



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

CARACTERIZACIÓN METATRANSCRIPCIONAL DE LA MICROBIOTA
INTESTINAL EN POBLACIÓN INFANTIL MEXICANA CON OBESIDAD Y
SÍNDROME METABÓLICO

TESIS

QUE PARA OPTAR POR EL GRADO DE:
Maestro en Ciencias

PRESENTA:
LUIGUI MICHEL GALLARDO BECERRA

TUTOR PRINCIPAL:
DR. ADRIÁN OCHOA LEYVA
Instituto de Biotecnología

MIEMBROS DEL COMITÉ TUTOR:
DR. ENRIQUE MERINO PÉREZ
Instituto de Biotecnología
DR. JUAN ENRIQUE MORETT SÁNCHEZ
Instituto de Biotecnología

Cuernavaca, Morelos. Enero, 2019



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

El presente trabajo fue realizado en el Laboratorio 22 del grupo del Dr. Adrián Ochoa Leyva del Departamento de Microbiología Molecular del Instituto de Biotecnología de la Universidad Autónoma de México (UNAM), bajo la dirección del Dr. Adrián Ochoa Leyva. El proyecto fue financiado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) mediante la beca de maestría 450940 y el proyecto No. SALUD-2014-C01-234188.

Dedicatoria

Dedico esta tesis a mi familia que fue mi primera escuela y actualmente es mi motor para llegar más lejos. A todos los integrantes del laboratorio del Dr. Adrián Ochoa, que más allá de ser mis compañeros son mis amigos y que en su compañía me divierto y aprendo todos los días. Al Dr. Adrián por 'adoptarme' en su laboratorio y por guiarme en el camino de la ciencia. A mis amigos: aquellos de toda la vida, los nuevos que me acompañaron en la maestría y los que conocí en el Instituto. Por último, a aquellos que implícita o explícitamente me dieron una oportunidad para continuar con mis estudios de posgrado: gracias por todo.

Agradecimientos

Primero quiero agradecer a los miembros de mi comité tutor el Dr. Enrique Merino Pérez y el Dr. Juan Enrique Morett Sánchez por enriquecer constantemente este proyecto y ayudarnos a darle la dirección correcta. A mi tutor principal el Dr. Adrián Ochoa Leyva por abrirme un lugar en su laboratorio y permitirme continuar con mis estudios de posgrado bajo su dirección, así como su guía en el mundo del quehacer científico.

De igual forma a los miembros del jurado evaluador de esta tesis: Dr. José Luis Puente García, Dra. Blanca Itzel Taboada Ramírez, Dr. Armando Hernández Mendoza, Dr. Samuel Canizales Quinteros y la Dra. Leonor Pérez Martínez, por su disposición y atenta corrección del manuscrito.

Me gustaría agradecer también a las personas que facilitaron la elaboración de esta tesis con su apoyo técnico: Alejandra Valdez Lara, Fernanda Cornejo Granados y al Biol. Filiberto Sánchez López, técnico de nuestro laboratorio, por su colaboración en la estandarización de la extracción de ARN y la elaboración de librerías simultáneas. Igualmente agradezco a Gloria Tanahiry Vazquez Castro y al Dr. Ricardo Alfredo Grande, ambos de la Unidad de Secuenciación Masiva y Bioinformática del Instituto de Biotecnología, por su apoyo en validación de la calidad del ARN y el desarrollo experimental para la secuenciación de las muestras. Agradezco también a Alejandro Mendoza-Vargas y la Unidad de Secuenciación Masiva del Instituto Nacional de Medicina Genómica por su soporte técnico para la secuenciación masiva. Al técnico de la Unidad de Cómputo del Instituto de Biotecnología M. en T. I. Juan Manuel Hurtado Ramírez por su apoyo en la administración del servidor y al Dr. Gamaliel López Leal por su consejo en el procedimiento bioinformático. Al Lic. J. Antonio Bolanos Guillen y a Gloria Villa Herrera de la Unidad de Docencia y Formación de Recursos Humanos por facilitar los múltiples trámites durante la maestría y finalmente los de titulación.

Por último agradezco al Concejo Nacional de Ciencia y Tecnología (CONACYT) por otorgarme una beca de maestría para dedicarme de tiempo completo al posgrado (número de beca 450940), así como su financiamiento para la elaboración de este proyecto (No. SALUD-2014-C01-234188). De igual forma al Programa de Apoyo a los Estudios del Posgrado (PAEP) por otorgarme financia-

miento para asistir al XVII Congreso Nacional de Biotecnología y Bioingeniería a presentar mis resultados.

Resumen

El microbioma intestinal tiene un papel crucial en la digestión, nutrición, resistencia a la colonización de patógenos, desarrollo del sistema inmune y comunicación con el sistema nervioso en el ser humano; aunado a esto, se han relacionado cambios en la microbiota intestinal (disbiosis) con patologías como la obesidad. Dicha enfermedad actualmente se considera la epidemia del siglo en países occidentales, donde México se encuentra en los primeros lugares en prevalencia en adultos e infantes. La obesidad es considerada como el primer factor para el desarrollo de Síndrome Metabólico (SM; que incluye también hiper-trigliceridemia, bajos niveles de HDL, hiperglucemia e hipertensión), que puede desembocar en patologías crónicas como *diabetes mellitus* de tipo II, enfermedades cardiovasculares, cáncer gastrointestinal e hígado graso no alcohólico. En este trabajo se caracterizó la microbiota intestinal de dos grupos de infantes mexicanos, el primero incluye a niños con obesidad metabólicamente sanos y el segundo a niños con obesidad y SM, esto con el fin de conocer cuales son las diferencias de la actividad funcional entre los dos grupos y dilucidar la aportación de la microbiota en el desarrollo del síndrome metabólico en individuos que ya padecen obesidad. Nuestros resultados demuestran que existen diferencias significativas entre la diversidad de especies y sus funciones entre los grupos comparados, con una riqueza mayor en los infantes con Obesidad pero con una diversidad más alta en los que presentan Obesidad y SM. Las funciones sobresalientes para los individuos con Obesidad estuvieron relacionadas con el control poblacional de la microbiota (biosíntesis de antibióticos, replicación de virus y regulación de la esporulación). Mientras que para el grupo de individuos con Obesidad y SM presentaron sobreexpresión de funciones relacionadas con producción y conversión de energía, biosíntesis y transporte de lipopolisacáridos, biosíntesis de metabolitos secundarios y metabolismo de polisacáridos; dichas funciones están relacionadas directamente con la degradación y acumulación de energía proveniente de la dieta lo cual está relacionado con el padecimiento del SM.

Índice general

	I
Índice de figuras	IX
Índice de tablas	XII
1. Marco teórico	3
1.1. Microbioma humano	3
1.2. Microbioma intestinal	4
1.2.1. Disbiosis y su relación con patologías	5
1.3. Obesidad y Síndrome Metabólico	5
1.4. Microbiota intestinal y obesidad	7
1.4.1. Disbiosis y obesidad	7
1.4.2. Herramientas para su análisis	8
1.4.3. Diversidad de especies, potencial funcional y actividad funcional	10
1.4.4. Metatranscriptómica y análisis de expresión diferencial	11
1.4.5. Metatranscriptómica humano	12
2. Metodología	16
2.1. Análisis bioinformático	18
2.1.1. Pretratamiento	18
2.1.2. Asignación de taxonomía	20
3. Análisis de Resultados	21
3.1. Librerías de RNA-seq	21
3.2. Análisis bioinformático	22
3.2.1. Pretratamiento	22
3.2.2. Asignación taxonómica con kraken2	23
3.2.3. Anotación funcional con HUMANN2	37
3.2.3.1. Clasificación funcional con COG	38

3.2.3.2.	Clasificación funcional con KEGG	41
3.2.3.3.	Clasificación funcional con GO	44
4.	Discusión	49
5.	Conclusiones y Perspectivas	52
A.	Anexos	54
A.1.	Procedimientos de biología molecular	54
A.1.1.	Extracción de ARN de muestras fecales	54
A.1.2.	Purificación de ARN	55
A.1.3.	Generación de librerías simultáneas	56
A.1.4.	Remoción de ARNr de bacteria	58
A.2.	Código	63
	Bibliografía	64

Índice de figuras

1.1.	El microbioma humano es diferente en cada parte del cuerpo (tomada y modificada de (17)).	3
1.2.	Factores que impactan en la microbiota intestinal (parte superior) y funciones que lleva a cabo (parte inferior).	4
1.3.	Microbioma intestinal y su relación con patologías (modificada de (20)).	5
1.4.	Prevalencia de sobrepeso y obesidad infantil a nivel mundial separada por sexo (gráficos tomados de www.worldobesity.org , visitada en enero del 2019).	6
1.5.	Gráfica que muestra la prevalencia de sobrepeso y obesidad de niños y niñas en México. Tomada y modificada de ENSANUT, 2016.	7
1.6.	Trasplante de microbiota fecal de paciente obesa que provoca el incremento de adiposidad en ratón recipiente (tomada y modificada de (22)).	8
1.7.	Estudio del microbioma (modificado de (1)).	9
1.8.	Diagrama de flujo que muestra los pasos generales para hacer un análisis de <i>RNA-seq</i> . Modificado de (15).	12
2.1.	Diagrama de la metodología resumida.	17
2.2.	Datos de las muestras utilizadas.	17
2.3.	Diagrama de flujo del pretratamiento.	19
2.4.	Diagrama de flujo de la anotación funcional (HUMAN2) y asignación taxonómica (kraken2).	20
3.1.	Bioanalizador de las muestras utilizadas.	21
3.2.	Resultados del pretratamiento.	22
3.3.	Diversidad alfa con la diversidad taxonómica con las muestras agrupadas en Obeso (O) u Obeso con Síndrome Metabólico (OSM); a) OTUs observados, b) Índice Chao1, c) Índice Shannon y d) Índice Simpson (****: Mann-Whitney $p < 0.005$).	24

3.4. Diagramas de Venn con los OTUs compartidas para cada grupo (a y b) y lo compartido entre grupos (c); Obeso (O), Obeso con Síndrome Metabólico (OSM).	25
3.5. Análisis de clasificación no supervisada con el coeficiente de correlación de Pearson sobre los <i>core</i> de OTUs de cada grupo (O=266 y OSM=286).	26
3.6. Clasificación taxonómica a nivel phylum, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).	27
3.7. LEfSe a nivel phylum.	28
3.8. Clasificación taxonómica a nivel clase, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).	28
3.9. LEfSe a nivel clase.	29
3.10. Clasificación taxonómica a nivel orden, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).	30
3.11. LEfSe a nivel orden.	30
3.12. Clasificación taxonómica a nivel familia, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).	31
3.13. LEfSe a nivel familia.	32
3.14. Clasificación taxonómica a nivel género, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).	32
3.15. LEfSe a nivel género.	34
3.16. Clasificación taxonómica a nivel especie, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).	35
3.17. LEfSe a nivel especie.	36
3.18. Diversidad alfa con la clasificación funcional. Las muestras están agrupadas en Obeso (O) u Obeso con Síndrome Metabólico (OSM); a) OTUs observados, b) Índice Chao1, c) Índice Shannon y d) Índice Simpson (****: Mann-Whitney $p > 0.005$).	37
3.19. <i>Core</i> de funciones compartidas en los grupos (a y b) y entre los grupos (c) con la clasificación de COG.	38
3.20. Clasificación funcional con la base de datos COG por muestra (a) y promediadas por grupo (b).	39
3.21. Análisis de expresión diferencial de las funciones con la base de datos COG con DESeq2 (a y b) y EdgeR (c y d); en rojo se muestran las funciones sobreexpresadas.	40
3.22. Clasificación funcional con la base de datos COG de las funciones con expresión diferencial.	41
3.23. <i>Core</i> de funciones compartidas en los grupos (a y b) y entre los grupos (c) con la clasificación de KEGG.	42
3.24. Clasificación funcional con la base de datos KEGG por muestra (a) y promediadas por grupo (b).	42

3.25. Análisis de expresión diferencial de las funciones con la base de datos KEGG con DESeq2 (a y b) y EdgeR (c y d); en rojo se muestran las funciones sobreexpresadas.	43
3.26. Clasificación funcional con la base de datos KEGG de las funciones con expresión diferencial.	44
3.27. <i>Core</i> de funciones compartidas en los grupos (a y b) y entre los grupos (c) con la clasificación de GO.	45
3.28. Análisis de expresión diferencial de las funciones con la base de datos KEGG con DESeq2 (a y b) y EdgeR (c y d); en rojo se muestran las funciones sobreexpresadas.	45
3.29. Clasificación con las funciones sobreexpresadas agrupadas en GO Biological Process (a=O, b=OSM).	46
3.30. Clasificación con las funciones sobreexpresadas agrupadas en GO Molecular Function (a=O, b=OSM).	47
3.31. Clasificación con las funciones sobreexpresadas en GO Biological Process (a=O, b=OSM).	48

Índice de tablas

3.1. Resultados de secuenciación por muestra.	22
3.2. Resultados de asignación con kraken2 con base de datos completa RefSeq.	23
3.3. Resultados de asignación con HUMANN2 con base de datos UniRef90.	38

Introducción

La microbiota es el conjunto de microorganismos (bacterias, arqueas, hongos y protozoarios) que habitan en un nicho determinado, incluyendo el cuerpo humano (1); junto con la microbiota oral, la microbiota intestinal es una de las comunidades microbianas más caracterizadas debido a su importancia, ya que tiene un papel crucial en la digestión de los alimentos, nutrición del hospedero, resistencia ante la invasión de patógenos, participación en el desarrollo del sistema inmune y una comunicación constante con el sistema nervioso (2). Se ha demostrado que promueve la salud o enfermedad en el hospedero (3) y en fechas recientes, con el avance de la investigación metagenómica y abaratamiento de las tecnologías de secuenciación masiva, se ha comenzado a caracterizar a la microbiota relacionada con la salud y diversas patologías, con el fin de poder utilizar a la microbiota con fines terapéuticos en múltiples hospederos incluyendo el ser humano (4). Actualmente la mayoría de los estudios se enfocan solamente en determinar la diversidad taxonómica en una condición específica, dejando de lado la caracterización de los genes transcritos y su regulación metatranscripcional, lo cual es una limitante para el conocimiento a profundidad de las funciones que lleva a cabo la microbiota intestinal y su relación con diversas enfermedades (5)(6).

Una de las patologías con las que se ha relacionado un desbalance en la composición microbiana (disbiosis) del intestino es la obesidad (7). Esta condición ha alcanzado proporciones epidémicas en las culturas occidentales y con ella enfermedades asociadas como *diabetes mellitus* de tipo II y enfermedades cardiovasculares (4). Uno de los grupos con una mayor prevalencia de obesidad en México es el infantil (edades de 5 a 11 años), lo que lo posiciona entre los primeros lugares a nivel mundial en dicho padecimiento (8). La obesidad es el factor principal del síndrome metabólico (SM), que incluye cinco criterios: circunferencia de cintura excesiva, hipertrigliceridemia, bajos niveles de HDL, hiperglucemia e hipertensión; se considera que una persona padece este síndrome cuando cumple con tres o más de estos criterios (9)(10). Para su uso pediátrico, se han extrapolado las mediciones de dichos criterios (9) y se estima que SM puede estar presente ~60 % de la población infantil mexicana (10). El SM es el primer factor de riesgo para el

desarrollo de enfermedades cardiovasculares, hígado graso no alcohólico y *diabetes mellitus* de tipo II, por lo que el estudio en etapas tempranas en población infantil que padecen de obesidad y SM es de gran importancia para formular programas de salud preventivos así como de tratamiento de dicho padecimiento (9).

Las primeras investigaciones para identificar a la microbiota intestinal en personas con obesidad muestran que hay un incremento en la proporción de Firmicutes y disminución de Bacteroidetes cuando se compara con individuos sanos (7)(11); sin embargo, investigaciones posteriores determinaron que dependiendo del diseño experimental utilizado se pueden encontrar o no cambios en la diversidad de especies entre individuos delgados y con obesidad (12). Por lo anterior es necesario considerar la expresión genética (metatranscriptómica), de proteínas generadas (metaproteómica) y/o de vías metabólicas compartidas (metabolómica) para establecer las funciones que lleva a cabo el microbioma intestinal y así como su efecto en diversas patologías, como la obesidad (5)(13). Como se mencionó, uno de los análisis que permite conocer la función del microbioma en una condición determinada es el metatranscriptómico, éste permite la identificación de los transcritos expresados en el microbioma, abordando el punto de vista de actividad funcional y brindando información complementaria con la proporcionada por la diversidad de especies y su potencial genómico, ambos provenientes del estudio metagenómico (5)(14). Para poder conocer los transcritos generados en el microbioma es necesario utilizar técnicas de biología molecular, con el fin de aislar y purificar el ARN, la preparación de librerías, así como su secuenciación y posteriormente, el uso de herramientas bioinformáticas para el análisis de las secuencias obtenidas (15).

Las preguntas de este proyecto de investigación son: ¿cuáles son las diferencias y similitudes entre los perfiles metatranscriptómicos de la microbiota intestinal entre un grupo de niños con obesidad y otro con obesidad aunado a SM? y ¿dichas diferencias tienen impacto en el desarrollo del SM? Para poder contestar a las premisas anteriores se realizó el análisis metatranscriptómico de muestras fecales de dos grupos de infantes mexicanos (de entre 5 y 11 años) con obesidad y con obesidad junto con SM. Con dicho análisis se buscará conocer los perfiles metatranscripcionales en cada uno de los grupos, así como sus similitudes y diferencias.

Conocer los factores, entre ellos el metatranscriptoma, que impactan en la composición del microbioma de obesos y con SM y compararlos con los de individuos sanos, es crítico para elucidar el papel del microbioma en los padecimiento y para el desarrollo de técnicas terapéuticas con objetivo en el microbioma.

1.1. Microbioma humano

A la colección de microorganismos que viven en asociación con el hospedero en diferentes regiones del cuerpo del humano se le conoce como microbioma humano (16)(17). Aquí se incluyen a las bacterias, arqueas, hongos, protozoarios y virus (1)(16), que varían dependiendo de la zona del cuerpo donde se encuentran (boca, nariz, piel, trato gastrointestinal y tracto urogenital) y condición del hospedero (edad y estado de salud) (Figura 1.1)(18).

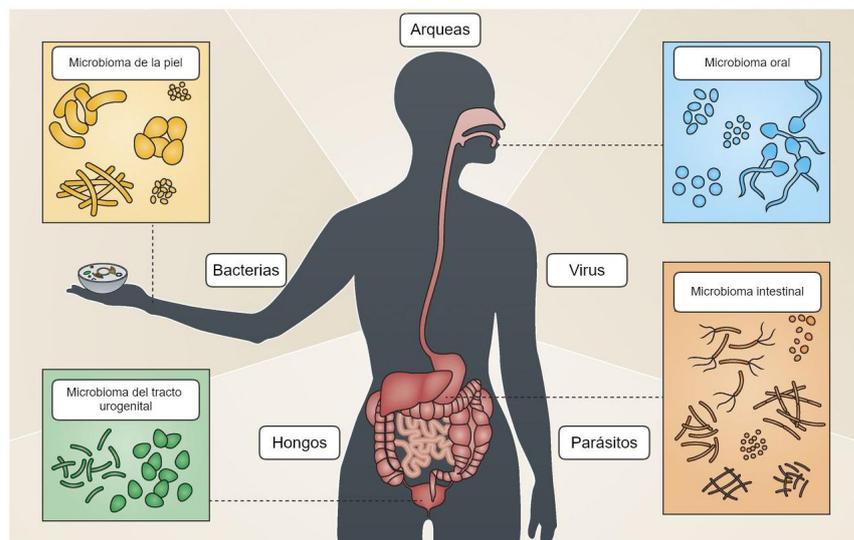


Figura 1.1: El microbioma humano es diferente en cada parte del cuerpo (tomada y modificada de (17)).

1.2. Microbioma intestinal

El microbioma intestinal tiene funciones de digestión del bolo alimenticio, nutrición del hospedero, protección ante agentes patógenos, desarrollo del sistema inmune e incluso comunicación con el sistema nervioso (Figura 1.2)(2)(3)(19)(16). La mayoría de estos microorganismos no son patógenos para nosotros e incluso están relacionados con la salud del hospedero (2), por lo cual, en fechas recientes existe un interés especial en conocer la función que lleva a cabo la microbiota en el intestino y saber cuáles son sus efectos en la salud humana, así como su relación con enfermedades (2)(16).

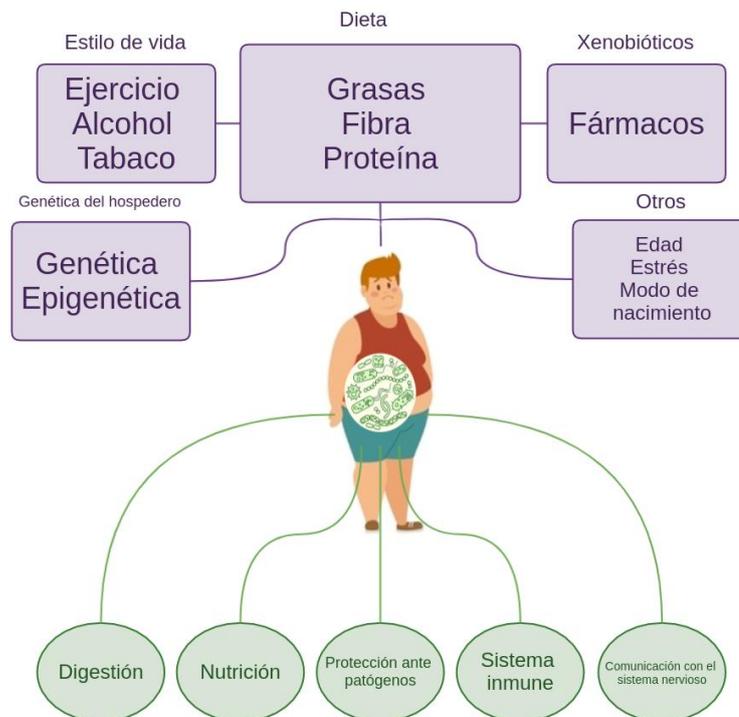


Figura 1.2: Factores que impactan en la microbiota intestinal (parte superior) y funciones que lleva a cabo (parte inferior).

La microbiota intestinal tiene diversas funciones que repercuten en la salud del hospedero como producción de metabolitos, hormonas y vitaminas, así como digestión de nutrientes que el hospedero no puede procesar (19). Es por esto que una modificación en su composición, y con ello su actividad, puede repercutir en la salud del ser humano.

1.2.1. Disbiosis y su relación con patologías

Como se menciona anteriormente la microbiota intestinal se ha relacionado con diversas enfermedades como: enfermedad inflamatoria intestinal, cáncer de colon, síndrome del intestino irritable, úlceras gástricas, hígado graso no alcohólico, así como obesidad y Síndrome metabólico (SM; Figura 1.3)(19).

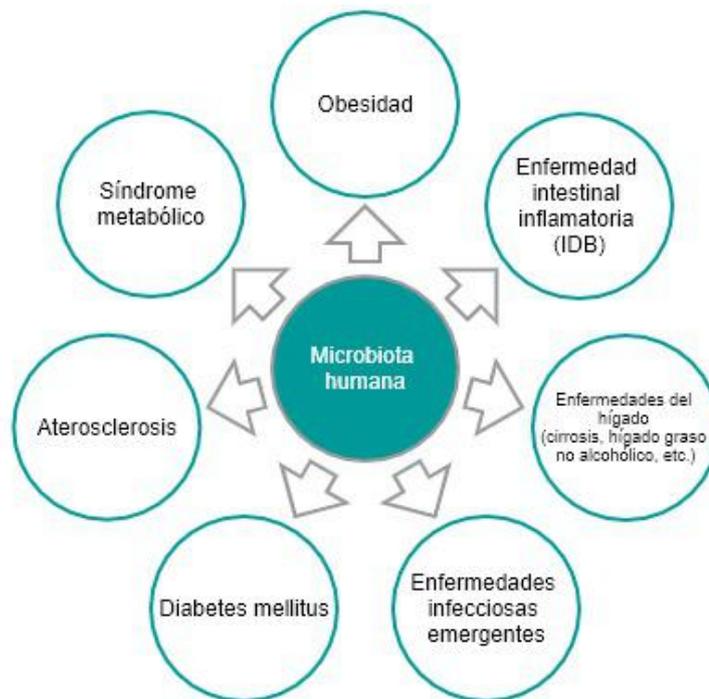


Figura 1.3: Microbioma intestinal y su relación con patologías (modificada de (20)).

Es importante mencionar que son pocas las investigaciones que han relacionado a una patología directamente con un género o especie presente en la microbiota de un individuo enfermo (13), por lo cual actualmente se estudia el efecto global de la microbiota y su relación con la enfermedad (5)

1.3. Obesidad y Síndrome Metabólico

La obesidad ha alcanzado proporciones epidémicas en la cultura occidental (4), incluyendo a México (Figura 1.4). Actualmente, nuestro país es el segundo

lugar a nivel mundial en obesidad en edad adulta, solamente debajo de Estados Unidos y en población infantil se clasifica entre los primeros lugares (21).



Figura 1.4: Prevalencia de sobrepeso y obesidad infantil a nivel mundial separada por sexo (gráficos tomados de www.worldobesity.org, visitada en enero del 2019).

Tres de cada diez infantes en nuestro país tiene sobrepeso u obesidad en México (Figura 1.5). En el último reporte del 2016 de la ENSANUT MC se observa un decremento no significativo respecto a la medición realizada en el 2012 (de 34.2 a 33.2%), por lo cual la obesidad infantil sigue siendo un problema de salud en nuestro país (Figura 1.5)(8).

La obesidad es primer factor causal del síndrome metabólico que incluye cinco complicaciones metabólicas, las cuales han sido extrapoladas para ser utilizadas en infantes por De Ferranti y colaboradores (9):

1. Obesidad central (circunferencia de cintura): $>$ al percentil 75 por edad y sexo
2. Bajo nivel de HDL (lipoproteínas de alta densidad, del inglés *High Density Lipoprotein*): < 1.3 mmol/L
3. Alto nivel de glucosa en sangre en ayunas: ≥ 6.1 mmol/L
4. Hipertrigliceridemia: ≥ 1.1 mmol/L
5. Hipertensión: $>$ al percentil 90 por edad, sexo y altura.

Aunque no se conoce la prevalencia del SM en la población infantil con obesidad, se estima que está presente en el 60% de la población infantil (10).

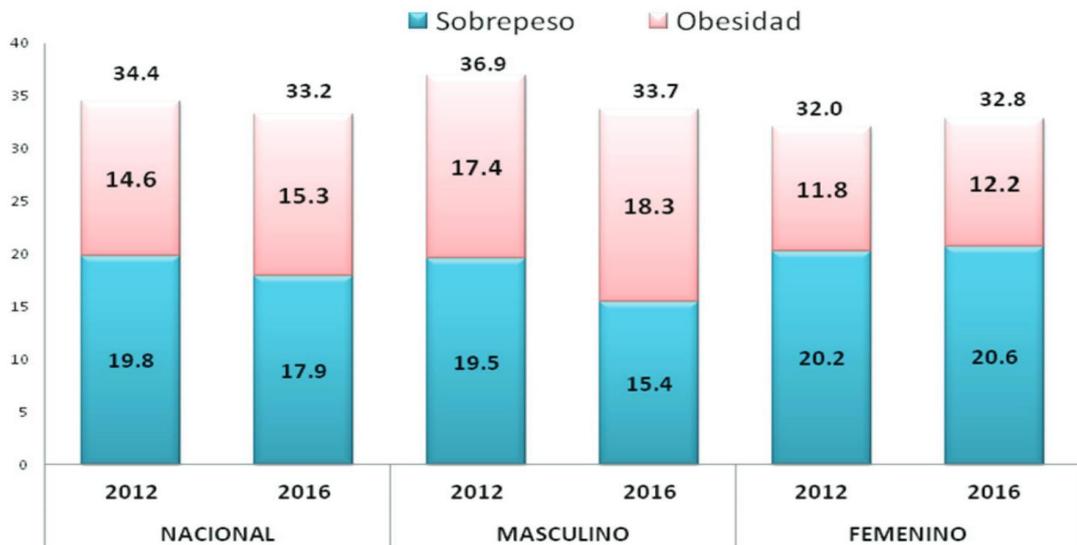


Figura 1.5: Gráfica que muestra la prevalencia de sobrepeso y obesidad de niños y niñas en México. Tomada y modificada de ENSANUT, 2016.

1.4. Microbiota intestinal y obesidad

1.4.1. Disbiosis y obesidad

Los trabajos más relevantes que relacionan a la microbiota intestinal y la obesidad son los realizados por el grupo del Dr. Jeffrey Gordon (Washington University, St. Louis, Estados Unidos). Dichas investigaciones han encontrado que hay un cambio de la diversidad entre los phyla más abundantes en el intestino, que juntos suman aproximadamente el 90 %, Firmicutes y Bacteroidetes. En un individuo sano se tiene una proporción del 20 al 40 % de Bacteroidetes y del 60 al 80 % de Firmicutes, al contrario con un individuo obeso se observa un aumento en los Firmicutes ($\approx 50\%$) y una disminución de Bacteroidetes ($\approx 50\%$). En esta investigación realizada por Turnbaugh y colaboradores se analizaron además los genes enriquecidos para la obesidad, donde 75 % fueron de Actinobacteria (comparados con 0 % de los genes enriquecidos para los sanos; el otro 25 % fue de Firmicutes) donde el 42 % de los genes enriquecidos para los delgados fueron de Bacteroidetes (comparados con el 0 % de los genes enriquecidos para los obesos) (11).

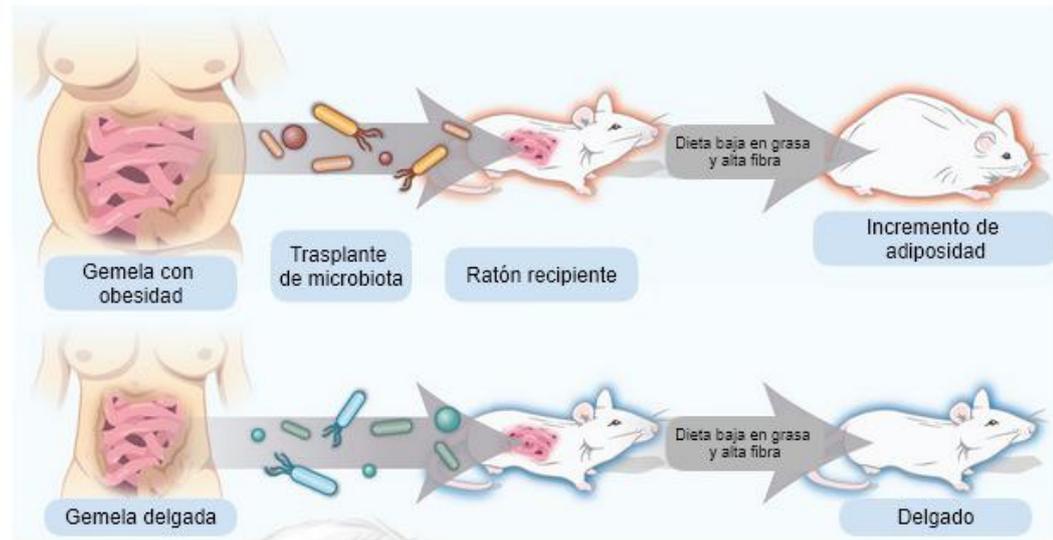


Figura 1.6: Trasplante de microbiota fecal de paciente obesa que provoca el incremento de adiposidad en ratón recipiente (tomada y modificada de (22)).

La primera explicación que se le daba al cambio entre estos phyla y la incidencia de la obesidad es que los Firmicutes contaban con mayor número de enzimas que participan en la degradación de lípidos y carbohidratos complejos y no digeribles (7)(11). Sin embargo, no solamente la cosecha de energía por parte de los microorganismos aumentados en proporción podía explicar el aumento de peso en el hospedero (4). Por ello se propuso que la microbiota intestinal relacionada con la obesidad tenía efectos de señalización con el hospedero para la acumulación tejido adiposo, así como de disminución de la saciedad al comunicarse con el sistema nervioso (18), aunque estos mecanismos no han sido comprendidos por completo.

1.4.2. Herramientas para su análisis

Actualmente se utilizan múltiples métodos independientes de cultivo para investigar al microbioma y la herramienta más óptima para analizarlo depende de la pregunta que se quiera resolver (Figura 1.7). Dichos análisis pueden enfocarse en conocer el ADN (metagenómica), el ARN (metatranscriptómica), las proteínas (metaproteómica) así como los metabolitos (metabolómica) que provienen del microbioma (23). Es importante mencionar que cada una de estas herramientas brinda información diferente del microbioma por lo cual cada una de ellas es

importante para conocer a fondo el microbioma (1)(23).

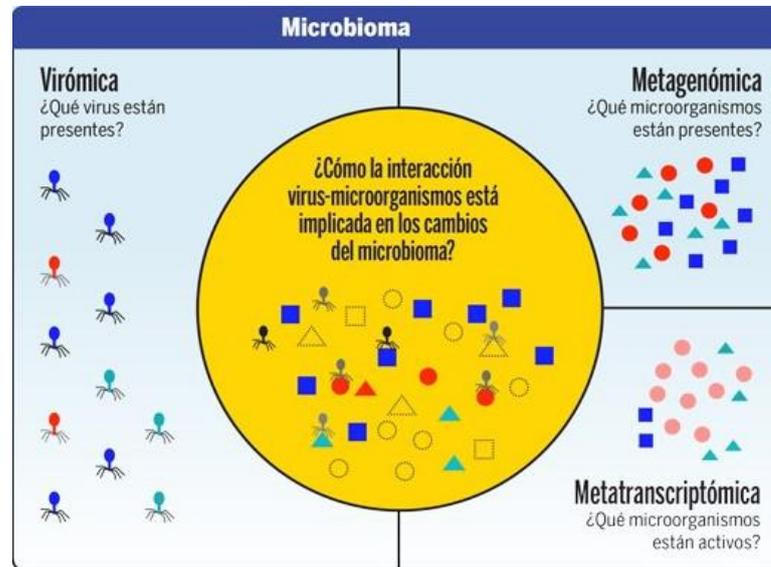


Figura 1.7: Estudio del microbioma (modificado de (1)).

Una de las herramientas más utilizadas para caracterizar a la diversidad de la microbiota intestinal es la secuenciación de genes marcadores, entre los cuales se encuentran el ARN ribosomal 16S (presente en bacterias y arqueas) y el ARN ribosomal 18S (de organismos eucarióticos) (1)(23). Sin embargo la secuenciación de genes marcadores generalmente no permite diferenciar entre géneros o especies, ya que algunas pueden llegar a compartir hasta el 99.5 % de la secuencia del ARN ribosomal 16S (23), además de que la diversidad se ve sesgada por el número de copias de este gen que pueden existir en los diferentes grupos de bacterias (1).

La metagenómica nos permite determinar el potencial funcional codificado en el microbioma (1)(23). A grandes rasgos este análisis se basa en la fragmentación aleatoria del ADN del microbioma, posteriormente su secuenciación y por último la identificación de los genes encontrados (1). Dicha identificación se puede realizar a partir de mapeo a genomas de referencia, de los cuales existen múltiples bases de datos libres, así como el ensamblado *de novo*, donde el total de lecturas son utilizadas para ensamblar genomas que posteriormente permiten la anotación funcional cuando se comparan con bases de datos de referencia (1).

El análisis metatranscriptómico nos permite describir la expresión de genes en el microbioma en un tiempo y condición determinada, esto es, la actividad funcional (1)(5)(24)(23). La metatranscriptómica nos puede revelar información acerca de cómo están interactuando los organismos con el hospedero, tanto en la salud como cuando se padece alguna enfermedad, y con esto conocer la función

del microbioma en dichas condiciones (23). Pocas investigaciones han complementado la información brindada por el análisis del gen ARNr 16S o el metagenómico con datos de metatranscriptómico para conocer la función del microbioma intestinal cuando se tiene alguna patología. Este análisis es el que se utiliza en esta investigación, por lo cual se explicará con más detalle posteriormente.

La metaproteómica estudia las proteínas que son producidas por la microbiota así como las que son secretadas por el hospedero durante respuestas inflamatorias (23). Cabe resaltar que se han encontrado proteínas únicas que no habían sido predichas por metatranscriptómica (23). Por último la metabolómica estudia los productos y secreciones de los microorganismos y su función dentro del microbioma. Brinda información que relaciona la interacción microbioma-hospedero y es de importancia porque muchas patologías se pueden relacionar más con la producción de metabolitos o con proteínas no funcionales o defectuosas que son resultado de la interacción del hospedero con su microbiota, y que va más allá de la variación de la diversidad de las comunidades microbianas(23).

1.4.3. Diversidad de especies, potencial funcional y actividad funcional

Como se menciona anteriormente el perfilamiento del gen ARNr 16S nos permite conocer la diversidad de especies, al igual que la metagenómica, y que además nos brinda información de potencial funcional de la microbiota, esto es: cuáles son los genes que se encuentran en los genomas de los microorganismos del microbioma y cuáles son sus funciones (1). Sin embargo, pocas investigaciones relacionan directamente el cambio de diversidad de especies con algún efecto en la salud del hospedero (13) y trabajos previos han demostrado que la diversidad de especies puede cambiar sin que se modifique el potencial funcional que puede existir en el microbioma (5)(13). Ahora bien, dado que el potencial funcional no muestra cuáles son los genes que se están expresando por parte de los microorganismos, es por lo que se requiere de la metatranscriptómica para conocer la actividad funcional en un tiempo y condición determinada (5)(6). Actualmente se propone relacionar la información de diversidad de especies, potencial y actividad funcional con el fin de determinar el papel que lleva a cabo la microbiota, conocer cuáles son los microorganismos relacionados y con esto establecer terapias para prevenir o controlar enfermedades relacionadas con la microbiota humana (1)(5)(13)(23).

1.4.4. Metatranscriptómica y análisis de expresión diferencial

Para el análisis de metatranscriptómica es necesario realizar primero la extracción de ARN, la depleción de ARNr, la eliminación de ADN utilizando ADNasas y la síntesis del ADNc, el cual posteriormente es secuenciado en equipos de secuenciación masiva; a pesar de que la secuenciación ha disminuido sus costos, realizar análisis de *RNA-seq* sigue teniendo un precio elevado. Una de las metodologías que disminuyen el tiempo y el costo de la elaboración de librerías de RNA-seq es la propuesta por Shishkin y colaboradores en el 2015 (24). Esta metodología propone la elaboración de librerías simultáneas utilizando *tags* de ARN diferentes para cada una de las librerías y que nos permitan juntarlas reacciones posteriores. Este *tag* también nos permitirá identificar a las muestras después de la secuenciación y así separar cada librería para su análisis individual (24).

El análisis de *RNA-seq* se puede llevar a cabo de diferentes maneras (Figura 1.8)(15). En los datos de metatranscriptómica no se tiene un genoma como referencia, ya que se están considerando los transcritos de múltiples organismos. Sin embargo se pueden buscar los transcritos en bases de datos que contienen múltiples genomas y con esto darle función a cada transcrito (1). Otra aproximación también es la de formar un metatranscriptoma *de novo*, en el cual con los fragmentos obtenidos de la secuenciación se ensamblan metatranscriptomas (usando SOAPdenovo, ABySS, Velvet-Oases y Trinity) y posteriormente se realiza la anotación de los transcritos utilizando bases de referencia y el análisis de expresión diferencial (1)(15).

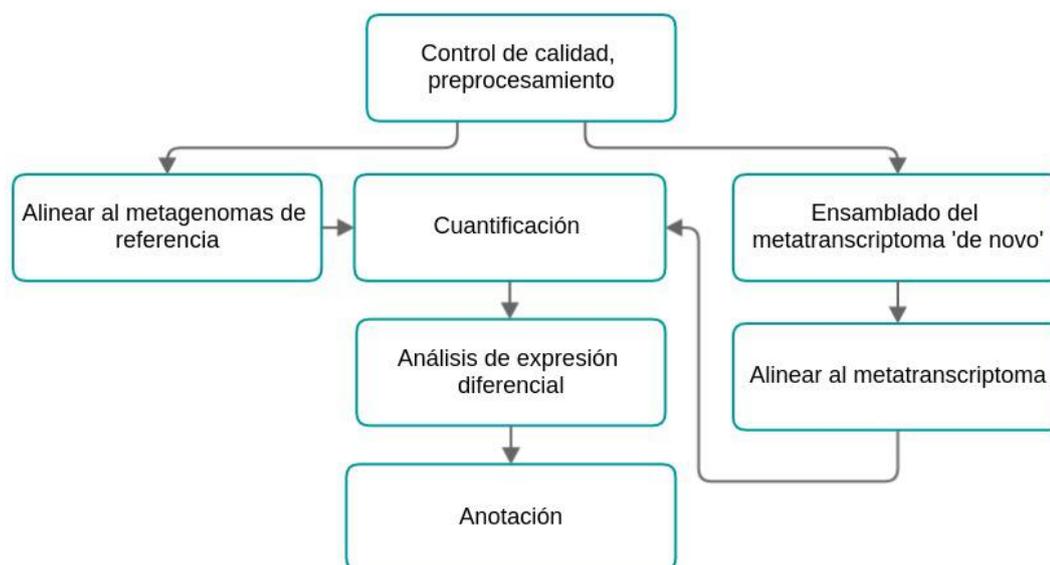


Figura 1.8: Diagrama de flujo que muestra los pasos generales para hacer un análisis de *RNA-seq*. Modificado de (15).

1.4.5. Metatranscriptómica humano

Uno de los primeros estudios en los cuales se buscó determinar las diferencias entre la información del metatranscriptoma y el metagenoma humano fue el realizado por Franzosa y colaboradores en el 2014, donde se demuestra que los perfiles metatranscripcionales son menos variables que los perfiles taxonómicos, pero más variables que las funciones metagenómicas; esto se explica porque no todos los genes tienen una expresión proporcional a su número de copias, ya que hay genes que se sobreexpresan y subexpresan con respecto a su número de copias en el ADN (5).

Con el uso de *RNA-seq*, citometría de flujo y secuenciación del ARNr 16S Maurice y colaboradores en el 2013 realizaron una investigación cuyo objetivo principal era determinar el impacto de los xenobióticos en el microbioma. Ellos concluyen que el uso de xenobióticos (fármacos dirigidos al hospedero y antibióticos) no solamente altera la fisiología de los microorganismos, sino que también impacta en la comunidad microbiana y en su expresión génica. Además demuestran que no toda la microbiota intestinal está activa ya que solamente un conjunto de microorganismos cuenta con una fisiología óptima para transcribir sus genes (25).

En una investigación más reciente realizada por Abu-Ali y colaboradores en

el 2018, se consideraron muestras fecales de 308 adultos de las cuales se obtuvieron 372 metatranscriptomas y 929 metagenomas con el fin de determinar el metatranscriptoma compartido (*core*) en un humano sano. Se determinó que solo un 44 % del *core* del metagenoma se encuentra en el *core* del metatranscriptoma, lo cual sugiere que en el metagenoma podemos encontrar redundancia de genes que pueden cumplir los mismos nichos si un individuo cuenta con una comunidad bacteriana diferente. Además, en la investigación se concluye que la mayoría de las rutas metabólicas *core* son expresadas por un número limitado de microorganismos, a pesar de que dichas rutas se encuentren ampliamente distribuidas en el metagenoma de otros microorganismos, lo cual sugiere que hay especies dominantes a nivel de expresión en los individuos sanos (26).

Hipótesis

Existen diferentes perfiles metatranscriptómicos en la microbiota intestinal entre infantes con obesidad y con Síndrome Metabólico.

Objetivos

Objetivo General

Caracterizar los perfiles metatranscriptómicos del microbioma intestinal de niños obesos y obesos con Síndrome Metabólico.

Objetivos específicos

1. Estandarización de un protocolo para la extracción de ARN de muestras fecales con calidad óptima para secuenciación masiva.
2. Construcción de librerías simultáneas de *RNA-seq* para la secuenciación masiva de los transcritos.
3. Comparación el metatranscriptoma compartido entre los grupos experimentales.
4. Análisis de los perfiles metatranscripcionales diferenciales, sobre y sub-expresados entre los grupos experimentales.

Metodología

En este trabajo se consideraron muestras fecales de dos grupos de infantes: 1) obesos metabólicamente sanos y 2) obesos con SM (al menos 3 de los criterios del SM). Cada grupo incluyó a 3 individuos (Figura 2.2), cuyas muestras fecales fueron colectadas en el 2014 en colaboración con el Instituto de Medicina Genómica y se encuentran conservadas en *RNAlater* a -70°C .

Los criterios para la selección de muestras de niños obesos con complicaciones fueron:

1. Obesidad central (circunferencia de cintura): $>$ al percentil 75 por edad y sexo
2. Nivel de triglicéridos en ayuno: ≥ 1.1 mmol/L
3. Bajo nivel de HDL: < 1.3 mmol/L
4. Presión sistólica: $>$ al percentil 90 por edad, sexo y altura.

En las siguientes secciones se describen en detalle cada uno de los pasos de la metodología. El resumen de la misma se puede observar en la Figura 2.1, donde se incluyen los procedimientos de biología molecular y el análisis bioinformático de la información obtenida en la secuenciación. La primera parte incluyó la extracción de RNA de las muestras fecales, depleción de ADN y ARNr, la construcción de librerías simultáneas de RNA-seq (24) y por último la secuenciación de estas librerías (los procedimientos de biología molecular se encuentran en el Anexo A.1); ésta se realizó con Illumina con extremos pareados (*paired-end*) utilizando longitudes de 75 y 150 bases por lectura.

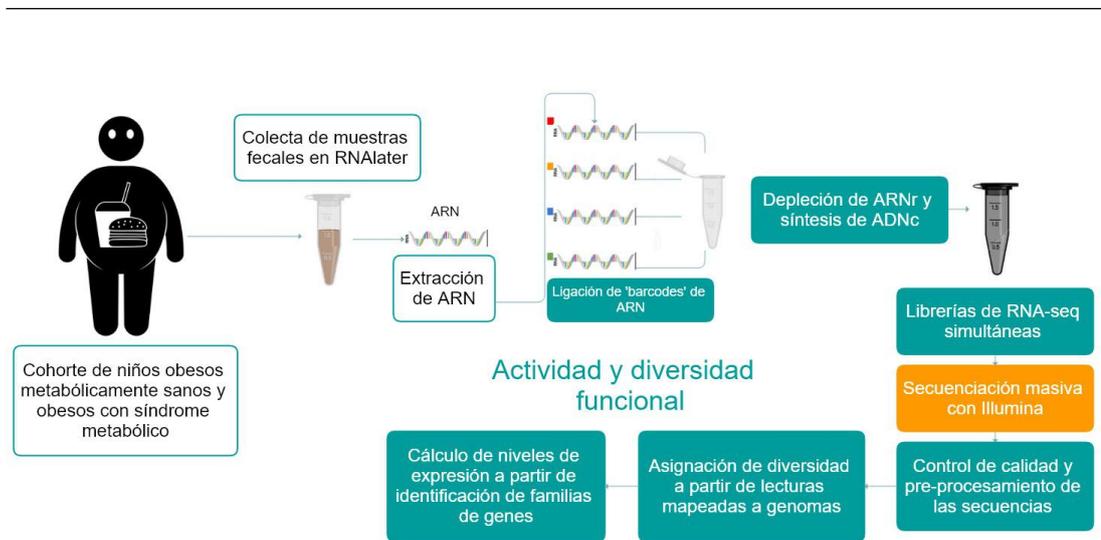


Figura 2.1: Diagrama de la metodología resumida.

Se realizó enseguida el pretratamiento de los datos obtenidos de la secuenciación que inició con la limpieza de los mismos, es decir, retirar las secuencias de baja calidad y los nucleótidos no asignados. Posteriormente, dentro de las lecturas que pasaron los filtros de calidad se reconocieron a las secuencias pertenecientes a ARNr y los genes que se alinearon al genoma humano posterior a la secuenciación. Se consideró que las lecturas resultantes del paso anterior fuesen de ARNm de la microbiota intestinal; con estas se elaboraron los perfiles metatranscripcionales por muestra (de cada individuo considerado) y por grupo de muestras (niños con obesidad y niños con obesidad y SM) para realizar las comparaciones entre los perfiles metatranscripcionales (Figura 2.2). El metadata de las muestras utilizadas se puede observar en la Figura 2.2.

Muestra	Grupo	Grupo Simplificado	Edad	Sexo	Número de complicaciones
024	Obeso	O	10	M	1 (Circunferencia de cintura > p75)
074	Obeso	O	9	M	1 (Circunferencia de cintura > p75)
090	Obeso	O	9	M	1 (Circunferencia de cintura > p75)
146	Obeso con síndrome metabólico	OSM	9	M	4 (Circunferencia de cintura > p75, TG > 100 mg/dL, HDL < 50mg/dL y TA > p90)
153	Obeso con síndrome metabólico	OSM	7	M	3 (Circunferencia de cintura > p75, TG > 100 mg/dL y HDL < 50mg/dL)
258	Obeso con síndrome metabólico	OSM	9	F	3 (Circunferencia de cintura > p75, TG > 100 mg/dL y HDL < 50mg/dL)

Figura 2.2: Datos de las muestras utilizadas.

2.1. Análisis bioinformático

2.1.1. Pretratamiento

En la Figura 2.3 se muestra la metodología para el pretratamiento de las secuencias, en los siguientes párrafos se explica en detalle cada uno de los pasos; el código utilizado para cada uno de los pasos se puede consultar en el Anexo A.2.

El primer paso que se hizo con las secuencias provenientes de librerías simultáneas fue separar a las muestras por su *barcode* de ARN, los cuáles fueron:

1. B1: TAATCA
2. B2: TACATA
3. B3: TATACT
4. B4: CCGATT

Para separarlos se utilizó *fastx_barcode_splitter* (versión: 11sep2008), el cual buscó en al inicio de la secuencia el *barcode* correspondiente para cada muestra y las separó en archivos independientes. Posteriormente se usó el *seqtk subseq* (versión: 1.3) para obtener las lecturas del archivo R2 correspondiente de cada muestra, basándose en los IDs previamente identificados. Al finalizar, se quitó el *barcode* para que no tuviese influencia en los pasos siguientes (Trimmomatic, versión: 0.36)(27).

Posteriormente se eliminaron las lecturas duplicadas (posibles efectos de PCR en la formación de las librerías) conservando un máximo permitido de 5% del total de la muestra (*cd_hit_dup*, versión: 4.6.8)(28). Se utilizaron ventanas de 4 nucleótidos con una calidad mínima de Q20 (Trimmomatic, versión: 0.36)(27). Además, se quitaron los adaptadores remanentes de la preparación de la librería para su secuenciación (Trimmomatic, versión: 0.36)(27).

Para quitar las lecturas remanentes de ARNr se utilizó *ribopicker* (versión: 0.4.3)(29) el cual consideró como referencia las bases de datos ribosomales SILVA de procariotes y eucariotes. Posteriormente a las lecturas no ribosomales se les alineó con *Hisat2* (versión: 2.1.0)(30) contra el genoma y transcriptoma humano, con el fin de conservar solamente las lecturas microbianas. Éstas fueron las que se consideraron como pretratadas y con las que se realizaron los análisis posteriores.

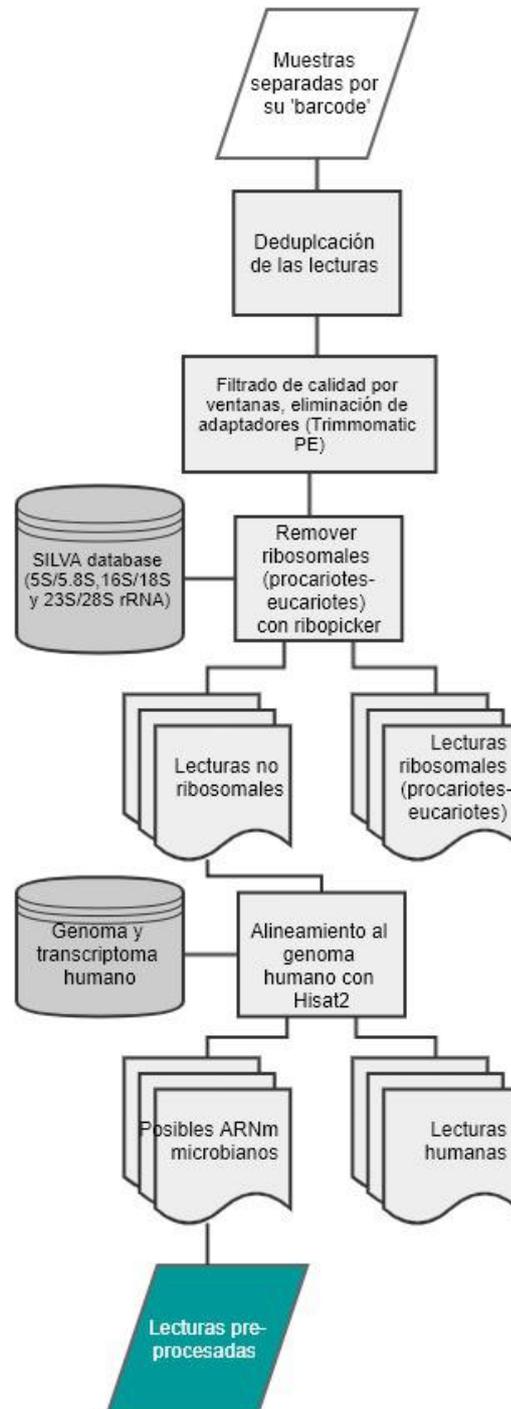


Figura 2.3: Diagrama de flujo del pretratamiento.

2.1.2. Asignación de taxonomía

Se utilizó `kraken2` (versión: 2.0.7) (31) para la asignación de taxonomía utilizando como referencia la base completa RefSeq y HUMANn2 (versión: 0.11.2)(32) para la anotación funcional con la base de datos UniRef90 (Figura 2.4); el código utilizado se puede consultar en el Anexo A.2.

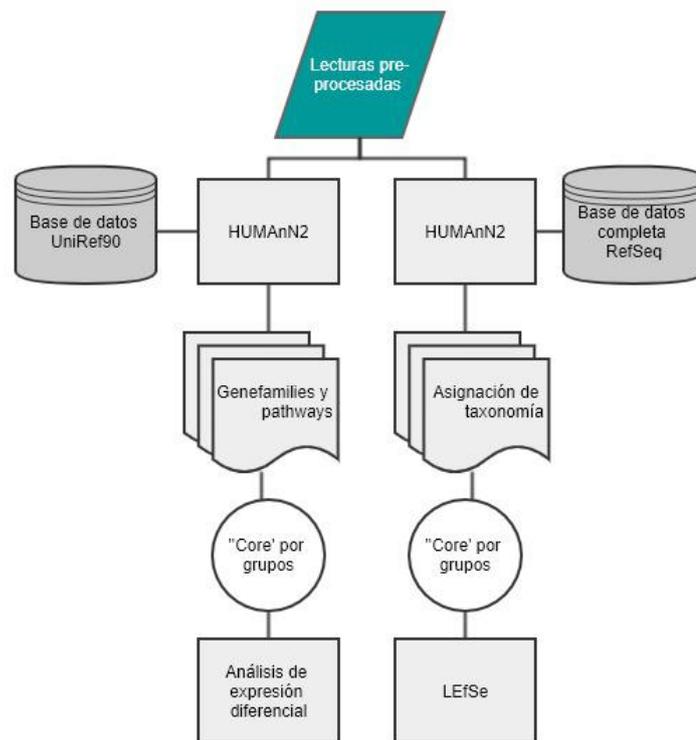


Figura 2.4: Diagrama de flujo de la anotación funcional (HUMANn2) y asignación taxonómica (kraken2).

Posteriormente se obtuvieron los grupos taxonómicos y actividades funcionales compartidas (*core* taxonomómico y funcional, respectivamente) para cada uno de los grupos, esto con el fin de eliminar la variación entre individuos. El análisis estadístico para la asignación taxonómica fue LEfSe (versión: Galaxy Version 1.0) (33). Para identificar las diferencias entre los grupos se realizó un análisis de expresión diferencial (`edgeR` (34) y `deseq2` (35)) con los resultados de anotación funcional.

Análisis de Resultados

3.1. Librerías de RNA-seq

Las muestras de ARN extraídas tuvieron un RIN (*RNA Integrity Number* de 6.22 ± 0.5 , en la Figura 3.1 se muestran los resultados de los análisis de integridad de cada una de las muestras) adecuado para la elaboración de librerías simultáneas y su posterior secuenciación. Posterior a la secuenciación masiva con Illumina se obtuvieron entre $20\,721\,151 \pm 8\,517\,844$ millones de lecturas por muestra; en la tabla 3.1 se encuentra el número de lecturas obtenido para cada una de ellas.

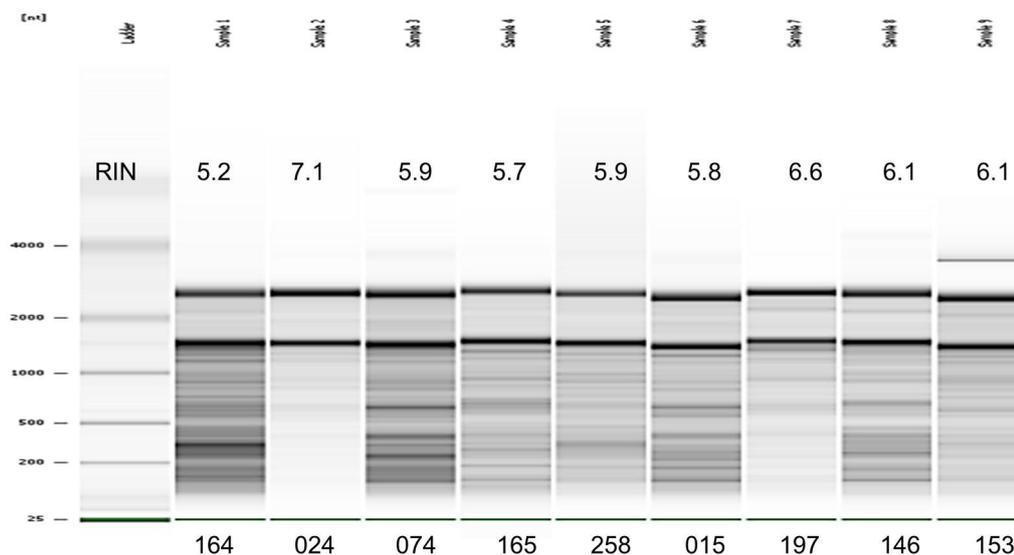


Figura 3.1: Bioanalizador de las muestras utilizadas.

Tabla 3.1: Resultados de secuenciación por muestra.

Muestra	Grupo	Grupo Simplificado	Lecturas crudas
024	Obeso	O	27 135 948
074	Obeso	O	12 032 290
090	Obeso	O	26 000 000
146	Obeso con síndrome metabólico	OSM	31 698 480
153	Obeso con síndrome metabólico	OSM	13 681 390
258	Obeso con síndrome metabólico	OSM	13 778 802

3.2. Análisis bioinformático

3.2.1. Pretratamiento

El primer paso del análisis bioinformático fue el pretratamiento donde se incluye la separación de las muestras de las librerías simultáneas por su *barcode*, filtrado de calidad, remoción de ARNr remanente y separación de lecturas que alinean al genoma y transcriptoma humano (Figura 3.2).

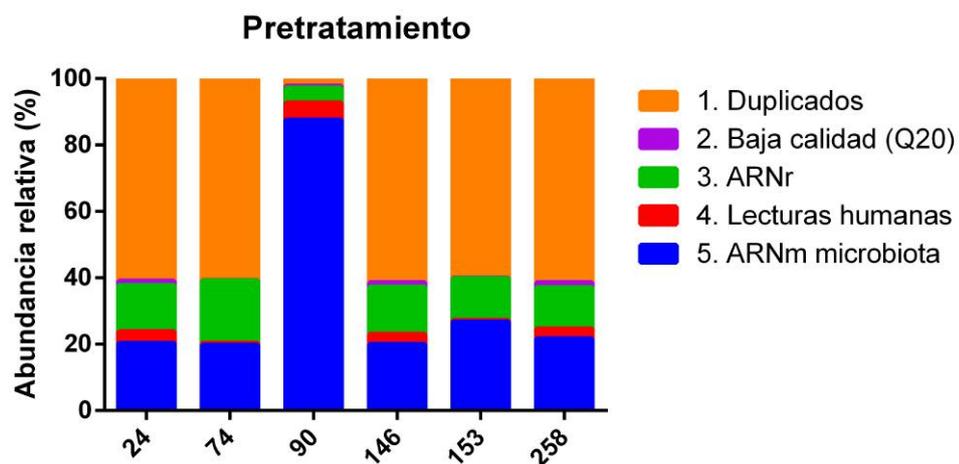


Figura 3.2: Resultados del pretratamiento.

El paso que removi6 más lecturas fue la deduplicación, la cual se realiz6 para disminuir el sesgo de sobreamplificaci6n con PCR durante la preparaci6n de la librería simultánea; para un análisis de *RNA-seq* se considera como aceptable entre un 5 y 10 %. Se puede observar que la muestra 90 es la que tiene un menor número de lecturas duplicadas, por lo cual al finalizar se cuenta con un mayor porcentaje de lecturas tentativas a ARNm del microbioma; esto se explica porque se prepar6 independientemente a las demás muestras. Al final se conservan en promedio 21.67 %, siendo la muestra 90 la que conserva mayor número de lecturas 87.56 %.

3.2.2. Asignaci6n taxon6mica con kraken2

En la tabla 3.2 se puede observar el porcentaje de asignaci6n taxon6mica con kraken2 por muestra (29.94 ± 12.62 %) utilizando como referencia la base de datos RefSeq, lo cual implica que el ~ 70 % de la informaci6n taxon6mica es desconocida.

Tabla 3.2: Resultados de asignaci6n con kraken2 con base de datos completa RefSeq.

Muestra	% asignado
24 (O)	17
74 (O)	46
90 (O)	46
146 (OSM)	42
153 (OSM)	17
258 (OSM)	28

El cálculo de la diversidad alfa a una profundidad de 315,272 lecturas por muestra se puede observar en la Figura 3.3; se encontraron diferencias significativas en las métricas consideradas (Observed OTUs, Chao1, Shannon, Simpson; Mann-Whitney $p < 0.005$). La riqueza de especies result6 ser mayor en las muestras del grupo O (índices Observed OTUs y Chao1); sin embargo, el grupo OSM cont6 con valores de Shannon y Simpson significativamente mayores, lo cual sugiere que a pesar de contar con un número menor de especies la aportaci6n de cada una de ellas es más equitativa. Lo anterior se puede explicar porque el grupo OSM cuenta con una pérdida de diversidad taxon6mica, tras la cual posiblemente no se han podido establecer grupos dominantes.

Posteriormente se obtuvieron los *core* taxon6micos por cada grupo de muestras (con el fin de obtener solamente la informaci6n taxon6mica compartida entre

los grupos y con ello delimitar lo común para cada fenotipo) donde se obtuvieron 266 y 286 especies para los grupos O y OSM, respectivamente (Figura 3.4). Se determinó también que ambos grupos compartían 210 especies (~75%).

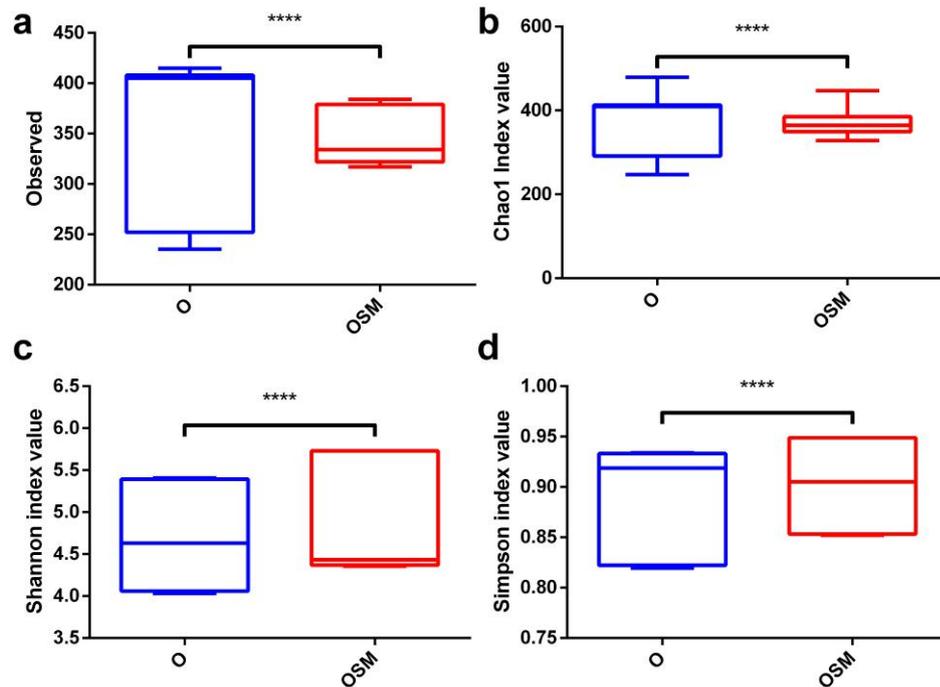


Figura 3.3: Diversidad alfa con la diversidad taxonómica con las muestras agrupadas en Obeso (O) u Obeso con Síndrome Metabólico (OSM); a) OTUs observados, b) Índice Chao1, c) Índice Shannon y d) Índice Simpson (****: Mann-Whitney $p < 0.005$).

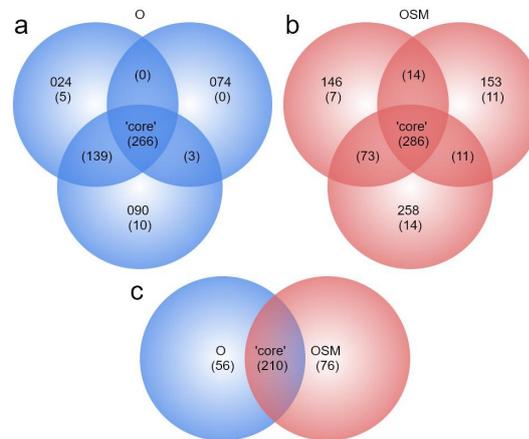


Figura 3.4: Diagramas de Venn con los OTUs compartidas para cada grupo (a y b) y lo compartido entre grupos (c); Obeso (O), Obeso con Síndrome Metabólico (OSM).

Debido a que la mayoría de los OTUs se comparten entre los grupos se hizo un análisis de clasificación no supervisada con el coeficiente de correlación de Pearson utilizando la taxonomía compartida entre los grupos (figure 3.4 módulo c) y como se puede observar en la Figura 3.5 las muestras se agruparon por fenotipo (O u OSM). Esto sugiere que a pesar de que la mayoría de los OTUs se comparten entre las muestras, lo que nos permite separar entre los grupos es su abundancia relativa.

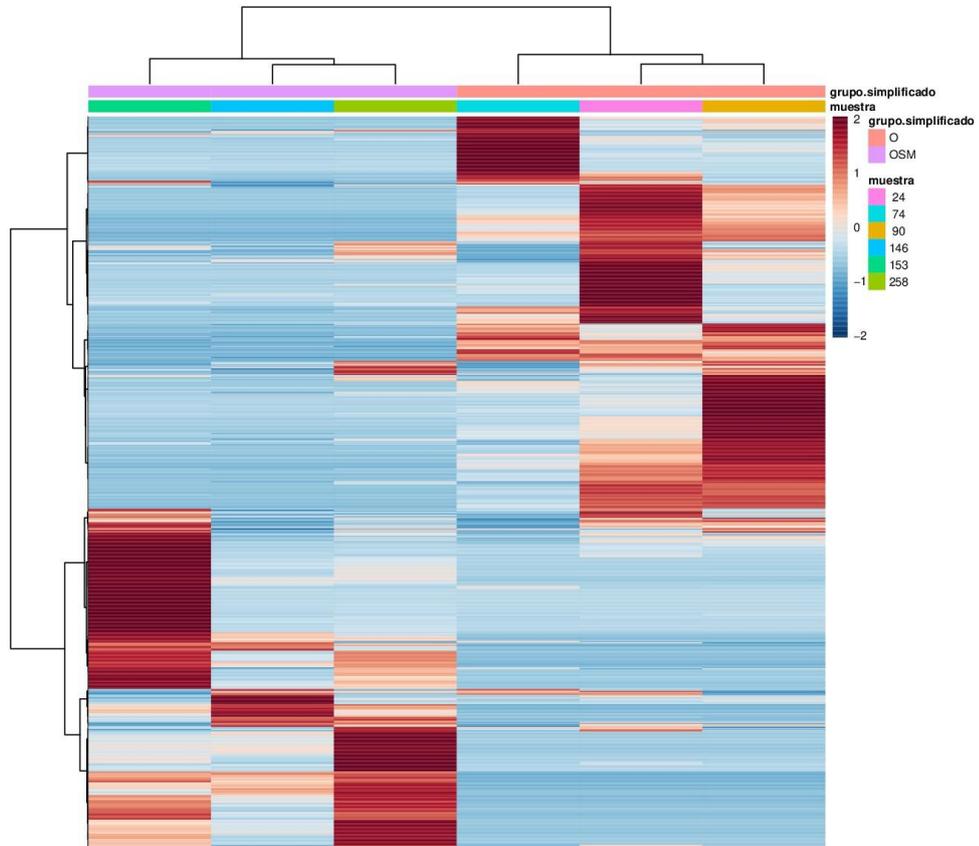


Figura 3.5: Análisis de clasificación no supervisada con el coeficiente de correlación de Pearson sobre los *core* de OTUs de cada grupo (O=266 y OSM=286).

La taxonomía agrupada por niveles taxonómicos (phylum, clase, orden, familia, género y especie) se puede observar de las Figuras 3.6, 3.8, 3.10, 3.12, 3.14 y 3.16. Para el nivel de phyla se puede observar que los más abundantes son Bacteroidetes y Firmicutes que juntos suman $\sim 90\%$ de la abundancia relativa de todas las muestras; los phyla Proteobacteria, Actinobacteria, Saccharibacteria, Cyanobacteria, Fusobacteria, Lentisphaerae, Spirochaetes y Verrucomicrobia son otros phyla que se encuentran en las muestras en abundancias más bajas ($>10\%$) (Figura 3.6).

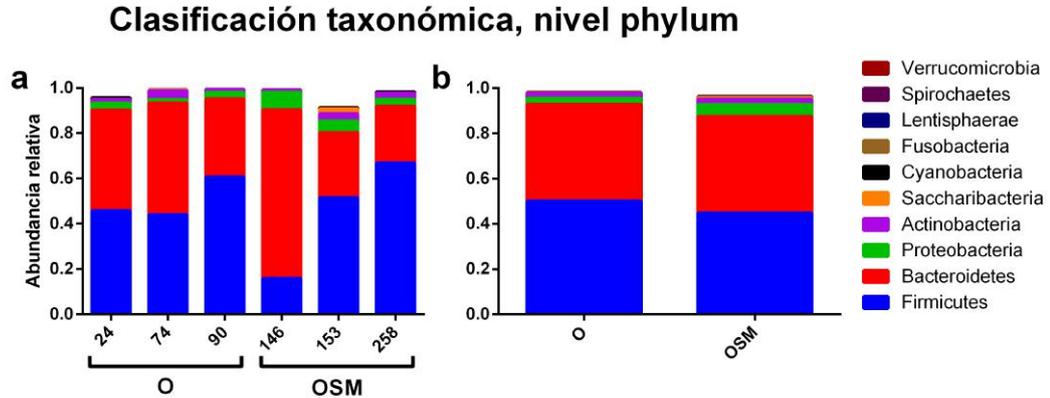


Figura 3.6: Clasificación taxonómica a nivel phylum, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).

El análisis LEfSe para determinar diferencias significativas entre los grupos, muestra que Proteobacteria tiene un aumento significativo en las muestras del grupo OSM; este phylum ha sido reportado cuando hay una disbiosis microbiana y es común en enfermedades humanas (36). Para el grupo O los phyla sobreabundantes fueron Armatimonadetes, Tenericutes (encontrado también en sobreabundante en ratones con obesidad (37)) y Lentisphaerae, el cuál previamente fue reportado en individuos con hígado graso no alcohólico y relacionado a la inflamación intestinal (38)(Figura 3.7).

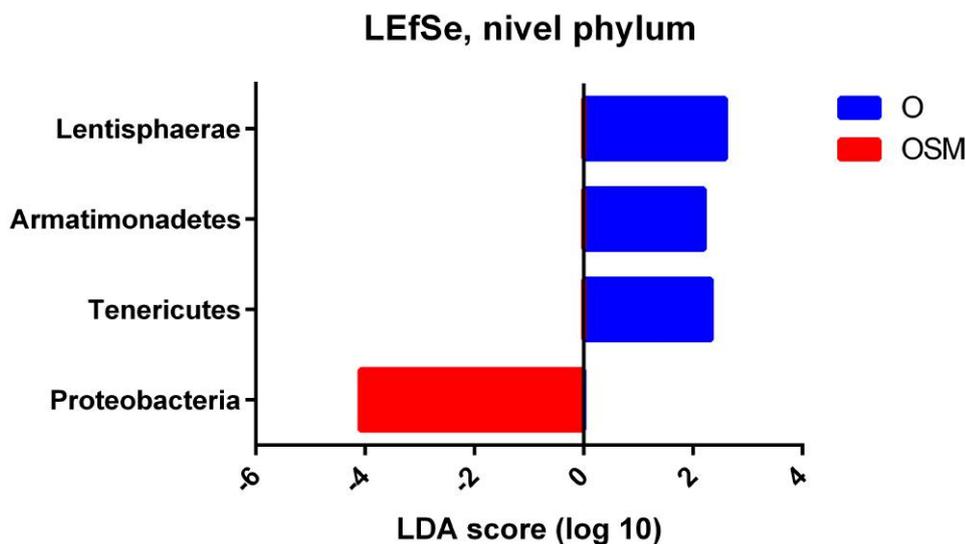


Figura 3.7: LEfSe a nivel phylum.

Las 10 clases con mayor abundancia se muestran en la Figura 3.8. Se obtuvo que las clases que dominan en todas las muestras son Aclorhidria y Bacteroidia con cerca de 80 % de las abundancias relativas totales, que se observan en colores azul y rojo, respectivamente. Las otras clases que se encuentran más abundantes son Gammaproteobacteria, Actinobacteria, Bacilli, Coriobacteria, Negativicutes, Erysipelotrichia, Deltaproteobacteria y Betaproteobacteria.

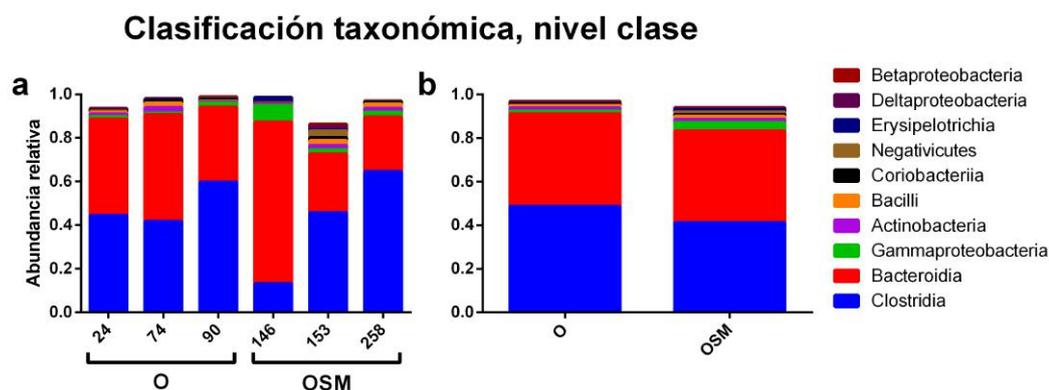


Figura 3.8: Clasificación taxonómica a nivel clase, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).

Con LEfSe se obtuvo que para el grupo OSM no existían diferencias signifi-

cativas a nivel de clase; para el grupo O se encontró que Chitinophagia, Fimbriimonadia, Lentisphaeria y Mollicutes estaban sobreabundantes en comparación al grupo OSM (Figura 3.9); éste último se ha reportado en ratones con obesidad inducida por alimentación (39).

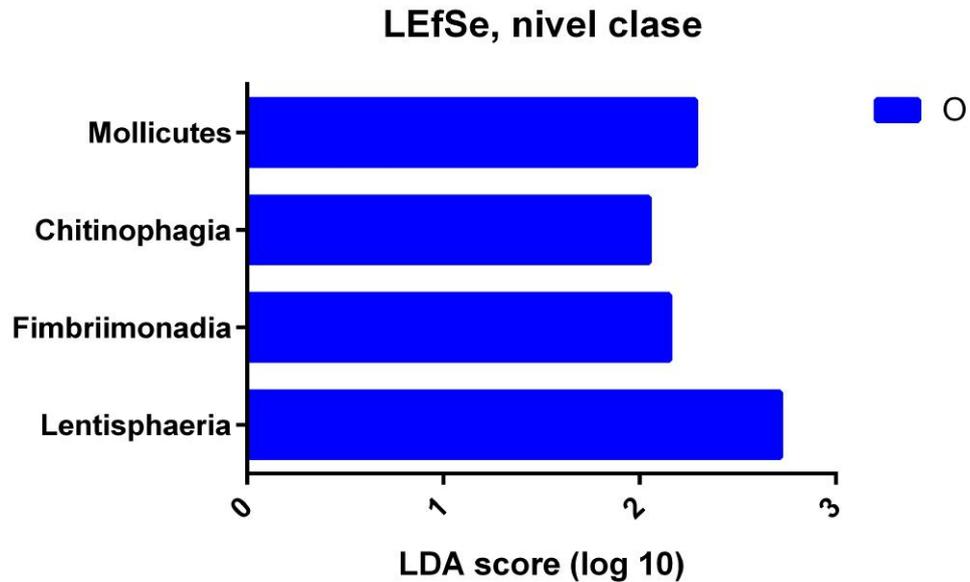


Figura 3.9: LEfSe a nivel clase.

A nivel de orden se obtuvo que los Clostridiales y los Bacteroidales fueron los más abundantes para ambos grupos. Clases con una menor abundancia fueron Enterobacterales, Bifidobacteriales, Lactobacillales, Veillonellales, Erysipelotrichales, Coriobacteriales y Baciliales, así como un orden desconocido del phylum Saccharibacteria (Figura 3.10).

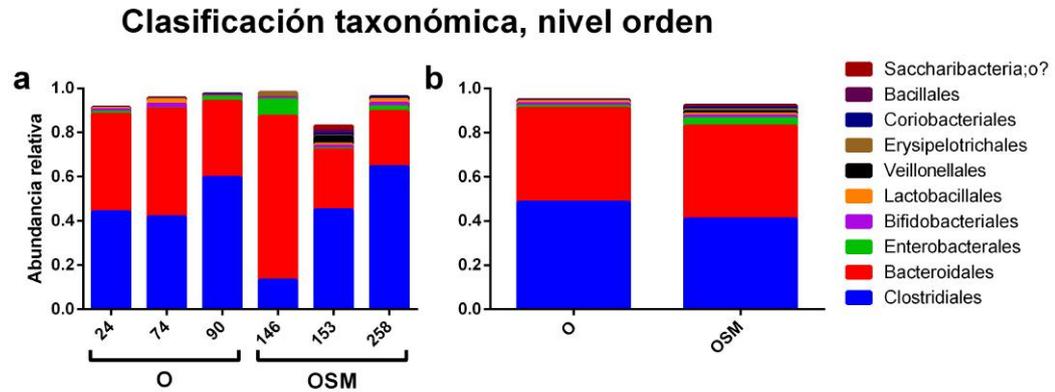


Figura 3.10: Clasificación taxonómica a nivel orden, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).

Al igual que a nivel de clase, no se encontraron diferencias significativas en los órdenes para el grupo OSM; para el grupo O se encontraron diferencias significativas: Victivallales, Actinomycetales, Chitinophagales, Fimbriimonadales, Pseudonocardiales y Rhodobacteriales (Figura 3.11); este último fue previamente reportado en individuos con hígado graso no alcohólico (40)

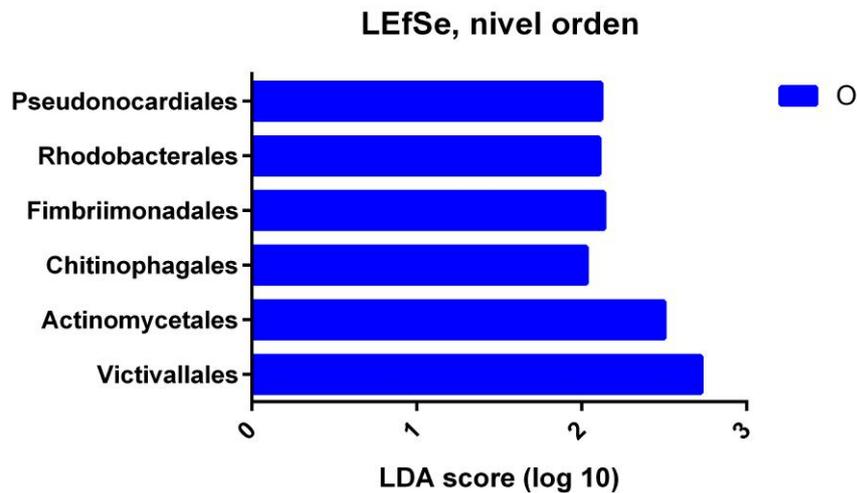


Figura 3.11: LEfSe a nivel orden.

La clasificación de nivel Familia nos muestra que las tres familias más abundantes fueron Bacteroidaceae, Ruminococcaceae y Lachnospiraceae sumando más

del 50 % en todas las muestras de ambos grupos; otras familias con menor abundancia fueron Tannerellaceae, Eubacteriaceae, Rikenellaceae, Prevotellaceae, Peptostreptococcaceae, Enterobacteriaceae y Clostridiaceae (Figura 3.12).

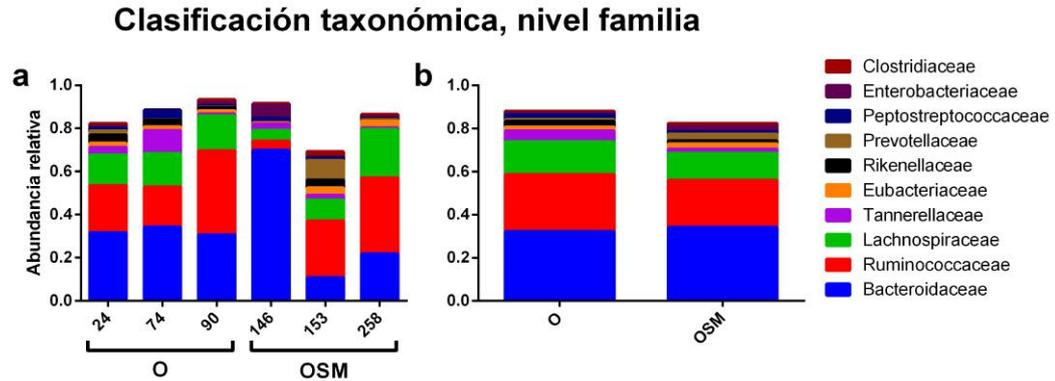


Figura 3.12: Clasificación taxonómica a nivel familia, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).

Las familias con diferencias significativas para el grupo OSM fueron Gallionellaceae, Geobacteraceae, Archangiaceae y Aerococcaceae. Mientras que para el grupo O las familias Rhodobacteraceae, Chitinophagaceae, Alcaligenaceae, Pseudonocardaceae y Fimbriimonadaceae, así como una familia no clasificada del orden Victivallales (Figura 3.12). Actinomycetaceae ha sido encontrado con mayores abundancias en individuos obesos (41), así como Pseudonocardaceae se ha relacionado con prediabetes (42).

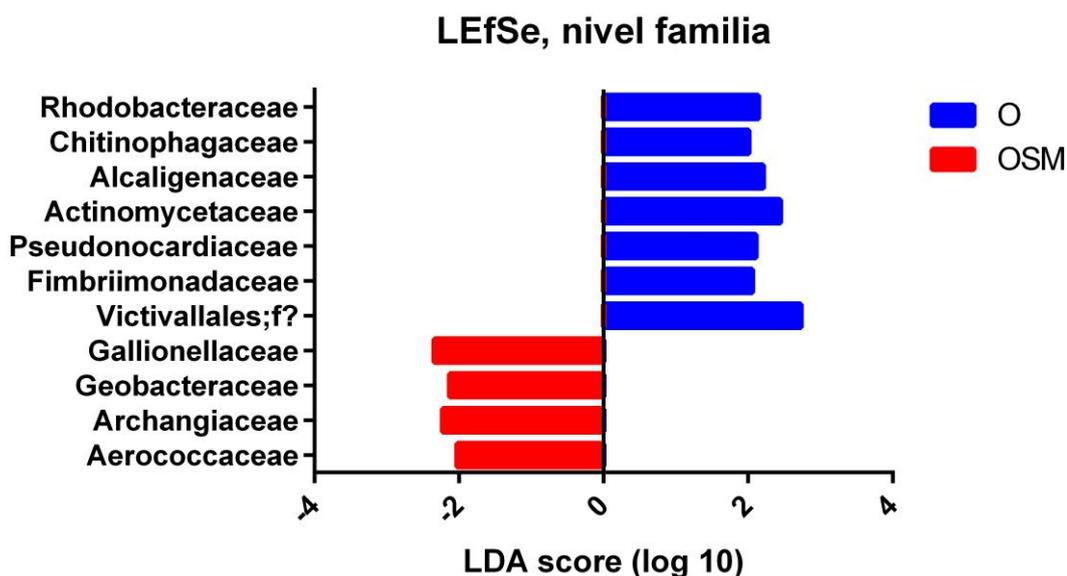


Figura 3.13: LEfSe a nivel familia.

Los 10 géneros más abundantes fueron *Bacteroides*, *Faecalibacterium*, *Parabacteroides*, *Alistipes*, *Eubacterium*, *Prevotella*, *Roseburia*, *Lachnospiraceae* y *Blautia*, así como un género no determinado dentro de la familia Lachnospiraceae. A diferencia de niveles superiores la variación entre las muestras son mayores a este nivel y al nivel de especie.

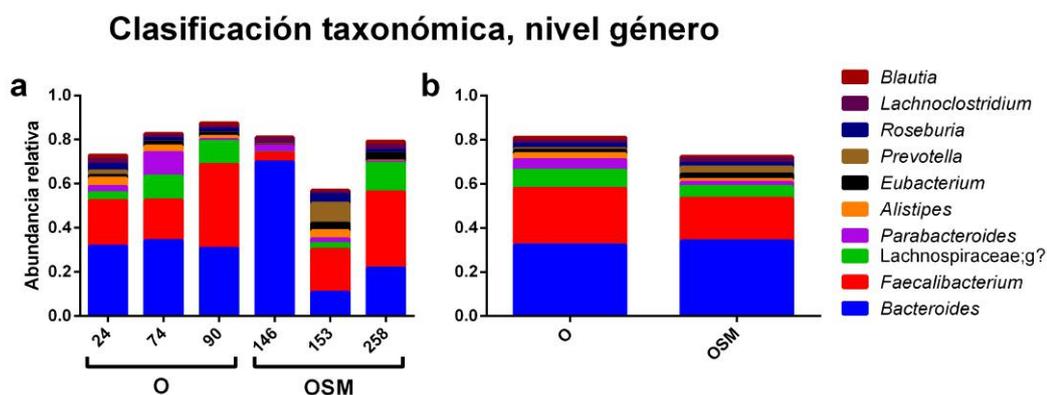


Figura 3.14: Clasificación taxonómica a nivel género, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).

Los géneros con diferencias significativas para el grupo OSM fueron 19: *Leu-*

conostoc, *Azobacteroides*, *Tumeriella*, *Zymomonas*, *Luteimonas*, *Desulforomonas*, *Enterococcus* (sobreaabundante en individuos con obesidad de población china (43).), *Geobacter*, *Veillonella* (encontrada en niños con *diabetes mellitus* de tipo I (44)), *Erysipelothrix*, *Pectobacterium*, *Stigmatella*, *Haemophilus*, *Paeniclostridium*, *Sulfuriferula*, *Aerococcus*, *Fronidhabitans*, *Alkaliphilus* y *Capnocytophaga*. Mientras que para el grupo O se pudieron diferenciar 13 géneros: *Scardovia*, *Adlercreutzia*, *Methylibium*, *Turicibacter*, *Faecalitalea*, *Desulfovibrio* (relacionado con la degradación de lípidos y carbohidratos junto con *Prevotella* (37)), *Muribaculum*, *Azotobacter*, *Acetobacterium* (encontrado en ratones con dietas altas en grasa (45)), *Siansivirga*, *Sinorhizobium* y *Fibriimonas*, así como un género no determinado de Victivallales (Figura 3.15).

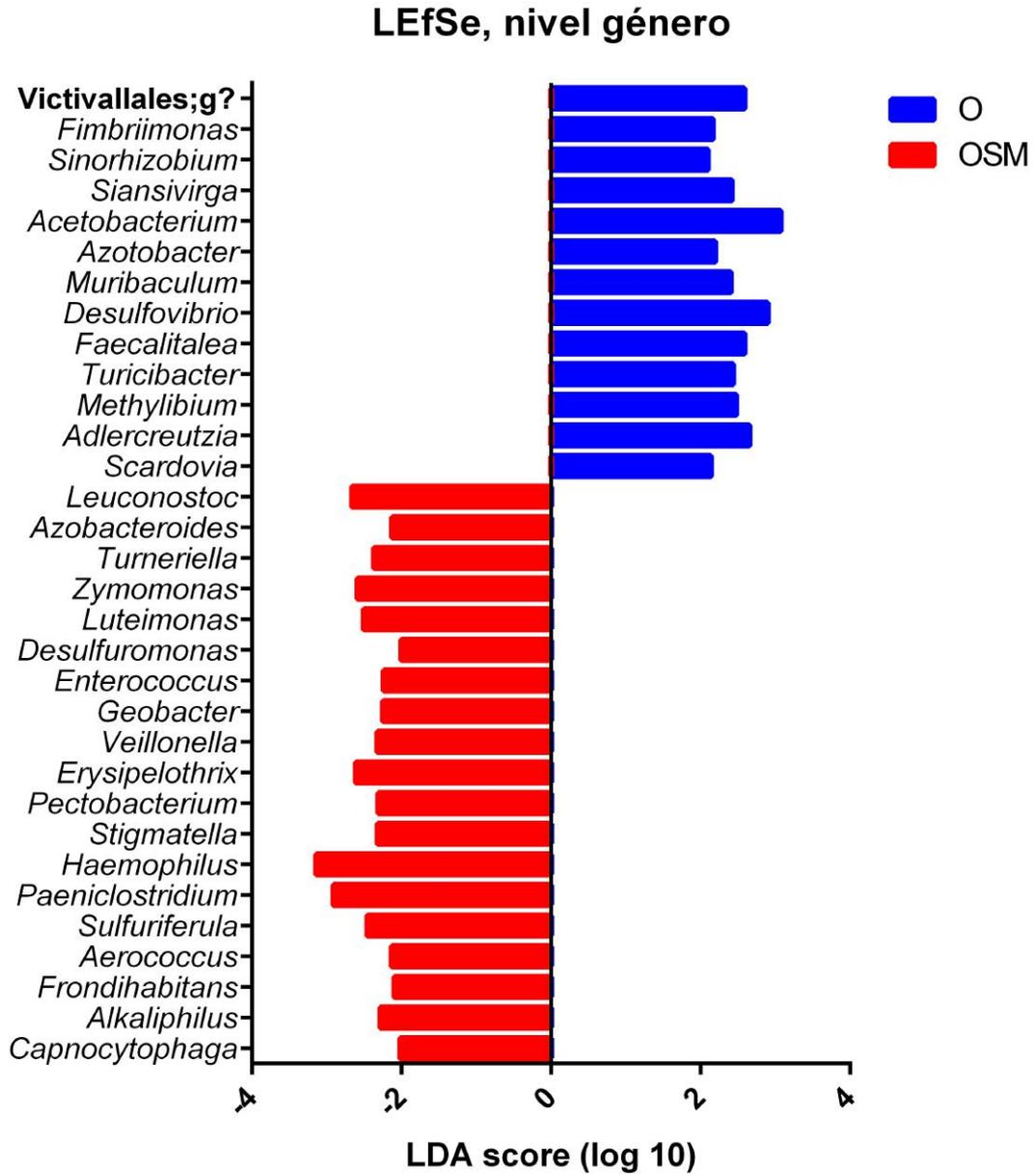


Figura 3.15: LEfSe a nivel género.

Las 10 especies con mayor abundancia fueron *Faecalibacterium prausnitzii*, *Bacteroides vulgatus*, *Eubacterium rectale*, *Bacteroides cellulosilyticus*, *Bacteroides ovatus*, *Bacteroides dorei*, *Bacteroides thetaiotaomicron* y *Roseburia hominis*, así como dos especies no clasificadas de los géneros *Bacteroides* sp. y *Parabacteroides* sp. (Figura 3.8).

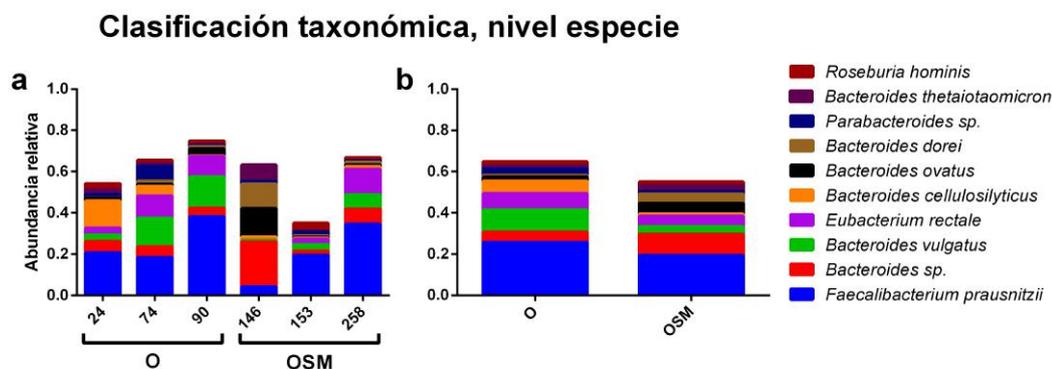


Figura 3.16: Clasificación taxonómica a nivel especie, considerando los primeros 10 lugares; graficados por muestra (a) y por grupo (b).

Las especies con diferencias significativas para el grupo OSM fueron 35: *Erysipelothrix sp.*, *Thermoanaerobacterium sp.*, *Sphaerochaeta pleomorpha*, *Stigmatella aurantiaca*, *Paenibacillus physcomitrellae*, *Enterococcus faecium*, *Clostridium perfringens*, *Turneriella parva*, *Aerococcus sanguinicola*, *Clostridium sp.*, *Clostridium chauvoei*, *Sulfuriferula sp.*, *Alkaliphilus oremlandii*, *Herbaspirillum hiltneri*, *Frondehabitans sp.*, *Prevotella scopos*, *Geobacter sp.*, *Paeniclostridium sordehlii*, *Prevotella jejuni*, *Luteimonas sp.*, *Marivirga tractuosa*, *Haemophilus pittmaniae*, *Ruminiclostridium thermocellum*, *Leuconostoc sp.*, *Prevotella sp.*, *Haemophilus parainfluenzae*, *Zymomonas mobilis*, *Paenibacillus ihbetae*, *Porphyromonas crevioricanis*, *Fusobacterium periodonticum*, *Clostridium bornimense*, *Pectobacterium carotovorum*, *Clostridium botulinum*, *Paenibacillus sp.* y *Cyclobacterium marinum*. Mientras que para el grupo O se observaron 25 especies con diferencias significativas: *Aeromonas sp.*, *Pelobacter sp.*, *Streptococcus salivarius*, *Faecalitalea cylindroides*, *Lactococcus piscium*, *Turicibacter sp.*, *Siansivirga zeaxanthinifaciens*, *Polyangium brachysporum*, *Prauserella marina*, *Pseudomonas sp.*, *Desulfovibrio sp.*, *Fimbriimonas ginsengisoli*, *Acetobacterium woodii*, *Streptococcus equinus*, *Azotobacter chroococcum*, *Muribaculum intestinale*, *Desulfosporosinus orientis*, *Adlercreutzia equolifaciens*, *Scardovia inopinata*, *Desulfovibrio piger*, *Ezakiella massiliensis*, *Arthrobacter sp.*, *Pseudomonas resinovorans*, *Burkholderiales sp.* y *Victivallales sp.* (Figura 3.16).

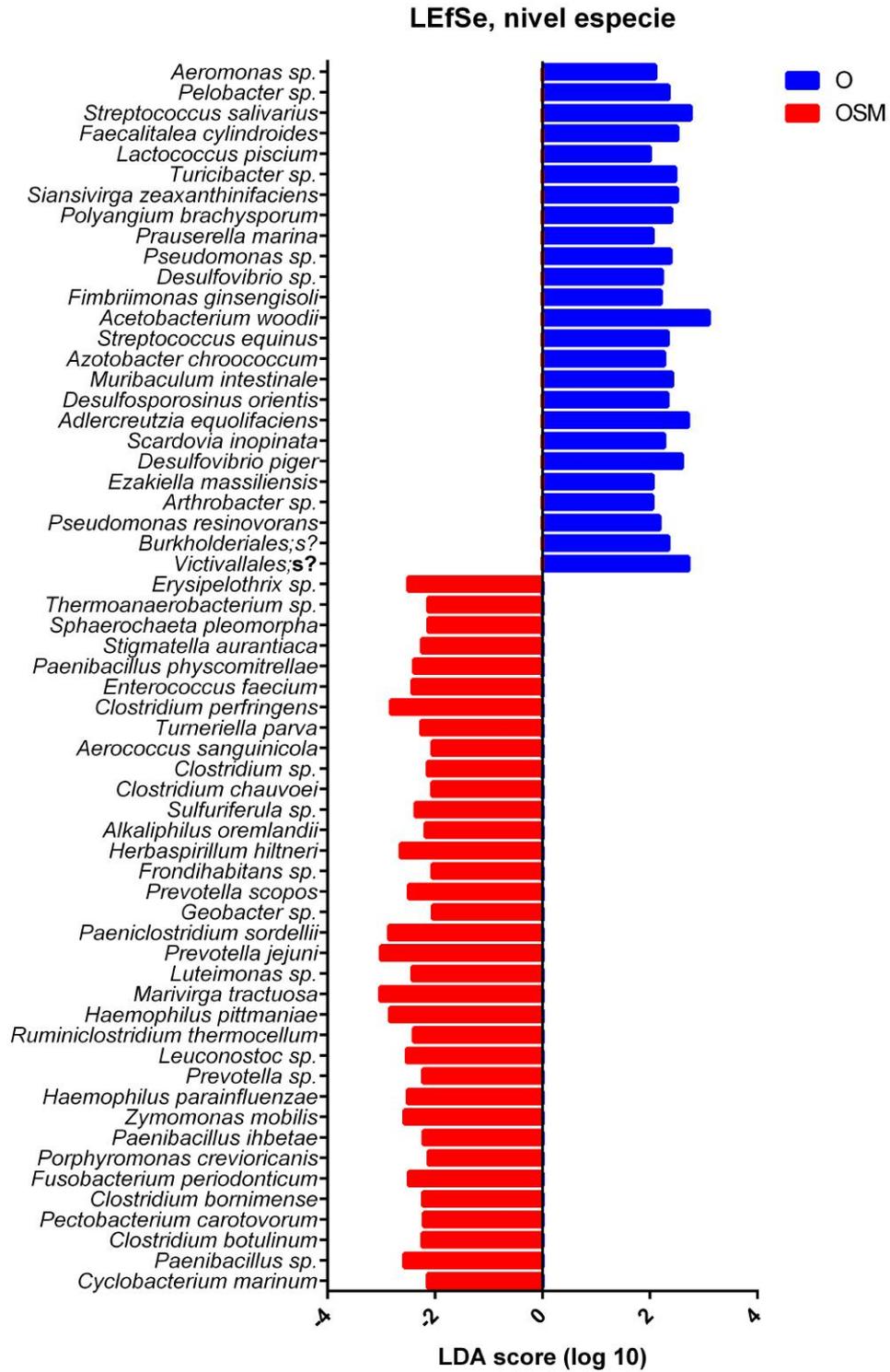


Figura 3.17: LEfSe a nivel especie.

3.2.3. Anotación funcional con HUMANn2

En la tabla 3.3 se puede observar el porcentaje de lecturas con anotación funcional con HUMANn2 y la base de datos UniRef90, donde la asignación varió de $28.16 \pm 17.54\%$. Se calculó la diversidad alfa entre los grupos de muestras (Figura 3.18) donde se obtuvieron diferencias significativas (prueba Mann-Whitney $p < 0.005$). Hay una mayor diversidad de funciones en el O, según los resultados de Observed y Chao1, sin embargo en el OSM es más equitativa la contribución según los índices de Shannon y Simpson.

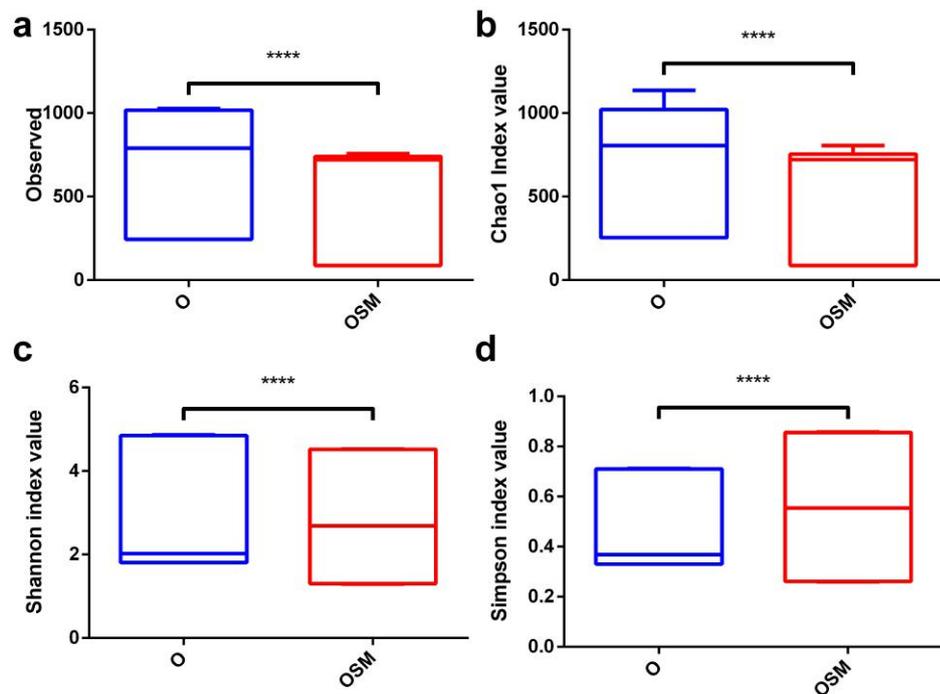


Figura 3.18: Diversidad alfa con la clasificación funcional. Las muestras están agrupadas en Obeso (O) u Obeso con Síndrome Metabólico (OSM); a) OTUs observados, b) Índice Chao1, c) Índice Shannon y d) Índice Simpson (****: Mann-Whitney $p > 0.005$).

Para determinar su funcionalidad se agruparon los resultados en grupos funcionales de COG (*Clusters of Orthologous Groups*), KEGG (*Kyoto Encyclopedia of Genes and Genomes*) y GO (*Gene Ontology*).

Tabla 3.3: Resultados de asignación con HUMANn2 con base de datos Uni-Ref90.

Muestra	% asignado
24 (O)	23
74 (O)	14
90 (O)	56
146 (OSM)	41
153 (OSM)	9
258 (OSM)	26

3.2.3.1. Clasificación funcional con COG

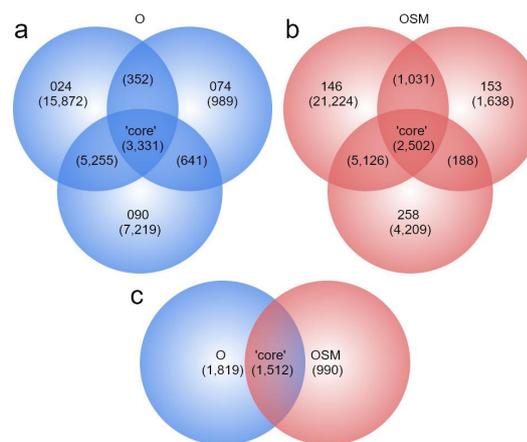


Figura 3.19: Core de funciones compartidas en los grupos (a y b) y entre los grupos (c) con la clasificación de COG.

En la Figura 3.19 se muestran los *core* de la clasificación funcional por COG de los grupos O (*core*=3,331) y OSM (*core*=2,502). En el módulo c se muestran las funciones compartidas entre los grupos: O=45.4% y OSM=60.4%.

Con las funciones *core* entre las muestras del mismo grupo se graficó la clasificación funcional por muestra y por grupo (Figura 3.20). Se puede observar que existe variación entre las muestras del mismo grupo (3.20.a), más no cuando se promedian por grupo (3.20.b).

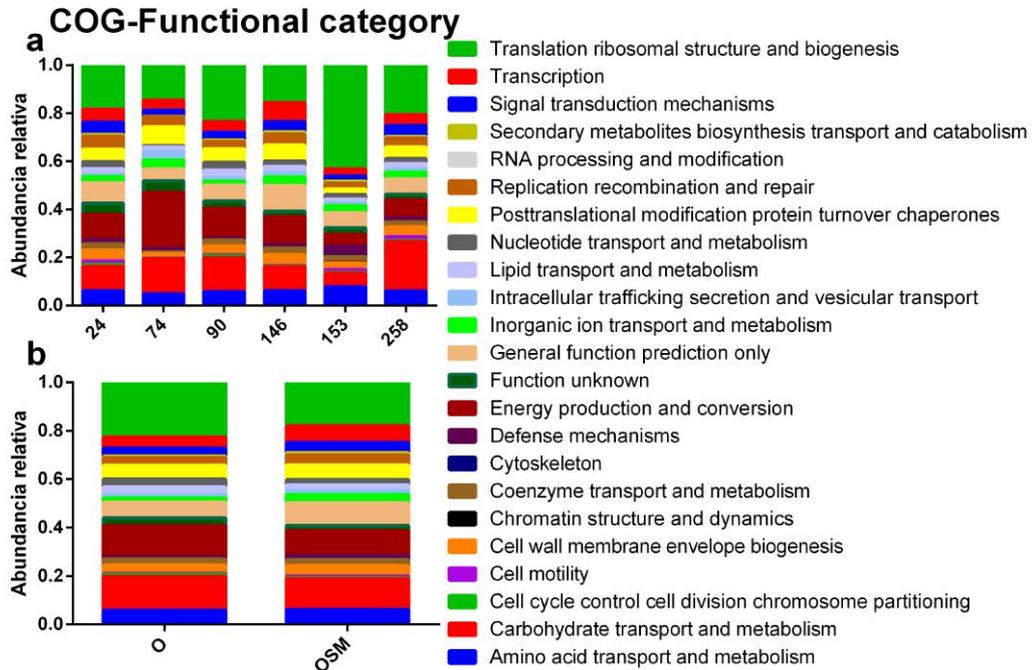


Figura 3.20: Clasificación funcional con la base de datos COG por muestra (a) y promediadas por grupo (b).

Para determinar las diferencias significativas entre las muestras se utilizaron los análisis de expresión diferencial DESeq2 y EdgeR comparando los COGs de cada grupo de muestras (Figura 3.21). Para el análisis con DESeq2 se obtuvieron 1248 funciones con diferencias significativas (545 sobre-expresadas en OSM y 703 en O; $\text{padj} < 0.005$), mientras que con EdgeR 1576 tuvieron diferencias significativas (para OSM 498 y para O 1069; $\text{FDR} < 0.005$). Posteriormente se buscaron las funciones que coincidían en ambos análisis de expresión diferencial dando un número total de 1175, 473 para el grupo OSM y 702 para el grupo O.

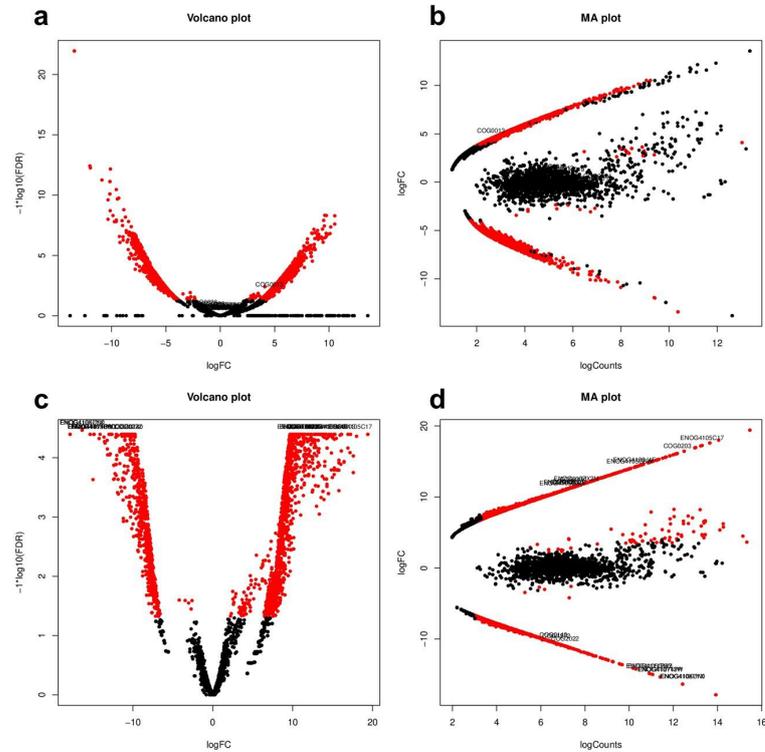


Figura 3.21: Análisis de expresión diferencial de las funciones con la base de datos COG con DESeq2 (a y b) y EdgeR (c y d); en rojo se muestran las funciones sobreexpresadas.

Las gráficas con las funciones diferencialmente sobreexpresadas agrupadas por COG se muestran en la Figura 3.22. Para el grupo OSM se puede observar un mayor número de funciones relacionadas con la modificación postraduccional, tráfico intracelular, secreción y transporte vesicular, así como funciones relacionadas a la producción de energía y conversión. Para el grupo O se encuentran sobre-expresadas en mayor abundancia funciones relacionadas con la traducción, estructura ribosomal y biogénesis, transcripción, replicación, biogénesis de membrana y pared celular, así como mecanismos de defensa.

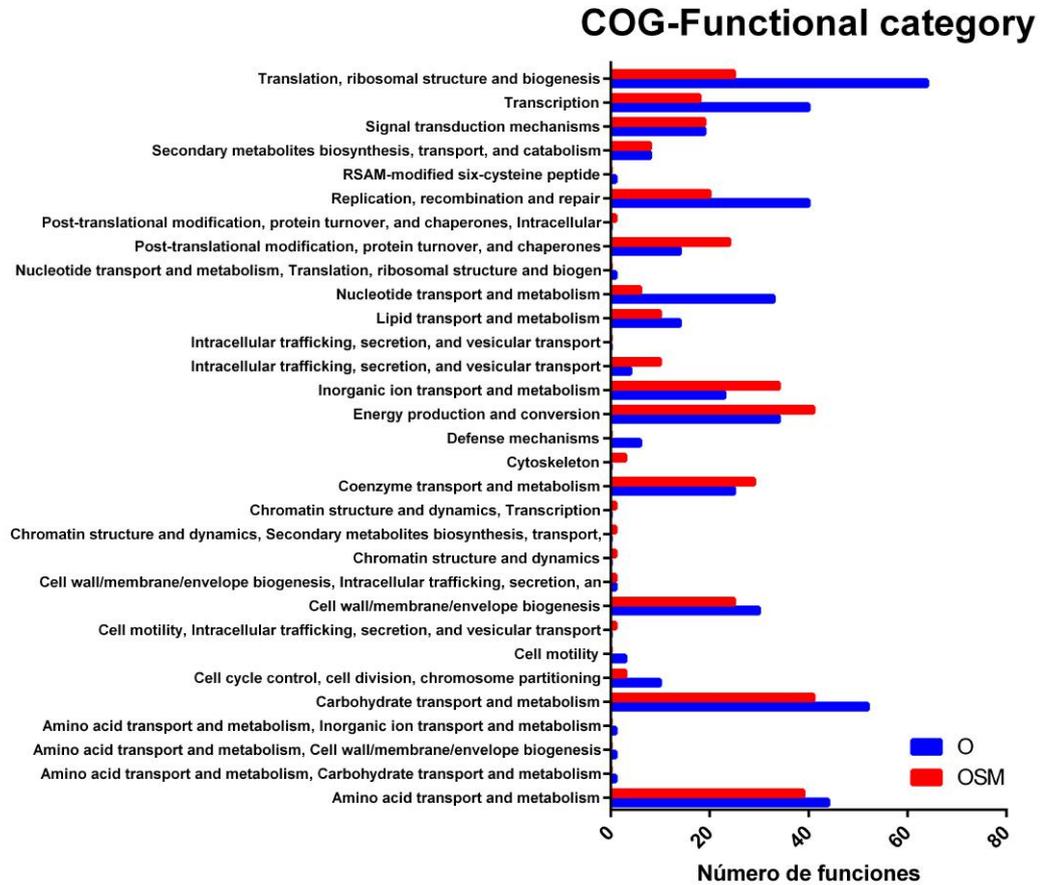


Figura 3.22: Clasificación funcional con la base de datos COG de las funciones con expresión diferencial.

3.2.3.2. Clasificación funcional con KEGG

Los *core* de KO por grupos se encuentran en la Figura 3.23 en los módulos a y b, mientras que en el módulo c se encuentran los KO que se comparten entre los grupos.

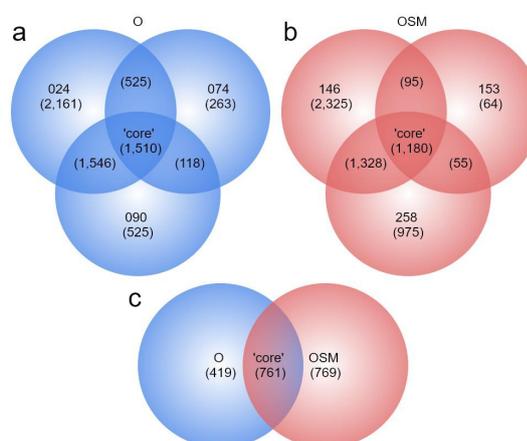


Figura 3.23: Core de funciones compartidas en los grupos (a y b) y entre los grupos (c) con la clasificación de KEGG.

Con los *core* por grupo se realizó una gráfica con la clasificación funcional de KEGG Metabolism (Figura 3.24), en donde se puede observar que las funciones conservadas entre las muestras son el metabolismo de carbohidratos, aminoácidos y de energía.

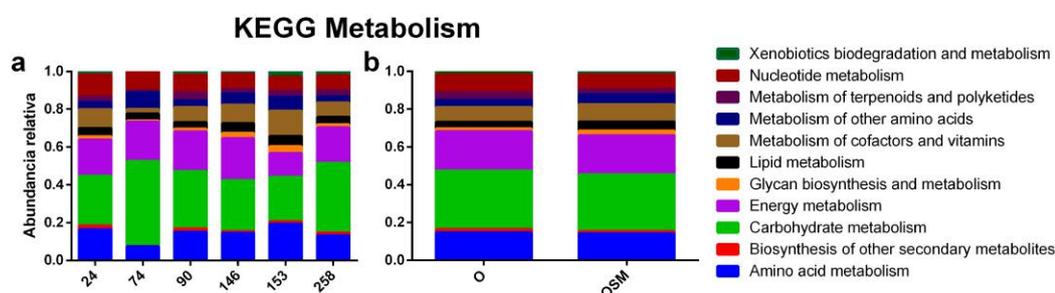


Figura 3.24: Clasificación funcional con la base de datos KEGG por muestra (a) y promediadas por grupo (b).

Para determinar los KOs con expresión diferencial entre los grupos de muestras se utilizaron DESeq2 y EdgeR. Para el primero se encontraron 879 funciones con expresión diferencial ($p_{adj} < 0.005$), 575 para el grupo OSM y 304 para el grupo O. Con EdgeR se obtuvieron 605 funciones con expresión diferencial ($FDR < 0.005$) de las cuales 313 fueron para OSM y 292 para el grupo O (Figura 3.25). Se contrastaron los resultados para ambos análisis de expresión diferencial y se obtuvo que 574 funciones son conservadas para ambos análisis, 308 para los OSM y 266

para O. Con dichas funciones agrupadas por KEGG-pathways se graficó la figura 3.26.

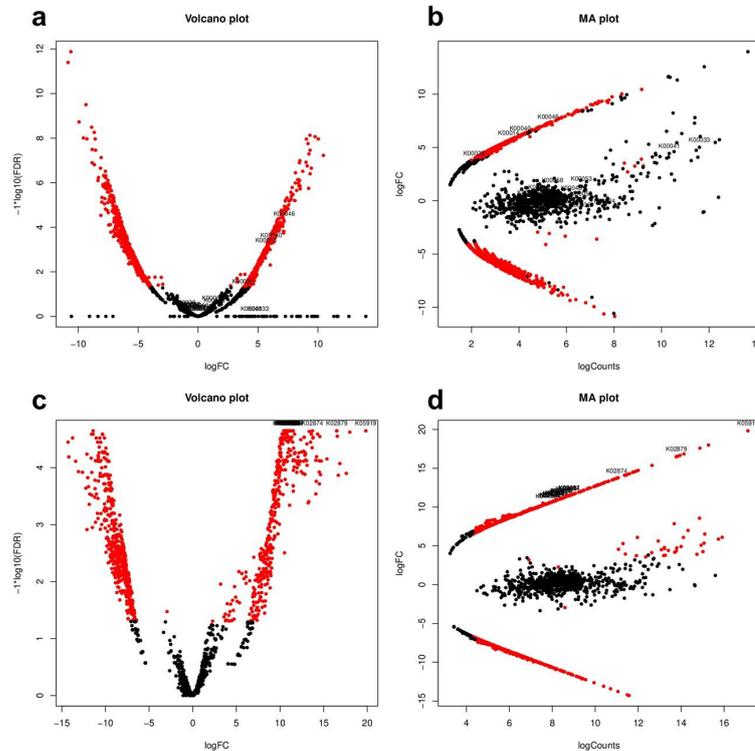


Figura 3.25: Análisis de expresión diferencial de las funciones con la base de datos KEGG con DESeq2 (a y b) y EdgeR (c y d); en rojo se muestran las funciones sobreexpresadas.

Entre las funciones sobre-expresadas en mayor número para el grupo OSM se encontraron el metabolismo de biotina, biosíntesis de lipopolisacáridos, ruta de las pentosas fosfato, biosíntesis de ubiquinona, metabolismo de fructosa y manosa, sistema de dos componentes, metabolismo de almidón y sucrosa, así como sistemas fosfotransferasa. Para el grupo O se encontró sobre expresión de funciones como biosíntesis de folato, metabolismo de histina, glicina, serina y treonina, así como de glioxilato y dicarboxilato. De igual forma funciones en relación a la biosíntesis de antibióticos y con relación a la estructura ribosomal.

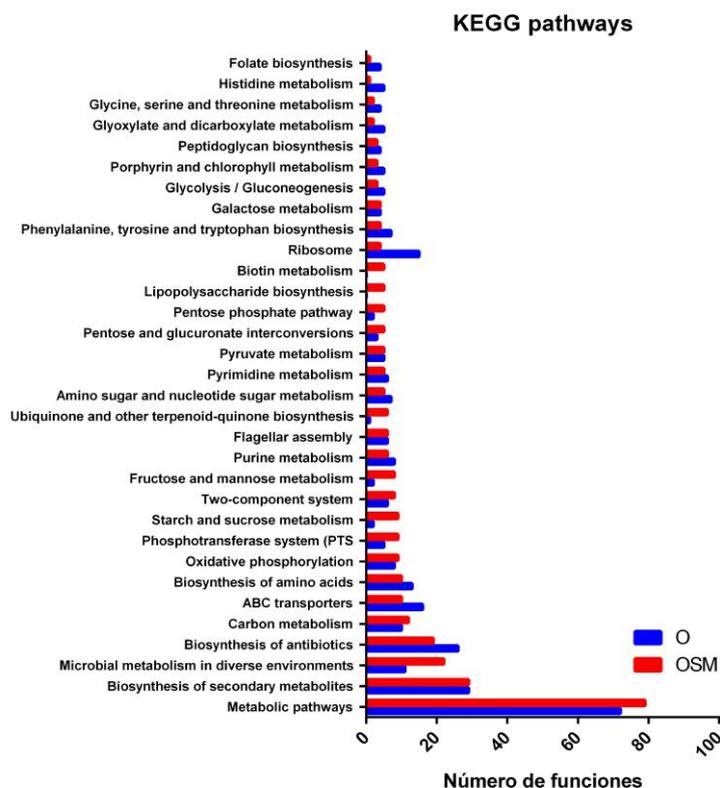


Figura 3.26: Clasificación funcional con la base de datos KEGG de las funciones con expresión diferencial.

3.2.3.3. Clasificación funcional con GO

Para la clasificación funcional con GO se obtuvieron los *core* por grupo (Figura 3.27); para el grupo O se obtuvieron 1,559 GO funciones compartidas mientras que 2,520 para el OSM. Las funciones compartidas entre los dos grupos fueron 1,365, el 87.5 % del grupo O y 54.2 % del grupo OSM.

La expresión diferencial entre los *core* de cada grupo se muestra en la Figura 3.28. Con el análisis con DESeq2 se obtuvieron un total de 618 funciones sobreexpresadas, siendo 111 para el grupo O y 506 para el OSM. Utilizando EdgeR se obtuvieron un total de 400 funciones sobreexpresadas, de las cuales 116 fueron par el grupo O y 284 para el grupo OSM. Comparando los dos análisis se coinciden en 381 funciones (105 para el grupo O y 276 para OSM) con las cuales se realizaron los gráficos de *Biological Process* (Figura 3.29), *Molecular Function* (Figura 3.30) y *Cellular Component* (Figura 3.31).

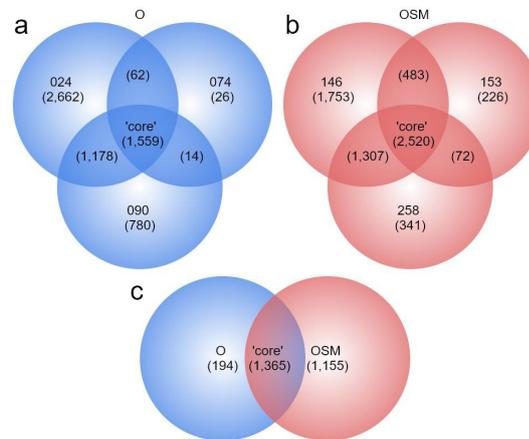


Figura 3.27: Core de funciones compartidas en los grupos (a y b) y entre los grupos (c) con la clasificación de GO.

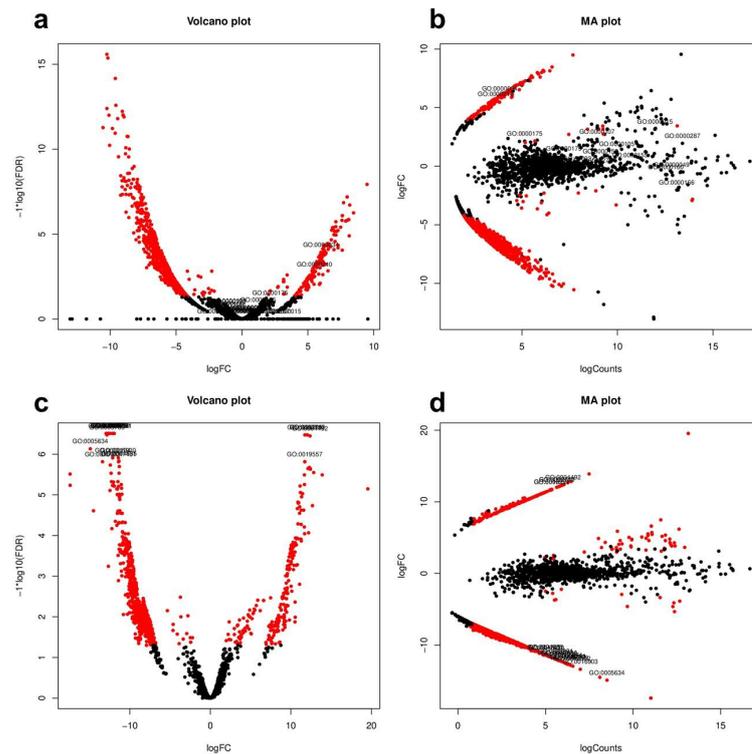


Figura 3.28: Análisis de expresión diferencial de las funciones con la base de datos KEGG con DESeq2 (a y b) y EdgeR (c y d); en rojo se muestran las funciones sobreexpresadas.

En la Figura 3.29 se muestran las funciones agrupadas por *Biological process*, donde se pueden observar las funciones sobreexpresadas en el grupo O (módulo a) y para el grupo OSM (módulo b). Funciones relacionadas al catabolismo de polisacáridos, transporte de iones, transporte de lipopolisacáridos, catabolismo de cuerpos cetónicos, transporte de bacteriocinas y fosforilación de carbohidratos fueron algunas de las funciones sobreexpresadas en el grupo OSM, lo cual nos permite inferir que hay una mayor cosecha de energía por parte de la microbiota intestinal en comparación al grupo O. Para el grupo O se sobreexpresaron funciones interesantes como replicación de virus de RNA, regulación de la esporulación, mantenimiento de elementos CRISPR y respuesta celular al calor.

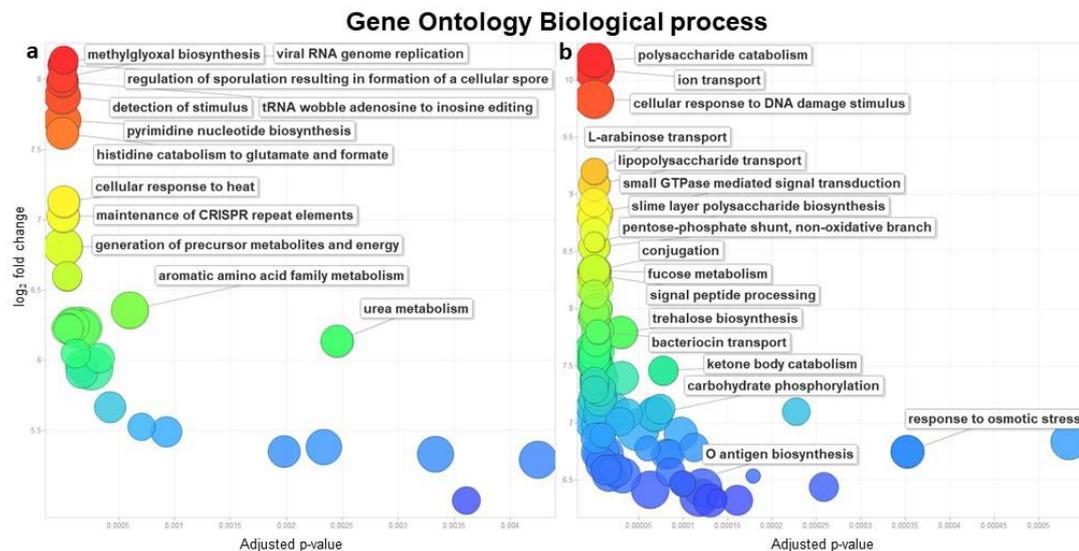


Figura 3.29: Clasificación con las funciones sobreexpresadas agrupadas en GO Biological Process (a=O, b=OSM).

En la gráfica 3.30 se muestran las funciones agrupadas en *Molecular Function*, donde se pueden observar las funciones sobreexpresadas en el grupo O (módulo a) y para el grupo OSM (módulo b). Para el grupo O funciones como 3-hidroxi-butiril-CoA, transportadores transmembranales de sideróforos, metilmalonil-CoA descarboxilasa, metilmalonil-CoA descarboxilasa, metilglioxal sintasa, diacilglicerol quinasa, oleato hidratasa y reservorio de nutrientes. Considerando a las muestras del grupo OSM se observó la sobreexpresión de funciones como ribosa-5-fosfato isomerasa, oxidoreductasa, ferredoxina hidrogenasa, 6-fosfogluconolactonasa, catabolismo de polisacáridos y actividad celulasa. En ambos grupos se encontraron funciones relacionadas en la conversión de energía.

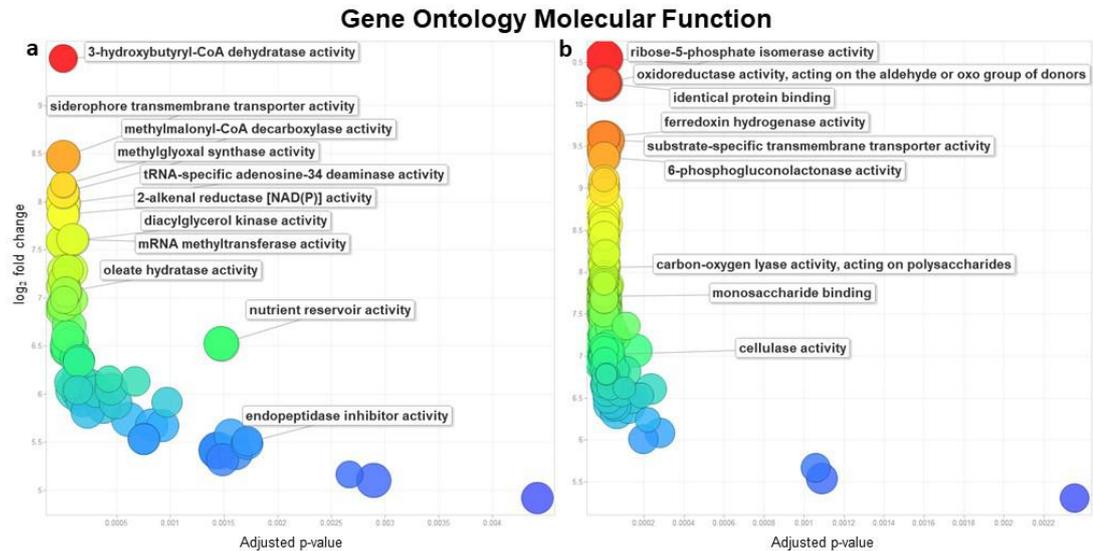


Figura 3.30: Clasificación con las funciones sobreexpresadas agrupadas en GO Molecular Function (a=O, b=OSM).

Considerando los Componentes Celulares se pudo observar que en el grupo O se encuentran sobrerrepresentados como complejo beta-galactosidasa, organelo intracelular unido a la membrana, complejo quinasa polifosfato y cuerpo basal de flagelo bacteriano (Figura 3.30 módulo a). Para el grupo OSM se encontraron componentes como mitocondria (no presente en microbiota pero se agrupan en este componente por las funciones de respiración celular relacionadas), complejo de poro, nucleolo (no presente en microbiota pero se agrupan en este componente funciones de transcripción-traducción), retículo endoplasmático (no presente en microbiota pero se agrupan en este componente funciones de metabolismo de lípidos y transporte celular), pilus y complejo de ferredoxin hidrogenasa; estas componentes están involucrados en la respiración celular y la degradación de lípidos, lo cual indica que hay una mayor cosecha de energía.

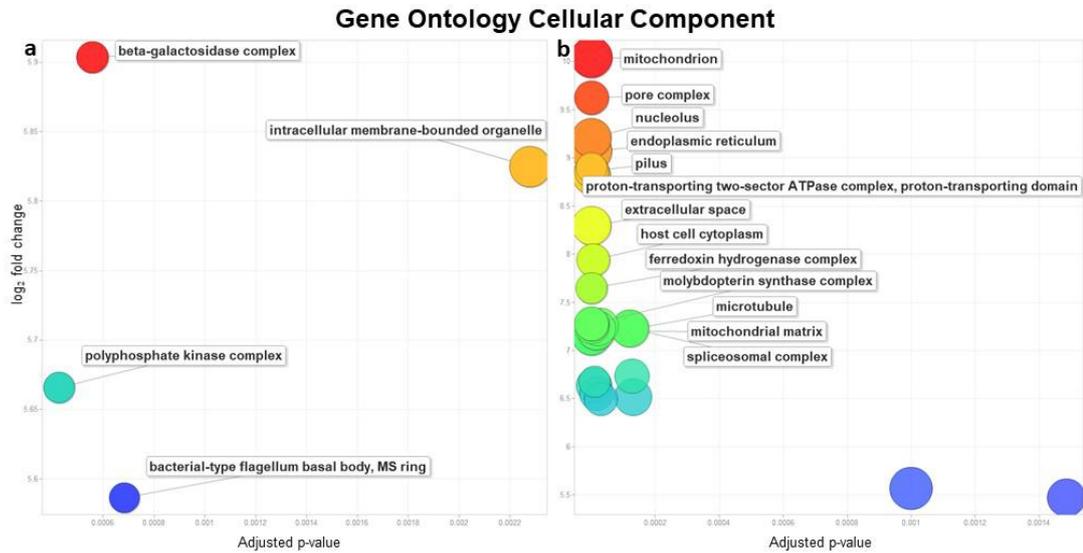


Figura 3.31: Clasificación con las funciones sobreexpresadas en GO Biological Process (a=O, b=OSM).

La diversidad taxonómica resultó tener diferencias significativas entre los grupos de muestras, Obeso (O) y Obeso con Síndrome Metabólico (OSM), porque con las métricas de Observed OTUs y Chao1 se obtuvo un incremento en riqueza de especies para el grupo O, sin embargo considerando los índices de Shannon y Simpson, el grupo OSM tuvo una mayor diversidad; esto se puede explicar por que aunque el número de especies es mayor en los obesos, en el grupo OSM se cuenta con una mayor equidad (determinada con un valor más alto del índice de Simpson) y una relación equidad-riqueza más alta (índice de Shannon) (figura 3.3). Considerando la diversidad alfa de las funciones encontradas (figura 3.18) se pudo determinar que existe una riqueza de funciones mayor en el grupo OSM (índices Observed y Chao1), pero al igual que en la taxonomía existe en este grupo una relación equidad-riqueza más alta, lo cual nos habla de que hay una distribución más homogénea de las funciones. Con los análisis de diversidad anteriores se puede inferir que los individuos con SM cuentan con un grado de enfermedad más avanzado, lo cual impacta en la riqueza de especies y funciones; ahora bien que al considerar la diversidad el grupo OSM cuenta con valores más altos de Shannon y Simpson, lo cual indica que los grupos taxonómicos y las funciones distribuidas más homogéneamente.

Con el fin de disminuir el impacto de la alta variación entre individuos, y que no tiene relación con el fenotipo O y OSM, se consideraron los *core* entre cada grupo de muestras para la asignación taxonómica y la anotación funcional. Esto resultó exitoso, ya que como se puede observar en la figura 3.5 las muestras tienen un grado de correlación más alto con las que pertenecen al mismo grupo. En los diagramas de Venn se puede observar también que la gran mayoría de los OTUs se comparten entre los grupos O y OSM, lo cual nos permite inferir de que no es una cuestión de presencia-ausencia sino más bien de abundancia relativa lo que nos permite discernir entre fenotipos.

Los phyla más dominantes son los Firmicutes y Bacteroidetes con cerca de

90 % de las abundancias relativas de todas las muestras lo cual es consistente con otros estudios de metagenoma y metatranscriptoma. Para los OSM se obtuvieron diferencias significativas con el análisis de LEfSe para el phylum Proteobacteria el cuál es considerada como un factor común en enfermedades humanas; ahora bien para el grupo O se obtuvieron los phyla Lentisphaerae y Tenericutes previamente relacionado con inflamación intestinal y obesidad, respectivamente (figura 3.7). A nivel de clase se puede observar que se encuentra dominado por Clostridia (Firmicutes) y Bacteroidia (Bacteroidetes) y que cuando se considera el análisis LEfSe no existen Clases sobre-expresadas en OSM pero sí en el grupo O: Lentisphaeria, Chitinophagia, Fimbriimonadia y Mollicutes, el último previamente relacionado con obesidad (39). Los órdenes dominantes fueron Clostridiales y Bacteroidales y se obtuvieron diferencias significativas solamente para el grupo O: Victivallales, Actinomycetales, Chitinophagales, Fimbriimonadales, Pseudonocardiales y Rhodobacterales, que previamente ha sido relacionado con hígado graso no alcohólico (40). Las tres familias más abundantes fueron Bacteroidaceae, Ruminococcaceae y Lachnospiraceae; se encontraron diferencias significativas para ambos grupos. Para OSM se encontró que Gallionellaceae, Geobacteraceae, Archangiaceae y Aerococcaceae están sobre-representadas en dicho grupo. Los géneros más abundantes en todas las muestras fueron *Bacteroides* y *Faecalibacterium*, lo cual corresponde con lo previamente reportado en análisis de metatranscriptoma humano (5). Con la asignación de taxonomía se determinó que existían bacterias con abundancias diferenciales entre los grupos algunas de ellas asociadas a la obesidad como *Turicibacter* sp., *Pseudomonas* sp., *Desulfovibrio* sp. y *Acetobacterium woodii* en el grupo O (37). Para los OSM encontramos especies diferenciales como *Enterococcus faecium* (probiótico (46)), *Clostridium perfringens* y *Prevotella* sp., anteriormente relacionados con obesidad (37). Además, se encontraron diferenciales *Prevotella jejuni* que ha sido reportado en individuos con enfermedad celiaca (37), *Paeniclostridium sordelli* relacionada con colitis y *Aerococcus sanguinicola* que causa infecciones urogenitales.

La actividad funcional resultó diferencial entre los grupos OSM y O, donde encontraron 1175 COGs (473 para OSM y 702 para O), 574 KOs (308 para OSM y 266 en O) y 381 GOs (276 para OSM y 105 para O). Es interesante mencionar que para las muestras del grupo OSM se obtuvo una mayor diversidad de funciones sobre-expresadas (con KEGG y GO, figuras 3.25 y 3.28, respectivamente), lo cual corresponde con una mayor diversidad taxonómica y de funciones en comparación al grupo O (figuras 3.3 y 3.18, respectivamente); lo anterior sugiere que las especies del grupo O pueden tener una gama más variada de funciones, pero con dominancia de algunas pocas; en comparación las del grupo OSM tienen abundancias relativas más homogéneas de sus funciones lo cual indica que no hay dominancia de ninguna de ellas.

Para el grupo OSM se puede observar un mayor número de funciones re-

lacionadas con la modificación postraducciona, tráfico intracelular, secreción y transporte vesicular, así como funciones relacionadas a la producción y conversión de energía. De igual forma se sobre-expresaron genes relacionados para el metabolismo de biotina, biosíntesis de lipopolisacáridos, biosíntesis de ubiquinona, metabolismo de fructosa y manosa, almidón y sucrosa. Para el grupo O se encuentran sobre-expresadas en mayor abundancia funciones relacionadas con la traducción, estructura ribosomal y biogénesis, transcripción, replicación, biogénesis de membrana y pared celular, así como mecanismos de defensa; al igual que funciones como biosíntesis de folato, metabolismo de histina, glicina, serina y treonina, así como de glioxilato y dicarboxilato. Interesantemente también se encontraron funciones relacionadas con la biosíntesis de antibióticos en este grupo, los cuales han sido previamente reportadas con la incidencia de obesidad (47) más no con el SM.

Conclusiones y Perspectivas

En esta tesis se muestran los resultados preliminares de un trabajo de investigación que busca conocer el impacto funcional de la microbiota intestinal desde el punto de vista del metatranscriptoma. Como conclusiones relevantes se puede mencionar que se logró con éxito la estandarización de la extracción de ARN proveniente de muestras fecales humanas, con una calidad suficiente (RIN 5 a 7) para poder generar librerías simultáneas de *RNA-seq*; de igual forma se implementó un protocolo para poder generar las librerías simultáneas de *RNA-seq* con lo cual se ahorró tiempo y costos en la elaboración de las mismas, haciendo viable el uso de las mismas en la continuación de este proyecto.

Con el análisis bioinformático se logró determinar que existen diferencias significativas en los perfiles metatranscripcionales de los grupos obeso y obeso con SM, tanto en microbiota sobreabundante como en actividad funcional. Es interesante mencionar que nos limitamos a comparar la taxonomía y funcionalidad compartida entre los grupos (*core*), con el fin de delimitar en lo posible la alta variabilidad que puede existir a nivel metatranscripcional proveniente de muestras fecales. Las diferencias encontradas, tanto en función como en diversidad microbiana, son consistentes entre las muestras de un mismo grupo.

Como perspectiva principal se puede mencionar que es necesario considerar un grupo de individuos más amplio para poder establecer con análisis estadísticos más robustos los marcadores funcionales relacionados a la obesidad y el SM. De igual forma se requiere que en el proyecto futuro se incluyan a individuos delgados (magros) metabólicamente sanos para poder comparar sus perfiles metatranscripcionales contra los de los grupos obeso y obeso con SM; esto permitirá dilucidar más a fondo el papel del microbioma en la evolución de obesidad y el SM. Para la continuación de este proyecto también se puede considerar una profundidad de secuenciación más alta para cada grupo de individuos, lo cual nos permitiría hacer ensamblados 'de novo' con el fin de poder descubrir nuevos genes que pudiesen estar relacionados con la promoción de obesidad y el SM, y que en este proyecto

tal vez no se detectaron por falta de información en bases de datos públicas. De igual forma sería interesante considerar la aportación que tienen los bacteriófagos en la modulación de las poblaciones microbianas, así como integración de nuevas funciones a la microbiota, lo cuál podría tener un impacto en el desarrollo de obesidad y SM.

A.1. Procedimientos de biología molecular

A.1.1. Extracción de ARN de muestras fecales

Extracción de RNA con RNeasy Mini Prep de Qiagen

1. Centrifugar 300 μL de muestra durante 1 min a velocidad máxima junto con 500 μL de RNA Later.
2. Transferir 700 μL del sobrenadante a un tubo de 1.5 mL.
3. Añadir 1 volumen (700 μL) de etanol al 70 %.
4. Transferir inmediatamente a la columna **RNeasy spin**, centrifugar durante 30 s a 10 000 rpm. Descartar el sobrenadante. Repetir hasta terminar la muestra.
5. Añadir 350 μL **RW1 buffer** a la columna. Centrifugar por 30 s a 13 000 rpm. Descartar sobrenadante.
6. Añadir 500 μL **RPE buffer** a la columna. Centrifugar por 30 s a 13 000 rpm. Descartar sobrenadante.
7. Añadir 500 μL **RPE buffer** a la columna. Centrifugar por 2 min a 13 000 rpm. Descartar sobrenadante.
8. Centrifugar por 1 min la columna a 13 000 rpm.

9. Colocar la columna en un nuevo tubo eppendorf de 1.5 mL (frío previamente). Añadir 30 μL de agua libre de ARNasas (fría). Centrifugar por 1 min a 13 000 rpm y retirar el sobrenadante. Repetir este paso para obtener dos eluciones finales.
10. Almacenar a -70°C .

A.1.2. Purificación de ARN

Tratamiento con ADNasa

1. Preparar la siguiente reacción:
 - a) 5 μL 10X Buffer DNase
 - b) 1 μL DNase (2 U/ μL)
 - c) 46 μL de ARN
 - d) 52 μL total
2. Incubar durante 30 min a temperatura ambiente.

Purificar con perlas magnéticas AgencourtR RNA Clean XP

1. **Nota: antes de utilizar, colocar las perlas a temperatura ambiente durante 2 horas. El volumen de reacción de las perlas es 1.8x.**
2. Homogeneizar las perlas magnéticas y la reacción. Se puede hacer con vórtex o moviendo el tubo hacia arriba y abajo.
3. Incubar durante 5 min a temperatura ambiente.
4. Colocar la reacción en una placa magnética e incubar de 5 a 10 minutos (hasta que se separen las perlas magnéticas de la solución).
5. Retirar el sobrenadante evitando no remover las perlas magnéticas.
6. Añadir 200 μL de etanol al 70 % a las perlas unidas al magneto, evitando despegarlas. Incubar durante 30 s. Retirar y desechar el sobrenadante. Repetir 3 veces este paso.
7. Al finalizar el paso anterior, dejar secar de 10 a 15 min.
8. Añadir 30 μL de agua DEPC y homogeneizar. Se puede hacer con vórtex o moviendo el tubo hacia arriba y abajo. Posteriormente colocar de nuevo en el magneto y retirar el sobrenadante.
9. Almacenar a -20°C

A.1.3. Generación de librerías simultáneas

Fragmentación de ARN

1. Preparar la siguiente reacción en tubos para PCR:
 - a) 10 μL FastAP Buffer 10X
 - b) 40 μL ARN
 - c) 50 μL total
2. Incubar durante 3 min a 94°C en un termociclador.
3. Llevar al hielo.

Eliminación de ADN con ADNasa

1. Preparar la siguiente reacción en tubos para PCR:
 - a) 50 μL de ARN fragmentado
 - b) 1 μL de inhibidor de RNase Murine (40U/ μL)
 - c) 4 μL DNase TURBO (2 U/ μL)
 - d) 10 μL FastAP (1 U/ μL)
 - e) 35 μL Agua DEPC
 - f) 100 μL total
2. Incubar durante 10 min a 37°C en un termociclador.
3. Lavado de reacción con perlas magnéticas AgencourtR RNA Clean XP al igual que se explica en la sección anterior. **Nota:** al final eluir en 6 μL de agua DEPC.

Primera ligación

1. Añadir 1 μL del adaptador con código de barras (100 pmoles; *tag*), diferente para cada muestra, a μL de ARN desfosforilado .
2. Incubar durante dos minutos a 70°C en un termociclador y al finalizar pasar a hielo.
3. **Nota:** la siguiente reacción se realiza a temperatura ambiente. El DMSO precipita cuando es expuesto al frío. Tomar 25 % extra de PEG y realizar el pipeteo lentamente.

- a) 2 μL de buffer 10X de ligasa de ARN T4
- b) 1.8 μL de DMSO (al 100 %)
- c) 0.2 μL de ATP (100 mM)
- d) 8 μL de PEG 8000 (al 50 %)
- e) 0.3 μL de inhibidor RNase Murine (40 U/ μL)
- f) 1.8 μL de ligasas de ARN T4 (30 000 U/ μL)
- g) 14.1 μL total

4. Mezclar la reacción completamente.
5. Centrifugar rápidamente
6. Añadir el volumen total de la reacción anterior a los 6 μL del ARN junto con el adaptador.
7. Homogeneizar por pipeteo lento.
8. Incubar durante 90 min a 22°C (o temperatura ambiente).

Agrupación de muestras independientes

1. Añadir 60 μL de buffer RLT a cada muestra para inhibición de la ligasa (volumen total de 80 μL)
2. Agrupar las muestras y realizar lavado utilizando el **Zymo & Concentrator TM- 5 column**
3. Realizar la siguiente reacción en un tubo de 1.5 mL (se pueden juntar hasta 4 muestras):
 - a) 80 μL de cada muestra
 - b) 160 μL por muestra de Binding Buffer
 - c) 240 μL por muestra de EtOH (al 100 %)
 - d) 480 μL por muestra
4. El volumen total de la reacción anterior para cuatro muestras (1920 μL) se transfiere a la Zymo-Spin TM IC en un tubo colector y se centrifuga durante 30 s. Desechar el sobrenadante.
5. Añadir 400 μL de buffer RNA Prep a la columna y centrifugar por 30 s. Desechar el sobrenadante.

6. Añadir 700 μL de buffer RNA Wash a la columna y centrifugar por 30 s. Desechar el sobrenadante.
7. Añadir 400 μL de buffer RNA Wash a la columna y centrifugar por 2 min para terminar de desechar completamente los buffers usados. Al terminar transferir la columna a un nuevo tubo de 1.5 mL.
8. Añadir 10 μL de agua DEPC directamente a la matriz y centrifugar por 30 s.
9. Repetir el paso anterior.
10. Obtener un volumen total de 20 μL en total de las dos eluciones de RNA.

A.1.4. Remoción de ARNr de bacteria

Se utiliza el Ribo-Zero TM (Kit de bacteria)

Preparación de perlas magnéticas

1. Tomar 100 μL de perlas magnéticas y depositarlas en un tubo nuevo de 1.5 mL libre de ARNasas.
2. Incubar por 1 min en el magneto (a temperatura ambiente) y desechar el sobrenadante.
3. Lavar con 100 μL de agua libre de ARNasas.
4. Repetir el lavado.
5. Resuspender en 65 μL de solución para resuspensión.

Tratamiento para remoción de ARNr

1. Realizar la siguiente reacción:
 - a) 20 μL de ARN total
 - b) 4 μL de solución rRNA Removal Solution
 - c) 2 μL de buffer de reacción
 - d) 26 μL total
2. Incubar durante 10 min a 68°C y al finalizar a temperatura ambiente durante 5 min.

Remoción de ARNr

1. Homogenizar con vórtex las perlas magnéticas preparadas previamente.
2. Añadir la reacción de Remoción de ARNr (26 μL) a las perlas magnéticas, mezclar por pipeteo y vórtex rápido.
3. Incubar durante 5 minutos a temperatura ambiente, vórtex.
4. Incubar durante 5 minutos a 50°C.
5. Transferir la reacción al magneto y remover el sobrenadante (91 μL) a un nuevo tubo de 1.5 mL.

Purificación de ARNr previamente depletado

1. Añadir 160 μL de perlas Ampure XP al sobrenadante de la reacción anterior (91 μL). Homogenizar por pipeteo.
2. Incubar a temperatura ambiente por 15 min.
3. Transferir el tubo al magneto e incubar durante 5 min.
4. Remover y desechar el sobrenadante. Evitar agitar las perlas unidas al magneto.
5. Añadir 200 μL de EtOH al 80 % e incubar por 30 s. Remover y desechar el sobrenadante. Evitar agitar las perlas unidas al magneto.
6. Repetir el lavado con EtOH.
7. Secar a temperatura ambiente por 15 min.
8. Eluir en 12.5 μL de agua DEPC.

Síntesis de primer cadena de ADNc

1. Tomar 12 μL de ARN depleado de ARNr y añadir 2 μL del oligo AR2 (25 μM). Mezclar bien
2. Incubar la reacción en termociclador durante 2 min a 70°C y al finalizar pasar a hielo.
3. Realizar la siguiente reacción:
 - a) 2 μL de buffer 10X (AffinityScript RT)
 - b) 2 μL de DTT (0.1 M)
 - c) 0.8 μL de mezcla de dNTPS (25 mM)

d) 0.4 μL de inhibidor RNase Murine (40 U/ μL)

e) 0.8 μL de Enzima RT (AffinityScript RT)

f) 6 μL total

4. Añadir los 6 μL de la reacción RT a los 14 μL (ARN depletado + oligo AR2) en hielo.
5. Homogeneizar y centrifugar por 5 s.
6. Incubar durante 55 min a 55°C.

Degradación del ARN después a la transcripción inversa

1. Añadir 2 μL de NaOH 1 N.
2. Incubar por 12 minutos a 70°C.
3. Neutralizar con 4 μL de Ácido Acético al 0.5 M.
4. Volumen total= 26 μL .

Primer lavado de remoción después de la transcripción inversa

1. Añadir 2 μL de NaOH 1 N.
2. Añadir 14 μL de agua DEPC a la reacción (26 μL) para llevarla a un volumen total de 40 μL .
3. Transferir a un nuevo tubo de 1.5 mL (continua la acción degradadora del NaOH).
4. Añadir 2 volúmenes (80 μL) de perlas magnéticas RNAClean XP a la reacción (40 μL) en un nuevo tubo de 1.5 mL y homogeneizar.
5. Incubar a temperatura ambiente durante 15 min.
6. Colocar durante 5 min el tubo en el magneto.
7. Remover el sobrenadante evitando agitar las perlas.
8. Añadir 200 μL de EtOH al 70 % a las perlas unidas al magneto e incubar por 30 s.
9. Retirar y descartar el EtOH evitando agitar las perlas.
10. Añadir 200 μL de EtOH al 70 % fresco a las perlas unidas al magneto e incubar durante 30 s.

11. Retirar y descartar el EtOH evitando agitar las perlas y dejar secar durante 10 min.
12. Eluir con 5 μL de agua DEPC. No transferir a un nuevo tubo. Guardar las perlas magnéticas.

Segunda ligación

1. Añadir 2 μL del oligo 3Tr3 para el cDNA.
2. Calentar durante 3 minutos a 75°C.
3. **Reacción de ligación:** para varias muestras, la siguiente reacción puede ser preparada antes y añadir a el cDNA/3Tr3 en el hielo.
 - a) 2 μL de buffer de ligación T4 10x
 - b) 0.8 μL de DMSO (al 100%)
 - c) 0.2 μL de ATP (100 Mm)
 - d) 8.5 μL de PEG 8000 (al 50%)
 - e) 5 μL de ADNc
 - f) 2 μL de Oligo 3Tr3, 80 pmoles(2 μL de 40 uM de stock)
 - g) 1.5 μL de Ligasa T4 de ARN
 - h) 20 μL de Total
4. Mezclar por pipeteo 10 veces o por vórtex. **Nota:**la solución es viscosa.
5. Dar una centrifugación rápida.
6. Incubar toda la noche a 22°C.

Remoción de adaptadores

1. Incrementar 2 volúmenes (20 μL) de agua DEPC y llegar a un volumen final de 40 μL .
2. Añadir 2 volúmenes (80 μL) de perlas magnéticas.
3. Homogenizar con pipeteo.
4. Incubar a temperatura ambiente durante 15 min.
5. Colocar la reacción durante 5 min en el magneto, y al finalizar descartar el sobrenadante con precaución de no agitar las perlas del magneto.

6. Añadir 200 μL de EtOH al 70 % a las perlas aún unidas al magneto e incubar durante 30 s.
7. Remover el EtOH al 70 % con cuidado de no agitar las perlas.
8. Añadir 200 μL de EtOH al 70 % a las perlas aún unidas al magneto e incubar durante 30 s.
9. Remover el EtOH al 70 % con cuidado de no agitar las perlas.
10. Secar durante 10 min.
11. Eluir con 25 μL de agua DEPC y transferir a un nuevo tubo de 1.5 mL.

Segundo lavado para remoción de adaptadores

1. Añadir 1.5 volúmenes (37.5 μL) de perlas magnéticas RNAClean XP.
2. Homogeneizar con pipeteo.
3. Incubar la reacción durante 15 min a temperatura ambiente.
4. Colocar durante 5 min en el magneto.
5. Añadir 200 μL de EtOH al 70 % a las perlas unidas al magneto e incubar durante 30 s.
6. Remover el EtOH al 70 %.
7. Añadir 200 μL de EtOH al 70 % a las perlas unidas al magneto e incubar durante 30 s.
8. Remover el EtOH al 70 %.
9. Secar durante 2-3 min.
10. Eluir con 23 μL de agua DEPC.

PCR para secuenciación de Librería

1. Preparar la siguiente reacción:
 - a) 25 μL de NEBNext High Fidelity Master Mix 2x
 - b) 1 μL de oligo Universal para PCR Illumina
 - c) 1 μL de oligo Index (X) Illumina
 - d) 10 μL de ADNc

e) 13 μ L de Agua DEPC

f) 50 μ L de Total

2. Programación del termociclador:

a) 1 ciclo de 30 s a 98°C

b) 12 ciclos de:

1) 10 s a 98°C

2) 30 s a 60°C

3) 30 s a 72°C

c) 1 ciclo de 5 min a 72°C

d) Mantener a 4°C

A.2. Código

Para separar las muestras de cada librería simultánea:

```
$ cat sim_R1.fastq | fastx_barcode_splitter.pl - -partial 2 - -mismatches 1 -  
-bcfile barcodes.txt - -bol
```

Para obtener el R2 complementario para cada muestra:

```
$ seqtk subseq sim_R2.fastq ids_compl_R1.txt > R2.fastq
```

Deduplicación de lecturas:

```
$ cd-hit-dup -i R1.fq -i2 R2.fq -o output-R1.fq -o2 output-R2.fq
```

Para remoción de adaptadores y curado de calidad por ventanas:

```
$ java -jar trimmomatic-0.35.jar PE -phred33 input_foward.fastq input_reverse.fastq  
output_foward.fastq output_reverse.fastq ILLUMINACLIP:TruSeq3-PE.fa:2:30:10  
SLIDINGWINDOW:4:20
```

Quitar ribosomales:

```
$ ribopicker.pl -dbs ssr,lsr -out_dir
```

Quitar transcritos humanos:

```
$ hisat -x hg19_hisat.index -1 R1.fastq -2 R2.fastq
```

Asignación de taxonomía con kraken2:

```
$ kraken -db $standard_database sample.fasta
```

Anotación funcional con HUMAnN2:

```
$ humann2 - -input sample.fasta - -output sample_humann2 - -output-format  
biom - -remove-stratified-output - -output-max-decimals 0
```

Bibliografía

- [1] Shirley Bikel, Alejandra Valdez-Lara, Fernanda Cornejo-Granados, Karina Rico, Samuel Canizales-Quinteros, Xavier Soberón, Luis Del Pozo-Yauner, and Adrián Ochoa-Leyva. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. *Computational and Structural Biotechnology Journal*, 13:390–401, 2015. [ix](#), [1](#), [3](#), [9](#), [10](#), [11](#)
- [2] Andrew S. Neish. Microbes in Gastrointestinal Health and Disease. *Gastroenterology*, 136(1):65–80, jan 2009. [1](#), [4](#)
- [3] Jose C. Clemente, Luke K. Ursell, Laura Wegener Parfrey, and Rob Knight. The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell*, 148(6):1258–1270, mar 2012. [1](#), [4](#)
- [4] Matthias H Tschöp, Philip Hugenholtz, and Christopher L Karp. Getting to the core of the gut microbiome. *Nature Biotechnology*, 27(4):344–346, apr 2009. [1](#), [5](#), [8](#)
- [5] Eric A Franzosa, Xochitl C Morgan, Nicola Segata, Levi Waldron, Joshua Reyes, Ashlee M Earl, Georgia Giannoukos, Matthew R Boylan, Dawn Ciulla, Dirk Gevers, Jacques Izard, Wendy S Garrett, Andrew T Chan, and Curtis Huttenhower. Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences*, 111(22):E2329–E2338, jun 2014. [1](#), [2](#), [5](#), [9](#), [10](#), [12](#), [50](#)
- [6] María José Gosalbes, Ana Durbán, Miguel Pignatelli, Juan José Abellan, Nuria Jiménez-Hernández, Ana Elena Pérez-Cobas, Amparo Latorre, and Andrés Moya. Metatranscriptomic Approach to Analyze the Functional Human Gut Microbiota. *PLoS ONE*, 6(3):e17447, mar 2011. [1](#), [10](#)
- [7] Ruth E Ley, F. Backhed, Peter Turnbaugh, Catherine a Lozupone, Robin D Knight, and Jeffrey I Gordon. Obesity alters gut microbial ecology. *Procee-*

- dings of the National Academy of Sciences*, 102(31):11070–11075, aug 2005. [1](#), [2](#), [8](#)
- [8] Ensanut. Ensanut 2012. *Instituto Nacional de Salud Pública*, 68(5):200, feb 2012. [1](#), [6](#)
- [9] S. D. de Ferranti. Prevalence of the Metabolic Syndrome in American Adolescents: Findings From the Third National Health and Nutrition Examination Survey. *Circulation*, 110(16):2494–2497, oct 2004. [1](#), [2](#), [6](#)
- [10] Maria Lola Evia-Viscarra, Edel Rafael Rodea-Montero, Evelia Apolinar-Jimenez, and Silvia Quintana-Vargas. Metabolic syndrome and its components among obese (BMI \geq 95th) Mexican adolescents. *Endocrine Connections*, 2(4):208–215, nov 2013. [1](#), [6](#)
- [11] Peter J Turnbaugh, Micah Hamady, Tanya Yatsunencko, Brandi L Cantarel, Alexis Duncan, Ruth E Ley, Mitchell L Sogin, William J Jones, Bruce A Roe, Jason P Affourtit, Michael Egholm, Bernard Henrissat, Andrew C Heath, Rob Knight, and Jeffrey I Gordon. A core gut microbiome in obese and lean twins. *Nature*, 457(7228):480–484, jan 2009. [2](#), [7](#), [8](#)
- [12] William a Walters, Zech Xu, and Rob Knight. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Letters*, 588(22):4223–4233, nov 2014. [2](#)
- [13] Nathalie M. Delzenne, Audrey M. Neyrinck, Fredrik Bäckhed, and Patrice D. Cani. Targeting gut microbiota in obesity: effects of prebiotics and probiotics. *Nature Reviews Endocrinology*, 7(11):639–646, aug 2011. [2](#), [5](#), [10](#)
- [14] Ran Blekhman, Julia K. Goodrich, Katherine Huang, Qi Sun, Robert Bukowski, Jordana T. Bell, Timothy D. Spector, Alon Keinan, Ruth E. Ley, Dirk Gevers, and Andrew G. Clark. Host genetic variation impacts microbiome composition across human body sites. *Genome Biology*, 16(1):191, dec 2015. [2](#)
- [15] Eija Korpelainen, Jarno Tuimala, Panu Somervuo, Mikael Huss, and Garry Wong. *RNA-seq Data Analysis A Practical Approach*. 2014. [ix](#), [2](#), [11](#), [12](#)
- [16] Jane Peterson, Susan Garges, Maria Giovanni, Pamela McInnes, Lu Wang, Jeffery a. Schloss, Vivien Bonazzi, Jean E. McEwen, Kris a. Wetterstrand, Carolyn Deal, Carl C. Baker, Valentina Di Francesco, T. Kevin Howcroft, Robert W. Karp, R. Dwayne Lunsford, Christopher R. Wellington, Tsegahiwot

-
- Belachew, Michael Wright, Christina Giblin, Hagit David, Melody Mills, Rachele Salomon, Christopher Mullins, Beena Akolkar, Lisa Begg, Cindy Davis, Lindsey Grandison, Michael Humble, Jag Khalsa, A. R. Little, Hannah Peavy, Carol Pontzer, Matthew Portnoy, Michael H. Sayre, Pamela Starke-Reed, Samir Zakhari, Jennifer Read, Bracie Watson, and Mark Guyer. The NIH Human Microbiome Project. *Genome Research*, 19(12):2317–2323, dec 2009. [3](#), [4](#)
- [17] Wendy S. Garrett. From cell biology to the microbiome: An intentional infinite loop. *The Journal of Cell Biology*, 210(1):7–8, jul 2015. [ix](#), [3](#)
- [18] Curtis Huttenhower, Dirk Gevers, Rob Knight, Sahar Abubucker, Jonathan H. Badger, Asif T. Chinwalla, Heather H. Creasy, Ashlee M. Earl, Michael G. FitzGerald, Robert S. Fulton, Michelle G. Giglio, Kymberlie Hallsworth-Pepin, Elizabeth a. Lobos, Ramana Madupu, Vincent Magrini, John C. Martin, Makedonka Mitreva, Donna M. Muzny, Erica J. Sodergren, James Versalovic, Aye M. Wollam, Kim C. Worley, Jennifer R. Wortman, Sarah K. Young, Qiandong Zeng, Kjersti M. Aagaard, Olukemi O. Abolude, Emma Allen-Vercoe, Eric J. Alm, Lucia Alvarado, Gary L. Andersen, Scott Anderson, Elizabeth Appelbaum, Harindra M. Arachchi, Gary Armitage, Cesar a. Arze, Tulin Ayvaz, Carl C. Baker, Lisa Begg, Tsegahiwot Belachew, Veena Bhonagiri, Monika Bihan, Martin J. Blaser, Toby Bloom, Vivien Bonazzi, J. Paul Brooks, Gregory a. Buck, Christian J. Buhay, Dana a. Busam, Joseph L. Campbell, Shane R. Canon, Brandi L. Cantarel, Patrick S. G. Chain, I-Min a. Chen, Lei Chen, Shaila Chhibba, Ken Chu, Dawn M. Ciulla, Jose C. Clemente, Sandra W. Clifton, Sean Conlan, Jonathan Crabtree, Mary a. Cutting, Noam J. Davidovics, Catherine C. Davis, Todd Z. DeSantis, Carolyn Deal, Kimberley D. Delehaunty, Floyd E. Dewhirst, Elena Deych, Yan Ding, David J. Dooling, Shannon P. Dugan, Wm Michael Dunne, a. Scott Durkin, Robert C. Edgar, Rachel L. Erlich, Candace N. Farmer, Ruth M. Farrell, Karoline Faust, Michael Feldgarden, Victor M. Felix, Sheila Fisher, Anthony a. Fodor, Larry J. Forney, Leslie Foster, Valentina Di Francesco, Jonathan Friedman, Dennis C. Friedrich, Catrina C. Fronick, Lucinda L. Fulton, Hongyu Gao, Nathalia Garcia, Georgia Giannoukos, Christina Giblin, Maria Y. Giovanni, Jonathan M. Goldberg, Johannes Goll, Antonio Gonzalez, Allison Griggs, Sharvari Gujja, Susan Kinder Haake, Brian J. Haas, Holli a. Hamilton, Emily L. Harris, Theresa a. Hepburn, Brandi Herter, Diane E. Hoffmann, Michael E. Holder, Clinton Howarth, Katherine H. Huang, Susan M. Huse, Jacques Izard, Janet K. Jansson, Huaiyang Jiang, Catherine Jordan, Vandita Joshi, James a. Katancik, Wendy a. Keitel, Scott T. Kelley, Cristyn Kells, Nicholas B. King, Dan Knights, Heidi H. Kong, Omry
-

-
- Koren, Sergey Koren, Karthik C. Kota, Christie L. Kovar, Nikos C. Kyrpides, Patricio S. La Rosa, Sandra L. Lee, Katherine P. Lemon, Niall Lennon, Cecil M. Lewis, Lora Lewis, Ruth E. Ley, Kelvin Li, Konstantinos Liolios, Bo Liu, Yue Liu, Chien-Chi Lo, Catherine a. Lozupone, R. Dwayne Lunsford, Tessa Madden, Anup a. Mahurkar, Peter J. Mannon, Elaine R. Mardis, Victor M. Markowitz, Konstantinos Mavromatis, Jamison M. McCorrison, Daniel McDonald, Jean McEwen, Amy L. McGuire, Pamela McInnes, Teena Mehta, Kathie a. Mihindukulasuriya, Jason R. Miller, Patrick J. Minx, Irene Newsham, Chad Nusbaum, Michelle O’Laughlin, Joshua Orvis, Ioanna Paganis, Krishna Palaniappan, Shital M. Patel, Matthew Pearson, Jane Peterson, Mircea Podar, Craig Pohl, Katherine S. Pollard, Mihai Pop, Margaret E. Priest, Lita M. Proctor, Xiang Qin, Jeroen Raes, Jacques Ravel, Jeffrey G. Reid, Mina Rho, Rosamond Rhodes, Kevin P. Riehle, Maria C. Rivera, Beltran Rodriguez-Mueller, Yu-Hui Rogers, Matthew C. Ross, Carsten Russ, Ravi K. Sanka, Pamela Sankar, J. Fah Sathirapongsasuti, Jeffery a. Schloss, Patrick D. Schloss, Thomas M. Schmidt, Matthew Scholz, Lynn Schriml, Alyxandria M. Schubert, Nicola Segata, Julia a. Segre, William D. Shannon, Richard R. Sharp, Thomas J. Sharpton, Narmada Shenoy, Nihar U. Sheth, Gina a. Simone, Indresh Singh, Christopher S. Smillie, Jack D. Sobel, Daniel D. Sommer, Paul Spicer, Granger G. Sutton, Sean M. Sykes, Diana G. Tabbaa, Mathangi Thiagarajan, Chad M. Tomlinson, Manolito Torralba, Todd J. Treangen, Rebecca M. Truty, Tatiana a. Vishnivetskaya, Jason Walker, Lu Wang, Zhengyuan Wang, Doyle V. Ward, Wesley Warren, Mark a. Watson, Christopher Wellington, Kris a. Wetterstrand, James R. White, Katarzyna Wilczek-Boney, YuanQing Wu, Kristine M. Wylie, Todd Wylie, Chandri Yandava, Liang Ye, Yuzhen Ye, Shibu Yooseph, Bonnie P. Youmans, Lan Zhang, Yanjiao Zhou, Yiming Zhu, Laurie Zoloth, Jeremy D. Zucker, Bruce W. Birren, Richard a. Gibbs, Sarah K. Highlander, Barbara a. Methé, Karen E. Nelson, Joseph F. Petrosino, George M. Weinstock, Richard K. Wilson, and Owen White. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207–214, 2012. [3](#), [8](#)
- [19] J. K. Nicholson, E. Holmes, J. Kinross, R. Burcelin, G. Gibson, W. Jia, and S. Pettersson. Host-Gut Microbiota Metabolic Interactions. *Science*, 336(6086):1262–1267, jun 2012. [4](#), [5](#)
- [20] Baohong Wang, Mingfei Yao, Longxian Lv, Zongxin Ling, and Lanjuan Li. The Human Microbiota in Health and Disease. *Engineering*, 3(1):71–82, feb 2017. [ix](#), [5](#)
- [21] Oecd Indicators. *Health at a Glance 2013*. Health at a Glance. OECD Publishing, nov 2013. [6](#)
-

-
- [22] Alan W Walker and Julian Parkhill. Fighting Obesity with Bacteria. *Science*, 341(6150):1069–1070, sep 2013. [ix](#), [8](#)
- [23] Lieke WJ van den Elsen, Hazel C Poyntz, Laura S Weyrich, Wayne Young, and Elizabeth E Forbes-Blom. Embracing the gut microbiota: the new frontier for inflammatory and infectious diseases. *Clinical & Translational Immunology*, 6(1):e125, 2017. [8](#), [9](#), [10](#)
- [24] Alexander A Shishkin, Georgia Giannoukos, Alper Kucukural, Dawn Ciulla, Michele Busby, Christine Surka, Jenny Chen, Roby P Bhattacharyya, Robert F Rudy, Miles M Patel, Nathaniel Novod, Deborah T Hung, Andreas Gnirke, Manuel Garber, Mitchell Guttman, and Jonathan Livny. Simultaneous generation of many RNA-seq libraries in a single reaction. *Nature Methods*, 12(4):323–325, mar 2015. [9](#), [11](#), [16](#)
- [25] Corinne Ferrier Maurice, Henry Joseph Haiser, and Peter James Turnbaugh. Xenobiotics Shape the Physiology and Gene Expression of the Active Human Gut Microbiome. *Cell*, 152(1-2):39–50, jan 2013. [12](#)
- [26] Galeb S. Abu-Ali, Raaj S. Mehta, Jason Lloyd-Price, Himel Mallick, Tobyn Branck, Kerry L. Ivey, David A. Drew, Casey DuLong, Eric Rimm, Jacques Izard, Andrew T. Chan, and Curtis Huttenhower. Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nature Microbiology*, 3(3):356–366, mar 2018. [13](#)
- [27] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, aug 2014. [18](#)
- [28] Beifang Niu, Limin Fu, Shulei Sun, and Weizhong Li. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics*, 11(1):187, 2010. [18](#)
- [29] Robert Schmieder, Yan Wei Lim, and Robert Edwards. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics*, 28(3):433–435, feb 2012. [18](#)
- [30] Daehwan Kim, Ben Langmead, and Steven L. Salzberg. HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4):357–360, 2015. [18](#)
- [31] Derrick E. Wood and Steven L. Salzberg. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, 15(3):R46, 2014. [20](#)
-

- [32] Eric A. Franzosa, Lauren J. McIver, Gholamali Rahnavard, Luke R. Thompson, Melanie Schirmer, George Weingart, Karen Schwarzberg Lipson, Rob Knight, J. Gregory Caporaso, Nicola Segata, and Curtis Huttenhower. Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, 15(11):962–968, nov 2018. [20](#)
- [33] Nicola Segata, Jacques Izard, Levi Waldron, Dirk Gevers, Larisa Miropolsky, Wendy S Garrett, and Curtis Huttenhower. Metagenomic biomarker discovery and explanation. *Genome biology*, 12(6):R60, jun 2011. [20](#)
- [34] Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, jan 2010. [20](#)
- [35] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, dec 2014. [20](#)
- [36] G Rizzatti, L R Lopetuso, G Gibiino, C Binda, and A Gasbarrini. Proteobacteria: A Common Factor in Human Diseases. *BioMed Research International*, 2017:1–7, 2017. [27](#)
- [37] Siobhan F. Clarke, Eileen F. Murphy, Kanishka Nilaweera, Paul R. Ross, Fergus Shanahan, Paul W. O’Toole, and Paul D. Cotter. The gut microbiota and its relationship to diet and obesity. *Gut Microbes*, 3(3):186–202, may 2012. [27](#), [33](#), [50](#)
- [38] Weiwei Jiang, Na Wu, Xuemei Wang, Yujing Chi, Yuanyuan Zhang, Xinyun Qiu, Ying Hu, Jing Li, and Yulan Liu. Dysbiosis gut microbiota associated with inflammation and impaired mucosal immune function in intestine of humans with non-alcoholic fatty liver disease. *Scientific Reports*, 5(1):8096, jul 2015. [27](#)
- [39] Peter J. Turnbaugh, Fredrik Bäckhed, Lucinda Fulton, and Jeffrey I. Gordon. Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host and Microbe*, 3(4):213–223, 2008. [29](#), [50](#)
- [40] Stavros Bashiardes, Hagit Shapiro, Shachar Rozin, Oren Shibolet, and Eran Elinav. Non-alcoholic fatty liver and the gut microbiota. *Molecular Metabolism*, 5(9):782–794, sep 2016. [30](#), [50](#)
- [41] Brandilyn A. Peters, Jean A. Shapiro, Timothy R. Church, George Miller, Chau Trinh-Shevrin, Elizabeth Yuen, Charles Friedlander, Richard B. Hayes,

- and Jiyoung Ahn. A taxonomic signature of obesity in a large study of American adults. *Scientific Reports*, 8(1):9749, dec 2018. [31](#)
- [42] Vallabh Shah, Stacey M Lambeth, Trechelle Carson, Janae Lowe, Thiruvarangan Ramaraj, Jonathan W. Leff, Li Luo, and Callum J Bell. Composition Diversity and Abundance of Gut Microbiome in Prediabetes and Type 2 Diabetes. *Journal of Diabetes and Obesity*, 2(2):108–114, dec 2015. [31](#)
- [43] Ya-Ping Hou, Qing-Qing He, Hai-Mei Ouyang, Hai-Shan Peng, Qun Wang, Jie Li, Xiao-Fei Lv, Yi-Nan Zheng, Shao-Chuan Li, Hai-Liang Liu, and Ai-Hua Yin. Human Gut Microbiota Associated with Obesity in Chinese Children and Adolescents. *BioMed Research International*, 2017:1–8, 2017. [33](#)
- [44] Mora Murri, Isabel Leiva, Juan Miguel Gomez-Zumaquero, Francisco J. Tinahones, Fernando Cardona, Federico Soriguer, and María Isabel Queipo-Ortuño. Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. *BMC Medicine*, 11(1):46, dec 2013. [33](#)
- [45] Janet C. Onishi, Sara Campbell, Michael Moreau, Falshruti Patel, Andrew I. Brooks, Yin Xiu Zhou, Max M. Häggblom, and Judith Storch. Bacterial communities in the small intestine respond differently to those in the caecum and colon in mice fed low- and high-fat diets. *Microbiology*, 163(8):1189–1197, aug 2017. [33](#)
- [46] Aarti Kathrani, Jennifer A. Larsen, Philip H. Kass, and Andrea J. Fascetti. Effect of short-term probiotic *Enterococcus faecium* SF68 dietary supplementation in overweight and obese cats without comorbidities. *Veterinary Record Open*, 3(1):e000164, apr 2016. [50](#)
- [47] Olli Turta and Samuli Rautava. Antibiotics, obesity and the link to microbes - what are we doing to our children? *BMC Medicine*, 14(1):57, dec 2016. [51](#)