técnicas de biología molecular, se ha logrado analizar la diversidad microbiana a través del uso de métodos independientes de cultivo, obteniendo información más precisa de los genomas bacterianos. Uno de estos métodos más usados es la amplificación por PCR de fragmentos del gen 16S rRNA y en algunos casos seguida de electroforesis en gel de gradiente desnaturante (DGGE). Estas técnicas se han utilizado para analizar la diversidad bacteriana ruminal, los cambios que se producen en la comunidad microbiana y la expresión génica después de los cambios en la dieta del rumiante^(1,2). Otro avance que ha permitido analizar de forma más amplia la diversidad microbiana del rumen, es la secuenciación dirigida de las regiones variables del gen 16S para diferenciar los microorganismos filogenéticamente muy cercanos, analizar los genes y genomas que degradan la biomasa en el rumen, caracterizar la microbiota ruminal y estudiar los efectos de levaduras sobre la diversidad bacteriana en el rumen^(3,4,5).

El reciente desarrollo de la metagenómica ha permitido el estudio de la diversidad microbiana de muestras ambientales aislando y analizando el material genético total presente en una

muestra ambiental^(6,7). En sus inicios esta estrategia se utilizó para la búsqueda de nuevas enzimas con potencial biotecnológico, extrayendo el ADN total contenido en una muestra ambiental, fragmentándolo y clonando genes de diferente tamaño en vectores como plásmidos (15 kb), fagos (hasta 20 kb), fósmidos y cósmidos (hasta 40 kb) y Cromosomas Artificiales Bacterianos (para fragmentos mayores). Estos vectores se insertaban en diferentes cepas huésped y se usaban sustratos fluorogénicos como indicadores de expresión. Sin embargo, en la búsqueda funcional de genes a través de los clones, la expresión de proteínas y la actividad enzimática eran de pequeña magnitud^(8,9,10).

Una parte crucial en la construcción de librerías metagenómicas es la extracción de los ácidos nucleicos a partir de la muestra. Existen dos estrategias principales para la extracción de ADN metagenómico: el tratamiento químico y la lisis directa con métodos mecánicos. Ambos métodos presentan ventajas y desventajas, con la lisis mecánica se recupera ADN de mayor diversidad microbiológica que con el tratamiento químico, sin embargo, con el tratamiento químico se obtiene ADN de mayor peso molecular. Respecto a la extracción de ARN se utilizan los mismos métodos de extracción para cualquier análisis de expresión, en los cuales se incluyen inhibidores de RNAsas y se recomienda congelar las muestras a -80 °C inmediatamente después de su recolección para evitar la degradación del ARN⁽⁹⁾.

Para seleccionar el método ideal de extracción se debe de tomar en cuenta el tipo de muestra, el ácido nucleico que se quiera purificar y el tipo de análisis que se pretenda realizar. Para el análisis metagenómico se han utilizado diferentes estrategias. Dentro de los métodos mecánicos se han usado perlas magnéticas para muestras de origen fecal, oral, piel, suelo y agua donde se han obtenido secuencias de alta calidad⁽¹¹⁾. Para el análisis de microbiomas ruminales, se han utilizado métodos que combinan extracción por perlas magnéticas (lisis mecánica) con columnas de extracción (tratamiento químico) para purificar ADN de microbioma ruminal^(11,12). Esta combinación aumentó el rendimiento de extracción respecto al uso de perlas magnéticas y columnas de extracción por separado. Otros métodos de identificación utilizan sondas de isótopos estables (Stable-isotope probing SIP) mediante las que se identifican los microorganismos que incorporan dichos isótopos a través del uso sustratos marcados. En particular, la técnica de sondas de isótopos estables de ácidos nucleicos (Nucleic acids-SIP) utiliza sustratos con isótopos de ¹³C y/o ¹⁵N, los cuales se incorporan a los genomas bacterianos y de esta manera pueden ser rastreados⁽¹³⁾. Otros sustratos con sondas de isótopos estables son ¹³CH₃-OH, ¹³C-fenol y 5-bromo-2-deoxiuridina. Sin embargo, las limitaciones del uso de sustratos marcados con isótopos estables incluyen el entrecruzamiento y el reciclaje de los isótopos dentro de la comunidad microbiana, lo que resulta en la pérdida del enriquecimiento específico de los microorganismos analizados⁽¹³⁾.

También se han desarrollado técnicas para identificar genes que cambian sus niveles de expresión durante distintos procesos biológicos. Por ejemplo, la hibridación sustractiva supresora (Suppression Subtractive Hybridization SSH) se ha utilizado para identificar variaciones entre muestras complejas de ADN como las del ambiente ruminal^(9,13). Los análisis de expresión diferencial permiten comparar el perfil de expresión genética de una comunidad microbiana antes y después de estar expuesta a una condición y/o sustrato específico y de esta forma identificar genes de importancia que presentan cambios en los perfiles de expresión génica por efecto de dicha condición y/o sustrato⁽¹³⁾. Otra técnica ampliamente utilizada en estudios de expresión génica son los *microarrays*, que ofrecen como ventaja una rápida identificación y caracterización de un número elevado de clones. Aunque los *microarrays* se pueden usar para la identificación de un gran número de genes conservados, dependen de secuencias conocidas reportadas previamente en bases de datos, por lo que se elimina la posibilidad de identificar genes nuevos^(8,10). Más recientemente se ha usado la secuenciación masiva para obtener toda la información posible del metagenoma presente en una muestra. Uno de los primeros trabajos con secuenciación masiva fue la identificación del metagenoma del Mar de los Sargazos en donde se generaron, anotaron y analizaron 1.045 billones de pares de bases de secuencias no redundantes para identificar el contenido genético, diversidad y abundancia relativa de los microorganismos. Se estimó que los datos obtenidos provenían de al menos 1,800 especies genómicas que incluyeron 148 filotipos de bacterias desconocidas y más de 782 genes nunca antes descritos que codifican para fotorreceptores parecidos a rodopsinas^(10,14).

La secuenciación masiva del metagenoma por “*shotgun*” tiene la característica de secuenciar todo el ADN presente en la muestra por lo que se pueden clasificar a los microorganismos taxonómicamente hasta el nivel de especie. Además, con las secuencias obtenidas por este tipo de secuenciación se pueden llegar a descubrir genes con funciones nunca antes descritas e incluso se pueden seleccionar las secuencias que pertenecen al gen 16S rRNA para realizar anotaciones taxonómicas. Estas clasificaciones se realizan con el uso de herramientas bioinformáticas que busquen homología con las secuencias analizadas en diferentes bases de datos ya existentes⁽¹⁵⁾. Específicamente en ambientes ruminales, se han analizado librerías metagenómicas para evaluar los efectos de dietas en el microbioma ruminal mediante perfiles metagenómicos y se ha utilizado el marcador del gen 16S rRNA para la determinación y clasificación de la diversidad microbiana de las secuencias^(3,5). Sin embargo, algunas de las secuencias de estas muestras no se han logrado clasificar adecuadamente, por lo que utilizar al menos otro marcador filogenético molecular diferente al gen 16S rRNA podría mejorar la clasificación taxonómica⁽¹⁵⁾.

En el presente trabajo se hace una revisión de las herramientas utilizadas para el análisis de metagenomas que van desde los marcadores moleculares clásicos hasta los utilizados con datos obtenidos a partir de secuenciaciones masivas, con énfasis en los metagenomas de ambientes ruminales.

Marcadores moleculares para el análisis metagenómico

Un marcador molecular es un segmento de ADN que corresponde a un gen o regiones no codificantes del genoma, estos segmentos de ADN permiten identificar diferentes variantes (alelos) y se localizan en un sitio determinado en los cromosomas (locus). Las diferencias que se obtienen en estos fragmentos de ADN se conocen como polimorfismos y se pueden detectar por hibridación de secuencias de ácidos nucleicos, secuenciación de nucleótidos, comparación de la longitud de los fragmentos producidos por la reacción en cadena de la polimerasa (PCR) y a través de sitios reconocidos por enzimas de restricción. Los marcadores moleculares se pueden utilizar para clasificar grupos taxonómicos, poblaciones, familias o individuos, tanto en eucariotas como en procariotas^(16,17). En estudios genéticos se han utilizado diversos marcadores moleculares en animales domésticos, en fauna silvestre, en especies en peligro de extinción, así como en pruebas forenses y de paternidad. Los más conocidos son RFLPs, minisatélites, AFLPs, RAPDs, microsátélites y SNPs (Cuadro 1).

Las características más relevantes que deben tener los marcadores moleculares para optimizar los estudios metagenómicos incluyen: (1) que sean genes de copia única (genes que sólo tienen una o dos copias en todo el genoma) ya que proporcionan menor incertidumbre que los marcadores de genes que presentan copias múltiples (gen con copias repetidas en el genoma); (2) que la secuencia del gen del marcador sea de fácil alineamiento para facilitar el análisis filogenético; (3) que la proporción de la región de sustitución del gen sea la suficiente para proporcionar información necesaria para su clasificación; (4) que los cebadores sean selectivos para amplificar el gen del marcador, pero no universales, para evitar falsos positivos; (5) que no haya excesiva variación en la secuencia del marcador que limite la determinación de la ancestría. Los genes que se utilizan como marcadores moleculares para clasificar microorganismos se describen a continuación.

Genes ribosomales

Los genes de RNA ribosomal se consideran la herramienta idónea para la clasificación taxonómica ya que son genes altamente conservados y evolutivamente estables, pero que contienen regiones hipervariables. La secuenciación de dichas regiones ha generado grandes bases de datos que ayudan a la clasificación taxonómica⁽¹⁸⁾. Los ribosomas de bacterias y arqueas constan de dos subunidades, la subunidad pequeña contiene un solo tipo de ARN (16S) y una subunidad grande que contiene dos tipos de ARN (5S y 23S)⁽¹⁷⁾.

16S rRNA. Este gen también se designa 16S rDNA, pero la Sociedad Americana de Microbiología (American Society for Microbiology, ASM) ha decidido utilizar el término “16S rRNA” para uniformizar la información. Tiene una secuencia aproximada de 1.550 pb de longitud y contiene regiones variables y conservadas con secuencias de oligonucleótidos

únicas para cada grupo filogenético^(18,19). La comparación de las secuencias del gen 16S rRNA de bacterias desconocidas con secuencias conocidas en las bases de datos es de gran ayuda para clasificar a las bacterias a nivel de género e incluso se ha llegado a identificar especies en algunos casos^(20,21).

5S rDNA. Es un gen de aproximadamente 120 nucleótidos de longitud, se encuentra prácticamente en todos los ribosomas con excepción de los mitocondriales, de algunos hongos, de animales superiores y de la mayoría de protistas. Aunque la secuencia de este gen está altamente conservada, la fiabilidad de este gen como marcador está cuestionada debido a que su longitud es muy pequeña y por lo tanto no ofrece la suficiente resolución para contribuir significativamente a comprender relaciones filogenéticas entre taxones⁽¹⁷⁾.

23S rDNA. Es un gen de aproximadamente 3,000 nucleótidos de longitud que se localiza en la subunidad grande de los ribosomas en procariontes. Este gen presenta inserciones y deleciones más grandes que el gen 16S rRNA. Las inserciones estables y deleciones de algunas bases en el gen 23S rDNA son características comunes en algunas clases y subclases de bacterias. Estos cambios complican los análisis, ya que las diferentes posiciones no pueden considerarse para realizar clasificaciones filogenéticas correctas⁽²²⁾. El gen 23S rDNA se ha utilizado en conjunto con el 16S rRNA para la clasificación taxonómica de bacterias no cultivables. Además se ha utilizado el espaciador intergénico (ITS) localizado en la región 16S-23S, la cual es una región muy variable para diferenciar entre dos cepas pertenecientes a la misma subespecie^(22,23).

Genes que codifican subunidades del citocromo c

Citocromo oxidasa I/II (COI/II). La enzima citocromo c oxidasa es una proteína de la cadena transportadora de electrones que se encuentra tanto en bacterias como en las mitocondrias de organismos eucariotas. Los genes COI y COII codifican para dos de las siete subunidades polipeptídicas del complejo citocromo c oxidasa. El gen COI evoluciona más lentamente en comparación con otros genes mitocondriales y es ampliamente utilizado en estudios filogenéticos⁽¹⁷⁾.

Genes que codifican proteínas con funciones conservadas

En estudios donde se ha encontrado mayor diversidad de microorganismos se han utilizado técnicas de análisis de comunidades moleculares basadas en el gen 16S rRNA respaldado por estudios de análisis de secuencias multilocus (MLSA), que implican la secuenciación de varios genes que codifican proteínas con funciones conservadas (*housekeeping genes*) para evaluar la diversidad en colecciones de cepas aisladas⁽²⁴⁾. En estos estudios las secuencias parciales de genes que codifican para proteínas con funciones conservadas se utilizan para

generar árboles filogenéticos y posteriormente resolver filogenias. La principal desventaja del uso del gen 16S rRNA como marcador filogenético es su resolución insuficiente a nivel de especie. No obstante, el empleo de un análisis filogenético complementario basado en genes codificadores de proteínas⁽²⁵⁾ permite incrementar la resolución de las filogenias a nivel infragenérico y determinar nuevas cepas. Hay más de 50 esquemas individuales de MLSA disponibles y las bases de datos de MLSA (<http://www.mlst.net/> y <http://www.pubmlst.org>) también se pueden usar para identificar secuencias microbianas no conocidas a nivel de especie^(24,26).

Los genes que se han utilizado en MLSA son aquellos que codifican subunidades de enzimas ubicuas, como la subunidad β de la ADN girasa (*gyrB*), la subunidad β de la ARN polimerasa (*rpoB*), el factor sigma 70 (sigma D) de la ARN polimerasa (*rpoD*), la recombinasa A (*recA*), la subunidad β de ATP sintasa FOF1 (*atpD*), el factor de iniciación de traducción IF-2 (*infB*), la modificación del tRNA GTPasa ThdF o TrmE (*thdF*) y la chaperonina GroEL (*groEL*)^(24,26).

La subunidad β de la metano-monooxigenasa en partículas (*pmoA*) se ha utilizado como marcador funcional para la detección de metanótrofos aerobios. La metano-monooxigenasa es la enzima responsable de la etapa de conversión inicial de metano a metanol. Se conocen dos formas de esta enzima, la metano-monooxigenasa soluble (sMMO) y una enzima unida a la membrana, la metano-monooxigenasa en partículas (pMMO). El gen *pmoA* es el marcador utilizado con mayor frecuencia, ya que está presente en la mayoría de las bacterias aeróbicas metanotróficas. También está presente en bacterias desnitrificadoras anaeróbicas⁽²⁷⁾. Otro marcador que se puede usar para la detección de metanótrofos es el gen *mxhF* que codifica la subunidad mayor de la metanol deshidrogenasa^(27,28).

Como ejemplo de esta aproximación, se puede citar el trabajo de Sánchez-Herrera *et al*⁽²⁶⁾, que han utilizado el gen 16S rRNA como marcador molecular de referencia para identificar y clasificar cepas del género *Nocardia* a nivel de género. Sin embargo, por ser un gen con múltiples copias genera problemas en la identificación de cepas aisladas de casos clínicos. Tras testar otros genes a través de la amplificación por PCR de sus segmentos: *sodA* (gen que codifica la enzima superóxido dismutasa), *hsp65* (proteína de choque térmico), *secA1* (subunidad de translocasa de la preproteína secA), *gyrB* (subunidad β de la ADN girasa), *rpoB* (subunidad β de la ARN polimerasa) y el espaciador intergénico 16S-23S, los autores únicamente pudieron discriminar entre especies estrechamente relacionadas de *Nocardia* mediante el gen *sodA*. El fragmento de 386 pb del gen *sodA* presenta regiones variables, con segmentos de 4 y 5 pb, y tiene potencial para ser utilizado como marcador molecular.

En conclusión, si bien existe una gran diversidad de marcadores moleculares para analizar comunidades microbianas, hasta el momento el estándar de oro para la clasificación de secuencias obtenidas a partir de muestras sigue siendo el gen 16S rRNA.

El uso de secuenciación masiva en la metagenómica

Si bien el análisis metagenómico se inició con el uso de distintos marcadores moleculares como AFLP, RAPDs, 16S rRNA etc. (Cuadro 1), se ha visto que algunos de estos marcadores mejoran su eficiencia cuando la técnica que se utiliza para identificarlos incluye su secuenciación en lugar de caracterizarlos por medio de reacciones con enzimas de restricción y/o amplificación por PCR. Desde sus inicios, la secuenciación del ADN con la tecnología de Sanger generó un gran impacto prácticamente en todas las ramas de las ciencias biológicas dentro de las cuales se encuentran los estudios de comunidades microbianas. Actualmente, el uso de la secuenciación Sanger puede generar hasta 96 secuencias por corrida con una longitud promedio de 650 pb, lo que podría ser suficiente para un análisis de marcadores filogenéticos⁽¹⁵⁾. Este tipo de estudios se conocen como secuenciación de primera generación y da como resultado secuencias de alta calidad de una longitud entre 500 a 1000 pb. Sin embargo su desventaja es que la proporción de marcadores moleculares que pueden ser secuenciados en una corrida, comparado con el total de microorganismos presentes en una muestra metagenómica es muy baja⁽¹¹⁾.

Cuadro 1: Métodos moleculares utilizados en estudios genéticos

Marcador molecular	Características	Referencia
RFLP (polimorfismo de longitud de fragmentos de restricción)	Se basa en los cambios de nucleótidos en un genoma que se dan en un sitio de reconocimiento de enzimas de restricción. En ciencias forenses se ha utilizado para probar si los tejidos provenientes de las escenas del crimen (sangre, piel, esperma, etc.) pertenecen a algún sospechoso. En manejo de razas de animales para rastrear la progenie además de usarse para pruebas de paternidad y para el diagnóstico de enfermedades.	Khlestkina ⁽¹⁶⁾ Wakchaure <i>et al</i> ⁽⁵⁰⁾
Minisatélites o VNTR (variaciones en el número de repeticiones en tándem)	Son secuencias cortas de 10 a 60 pb, repetidas en número variable en uno o más sitios del genoma. Se han utilizado para identificar linajes paternos en individuos y evaluar la diversidad genética en poblaciones de animales domésticos, de fauna silvestre y de gramíneas.	Kumar <i>et al</i> ⁽⁵¹⁾ Lang <i>et al</i> ⁽⁵²⁾

AFLP (polimorfismo de longitud de fragmento amplificado)	Es la amplificación de fragmentos genómicos digeridos con enzimas de restricción que reconocen secuencias dispersas a lo largo del genoma. Se ha utilizado para estudios de ADN “fingerprinting”, para clonar y mapear secuencias de ADN específicas y para hacer mapas genéticos.	Khlestkina ⁽¹⁶⁾ Kumar <i>et al</i> ⁽⁵¹⁾
RAPD (ADN polimórfico amplificado al azar)	Utilizan cebadores cortos de secuencia arbitraria para dirigir una reacción de amplificación en regiones discretas del genoma y se obtienen fragmentos de diversos tamaños. Se han utilizado para estudio de ADN “fingerprinting”, para relacionar especies cercanas, en mapeo genético, en genética de poblaciones, en genética evolutiva molecular y en estudios de razas genéticas en animales y plantas.	Beuzen <i>et al</i> ⁽⁵³⁾ Vignal <i>et al</i> ⁽⁵⁴⁾ Wakchaure <i>et al</i> ⁽⁵⁰⁾
Microsatélites o SSR (secuencias simples repetidas)	Son secuencias de 2 a 6 pb repetidas en tándem en todo el genoma y presentan un elevado polimorfismo en función del número de repeticiones que se encuentran en regiones de genes no codificantes. Se han utilizado en estudios de identificación de animales, evaluación de recursos genéticos, pruebas de paternidad, investigación de enfermedades, determinación de la variación genética dentro y entre razas, genética de poblaciones, migración mapeo de genes y genomas y para detectar polimorfismos incluso en estudios <i>in silico</i> .	Khlestkina ⁽¹⁶⁾ Beuzen <i>et al</i> ⁽⁵³⁾ Kumar <i>et al</i> ⁽⁵¹⁾ Duran <i>et al</i> ⁽⁵⁵⁾
SNP (polimorfismo de un solo nucleótido)	Son regiones de ADN en las que se observa la sustitución de un nucleótido por otro, o la adición o eliminación de uno o pocos nucleótidos. Se ha utilizado en el análisis de genes de herencia biparental y en el análisis de diferencias genéticas, para hacer mapas genéticos y para detectar variaciones genéticas dentro de especies.	Khlestkina ⁽¹⁶⁾ Yu <i>et al</i> ⁽⁵⁶⁾ Beuzen <i>et al</i> ⁽⁵³⁾ Kumar <i>et al</i> ⁽⁵¹⁾

Con el surgimiento de las tecnologías de secuenciación masiva, conocidas como “Next Generation Sequencing technologies (NGS)” se pueden secuenciar millones de moléculas de ADN de manera simultánea, lo que facilita en gran medida el estudio de la diversidad microbiana⁽¹⁵⁾. Una de las primeras tecnologías de secuenciación de alto rendimiento fue la pirosecuenciación 454, que se utilizó para la secuenciación dirigida de amplicones de genes de RNA ribosomal⁽²⁹⁾. Esta técnica tenía la ventaja de que se podían obtener secuencias de hasta 1,200 pb, pero con un error significativamente mayor al de otras plataformas de

secuenciación (1%) y a un mayor costo⁽¹⁵⁾. La secuenciación de segunda generación, también conocida como secuenciación de lecturas cortas (50 a 400 pb) utiliza principalmente la plataforma Illumina⁽¹¹⁾. Entre sus ventajas, cabe mencionar que se pueden obtener una mayor cantidad de lecturas, tienen un porcentaje de error aproximado de 0.1% y comparativamente su costo es reducido⁽¹⁵⁾. Actualmente es la tecnología más popular, pero requiere de una fase de análisis bioinformático más complejo que el de otras plataformas.

Tradicionalmente, cuando se usan estas dos plataformas (pirosecuenciación 454 e Illumina) para el análisis metagenómico con el uso del marcador 16S rRNA, se realiza un paso previo de amplificación por PCR, limitando las especies identificadas únicamente a bacterias y arqueas ya que los cebadores serán siempre para amplificar fragmentos del gen 16S rRNA. Si la población también incluye a microorganismos eucariotas como levaduras y protozoarios, estos no podrán ser detectados. Por otro lado, este paso de amplificación por PCR conlleva un enriquecimiento del ADN lo que produce un sesgo hacia las especies que se encuentran en mayor proporción provocando que las especies que se encuentran en menor porcentaje difícilmente puedan ser detectadas. Finalmente, este tipo de análisis identifica microorganismos hasta nivel de género⁽²⁹⁾.

Una alternativa para el aumento de la resolución a nivel taxonómico radica en el estudio metagenómico con las técnicas de secuenciación masiva llamadas “Whole Genome Shotgun sequencing” (WGS) y “Shotgun metagenomics sequencing (SMS)”, en las cuales el ADN metagenómico total es secuenciado^(30,31). La mayor ventaja de estos métodos es que los microorganismos se pueden clasificar hasta nivel de especie además de que se pueden identificar no solo procariotas sino también eucariotas y de que no requiere el paso previo de amplificación por PCR por lo que se elimina el sesgo. Otra ventaja de estas secuenciaciones es que al tener secuencias de todo el ADN presente en la muestra se pueden seleccionar las correspondientes al gen 16S rRNA para utilizarlo como marcador molecular taxonómico y además se pueden buscar secuencias de genes de otros marcadores polimórficos constitutivos (MLSA) que ayudarían a hacer una mejor clasificación de los microorganismos. Las principales desventajas son el mayor costo que la secuenciación dirigida del gen 16S rRNA y que requieren de análisis bioinformáticos de datos más complejos⁽³²⁾. Se han realizado varios estudios para identificar metagenomas en una amplia gama de ambientes poblacionales y se ha utilizado tanto la secuenciación dirigida del gen 16S rRNA como la secuenciación completa de metagenomas mediante WGS y/o SMS.

Herramientas bioinformáticas para el análisis metagenómico

Es importante señalar que los datos obtenidos de secuenciación masiva requieren utilizar herramientas bioinformáticas para poder ser analizados. Cuanto mayor sea la cantidad de

datos generados se requerirá de mayores recursos bioinformáticos⁽¹⁵⁾, tanto de aplicaciones que implementen algoritmos de análisis, como de bases de datos con información sobre genomas microbianos (Cuadro 2).

Cuadro 2: Programas bioinformáticos para el análisis de secuencias metagenómicas

Aplicación bioinformática	Método de análisis	Referencia
MG-RAST	Asignación de anotaciones estructurales y funcionales de acuerdo con bases de datos de nucleótidos y proteínas por homología.	Meyer <i>et al</i> ⁽³³⁾
MOTHUR	Analiza secuencias del gen 16S rRNA, cuantifica parámetros ecológicos para medir diversidad α y β , visualiza el análisis mediante diagramas de Venn, heat maps y dendogramas, selecciona colecciones de secuencias basadas en su calidad y calcula la distancia de secuencia por pares.	Schloss <i>et al</i> ⁽³⁴⁾
QUIIME	Analiza secuencias microbianas del gen 16S rRNA, realiza perfiles taxonómicos y filogenéticos, y comparaciones entre las muestras.	Kuczynski <i>et al</i> ⁽³⁵⁾
PhaMe	Realiza comparaciones basados en SNPs de genomas completos, secuencias ensambladas y secuencias son procesar para el análisis filogenético y de evolución molecular.	Ahmed <i>et al</i> ⁽³⁶⁾
VITCOMIC1	Hace un análisis del gen 16S rRNA y secuencias de alto rendimiento para visualizar la composición filogenética de muestras metagenómicas.	Mori <i>et al</i> ⁽³⁷⁾
16SPIP	Detección rápida de microorganismos patógenos en muestras clínicas basadas en secuencias metagenómicas del gen 16S rRNA.	Miao <i>et al</i> ⁽³⁸⁾
PICRUSt	Algoritmo que tiene un enfoque de metagenómica predictiva a partir de datos del gen 16S rRNA y de una base de datos de genomas de referencia.	Langille <i>et al</i> ⁽³⁹⁾
CowPI	Utiliza PICRUSt para analizar datos del gen 16S rRNA del microbioma ruminal.	Wilkinson <i>et al</i> ⁽⁵⁷⁾

Kraken	Asignación de etiquetas taxonómicas en secuencias de DNA metagenómico utilizando la alineación de k-mers logrando una clasificación más precisa en comparación con BLAST.	Wood <i>et al</i> ⁽⁵⁸⁾
Kaiju	Clasificador de metagenomas que encuentra coincidencias máximas a nivel de proteína utilizando la transformación de Burrows-Wheeler, clasifica lecturas con mayor sensibilidad y precisión similar en comparación con los clasificadores basados en k-mers, especialmente en los géneros que están subrepresentados en las bases de datos de referencia.	Menzel <i>et al</i> ⁽⁵⁹⁾

Una de las aplicaciones más usada desde su lanzamiento es el servidor MG-RAST⁽³³⁾ que asigna anotaciones funcionales a las secuencias analizadas comparando dichas secuencias con bases de datos de proteínas y de nucleótidos por homología, además de permitir la realización de análisis filogenéticos. Esta herramienta es gratuita, de fácil acceso y se alimenta con la información proporcionada por los investigadores por lo que ayuda a terminar con el principal cuello de botella en el análisis de secuencias de metagenomas, que radica en la disponibilidad de información para asignar anotaciones genómicas⁽³³⁾. Otras dos herramientas bioinformáticas de amplio uso en metagenómica son MOTHUR⁽³⁴⁾, que también es de libre acceso y que se alimenta de la información metagenómica que los usuarios van agregando a una base de datos que se actualiza mensualmente y QUIIME⁽³⁵⁾, que se utiliza para el análisis de la comunidades microbianas de datos bacterias y arqueas.

Otro programa de amplio uso para el análisis de metagenomas es PhaME⁽³⁶⁾ (Phylogenetic and Molecular Evolutionary), que utiliza SNP de genomas completo para medir la diversidad interespecífica mediante análisis filogenético. PhaME⁽³⁶⁾ se puede utilizar para medir la divergencia entre las especies y entre cepas aisladas, además de minimizar los errores de secuenciación y ensamblaje. La genómica comparativa, incluidos los análisis filogenéticos basados en genes ortólogos y SNP requieren de genomas ensamblados o terminados. PhaME utiliza el enfoque basado en SNP de genomas completos disponibles en las bases de datos, secuencias ensambladas (contigs) y secuencias sin procesar para realizar análisis filogenéticos y de evolución molecular. Este programa combina algoritmos para hacer alineamientos de todo el genoma, mapeo de lecturas y construcción filogenética; utiliza comandos internos para inferir el genoma principal y SNP, inferir árboles y realizar otros análisis de evolución molecular. PhaME es especialmente útil para el análisis y detección de organismos poco abundantes en muestras de metagenomas y se ha utilizado en datos de muestras bacterianas, de virus, como el del ébola en Zaire y levaduras, entre otros⁽³⁶⁾.

Otras herramientas se focalizan en el análisis de las regiones hipervariables del gen 16S rRNA, como por ejemplo VITCOMIC1⁽³⁷⁾, que combina la información obtenida de la secuenciación dirigida del gen 16S rRNA así como de la secuenciación masiva WGS o SMS para visualizar mejor la composición filogenética de muestras metagenómicas, además de generar un registro más exacto de la comunidad microbiana. De la misma manera, la aplicación 16SPIP⁽³⁸⁾ también se ha utilizado para la detección rápida de microorganismos patógenos en muestras clínicas basada en datos de secuencias metagenómicas del 16S rRNA. En cuanto a los enfoques de "metagenómica predictiva", cabe destacar el algoritmo PICRUST⁽³⁹⁾, que utiliza modelos evolutivos para predecir metagenomas a partir de datos del gen 16S rRNA y una base de datos de genomas de referencia. Esta herramienta se ha utilizado con datos de muestras de microbiomas de suelos, de intestinos de mamíferos, de tapetes microbianos y de humanos⁽³⁹⁾, como por ejemplo el estudio de la microbiota bucal de humanos en el que se analizaron 6,431 muestras del gen 16S rRNA del Proyecto del Microbioma Humano^(39,40).

Ejemplos de caracterización metagenómica con metodologías de alto rendimiento

Existen diversos trabajos de caracterización metagenómica para identificar microorganismos que viven en ambientes de interés por su gran variabilidad e importancia ecológica (Cuadro 3). A continuación, se presentan algunos ejemplos de estos trabajos sin ánimo de ser exhaustivos. Por ejemplo, a partir de muestras de tres estaciones marinas que son parte de la expedición global Tara, se realizó una secuenciación masiva de 29 metagenomas⁽²⁹⁾. Al realizar el análisis taxonómico con los datos de las secuencias correspondientes al gen 16S rRNA se lograron identificar todas las regiones variables del gen (V1 a V9). Con fines comparativos también se realizó la secuenciación dirigida de gen 16S rRNA. Los resultados que obtuvieron indicaron que la eficiencia en la clasificación taxonómica con el uso de base de datos ribosomales RDP (Ribosomal Database Project) es similar para ambos tipos de secuenciación. No obstante, la secuenciación masiva ofrece dos grandes ventajas, reduce el error ocasionado en la PCR de amplicones y se genera una gran cantidad de datos funcionales que se pueden analizar a la par del análisis taxonómico.

Cuadro 3: Ejemplos de caracterización metagenómica

Muestra	Tipo de análisis	Referencia
Plancton marino de estaciones marinas de la expedición <i>Tara Oceans</i>	Perfiles taxonómicos y estructura de comunidades procariotas mediante secuenciación masiva dirigida de 16S rRNA	Logares <i>et al</i> ⁽²⁹⁾
Sedimentos del Manglar Sundarban	Análisis de diversidad y distribución de bacterias a través de la secuenciación dirigida de 16S rRNA	Basak <i>et al</i> ⁽⁴¹⁾
Sedimentos del mar Árabe	Análisis de la estructura y diversidad bacteriana en base a la secuenciación de una librería de 16S rRNA	Nair <i>et al</i> ⁽⁴²⁾
Aguas termales de Sungai Klah de Malasia	Análisis de diversidad mediante la secuenciación dirigida de la región V3-V4 de 16S rRNA	Chan <i>et al</i> ⁽⁴³⁾
Aguas termales de Mushroom Spring del Parque Nacional de Yellowstone	Diversidad microbiana basada en la secuenciación dirigida del gen 16S rRNA y secuenciación metagenómica.	Thiel <i>et al</i> ⁽⁴⁴⁾
Hielo basal del Glaciar Matanuska, Alaska	Análisis de diversidad microbiana de secuenciación dirigida del gen 16S rRNA y secuenciación metagenómica	Kayani <i>et al</i> ⁽⁴⁵⁾
Sangre de donantes sanos	Análisis del microbioma mediante amplificación por PCR y secuenciación dirigida de 16S rRNA	Païsse <i>et al</i> ⁽⁴⁶⁾
Microbioma fecal humano	Estudio comparativo del genoma completo por secuenciación masiva y secuenciación dirigida de 16S rRNA	Ranjan <i>et al</i> ⁽³²⁾
Queso Gouda pasteurizado y no pasteurizado	Análisis de diversidad mediante la secuenciación dirigida del gen 16S rRNA	Salazar <i>et al</i> ⁽⁴⁷⁾
Microbiota ileal y cecal de pollos de engorda	Análisis de diversidad mediante la amplificación de la región V3 del gen 16S rRNA	Mohd-Shaufi <i>et al</i> ⁽⁴⁸⁾
Microbiota adherida a la fibra en rumen bovino	Caracterización de los genes y genomas de ADN metagenómico	Hess <i>et al</i> ⁽³⁾

Rumen de bovinos productores de leche y carne	Análisis taxonómico del microbioma del rumen a través de la pirosecuenciación dirigida del gen 16S rRNA	Wu <i>et al</i> ⁽²⁰⁾
Microbiota ruminal en bovinos suplementados con levaduras	Análisis de la diversidad microbiana del rumen a través de la pirosecuenciación dirigida de la región ribosomal V1 del gen 16S rRNA	Pinloche <i>et al</i> ⁽⁵⁾
Microbiota ruminal en bovinos suplementados con tiamina	Análisis de la diversidad bacteriana a través de la secuenciación dirigida del gen 16S rRNA	Pan <i>et al</i> ⁽⁴⁹⁾
Microbioma de piel sana y con dermatitis digital bovina	Caracterización del microbioma y composición de genes funcionales de la piel sana, piel en etapas de lesión activa e inactiva mediante secuenciación masiva del genoma completo y anotación de las muestras mediante MG-RAST	Zinicola <i>et al</i> ⁽³⁰⁾
Líquido ruminal de tres fracciones del rumen bovino	Perfiles metagenómicos del rumen mediante la secuenciación masiva paralela no dirigida en ADN metagenómico.	Ross <i>et al</i> ⁽³¹⁾

Otro trabajo metagenómico en el ámbito de la secuenciación masiva se centró en analizar la diversidad y la distribución bacteriana presente en sedimentos del manglar tropical de Sundarban⁽⁴¹⁾. Para esta identificación se hizo uso de la secuenciación dirigida del 16S rRNA a través de pirosecuenciación 454, obteniendo un total de 153,926 secuencias. El análisis con el software MG-RAST permitió la identificación de 56,547 especies pertenecientes a 44 filotipos diferentes, siendo el más dominante el filotipo *Proteobacteria*. Por otro lado, el análisis metagenómico de sedimentos del mar Arábigo⁽⁴²⁾ con secuenciación Sanger dirigida a 16S rRNA clasificó las secuencias obtenidas en siete filotipos distintos donde también predominó el filotipo *Proteobacteria*.

Un gran número de trabajos se han centrado en la caracterización de metagenomas de ambientes extremos. Por ejemplo, la secuenciación del 16S rRNA y de genomas completos se ha utilizado para identificar la diversidad de bacterias termófilas presentes en aguas termales de Malasia cuya temperatura varía entre 50 y 110 °C⁽⁴³⁾. Al analizar los datos del 16S rRNA se encontraron cerca de 35 filotipos de los cuales *Firmicutes* y *Proteobacteria* representaron el 57% del microbioma. Respecto a las termófilas el 70% de las detectadas fueron anaerobias estrictas, sin embargo, *Hydrogenobacter spp.* (termofilos quimiolitotróficos obligados) representaron uno de los taxones con mayor frecuencia, encontrando también muchos microorganismos fotosintéticos termofílicos. La mayoría de

los filotipos identificados coincidieron con lo encontrado en la secuenciación de genomas completos. Gracias a este tipo de análisis se pudieron identificar y clasificar microorganismos extremos como lo son los termófilos, los anaerobios y los quimiolitotróficos que difícilmente se hubieran podido caracterizar con los métodos microbiológicos clásicos⁽⁴³⁾.

Otro estudio para identificar microbiota de ambientes extremos fue realizado a partir de muestras de microorganismos que crecen en los hongos que habitan el parque Yellowstone a través de la secuenciación dirigida de 16S rRNA⁽⁴⁴⁾. A lo largo de los años el estudio de los microorganismos de este hábitat se ha enfocado en bacterias clorofototróficas pertenecientes a los filotipos *Cyanobacteria* y *Chloroflexi*. Sin embargo, los resultados del estudio dieron a conocer que la variación microbiana está dominada por un solo taxón: *Roseiflexus spp.* que pertenece al grupo de los microorganismos fotótrofos anoxigénicos⁽⁴⁴⁾. La secuenciación dirigida del 16S rRNA junto con la de genomas completos se ha utilizado también en glaciares, para los que la información microbiana también está muy limitada. El primer estudio metagenómico de glaciares reportado⁽⁴⁵⁾ permitió identificar nueve diferentes genomas entre los cuales se encuentran *Anaerolinea*, *Synthrophus* y *Thiobacillus* y se encontraron rutas metabólicas involucradas en la oxidación del azufre y en la nitrificación. Existen ejemplos del uso de la secuenciación masiva en poblaciones metagenómicas dentro del ámbito de la salud, y agroalimentario. A modo de ejemplo en el ámbito de la salud humana, estudios de secuenciaciones dirigidas del 16S rRNA para describir la microbiota presente en la sangre de individuos sanos han mostrado que la sangre de personas sanas no es un tejido estéril⁽⁴⁶⁾. A nivel de filotipos, más del 80% de los microorganismos presentes en sangre pertenecían a *Proteobacterias* aunque también se encontraron filotipos de *Actinobacteria*, *Firmicutes* y *Bacteriodetes*. Ranjan *et al*⁽³²⁾ utilizaron diferentes estrategias para caracterizar el microbioma fecal humano. A partir de una sola muestra obtuvieron 194.1 x10⁶ lecturas provenientes de distintas estrategias de secuenciación (secuenciación dirigida de 16S rRNA, Illumina HiSeq, Illumina MiSeq) al compararlas, especialmente la secuenciación dirigida del gen 16S rRNA respecto a la secuenciación WGS concluyeron que esta última presenta más ventajas, ya que se aumenta la capacidad de identificar especies bacterianas, se aumenta la detección de la diversidad y de la predicción de genes y también se mejora la precisión en la detección de especies al aumentar la longitud de las secuencias.

En el ámbito agroalimentario la secuenciación dirigida de 16S rRNA también se ha utilizado para identificar microorganismos presentes en queso Gouda⁽⁴⁷⁾ cuando se preparó ya sea con leche de pasteurizada o sin pasteurizar, evaluando además los cambios por efecto del añejamiento. Dicho estudio permitió identificar 120 géneros en el queso sin pasteurizar y 92 en el queso pasteurizado. Además, en función del tiempo de añejamiento tenía una influencia significativa en la presencia de microbiota. Los géneros más abundantes en todas las muestras fueron *Bacillaceae*, *Lactococcus*, *Lactobacillus*, *Streptococcus* y *Staphylococcus*.

En el caso de pollos de engorda en crecimiento, se ha estudiado la variación de la microbiota ileal y cecal a través del tiempo⁽⁴⁸⁾. Para ello, se amplificó y secuenció la región hipervariable V3 del gen 16S rRNA. Los resultados mostraron que las comunidades microbianas cecales tuvieron más diversidad que las ileales. Además, se observó que la presencia de la bacteria *Clostridium* (potencialmente patógena) aumentaba conforme los animales iban creciendo y que la población de microorganismos benéficos como *Lactobacillus* era baja en todos los intervalos⁽⁴⁸⁾.

En el caso de los metagenomas ruminales, cabe destacar que uno de los primeros estudios de secuenciación se realizó para la búsqueda de enzimas celulolíticas nunca antes descritas⁽³⁾. En dicho estudio, se realizó pirosecuenciación 454, obteniendo 268 gigabases de información de ADN metagenómico. A partir de esta información, se lograron identificar 27,755 supuestos genes de enzimas carbohidrato-activas de los cuales 90 codificaban para posibles proteínas y de ellas el 57% eran enzimáticamente activadas por sustratos celulósicos. Otro estudio centrado en el metagenoma ruminal de becerros de ganado lechero y de novillos de ganado de carne⁽²⁰⁾ utilizó pirosecuenciación dirigida de 16S rRNA para evaluar la variación entre las poblaciones respecto al tipo de ganado. En este estudio se encontraron 8 filotipos, 11 clases, 15 familias y 17 géneros distintos, y diferencias en la abundancia de los filotipos encontrados entre ganado lechero y de carne. Los filotipos más abundantes fueron *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, *Fibrobacteres* y *Spirochaetes* en ambos tipos de ganado, pero con una menor abundancia de *Bacteroidetes* y *Proteobacteria* en ganado de carne. Es conocido que el uso de levaduras como aditivos nutricionales en bovinos trae mejoras en la producción de leche y en la ganancia de peso. Sin embargo, no se sabe si el efecto que provocan las levaduras es un estímulo general a todas las especies microbianas o solo afecta algunas del ambiente ruminal. Debido a lo anterior se realizó un estudio para evaluar los cambios en la microbiota ruminal cuando los animales se alimentaban con aditivo de levaduras y comparándolos cuando consumían solo la dieta basal⁽⁵⁾. En este trabajo se utilizó pirosecuenciación 454 de la región V1 del gen 16S rRNA para identificar la población de los microorganismos ruminales. Los resultados mostraron que se observaba un cambio en las principales bacterias fibrolíticas (*Fibrobacter* y *Ruminococcus*) y en bacterias utilizadoras de lactato (*Megasphaera* y *Selenomonas*) cuando se adicionaba el aditivo de levaduras. La secuenciación dirigida del gen 16S rRNA en la población de microorganismos ruminales de bovinos lecheros adultos cuando se combinaba con dietas de alto contenido de grano, se ha utilizado para evaluar el efecto de la tiamina como aditivo en la nutrición animal⁽⁴⁹⁾. Los resultados confirmaron que la suplementación de tiamina puede mejorar la función ruminal ya que se aumentó el número de bacterias celulolíticas cuando se les administró dicho aminoácido.

En el campo de la sanidad animal, también se ha utilizado la secuenciación de metagenomas completos. Por ejemplo, se ha comparado el metagenoma de piel con dermatitis digital bovina activa y en recesión con la piel de bovinos sanos para ver si se detectaban patógenos

que estuvieran involucrados con la patogénesis de la enfermedad⁽³⁰⁾. Las secuencias obtenidas se analizaron con MG-RAST y se lograron identificar seis filotipos principales, de entre los cuales, *Firmicutes* y *Actinobacteria* fueron los predominantes en el microbioma de los sanos mientras de *Spirochetes*, *Bacteroidetes* y *Proteobacteria* fueron los más abundantes en los enfermos activos y de recesión, confirmando así que la presencia de la enfermedad cambia la población del metagenoma.

La obtención de perfiles metagenómicos ruminales se ha realizado mediante la secuenciación de metagenomas completos a partir de muestras de líquido ruminal provenientes de tres diferentes bovinos y entre localizaciones diferentes del rumen⁽³¹⁾. Además de comparar con el metagenoma proveniente de heces de los mismos animales, los resultados indicaron que la variación de los perfiles metagenómicos era menor entre las muestras tomadas del mismo animal, aunque fueran tomadas de diferentes regiones del rumen. Contrariamente a lo esperado, no se encontró relación con el perfil metagenómico de heces y de líquido ruminal del mismo animal.

Conclusiones

Tradicionalmente el análisis metagenómico utilizaba metodologías laboriosas, como el uso de electroforesis en gradiente desnaturizante, la digestión de los genomas con enzimas de restricción y su visualización mediante geles de agarosa y/o de acrilamida. El desarrollo de las metodologías de secuenciación de ácidos nucleicos, sobre todo de las nuevas tecnologías de secuenciación masiva, han ayudado a disminuir esta problemática.

El gen 16S rRNA se ha considerado tradicionalmente el estándar de oro para clasificar microorganismos procariotas (bacterias y arqueas), ya que cumple con todas las características necesarias para ser un marcador molecular. No obstante, pese al gran número de trabajos que han empleado la secuenciación de las regiones hipervariables de este marcador, presenta la desventaja de no poder determinar taxones a nivel infragenérico. Una estrategia utilizada para mejorar la clasificación taxonómica ha sido la combinación del marcador 16S rRNA con algún otro gen de expresión constitutiva como los genes *sodA*, *hps65*, *gyrB*, entre otros, e incluso también se han usado genes que codifican para subunidades del complejo enzimático del citocromo c para llegar a clasificar a los microorganismos hasta especie.

En la última década, las tecnologías de secuenciación masiva han hecho posible que las poblaciones microbianas puedan ser analizadas con más profundidad, bien sea secuenciando el gen completo 16S rRNA, incrementando así la resolución de dicho marcador, o bien combinando la información de ese gen con la secuenciación de metagenomas completos. En este último tipo de análisis se obtienen secuencias de todo el material genómico presente en la muestra, lo que ofrece la gran ventaja de que además de hacer la clasificación taxonómica

también se puede obtener información funcional de los genes detectados. Así, a pesar de las limitaciones del análisis bioinformático requerido, el empleo de estas metodologías permite realizar análisis más completos.

No obstante, pese al desarrollo de las técnicas de secuenciación de alto rendimiento, la secuenciación dirigida del 16S rRNA en la plataforma de Sanger no está del todo obsoleta y la selección de la estrategia de análisis dependerá de los objetivos del estudio, del grado de precisión que se desee, del tamaño muestral y de los recursos económicos que se puedan destinar por parte del equipo investigador. Por ejemplo, si lo que se está buscando es la presencia y/o ausencia de un solo género bacteriano en particular, la secuenciación Sanger sería lo ideal ya que tiene la capacidad de secuenciar fragmentos relativamente grandes con mayor precisión que la de cualquier plataforma de secuenciación masiva. Si lo que se quiere es discriminar entre especies de un solo género bacteriano se pueden usar dos estrategias: la secuenciación de alguna región hipervariable del 16S rRNA con el uso de algún otro gen constitutivo (MLSA); o bien la secuenciación de todo el gen para obtener la información de todas las regiones hipervariables.

Actualmente, la metagenómica se enfrenta a numerosos retos derivados de la gran cantidad de información generada, su almacenamiento y la forma en la que esta debe ser tratada. A pesar de que se han diseñado muchas herramientas y aplicaciones para el análisis bioinformático de metagenomas, no existe un solo “protocolo” de análisis, por lo que cada estudio debe ser adaptado a la naturaleza de las muestras y a los objetivos del experimento. En conclusión, en estudios de diversidad microbiana el uso del marcador molecular 16S rRNA siempre estará vigente para hacer clasificaciones taxonómicas, ya sea a través de la secuenciación de una o dos de sus regiones hipervariables o del gen completo, e incluso se puede combinar con el uso de otro gen constitutivo como marcador molecular para realizar una mejor clasificación taxonómica. Por otro lado, las tecnologías de secuenciación masiva han mejorado mucho la capacidad de estudio y la velocidad de análisis de metagenomas. Esto se ha producido de forma particular en muestras ambientales con importancia ecológica, en sanidad tanto humana como animal, en estudio simbióticos de plantas con hongos endófitos, en la evaluación de metagenomas ruminales, por mencionar algunas.

Literatura citada:

1. Li W, Huan X, Zhou Y, Ma Q, Chen Y. Simultaneous cloning and expression of two cellulase genes from *Bacillus subtilis* newly isolated from Golden Takin (*Budorcas taxicolor* Bedfordi). *Biochem Biophys Res Commun* 2009;383(4):397-400.
2. Sadet S, Martin C, Meunier B, Morgavi DP. PCR-DGGE analysis reveals a distinct diversity in the bacterial population attached to the rumen epithelium. *Animal* 2007;1(7):939-944.

3. Hess M, Sczyrba A, Egan R, *et al.* Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 2011;331(6016):463-467.
4. Li RW, Connor EE, Li C, Ransom L, Baldwin VI, Sparks ME. Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. *Environ Microbiol* 2012;14(1):129-139.
5. Pinloche E, McEwan N, Marden JP, Bayourthe C, Auclair E, Newbold CJ. The Effects of a probiotic yeast on the bacterial diversity and population structure in the rumen of cattle. *PLoS One* 2013;8(7):e67824.
6. Kumar M, Shrivastava N, Teotia P, *et al.* Omics: Tools for assessing environmental microbial diversity and composition. In: Varma A. SA, ed. *Modern tools and techniques to understand microbes*. Springer, Cham; 2017:273-283.
7. Marshall IPG, Karst SM, Nielsen PH, Jørgensen BB. Metagenomes from deep Baltic Sea sediments reveal how past and present environmental conditions determine microbial community composition. *Mar Genomics* 2018;37:58-68.
8. Simon C, Daniel R. Metagenomic analyses: Past and future trends. *Appl Environ Microbiol* 2011;77(4):1153-1161.
9. Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R, Wittwer P. Metagenomic gene discovery: Past, present and future. *Trends Biotechnol* 2005;23(6):321-329.
10. Streit WR, Schmitz RA. Metagenomics - The key to the uncultured microbes. *Curr Opin Microbiol* 2004;7(5):492-498.
11. Pacheco-Arjona JR, Sandoval-Castro CA. Tecnologías de secuenciación del metagenoma del rumen. *Trop Subtrop Agroecosyst* 2018;21:587-598.
12. Yu Z, Morrison M. Improved extraction of PCR-quality community DNA from digesta and fecal samples. *BioTechniques* 2004;36:808-812.
13. Singh B, Bhat TK, Kurade NP, Sharma OP. Metagenomics in animal gastrointestinal ecosystem: a microbiological and biotechnological perspective. *Indian J Microbiol* 2008;48:216-227.
14. Venter JC, Remington K, Heidelberg JF, *et al.* Environmental genome shotgun sequencing of the Sargasso sea. *Science* 2004;304(5667):66-74.
15. Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Front Genet* 2015;6:348.

16. Khlestkina EK. Molecular markers in genetic studies and breeding. *Russ J Genet Appl Res* 2014;4(3):236-244.
17. Patwardhan A, Samit R, Roy A. Molecular markers in phylogenetic studies-A Review. *J Phylogenetics Evol Biol* 2014;02(02).
18. D'Amore R, Ijaz UZ, Schirmer M, Kenny JG, Gregory R, Darby AC, *et al.* A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics*. 2016;17:55.
19. Clarridge JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* 2004;17(4):840-862.
20. Wu S, Baldwin RL, Li W, Li C, Connor EE, Li RW. The bacterial community composition of the bovine rumen detected using pyrosequencing of 16S rRNA genes. *Metagenomics* 2012;1:235571.
21. Valenzuela-Gonzalez F, Martínez-Porchas M, Villalpando-Canchola E, Vargas-Albores F. Studying long 16S rDNA sequences with ultrafast-metagenomic sequence classification using exact alignments (Kraken). *J Microbiol Methods* 2016;122:38-42.
22. Ludwig W, Schleifer KH. Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. *FEMS Microbiol Rev* 1994;15(2-3):155-173.
23. Osorio CR, Collins MD, Romalde JL, Toranzo AE. Variation in 16S-23S rRNA intergenic spacer regions in *Photobacterium damsalae*: a Mosaic-Like structure. *Appl Environ Microbiol* 2005;71(2):636-645.
24. Glaeser SP, Kämpfer P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol* 2015;38(4):237-245.
25. Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF, Kjelleberg S. Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol* 2007;73(1):278-288.
26. Sánchez-Herrera K, Sandoval H, Mouniee D, *et al.* Molecular identification of *Nocardia* species using the *sodA* gene: Identificación molecular de especies de *Nocardia* utilizando el gen *sodA*. *New Microbes New Infect* 2017;19:96-116.
27. Dumont MG. Primers: Functional marker genes for Methylotrophs and Methanotrophs. In: McGenity T, Timmis KNB, editors. *Hydrocarbon and lipid microbiology protocols - Springer Protocols Handbooks*. Berlin: Springer Protocols Handbooks; 2014.

28. Kolb S, Stacheter A. Prerequisites for amplicon pyrosequencing of microbial methanol utilizers in the environment. *Front Microbiol* 2013;4(SEP):1-12.
29. Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmiento H, *et al.* Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol.* 2014;16(9):2659-2671.
30. Zinicola M, Higgins H, Lima S, Machado V, Guard C, Bicalho R. Shotgun metagenomics sequencing reveals functional genes and microbiome associated with bovine digital dermatitis. *PLoS ONE* 2015;10(7)e0133674.
31. Ross EM, Moate PJ, Bath CR, Davidson SE, Sawbridge TI, Guthridge KM, Cocks BG, Hayes BJ. High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genetics* 2012;13:53.
32. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of microbiome: Advantages of whole genome shotgun *versus* 16S amplicon sequencing. *Biochem Biophys Res Commun* 2016;469:967-977.
33. Meyer F, Paarmann D, D`Souza M, Olson R, Glass EM, Kubal M, *et al.* The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 2008;9:386.
34. Schloss PD, Westcott S. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75(23):7537-7541.
35. Kuczynski J, Stombauhg, Walters WA, Gonzalez A, Caporaso JG, Knight R. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Curr Protoc Bioinformatics* 2011;10:7.
36. Ahmed SA, Lo CC, Li PE, Davenport KW, Chain PSG. From raw reads to trees: Whole genome SNP phylogenetics across the tree of life. *bioRxiv* 2015:032250.
37. Mori H, Maruyama T, Yano M, Yamada T, Kurokawa K. VITCOMIC2: visualization tool for the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing. *BMC Systems Biol* 2018;12(2):30.
38. Miao J, Han N, Qiang Y, Zhang T, Li X, Zhang W. 16SPIP: a comprehensive analysis pipeline for rapid pathogen detection in clinical samples based on 16S metagenomic sequencing. *BMC Bioinformatics* 2017;18(16):568.

39. Langille MGI, Zaneveld J, Caporaso JG, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 2013;31(9):814-82.
40. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* 2012;13;486(7402):215-21. doi: 10.1038/nature11209.
41. Basak P, Pramanik A, Sengupta S, Nag S, Bhattacharyya A, Roy D, Pattanayak R, Ghosh A, Chattopadhyay D, Bhattacharyya M. Bacterial diversity assessment of pristine mangrove microbial community from Dhulibhashani, Sundarbans using 16S rRNA gene tag sequencing. *Genomics Data* 2016;7:76-78.
42. Nair HP, Puthusseri RM, Vincent H, Bhat SG. 16S rDNA-based bacterial diversity analysis of Arabian Sea sediments: A metagenomic approach. *Ecol Genet Genomics* 2017;3(5):47-51.
43. Chan CS, Chan KG, Tay YL, Chua TH, Goh KM. Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front Microbiol* 2015;6:177.
44. Thiel V, Wood JM, Olsen MT, Tank M, Katt CG, Ward DM, Bryant DA. The dark side of the mushroom spring microbial mat: Life in the shadow of chlorophototrophs. I. Microbial diversity based on 16S rRNA gene amplicons and metagenomic sequencing. *Front Microbiol* 2016;7:919.
45. Kayani MR, Doyle SM, Sangwan N, Wang G, Gilbert JA, Christner BC, Zhu TF. Metagenomic analysis of basal ice from an Alaskan glacier. *Microbiome* 2018;6:123.
46. Pâisse S, Valle C, Servant F, Courtney M, Burcelin R, Amar J, Lelouvier B. Comprehensive description of blood microbiome from healthy donors assessed by 16S targeted metagenomic sequencing. *Transfusion* 2016;56:1138-1147.
47. Salazar JK, Carstens CK, Ramachandran P, Shazer AG, Narula SS, Reed E, Ottesen A, Schill KM. Metagenomics of pasteurized and unpasteurized gouda cheese using targeted 16S rDNA sequencing. *BMC Microbiology* 2018;18:189.
48. Mohd-Shaufi MA, Sieo CC, Chong CW, Gan HM, Ho YW. Deciphering chicken gut microbial dynamics based on high-throughput 16S rRNA metagenomics analyses. *Gut Pathog* 2015;7:4
49. Pan X, Xue F, Nan X, Tang Z, Wang K, Beckers Y, Jiang L, Xiong B. Illumina sequencing approach to characterize thiamine metabolism related bacteria and the impacts of thiamine supplementation on ruminal microbiota in dairy cows fed high-grain diets. *Front Microbiol* 2017;8:1818.

50. Wakchaure R, Ganguly S, Para PA, Praveen PK, Qadri K. Molecular markers and their applications in farm animals: A Review. *Int J Recent Biotechnol* 2015;3(January 2016):23-29.
51. Kumar A, Tomar SS, Kumar A, Singh J. Importance of molecular markers in livestock improvement: a review. *Int J Agric Res Innov Technol* 2017;5(4):614-621.
52. Lang T, Li G, Yu Z, Ma J, Chen Q, Yang E, Yang Z. Genome-wide distribution of novel Ta-3A1 mini-satellite repeats and its use for chromosome identification in wheat and related species. *Agronomy* 2019;9(2):60.
53. Beuzen ND, Stear MJ, Chang KC. Molecular markers and their use in animal breeding. *Vet J* 2000;160(1):42-52.
54. Vignal A, Milan D, San Cristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* 2002;34(2002):275-305.
55. Duran C, Singhania R, Raman H, Batley J, Edwards D. Predicting polymorphic EST-SSRs in silico. *Mol Ecol Resour* 2013;13(3):538-545.
56. Yu H, Xie W, Wang J, *et al.* Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS One* 2011;6(3).
57. Wilkinson TJ, Huws SA, Edwards JE, Kingston-Smith A, Siu Ting K, *et al.* CowPI: a rumen microbiome focused version of the PICRUST functional inference software. *Frontiers in Microbiol* 2018;(9):1095.
58. Wood DE, Salzberg SL. Kraken: ultrafast metagenomics sequence classification using exact alignments. *Genome Biology*. 2014;15(3):R46.
59. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature communications* 2016;7:11257.