UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

IDENTIFICACIÓN DE PÉPTIDOS ANTIMICROBIANOS (AMPs) SOBREEXPRESADOS POR LA MICROBIOTA INTESTINAL ASOCIADA A OBESIDAD Y SÍNDROME METABÓLICO

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:
LUIGUI MICHEL GALLARDO BECERRA

TUTOR PRINCIPAL:
DR. ADRIÁN OCHOA LEYVA
Instituto de Biotecnología, UNAM

MIEMBROS DEL COMITÉ TUTOR:
DR. SAMUEL CANIZALES QUINTEROS
Facultad de Química, UNAM
DR. LUIS DAVID ALCARÁZ
Facultad de Ciencias, UNAM

Cuernavaca, Mor. Marzo, 2026

Índice

Índice de Figuras

Índice de Tablas

Miembros del Jurado

PRESIDENTE

Dra. Liliana Pardo López

Instituto de Biotecnología, UNAM

SECRETARIO

Dra. Luz de María Breton Deval

Instituto de Biotecnología, UNAM

VOCAL

Dr. Gabriel Del Río Guerra

Instituto de Fisiología Celular, UNAM

VOCAL

Dra. Daniela Elizabeth Ledezma Tejeida

Centro de Ciencias Genómicas, UNAM

VOCAL

Dr. Gamaliel López Leal

Centro de Investigación en Dinámica Celular, UAEM

El presente trabajo se realizó bajo la dirección del Dr. Adrián Ochoa Leyva en el Laboratorio 22 del Departamento de Microbiología Molecular del Instituto de Biotecnología de la Universidad Nacional Autónoma de México.

Para su realización, contó con el financiamiento del Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) con la beca de doctorado (CVU 778192) y los proyectos SALUD-2014-C01-234188 y Ciencia de Frontera-2019–263986. Así mismo contó con financiamiento de la Universidad Nacional Autónoma de México con el programa DGAPA PAPIIT UNAM (IN222526 e IN219723), del Programa de Doctorado en Ciencias Bioquímicas y del Programa de Apoyo a los Estudios de Posgrado (PAEP).

Agradecimientos

Primero me gustaría agradecer a mi tutor, el Dr. Adrián Ochoa Leyva, por aceptarme en su laboratorio, por su confianza y por guiarme en mi formación científica. Gracias por enseñarme que esta profesión va más allá de los experimentos y los números; que las historias detrás de nuestro quehacer científico y su impacto social son fundamentales para construir un país mejor.

Agradezco a los miembros de mi comité tutor, el Dr. Samuel Canizales Quinteros y el Dr. Luis David Alcaráz, quienes me guiaron para dar forma a este proyecto con sus valiosas contribuciones en cada uno de mis tutorales. De igual manera, expreso mi gratitud a los miembros del jurado evaluador: Dra. Liliana Pardo López, Dra. Luz de María Breton Deval, Dr. Gabriel Del Río Guerra, Dra. Daniela Elizabeth Ledezma Tejeida y Dr. Gamaliel López Leal, por su retroalimentación en mi examen de candidatura y la atenta revisión de este manuscrito.

Gracias a todos los integrantes y exintegrantes del laboratorio del Dr. Ochoa, que más allá de ser compañeros se han convertido en amigos con quienes crecí, aprendí y compartí grandes momentos. A mis mentores "no oficiales", de quienes aprendí mucho en el día a día: los postdocs del laboratorio, Fer, Gama y Rodrigo. A mis amigos: los de toda la vida, los que me acompañaron desde la maestría y aquellos que se sumaron durante el doctorado en el Instituto. Una mención especial a los miembros de nuestro club: Diana, Nori, Tonalli, Víctor, Rodo, Ricardo y Cristina.

Deseo reconocer también a quienes apoyaron la parte técnica de este trabajo: Dra. Alejandra Valdez Lara, Dra. Fernanda Cornejo Granados y al Biol. Filiberto Sánchez López —a quien agradezco también su apoyo personal—, por su colaboración en la estandarización de la extracción de ARN y la preparación de librerías. Igualmente, a Gloria Tanahiry Vázquez Castro y al Dr. Ricardo Alfredo Grande, de la Unidad de Secuenciación Masiva y Bioinformática del Instituto de Biotecnología, por su asistencia en la validación de calidad del ARN y la secuenciación de las muestras.

Asimismo, agradezco a Alejandro Mendoza-Vargas y a la Unidad de Secuenciación Masiva del Instituto Nacional de Medicina Genómica por el soporte técnico brindado. Al M. en T. I. Juan Manuel Hurtado Ramírez, de la Unidad de Cómputo del Instituto de Biotecnología, por su ayuda en la administración del servidor. Al Lic. J. Antonio Bolanos Guillen y a la M. en C. Lidia Judith Martínez Carrera, de la Unidad de Docencia, por facilitar los trámites durante el doctorado y el proceso de titulación.

Por último, agradezco al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por otorgarme la beca de doctorado (CVU: 778192) para dedicarme de tiempo completo al posgrado, y por el financiamiento de los proyectos SALUD-2014-C01-234188 y Ciencia de Frontera-2019–263986. También agradezco a la Universidad Nacional Autónoma de México por el apoyo mediante el programa DGAPA PAPIIT UNAM (IA203118, IN215520 e IN219723) y al Programa de Apoyo a los Estudios de Posgrado (PAEP) por el financiamiento otorgado para asistir a un congreso a presentar mi trabajo.

Dedicatoria

Dedico esta tesis a mi familia, que fue mi primera escuela. A mi madre, quien es y será siempre mi faro en los momentos de incertidumbre: ¡por fin lo logramos! A mi esposa Jasmin, por aceptar estar conmigo, por su comprensión y apoyo para ayudarme a llegar más lejos. Fuiste mi apoyo incondicional en esta complicada recta final: gracias por seguir creyendo en mí.

Resumen

La microbiota intestinal es esencial para la salud del hospedero, y los desequilibrios en su composición o función se han asociado con la obesidad infantil y el Síndrome Metabólico (SM), ambos considerados problemas de salud pública a nivel mundial, ocupando México los primeros lugares. Entre los mecanismos que regulan la estructura de la comunidad microbiana destacan los péptidos antimicrobianos (AMPs), producidos por la propia microbiota, los cuales desempeñan un papel clave en la competencia microbiana y en el mantenimiento del equilibrio ecológico en el intestino humano.

El objetivo principal de este estudio fue determinar la expresión de AMPs en la microbiota intestinal de niñas y niños con obesidad (O) y con obesidad asociada al Síndrome Metabólico (OMS), utilizando un enfoque metatranscriptómico basado en *RNA-seq*. Primero caracterizamos la microbiota de ambos grupos mediante el marcador 16s, confirmando la presencia de cambios en la composición bacteriana entre ambos. Con ello, se identificaron biomarcadores microbianos asociados a cada grupo que mostraron correlaciones significativas con los parámetros bioquímicos y antropométricos de los participantes. Posteriormente, mediante el análisis *RNA-seq* se generó un ensamblado *de novo* y, con el uso de diferentes herramientas de predicción, se identificaron 51 candidatos a AMPs derivados de marcos de lectura pequeños (smORFs). La mayoría presentó homólogos bacterianos, mientras que ocho se asociaron con elementos genéticos móviles (siete con bacteriófagos y uno con un plásmido), destacando a estos elementos como reservorios de funciones antimicrobianas en la microbiota. El análisis conjunto del viroma de las mismas muestras confirmó que estos AMPs son producidos por virus activos y no por bacterias hospedadoras. De igual forma, se detectó que el plásmido que codifica uno de los AMPs de interés presentaba expresión del resto de sus genes. Ambos resultados evidencian el origen móvil de los AMPs y su actividad en el ecosistema intestinal. Particularmente, se detectaron seis AMPs diferencialmente expresados entre O y OMS, cuya expresión se correlacionó inversamente con la abundancia de taxones bacterianos específicos, lo que sugiere un papel potencial en las alteraciones microbianas vinculadas con la enfermedad. Dichos AMPs también se detectaron en un conjunto independiente de 372 metatranscriptomas intestinales humanos, lo que sugiere que su presencia es común y conservada en el microbioma intestinal humano.

La validación experimental de uno de los AMPs codificados por un bacteriófago, el AMP 3020, en sus dos variantes (ADR1 y ADR2), cuyo origen viral fue confirmado mediante el análisis del viroma intestinal, demostró su actividad antibacteriana *in vitro* frente a bacterias Gram positivas y Gram negativas, sin evidenciar citotoxicidad en células T humanas. En conjunto, los hallazgos de este estudio evidencian el papel funcional de los AMPs, incluidos aquellos codificados en elementos genéticos móviles, en la microbiota intestinal asociada a la obesidad y al SM, y los posicionan como candidatos prometedores para el desarrollo de biomarcadores y terapias dirigidas al microbioma, con potencial para contrarrestar la disbiosis vinculada a estas condiciones.

Abstract

The gut microbiota is essential for host health, and imbalances in its composition or function have been associated with childhood obesity and Metabolic Syndrome (MS), both considered major global public health concerns, particularly in Mexico, where prevalence rates are among the highest in the world. Among the mechanisms that regulate the structure of the microbial community, antimicrobial peptides (AMPs) produced by the microbiota play a key role in microbial competition and in maintaining the ecological balance of the human intestine.

The main objective of this study was to determine the expression of AMPs in the gut microbiota of children with obesity (O) and obesity associated with MS (OMS) using an RNA-seq–based metatranscriptomic approach. First, we characterized the microbiota of both groups through 16S rRNA profiling, confirming compositional differences between them. Microbial biomarkers associated with each group were identified, showing significant correlations with participants' biochemical and anthropometric parameters. Subsequently, de novo assembly and multiple prediction tools were used to identify 51 candidate AMPs derived from small open reading frames (smORFs). Most exhibited bacterial homologs, while eight were associated with mobile genetic elements, seven with bacteriophages and one with a plasmid, highlighting these elements as reservoirs of antimicrobial functions within the microbiota.

Combined analysis of the virome from the same samples confirmed that these AMPs are produced by active viruses rather than bacterial hosts. Similarly, the plasmid encoding one AMP showed gene expression across the rest of its genome. Both results support the mobile origin of these AMPs and their transcriptional activity within the intestinal ecosystem. Six AMPs were differentially expressed between O and OMS, whose expression correlated inversely with the abundance of specific bacterial taxa, suggesting a potential role in the microbial alterations associated with disease. These AMPs were also detected in an independent dataset of 372 human gut metatranscriptomes, indicating that their presence is widespread and conserved in the human microbiome.

Experimental validation of one phage-encoded AMP, AMP 3020, in its two variants (ADR1 and ADR2), whose viral origin was confirmed by virome analysis, demonstrated in vitro antibacterial activity against both Gram-positive and Gram-negative bacteria, with no evidence of cytotoxicity in human T cells. Collectively, these findings demonstrate the functional role of AMPs, including those encoded in mobile genetic elements, within the gut microbiota associated with obesity and MS. Moreover, they highlight these peptides as promising candidates for the development of microbiome-based biomarkers and therapeutic strategies aimed at counteracting dysbiosis linked to these metabolic conditions.

Introducción

Microbiota intestinal y su relación con la obesidad y el Síndrome Metabólico (SM)

La obesidad infantil y el Síndrome Metabólico (SM) representan un problema creciente de salud pública a nivel mundial [1], y México es uno de los países más afectados. En 2023, la Organización para la Cooperación y el Desarrollo Económicos (OCDE) reportó que México se mantiene entre los países con mayores niveles de obesidad dentro de sus estados miembro, solo por debajo de Estados Unidos, con una prevalencia de obesidad en adultos superior al 36% [2]. De acuerdo con la Encuesta Nacional de Salud y Nutrición (ENSANUT) Continua 2020–2022, publicada en 2023, cerca del 38% de los niños y adolescentes mexicanos presentan sobrepeso u obesidad, lo que confirma una tendencia ascendente respecto a reportes previos [3]. Una parte de estos casos podría cursar con SM no diagnosticado [4], lo que resalta la urgencia de implementar estrategias específicas de prevención y tratamiento para esta población.

El SM se define por la presencia conjunta de al menos tres de las siguientes alteraciones cardiometabólicas: hiperglucemia, hipertrigliceridemia, dislipidemia (caracterizada por niveles bajos de colesterol HDL) e hipertensión [5], [6] (Figura 1). En población pediátrica, su diagnóstico se establece mediante puntos de corte específicos ajustados por edad, sexo y talla [7]. La acumulación de grasa corporal durante la infancia aumenta la probabilidad de desarrollar, en la adultez, diabetes tipo 2, enfermedades cardiovasculares y hepáticas, entre otras complicaciones metabólicas [1], [8]. Estas condiciones no solo impactan la calidad de vida, sino que también aumentan el riesgo de discapacidad o muerte prematura [1].

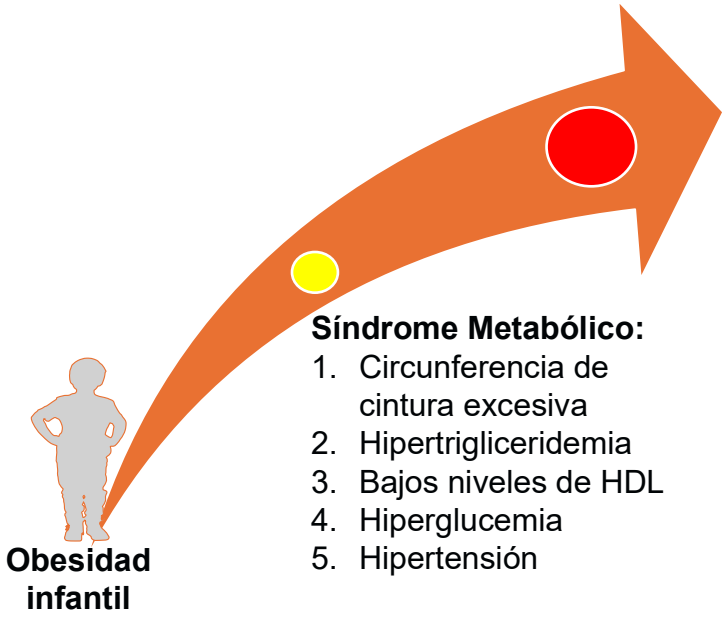


Figura 1: Progresión de la obesidad infantil hacia el Síndrome Metabólico (SM) y enfermedades crónicas. La obesidad durante la infancia favorece el desarrollo del SM, caracterizado por alteraciones metabólicas como el exceso de grasa abdominal, la hipertrigliceridemia, los bajos niveles de colesterol HDL, la hiperglucemia y la hipertensión. A su vez, este síndrome incrementa el riesgo de padecer enfermedades crónicas en la edad adulta, incluyendo el hígado graso no alcohólico, la diabetes mellitus tipo 2, el cáncer gastrointestinal, hepático o renal, así como enfermedades cardiovasculares y trastornos musculoesqueléticos.

La obesidad y el SM son condiciones multifactoriales influenciadas por factores biológicos, genéticos, ambientales y del estilo de vida [9], [10]. Entre estos, la microbiota intestinal destaca como un elemento central [11], [12], [13], [14]. Para mantener la salud del hospedero, se requiere un ecosistema intestinal diverso y equilibrado, en el que la microbiota cumple funciones esenciales, como modular la respuesta inmune, favorecer la absorción de nutrientes, producir vitaminas y regular el metabolismo energético del hospedero [15], [16] (Figura 2). Cuando este equilibrio se altera, se presenta un estado de disbiosis, definida como un desequilibrio en la composición o en la función de la microbiota respecto a su estado normal, el cual se ha asociado de forma consistente con la obesidad y el SM [17], [18].

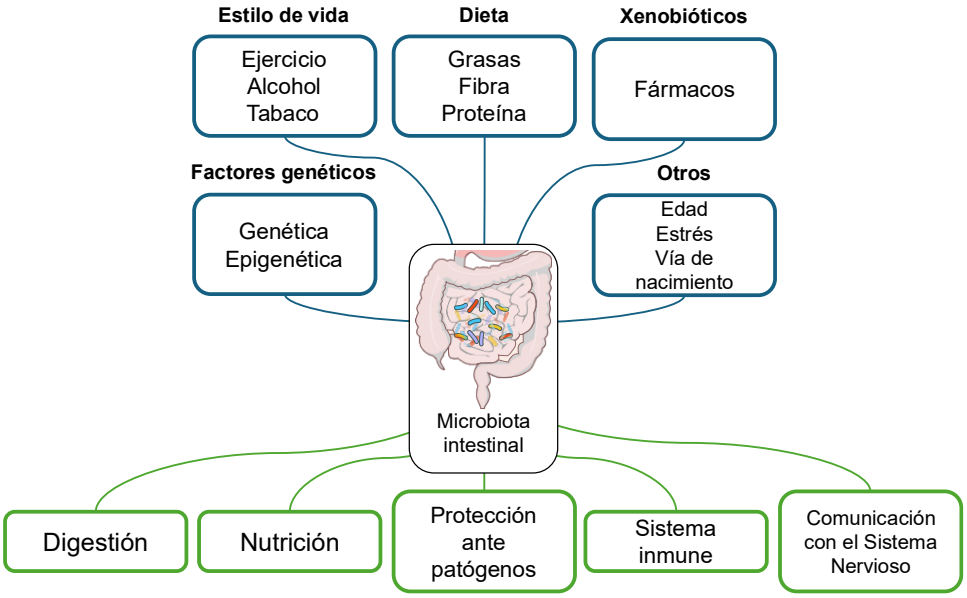


Figura 2: Factores que influyen en la composición y función de la microbiota intestinal. Diversos elementos modulan su estructura y actividad, incluyendo el estilo de vida (ejercicio, consumo de alcohol y tabaco), la dieta (grasas, fibra y proteínas), la exposición a xenobióticos (fármacos) y factores genéticos y epigenéticos del hospedero. Otros aspectos, como la edad, el estrés y la vía de nacimiento, también modelan su composición. En conjunto, la microbiota intestinal participa en funciones clave como la digestión, la nutrición, la defensa ante patógenos, la regulación inmune y la comunicación con el sistema nervioso.

La disbiosis puede aumentar el aprovechamiento de energía de la dieta, modificar el perfil de lípidos y favorecer un estado de inflamación crónica de bajo grado, rasgos observados en personas con obesidad y SM [11], [19]. Estudios iniciales reportaron que, en individuos con obesidad, aumenta la proporción del filo Firmicutes y disminuye Bacteroidetes, lo que se ha asociado a una mayor capacidad para degradar carbohidratos y lípidos complejos, y con ello a un mayor almacenamiento de energía [12], [13]. Además de este efecto, la microbiota intestinal participa en la regulación del metabolismo energético, enviando señales que influyen en la acumulación de tejido adiposo y en la sensación de saciedad a través de su interacción con el sistema nervioso [20], [21].

En años recientes, los estudios metagenómicos y metatranscriptómicos han mostrado que, además de los cambios en la composición microbiana, existen alteraciones funcionales en la microbiota asociadas con enfermedades metabólicas [22],

[23], [24], [25]. Estas incluyen la sobreexpresión de genes vinculados con la obtención y conversión de energía, la biosíntesis de lípidos y el transporte de lipopolisacáridos, procesos que pueden favorecer tanto el almacenamiento energético como la inflamación sistémica [11], [26], [27], [28]. Estas diferencias funcionales subrayan la importancia de evaluar la actividad transcripcional del microbioma para comprender su papel en el desarrollo de enfermedades metabólicas [22], [24], [27], [29]. En este contexto, resulta esencial no solo conocer la composición microbiana, sino también caracterizar su actividad funcional, es decir, los genes que se expresan y participan en los procesos metabólicos relacionados con la obesidad y el SM.

Enfoques funcionales para el estudio del microbioma

El estudio del microbioma intestinal ha evolucionado desde enfoques descriptivos hacia aproximaciones funcionales que permiten caracterizar la actividad metabólica y regulatoria de las comunidades microbianas [30]. Mientras que la metagenómica refleja el potencial genético de una comunidad, la metatranscriptómica identifica los genes que se expresan activamente bajo condiciones fisiológicas o patológicas específicas [22], [24]. Este enfoque, basado en la secuenciación masiva de ARN (*RNA-seq*, por sus siglas en inglés), proporciona una visión dinámica del funcionamiento microbiano, al mostrar la respuesta transcripcional frente a estímulos del hospedero, la dieta o el entorno [27].

A diferencia de los estudios metagenómicos, que solo infieren la capacidad funcional, la metatranscriptómica ofrece evidencia directa de las rutas metabólicas activas y de los mecanismos moleculares implicados en la interacción microbio–huésped. Técnicamente, este proceso requiere la estabilización inmediata del ARN tras la toma de muestra para evitar la degradación de los transcritos [29]. Posteriormente, se realiza la extracción del ARN total, seguida de etapas críticas de depleción del ARN ribosomal y eliminación de los transcritos del hospedero mediante kits de afinidad [31]. Este enriquecimiento del ARN mensajero bacteriano es esencial para obtener una resolución adecuada de los genes expresados. Finalmente, el ARNm se convierte en ADN complementario para su secuenciación masiva [31]. Esto la convierte en una herramienta esencial para comprender cómo las bacterias modulan procesos como la digestión de nutrientes, la producción de metabolitos, la regulación inmunitaria y la tolerancia a patógenos [32]. En el contexto de las enfermedades metabólicas, la metatranscriptómica ha revelado alteraciones en la expresión de genes asociados con el metabolismo energético, la síntesis lipídica y la inflamación, los cuales pueden afectar la homeostasis intestinal y favorecer el desarrollo de disbiosis [17], [28].

La integración de la metatranscriptómica con otras aproximaciones, como la metagenómica, la proteómica y la virómica, permite una visión más completa de la ecología funcional del intestino humano, al vincular la presencia de genes con su expresión y con la posible traducción de estos en productos bioactivos [24], [30], [33]. Específicamente, la virómica se encarga del estudio del conjunto de genomas virales (viroma) en un ecosistema [30]; en el contexto intestinal, esta disciplina es fundamental para identificar genes de bacteriófagos, los cuales actúan como reservorios de diversidad genética que permiten regular la estructura y función de las comunidades bacterianas [34]. Esto incluye no solo mecanismos de lisis, sino también la transferencia de genes

que pueden modificar el metabolismo bacteriano o la producción de péptidos con actividad biológica [35], [36]. Este tipo de análisis integrativo resulta especialmente útil para explorar la relación entre la microbiota y condiciones metabólicas complejas, como la obesidad y el SM, en las que participan tanto factores del hospedero como mecanismos microbianos aún poco caracterizados [30].

En síntesis, la metatranscriptómica constituye una herramienta clave para identificar los genes expresados por la microbiota y evaluar sus posibles efectos sobre la fisiología del hospedero, aportando información esencial para comprender los procesos funcionales que sustentan la disbiosis metabólica [35]. Dentro de estos procesos, resulta fundamental caracterizar el Secrebioma, definido como el conjunto de proteínas y péptidos con señales de secreción producidos por la comunidad microbiana que median las interacciones tanto entre los miembros de la microbiota como con el hospedero [37]. Entre los transcritos que tienen relevancia funcional en este contexto, destacan los péptidos antimicrobianos (AMPs), al igual que su posible relación con las alteraciones metabólicas asociadas al SM.

AMPs y su papel en la competencia microbiana intestinal

Los AMPs son péptidos de bajo peso molecular, por lo general catiónicos y anfifílicos [38]. Bajo un criterio funcional, se definen como péptidos de 5 a 50 aminoácidos que no poseen actividad enzimática [38], [39], [40]. Son producidos tanto por macroorganismos, incluidos animales y plantas, como por microorganismos, y participan en la defensa frente a patógenos y en la competencia microbiana [41]. En el contexto intestinal, los AMPs desempeñan un papel esencial en el mantenimiento del equilibrio microbiano, actuando como moléculas efectoras del sistema inmunitario innato y modulando la composición microbiana al limitar el crecimiento de patógenos y favorecer la colonización de microorganismos comensales [42]. Estos péptidos son sintetizados tanto por las células epiteliales del hospedero, que evitan la colonización de la mucosa interna, como por la microbiota, en la mucosa externa, donde la abundancia y diversidad microbiana son mayores [42], [43], [44] (Figura 3).

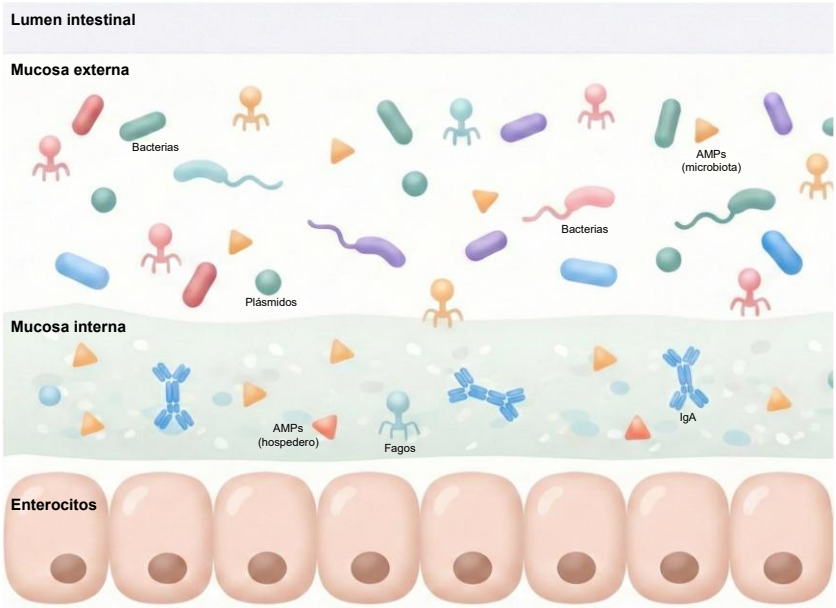


Figura 3: Papel de los AMPs en la dinámica del microbioma intestinal. Los AMPs, producidos tanto por el hospedero como por la microbiota, mantienen el equilibrio microbiano intestinal. Los péptidos del hospedero protegen la mucosa interna al evitar la colonización por bacterias potencialmente dañinas, mientras que los producidos por la microbiota modulan la competencia entre especies y la respuesta del hospedero. Algunos se encuentran codificados en elementos genéticos móviles, como fagos y plásmidos, lo que les confiere ventajas adaptativas dentro del ecosistema intestinal [45].

Recientes estudios metagenómicos y metatranscriptómicos han revelado una notable diversidad de AMPs bacterianos expresados en el intestino humano, muchos de ellos con propiedades estructurales atípicas derivadas de pequeños marcos de lectura abiertos (smORFs) [41], [46]. Estos péptidos no solo cumplen funciones antimicrobianas, sino también inmunomoduladoras y de señalización local, reforzando su papel en la regulación dinámica de la comunidad intestinal [39].

Mecanismos de acción de los AMPs

En cuanto a su mecanismo de acción, los AMPs pueden: a) modular la respuesta inmune del hospedero, ya sea activando y reclutando células del sistema inmune o ajustando el reconocimiento mediado por receptores tipo Toll; b) desestabilizar o romper la membrana o la pared celular del microorganismo blanco; c) inhibir funciones intracelulares esenciales lo cual desemboca en la muerte del organismo blanco [39], [47], [48] (Figura 4). Además, algunos AMPs participan en la comunicación bacteriana (*quorum sensing*) y en la regulación de factores de virulencia, ampliando su papel en las interacciones microbianas [41], [49].

El microbioma constituye una reserva natural de AMPs aún por descubrir, con gran potencial terapéutico contra infecciones bacterianas, incluidas las causadas por patógenos multirresistentes [50], [51]. Sin embargo, la mayoría de estos péptidos permanecen ocultos debido a su pequeño tamaño, baja abundancia y a la falta de anotaciones precisas en los genomas microbianos, lo que dificulta su detección mediante enfoques convencionales [52].

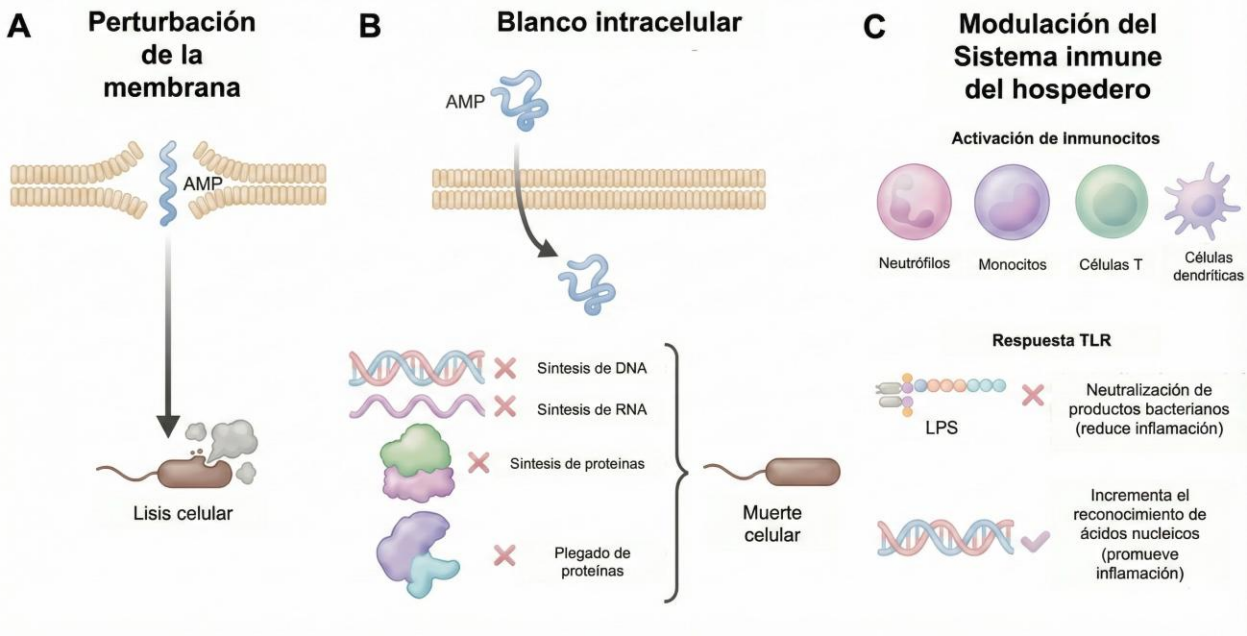


Figura 4: Mecanismos de acción de los AMPs del microbioma. A) Inducen la muerte celular al dañar la membrana del microorganismo. B) Inhiben funciones celulares esenciales, afectando procesos vitales. C) Modulan la respuesta inmunitaria del hospedero al activar células inmunes o receptores tipo Toll (TLR), neutralizar productos bacterianos como LPS o reconocer ácidos nucleicos microbianos [45].

Identificación de AMPs mediante enfoques multi-ómicos

El estudio funcional de los AMPs requiere aproximaciones integrativas que combinen distintas capas de información ómica. La metagenómica identifica los genes potenciales, la metatranscriptómica determina cuáles están activos, y la proteómica y metabolómica validan su traducción y efecto funcional [24], [29] (Figura 5). La integración de estos datos multi-ómicos, junto con modelos computacionales basados en aprendizaje automático, ha mejorado la capacidad para vincular la expresión génica con su función ecológica, en particular para los AMPs expresados en el intestino [50]. Asimismo, la incorporación de la virómica como componente de los estudios multiómicos ha permitido identificar AMPs codificados por bacteriófagos, revelando la participación de estos virus en la competencia microbiana y en la evolución del microbioma intestinal [53]. En conjunto, los enfoques multi-ómicos ofrecen una visión global de la actividad antimicrobiana del intestino humano, integrando información genómica, funcional y ecológica [30].

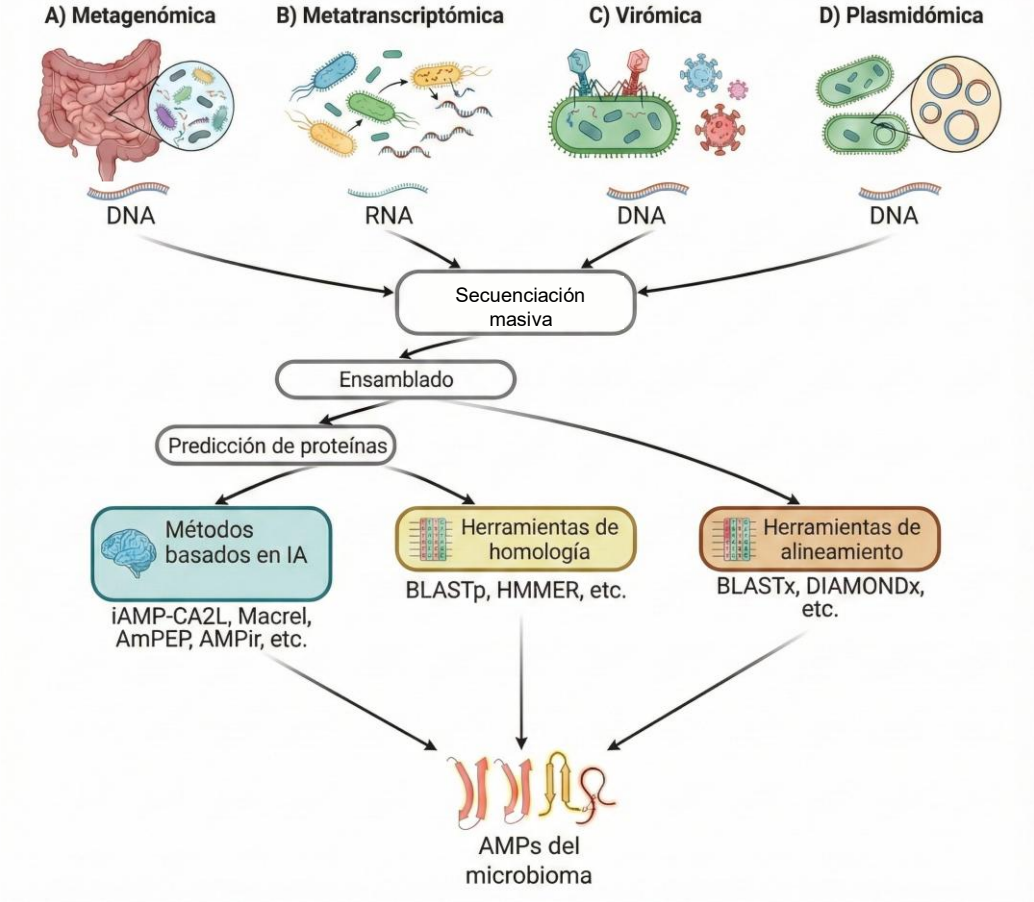


Figura 5: Enfoque multi-ómico y bioinformático para la identificación de AMPs del microbioma. Las aproximaciones de A) metagenómica, B) metatranscriptómica, C) virómica y D) plasmidómica permiten obtener secuencias de ADN o ARN provenientes de distintos componentes del microbioma intestinal. Tras el ensamblaje y la predicción de proteínas, las secuencias son analizadas mediante herramientas basadas en inteligencia artificial o métodos de homología y alineamiento. La integración de estos enfoques permite detectar y caracterizar los AMPs expresados o codificados por bacterias, fagos y plásmidos, proporcionando una visión funcional del potencial antimicrobiano del microbioma [45].

El acceso cada vez mayor a datos metagenómicos y metatranscriptómicos ha permitido identificar nuevos AMPs del microbioma de manera *in silico*, a partir de smORFs o secuencias no anotadas previamente, ampliando así el conocimiento del repertorio funcional del microbioma [50], [51], [54]. Los modelos de aprendizaje automático y los predictores de consenso, que combinan múltiples algoritmos de predicción de actividad antimicrobiana, se han convertido en herramientas esenciales para filtrar secuencias candidatas y priorizar aquellas con mayor probabilidad de ser AMPs [52], [55]. En este contexto, la integración de datos multi-ómicos, como la metatranscriptómica y la virómica, resulta clave para detectar los AMPs más relevantes y activos en el intestino [45].

AMPs codificados en elementos genéticos móviles

Una fracción considerable de los AMPs del microbioma se encuentra codificada en elementos genéticos móviles (EGMs), como plásmidos, fagos, profagos o islas

genómicas, lo que facilita su transferencia horizontal y su diseminación dentro del ecosistema intestinal [44], [56], [57], [58], [59]. Estos elementos no solo contribuyen a la propagación de genes de resistencia o virulencia, sino que también albergan funciones antimicrobianas que modulan la competencia bacteriana y favorecen la estabilidad de las comunidades microbianas [60], [61].

Además, muchos AMPs provienen de smORFs de origen móvil, que históricamente pasaban desapercibidos en las anotaciones genómicas y hoy pueden detectarse mediante algoritmos especializados basados en aprendizaje automático [54]. Aunque existen diversas bases de datos y servidores para predecir AMPs, aún son limitados los protocolos integrales que permiten identificarlos directamente en datos multi-ómicos y validar experimentalmente su función. El descubrimiento de AMPs móviles ha cobrado relevancia por su posible impacto en la estructura y estabilidad del microbioma intestinal, ya que su movilidad genética permite a las bacterias adquirir nuevas estrategias de defensa o ataque frente a presiones ambientales y cambios en la comunidad [36], [59].

Brecha de conocimiento y justificación del presente estudio

A pesar de los avances en la caracterización funcional del microbioma, persisten vacíos importantes en la identificación y validación de AMPs expresados en el intestino humano, particularmente aquellos codificados en elementos genéticos móviles y en smORFs. La mayoría de los estudios se ha centrado en describir la composición microbiana, sin profundizar en los mecanismos funcionales que regulan la competencia bacteriana y la homeostasis intestinal. En consecuencia, aún se desconoce la amplitud del repertorio antimicrobiano expresado por la microbiota intestinal y su asociación con enfermedades como la obesidad y el SM. En este contexto, el presente trabajo de tesis doctoral integra enfoques multi-ómicos para estudiar la expresión y el origen de los AMPs en la microbiota intestinal infantil, con énfasis en su asociación con obesidad y SM.

Hipótesis

La expresión diferencial de péptidos antimicrobianos (AMPs) de la microbiota se asocia con la disbiosis en la Obesidad y el Síndrome Metabólico principalmente mediada por elementos génicos móviles.

Objetivos

Objetivo principal:

- Determinar la expresión diferencial de AMPs en la microbiota de población infantil con Obesidad y SM.

Objetivos específicos:

- Caracterizar la estructura taxonómica de la microbiota intestinal mediante el perfilamiento del gen ARNr 16S.
- Analizar el metatranscriptoma intestinal para identificar genes expresados diferencialmente y determinar el Secrebioma asociado a cada condición metabólica.
- Identificar y cuantificar la expresión diferencial de AMPs codificados por la microbiota intestinal en los grupos con obesidad y SM.
- Determinar el origen genético de los AMPs, distinguiendo aquellos codificados en elementos génicos móviles (fagos o plásmidos), mediante la integración de datos metatranscriptómicos y virómicos.
- Evaluar experimentalmente la actividad antibacteriana *in vitro* de los AMPs seleccionados frente a bacterias representativas de distintos filos.
- Comprobar la ausencia de citotoxicidad o reactividad inmunológica de los AMPs activos frente a células T humanas.

Materiales y métodos

Población de estudio y colecta de las muestras

El estudio se realizó a partir de una submuestra de 27 niños seleccionados de una cohorte infantil mexicana de 750 participantes, con edades entre 5 y 11 años. Las muestras fecales fueron recolectadas en 2014 por el grupo del Dr. Samuel Canizales Quinteros (Facultad de Química/INMEGEN), siguiendo protocolos estandarizados y aprobados por el comité de ética correspondiente. Las muestras se preservaron en *RNAlater* y se almacenaron a -70 °C hasta su procesamiento, garantizando la integridad del ARN para el análisis metatranscriptómico. Además, se registraron variables antropométricas y bioquímicas con el fin de comparar las características metabólicas entre los grupos.

El objetivo fue comparar las diferencias microbianas y funcionales entre tres grupos definidos según su estado metabólico: normopeso (NW, n=7), obesidad (O, n=10) y obesidad con síndrome metabólico (OMS, n=10) (Figura 6). El grupo NW incluyó infantes metabólicamente sanos, sin alteraciones bioquímicas ni antropométricas, cuyos parámetros clínicos se encontraban dentro de los rangos normales para su edad y sexo. Este grupo sirvió como referencia para identificar las variaciones microbianas y funcionales asociadas con alteraciones metabólicas.

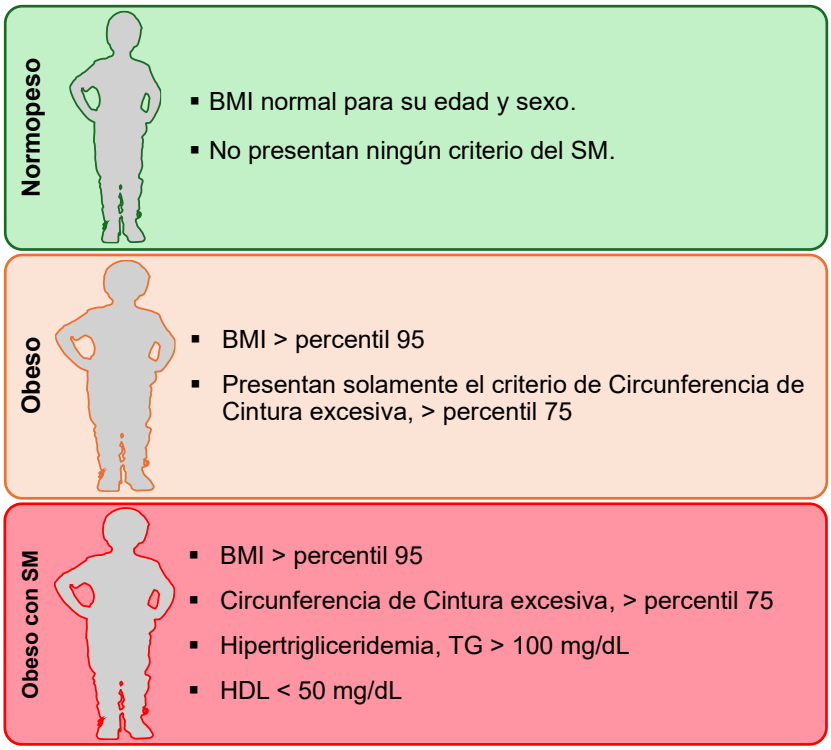


Figura 6: Criterios de selección y clasificación de los grupos de estudio. El grupo normopeso (NW) incluyó infantes con IMC normal y sin criterios del síndrome metabólico (SM). El grupo obesidad (O) presentó IMC>percentil 95 y cintura>percentil 75. El grupo obesidad con síndrome metabólico (OMS) cumplió con obesidad central más al menos tres de los siguientes criterios: hipertrigliceridemia (TG>100 mg/dL), colesterol HDL<50 mg/dL y presión arterial elevada.

El grupo O estuvo conformado por infantes con obesidad central, definida como una circunferencia de cintura superior al percentil 75 (P75) para edad y sexo, pero sin otras alteraciones metabólicas. Estos participantes presentaban un exceso de adiposidad sin manifestar aún las alteraciones bioquímicas o cardiovasculares características del SM. Por su parte, el grupo OMS se definió por la presencia de obesidad central acompañada de tres o más de los siguientes criterios: triglicéridos en ayuno ≥1.1 mmol/L, colesterol HDL<1.3 mmol/L y presión arterial sistólica >P90 por edad, sexo y talla. Para el análisis metatranscriptómico (*RNA-seq*), se consideró una submuestra de ocho participantes distribuidos de la siguiente manera: NW=2, O=3 y OMS=3, representando proporcionalmente a cada grupo metabólico para la evaluación funcional del microbioma.

Extracción de ADN y ARN para secuenciación masiva

La preparación del material genético se dividió en dos flujos paralelos: ADN, destinado al análisis taxonómico mediante el gen ARNr 16S, y ARN total, empleado para el análisis metatranscriptómico (Figura 7).

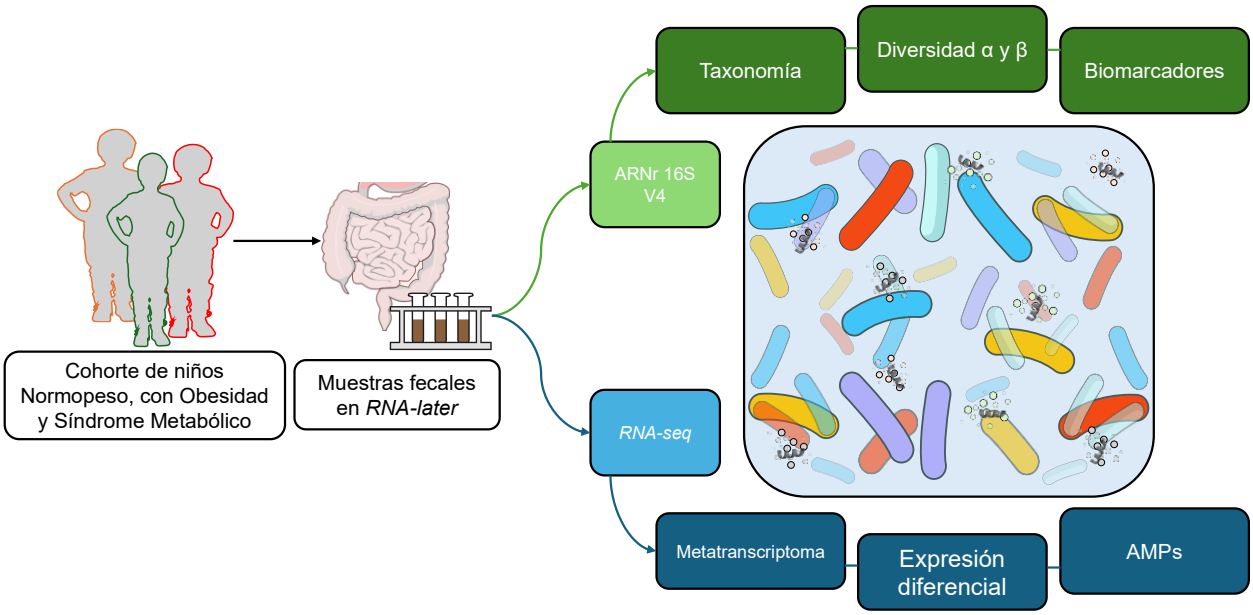


Figura 7: Esquema general del diseño experimental y análisis realizados. A partir de muestras fecales preservadas en RNAlater, se realizaron análisis 16S (región V4) para determinar la composición y diversidad microbiana, y RNA-seq para obtener perfiles metatranscriptómicos. Estos datos permitieron evaluar la diversidad alfa y beta, identificar genes diferencialmente expresados y detectar biomarcadores, incluidos los AMPs.

Para el análisis 16S, se extrajo ADN a partir de 200 mg de heces por participante utilizando el *QIAamp DNA Stool Mini Kit* (Qiagen, Alemania). Se amplificó la región hipervariable V4 con los cebadores 515F/806R, y los amplicones se secuenciaron en plataforma Illumina MiSeq con lecturas pareadas de 2×250 pb.

El ARN total se extrajo a partir de 200 mg de muestra fecal empleando una combinación de los kits *ZR Soil/Fecal RNA MicroPrep* (Zymo Research, USA) y *RNeasy Mini Kit* (Qiagen, Alemania), siguiendo las instrucciones de los fabricantes. La calidad y

cantidad del ARN se evaluaron mediante *Bioanalyzer 2100* (Agilent) y *Qubit 2.0 Fluorometer* (Invitrogen), confirmando perfiles ribosomales bien definidos. Posteriormente, se realizó la depleción de ARN ribosomal humano y bacteriano utilizando el *Ribo-Zero Gold rRNA Removal Kit* (Illumina, USA). Las librerías se prepararon siguiendo el protocolo de Shishkin et al., 2015 [31], adaptado para librerías simultáneas, con los siguientes pasos: depleción de ADN (*TURBO DNase*, 30 min a temperatura ambiente), desfosforilación de extremos 5′/3′ con FastAP, ligación inicial de adaptadores con códigos de barras, fragmentación controlada por hidrólisis alcalina (NaOH 1 N), neutralización, segunda ligación de adaptadores 3′ y amplificación por PCR (15 ciclos). La secuenciación se realizó en la plataforma Illumina HiSeq, con lecturas paired-end (2×150 pb), alcanzando una profundidad de 20–30 millones de lecturas por muestra.

Análisis bioinformático para el perfilamiento de 16S

El análisis taxonómico del gen ARNr 16S (región V4) se realizó empleando dos enfoques complementarios: el primero, utilizado en el primer artículo, con QIIME v1.9.1, y el segundo, aplicado como actualización metodológica en el segundo artículo, con QIIME2 v2024.5 (Figura 8). Estas aproximaciones se aplicaron de forma secuencial con el objetivo de mantener la reproducibilidad del análisis original e incorporar las mejoras en la inferencia y clasificación taxonómica disponibles en versiones recientes de QIIME.

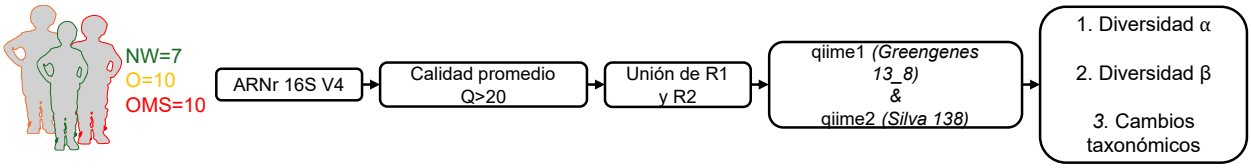


Figura 8: Esquema general del análisis taxonómico basado en el gen ARNr 16S (región hipervariable V4). Las muestras fecales correspondientes a los grupos NW (n=7), O (n=10) y OMS (n=10) fueron procesadas para la amplificación de la región V4 del gen ARNr 16S. Las lecturas obtenidas se evaluaron en su calidad promedio (Q> 20), se unieron los pares R1 y R2 y posteriormente se analizaron con QIIME v1.9.1 (base Greengenes 13_8) y QIIME2 v2024.5 (base SILVA v138). Con ambos enfoques se calcularon los índices de diversidad alfa y beta, así como las diferencias taxonómicas entre los grupos.

En el análisis inicial, las lecturas crudas fueron evaluadas con FastQC y sometidas a filtrado de calidad (Phred>Q20) y eliminación de bases ambiguas utilizando Trimmomatic. Las lecturas pareadas se unieron y se importaron a QIIME v1.9.1. Las secuencias se agruparon en unidades taxonómicas operativas (OTUs) al 97% de identidad mediante UCLUST, empleando un enfoque de agrupamiento por referencia cerrada (*closed-reference*) con la base de datos Greengenes 13_8 para la asignación taxonómica. Para conservar únicamente los OTUs comunes entre grupos, se excluyeron aquellas secuencias con abundancia relativa≤0.005% del total de lecturas. La diversidad alfa se estimó mediante 10 000 iteraciones, utilizando los índices Chao1 (riqueza), Shannon (diversidad) y Simpson (dominancia), mientras que la diversidad beta se evaluó con distancias UniFrac ponderadas y no ponderadas, visualizadas mediante análisis de coordenadas principales (PCoA). Las diferencias en la estructura microbiana entre grupos se determinaron mediante PERMANOVA, y la identificación de taxones diferencialmente abundantes se realizó con LEfSe, considerando un valor de LDA≥1.0 como para la detección de biomarcadores taxonómicos.

Posteriormente, se efectuó una actualización del análisis taxonómico utilizando QIIME2 v2024.5, con el fin de incorporar mejoras en la inferencia de secuencias y la clasificación filogenética. En esta versión, las lecturas fueron importadas, desreplicadas y agrupadas de novo al 97% de identidad, generando un conjunto refinado de secuencias representativas. La asignación taxonómica se llevó a cabo mediante un clasificador Naive Bayes entrenado con la base de datos SILVA v138. Con estos resultados se calcularon los mismos índices y métricas de diversidad empleados en el análisis previo, manteniendo la coherencia metodológica entre ambas aproximaciones.

Ensamblado *de novo* de *RNA-Seq*

Las lecturas crudas obtenidas por *RNA-seq* fueron sometidas a control de calidad con FastQC y a filtrado por ventanas de seis nucleótidos (Q>20) utilizando Trimmomatic v0.36. Las secuencias ribosomales de procariotas y eucariotas se eliminaron con Ribopicker v0.4.3, empleando la base de datos SILVA v138. Posteriormente, las secuencias de origen humano se descartaron mediante Kneaddata, reteniendo únicamente las lecturas de origen microbiano. El ensamblado de novo se realizó con Trinity, ajustando los parámetros para conservar transcritos cortos potencialmente codificantes de marcos de lectura pequeños (smORFs). Las lecturas originales se alinearon nuevamente al ensamblado con Bowtie2 para evaluar la cobertura, y la predicción de marcos de lectura abiertos (ORFs) se efectuó con TransDecoder, aplicando el parámetro -m 5 para detectar smORFs menores de 100 aminoácidos (Figura 9).

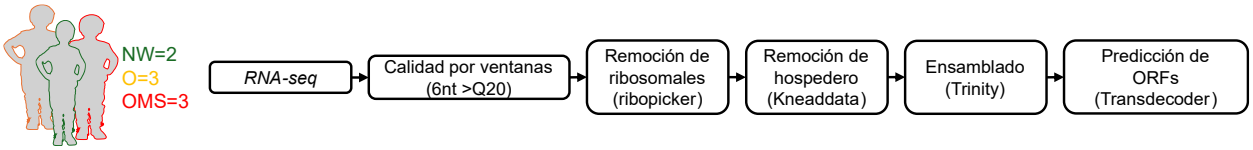


Figura 9: Esquema general del ensamblado de novo del metatranscriptoma. Las lecturas crudas se filtraron para eliminar secuencias ribosomales con Ribopicker y de origen humano con Kneaddata, reteniendo únicamente lecturas microbianas. El ensamblado de novo se realizó con Trinity, y la predicción de smORFs con TransDecoder.

Predicción y análisis funcional del Secrebioma intestinal

El Secrebioma microbiano se definió como el conjunto de proteínas potencialmente secretadas por la microbiota intestinal. Para su predicción, se utilizaron las proteínas predichas de los transcritos obtenidos a partir del ensamblado *de novo*. Las secuencias de proteínas codificadas fueron analizadas mediante una combinación de herramientas bioinformáticas enfocadas en la detección de péptidos señal y la predicción de localización subcelular.

En primer lugar, se emplearon las herramientas SignalP, TatP, LipoP y SecretomeP para predecir las rutas clásicas y no clásicas de secreción, identificando proteínas con péptidos señal asociados a los sistemas Sec, Tat, lipoproteínas y secreción no clásica, respectivamente. Las secuencias candidatas a proteínas secretadas fueron posteriormente filtradas mediante TMHMM v2.0, con el fin de eliminar aquellas que presentaran dominios transmembrana. Finalmente, se utilizó Phobius para confirmar la

señal N-terminal y descartar falsos positivos. Las proteínas que presentaron péptido señal y ausencia de múltiples dominios transmembrana fueron consideradas como parte del Secrebioma intestinal (Figura 10).

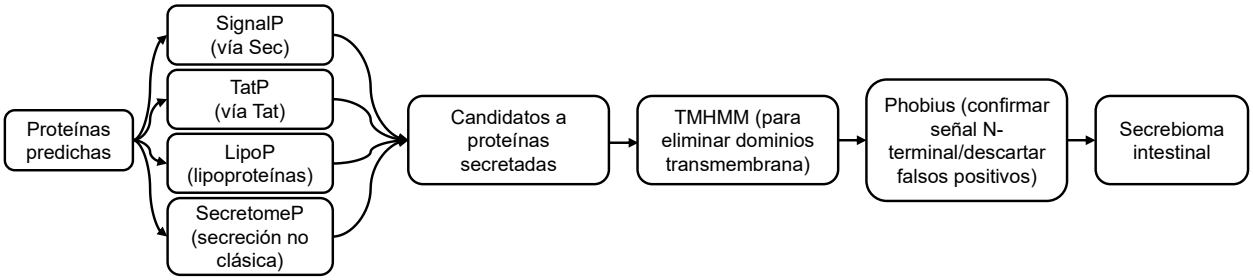


Figura 10: Esquema para la predicción del Secrebioma intestinal. Primero se evaluaron las rutas clásicas y no clásicas mediante SignalP (vía Sec), TatP (vía Tat), LipoP (lipoproteínas) y SecretomeP (secreción no clásica). Las proteínas candidatas fueron posteriormente filtradas con TMHMM para eliminar aquellas que presentaran dominios transmembrana, y analizadas con Phobius para confirmar la presencia de señal N-terminal y descartar posibles falsos positivos. Las secuencias resultantes conformaron el conjunto de proteínas potencialmente secretadas definido como el Secrebioma intestinal.

Las anotaciones funcionales se realizaron mediante la asignación de categorías KEGG Orthology (KO) y Gene Ontology (GO), permitiendo identificar rutas metabólicas, funciones enzimáticas y procesos biológicos enriquecidos en cada grupo. Adicionalmente, se empleó la base de datos CAZy (Carbohydrate-Active enZYmes) para identificar familias de enzimas relacionadas con la degradación, modificación o síntesis de polisacáridos, así como proteínas implicadas en la interacción hospedero–microbiota.

Identificación, filtrado y validación de AMPs

La búsqueda de AMPs expresados por la microbiota intestinal se realizó a partir de los transcritos ensamblados derivados del análisis metatranscriptómico (Figura 11). La detección de smORFs se efectuó con TransDecoder, ajustando el parámetro -m 5 para permitir la identificación de marcos de lectura cortos potencialmente codificantes, como los AMPs. Posteriormente, se aplicó un filtrado de calidad para eliminar transcritos con baja o inconsistente expresión, conservando únicamente aquellos con FPKM≥1 en al menos tres muestras.

La predicción de AMPs se realizó mediante una estrategia que integró tres herramientas complementarias: Macrel [62], AxPEP [63] y AMP Scanner V2 [52]. Macrel utiliza modelos de aprendizaje automático basados en características composicionales y fisicoquímicas de las secuencias; AxPEP evalúa propiedades asociadas a la actividad antimicrobiana, como la carga neta, hidrofobicidad y momento anfipático; mientras que AMP Scanner V2 emplea redes neuronales convolucionales (CNN) entrenadas con bases de datos de AMPs conocidos y péptidos inactivos. Los péptidos identificados por al menos una de las tres herramientas se consideraron candidatos potenciales. Posteriormente, estos candidatos fueron sometidos a un proceso de curación y filtrado secuencial. Se conservaron únicamente los péptidos con una longitud ≥10 aminoácidos, codificados por ORFs completos y no sobrelapados, con homólogos detectables en la base NCBI y secuencia en aminoácidos del AMP confirmada en genomas homólogos.

Adicionalmente, se descartaron péptidos con longitudes >100 aminoácidos o reclasificados como smORFs falsos positivos durante la validación. El conjunto final se consideró como AMPs de alta confiabilidad, garantizando la plausibilidad funcional de dichos péptidos.

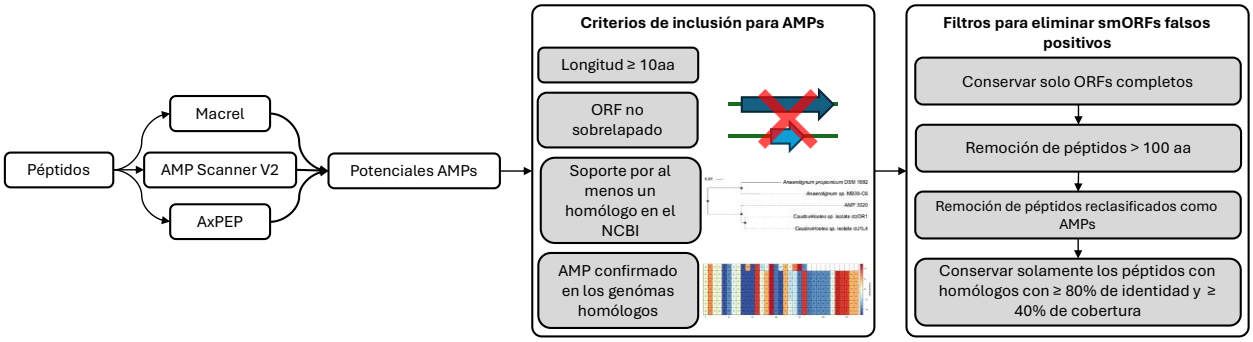


Figura 11: Esquema de predicción, filtrado y validación de AMPs. Los péptidos derivados de los transcritos ensamblados se evaluaron mediante tres predictores (Macrel, AxPEP y AMP Scanner V2) integrados en un análisis de consenso. Los AMPs potenciales se depuraron según criterios de inclusión (longitud ≥ 10 aa, ORFs no sobrelapados y presencia de homólogos en NCBI) y se filtraron para eliminar smORFs falsos positivos, conservando únicamente aquellos con homología ≥ 80 % de identidad y ≥ 40 % de cobertura.

El origen taxonómico de los AMPs se determinó mediante BLASTN contra la base NCBI NT (E-value≤1e−5; identidad≥80%; cobertura≥80%), y la asignación taxonómica se realizó con el algoritmo del último ancestro común (LCA) en MEGAN6 v6.25.10. Los contextos genómicos se representaron con AnnotationSketch, mientras que las comparaciones de sintenia se efectuaron con Easyfig, ambos basados en alineamientos BLASTN. Los árboles filogenéticos se generaron con iTOL, incorporando AMPs de referencia previamente descritos, y se analizaron las regiones genómicas flanqueantes con Easyfig para identificar posibles clústeres génicos o elementos regulatorios asociados.

Análisis de expresión diferencial

Primero se calcularon los valores de abundancia transcripcional normalizados con RSEM, utilizando un enfoque de cuantificación libre de referencia implementado dentro del flujo de trabajo de Trinity. Los transcritos con valores FPKM<1 fueron descartados, y solo se consideraron aquellos presentes en al menos tres muestras. El análisis diferencial de los AMPs se realizó con DESeq2 v1.26.0 en R, empleando un modelo de distribución binomial negativa para estimar la variabilidad entre muestras. Los valores de significancia se ajustaron mediante el método de Benjamini–Hochberg, considerando diferencialmente expresados los AMPs con padj≤0.05 y $\log_2$FC≥0.5.

Asociaciones entre taxones y expresión de AMPs

Para integrar los resultados del perfil taxonómico (ARNr 16S) con la expresión funcional del metatranscriptoma, se realizaron correlaciones de Spearman entre la abundancia relativa de los principales taxones microbianos (género y familia) obtenidos mediante

QIIME2 y los valores FPKM normalizados de los AMPs identificados, considerando únicamente asociaciones con $p \leq 0.05$ y $\rho \geq 0.6$. Las correlaciones significativas se representaron en mapas de calor generados en R (paquete ggplot2), donde se visualizaron las relaciones positivas y negativas entre taxones y AMPs. Este enfoque permitió identificar microorganismos potencialmente implicados en la producción o regulación de AMPs dentro de la microbiota intestinal, proporcionando una conexión funcional entre la composición microbiana y la expresión de dichos AMPs.

Validación experimental del AMP 3020

La validación experimental del péptido AMP 3020, identificado en el metatranscriptoma intestinal, se realizó para confirmar su actividad antimicrobiana y biocompatibilidad celular. El péptido, de origen viral (100% de identidad con una proteína del fago Caudoviricetes ctJ1L4), fue sintetizado en dos variantes: ADR1, con metionina N-terminal, y ADR2, sin metionina inicial. Ambas se purificaron por HPLC en columna C18 y se verificaron por espectrometría de masas.

La actividad antimicrobiana se evaluó mediante microdilución en caldo frente a *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Staphylococcus aureus* y *Streptococcus pneumoniae*. Las bacterias se ajustaron a ~10^5–10^6 UFC/mL y se incubaron con diluciones seriadas del péptido durante 18–24 h a 37 °C. La concentración inhibitoria mínima (CIM) se definió como la menor concentración sin crecimiento visible (OD600). Se incluyeron controles de crecimiento, BSA como control de especificidad y antibióticos de referencia (Kanamicina y Ampicilina), todos a una concentración final de 100 µg/mL.s

Adicionalmente la citotoxicidad se evaluó por citometría de flujo en linfocitos T ($CD3^+$, $CD4^+$ y $CD8^+$) aislados de PBMCs de donadores sanos, incubados 24 h con 20 µg/mL de cada variante. La viabilidad celular se determinó mediante tinción de exclusión, permitiendo identificar posibles efectos citotóxicos.

Resultados

Los resultados de esta tesis se componen de tres estudios complementarios que abordan la estructura, funcionalidad y mecanismos de regulación de la microbiota intestinal en el contexto de obesidad y SM. Mediante un enfoque multi-ómico se analizó primero la composición bacteriana mediante la secuenciación del gen ARNr 16S, posteriormente la expresión funcional del Secrebioma a partir de datos de *RNA-seq*, y finalmente la identificación y caracterización de AMPs expresados por la microbiota intestinal, integrando información metatranscriptómica y virómica.

El Capítulo 1 reúne los resultados del primer artículo, publicado en *Microbial Cell Factories* bajo el título "*Metatranscriptomic analysis to define the Secrebiome, and 16S rRNA profiling of the gut microbiome in obesity and metabolic syndrome of Mexican children*". Este capítulo integra el análisis de la composición, estructura y función de la microbiota intestinal, basado en la secuenciación del gen ARNr 16S y en el estudio de las proteínas secretadas por el microbioma a partir del análisis del metatranscriptoma.

El Capítulo 2 corresponde al segundo estudio, publicado en *Microbial Ecology* bajo el título "*Bioactive Plasmid- and Phage-Encoded Antimicrobial Peptides (AMPs) in the Human Gut: A Metatranscriptome–Virome Profiling Reveals Exploratory Links to Metabolic Human Diseases*". Este capítulo describe la identificación y análisis de AMPs expresados por la microbiota intestinal, incluyendo aquellos codificados en elementos genéticos móviles como fagos y plásmidos. Además, integra datos de metatranscriptómica y viromas obtenidos de las mismas muestras, junto con correlaciones con los perfiles 16S.

Como anexo se incluye un artículo de revisión, publicado también en *Microbial Ecology* bajo el título "*Perspectives in Searching Antimicrobial Peptides (AMPs) Produced by the Microbiota*", que presenta una revisión de las metodologías empleadas para la búsqueda e identificación de AMPs en datos de secuenciación masiva obtenidos de la microbiota. Este trabajo fue fundamental y sirvió como marco conceptual y metodológico para orientar las estrategias bioinformáticas aplicadas en la búsqueda de AMPs en nuestros datos.

En conjunto, estos tres estudios ofrecen una visión integral, multi-ómica y ecológica de la microbiota intestinal en la obesidad y el SM, al caracterizar su composición, funcionalidad y mecanismos de regulación mediados por AMPs. Los resultados contribuyen a comprender los procesos microbianos asociados con la disbiosis metabólica y destacan el potencial biotecnológico de los AMPs como base para el desarrollo de nuevas estrategias antimicrobianas y terapéuticas derivadas del microbioma.

Capítulo 1. Disbiosis asociada a la obesidad y SM en población infantil

1.1 Composición taxonómica de la microbiota intestinal en los grupos de estudio

La composición taxonómica del gen ARNr 16S mostró que la microbiota intestinal de los grupos NW, O y OMS estuvo dominada por tres filos principales: Firmicutes, Bacteroidetes y Proteobacteria, que en conjunto representaron más del 95% de las lecturas totales. Se observaron variaciones significativas en la abundancia relativa de los dos primeros filos. Firmicutes presentó un incremento progresivo de 26.7% en NW a 40.3% en O y 46.4% en OMS (p=0.0431), mientras que Bacteroidetes disminuyó de 66.9% a 55.3% y 45.5%, respectivamente (p=0.0249). En consecuencia, la relación Firmicutes/Bacteroidetes (F/B) aumentó de 0.55 en NW a 1.20 en OMS, reflejando un cambio estructural asociado con el estado metabólico (Figura 12A).

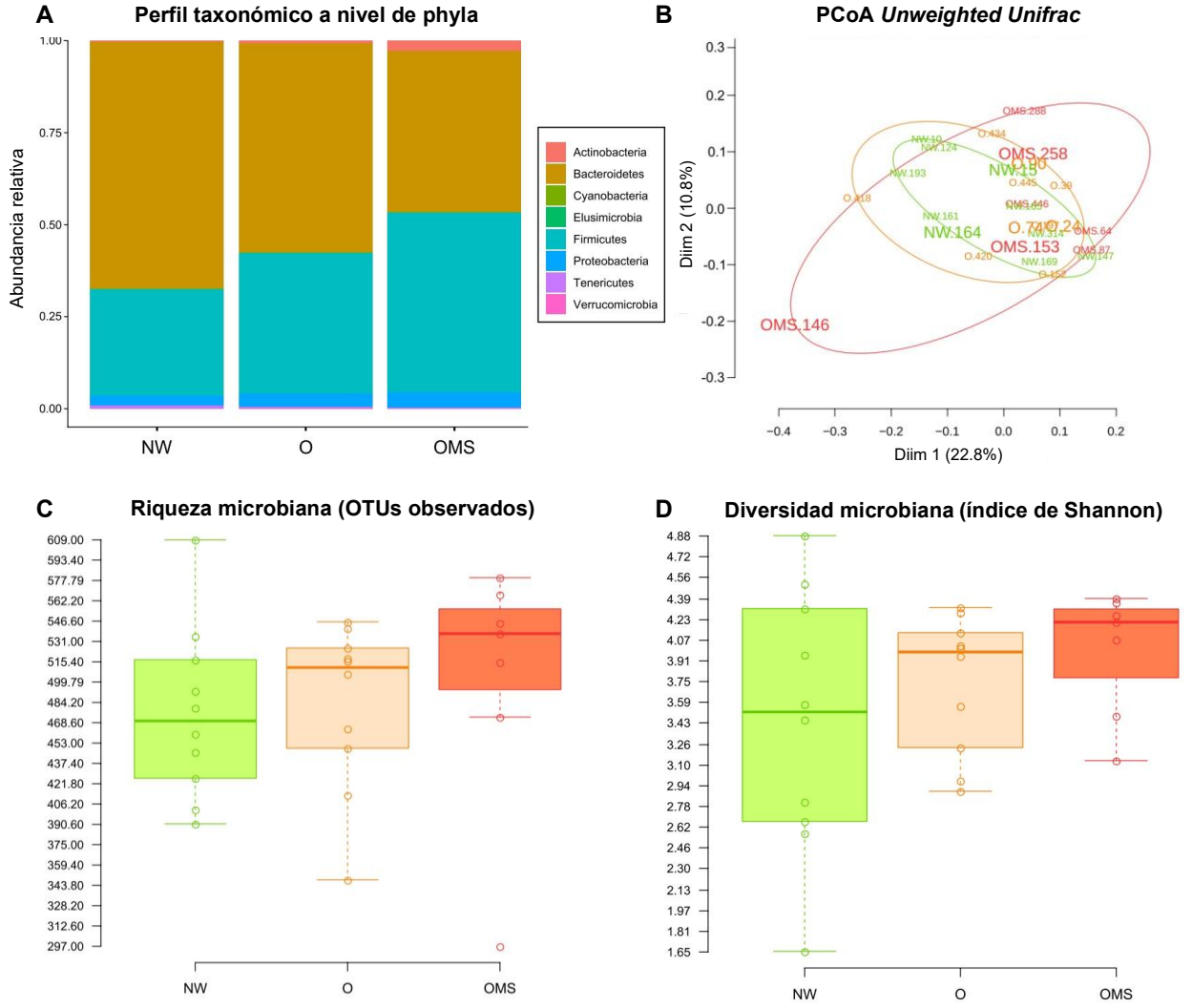


Figura 12: Composición taxonómica y diversidad microbiana asociada a obesidad y SM en población infantil. A) Abundancia relativa promedio de los principales filos bacterianos en los grupos NW, O y OMS. B) Análisis de coordenadas principales (PCoA) basado en distancias UniFrac no ponderadas. C) Número OTUs observados por grupo. D) Índice de diversidad de Shannon.

1.2 Estructura beta y alfa de la comunidad microbiana intestinal

El análisis de la diversidad beta, la cual mide la diferencia en la composición de especies entre las comunidades, se basó en distancias *UniFrac* no ponderadas (*Unweighted*) y se visualizó mediante PCoA (Figura 12B). Los resultados mostraron centroides parcialmente sobrelapados entre los grupos, lo que sugiere que no existe una separación clara en la estructura global de las comunidades. No obstante, se observaron distintos grados de heterogeneidad en la variación de los datos: las muestras NW formaron un clúster más compacto, mientras que las muestras O y OMS presentaron una dispersión notablemente mayor. Si bien esta heterogeneidad podría reflejar una transición en el espectro metabólico, no se puede descartar que parte de esta variabilidad sea atribuible a factores intrínsecos de la toma de muestras o a la naturaleza diversa de la población estudiada.

En cuanto a la diversidad alfa, la cual mide la riqueza y uniformidad de especies dentro de cada muestra, tanto la riqueza (Figura 12C, número de OTUs observados) como la diversidad (Figura 12D, índice de Shannon) fueron mayores en O y OMS en comparación con NW, alcanzando los valores más altos en OMS. Estos resultados son consistentes con observaciones previas en cohortes pediátricas y adultas que reportan un aumento en la diversidad microbiana asociado con obesidad y SM.

1.3 Abundancia diferencial y biomarcadores microbianos entre los grupos metabólicos

El análisis de abundancia diferencial mediante LEfSe identificó 41 taxones con diferencias significativas entre los grupos (Figura 13). Para el grupo O, destacaron *Porphyromonas* y *Faecalibacterium prausnitzii* como géneros sobreabundantes, este último reconocido por sus propiedades antiinflamatorias y su participación en la producción de butirato. En contraste, el grupo OMS se caracterizó por una sobreabundancia de miembros de la clase Coriobacteriia, incluyendo el género *Collinsella* y la especie *Collinsella aerofaciens*, los cuales mostraron correlaciones positivas con triglicéridos ($r=0.62$, $p=0.00053$) y negativas con colesterol HDL ($r=-0.4$, $p=0.039$). Asimismo, la familia Erysipelotrichaceae y el género *Catenibacterium* estuvieron significativamente aumentados en OMS, mientras que *Parabacteroides distasonis* presentó una reducción notable en este grupo respecto a O. En conjunto, estos resultados evidencian una reestructuración gradual de la microbiota intestinal, caracterizada por el incremento de Firmicutes, la disminución de Bacteroidetes, la expansión de taxones como *C. aerofaciens* y *Catenibacterium*, y la pérdida de especies beneficiosas como *P. distasonis*.

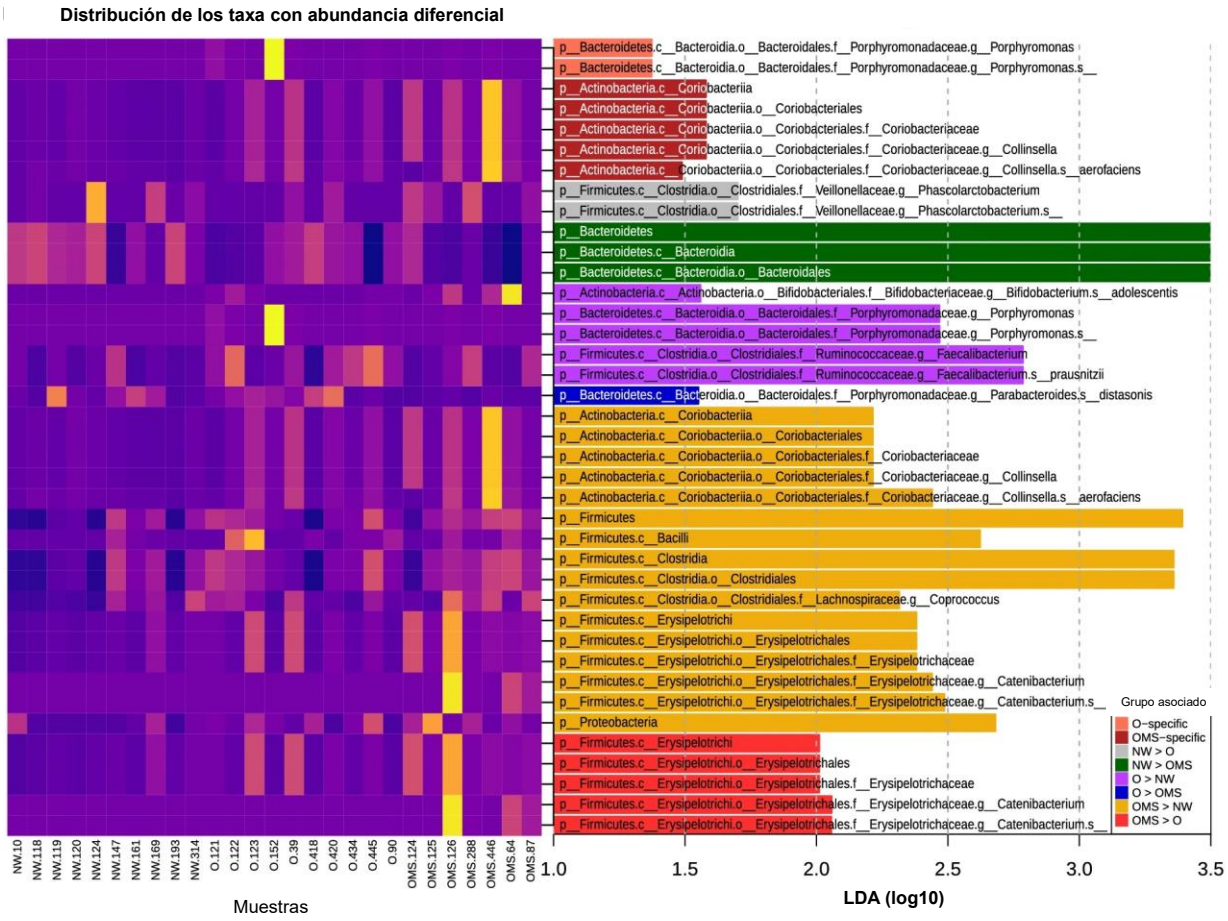


Figura 13: Análisis de biomarcadores mediante LEfSe mostrando los taxones con diferencias significativas entre los grupos NW, O y OMS. A la izquierda, el mapa de calor representa la distribución relativa de los taxones diferencialmente abundantes en las muestras; a la derecha, el gráfico de barras muestra el tamaño del efecto (LDA, log_{10}) de cada taxón asociado a cada condición.

1.4 Análisis funcional del Secrebioma intestinal

El análisis del Secrebioma intestinal se realizó a partir del conjunto de transcritos obtenidos mediante el ensamblado de novo del metatranscriptoma. En total, se obtuvieron aproximadamente 115,000 transcritos codificantes. El primer paso fue la validación del ensamblado y la predicción de proteínas codificantes, verificando la calidad de los transcritos mediante la cobertura de lecturas y la ausencia de contaminantes. Posteriormente, las proteínas candidatas fueron evaluadas con SignalP, TatP, LipoP y SecretomeP para identificar péptidos señal de los sistemas de secreción Sec, Tat y lipoproteínas, respectivamente. Adicionalmente, se analizaron con TMHMM y Phobius para descartar proteínas con múltiples regiones transmembrana y predecir su localización subcelular. Con base en estos análisis, se confirmaron aquellas con péptido señal, ausencia de más de una hélice transmembrana y localización extracelular o periplásmica, características consistentes con proteínas secretadas del Secrebioma. Finalmente, se identificaron 30,024 proteínas (26%) potencialmente secretadas, tras la detección del péptido señal y la exclusión de aquellas con múltiples regiones

transmembrana. Dentro de la clasificación funcional de dichas proteínas mostró que la mayoría correspondía a proteínas extracelulares (47%) y de membrana externa (32%), seguidas de proteínas periplásmicas (12%), mientras que una fracción menor se asoció a la membrana interna (9%).

Adicionalmente, se evaluó la presencia de las proteínas del Secrebioma en metatranscriptomas intestinales públicos mediante análisis de homología, con el propósito de verificar su replicabilidad y expresión en diferentes poblaciones humanas. Aproximadamente el 74% de las proteínas identificadas presentaron homólogos expresados en otros metatranscriptomas intestinales. Este hallazgo respalda la consistencia de estas secuencias en distintos contextos poblacionales y sugiere la relevancia biológica del Secrebioma predicho en el ecosistema intestinal humana.

El análisis de expresión diferencial evidenció un conjunto de transcritos del Secrebioma con diferencias significativas entre los grupos O y OMS (Figura 14). Los transcritos se asociaron con distintos taxones bacterianos, incluidos *F. prausnitzii*, *Paraprevotella xylaniphila*, *Bacteroides spp.*, *Clostridium* y *Escherichia coli*. Las proteínas codificadas por estos transcritos estuvieron principalmente relacionadas con procesos de transporte y ensamblaje de complejos proteicos, así como con rutas de metabolismo de carbohidratos, fosfatos y cofactores redox, indicando funciones asociadas a la actividad metabólica y estructural del microbioma intestinal.

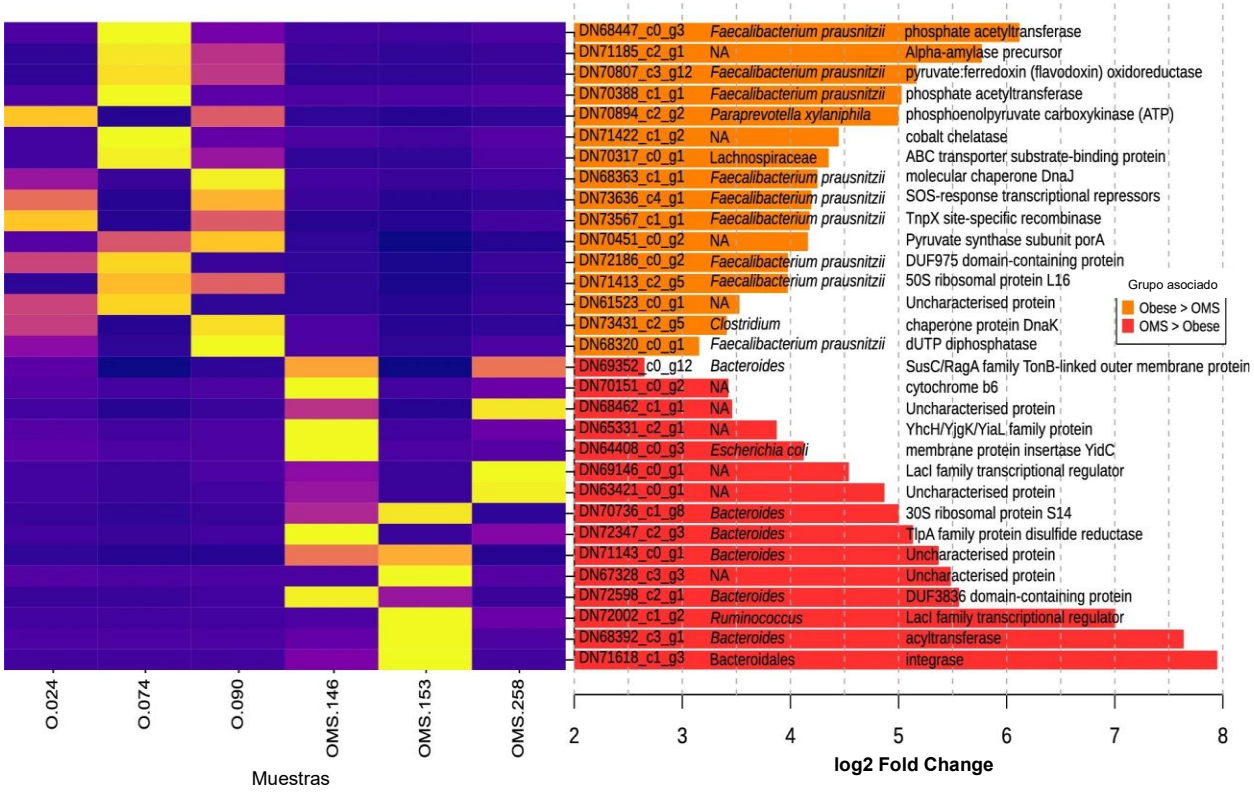


Figura 14: Transcritos diferencialmente expresados del Secrebioma intestinal entre los grupos con O y OMS). El mapa de calor muestra la abundancia normalizada de los transcritos en cada muestra, donde los colores amarillos indican una mayor expresión relativa. A la derecha, el gráfico de barras representa los valores de cambio en la expresión (log_2 Fold Change) determinados mediante DESeq2.

Debido a que las enzimas activas sobre carbohidratos (CAZy) representan un componente esencial para la obtención de energía a partir de la dieta por parte de la microbiota, resulta fundamental evaluar su distribución en el contexto de la obesidad. Con el objetivo de determinar si existían diferencias específicas en estas familias, se comparó la composición del Secrebioma predicho con el conjunto no secretado (Figura 15). El análisis reveló una mayor representación de las categorías SLH (*S-layer homology domain* o dominios de homología de capa S), cohesinas y dockerinas en el Secrebioma, junto con una proporción significativamente superior de proteínas con dominios CBM (*Carbohydrate-Binding Modules* o módulos de unión a carbohidratos) y PL (*Polysaccharide Lyases* o liasas de polisacáridos). Estas familias están asociadas con la degradación de polisacáridos, la unión a carbohidratos y la formación de complejos multienzimáticos, y presentaron diferencias estadísticamente significativas ($p<0.05$).

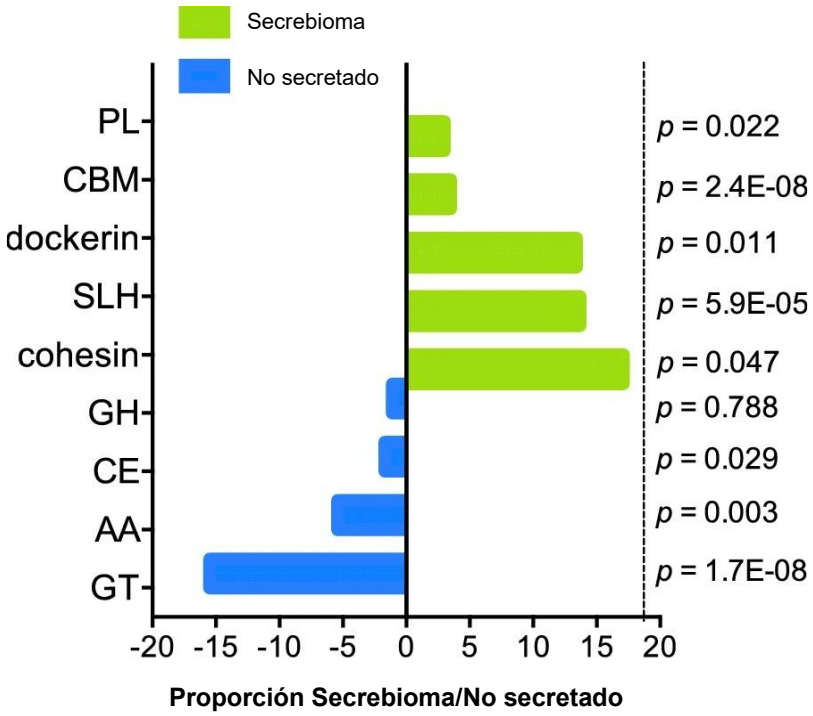


Figura 15: Familias de CAZy entre las proteínas secretadas (verde) y no secretadas (azul). Las barras muestran el valor del cociente secretadas/no-secretadas para cada familia CAZy. Las categorías PL, CBM, dockerina, SLH y cohesina presentaron diferencias significativas ($p<0.05$), mientras que GH y CE no mostraron diferencias significativas.

Es importante mencionar que, a partir del análisis del Secrebioma descrito en este capítulo, no se identificaron proteínas con características compatibles con AMPs. Este resultado llevó a plantear una estrategia metodológica específica con el fin de buscar, predecir y caracterizar dichos péptidos en la microbiota intestinal. Como parte de este proceso, se realizó una revisión del estado del arte sobre los métodos computacionales y experimentales empleados para la detección de AMPs en datos de secuenciación masiva, la cual fue publicada como un artículo de revisión y que se incluye en esta tesis como anexo. Los resultados y la aplicación de la nueva metodología se presentan en el Capítulo 2, centrado en la identificación y análisis funcional de AMPs del microbioma

expresados por la microbiota intestinal, así como en su origen, correlaciones y validación experimental.

Capítulo 2. Predicción, origen y validación de AMPs expresados por la microbiota intestinal

2.1 Predicción de AMPs a partir del metatranscriptoma

El ensamblado *de novo* del metatranscriptoma generado a partir de ocho muestras fecales (NW=2, O=3, OMS=3) produjo 51 087 transcritos no humanos, de los cuales 35 803 correspondieron a smORFs menores de 100 aminoácidos. Tras aplicar un filtro de expresión con FPKM≥1 y presencia en al menos tres muestras, se conservaron 1,095 smORFs, que se utilizaron como base para la búsqueda de AMPs expresados por la microbiota intestinal (Figura 16A).

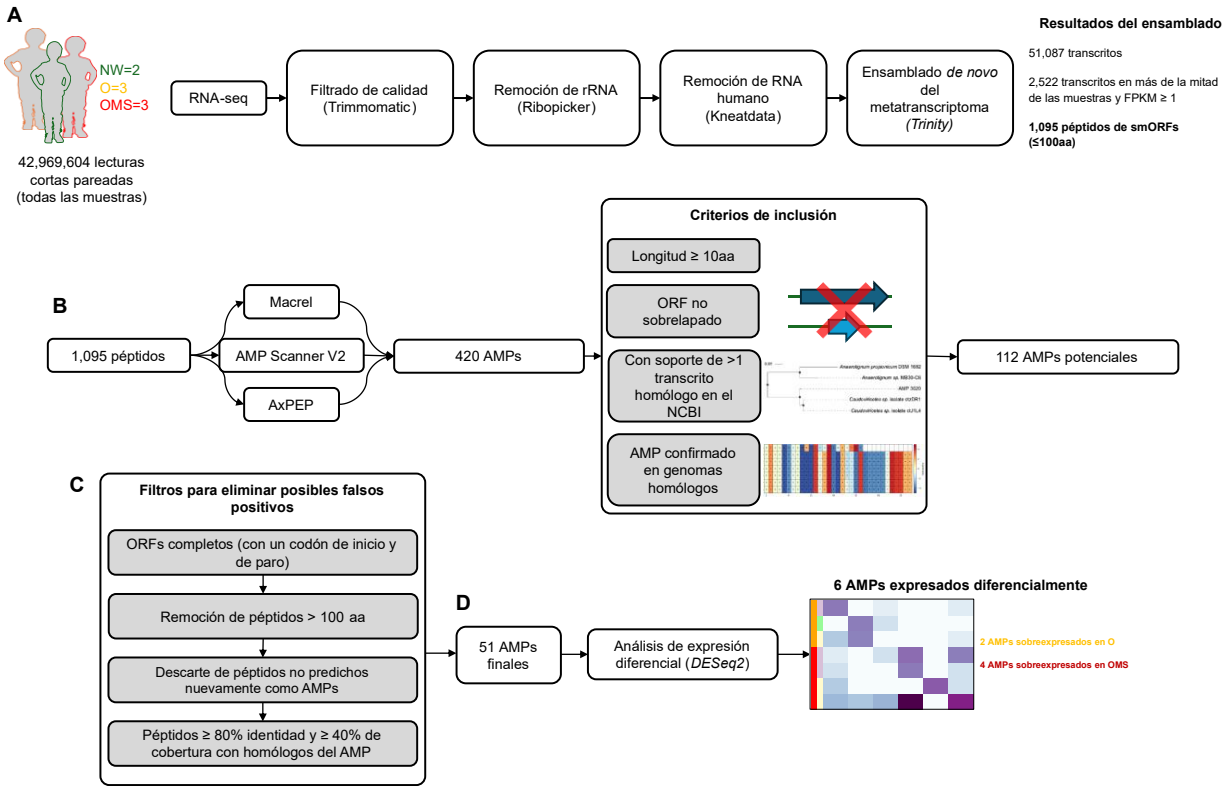


Figura 16: Esquema de identificación y clasificación taxonómica de AMPs. A) Ensamblado y filtrado del metatranscriptoma y detección de smORFs expresados. B) Predicción de péptidos candidatos mediante herramientas integradas. C) Selección y clasificación de 51 AMPs de alta confianza. D) AMPs diferencialmente expresados entre los grupos O y OMS.

A partir de la integración de tres herramientas de predicción complementarias (Macrel, AxPEP y AMP Scanner V2), se identificaron 420 péptidos candidatos con potencial antimicrobiano, que fueron depurados considerando su secuencia y evidencia de homología, obteniéndose un conjunto final de 51 AMPs expresados en el metatranscriptoma intestinal (Figura 16B–C). El análisis taxonómico de estos AMPs mostró que la mayoría se originaron en genomas bacterianos (45/51; 88.2%), mientras que los restantes se asociaron con bacteriófagos (6/51; 11.8%). Entre las bacterias, los principales contribuyentes fueron *F. prausnitzii* (16 AMPs), seguido de *Bacteroides* (4), *Parabacteroides* (3), *Blautia wexlerae* (3), *Ruminococcus bicirculans* (2), *Pseudoclostridium thermosuccinogenes* (2) y *Phocaeicola* (2), además de otras taxa

minoritarias como *Ruminococcus torques*, *Romboutsia ilealis*, *Phocaeicola vulgatus*, *Dorea*, *Coprococcus* y *Escherichia coli.* Dentro de los elementos móviles, se identificaron AMPs asociados a fagos de los órdenes Caudovirales y Caudoviricetes, incluyendo miembros de la familia Siphoviridae y virus intestinales no cultivados. El análisis de expresión diferencial entre los grupos O y OMS reveló seis AMPs expresados diferencialmente, de los cuales cuatro se encontraron sobreexpresados en OMS y dos en O (Figura 16D).

2.2 Expresión diferencial de AMPs entre obesidad y Síndrome Metabólico

El análisis de expresión diferencial permitió identificar seis AMPs expresados diferencialmente entre los grupos O y OMS, cuatro se encontraron sobreexpresados en OMS y dos en O ($(\log_2 FC \geq 0.5$ y padj≤ 0.05, Figura 17). Los AMPs diferencialmente expresados presentaron distintos orígenes genómicos: tres de origen viral (Caudovirales, Caudoviricetes y un virus intestinal no cultivado), dos de origen cromosómico (*Escherichia coli* y *Romboutsia ilealis*) y uno plasmídico (plásmido de *Bacteroides dorei*).

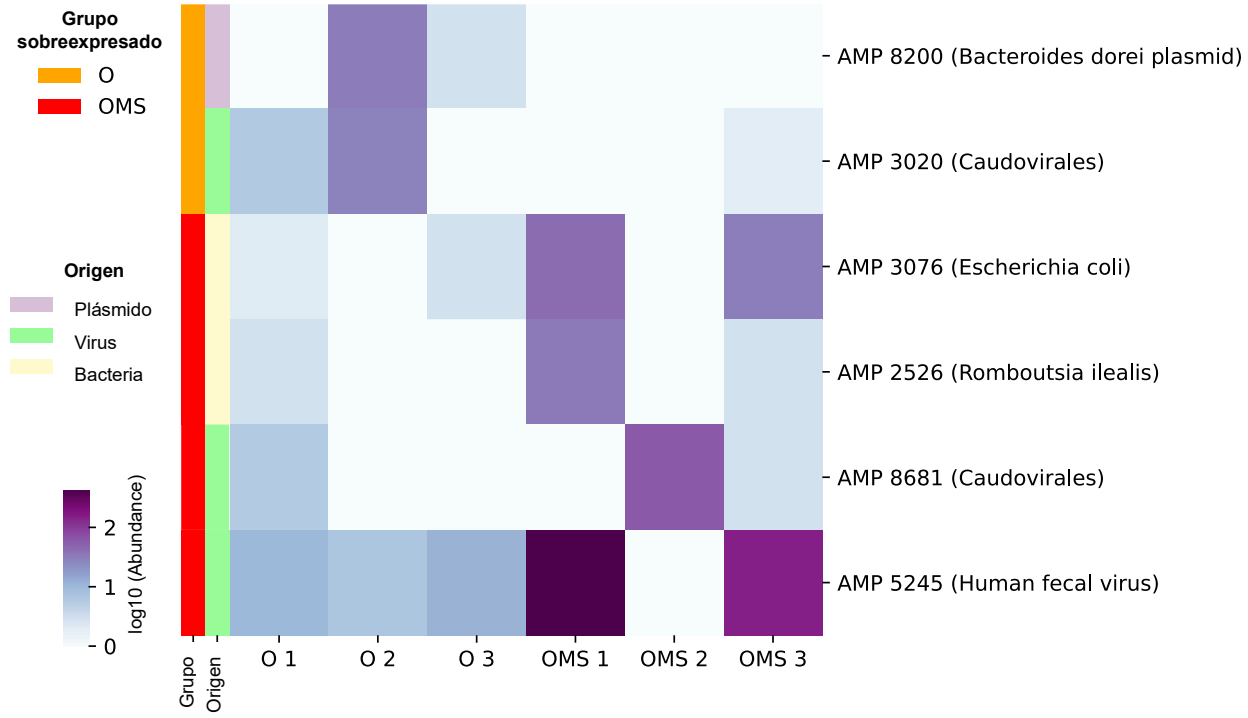


Figura 17: Heatmap de expresión diferencial de los AMPs identificados entre los grupos O y OMS. Los AMPs se agrupan según su origen genómico (bacteriano, viral o plasmídico).

Para determinar la prevalencia de los seis AMPs sobreexpresados, se analizó una cohorte externa de 372 metatranscriptomas intestinales humanos (BioProject PRJNA354235) [32], compuesta por individuos adultos sanos (Figura 18).. El objetivo de este análisis fue evaluar la distribución de estos péptidos en una población sin las patologías metabólicas estudiadas para determinar si su presencia es común y conservada en el microbioma humano. Los resultados mostraron que todos los AMPs se

encontraron presentes en una fracción considerable de las muestras, con una frecuencia de detección que varió entre 9.8% y 98.4% (media=68%).

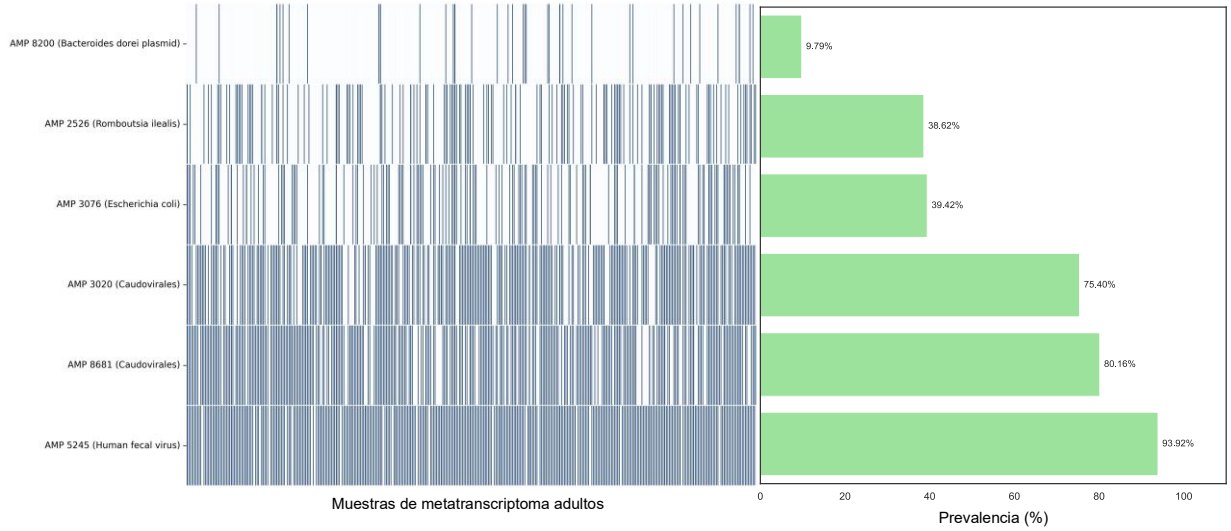


Figura 18: Prevalencia de los seis AMPs sobreexpresados en una cohorte independiente de 372 metatranscriptomas intestinales humanos (BioProject PRJNA354235 [32]). Cada barra representa el porcentaje de muestras en las que se detectó cada AMP, clasificados según su origen genómico (bacteriano, viral o plasmídico).

2.3 Análisis filogenómico y contextual de los AMPs expresados diferencialmente

El análisis filogenómico se realizó con el propósito de caracterizar la diversidad evolutiva y el contexto genómico de los AMPs expresados diferencialmente en el metatranscriptoma intestinal. Para todos los AMPs expresados diferencialmente, se efectuó un análisis comparativo a nivel de secuencia de ADN y proteína, complementado con la exploración de su contexto genómico en diferentes especies relacionadas. Este enfoque permitió evaluar la conservación evolutiva y la posible integración funcional de estos genes dentro de los genomas bacterianos.

El AMP 3076 presentó una alta similitud con secuencias homólogas de distintas cepas de *Escherichia coli*, mostrando una relación filogenética cercana y la formación de un clado monofilético con baja divergencia evolutiva (Figura 19A). La alineación múltiple de secuencias peptídicas evidenció una conservación completa del péptido, sin variaciones en la secuencia de aminoácidos entre las cepas analizadas (Figura 19B). Por su parte, el análisis del contexto genómico mostró una organización altamente conservada de los genes adyacentes entre las cepas bacterianas analizadas, manteniendo un patrón estructural homogéneo alrededor del gen que codifica el AMP (Figura 19C).

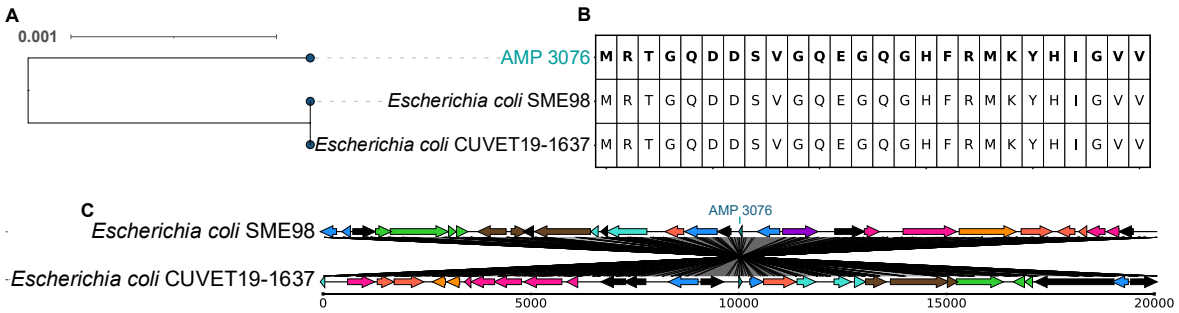


Figura 19: Análisis filogenómico y contextual del AMP 3076. A) Árbol filogenético basado en secuencias de ADN que muestra la relación del transcrito con sus genomas homólogos. B) Alineación múltiple de las secuencias peptídicas. C) Comparación del contexto genómico entre E. coli SME98 y E. coli CUVET19-1637, mostrando los genes adyacentes al AMP 3076.

De manera similar, el AMP 2526 presentó homología con proteínas bacterianas conservadas en distintas especies del género *Rombustia*, formando un agrupamiento filogenético estrechamente relacionado (Figura 20A). La alineación de secuencias peptídicas mostró una conservación elevada de residuos hidrofóbicos y cargados positivamente, típicos de péptidos con potencial antimicrobiano (Figura 20B). El análisis del contexto genómico indicó que el gen del AMP se encuentra dentro de regiones genómicas conservadas, flanqueado por genes estructurales y regulatorios compartidos entre los genomas comparados (Figura 20C).

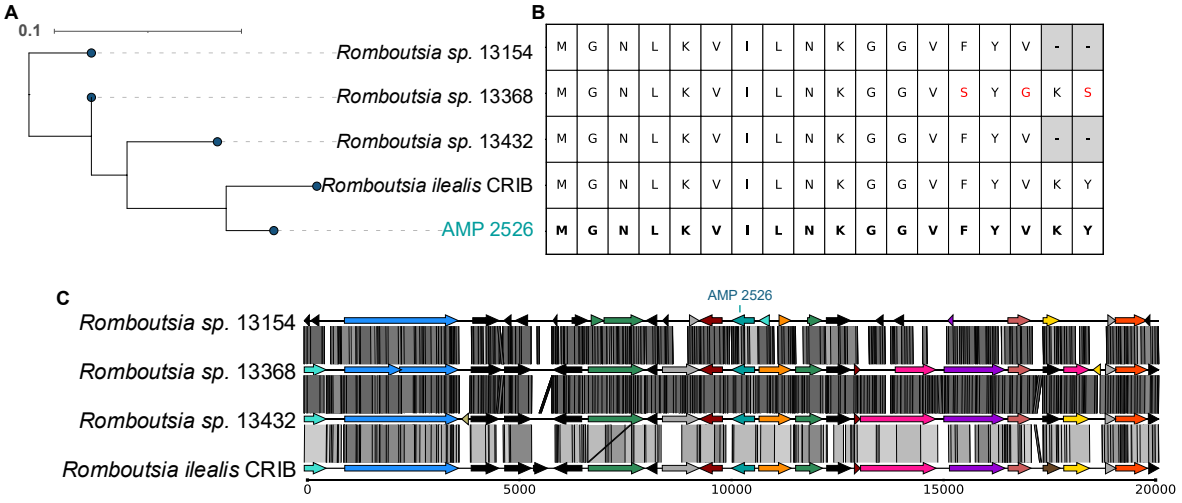


Figura 20: Análisis filogenómico y contextual del AMP 2526. A) Árbol filogenético basado en secuencias de ADN que muestra la relación del transcrito con sus genomas homólogos. B) Alineación múltiple de las secuencias peptídicas. C) Comparación del contexto genómico entre las especies del género Bacteroides, mostrando los genes adyacentes al AMP 2526.

El AMP 8200 se identificó en un plásmido asociado a *Bacteroides dorei*, y su análisis filogenómico permitió establecer su relación con secuencias homólogas presentes en genomas bacterianos y plasmídicos estrechamente relacionados. El análisis de agrupamiento mostró la agrupación del AMP 8200 dentro de un clado que incluye secuencias compartidas entre distintas especies del género *Bacteroides* y el plásmido anteriormente mencionado (Figura 21A). La alineación múltiple de secuencias peptídicas evidenció un alto grado de similitud entre las variantes, con una conservación

completa en las ramas filogenéticamente más cercanas al transcrito identificado en nuestro metatranscriptoma (Figura 21B). El análisis del contexto genómico mostró la organización de genes adyacentes al AMP en el plásmido de *Bacteroides dorei*, incluyendo marcos de lectura asociados con funciones de replicación y transferencia conjugativa, características comunes de elementos plasmídicos activos (Figura 21C).

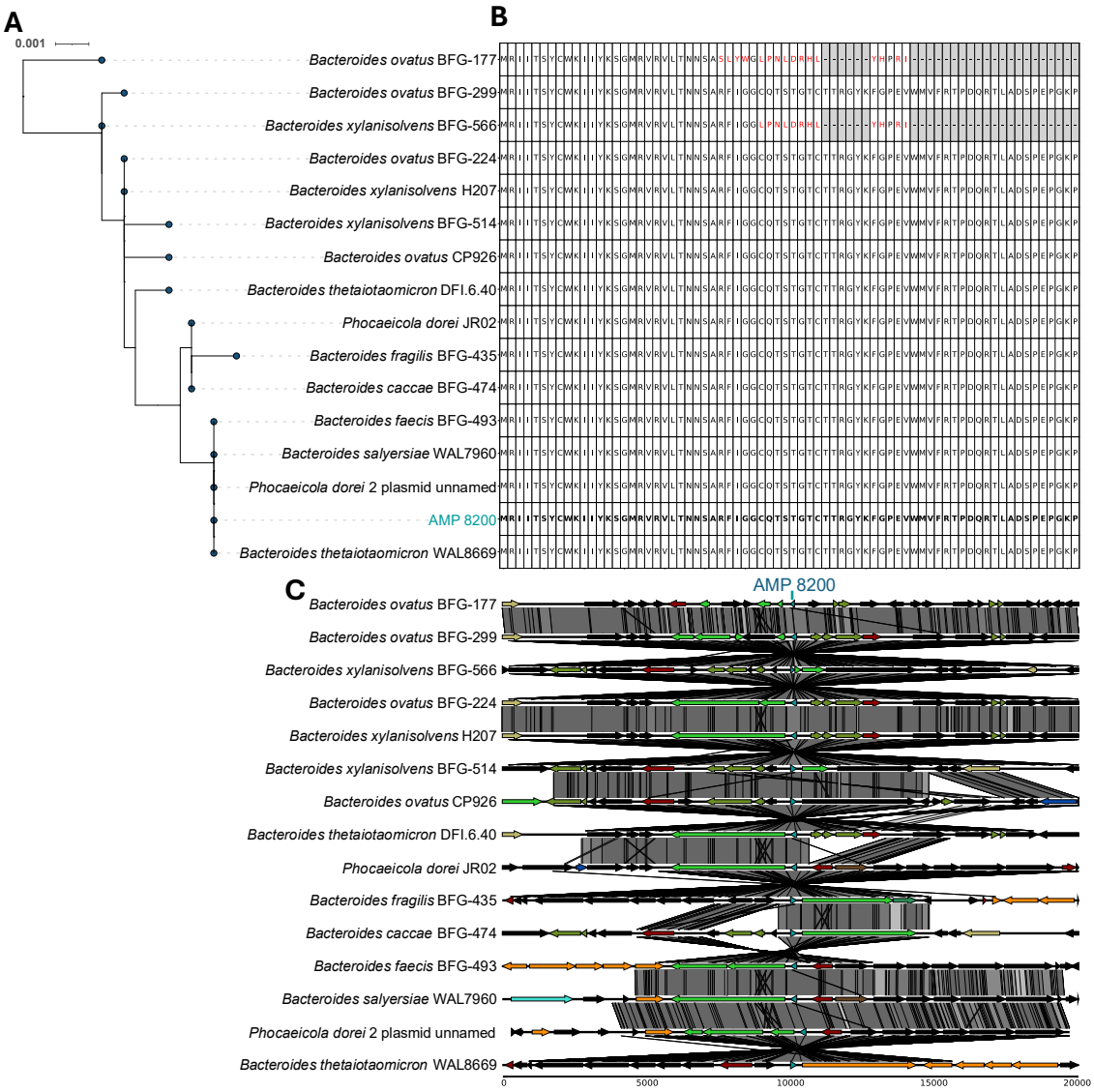


Figura 21: Análisis filogenómico y contextual del AMP 8200 de origen plasmídico. A) Árbol filogenético basado en secuencias de ADN que muestra la relación del transcrito con sus genomas homólogos. B) Alineación múltiple de las secuencias peptídicas. C) Comparación del contexto genómico entre el plásmido y especies del género Bacteroides, mostrando los genes adyacentes al AMP 8200.

En el caso del AMP 3020, identificado dentro del genoma del fago Caudoviricetes sp. ctJ1L4, con alta similitud respecto a dos especies del género *Anaerotignum* (Figura 22A). La alineación múltiple de las secuencias peptídicas mostró una alta similitud entre las variantes identificadas, con diferencias localizadas principalmente en la región C-terminal del péptido (Figura 22B). El análisis del contexto genómico mostró que el gen del AMP se encuentra próximo a genes estructurales del fago, incluyendo proteínas de

cápside y ensamblaje, lo que resulta coherente con su ubicación dentro de un genoma viral completo y funcional (Figura 22C).

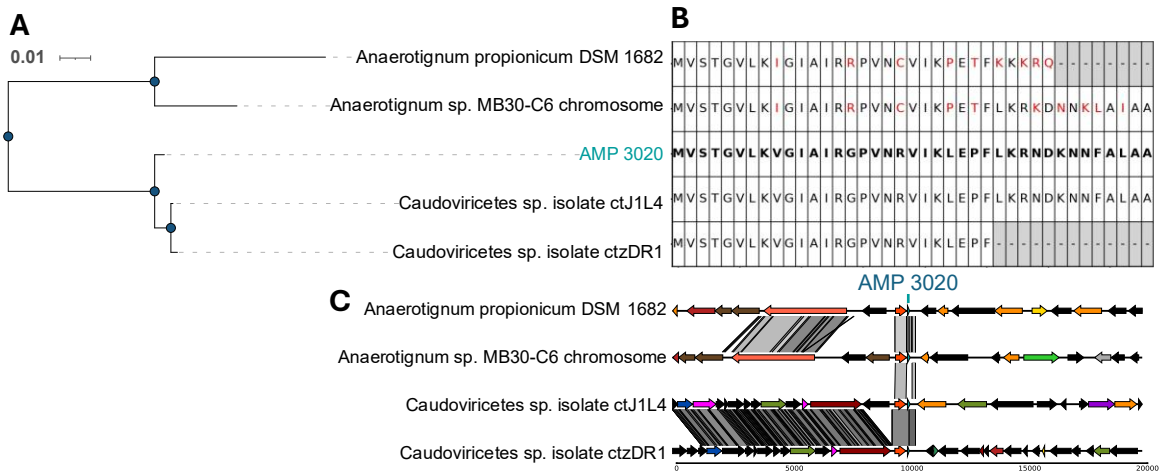


Figura 22: Análisis filogenómico y contextual del AMP 3020 de origen viral. A) Árbol filogenético basado en secuencias de ADN que muestra la relación del transcrito con sus genomas homólogos. B) Alineación múltiple de las secuencias peptídicas. C) Comparación del contexto genómico de los fagos y las bacterias homólogas, mostrando los genes adyacentes al AMP 3020.

El AMP 8681, identificado dentro del genoma del fago Caudoviricetes sp. ctIN07, mostró una relación filogenética cercana con otros fagos de este clado (Figura 23A). El árbol filogenético basado en secuencias de ADN evidenció una agrupación consistente con genomas virales del orden Caudoviricetes, lo que respalda su origen viral. La alineación múltiple de secuencias peptídicas reveló una alta similitud entre las variantes del AMP, con ligeras variaciones (Figura 23B). El análisis del contexto genómico mostró que el gen del AMP se encuentra flanqueado por marcos de lectura relacionados con funciones estructurales virales y proteínas implicadas en el ensamblado del fago, en concordancia con su ubicación dentro de un genoma viral activo (Figura 23C).

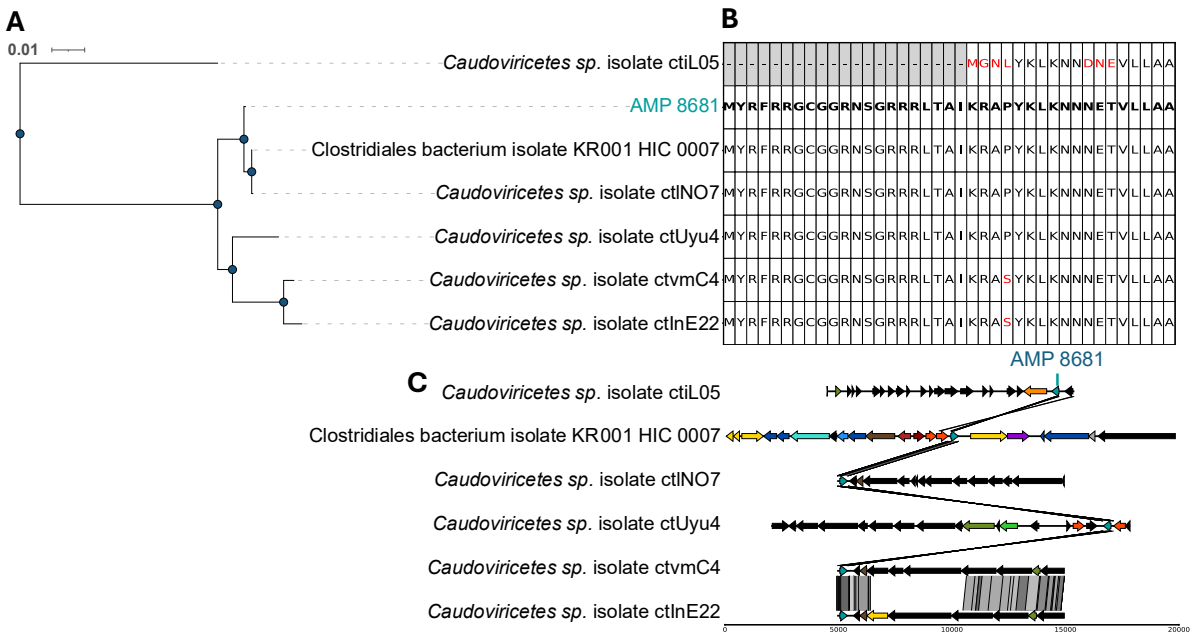


Figura 23: Análisis filogenómico y contextual del AMP 8681 de origen viral. A) Árbol filogenético basado en secuencias de ADN que muestra la relación del transcrito con sus genomas homólogos. B) Alineación múltiple de las secuencias peptídicas. C) Comparación del contexto genómico de los fagos y bacterias homólogas, mostrando los genes adyacentes al AMP 8681.

El AMP 5245 fue identificado en el genoma de un *Uncultured human fecal virus* y en bacterias del género *Blautia wexlerae*, lo que sugiere a esta especie como posible hospedera. El análisis filogenético basado en ADN mostró que el AMP 5245 forma un clado definido junto con secuencias de fagos intestinales no cultivados, indicando su pertenencia a linajes virales presentes en el microbioma humano (Figura 24A). La alineación múltiple de secuencias peptídicas reveló una alta conservación entre las variantes identificadas, sin diferencias en la secuencia de aminoácidos principales (Figura 24B). El análisis del contexto genómico evidenció que el gen del AMP se localiza dentro de un bloque conservado de genes virales, incluyendo marcos de lectura adyacentes asociados con funciones estructurales, ensamblaje y empaquetamiento, características típicas de genomas de fagos completos (Figura 24C).

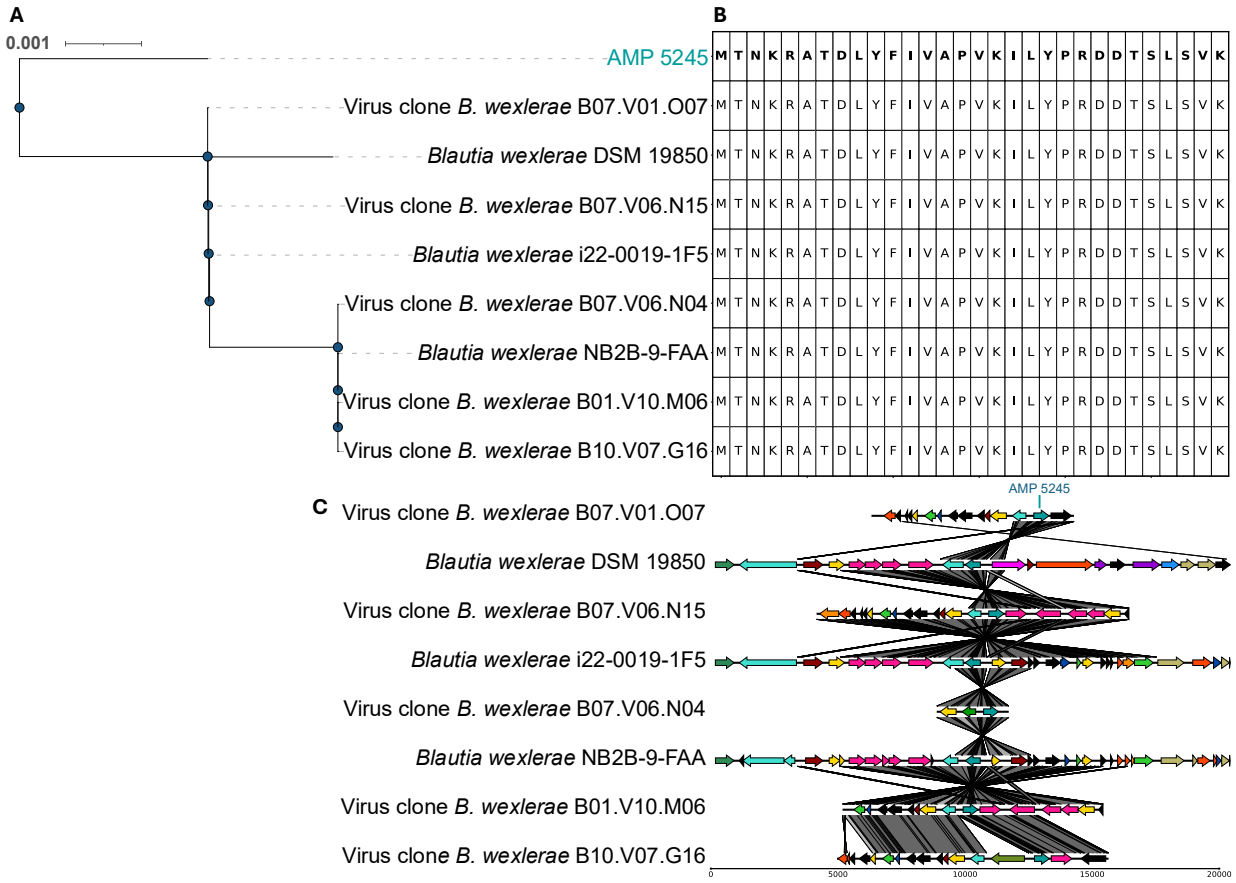


Figura 24: Análisis filogenómico y contextual del AMP 5245 de origen viral. A) Árbol filogenético basado en secuencias de ADN que muestra la relación del transcrito con sus genomas homólogos. B) Alineación múltiple de las secuencias peptídicas. C) Comparación del contexto genómico de los virus y sus bacterias homólogas, mostrando los genes adyacentes al AMP 5245.

2.4 Comprobación del origen plasmídico y viral de los AMPs expresados diferencialmente

Con el objetivo de comprobar el origen genético de los AMPs codificados en elementos móviles, se integraron los resultados del análisis metatranscriptómico con la información genómica del plásmido y del cromosoma bacteriano, así como con los datos del viroma obtenidos a partir de las mismas muestras. Este enfoque permitió confirmar que una parte de los AMPs se encuentra codificada en elementos móviles activos, como fagos y plásmidos, lo que respalda su participación en procesos funcionales de origen móvil dentro del ecosistema intestinal.

En el caso del AMP 8200 de origen plasmídico, este se identificó dentro de un transcrito con alta similitud al plásmido de *Bacteroides dorei*, mostrando una identidad de secuencia superior al 90% y una cobertura de alineamiento del 95.5%. El mapeo de lecturas de *RNA-seq* confirmó la expresión activa del plásmido completo, incluyendo las regiones adyacentes al gen del AMP, lo que respalda su origen plasmídico y actividad transcripcional en las muestras analizadas (Figura 25A), mientras que la cobertura de

RNA-seq (95.55%) a lo largo del plásmido muestra la posición del gen que codifica el AMP, señalada en rojo (Figura 25B).

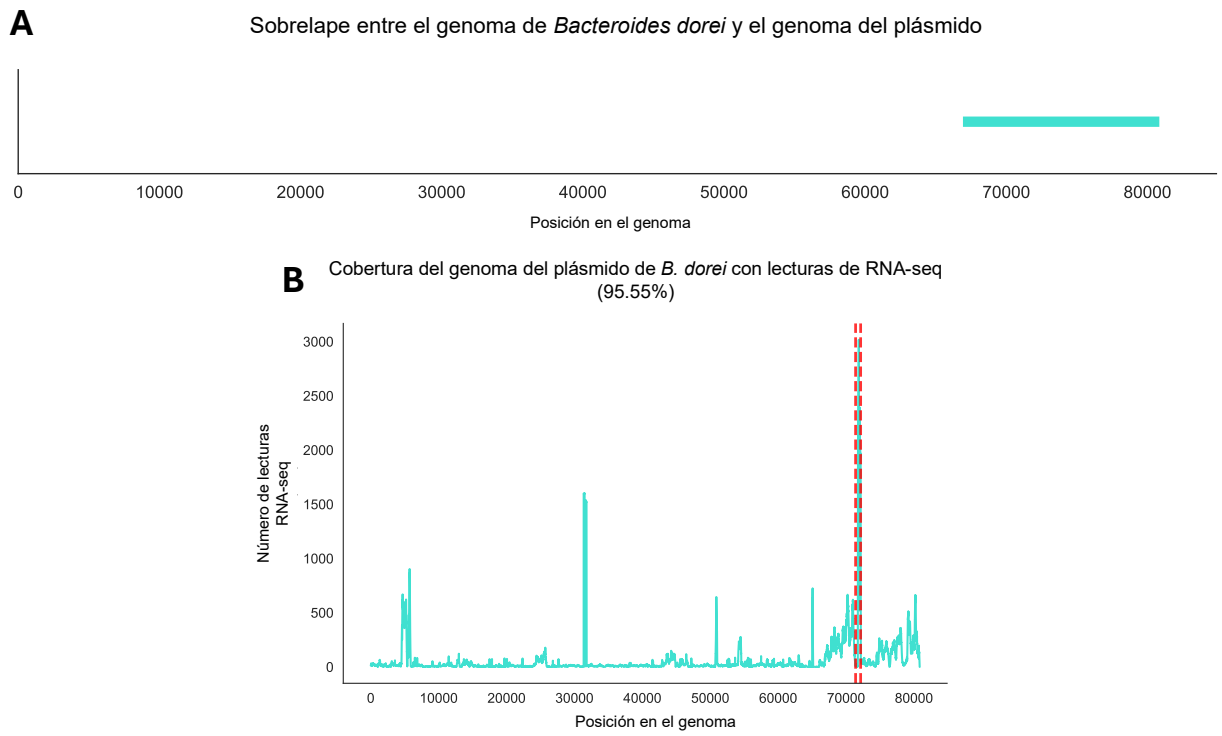


Figura 25: Análisis del origen del AMP 8200. A) Sobrelape entre el genoma de Bacteroides dorei y el plásmido identificado con el transcrito codificante del AMP. B) Cobertura de lecturas de RNA-seq a lo largo del plásmido, con una cobertura total del 95.55%. La posición del gen que codifica el AMP se muestra en rojo.

Por otro lado, la verificación del origen viral de los AMPs se realizó mediante el mapeo de lecturas del viroma y del metatranscriptoma sobre los genomas virales que contenían los péptidos correspondientes. Este análisis permitió evidenciar su existencia como fagos completos en el viroma de las mismas muestras, además de identificar la expresión activa de los genes que codifican los AMPs. En el caso del AMP 3020, localizado dentro del genoma del fago Caudoviricetes sp. ctJ1L4, el mapeo del viroma mostró una cobertura total del 70.46% a lo largo del genoma viral, evidenciando la presencia de partículas virales activas en las mismas muestras analizadas pertenecientes a este fago (Figura 26A). De manera complementaria, el mapeo de lecturas del *RNA-seq* presentó una cobertura del 14.96% sobre el mismo genoma, lo que confirma la expresión del AMP en un contexto de transcripción viral activa (Figura 26B). Finalmente, el sobrelape entre el genoma del fago y el de *Anaerotignum sp.* reveló una integración parcial de ambas secuencias, indicando una posible interacción fago-hospedero que favorecería la conservación y expresión del gen codificante del AMP dentro del microbioma intestinal (Figura 26C).

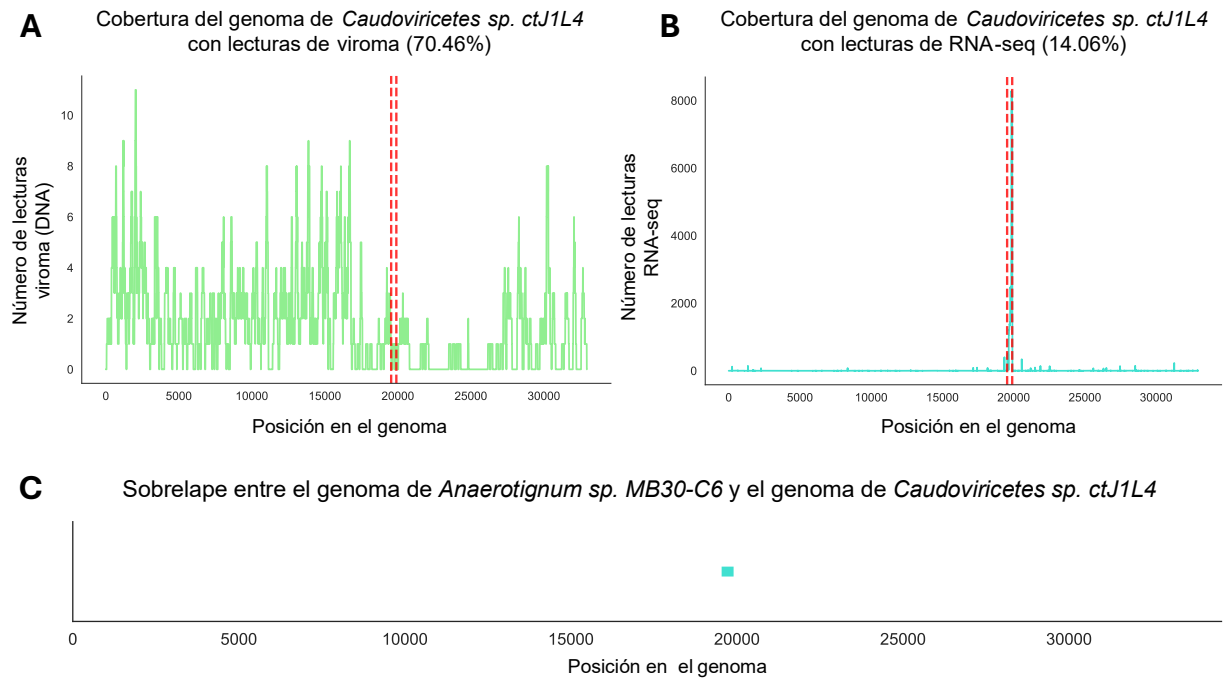


Figura 26: Análisis del origen del AMP 3020. A) Cobertura del viroma a lo largo del genoma del fago Caudoviricetes sp. ctJ1L4. B) Cobertura de lecturas de RNA-seq sobre el mismo genoma, mostrando la posición del gen que codifica el AMP marcada en rojo. C) Sobrelape del genoma de Anaerotignum sp. con la región del fago que contiene el AMP.

De manera similar, el AMP 8681 fue identificado dentro del genoma del fago Caudoviricetes sp. ctIN07. El mapeo de lecturas del viroma mostró cobertura a lo largo del genoma viral, evidenciando la presencia de partículas virales en las mismas muestras (Figura 27A). El mapeo de lecturas de *RNA-seq* reveló una señal de expresión sobre la misma región, confirmando la transcripción activa del gen que codifica este AMP (Figura 27B). Finalmente, el sobrelape con el genoma de Clostridiales bacterium isolate KR001 mostró una región homóloga adyacente al gen del AMP, indicando una relación genómica entre el fago y la bacteria correspondiente (Figura 27C).

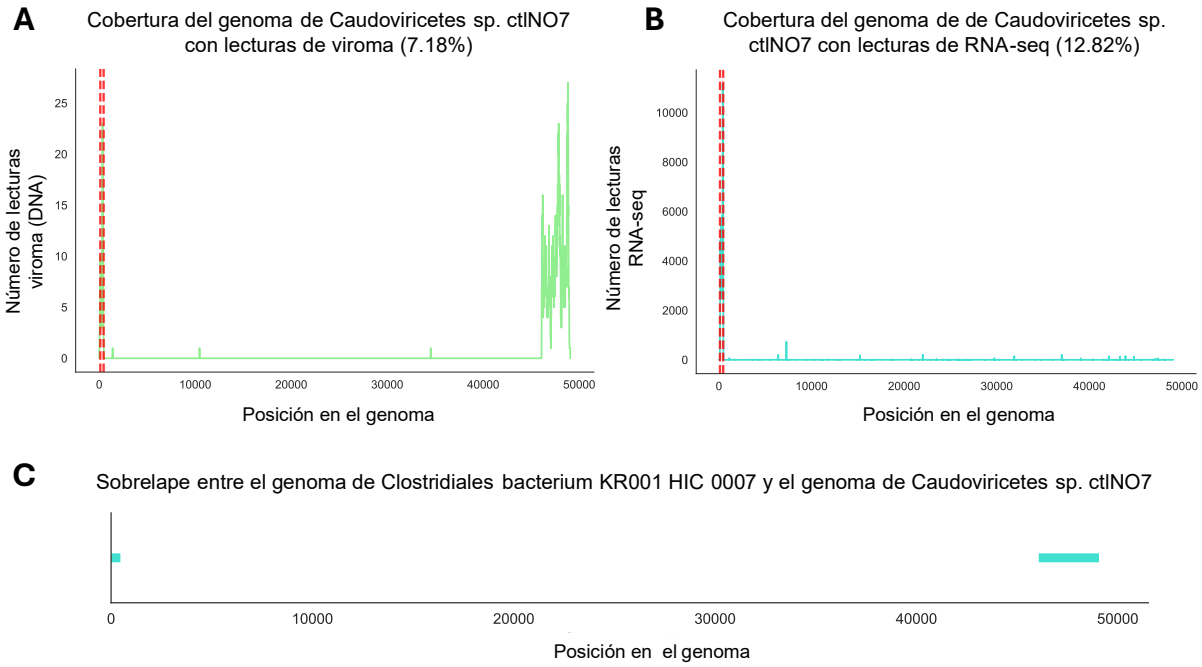


Figura 27: Análisis del origen del AMP 8681. A) Cobertura del viroma sobre el genoma del fago Caudoviricetes sp. ctlN07. B) Cobertura de lecturas de RNA-seq sobre el mismo genoma, mostrando la posición del gen que codifica el AMP marcada en rojo. C) Sobrelape del genoma de Clostridiales bacterium isolate KR001 con la región del fago que contiene el AMP.

Finalmente, el AMP 5245 fue identificado dentro del genoma de un Uncultured human fecal virus. El mapeo de lecturas del viroma mostró cobertura a lo largo del genoma viral, mientras que el análisis de *RNA-seq* evidenció una señal coincidente en la misma región, indicando la expresión del gen que codifica este AMP (Figura 28A–B). El sobrelape con el genoma de *Blautia wexlerae* DSM 19850 mostró regiones homólogas entre ambos genomas, lo que confirma la relación entre el fago y su hospedero bacteriano (Figura 28C).

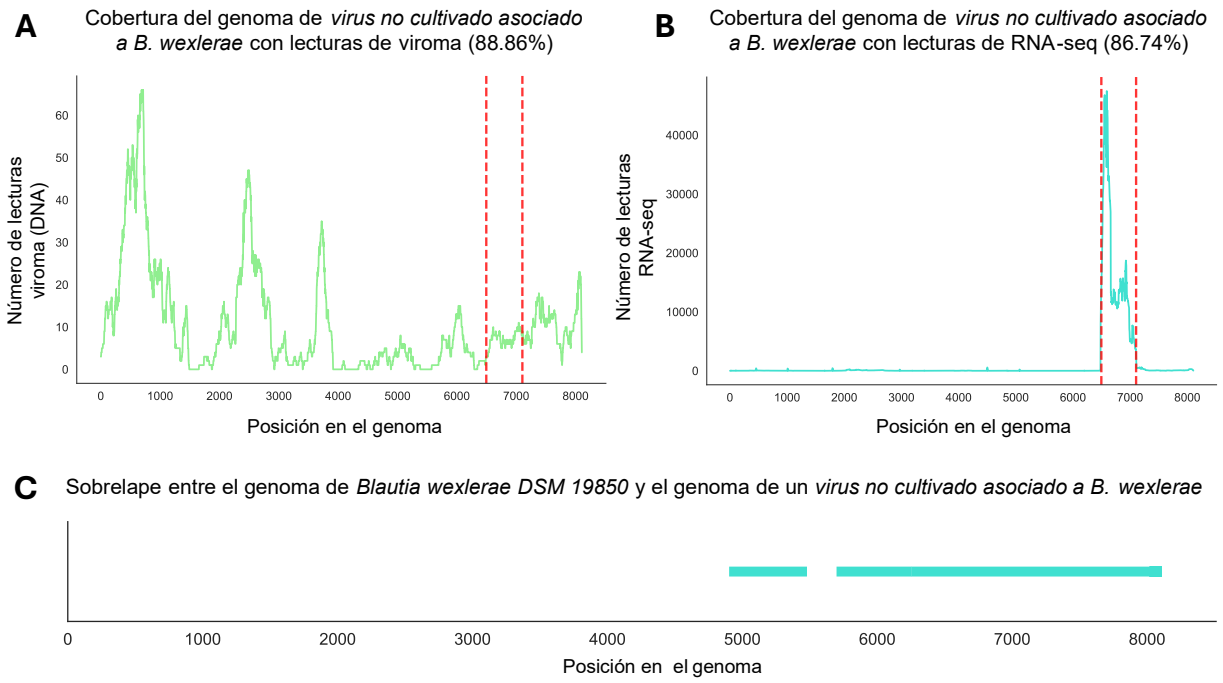


Figura 28: Análisis del origen del AMP 5245. A) Cobertura del viroma sobre el genoma del Uncultured human fecal virus. B) Cobertura de lecturas de RNA-seq sobre el mismo genoma, mostrando la posición del gen que codifica el AMP marcada en rojo. C) Sobrelape del genoma de Blautia wexlerae DSM 19850 con la región del fago que contiene el AMP.

2.5 Correlaciones entre la expresión de AMPs y la composición microbiana intestinal

Con el objetivo de analizar la relación entre la expresión de los AMPs diferencialmente expresados y la composición microbiana intestinal, se integraron los niveles de expresión obtenidos por metatranscriptómica con la abundancia relativa de taxones bacterianos derivada del análisis del gen ARNr 16S. Las correlaciones se estimaron mediante el coeficiente de Spearman, utilizando los valores de expresión normalizados (FPKM) de los AMPs y la abundancia relativa de los taxones detectados en las muestras. Se identificaron asociaciones significativas ($p \leq 0.05$) entre la expresión de los AMPs y distintos grupos bacterianos, evidenciando correlaciones negativas (Figura 29A) como positivas (Figura 29B).

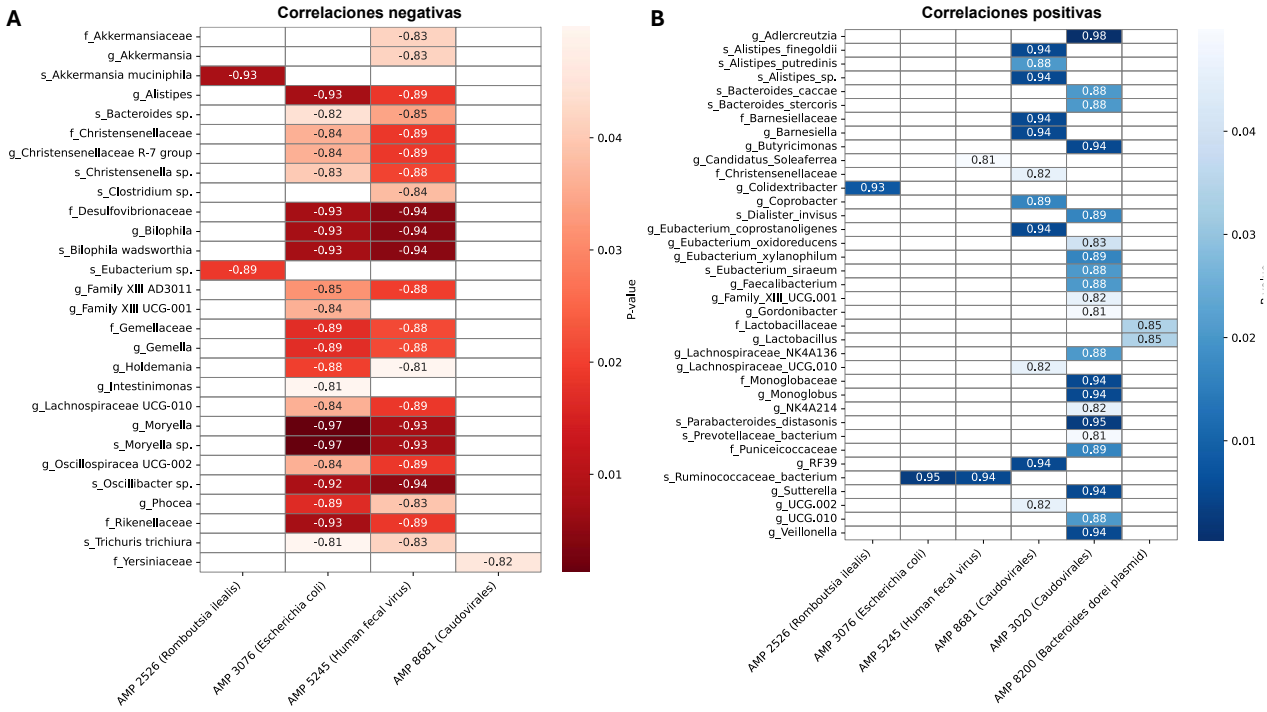


Figura 29: Correlaciones entre la expresión de los AMPs expresados diferencialmente y la microbiota intestinal. A) Correlaciones negativas (rojo) entre los AMPs y taxones asociados con salud intestinal. B) Correlaciones positivas (azul) entre los AMPs y taxones asociados con disbiosis o alteraciones metabólicas. Los coeficientes de Spearman (ρ) se muestran por intensidad de color (p≤0.05).

Entre las correlaciones negativas se identificaron asociaciones con taxones vinculados a un perfil intestinal saludable, como *Akkermansia muciniphila*, Christensenellaceae R-7 group, Oscillospiraceae y *Alistipes*, con coeficientes de correlación entre –0.8 y –0.9 respecto a los AMPs 2526, 3076 y 5245. Asimismo, varios géneros de Lachnospiraceae y *Eubacterium* mostraron correlaciones negativas consistentes, indicando que la expresión de estos AMPs coincide con una menor abundancia de bacterias consideradas benéficas. En contraste, las correlaciones positivas se observaron principalmente en taxones asociados con disbiosis o alteraciones metabólicas, como *Adlercreutzia*, *Dialister*, *Veillonella*, *Sutterella* y *Colidextribacter*, con coeficientes entre 0.8 y 0.95. En particular, los AMPs 3020, 8681 y 8200 mostraron correlaciones positivas con *Ruminococcaceae bacterium*, *Lactobacillus* y *Eubacterium coprostanoligenes*, respectivamente, reflejando distintos patrones de coexpresión entre los grupos O y OMS.

Los resultados detallados de estas correlaciones, que incluyen los taxones asociados, la dirección de la relación y su correspondencia con reportes previos, se presentan en la Tabla 1, donde se agrupan las asociaciones de cada AMP con taxones vinculados tanto a perfiles metabólicamente saludables como a condiciones relacionadas con la obesidad.

Tabla 1: Correlaciones significativas entre los AMPs expresados diferencialmente y la composición bacteriana intestinal. Se indican la dirección de la correlación (positiva o negativa), los taxones asociados (a nivel de familia,

género o especie), su relación reportada con perfiles metabólicamente saludables o vinculados a obesidad, y las referencias correspondientes.

AMP (ID)	**Tipo de correlación**	**Taxa correlacionada (Familia/Género/Especie)**	**Asociación con Salud o con Obesidad**	**Referencias**
AMP 2526	↓ Negativa	*A. muciniphila*	Asociado a Salud	[64], [65]
AMP 2526	↓ Negativa	*Eubacterium sp.*	Asociación mixta	[66], [67]
AMP 3076 y AMP 5245	↓ Negativa	*Bacteroides, Oscillibacter, Alistipes, Holdemia, Christensenella spp, Christensenellaceae R-7 group, Lachnospiraceae UCG-010,*	Asociado a Salud	[14], [68], [69], [70], [71], [72], [73], [74], [75], [76]
AMP 3076	↓ Negativa	*Intestimonas*	Asociado a Salud	[77]
AMP 5245	↓ Negativa	*Clostridium spp*	Asociación mixta	[78], [79]
AMP 5245	↓ Negativa	*Akkermansia*	Asociado a Salud	[64], [80]
AMP 8681	↓ Negativa	*Yersiniaceae*	Asociado a Obesidad	[81]
AMP 2526	↑ Positiva	*Colidextribacter*	Asociado a Obesidad	[82], [83]
AMP 3076 y AMP 5245	↑ Positiva	*Ruminococcaceae bacterium*	Asociado a Obesidad	[84]
AMP 5245	↑ Positiva	*Candidatus Soleaferrea*	Asociación mixta	[65]
AMP 8681	↑ Positiva	*Alistipes spp., Barnesiella, Coprobacter, Eubacterium coprostanoligenes, Lachnospiraceae UCG-010, RF39*	Asociado a Salud	[14], [68], [69], [70], [71], [72], [73], [74], [75], [76]
AMP 3020	↑ Positiva	*Adlercreutzia, Dialister invisus, Eubacterium oxidoreducens, Monoglobus, Family XIII UCG-001, Sutterella, Veillonella*	Asociado a Obesidad	[85], [86], [87], [88], [89], [90], [91]
AMP 3020	↑ Positiva	*Bacteroides caccae, Butyricimonas, Prevotellaceae bacterium, Veillonella*	Asociación mixta	[92], [93], [94]
AMP 3020	↑ Positiva	*Bacteroides stercoris, Eubacterium coprostanoligenes, Eubacterium xylanophilum, Eubacterium siraeum, Faecalibacterium, Parabacteroides distasonis*	Asociado a Salud	[66], [73], [95], [96], [97], [98]

AMP 8200	↑ Positiva	*Lactobacillus*	Asociado a Salud	[99]

2.6 Validación funcional del AMP 3020

Para evaluar experimentalmente el potencial antimicrobiano de los péptidos identificados, se seleccionó el AMP 3020, derivado de un bacteriófago de la clase Caudoviricetes. Este péptido fue elegido por su expresión diferencial y su origen viral confirmado mediante análisis genómicos y de homología. El AMP 3020 se sintetizó químicamente en dos variantes: ADR1, que conserva el residuo metionina inicial, y ADR2, que carece de este residuo y comienza con valina (Figura 30A).

Ambas variantes se purificaron por HPLC y se confirmaron por espectrometría de masas. La actividad antimicrobiana se evaluó mediante ensayos de microdilución en caldo frente a bacterias Gram negativas (*Pseudomonas aeruginosa* ATCC 27853, *Klebsiella pneumoniae* ATCC 700603) y Gram positivas (*Staphylococcus aureus* ATCC 29213, *Streptococcus pneumoniae* ATCC 46916). En todos los casos, ADR1 y ADR2 redujeron de manera significativa el crecimiento bacteriano en comparación con los controles no tratados al utilizar una concentración de 100 µg/mL ($p \leq 0.05$). ADR1 mostró una mayor inhibición frente a *P. aeruginosa*, mientras que ADR2 presentó mayor efecto sobre *K. pneumoniae*. Ambas variantes inhibieron el crecimiento de *S. pneumoniae*, aunque sólo ADR1 lo hizo en *S. aureus* (Figura 30B-E).

La posible citotoxicidad de los péptidos se evaluó en células mononucleares de sangre periférica (PBMCs) de donadores sanos expuestas a 20 µg/mL de cada variante. La viabilidad de las subpoblaciones T ($CD3^+$, $CD4^+$ y $CD8^+$) se determinó por citometría de flujo, sin observar incrementos significativos en la muerte celular respecto a los controles (Figura 30F).

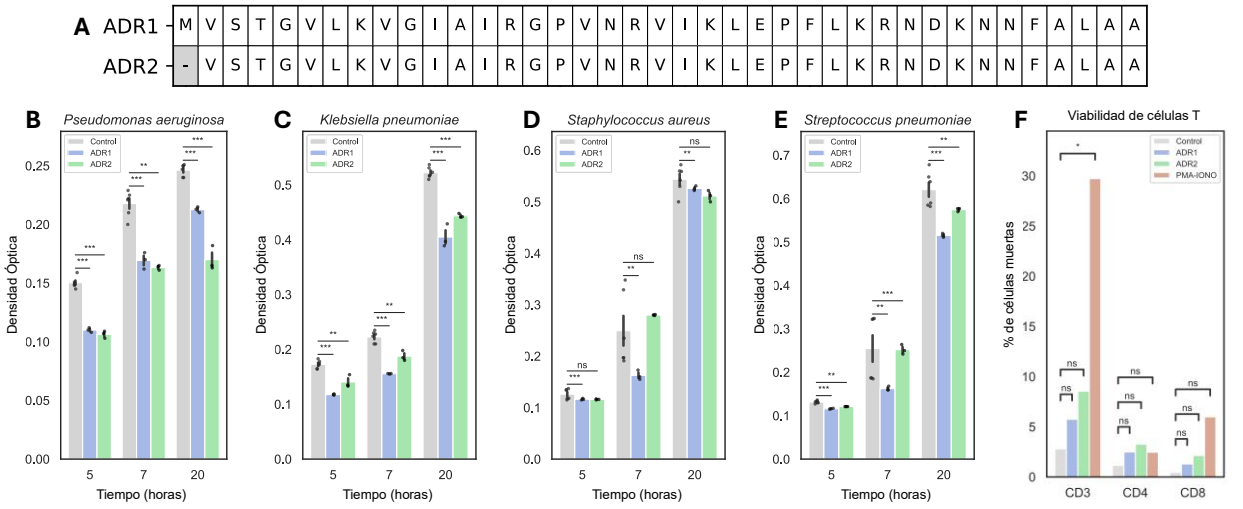


*Figura 30: Evaluación experimental de la actividad antimicrobiana y citotoxicidad de las variantes ADR1 y ADR2 derivadas del AMP 3020 (100 µg/mL). A) Secuencias en aminoácidos de ADR1 (con metionina inicial) y ADR2 (sin metionina). Inhibición del crecimiento de B) Pseudomonas aeruginosa, C) Klebsiella pneumoniae, D) Staphylococcus aureus y E) Streptococcus pneumoniae. F) Viabilidad de linfocitos T humanos ($CD3^+$, $CD4^+$ y $CD8^+$) determinada por citometría de flujo. (Significancia estadística: *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001; ns: no significativo).*

Adicionalmente, con el fin de descartar efectos inespecíficos derivados de la concentración proteica en el ensayo y evaluar la persistencia de la actividad antimicrobiana, se realizaron mediciones de densidad óptica en intervalos de 5, 7 y 20 horas utilizando BSA en una concentración de 100 µg/mL como control adicional (Figura 31). En *P. aeruginosa*, la densidad óptica se mantuvo significativamente menor en las muestras tratadas con ADR1 en comparación con ADR2 y BSA durante todo el ensayo (Figura 31A). Para *K. pneumoniae*, ambas variantes presentaron una reducción de la densidad óptica constante hasta las 20 horas (Figura 31B). En las bacterias Gram positivas, *S. aureus* mostró una inhibición inicial que perdió significancia estadística para la variante ADR2 en el punto final del experimento (Figura 31C). Por último, en *S. pneumoniae*, ambas variantes redujeron el crecimiento en comparación con el control y BSA, aunque se observaron variaciones en los niveles de significancia para ADR2 a las 7 y 20 horas (Figura 31D). En todos los casos, el control de BSA presentó una cinética de crecimiento similar a la del control no tratado.

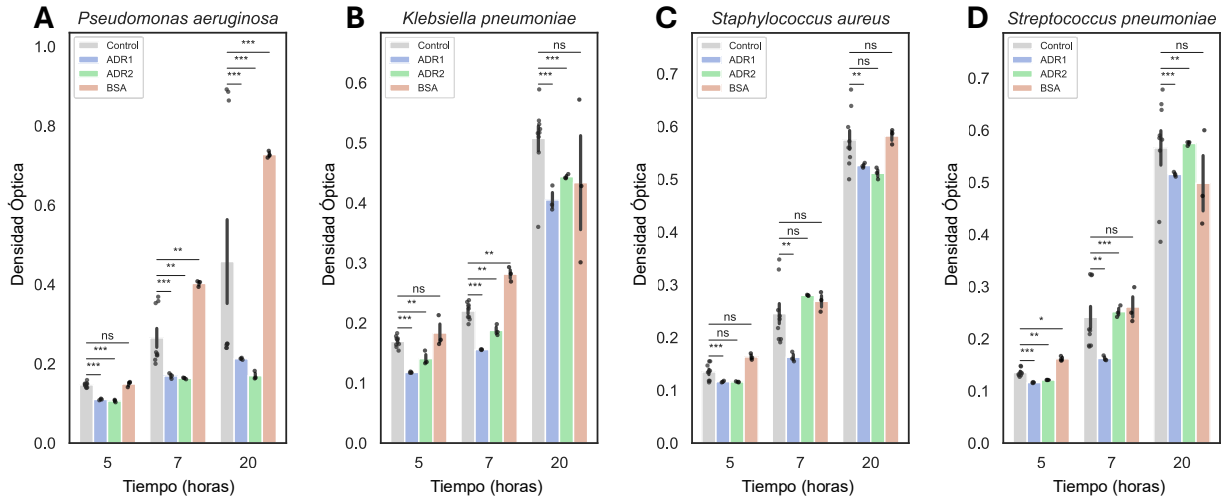


*Figura 31: Evaluación de la especificidad proteica en la inhibición del crecimiento bacteriano mediante el contraste de las variantes ADR1 y ADR2 frente a BSA. Comparativa de la cinética de crecimiento medida por densidad óptica a las 5, 7 y 20 horas en presencia de los AMPs ADR1 y ADR2 y albúmina de suero bovino (BSA) a 100 µg/mL, utilizada como control para descartar efectos inhibitorios no específicos. Inhibición del crecimiento de A) Pseudomonas aeruginosa, B) Klebsiella pneumoniae, C) Staphylococcus aureus y D) Streptococcus pneumoniae. (Significancia estadística: *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001; ns: no significativo).*

Finalmente, con el propósito de contrastar la eficacia de los péptidos identificados frente a tratamientos estándares, se evaluó la cinética de crecimiento bacteriano en presencia de los antibióticos Kanamicina y Ampicilina, ambos con una concentración de 100 µg/mL (Figura 32). Los resultados muestran que los antibióticos convencionales presentan una inhibición casi total de la densidad óptica en todas las cepas evaluadas desde las primeras 5 horas de incubación. Si bien los antibióticos convencionales presentan una inhibición casi total de la densidad óptica en todas las cepas evaluadas, las variantes ADR1 y ADR2 mantienen una reducción significativa del crecimiento bacteriano en comparación con el control no tratado (Figura 32A-D).

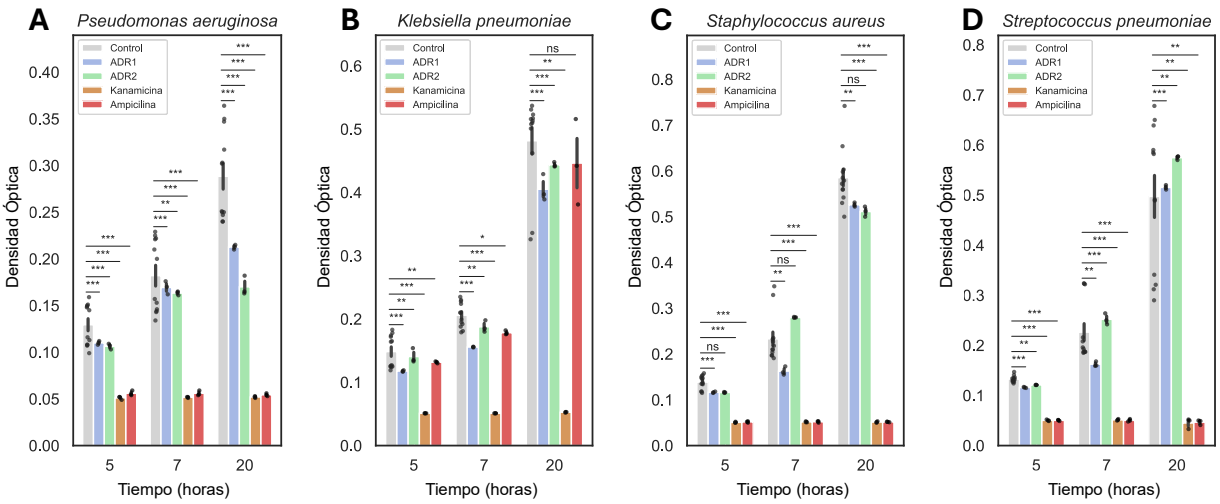


*Figura 32: Evaluación de la potencia antimicrobiana de las variantes ADR1 y ADR2 mediante el contraste con antibióticos de referencia clínica. Comparativa de la cinética de crecimiento medida por densidad óptica a las 5, 7 y 20 horas en presencia de los ADR1, ADR2, Kanamicina y Ampicilina. Inhibición del crecimiento de A) Pseudomonas aeruginosa, B) Klebsiella pneumoniae, C) Staphylococcus aureus y D) Streptococcus pneumoniae. (Significancia estadística: *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001; ns: no significativo).*

Discusión

Los resultados de esta tesis muestran que la microbiota intestinal infantil presenta alteraciones estructurales y funcionales estrechamente asociadas con el estado metabólico de los individuos. Primero con la información del gen ARNr 16S, observamos que las niñas y niños con obesidad y con SM presentan una composición microbiana distinta respecto al grupo normopeso, caracterizada por un aumento en Firmicutes y una disminución en Bacteroidetes. Esta disbiosis ha sido ampliamente descrita como un rasgo común en individuos con obesidad y SM, donde la modificación de la razón Firmicutes/Bacteroidetes se asocia con una mayor eficiencia en la extracción de energía y un incremento en la acumulación de grasa corporal [12], [13], [17]. En nuestras muestras, la tendencia ascendente de esta relación refleja un cambio funcional del ecosistema intestinal que parece acompañar la progresión hacia la disfunción metabólica.

La sobreabundancia de taxones como *Collinsella aerofaciens*, *Catenibacterium* y miembros de la familia Erysipelotrichaceae, junto con la disminución de *Faecalibacterium prausnitzii* y *Parabacteroides distasonis*, coincide con lo reportado en estudios que vinculan estas bacterias con inflamación intestinal, dislipidemia y reducción en la producción de butirato [100], [101], [102]. Estos resultados sugieren un reacomodo funcional de la microbiota en el que los taxones beneficiosos disminuyen y se favorece la expansión de grupos oportunistas con posibles efectos proinflamatorios [19], [103], [104]. La progresión observada desde el estado normopeso hasta OMS apoya la idea de una disbiosis gradual, donde la estructura microbiana se modifica de manera paralela a los cambios metabólicos del hospedero.

El análisis del Secrebioma complementó esta perspectiva al identificar el conjunto de proteínas potencialmente secretadas por la microbiota y sus posibles funciones en la comunicación microbiana y la interacción con el hospedero. Este análisis mostró que aproximadamente una cuarta parte del total de proteínas codificadas correspondía a proteínas secretadas, principalmente extracelulares y de membrana externa. Las proteínas del Secrebioma mostraron enriquecimiento en funciones metabólicas, de transporte y de interacción célula-célula, junto con diferencias significativas en las familias de enzimas activas sobre carbohidratos especialmente CBM, PL, SLH, cohesinas y dockerinas en comparación con las proteínas no secretadas. Estas categorías funcionales sugieren un papel importante del Secrebioma en la degradación de carbohidratos complejos y la organización de complejos multienzimáticos extracelulares, consistentes con observaciones previas en microbiomas intestinales [105], [106].

Adicionalmente, se identificaron 31 transcritos específicos del Secrebioma con abundancia diferencial entre los grupos O y OMS. La mayoría de los transcritos sobreexpresados en el grupo O presentaron homología con genes del genoma de *Faecalibacterium prausnitzii*, una especie considerada probiótica y asociada con un metabolismo intestinal saludable [97], mientras que los transcritos diferencialmente expresados en el grupo OMS se relacionaron con genes del pangenoma de *Bacteroides*, un género previamente vinculado con individuos con sobrepeso y con la presencia de CAZymes implicadas en la sobreextracción de energía [107], [108]. Además, dos

proteínas chaperonas, DnaK (HSP70) y DnaJ (HSP40), mostraron una correlación negativa con los niveles séricos de LDL, siendo más abundantes en el grupo O y menos en OMS, lo que sugiere un papel potencial en procesos de homeostasis proteica y metabolismo energético. Finalmente, el análisis comparativo con metatranscriptomas públicos confirmó la presencia de estas proteínas en distintas poblaciones humanas [32], lo que respalda su relevancia biológica y su conservación funcional dentro del microbioma intestinal.

Si bien el análisis del Secrebioma permitió obtener un primer vistazo funcional más allá de la información proporcionada por el perfil taxonómico del 16S, en esta primera aproximación no fue posible identificar AMPs dentro del conjunto de proteínas secretadas. Esto evidenció la necesidad de emplear estrategias bioinformáticas específicas para su detección, por lo que se utilizaron herramientas de predicción especializadas, las cuales permitieron identificar y caracterizar dichos péptidos. En total, se detectaron 51 AMPs expresados por la microbiota intestinal, la mayoría de origen bacteriano, destacando *Faecalibacterium prausnitzii*, *Bacteroides*, *Parabacteroides* y *Blautia wexlerae* como los principales contribuyentes. Aunque estos 51 candidatos representan una pequeña proporción del total de transcritos microbianos, constituyen una fracción funcionalmente específica cuya detección sugiere que la producción de AMPs es un componente constante en la dinámica ecológica intestinal. Además, la detección de AMPs codificados en elementos genéticos móviles, como fagos y plásmidos, amplía el marco conceptual sobre el papel del viroma y del mobiloma en la microbiota intestinal [109], [110], [111]. En trabajos previos de nuestro grupo demostramos que el viroma intestinal de niños con obesidad y SM presenta alteraciones en diversidad y riqueza [35], lo cual podría estar relacionado con la capacidad de estos fagos para portar y expresar genes antimicrobianos que contribuyen a la competencia y estabilidad microbiana. En este estudio confirmamos que dichos elementos móviles no sólo actúan como reservorios genéticos, sino que participan activamente en la expresión de genes funcionales dentro del ecosistema intestinal.

Entre los AMPs identificados, seis mostraron expresión diferencial entre los grupos O y OMS, cuatro de ellos sobreexpresados en OMS y dos en O. Este resultado indica que el repertorio antimicrobiano se modula conforme avanza la alteración metabólica. La diversidad de orígenes genómicos, bacteriano, viral y plasmídico, pone de manifiesto la complejidad de las interacciones dentro del microbioma y su capacidad para ajustar la producción de AMPs frente a condiciones fisiológicas cambiantes [38], [112]. En este sentido, proponemos que los AMPs reflejan no sólo la respuesta de la comunidad microbiana ante el estrés metabólico, sino también un mecanismo de reorganización ecológica que puede influir en la composición bacteriana observada en estados de disbiosis.

La detección de estos mismos AMPs en una cohorte externa de más de 300 metatranscriptomas humanos refuerza su importancia biológica y su posible papel conservado en el microbioma intestinal humano. La amplia distribución de estos péptidos sugiere que forman parte de un conjunto estable de moléculas bioactivas cuya expresión se ajusta de manera flexible a las condiciones del hospedero [24]. En conjunto, estos datos respaldan la idea de que los AMPs podrían servir como biomarcadores funcionales

de diferentes estados metabólicos, además de representar potenciales herramientas terapéuticas para modular la microbiota [50], [51].

Al integrar los datos de expresión génica con los perfiles taxonómicos, observamos correlaciones significativas entre la expresión de los AMPs y la abundancia de distintos taxones microbianos. Los patrones de correlación indicaron tanto relaciones negativas como positivas, reflejando posibles interacciones competitivas o cooperativas dentro del ecosistema intestinal. Las correlaciones negativas involucraron principalmente taxones asociados con un perfil metabólico saludable, como *Akkermansia muciniphila*, Christensenellaceae R-7 group, Oscillospiraceae y *Alistipes*, todos previamente descritos como indicadores de un intestino sano. Por el contrario, las correlaciones positivas se concentraron en géneros vinculados a estados inflamatorios o dismetabólicos, como *Adlercreutzia*, *Dialister*, *Sutterella*, *Veillonella* y *Colidextribacter*. Estas asociaciones sugieren que los AMPs podrían actuar como mediadores ecológicos que modulan la abundancia de ciertos grupos bacterianos en función de la condición metabólica del hospedero. En particular, observamos que el AMP 3020, de origen viral, presentó correlaciones con múltiples taxones asociados con obesidad, lo que refuerza su potencial relevancia ecológica y su papel en la competencia microbiana dentro del intestino.

La validación experimental del AMP 3020 proporcionó evidencia funcional de su actividad antimicrobiana. La síntesis de dos variantes, ADR1 y ADR2, permitió confirmar su capacidad inhibitoria frente a bacterias Gram positivas y Gram negativas de relevancia clínica. Ambas variantes mostraron una inhibición significativa del crecimiento bacteriano, aunque con diferencias específicas en el rango de susceptibilidad, lo que indica que pequeñas modificaciones en la secuencia pueden alterar la eficacia o el espectro de acción. Asimismo, no se observaron efectos citotóxicos en linfocitos T humanos expuestos a las mismas concentraciones, lo que sugiere una alta selectividad del péptido [113], [114]. Este resultado apoya la idea propuesta en el artículo de revisión incluido en esta tesis, donde discutimos que los AMPs derivados del microbioma humano representan una fuente prometedora de moléculas bioactivas con potencial terapéutico debido a su baja toxicidad y su actividad de amplio espectro [45].

La evaluación de las cinéticas de crecimiento permitió observar que la actividad antimicrobiana del AMP 3020 es un fenómeno dependiente no solo de la concentración, sino también del tiempo de exposición y de la composición de la pared celular bacteriana, lo cual es consistente con lo reportado para péptidos antimicrobianos en distintos sistemas bacterianos [38]. La inclusión de BSA como control de especificidad en distintos intervalos fue fundamental para demostrar que la inhibición observada es una propiedad intrínseca de la actividad biológica de las variantes ADR1 y ADR2. Este control permitió descartar que el decremento en la densidad óptica fuera causado por efectos físicos inespecíficos o por la alta carga proteica en el medio de cultivo, confirmando así la selectividad de los péptidos sintetizados [115]

Asimismo, el contraste con Kanamicina y Ampicilina permitió situar el desempeño de estos péptidos frente a fármacos de referencia clínica. Los resultados demuestran que, si bien los antibióticos convencionales ejercen una inhibición bactericida total desde etapas tempranas, las variantes ADR1 y ADR2 poseen una capacidad funcional para contener significativamente el crecimiento bacteriano de manera sostenida, aunque sin

alcanzar la potencia absoluta de los tratamientos estándares, lo cual es consistente con reportes que indican que los péptidos antimicrobianos pueden presentar una eficacia menor o más dependiente del contexto experimental en comparación con antibióticos convencionales [112], [116].

Considerando la capacidad antimicrobiana de este péptido, las correlaciones positivas observadas *in silico* podrían parecer contradictorias; sin embargo, dichas correlaciones sugieren una dinámica de exclusión competitiva [117]. Es probable que los taxones que aumentan su abundancia sean resistentes al AMP, prosperando en el nicho liberado tras la eliminación de competidores sensibles [44]. De esta manera, el péptido modularía el equilibrio ecológico favoreciendo a ciertos grupos mientras limita a otros, un mecanismo de interferencia competitiva previamente descrito como determinante para la colonización en el intestino de mamíferos [118]. Esta hipótesis plantea la necesidad de una validación experimental in vivo en modelos murinos, donde el análisis de 16S permitiría validar experimentalmente los cambios en la microbiota intestinal inducidos por dicho péptido.

Cabe mencionar que una de las limitantes de este trabajo es el número de muestras procesadas, ya que el análisis metatranscriptómico de los 8 individuos no nos permitió comparaciones con el grupo normopeso. Si bien esta n es reducida para establecer conclusiones generalizables, este análisis tuvo un carácter exploratorio que permitió identificar tendencias funcionales. Para fortalecer estos hallazgos, comparamos nuestros resultados con una cohorte externa de 372 metatranscriptomas públicos [32], confirmando que la presencia y expresión de estos AMPs es un fenómeno conservado en distintas poblaciones humanas y no un evento aislado de nuestra población.

En conjunto, nuestros resultados proponen un modelo en el que la disbiosis intestinal, la expresión diferencial de AMPs y la participación de elementos genéticos móviles están estrechamente interconectadas. Durante la progresión hacia la obesidad y el SM, el microbioma intestinal parece responder mediante la activación de un repertorio de genes antimicrobianos que alteran la composición microbiana, favoreciendo la expansión de taxones adaptados a condiciones inflamatorias o disbióticas. En este contexto, los fagos y plásmidos no actúan únicamente como vehículos de transferencia genética, sino como participantes activos en la regulación ecológica y la adaptación del microbioma [109], [110], [119]. Esta evidencia refuerza la noción de que el intestino humano debe entenderse como un sistema holobiótico donde las bacterias, los virus y el hospedero coevolucionan de forma dinámica, manteniendo un equilibrio funcional que puede alterarse ante presiones metabólicas o ambientales.

Conclusiones

El presente trabajo de tesis doctoral integró enfoques metagenómicos, metatranscriptómicos y virómicos para caracterizar las alteraciones estructurales y funcionales de la microbiota intestinal asociadas con la obesidad y el Síndrome Metabólico en la población infantil mexicana. Los resultados obtenidos demuestran que la disbiosis observada en estas condiciones no se limita a cambios en la composición bacteriana, sino que implica una reorganización funcional profunda del ecosistema intestinal, reflejada en variaciones en la expresión génica y en la producción de AMPs de origen bacteriano y viral.

A nivel taxonómico, se observó un aumento en Firmicutes y una disminución en Bacteroidetes, junto con la expansión de taxones oportunistas y la reducción de bacterias productoras de butirato, como *Faecalibacterium prausnitzii* y *Parabacteroides distasonis*. Este patrón refleja un proceso de disbiosis progresiva vinculado con estados inflamatorios y alteraciones metabólicas del hospedero.

El análisis metatranscriptómico permitió identificar 51 AMPs expresados por la microbiota intestinal, la gran mayoría (50 péptidos) derivados de smORFs no descritos previamente. Una proporción significativa de estos péptidos se originó en elementos genéticos móviles, incluyendo fagos y plásmidos, lo que demuestra que el mobiloma intestinal contribuye activamente a la regulación ecológica mediante la producción de moléculas bioactivas. Estos hallazgos amplían la comprensión del papel del viroma intestinal como fuente de diversidad funcional y como un componente dinámico de la competencia microbiana.

El análisis de expresión diferencial reveló seis AMPs modulados entre los grupos O y OMS, cuyos patrones de correlación con distintos taxones bacterianos sugieren una interacción funcional entre la expresión antimicrobiana y la estructura microbiana intestinal. En particular, el AMP 3020, de origen viral, mostró un comportamiento diferencial y una correlación positiva con taxones asociados con disbiosis.

La validación experimental del AMP 3020 confirmó su actividad antimicrobiana de amplio espectro frente a bacterias Gram positivas y negativas, sin efectos citotóxicos detectables sobre linfocitos T humanos. Este resultado aporta evidencia funcional de que los virus intestinales no solo actúan como reservorios genéticos, sino también como productores activos de péptidos bioactivos con potencial terapéutico. En conjunto, los resultados de esta tesis proponen un modelo en el cual la microbiota intestinal, a través de la expresión de AMPs, incluidos los codificados en elementos móviles, puede modular activamente la composición y funcionalidad del ecosistema intestinal (Figura 33). Estos hallazgos aportan una base funcional y ecológica para comprender los mecanismos que vinculan la disbiosis con la O y el SM, y plantean el uso de los AMPs como biomarcadores o herramientas biotecnológicas para el estudio y manejo del microbioma humano.

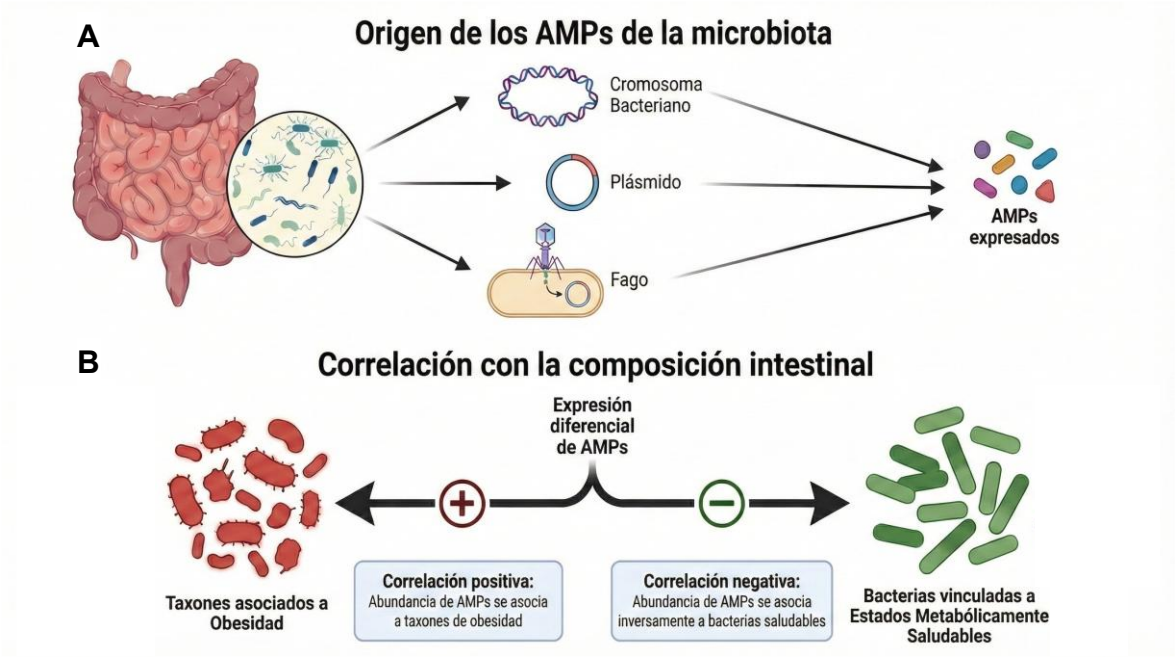


Figura 33: Resumen de las conclusiones. A) Los AMPs expresados por la microbiota intestinal provienen de distintos orígenes genómicos, incluyendo cromosomas bacterianos, plásmidos y fagos. B) Su expresión diferencial se relaciona con cambios en la composición bacteriana intestinal, mostrando correlaciones positivas con taxones asociados a obesidad y negativas con bacterias vinculadas a estados metabólicamente saludables.

Perspectivas

Los resultados de esta tesis abren nuevas líneas de investigación orientadas a profundizar en la función y aplicación biomédica de los AMPs del microbioma en el contexto de enfermedades importantes para la población y asociadas a la regulación de la microbiota intestinal. Será importante ampliar los análisis metatranscriptómicos y metagenómicos en cohortes más grandes y diversas, para confirmar los patrones de expresión diferencial observados y su relación con distintos grados de O y SM. La incorporación de herramientas de Inteligencia Artificial permitirá optimizar la detección de smORFs y la predicción de la función antimicrobiana, fortaleciendo la interpretación funcional del microbioma.

Sin duda alguna, la integración de datos multi-ómicos, como metatranscriptómica, metaproteómica y metabolómica, ofrecerá una visión más completa de las redes de regulación microbiana y su influencia en la fisiología del hospedero. En este contexto, se propone la evaluación in vivo del AMP 3020 en modelos murinos (Figura 34), para analizar su actividad antimicrobiana, efecto modulador sobre la microbiota intestinal y seguridad sistémica. Este enfoque permitirá determinar si el péptido conserva su eficacia en un sistema biológico complejo y si puede contribuir a restaurar la eubiosis intestinal alterada por O o SM.

Desde una perspectiva aplicada, los AMPs del microbioma identificados podrían emplearse como biomarcadores moleculares para el diagnóstico temprano de disbiosis y el monitoreo de tratamientos nutricionales o farmacológicos. A largo plazo, la modulación dirigida de estos péptidos, mediante fagoterapia, edición génica o administración controlada, podría representar una alternativa a los antibióticos tradicionales y favorecer estrategias personalizadas para la restauración del equilibrio metabólico.

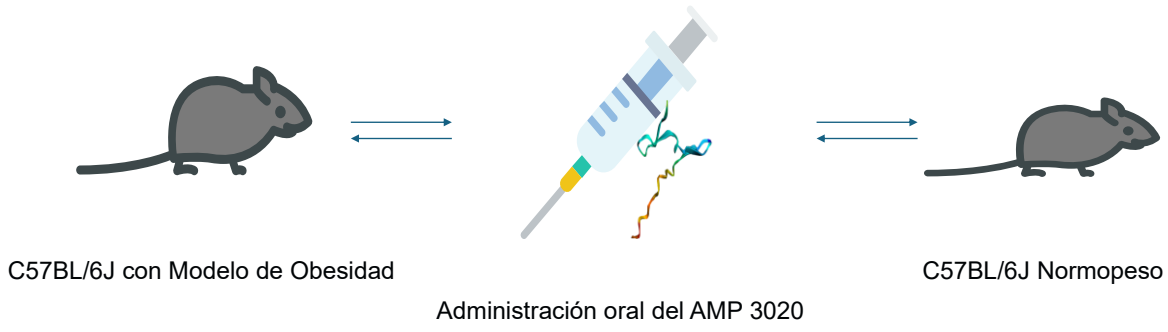


Figura 34: Evaluación in vivo del AMP 3020. Representación esquemática del modelo propuesto en ratones C57BL/6J con obesidad inducida, que recibirán administración oral del AMP 3020 para analizar su efecto sobre la microbiota intestinal, la inflamación y los parámetros metabólicos en comparación con animales normopeso.

Referencias

[1] P. W. Franks, R. L. Hanson, W. C. Knowler, M. L. Sievers, P. H. Bennett, and H. C. Looker, "Childhood Obesity, Other Cardiovascular Risk Factors, and Premature Death," *New England Journal of Medicine*, vol. 362, no. 6, pp. 485–493, Feb. 2010, doi: 10.1056/NEJMoa0904130.

[2] OECD, *Health at a Glance 2023: OECD Indicators*. in Health at a Glance. OECD Publishing, 2023. doi: 10.1787/7a7afb35-en.

[3] T. Shamah-Levy *et al.*, "Prevalencias de sobrepeso y obesidad en población escolar y adolescente de México. Ensanut Continua 2020-2022," *Salud Publica Mex*, vol. 65, pp. s218–s224, Jun. 2023, doi: 10.21149/14762.

[4] M. L. Evia-Viscarra, E. R. Rodea-Montero, E. Apolinar-Jiménez, and S. Quintana-Vargas, "Metabolic syndrome and its components among obese (BMI ≥95th) Mexican adolescents," *Endocrine Connections*, vol. 2, no. 4, pp. 208–215, Dec. 2013, doi: 10.1530/EC-13-0057.

[5] S. D. de Ferranti, K. Gauvreau, D. S. Ludwig, E. J. Neufeld, J. W. Newburger, and N. Rifai, "Prevalence of the Metabolic Syndrome in American Adolescents," *Circulation*, vol. 110, no. 16, pp. 2494–2497, Oct. 2004, doi: 10.1161/01.CIR.0000145117.40114.C7.

[6] D. Al-Hamad and V. Raman, "Metabolic syndrome in children and adolescents," *Translational Pediatrics*, vol. 6, no. 4, Art. no. 4, Oct. 2017, doi: 10.21037/tp.2017.10.02.

[7] K. G. M. Alberti, P. Zimmet, and J. Shaw, "The metabolic syndrome—a new worldwide definition," *The Lancet*, vol. 366, no. 9491, pp. 1059–1062, Sep. 2005, doi: 10.1016/S0140-6736(05)67402-8.

[8] F. M. Biro and M. Wien, "Childhood obesity and adult morbidities," *The American Journal of Clinical Nutrition*, vol. 91, no. 5, pp. 1499S-1505S, May 2010, doi: 10.3945/ajcn.2010.28701B.

[9] R. S. Ahima and M. A. Lazar, "The Health Risk of Obesity—Better Metrics Imperative," *Science*, vol. 341, no. 6148, pp. 856–858, Aug. 2013, doi: 10.1126/science.1241244.

[10] T. Lobstein *et al.*, "Child and adolescent obesity: part of a bigger picture," *The Lancet*, vol. 385, no. 9986, pp. 2510–2520, Jun. 2015, doi: 10.1016/S0140-6736(14)61746-3.

[11]P. J. Turnbaugh, R. E. Ley, M. A. Mahowald, V. Magrini, E. R. Mardis, and J. I. Gordon, "An obesity-associated gut microbiome with increased capacity for energy harvest," *Nature*, vol. 444, no. 7122, pp. 1027–1031, Dec. 2006, doi: 10.1038/nature05414.

[12] P. J. Turnbaugh *et al.*, "A core gut microbiome in obese and lean twins," *Nature*, vol. 457, no. 7228, pp. 480–484, Jan. 2009, doi: 10.1038/nature07540.

[13] R. E. Ley, P. J. Turnbaugh, S. Klein, and J. I. Gordon, "Human gut microbes associated with obesity," *Nature*, vol. 444, no. 7122, pp. 1022–1023, Dec. 2006, doi: 10.1038/4441022a.

[14] M. Duan, Y. Wang, Q. Zhang, R. Zou, M. Guo, and H. Zheng, "Characteristics of gut microbiota in people with obesity," *PLOS ONE*, vol. 16, no. 8, p. e0255446, Aug. 2021, doi: 10.1371/journal.pone.0255446.

[15] F. Sommer and F. Bäckhed, "The gut microbiota — masters of host development and physiology," *Nat Rev Microbiol*, vol. 11, no. 4, pp. 227–238, Apr. 2013, doi: 10.1038/nrmicro2974.

[16] E. Thursby and N. Juge, "Introduction to the human gut microbiota," *Biochem J*, vol. 474, no. 11, pp. 1823–1836, May 2017, doi: 10.1042/BCJ20160510.

[17] V. K. Ridaura *et al.*, "Gut Microbiota from Twins Discordant for Obesity Modulate Metabolism in Mice," *Science*, vol. 341, no. 6150, p. 1241214, Sep. 2013, doi: 10.1126/science.1241214.

[18] P. Gérard, "Gut microbiota and obesity," *Cell. Mol. Life Sci.*, vol. 73, no. 1, pp. 147–162, Jan. 2016, doi: 10.1007/s00018-015-2061-5.

[19] P. D. Cani *et al.*, "Changes in Gut Microbiota Control Metabolic Endotoxemia-Induced Inflammation in High-Fat Diet–Induced Obesity and Diabetes in Mice," *Diabetes*, vol. 57, no. 6, pp. 1470–1481, Jun. 2008, doi: 10.2337/db07-1403.

[20] S. O. Fetissov, "Role of the gut microbiota in host appetite control: bacterial growth to animal feeding behaviour," *Nat Rev Endocrinol*, vol. 13, no. 1, pp. 11–25, Jan. 2017, doi: 10.1038/nrendo.2016.150.

[21] P. D. Cani and C. Knauf, "How gut microbes talk to organs: The role of endocrine and nervous routes," *Molecular Metabolism*, vol. 5, no. 9, pp. 743–752, Sep. 2016, doi: 10.1016/j.molmet.2016.05.011.

[22] E. A. Franzosa *et al.*, "Gut microbiome structure and metabolic activity in inflammatory bowel disease," *Nat Microbiol*, vol. 4, no. 2, pp. 293–305, Feb. 2019, doi: 10.1038/s41564-018-0306-4.

[23] J. Lloyd-Price *et al.*, "Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases," *Nature*, vol. 569, no. 7758, pp. 655–662, May 2019, doi: 10.1038/s41586-019-1237-9.

[24] A. Heintz-Buschart and P. Wilmes, "Human Gut Microbiome: Function Matters," *Trends in Microbiology*, vol. 26, no. 7, pp. 563–574, Jul. 2018, doi: 10.1016/j.tim.2017.11.002.

[25] J. Qin *et al.*, "A metagenome-wide association study of gut microbiota in type 2 diabetes," *Nature*, vol. 490, no. 7418, pp. 55–60, Oct. 2012, doi: 10.1038/nature11450.

[26] F. Bäckhed *et al.*, “The gut microbiota as an environmental factor that regulates fat storage,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 44, pp. 15718–15723, Nov. 2004, doi: 10.1073/pnas.0407076101.

[27] M. Shakya, C.-C. Lo, and P. S. G. Chain, “Advances and Challenges in Metatranscriptomic Analysis,” *Front. Genet.*, vol. 10, Sep. 2019, doi: 10.3389/fgene.2019.00904.

[28] M. Valles-Colomer *et al.*, “The neuroactive potential of the human gut microbiota in quality of life and depression,” *Nat Microbiol*, vol. 4, no. 4, pp. 623–632, Apr. 2019, doi: 10.1038/s41564-018-0337-x.

[29] E. A. Franzosa *et al.*, “Relating the metatranscriptome and metagenome of the human gut,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 22, pp. E2329–E2338, Jun. 2014, doi: 10.1073/pnas.1319284111.

[30] S. Bikel *et al.*, “Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome,” *Computational and Structural Biotechnology Journal*, vol. 13, pp. 390–401, Jan. 2015, doi: 10.1016/j.csbj.2015.06.001.

[31] A. A. Shishkin *et al.*, “Simultaneous generation of many RNA-seq libraries in a single reaction,” *Nat Methods*, vol. 12, no. 4, pp. 323–325, Apr. 2015, doi: 10.1038/nmeth.3313.

[32] G. S. Abu-Ali *et al.*, “Metatranscriptome of human faecal microbial communities in a cohort of adult men,” *Nat Microbiol*, vol. 3, no. 3, pp. 356–366, Jan. 2018, doi: 10.1038/s41564-017-0084-4.

[33] R. Knight *et al.*, “Best practices for analysing microbiomes,” *Nat Rev Microbiol*, vol. 16, no. 7, pp. 410–422, Jul. 2018, doi: 10.1038/s41579-018-0029-9.

[34] S. Bikel, L. Gallardo-Becerra, F. Cornejo-Granados, and A. Ochoa-Leyva, “Protocol for the isolation, sequencing, and analysis of the gut phageome from human fecal samples,” *STAR Protocols*, vol. 3, no. 1, p. 101170, Mar. 2022, doi: 10.1016/j.xpro.2022.101170.

[35] S. Bikel *et al.*, “Gut dsDNA virome shows diversity and richness alterations associated with childhood obesity and metabolic syndrome,” *iScience*, vol. 24, no. 8, p. 102900, Aug. 2021, doi: 10.1016/j.isci.2021.102900.

[36] W. Figueroa, D. Cazares, and A. Cazares, “Phage-plasmids: missed links between mobile genetic elements,” *Trends in Microbiology*, vol. 32, no. 7, pp. 622–623, Jul. 2024, doi: 10.1016/j.tim.2024.04.014.

[37] L. Gallardo-Becerra *et al.*, “Metatranscriptomic analysis to define the Secrebiome, and 16S rRNA profiling of the gut microbiome in obesity and metabolic syndrome of Mexican children,” *Microb Cell Fact*, vol. 19, no. 1, p. 61, Dec. 2020, doi: 10.1186/s12934-020-01319-y.

[38] R. E. W. Hancock and H.-G. Sahl, “Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies,” *Nat Biotechnol*, vol. 24, no. 12, Art. no. 12, Dec. 2006, doi: 10.1038/nbt1267.

[39] L. Zhang and R. L. Gallo, “Antimicrobial peptides,” *Current Biology*, vol. 26, no. 1, pp. R14–R19, Jan. 2016, doi: 10.1016/j.cub.2015.11.017.

[40] A. Izadpanah and R. L. Gallo, “Antimicrobial peptides,” *Journal of the American Academy of Dermatology*, vol. 52, no. 3, pp. 381–390, Mar. 2005, doi: 10.1016/j.jaad.2004.08.026.

[41] E. Garcia-Gutierrez, M. J. Mayer, P. D. Cotter, and A. Narbad, “Gut microbiota as a source of novel antimicrobials,” *Gut Microbes*, vol. 10, no. 1, pp. 1–21, Jan. 2019, doi: 10.1080/19490976.2018.1455790.

[42] N. H. Salzman *et al.*, “Enteric defensins are essential regulators of intestinal microbial ecology,” *Nat Immunol*, vol. 11, no. 1, pp. 76–82, Jan. 2010, doi: 10.1038/ni.1825.

[43] M. J. Ostaff, E. F. Stange, and J. Wehkamp, “Antimicrobial peptides and gut microbiota in homeostasis and pathology,” *EMBO Mol Med*, vol. 5, no. 10, pp. 1465–1483, Oct. 2013, doi: 10.1002/emmm.201201773.

[44] S. Kommineni *et al.*, “Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract,” *Nature*, vol. 526, no. 7575, pp. 719–722, Oct. 2015, doi: 10.1038/nature15524.

[45] L. Gallardo-Becerra, M. Cervantes-Echeverría, F. Cornejo-Granados, L. E. Vazquez-Morado, and A. Ochoa-Leyva, “Perspectives in Searching Antimicrobial Peptides (AMPs) Produced by the Microbiota,” *Microb Ecol*, vol. 87, no. 1, p. 8, Dec. 2023, doi: 10.1007/s00248-023-02313-8.

[46] M. D. T. Torres, M. C. R. Melo, L. Flowers, O. Crescenzi, E. Notomista, and C. De La Fuente-Nunez, “Mining for encrypted peptide antibiotics in the human proteome,” *Nat Biomed Eng*, vol. 6, no. 1, pp. 67–75, Nov. 2021, doi: 10.1038/s41551-021-00801-1.

[47] S. Mukherjee and L. V. Hooper, “Antimicrobial Defense of the Intestine,” *Immunity*, vol. 42, no. 1, pp. 28–39, Jan. 2015, doi: 10.1016/j.immuni.2014.12.028.

[48] J. M. Ageitos, A. Sánchez-Pérez, P. Calo-Mata, and T. G. Villa, “Antimicrobial peptides (AMPs): Ancient compounds that represent novel weapons in the fight against bacteria,” *Biochemical Pharmacology*, vol. 133, pp. 117–138, Jun. 2017, doi: 10.1016/j.bcp.2016.09.018.

[49] P. D. Cotter, C. Hill, and R. P. Ross, “Bacteriocins: developing innate immunity for food,” *Nat Rev Microbiol*, vol. 3, no. 10, Art. no. 10, Oct. 2005, doi: 10.1038/nrmicro1273.

[50] C. D. Santos-Júnior *et al.*, “Discovery of antimicrobial peptides in the global microbiome with machine learning,” *Cell*, vol. 0, no. 0, Jun. 2024, doi: 10.1016/j.cell.2024.05.013.

[51] M. D. T. Torres *et al.*, “Mining human microbiomes reveals an untapped source of peptide antibiotics,” *Cell*, vol. 0, no. 0, Aug. 2024, doi: 10.1016/j.cell.2024.07.027.

[52] D. Veltri, U. Kamath, and A. Shehu, “Deep learning improves antimicrobial peptide recognition,” *Bioinformatics*, vol. 34, no. 16, pp. 2740–2747, Aug. 2018, doi: 10.1093/bioinformatics/bty179.

[53] H. Wu *et al.*, “ESKtides: a comprehensive database and mining method for ESKAPE phage-derived antimicrobial peptides,” *Database (Oxford)*, vol. 2024, p. baae022, Feb. 2024, doi: 10.1093/database/baae022.

[54] H. Sberro *et al.*, “Large-Scale Analyses of Human Microbiomes Reveal Thousands of Small, Novel Genes,” *Cell*, vol. 178, no. 5, pp. 1245-1259.e14, Aug. 2019, doi: 10.1016/j.cell.2019.07.016.

[55] G. Agüero-Chapin *et al.*, “Emerging Computational Approaches for Antimicrobial Peptide Discovery,” *Antibiotics*, vol. 11, no. 7, p. 936, Jul. 2022, doi: 10.3390/antibiotics11070936.

[56] K. Nakazono *et al.*, “Complete sequences of epidermin and nukacin encoding plasmids from oral-derived Staphylococcus epidermidis and their antibacterial activity,” *PLOS ONE*, vol. 17, no. 1, p. e0258283, Jan. 2022, doi: 10.1371/journal.pone.0258283.

[57] J. Du *et al.*, “Enhancing bacteriophage therapeutics through in situ production and release of heterologous antimicrobial effectors,” *Nat Commun*, vol. 14, no. 1, p. 4337, Jul. 2023, doi: 10.1038/s41467-023-39612-0.

[58] J. L. Tyagi, P. Gupta, M. M. Ghate, D. Kumar, and K. M. Poluri, “Assessing the synergistic potential of bacteriophage endolysins and antimicrobial peptides for eradicating bacterial biofilms,” *Arch Microbiol*, vol. 206, no. 6, p. 272, May 2024, doi: 10.1007/s00203-024-04003-6.

[59] R. F. Inglis, B. Bayramoglu, O. Gillor, and M. Ackermann, “The role of bacteriocins as selfish genetic elements,” *Biol Lett*, vol. 9, no. 3, p. 20121173, Jun. 2013, doi: 10.1098/rsbl.2012.1173.

[60] S. Arbulu and M. Kjos, “Revisiting the Multifaceted Roles of Bacteriocins,” *Microb Ecol*, vol. 87, no. 1, p. 41, Feb. 2024, doi: 10.1007/s00248-024-02357-4.

[61] S. Federici, S. P. Nobs, and E. Elinav, “Phages and their potential to modulate the microbiome and immunity,” *Cell Mol Immunol*, vol. 18, no. 4, pp. 889–904, Apr. 2021, doi: 10.1038/s41423-020-00532-4.

[62] C. D. Santos-Júnior, S. Pan, X.-M. Zhao, and L. P. Coelho, “Macrel: antimicrobial peptide screening in genomes and metagenomes,” *PeerJ*, vol. 8, p. e10555, Dec. 2020, doi: 10.7717/peerj.10555.

[63] P. Bhadra, J. Yan, J. Li, S. Fong, and S. W. I. Siu, “AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest,” *Sci Rep*, vol. 8, no. 1, Art. no. 1, Jan. 2018, doi: 10.1038/s41598-018-19752-w.

[64] Y. Xu, N. Wang, H.-Y. Tan, S. Li, C. Zhang, and Y. Feng, “Function of Akkermansia muciniphila in Obesity: Interactions With Lipid Metabolism, Immune Response and Gut Systems,” *Front. Microbiol.*, vol. 11, Feb. 2020, doi: 10.3389/fmicb.2020.00219.

[65] X. Lin and K. Xu, “Causal Link Between Gut Microbiota and Arterial Occlusive Diseases: Insights From Mendelian Randomization and Bioinformatics Analysis,” *Catheterization and Cardiovascular Interventions*, vol. n/a, no. n/a, doi: 10.1002/ccd.70047.

[66] X. Lai *et al.*, “Eubacterium siraeum suppresses fat deposition via decreasing the tyrosine-mediated PI3K/AKT signaling pathway in high-fat diet-induced obesity,” *Microbiome*, vol. 12, no. 1, p. 223, Oct. 2024, doi: 10.1186/s40168-024-01944-4.

[67] T. S. Dong *et al.*, “A Distinct Brain-Gut-Microbiome Profile Exists for Females with Obesity and Food Addiction,” *Obesity*, vol. 28, no. 8, pp. 1477–1486, 2020, doi: 10.1002/oby.22870.

[68] L. B. Thingholm *et al.*, “Obese Individuals with and without Type 2 Diabetes Show Different Gut Microbial Functional Capacity and Composition,” *Cell Host & Microbe*, vol. 26, no. 2, pp. 252-264.e10, Aug. 2019, doi: 10.1016/j.chom.2019.07.004.

[69] W. Mazier *et al.*, “A New Strain of Christensenella minuta as a Potential Biotherapy for Obesity and Associated Metabolic Diseases,” *Cells*, vol. 10, no. 4, p. 823, Apr. 2021, doi: 10.3390/cells10040823.

[70] M. A. Ahmad, M. Karavetian, C. A. Moubareck, G. Wazz, T. Mahdy, and K. Venema, “Association of the gut microbiota with clinical variables in obese and lean Emirati subjects,” *Front. Microbiol.*, vol. 14, Aug. 2023, doi: 10.3389/fmicb.2023.1182460.

[71] T. Tavella *et al.*, “Elevated gut microbiome abundance of Christensenellaceae, Porphyromonadaceae and Rikenellaceae is associated with reduced visceral adipose tissue and healthier metabolic profile in Italian elderly,” *Gut Microbes*, vol. 13, no. 1, p. 1880221, Jan. 2021, doi: 10.1080/19490976.2021.1880221.

[72] L. Wu, S.-H. Park, and H. Kim, “Direct and Indirect Evidence of Effects of Bacteroides spp. on Obesity and Inflammation,” *Int J Mol Sci*, vol. 25, no. 1, p. 438, Dec. 2023, doi: 10.3390/ijms25010438.

[73] S. W. Ryu *et al.*, “Anti-obesity activity of human gut microbiota Bacteroides stercoris KGMB02265,” *Arch Microbiol*, vol. 206, no. 1, p. 19, Dec. 2023, doi: 10.1007/s00203-023-03750-2.

[74] C. E. Orsso *et al.*, “Composition and Functions of the Gut Microbiome in Pediatric Obesity: Relationships with Markers of Insulin Resistance,” *Microorganisms*, vol. 9, no. 7, p. 1490, Jul. 2021, doi: 10.3390/microorganisms9071490.

[75] Y. C. Orbe-Orihuela *et al.*, “Association of Gut Microbiota with Dietary-dependent Childhood Obesity,” *Archives of Medical Research*, vol. 53, no. 4, pp. 407–415, Jun. 2022, doi: 10.1016/j.arcmed.2022.03.007.

[76] M. Alcazar *et al.*, “Gut microbiota is associated with metabolic health in children with obesity,” *Clin Nutr*, vol. 41, no. 8, pp. 1680–1688, Aug. 2022, doi: 10.1016/j.clnu.2022.06.007.

[77] E. Rampanelli *et al.*, “Gut bacterium Intestinimonas butyriciproducens improves host metabolic health: evidence from cohort and animal intervention studies,” *Microbiome*, vol. 13, no. 1, p. 15, Jan. 2025, doi: 10.1186/s40168-024-02002-9.

[78] A. Woting, N. Pfeiffer, G. Loh, S. Klaus, and M. Blaut, “Clostridium ramosum Promotes High-Fat Diet-Induced Obesity in Gnotobiotic Mouse Models,” *mBio*, vol. 5, no. 5, pp. e01530-14, Sep. 2014, doi: 10.1128/mBio.01530-14.

[79] J. Liao *et al.*, “Clostridium butyricum Strain CCFM1299 Reduces Obesity via Increasing Energy Expenditure and Modulating Host Bile Acid Metabolism,” *Nutrients*, vol. 15, no. 20, p. 4339, Jan. 2023, doi: 10.3390/nu15204339.

[80] Q. Zhou *et al.*, “Gut bacteria Akkermansia is associated with reduced risk of obesity: evidence from the American Gut Project,” *Nutr Metab (Lond)*, vol. 17, p. 90, 2020, doi: 10.1186/s12986-020-00516-1.

[81] J. Sun *et al.*, “The visceral adipose tissue bacterial microbiota provides a signature of obesity based on inferred metagenomic functions,” *Int J Obes*, vol. 47, no. 10, pp. 1008–1022, Oct. 2023, doi: 10.1038/s41366-023-01341-1.

[82] Z. Yu, X.-F. Yu, G. Kerem, and P.-G. Ren, “Perturbation on gut microbiota impedes the onset of obesity in high fat diet-induced mice,” *Front. Endocrinol.*, vol. 13, Aug. 2022, doi: 10.3389/fendo.2022.795371.

[83] X. Zhu *et al.*, “A High-Fat Diet Increases the Characteristics of Gut Microbial Composition and the Intestinal Damage Associated with Non-Alcoholic Fatty Liver Disease,” *International Journal of Molecular Sciences*, vol. 24, no. 23, p. 16733, Jan. 2023, doi: 10.3390/ijms242316733.

[84] E. Novikova, N. Belkova, A. Pogodina, A. Romanitsa, and L. Rychkova, “ODP377 Adolescents With Obesity Breastfed Until Four Months Age Have High Abundance Of Ruminococcaceae Bacteria In Gut Microbiota,” *J. Endocr. Soc.*, vol. 6, no. Supplement_1, p. A599, Dec. 2022, doi: 10.1210/jendso/bvac150.1242.

[85] M. Dekker Nitert, A. Mousa, H. L. Barrett, N. Naderpoor, and B. de Courten, “Altered Gut Microbiota Composition Is Associated With Back Pain in Overweight and Obese Individuals,” *Front Endocrinol (Lausanne)*, vol. 11, p. 605, Sep. 2020, doi: 10.3389/fendo.2020.00605.

[86] A. McLeod *et al.*, “Comparing the gut microbiome of obese, African American, older adults with and without mild cognitive impairment,” *PLoS One*, vol. 18, no. 2, p. e0280211, 2023, doi: 10.1371/journal.pone.0280211.

[87] M. Lu *et al.*, “Causal relationship between gut microbiota and childhood obesity: A Mendelian randomization study and case–control study,” *Clinical Nutrition ESPEN*, vol. 63, pp. 197–206, Oct. 2024, doi: 10.1016/j.clnesp.2024.05.012.

[88] S. M. Murga-Garrido *et al.*, “Virulence Factors of the Gut Microbiome Are Associated with BMI and Metabolic Blood Parameters in Children with Obesity,” *Microbiol Spectr*, vol. 11, no. 2, p. e0338222, Feb. 2023, doi: 10.1128/spectrum.03382-22.

[89] X. Song *et al.*, “Inulin Can Alleviate Metabolism Disorders in ob/ob Mice by Partially Restoring Leptin-related Pathways Mediated by Gut Microbiota,” *Genomics Proteomics Bioinformatics*, vol. 17, no. 1, pp. 64–75, Feb. 2019, doi: 10.1016/j.gpb.2019.03.001.

[90] T. Sen *et al.*, “Diet-driven microbiota dysbiosis is associated with vagal remodeling and obesity,” *Physiol Behav*, vol. 173, pp. 305–317, May 2017, doi: 10.1016/j.physbeh.2017.02.027.

[91] Y. Yun *et al.*, “Comparative analysis of gut microbiota associated with body mass index in a large Korean cohort,” *BMC Microbiol*, vol. 17, no. 1, p. 151, Jul. 2017, doi: 10.1186/s12866-017-1052-0.

[92] C. K. Chakraborti, “New-found link between microbiota and obesity,” *World J Gastrointest Pathophysiol*, vol. 6, no. 4, pp. 110–119, Nov. 2015, doi: 10.4291/wjgp.v6.i4.110.

[93] E. S. Chambers *et al.*, “Dietary supplementation with inulin-propionate ester or inulin improves insulin sensitivity in adults with overweight and obesity with distinct effects on the gut microbiota, plasma metabolome and systemic inflammatory responses: a randomised cross-over trial,” *Gut*, vol. 68, no. 8, pp. 1430–1438, Aug. 2019, doi: 10.1136/gutjnl-2019-318424.

[94] H. Lee *et al.*, “A Novel Bacterium, Butyricimonas virosa, Preventing HFD-Induced Diabetes and Metabolic Disorders in Mice via GLP-1 Receptor,” *Front Microbiol*, vol. 13, p. 858192, May 2022, doi: 10.3389/fmicb.2022.858192.

[95] J. Hu *et al.*, “Gut Microbiota Signature of Obese Adults Across Different Classifications,” *Diabetes Metab Syndr Obes*, vol. 15, pp. 3933–3947, 2022, doi: 10.2147/DMSO.S387523.

[96] C. Sabater *et al.*, “Enzymatic synthesis of β-galactosylated xylitol derivatives modulates gut microbiota and improves obesity-related metabolic parameters in mice,” *Food Funct.*, vol. 16, no. 13, pp. 5493–5510, Jul. 2025, doi: 10.1039/D5FO00978B.

[97] T. U. Maioli *et al.*, “Possible Benefits of Faecalibacterium prausnitzii for Obesity-Associated Gut Disorders,” *Front. Pharmacol.*, vol. 12, Dec. 2021, doi: 10.3389/fphar.2021.740636.

[98] K. Wang *et al.*, “Parabacteroides distasonis Alleviates Obesity and Metabolic Dysfunctions via Production of Succinate and Secondary Bile Acids,” *Cell Rep*, vol. 26, no. 1, pp. 222-235.e5, Jan. 2019, doi: 10.1016/j.celrep.2018.12.028.

[99] W. J. Choi *et al.*, “Lactobacillus plantarum LMT1-48 exerts anti-obesity effect in high-fat diet-induced obese mice by regulating expression of lipogenic genes,” *Sci Rep*, vol. 10, no. 1, p. 869, Jan. 2020, doi: 10.1038/s41598-020-57615-5.

[100] A. Purohit *et al.*, “Collinsella aerofaciens linked with increased ethanol production and liver inflammation contribute to the pathophysiology of NAFLD,” *iScience*, vol. 27, no. 2, Feb. 2024, doi: 10.1016/j.isci.2023.108764.

[101] P. Louis and H. J. Flint, “Formation of propionate and butyrate by the human colonic microbiota,” *Environmental Microbiology*, vol. 19, no. 1, pp. 29–41, 2017, doi: 10.1111/1462-2920.13589.

[102] A. Schwiertz *et al.*, “Microbiota and SCFA in Lean and Overweight Healthy Subjects,” *Obesity*, vol. 18, no. 1, pp. 190–195, 2010, doi: 10.1038/oby.2009.167.

[103] E. Le Chatelier *et al.*, “Richness of human gut microbiome correlates with metabolic markers,” *Nature*, vol. 500, no. 7464, pp. 541–546, Aug. 2013, doi: 10.1038/nature12506.

[104] P. D. Cani, A. Everard, and T. Duparc, “Gut microbiota, enteroendocrine functions and metabolism,” *Current Opinion in Pharmacology*, vol. 13, no. 6, pp. 935–940, Dec. 2013, doi: 10.1016/j.coph.2013.09.008.

[105] F. H. Karlsson *et al.*, “Gut metagenome in European women with normal, impaired and diabetic glucose control,” *Nature*, vol. 498, no. 7452, pp. 99–103, Jun. 2013, doi: 10.1038/nature12198.

[106] J. Qin *et al.*, “A human gut microbial gene catalogue established by metagenomic sequencing,” *Nature*, vol. 464, no. 7285, pp. 59–65, Mar. 2010, doi: 10.1038/nature08821.

[107] A. Ignacio *et al.*, “Correlation between body mass index and faecal microbiota from children,” *Clinical Microbiology and Infection*, vol. 22, no. 3, p. 258.e1-258.e8, Mar. 2016, doi: 10.1016/j.cmi.2015.10.031.

[108] A. G. Wexler and A. L. Goodman, “An insider’s perspective: Bacteroides as a window into the microbiome,” *Nat Microbiol*, vol. 2, no. 5, p. 17026, Apr. 2017, doi: 10.1038/nmicrobiol.2017.26.

[109] A. N. Shkoporov and C. Hill, “Bacteriophages of the Human Gut: The ‘Known Unknown’ of the Microbiome,” *Cell Host & Microbe*, vol. 25, no. 2, pp. 195–209, Feb. 2019, doi: 10.1016/j.chom.2019.01.017.

[110] M. Touchon, J. A. Moura de Sousa, and E. P. Rocha, “Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer,” *Current Opinion in Microbiology*, vol. 38, pp. 66–73, Aug. 2017, doi: 10.1016/j.mib.2017.04.010.

[111] V. R. Carr, A. Shkoporov, C. Hill, P. Mullany, and D. L. Moyes, “Probing the Mobilome: Discoveries in the Dynamic Microbiome,” *Trends in Microbiology*, vol. 29, no. 2, pp. 158–170, Feb. 2021, doi: 10.1016/j.tim.2020.05.003.

[112] M. Zasloff, “Antimicrobial peptides of multicellular organisms,” *Nature*, vol. 415, no. 6870, pp. 389–395, Jan. 2002, doi: 10.1038/415389a.

[113] H.-K. Kang, C. Kim, C. H. Seo, and Y. Park, “The therapeutic applications of antimicrobial peptides (AMPs): a patent review,” *J Microbiol.*, vol. 55, no. 1, pp. 1–12, Jan. 2017, doi: 10.1007/s12275-017-6452-1.

[114] N. Mookherjee, M. A. Anderson, H. P. Haagsman, and D. J. Davidson, “Antimicrobial host defence peptides: functions and clinical potential,” *Nat Rev Drug Discov*, vol. 19, no. 5, pp. 311–332, May 2020, doi: 10.1038/s41573-019-0058-8.

[115] M. Mahlapuu, J. Håkansson, L. Ringstad, and C. Björn, “Antimicrobial Peptides: An Emerging Category of Therapeutic Agents,” *Front. Cell. Infect. Microbiol.*, vol. 6, Dec. 2016, doi: 10.3389/fcimb.2016.00194.

[116] R. E. W. Hancock, E. F. Haney, and E. E. Gill, “The immunology of host defence peptides: beyond antimicrobial activity,” *Nat Rev Immunol*, vol. 16, no. 5, pp. 321–334, May 2016, doi: 10.1038/nri.2016.29.

[117] L. García-Bayona and L. E. Comstock, “Bacterial antagonism in host-associated microbial communities,” *Science*, vol. 361, no. 6408, p. eaat2456, Sep. 2018, doi: 10.1126/science.aat2456.

[118] M. Sassone-Corsi *et al.*, “Microcins mediate competition among Enterobacteriaceae in the inflamed gut,” *Nature*, vol. 540, no. 7632, pp. 280–283, Dec. 2016, doi: 10.1038/nature20557.

[119] N. A. Moran and D. B. Sloan, “The Hologenome Concept: Helpful or Hollow?,” *PLOS Biology*, vol. 13, no. 12, p. e1002311, Dec. 2015, doi: 10.1371/journal.pbio.1002311.

Anexos

Anexo 1: Artículo de investigación *Metatranscriptomic analysis to define the Secrebiome, and 16S rRNA profiling of the gut microbiome in obesity and metabolic syndrome of Mexican children*

Anexo 2: Artículo de revisión *Perspectives in Searching Antimicrobial Peptides (AMPs) Produced by the Microbiota*

Anexo 3: Artículo de investigación *Bioactive Plasmid- and Phage-Encoded Antimicrobial Peptides (AMPs) in the Human Gut: A Metatranscriptome–Virome Profiling Reveals Exploratory Links to Metabolic Human Diseases*

Microbial Cell Factories

RESEARCH

Open Access

Metatranscriptomic analysis to define the Secrebiome, and 16S rRNA profiling of the gut microbiome in obesity and metabolic syndrome of Mexican children

Luigui Gallardo-Becerra[1†], Fernanda Cornejo-Granados[1†], Rodrigo García-López[1†], Alejandra Valdez-Lara[1], Shirley Bikel[1], Samuel Canizales-Quinteros[2], Blanca E. López-Contreras[2], Alfredo Mendoza-Vargas[3], Henrik Nielsen[4] and Adrián Ochoa-Leyva[1*]

Abstract

Background: In the last decade, increasing evidence has shown that changes in human gut microbiota are associated with diseases, such as obesity. The excreted/secreted proteins (secretome) of the gut microbiota affect the microbial composition, altering its colonization and persistence. Furthermore, it influences microbiota-host interactions by triggering inflammatory reactions and modulating the host's immune response. The metatranscriptome is essential to elucidate which genes are expressed under diseases. In this regard, little is known about the expressed secretome in the microbiome. Here, we use a metatranscriptomic approach to delineate the secretome of the gut microbiome of Mexican children with normal weight (NW) obesity (O) and obesity with metabolic syndrome (OMS). Additionally, we performed the 16S rRNA profiling of the gut microbiota.

Results: Out of the 115,712 metatranscriptome genes that codified for proteins, 30,024 (26%) were predicted to be secreted, constituting the Secrebiome of the gut microbiome. The 16S profiling confirmed an increased abundance in Firmicutes and decreased in Bacteroidetes in the obesity groups, and a significantly higher richness and diversity than the normal weight group. We found novel biomarkers for obesity with metabolic syndrome such as increased Coriobacteraceae, *Collinsela*, and *Collinsella aerofaciens*; Erysipelotrichaceae, *Catenibacterium* and *Catenibacterium sp.*, and decreased *Parabacteroides distasonis*, which correlated with clinical and anthropometric parameters associated to obesity and metabolic syndrome. Related to the Secrebiome, 16 genes, homologous to *F. prausniitzi*, were overexpressed for the obese and 15 genes homologous to Bacteroides, were overexpressed in the obesity with metabolic syndrome. Furthermore, a significant enrichment of CAZy enzymes was found in the Secrebiome. Additionally, significant differences in the antigenic density of the Secrebiome were found between normal weight and obesity groups.

Conclusions: These findings show, for the first time, the role of the Secrebiome in the functional human-microbiota interaction. Our results highlight the importance of metatranscriptomics to provide novel information about the

*Correspondence: aochoa@ibt.unam.mx
†Luigui Gallardo-Becerra, Fernanda Cornejo-Granados, Rodrigo García López—Co-first authors
[1] Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, C.P. 62210 Cuernavaca, Morelos, Mexico
Full list of author information is available at the end of the article

gut microbiome's functions that could help us understand the impact of the Secrebiome on the homeostasis of its human host. Furthermore, the metatranscriptome and 16S profiling confirmed the importance of treating obesity and obesity with metabolic syndrome as separate conditions to better understand the interplay between microbiome and disease.

Keywords: Secrebiome, Microbiota, Microbiome, Obesity, Metabolic syndrome, Metatranscriptome, Metatranscriptomics, AAR, CAZY, Secretome

Background

The gut is an essential metabolic, endocrine, and immune organ inhabited by millions of microbes. Recently, our knowledge of the diversity of microorganisms in the human gut has exponentially increased with the development of high-throughput sequencing technologies [1]. The intestinal microorganism community (gut microbiota) is a dynamic ecosystem with critical functional roles in the development and physiology of its host, preventing the growth of pathogenic bacteria, modulating the immune response, affecting nutrient absorption, regulating metabolic processes, etc [1]. Meanwhile, the host's intestinal microenvironment also impacts the structure and function of the microorganisms which inhabit it [2, 3]. Although metagenomic approaches provide information on the microbiota composition and the potential functions of the codified genes, the expression profile of the community is needed to know what genes are expressed under certain conditions [4, 5].

High-throughput sequencing of RNA transcripts (RNA-seq) from microbial communities (metatranscriptome) allows an unprecedented opportunity to analyze the functional and taxonomical dynamics of the expressed microbiome, which can be associated with human health and disease [1]. The microbial encoded genes are not directly correlated with their transcription; up to 41% of microbial transcripts in the human gut have different relative abundances as compared to their genome content [4]. Metatranscriptome analysis has been applied to the human gut microbiota, revealing changes on the microbial community gene-expression profile during the exposure to dietary [6] and xenobiotic interventions [7] or with the presence of inflammatory diseases [8], showing divergent expression profiles of microbial community (subject-specific expression) [9], as well as a core of expressed functions [10, 11]. Recently, a metatranscriptomic 'core' universally transcribed over time and across participants, often by different microorganisms, has been identified [3]. Thus, metatranscriptomics emerges as a highly informative approach to analyze the expressed functional dynamics of microbial communities that can be associated with the presence of diseases [1].

The secretome is defined as the complete set of Excreted/Secreted (ES) proteins of a cell [12]. These proteins play a critical role in biological processes important for gut colonization, persistence, interaction with mucosal cells, activating signaling pathways, contributing to probiotic effects, etc. In pathogenic bacteria, the secretome plays a crucial role in parasitism, modulating the host immune response and promoting the proliferation of infection [13–15]. Millions of years of co-evolution between the human host and its gut microbial community have resulted in a continuous dialogue of the microbiota with the immune system promoting the gut homeostasis. Thus, the collection of secreted proteins represents proteins released by the microbiota into the intestinal lumen for bacteria-bacteria and bacteria-host interactions regulating the interaction within the gut. However, current knowledge of the molecular mechanisms responsible for the cross-talk between the gut microbiota and the human host remains incomplete.

There are several examples of the role that ES proteins from the microbiota play in the gut's health. Bile salt hydrolases of lactobacilli reduce blood cholesterol and diminish the risk for cardiovascular diseases [16, 17], and show activity against intestinal protozoan parasites [18]. Other enzymes such as N-acylated homoserine lactone (AHL)-lactonase help modulate the structure of the microbiota by decreasing the quorum-sensing of pathogenic bacteria [19]. Several secreted proteins act as essential mediators for the establishment of a bifidobacteria-host immune system dialogue [20]. However, no study has been conducted to examine the total secretome of the gut microbiota, and most importantly, what is the expressed secretome codified in the metatranscriptome of the human gut microbiome. Thus, the primary goal of this study was to determine the expressed secretome of the human gut microbiome associated with childhood obesity.

An essential step for the energy harvesting by the microbiota involves the carbohydrate degradation from food. Microbial enzymes conduct the breakdown of the diet's complex oligosaccharides into fermentable monosaccharides and their posterior transformation into components that can be absorbed by the intestinal cells and the microbiota [21]. Given that the human genome

lacks the enzymes required to cleave the glycosidic bonds of the complex dietary polysaccharides [22], this process is mainly performed by three types of Carbohydrate active enzymes (CAZymes) enzymes: the Carbohydrate Esterases (CEs), Glycoside Hydrolases (GH), and Polysaccharide Lyases (PL). CAZymes plays an essential role in determining the nutritional status of individuals. The carbohydrate degradation resulting from the bacterial fermentation of fiber-rich diets constitutes the first step in the production of short-chain fatty acids (SCFAs), essential molecules generally associated with multiple health-promoting and beneficial metabolic effects for the host. However, the increased energy harvest derived by this process has also been proposed to contribute to diet-induced obesity in mice [23].

Here, we report a deeply sequenced metatranscriptome study of human gut microbiome samples to elucidate the functional profile of the Secrebiome. We carried out a high throughput shotgun sequencing of the total RNA and a profiling of the 16S rRNA gene using the V4 region as a marker for the phylogenetic diversity of the expressed bacterial community. Then, we constructed a de novo metatranscriptome to assess the total bacterial genes expressed and secreted that are present in the different groups. Using high-throughput sequencing, we determined for the first time, that ~ 26% of the total genes expressed in the metatranscriptome corresponded to potentially secreted proteins, which may serve as a critical mechanism to regulate the structure–function of the microbiome and their relationship with the host. Furthermore, 31 of these genes were differentially expressed in the obese (O) and obese with metabolic syndrome (OMS) groups. We also observed some novel correlations between differentially abundant taxa and the clinical and anthropometric parameters of the children cohort. When characterizing the Secrebiome, we observed an elevated enzymatic activity, mostly from hydrolases and transferases, which suggests the critical role that these proteins have in the metabolism of nutrients present in the host system.

Results

Different microbiota structures were found in normal weight and obesity groups

The study population of this work consisted of 27 children, classified as follows: 7 normal weight (NW), 10 with obesity (O), and 10 with obesity and metabolic syndrome (OMS) selected from a cohort of 750 children around 9 years of age. From the cohort, 333 samples were collected for RNA extraction, where 65% had normal weight, and the remaining 35% were obese. Importantly, 52.94% of obese children also had metabolic syndrome. Anthropometric and biochemical characteristics of the population are shown in Additional file 1: Table S1. Most of anthropometric and clinical parameters related to the obesity and obesity with metabolic syndrome were statistically different among the NW, O and OMS groups (Additional file 1: Table S1).

For the 16S profiling of the microbiota associated with the three groups, we sequenced the corresponding V4 region amplicons of the 16S rRNA gene, and after the application of quality filters, we obtained 2,937,796 joined reads that were classified into 801 OTUs at 97% sequence identity. All samples successfully recovered most of the groups variation, as seen in the rarefaction curve, which shows saturation of diversity at a sequence depth of 20,000 reads (Additional file 2: Fig. S1a, b). Regarding the associated taxonomy of all samples, the bulk of OTU abundance showed that the three groups are dominated by the same three phyla Firmicutes, Bacteroidetes, and Proteobacteria, accounting for 39%, 56%, and 3% of the total reads, respectively (Fig. 1a and Additional file 3: Fig. S2a–f). Firmicutes was markedly increased in the OMS (46.44%), and O (40.29%) as compared to the NW group (26.65%), although only the difference between NW and OMS was statistically significant ($p = 0.0431$). The opposite effect was observed with Bacteroidetes, which was increased in NW (66.88%) as compared to the O (55.27%) and OMS (45.48%), although only the difference between the NW and OMS was significant ($p = 0.0249$). Interestingly, the ratio of Firmicutes/Bacteroidetes was NW = 0.55, O = 0.92, and OMS = 1.20, with significance difference only between NW Vs OMS ($p = 0.025$). Furthermore, both the OMS and O groups exhibited significantly higher richness and diversity than the NW group, with the larger values in the OMS group (Fig. 1b, c). The between-sample diversity comparison carried out with a PCoA ordination based on unweighted UniFrac distances showed no defined clusters by group ($R^2 = 0.066$, $p = 0.75$), although the most disperse one was the OMS (based on the within-group variation), followed by the O and NW groups (Fig. 1d). The weighted distances showed the same behavior (Additional file 4: Fig. S3).

We found significant differences in the abundance of 41 taxa at different taxonomic levels between the three groups (Fig. 2). The class Coriobacteria, order Coriobacteriales, family Coriobacteraceae, genus *Collinsella* and species *Collinsella aerofaciens* were significantly more abundant specifically in the OMS group (when compared against both the NW and O groups), thus suggesting them as potential biomarkers for obesity with metabolic syndrome. Likewise, the genus *Porphyromonas* and, more specifically, an undetermined species in the same genus were significantly more abundant specifically in the O group when compared against NW and OMS, suggesting them as potential biomarkers for obesity.

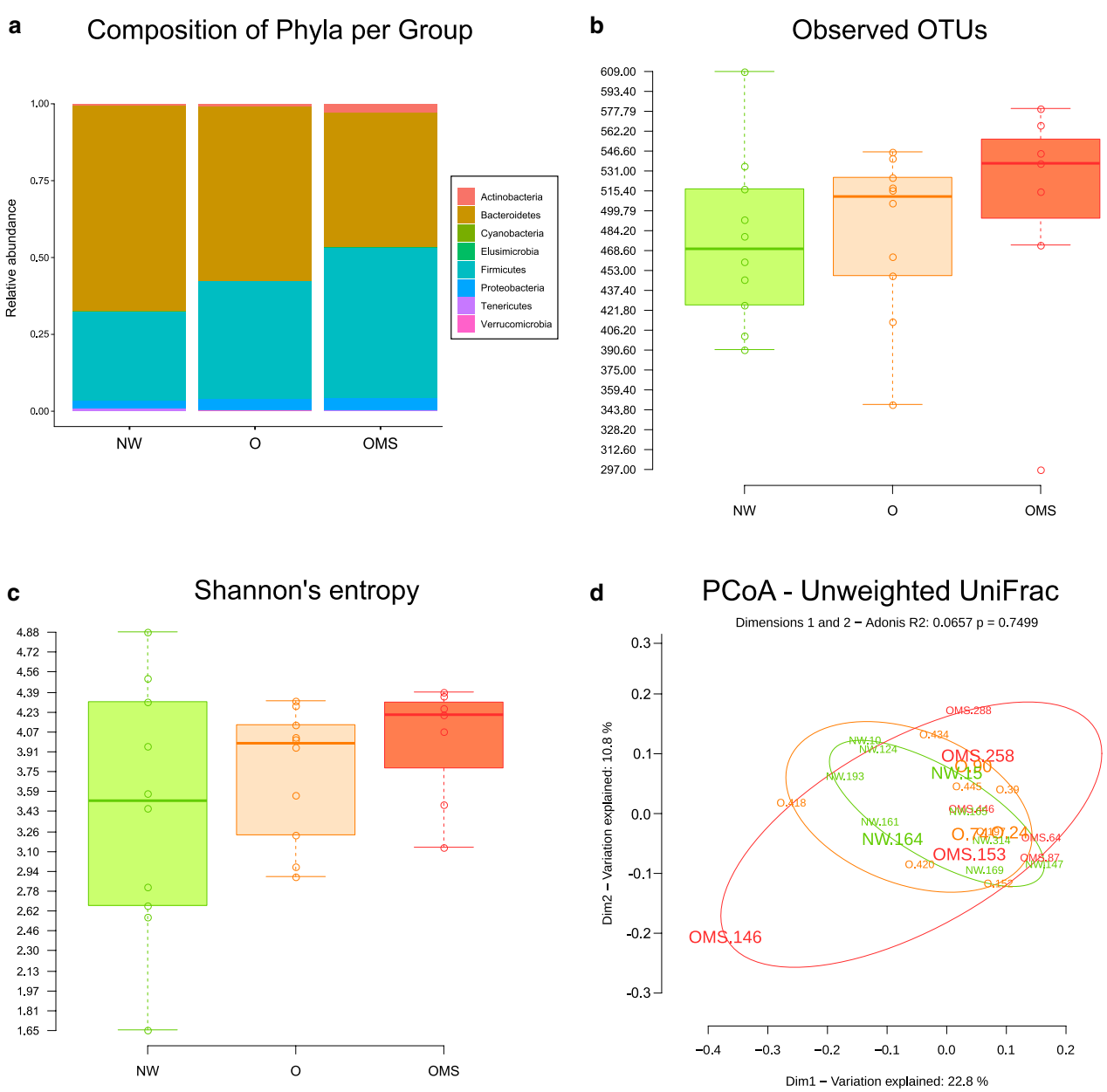


Fig. 1 Analysis of the 16S rRNA gene profiles. Each panel compares the NW, O, and OMS groups. **a** Mean Relative abundance of the phyla present in each group. **b** Mean observed OTUs per sample and group. The averaged total unique OTUs from 10,000 rarefactions per sample are shown as points with overlying boxplots showing the distribution within each group. **c** Mean Shannon's entropy per sample and group. The averaged Shannon's index value from 10,000 rarefactions per sample are shown as points with overlying boxplots showing the distribution within each group. **d** Principal coordinate analyses of Unweighted UniFrac distances. Elipses were calculated based on the most distant samples per group. Samples for which RNA-seq information is available are presented with a larger font

Regarding pairwise group statistical significant comparisons, the genus *Phascolarctobacterium* and one undetermined species within this genus were more prevalent in the NW as compared to O, while Bacteroidetes, Bacteroidia, and Bacteroidales were more abundant in the NW than in the OMS group (Fig. 2). The O group showed an over-abundance of bacteria from the *Porphyromonas* and *Faecalibacterium* genera, as well as *Faecalibacterium prausnitzii*, *Bifidobacterium adolescentis*, and an undetermined species of *Porphyromonas* as compared to NW, while *Parabacteroides distasonis* was over-abundant in the O than in the OMS (Fig. 2). The OMS group

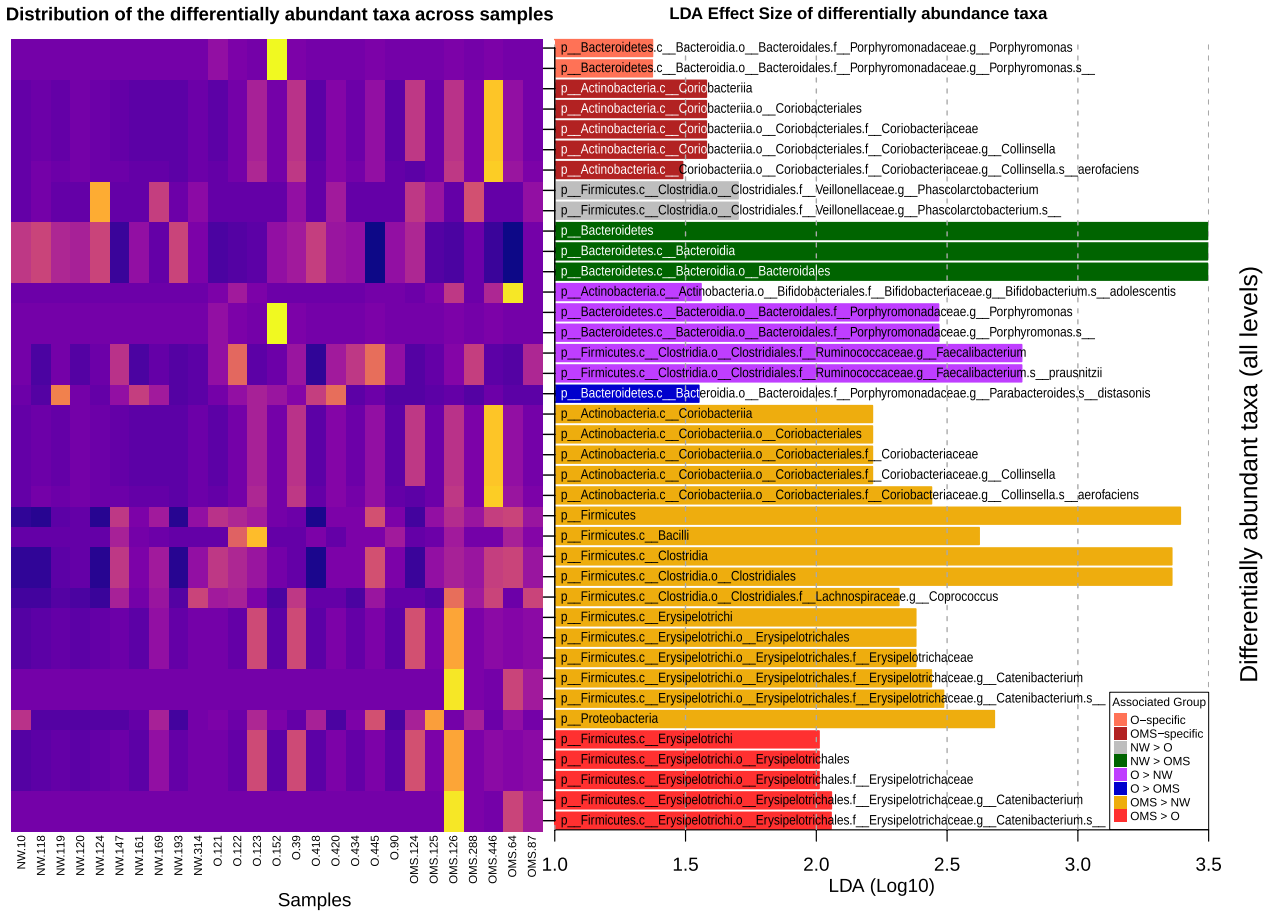


Fig. 2 Differentially abundant taxa in association with groups and their distribution in the samples. Bars show the effect size of taxa that are differentially more abundant in the specified groups after analysis with LEfSe. Those labeled as specific were found significantly more in their groups, whereas the rest are found in pairwise group comparisons. The heatmap shows the abundance of those taxa in the samples with more abundant taxa in yellow

showed a larger number of differentially abundant taxa at different levels, with the class Erysipelotrochi, order Erysipelotrichales, family Erysipelotrichaceae, genus *Catenibacterium* and an undetermined *Catenibacterium* species reported more over-abundant in the OMS group than in either the NW or O groups when considered separately (Fig. 2). Additionally, the OMS had over-abundance of phyla Firmicutes and Proteobacteria; class Coriobacteriia, Bacilli, and Clostridia; order Coriobacteriales and Clostridiales; family Coriobacteriaceae, genera *Colinssella*, and *Coprococcus*; and species *Colinsella aerofaciens* when compared to NW (Fig. 2).

Microbiota changes are associated with anthropometric and clinical parameters

We analyzed whether the 41 differentially abundant taxa were correlated with changes in the anthropometric and clinical parameters that are involved in obesity and metabolic syndrome (Additional file 5: Table S2). Interestingly, *Collinsella aerofaciens* which was over-abundant in OMS vs all groups, showed a significant positive correlation with triglycerides ($r = 0.62$, $p = 0.00053$) (Fig. 3a), while showing a negative correlation with HDL ($r = -0.4$, $p = 0.039$) (Fig. 3b). Interestingly, *C. aerofaciens* also showed a positive, albeit weak, correlation with waist circumference ($r = 0.28$, $p = 0.16$) and BMI ($r = 0.34$, $p = 0.078$). In the upper taxonomic levels, the genus *Collinsella* showed a positive significant correlation with triglycerides ($r = 0.64$, $p = 0.00029$) (Fig. 3c), and BMI ($r = 0.39$, $p = 0.044$) (Fig. 3d), and a negative correlation with HDL ($r = -0.4$, $p = 0.036$) (Fig. 3e), and a positive but weak correlation with waist circumference ($r = 0.35$, $p = 0.073$). The same behavior was observed with the corresponding family Coriobacteriaceae, order Coriobacteriales and class Coriobacteriia (Additional file 5: Table S2). We also found positive correlations of tryglicerides and BMI with Clostridiales, Clostridia (Additional file 5: Table S2), and Firmicutes (Additional

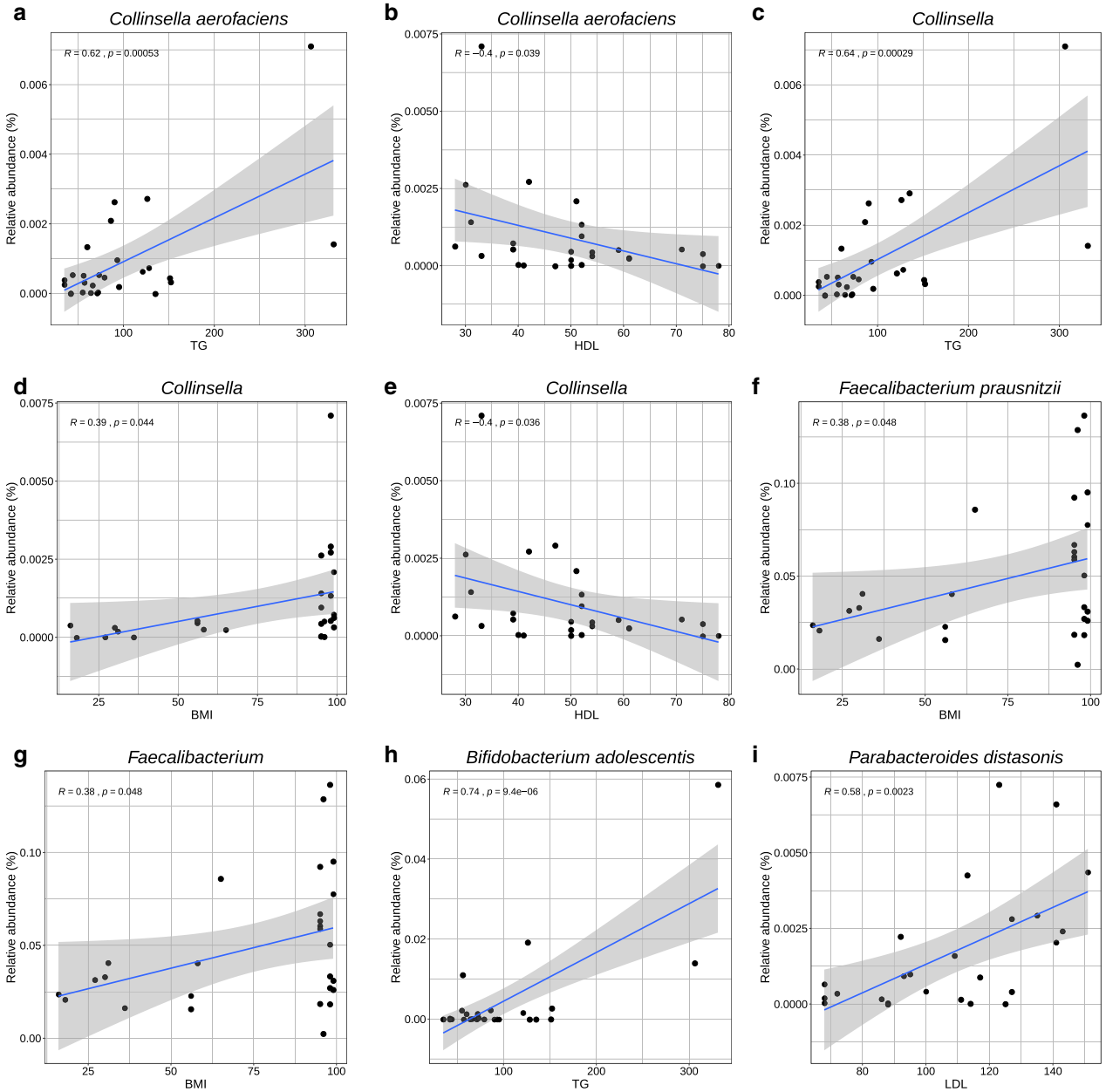


Fig. 3 Linear regression and Pearson correlation of LEfSe biomarkers. The title of each graph corresponds to the LEfSe biomarkers taxa. The x-axis shows the value of the clinical and anthropometrical parameters, and the y-axis shows the relative abundance for each taxa

file 5: Table S2), which were more abundant in the OMS vs NW groups. Contrary, Bacteroidetes, Bacteroidia and Bacteroidales, which were over-abundant in NW when compared to OMS were negatively correlated with tryglicerides and BMI (Additional file 5: Table S2). In addition, *Faecalibacterium prausnitzii* (Fig. 3f) and *Faecalibacterium* (Fig. 3g) showed a positive correlation with BMI (r = 0.38, p = 0.048) while *Bifidobacterium adolescentis* correlated positively with triglycerides (r = 0.74, p = 0.0000094) (Fig. 3h) and a negative but weak correlation with HDL (r = − 0.34, p = 0.082). Both species were significantly more abundant in O with respect to NW. Furthermore, *Parabacteroides distasonis* which was more abundant in O when compared to OMS showed a positive correlation with LDL (r = 0.58, p = 0.0023) (Fig. 3i). Importantly, we found a positive correlation with waist-circumference and taxa in different levels, including Erysipelotrichi, Erysipelotrichales

and Erysipelotrichaceae, and Bacilli (Additional file 5: Table S2), which were more abundant in OMS pairwise comparisons against NW and O.

RNAseq resulted in a representative assembly of the metatranscriptome

To explore the expressed taxonomy and functions associated with obesity and metabolic syndrome, we conducted a metatranscriptomic analysis. To this end, we obtained and sequenced the total RNA of eight fecal samples (NW = 2, OB = 3, and OMS = 3). The integrity of the total RNA showed the dominance of the 16S/23S and 18S/28S rRNA peaks (Additional file 6: Fig. S4). The samples were rRNA depleted and sequenced and after quality control and removal of sequencing artifacts, a total of 172,444,756 million reads with an average read length of 92 nt were obtained. After the removal of Eukaryotic and Prokaryotic ribosomal RNA and human transcripts, we obtained 110,014,240 RNA-seq reads, which were used for the de novo assembly of the metatranscriptome. Next, we obtained an assembly comprised of 224,427 contigs with an N50 of 702 bp and a total of 125,015,187 nucleotides assembled (Additional file 7: Table S3). From these contigs, we obtained 115,712 open reading frames (ORF) containing a protein sequence. Notably, > 54% of the total reads from the eight samples mapped back to the assembly, suggesting that our assembly covers a broad fraction of the sequence spectrum for all the samples

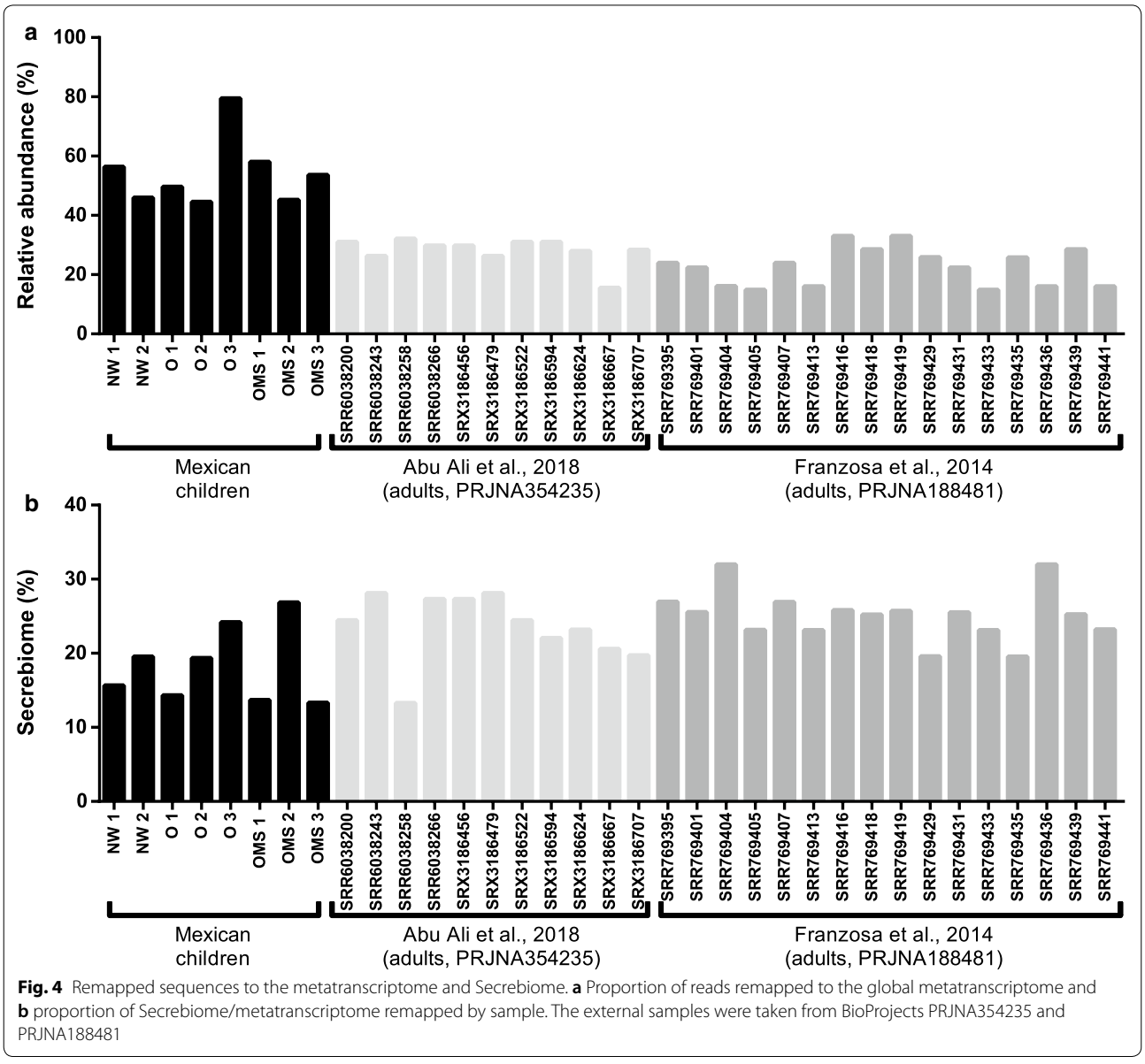


Fig. 4 Remapped sequences to the metatranscriptome and Secrebiome. **a** Proportion of reads remapped to the global metatranscriptome and **b** proportion of Secrebiome/metatranscriptome remapped by sample. The external samples were taken from BioProjects PRJNA354235 and PRJNA188481

(Fig. 4a). In the same manner, the read mapping to our assembly was performed with metatranscriptomic samples collected from twenty-seven healthy adults from two independent studies (NCBI BioProjects PRJNA354235 and PRJNA188481). Interestingly, > 25% of the reads mapped to our assembly (Fig. 4a), suggesting that our metatranscriptome also covered a good proportion of the sequence spectrum in adults, considering the known differences of the microbiota composition among children and adults.

Determination of the Secrebiome

Proteins can be secreted by bacteria through multiple secretory mechanisms. Thus, we used a bioinformatic strategy [14, 15] to predict the secreted proteins encoded in our metatranscriptome. Out of the 115,712 genes that codify for proteins, we predicted 30,024 as potentially secreted, which represented ~ 26% of the total proteins in our metatranscriptome. All these proteins were referred to as the secretome of the microbiome, the "Secrebiome". Notably, this represents 31.10% of the expression of the total metatranscriptome. From the 30,024 secreted proteins, 96.57% showed significant similarity against homologs in the NCBI's non-redundant protein database (nr), 64.26% were assigned a Gene Ontology (GO) term, and 19.84% were assigned an Enzyme Commission (EC) number. The GO terms distribution showed that the Secrebiome was mainly composed of Hydrolase activity (18.4%) at Molecular Function (Additional file 8: Fig. S5a), membrane (37%) at Cellular Component (Additional file 8: Fig. S5b); and organic substance metabolic process (21%) at Biological process (Additional file 8: Fig. S5c). Finally, the EC classification showed that Hydrolases and Transferases were the most abundant terms with 47.5% and 23.4% of the EC sequences, respectively (Additional file 8: Fig. S5d). The GO and EC recruitment plots showed a relatively complete collection of the total ontology and enzyme variability as most curves passed their inflection points after 10,000 observations in most samples (Additional file 9: Fig. S6a, b), suggesting an adequate functional coverage of our samples to the Secrebiome.

Importantly, when we mapped the RNA-seq reads of the eight samples to the Secrebiome > 18% of the reads that mapped to the total metatranscriptome, aligned back to the Secrebiome (Fig. 4b), suggesting that an important proportion of reads potentially represented secreted proteins and thus may be in direct contact with the host. Notably, the reads from the two external published BioProjects, mapped back > 24% of the reads (Fig. 4b), suggesting that approximately a quarter of the sequences in the metatranscriptome of adults, correspond to the expression of the Secrebiome. Interestingly, the read proportion that mapped to our samples is more variable than the percentage obtained for the studies on adults, suggesting that the Secrebiome in children is more variable among individuals than in adults (Fig. 4b).

Differential expression of the Secrebiome among obesity groups

A comparison of the predicted abundance values (RSEM) of the Secrebiome transcripts from the obese groups resulted in the identification of 31 transcripts differentially expressed in either the O (16 transcripts) and OMS (15 transcripts) groups (Fig. 5). Because RNA-seq data was only available for two samples in the NW group and their contribution to the set was limited, this group was excluded from the DESeq differential expression analysis to avoid a detrimental effect in the internal normalization. Regardless, the resulting 31 differentially expressed transcripts obtained for both the O and OMS groups were compared against an independent normalization carried out separately with the NW group to compare their expression in this group and the obese groups. As seen in Fig. 5 and Additional file 10: Figure S7, the fold change of expression of the 31 transcripts was mainly driven by a strong signal in the O and OMS groups, with a negligible impact from the NW group, suggesting that the comparison of the overexpressed genes is associated to the obesity phenotype.

The transcript DN71185_c2_g1, was one of the most strongly associated with the O group. It was cross-referenced to a carbohydrate-binding module (CBM26), and its sequence was homologous to an alpha-amylase precursor. Further, the taxonomy assigned to most of the differentially expressed transcripts corresponded to *Faecalibacterium prausnitzii,* mainly present in the O group (9 out of 16) and to the genera Bacteroides, mainly present in the OMS group (6 out of 15) (Fig. 5). Interestingly, nine of the 31 transcripts had significant correlations with anthropometric and biochemical measurements (Additional file 11: Fig. S8). The transcripts corresponding to a molecular chaperone DnaJ (DN68363_c1_g1) and a chaperone protein DnaK (DN73431_c2_g5) negatively correlated with LDL levels. One transcript corresponding to an integrase (DN71618_c1_g3) correlated positively with triglycerides and weight, and five transcripts correlated positively with diastolic and systolic blood pressure. Finally, two transcripts correlated negatively with glucose levels, TlpA family protein disulfide reductase (DN72347_c2_g3), and cytochrome b6 (DN70151_c0_g2). Furthermore, a gene coding for integrase (DN71618_c1_g3) was positively correlated with triglycerides and weight, while the 30S ribosomal protein S14 (DN70736_c1_g8) was also positively correlated

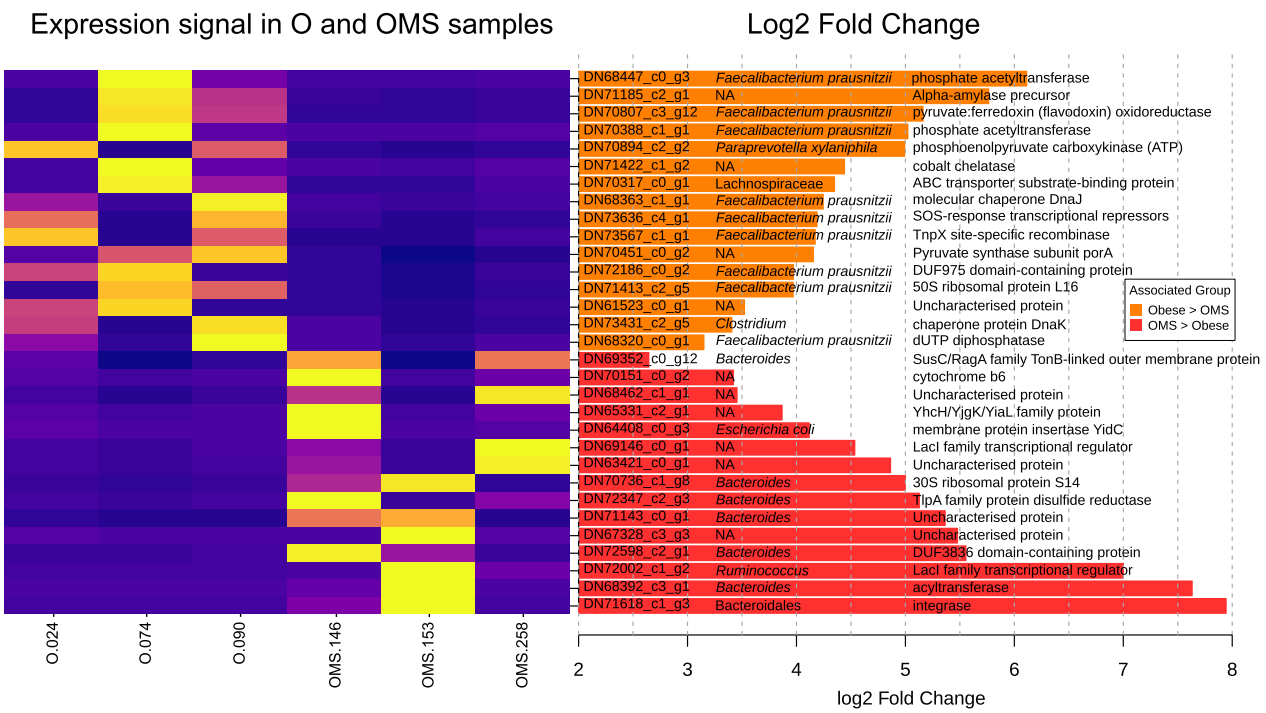


Fig. 5 Differentially expressed secretome transcripts. Transcripts significantly associated with either the O or OMS were determined with DESeq after expression signal standardization. The heatmap shows the normalized RSEM abundance of each transcript per sample with a higher abundance in yellow. Bars on the right depict the total fold transcripts detected in association to each group (in log2 scale). The overlying matrix contains the predicted protein name on the left, the putative taxonomic assignation based on an LCA approach on the center, and the associated functional annotation on the right

with weight. All these transcripts can be suggested as potential biomarkers to differentiate the microbiome expression profiles among obesity and obesity with metabolic syndrome.

CAZY enzymes were enriched in the Secrebiome

One of the most critical roles of the microbiota is its ability to utilize complex carbohydrate sources. Thus, classification according to the Carbohydrate-Active enzymes (CAZy) was performed to assess which expressed secreted proteins possessed the ability to enhance the carbohydrate-harvesting activity and analyze their association with obesity phenotypes. This analysis resulted in the identification of 2249 secreted proteins that had catalytic or carbohydrate-binding modules involved in the degradation and modification of carbohydrates. Interestingly, enrichment of CAZY enzymes was observed in the Secrebiome (7.5%) as compared to the non-secreted proteins (3.0%) of the metatranscriptome. The most abundant CAZY enzyme classes in the Secrebiome were Glycoside Hydrolases (GH; 35.1%), Carbohydrate-Binding Module (CBM; 33.8%), S-Layer Homology domain (SLH; 13.7%), and Carbohydrate Esterases (CE; 9.7%) (Additional file 12: Fig. S9).

We did not find significant differences in the distribution of CAZY enzymes within the Secrebiomes or within the non-secreted proteins of the three groups. However, when we compared the secreted against the non-secreted proteins, we found a significant abundance enrichment of CAZY enzymes from cohesin, SLH, dockerin, CBM, and Polysaccharide Lyases (PL) classes in the Secrebiome when compared to the non-secreted proteins of the metatranscriptome (Fig. 6). Contrastingly, a significant underrepresentation was observed in the Secrebiome for Glycosyltransferases (GT), Auxiliary Activities (AA), and CE CAZy classes. Only the abundance of the glycoside hydrolases was not significantly different among Secrebiome and non-secreted proteins (Fig. 6). These data suggest a differential CAZy activity in the Secrebiome. Interestingly, the families enriched in the Secrebiome seemed to be focused on binding and receptor functions, which suggest a role in the communication between the microbial communities and of the bacteria with the host.

The Secrebiome of obesity had more abundance of antigenic regions (AAR)

Because an important proportion of the expressed proteins may be secreted and therefore, be in direct contact

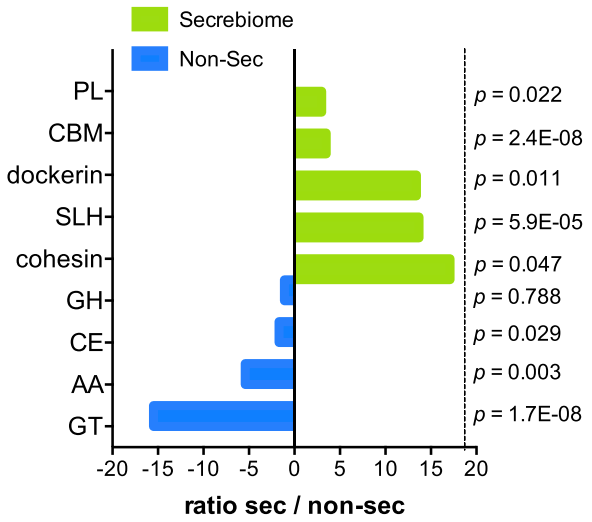
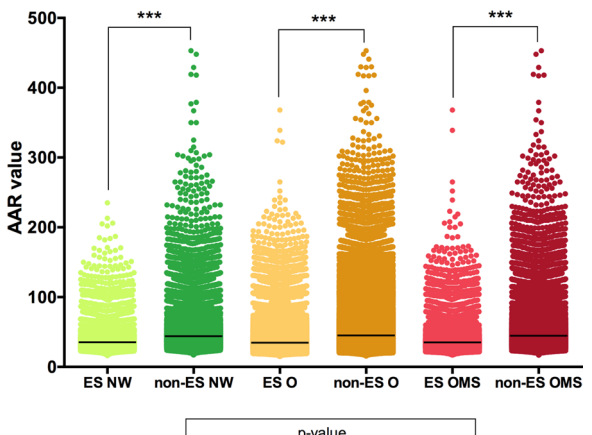


Fig. 6 Ratio of Carbohydrate-Active enzymes (CAZys) in secreted and non-secreted proteins. The bars show the ratio of each CAZy family present in the total set of secreted (green)and non-secreted (blue) proteins. We used multiple t-tests to analyze the significant differences among the different enzyme families



p-value	
ES NW vs non-ES NW	< 2.2E-16
ES O vs non-ES O	< 2.2E-16
ES OMS vs non-ES OMS	< 2.2E-16
ES NW vs ES O	4.6E-04
ES NW vs ES OMS	0.471
ES O vs ES OMS	0.002

Fig. 7 Comparison between the abundance of antigenic regions values (AAR) of secreted and non-secreted proteins in the different groups. The x-axis shows the protein set analyzed, and the Y-axis shows the AAR value obtained for each protein in the dataset. The line shows the mean value for each group. A Mann–Whitney test was performed to compare the AAR within each group with a confidence level of 99% (***$p \leq 0.001$)

with the host, we determined the antigenic potential of the Secrebiomes calculating the abundance of antigenic regions (AAR) value. We found that the Secrebiomes of NW, OB, and OMS had significantly more antigenic density (Mann–Whitney test $p \leq 0.0001$) than the corresponding sets of non-secreted proteins for each phenotype (Fig. 7). Interestingly, the Secrebiome of the NW (AAR = 40.33) was more antigenic than OMS (AAR = 40.5) and O (AAR = 40.66). However, significant differences among Secrebiome was only observed in comparisons of NW vs O ($p = 0.0004673$), and O vs OMS ($p = 0.002006$), while NW vs OMS was not significant ($p = 0.4709$). This result suggests that the Secrebiome associated with obesity decreased the expression of proteins associated with a higher antigenic density.

Discussion

The current study presents, to the best of our knowledge, the first insight into the expressed genes corresponding to the secreted proteins of the microbiome, the Secrebiome, associated with obesity with, and without metabolic syndrome, in children around 9 years of age. The current Mexican cohort reflects the unrelenting prevalence of the obese population worldwide, a growing problem that is more pronounced in developing and developed countries. In 2017, the Organization for Economic Cooperation and Development (OECD) reported that the United States of America and Mexico led obesity rankings. In these terms, children are a particularly susceptible population that has been often overlooked by governmental policies and scientific studies but comprise an essential economic and health issue for the future of these countries. In our cohort of 750 individuals, we found that 52.94% of the obese children have metabolic syndrome, which is similar to the 62% obtained in previous reports for Mexican obese children [24].

The main focus of the current study was to assess the actual expressed profiles of the microbiota, using a metatranscriptomic approach (RNA-Seq). Whereas the metagenome may show the full functional potential of the microbiota, only a small fraction of the bacterial genes are expressed [25]. Thus, a metatranscriptomic approach provides a more accurate description of the underlying dynamics of obesity, by capturing the community transcript population representing each taxonomic group in the study [25]. Although our results from metatranscriptomics should be interpreted with caution due to the limited sample size, this work provides a framework for further studies using metatranscriptomics to analyze the expressed genes of the Secrebiome and its close relationship with the host under diseases.

We used the V4 region of the 16S rRNA gene to establish the taxonomic profile of the microbiota from fecal samples of three children groups: normal weight (NW), obese (O), and obese with metabolic syndrome (OMS). We observed that the three groups are dominated by the

same three phyla, Firmicutes, Bacteroidetes, and Proteobacteria. However, we detected specific changes in the microbiota structure depending on the obesity type. The OMS patients had the highest bacterial diversity, followed by O and NW patients. This observation is in agreement with other studies with big cohorts of obese patients that also observe an increase in diversity in them [26]. Contrary, some studies revealed a decrease in diversity in obese patients [27, 28]. Regarding the Mexican children population, one study shows a higher diversity associated with obesity [29], and the other two studies showed no difference in diversity among obesity and normal-weight controls [30, 31]. Thus, the role of bacterial richness and diversity in children obesity remains unclear, suggesting that more studies are necessary. The PCoA analysis of the microbiota showed that the most compact cluster was formed among the NW samples, while the most disperse was the OMS group; the O group seemed to be an intermediate state of healthy and obese with metabolic syndrome. Similar clustering behavior was recently observed as a consequence of metabolic abnormalities in the adult population [26].

After analyzing the differentially abundant taxa, we observed that the genus *Porphyromonas* and an undetermined species within this genus were specifically over-abundant in the O group, suggesting that these taxa could be potentially used as biomarkers of obesity. Although we did not evaluate a concrete species, the presence of a typical member of this genus, *P. gingivalis* has been associated as a risk factor for developing obesity and diabetes [32]. Interestingly, we found that the class Coriobacteria, order Coriobacteriales, family Coriobacteraceae, genus *Collinsella* and species *Collinsella aerofaciens* were significantly more abundant, specifically in the OMS as compared to O and NW (Fig. 2), suggesting them as potential biomarkers for obesity with metabolic syndrome. Furthermore, this species also showed a significant positive correlation with triglycerides ($r=0.62$, $p=0.00053$), while showing a negative correlation with HDL ($r=-0.4$, $p=0.039$). Also, the genus *Collinsella* correlated positively with triglycerides ($r=0.64$, $p=0.00029$), BMI ($r=0.39$, $p=0.044$), and a negative correlation with HDL ($r=-0.4$, $p=0.036$), and had a weak positive correlation with waist circumference ($r=0.35$, $p=0.073$). The same behavior was observed with the corresponding family Coriobacteriaceae, order Coriobacteriales and class Coriobacteriia (Additional file 5: Table S2). We suggest that these taxa may be considered as biomarkers for OMS. In this regard, it was recently reported that an altered abundance of *Collinsella* genus changes the host's plasma cholesterol levels, showing a positive correlation with LDL cholesterol [33]. Additionally, another study with big cohorts of adults recently reported the finding of an increased *Collinsella* abundance in obese patients [26]. Also, it has been observed that low dietary fiber intake increases *Collinsella* abundance in the gut microbiota of obese pregnant women [34]. Evidence from literature strongly suggests that Coriobacteraceae are important constituents of gut microbiomes affecting the physiology of human and mice hosts [35]. Thus, it appears crucial to evaluate host-microbe interactions in finer detail, in the context of host lipid and cholesterol metabolism and their role with obesity and obesity with metabolic syndrome.

Importantly, the abundance of *Parabacteroides distasonis* was significantly decreased in OMS as compared to O but not with respect to NW. Interestingly this species correlated positively with LDL cholesterol levels ($r=0.58$, $p=0.0023$). Recently, a study showed that *P. distasonis* produces metabolites that can reduce weight gain and hyperglycemia as well as improve glucose metabolism and symptoms of obesity-related conditions such as liver disease [36]. In this regard, our findings highlight the important effects of this species as a biomarker of obese children with metabolic syndrome. Also, the class Erysipelotrichi, order Erysipelotrichales, family Erysipelotrichaceae, genus *Catenibacterium* and an undetermined *Catenibacterium* species were significantly more abundant in OMS groups when compared to the O and NW, and may prove interesting biomarkers for obesity with metabolic syndrome. This genus has only one species that has been properly described (*Catenibacterium mitsuokai*), and a study with humanized mice fed with a "western" diet revealed an increased representation of *Catenibacterium mitsuokai*, in the fecal samples of these animals when compared with mice fed with a low fat/plant polysaccharide diet [37].

There is evidence from literature documenting a potential role for the family Erysipelotrichaceae in host physiology. The increased abundance of this family has been associated with host dyslipidemia in the context of obesity, metabolic syndrome and hypercholesterolemia [38–40]. Additionally, nutritional studies also support the influence of dietary fat on the abundance of this family [41]. Our data highlights the importance of the family Erysipelotrichaceae in metabolic syndrome of obese children and suggest that *Catenibacterium* and its undetermined species may be interesting as novel biomarkers of this disease. On the other hand, we also found an increased abundance of Bacteroidales in NW as compared to the OMS group, and an increase in the abundance of Clostridiales in OMS as compared to NW group. Recently, a report in adults determined that the microbial gut community of obese people was characterized by higher Clostridiales whereas lean people tend to have higher Bacteroidales counts [42]. The microbiota

of obese adults recently revealed a high abundance of *Coprococcus*, a genus of Clostridiales [43], as well as in pregestational obesity [44], which is in accordance with our data, suggesting that these bacteria are important determinants in children obesity with metabolic syndrome.

The N50 of our metatranscriptome assembly was ~ 700 nt, which is congruent to the expected length of Prokaryotic mRNA molecules, (~ 900 pb long). Additionally, > 54% of the total reads from our samples mapped back to the assembly, suggesting an adequate assembly, even better than the ones obtained in other studies using a similar methodology [45]. Indeed, when we remapped sequencing data from different cohorts, we recovered ~ 25% of the reads despite them being of healthy adults from different populations, so this metatranscriptome and Secrebiome could be considered as a reference for further studies in different populations. Interestingly, the read proportion that mapped back to the Secrebiome from our children samples tended to be lower and more variable than the read proportion of the adult studies, suggesting that the adults tend to maintain a stable core of the Secrebiome, while children are more variable in read mapping abundance.

Bioinformatic prediction of the expressed genes corresponding to the Secrebiome is a novel methodology that could help us discover and understand novel communication mechanisms between the host and the microbiota. We found that ~ 26% of the total metatranscriptome expressed genes on the gut microbiota can be flagged as secreted proteins. The study of CAZymes in obesity studies is crucial due to their role in the degradation of complex carbohydrates that can be absorbed by the intestinal colonocyte. Contrary to a previous study of CAZymes in metagenome of obesity [46], we did not find differences in the Secrebiome between NW, O, and OMS, possibly due to the experimental design of our study. Still, we found that the abundance of *Faecalibacterium* and *Faecalibacterium prausnitzii* were significantly increased in O as compared to NW (Fig. 1e) and that this species had a significant positive correlation with BMI ($r = 0.38$, $p = 0.048$). It is well documented that *F. prausnutzii* can act as a probiotic due to its production of butyrate [47, 48], and it has been reported as depleted from the gut microbiota in individuals with metabolic syndrome [49], although we did not find a total depletion of this species in OMS group. However, a recent study with adult cohorts of obese patients found an increased abundance of *Faecalibacterium* [26], and also in a Mexican obese children cohort, an increase of abundance was reported [30]. Thus, further studies are necessary to clarify the role of *Faecalibacterium* in obesity.

Secreted proteins vary among the bacterial realm, as secretion systems are widely variable in complex communities such as the gut [50]. After establishing the characterization of the secreted proteins in our samples, the Gene Ontology and Enzyme Commission numbers showed a substantial prevalence of surface and membrane proteins in the Secrebiome. Furthermore, we also detected a marked prevalence of catabolic enzymatic activity and binding proteins that may contribute to nutrient uptake, polymer biodegradation, or cell attachment to the substratum or other cells. The Secrebiome presented a very different CAZY profile when compared to the rest of the metatranscriptome, with the Secrebiome being enriched in carbohydrate-catabolism proteins, namely glycoside hydrolases, carbohydrate-binding module, SLH domain bearing proteins, and carbohydrate esterases. Although the corresponding CAZy distribution was not group-associated, as a whole, the Secrebiome was particularly enriched in the production of SLH, and the dockerin-cohesin complex, congruent to binding requirements for the assembly of cellulosomes. Additionally, the abundance of antigenic regions (AAR), which evaluates antigenic density, found differences that were statistically significant between the NW and the O groups. As previous studies by our group have shown, the secretomes of more virulent pathogenic bacteria tend to have a higher antigenic density; this may be an indicator of pathogen activity in the obese subjects that may deserve further study [15, 51].

A total of 31 specific transcripts within the Secrebiome were deemed differentially abundant in association with either the O or OMS group. Since most were virtually absent from the NW (Fig. 4), we infer they may be related to the subjects' condition. Interestingly, most of those presenting a high fold change in the O group (Fig. 5) were homologous to the genome of *Faecalibacterium prausnitzii,* which, as mentioned previously, has been considered as probiotic. Contrastingly, most differentially expressed transcripts associated with the OMS group were homologous to the pangenome of *Bacteroides*, a genus that has been commonly reported to be associated with overweight subjects [52]. *Bacteroides* species were also predominant among urban gut microbiomes and contained unique gene clusters, which encode different CAZymes, whose functions could be related to an energy over-extraction of the microbiota form OMS patients [53]. Even so, it is essential to note that although we can evaluate the taxonomy of these specific marker transcripts, this is by no means a thorough representation of the whole taxonomy and cannot rule out their presence and activity in the O or NW groups. Only some correlations were significant with the clinical data associated with obesity. Of these, perhaps the most

interesting were those with LDL, commonly regarded as "bad cholesterol". Two chaperone proteins were found to be negatively correlated to LDL mg/dL, DnaK (HSP70), and DnaJ (HSP40), both more prevalent in the O group and less in the OMS group. These have been reported to interact with one another to carry out ATP-hydrolysis, which may be related to the thermogenesis and binding with unfolded polypeptide chains to prevent their aggregation.

Overall, this study showed that the definition of the Secrebiome provided new information on the gut microbiome functions that are directly related to the host communication, helping us to understand the functional interplay within the holobiont and open new insights in the study of the expressed microbiome. These findings may provide valuable insights to understand how the expressed bacterial genes and their respective proteins in the gut influence the metabolic response of the host to different nutrients and the alterations during obesity and obesity with metabolic syndrome.

Conclusions

One of the most exciting roles of the microbiota is its capacity to interact with the host and influence the overall health state. One of the mechanisms that the bacteria use to this end, is the secretion of proteins that interact directly with the host, metabolizing the nutrients for energy harvest and binding to host-cell receptors for signaling activities. The 16S rRNA profile of our samples showed the enrichment of taxa that have been previously reported as biomarkers for obesity such as Coriobacteraceae and *Collinsella* and suggest some new ones to be considered such as an increase of *Collinsella aerofaciens* and Erysipelotrichaceae, *Catenibacterium* and *Catenibacterium sp.*, and a decrease of *Parabacteroides distasonis*, which correlated with clinical and anthropometric parameters of obesity and metabolic syndrome. After comparing our metatranscriptome and Secrebiome with data from studies of other populations we observed that the proportion of mapped sequences was more variable for children samples than for adults, suggesting the establishment of a core Secrebiome in the adult population; further, given the high proportion of sequences that mapped back to our assembly, it could be considered as a reference for further studies in different populations. The differential expression analysis of the genes of Secrebiome proteins showed 31 significantly overexpressed genes for the O and the OMS group. Interestingly, these genes were homologous to *F. prausnutzii* in the O group and Bacteroides in the OMS group. Although this data does not establish the causal role of these taxa to the disease phenotypes, it cannot be discarded that some taxa could cause the dysbiosis present in a disease state.

Finally, the analysis of CAZy enzymes showed a differential distribution between the secreted and non-secreted proteins. The Secrebiome showed an increased presence of cohesin, SLH, dockerin, CBM, and PL domain bearing proteins, showing a potential critical role of the secreted proteins in the degradation and utilization of carbohydrates as nutrients. Overall, this study showed that the definition of the expressed genes of the Secrebiome provided new information on the gut microbiome functions that are directly related to the host health state and provide essential insights to understand how the bacterial proteins in the gut could influence the metabolic response of the host to different nutrients. The metatranscriptome and 16S profiling demonstrated the importance of the separation of obesity from obesity with metabolic syndrome patients for a better understanding of the microbiome role in the disease.

Methods

Study population, anthropometric and clinical parameters

We analyzed the stools from 10 normal weight (NW), 10 obese (O), and 7 obese with metabolic syndrome (OMS) children, aged 7–10 years old, from a summer-camp of children of Mexican Health Ministry employees. All children came from households with a middle economic class income and belonged to a similar sociocultural status. All of them lived in Mexico City at the time of collection and did not practice any sport regularly. The study groups were paired by gender and age. Samples were refrigerated at home at 4 °C and transported to the research facility within the following 12 h after collection in a portable cooler with ice packs to preserve the temperature. All samples were received at the research facility in the early morning; 200-mg aliquots were made and stored at − 70 °C in sterile plastic containers with RNA later. Obesity was defined by body mass index (BMI) ≥ 95th percentile, whereas NW was defined as BMI between the 15th and 75th percentiles for age and gender, based on the guidelines of the Centers for Disease Control and Prevention (CDC). Metabolic syndrome parameters were determined according to previous reports [24], and OMS were defined by the presence of waist circumference > 75th by age and gender, and at least two of the following metabolic traits: (1) triglycerides > 1.1 mmol/L (100 mg/dL); (2) HDL cholesterol < 1.3 mmol/L (50 mg/dL), (3) glucose > 6.1 mmol/L (110 mg/dL) and (4) systolic blood pressure > 90th percentile for gender, age, and height. Blood samples of 5 mL were drawn after 8–12 h of fasting on the same day of the feces collection. Children in the O group were selected so that they did not have more than one trait matching the metabolic syndrome traits. Exclusion criteria for all samples

included recent bodyweight loss > 10%, antibiotic intake 3 months before sample collection, and the occurrence of diarrhea or acute gastrointestinal illness during the same period. The Ethics Committee approved the study of the Instituto Nacional de Medicina Genómica (INMEGEN) in Mexico City, Mexico. The parents or legal guardians of each child signed the informed consent form for participation, and all children assented to participate. Anthropometric parameters (Additional file 1: Table S1), blood pressure, and body mass index were measured following standardized procedures, as previously described.

Sample collection and DNA/RNA extraction

Total bacterial metagenomic DNA for 16S rRNA amplicon sequencing was extracted from 200 mg of feces using the QIAamp® DNA Stool Mini Kit (Qiagen, Inc.; Hilden, Germany) following the manufacturer's protocol. The total RNA extraction was performed with a combination of the ZR Soil/Fecal RNA MicroPrep (Zymo Research; California, USA) and the RNeasy Mini Kit (Qiagen, Inc.; Hilden, Germany), according to manufacturer's protocol. The total RNA quality was assessed with an Agilent 2100 bioanalyzer and quantified with a Qubit 2.0 Fluorometer. Human and bacterial ribosomal RNA was removed using Ribo-Zero Gold rRNA Removal Kit (Illumina; California, USA) following manufacturer instructions.

High-throughput 16S rRNA profiling and RNA-Seq

The V4 hypervariable region was amplified using 515F and 806R primers following the protocol by Caporaso and collaborators [54]. The amplicons were prepared using 100 ng of total DNA, and products were confirmed by agarose gel and purified using Agencourt AMPure XP beads (Beckman Coulter). Fragment size and DNA concentration of each amplicon were determined using an Agilent D1000 ScreenTape for 4200 TapeStation System (Agilent Technologies) and a Qubit 2.0 fluorometer (Invitrogen), respectively. Amplicons were sequenced using an Illumina MiSeq platform at the INMEGEN using reagents for 2×250 paired-end sequencing.

The RNA-seq libraries were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs; Massachusetts, USA). In brief, total RNA previously depleted of rRNA was fragmented at 94 °C for 14 min, followed by first-strand retrotranscription and second cDNA synthesis with random primers using Klenow fragments. Next, we performed the ligation of Illumina compatible indices for multiplex after repairing DNA ends and adding a dA-tail to each strand. Finally, we performed the enrichment PCR amplification with 12 cycles following the manufacturer's instructions. The quality and quantity of the resulting libraries were assessed with a Qubit 2.0 Fluorometer and Agilent 2100 bioanalyzer. All RNAseq libraries were sequenced at the INMEGEN sequencing service using the Illumina NextSeq500 platform for 2×150 paired-end sequencing.

Bioinformatic analysis of the 16S rRNA profiling data

We applied a quality filtering (> Q20 Phred score) and ambiguous nucleotides removal. The resulting reads were joined and analyzed using the QIIME 1.9.1 package. Amplicon sequences were clustered at 97% identity into operational taxonomic units (OTUs) guided by the Greengenes Database (version 13_8) using UCLUST allowing for reverse strand matches following a closed reference-based clustering approach. We assigned the taxonomy to the resulting Operational Taxonomic Unit (henceforth OTUs) based on the one from 97% identity clusters of the Greengenes database. Sequences not aligning the references were not considered for downstream analyses. We eliminated the OTUs accounting for $\leq 0.005\%$ of the total read abundance (80 cumulative reads) from downstream analyses. A valid taxonomy was assigned to the reads, and it was collated and reported in terms of relative abundance. UniFrac distances were calculated with scripts from the QIIME v1.9 suite. Data ordination was carried out from the distance matrices using a principal coordinate analysis (PCoA) with vegan using in-house scripts. Differentially abundant taxa (group-specific) were identified with the LDA Effect Size algorithm (LEfSe Galaxy Version 1.0) using a standardized taxonomy table (mean of 10,000 rarefactions at a depth of the smallest sample) to cope with uneven sample size. No per-sample normalization of the sum of the values to 1 M was used, and a minimum LDA score of 1 was considered. Group pairwise comparisons were carried out as well (not group-specific).

Bioinformatic analysis of RNA-Seq data

The process is briefly described as follows. The first step was the quality assessment by FastQC (Andrews S. (2010). We filtered using a 20 Phred quality score trimming with Trimmomatic v 0.36 with a sliding window of 6 nt, following the removal of sequencing adapters and ambiguous bases. The resulting high-quality sequences were depleted of Prokaryotic and Eukaryotic rRNAs using Ribopicker v.0.4.3 and the SILVA rRNA database (132 release). The remaining non-rRNA sequences were aligned to the human genome and transcriptome (GRCh38_snp_tran) using HISAT2. Human-filtered sequences were then used as the input for the de novo transcriptome assembly with the Trinity metatranscriptomic assembler. After that, the

original reads were aligned to the transcriptome using Bowtie2 v2.3 as part of the Trinity pipeline. The abundance of expression was determined by FPKM for the normalization of sequencing depth and transcripts length, considering only the transcripts with FPKM > 1. TransDecoder (https://github.com/TransDecoder/) was used to identify the longest ORFs candidate of protein coding regions within transcript sequences. The resulted protein sequences were annotated with hmmscan from the HMMER suite v3.1 against the dbCAN to obtain the carbohydrate-active enzymes (CAZY). Protein sequence homology was annotated by BLASTP against the NCBI non-redundant (NR) and functionally mapped to Gene Ontology (GO) terms using Blast2GO. The E-value cut-off was set at 1.0E−3.

Secrebiome definition

To assess the set of secreted proteins of the metatranscriptome, we followed the bioinformatic strategy previously published in Cornejo-Granados et al. 2017. Briefly, the complete set of proteins was analyzed independently with six different feature-based tools to identify the excreted/secreted proteins (ES) by different secretion pathways and removing the ones that had transmembrane domains. All secreted proteins were analyzed with BLASTP against the NCBI's non-redundant (nr) database using Blast2GO with an E-value cut-off set at 1.0E−3 to identify homolog proteins. Additionally, all proteins were functionally mapped to GO terms and annotated by setting the following parameters: E-value-hi-filter: 1.0E−3; Annotation cut-off: 55; GO weight: 5 and Hsp-Hit Coverage cut-off: 0. Finally, we used Blast2GO to identify the over or under-represented GO and EC terms in the ES proteins, by setting the term filter p value to ≤ 0.05.

Differential gene expression analysis

We selected the genes of the Secrebiome with an FPKM > 1 for the differential gene expression analysis. In this manner, a gene is only considered if it is covered by reads in 1 kB long. These cut-offs were more stringent than those used in previous studies. Protein isomeres detected in the secretome were subjected to filtering for reducing data sparsity. The expression measurement was in RNA-Seq by Expectation–Maximization (RSEM), a reference-free transcript quantification method. Briefly, the expression below 1 in any sample was not considered, and isomeres with less than 2 cumulative observations across the samples, or those appearing in less than 3 samples or less than 2 groups were discarded. The RSEM of the resulting core isomeres in the O and OMS groups were standardized with the Differential expression of RNA-Seq (DESeq 2 v. 1.26.0) R package using default parameters, and those that were more abundant in either group were subsequently identified. Isomeres with a p-value below an α of 0.05 were considered as differentially abundant expression markers (given by log2 fold change) and were used for downstream analyses. Identified gene markers associated with either the O or the OMS groups were compared with the NW samples by standardizing the RSEM tables with DESeq using the three groups. Pearson/Spearman's rank correlations between the gene markers and clinical data were calculated with in-house scripts. Cross-referenced GO and EC tables were filtered to retain only proteins identified with a unique identifier. These were collated per sample and the tables were standardized following the same procedures for the protein isomeres. These were then used for the calculation of recruitment plots using vegan with in-house scripts. The sequences corresponding to the differentially abundant transcripts identified with DESeq were used for homology search against NCBI's NT database using the blastn algorithm from the BLAST + suite v 2.10 using with expected value 0.0001, id = 0.80, coverage 75%, Match/Mismatch 2,-3, and Gap/ext 5,2. If trailing hsps had a difference of > 7.5% identity, only the former was considered, and hits with no taxonomic information were ignored. The associated taxonomy was determined with a last common ancestor approach of all results.

Search of CAZymes in the metatranscriptome and Secrebiome

We downloaded the CAZy database from http://bcb.unl.edu/dbCAN2/download/Databases/dbCAN-old@UGA/ containing 921,174 sequences as of September 2017. The detection of Carbohydrate-Active enzymes was performed using HMMER v3.1. Briefly, we prepared the CAZy HMM database with hmmpress using default parameters, and searches were carried out with hmmscan using the predicted proteins of the metatranscriptome and Secrebiome as inputs. Finally, we parsed the results using hmmscan-parser.sh and in-house scripts.

Calculation of the abundance of antigenic regions (AAR)

The AAR is a value used to estimate the antigenic density of a protein, calculating the number of antigenic regions and normalized by the sequence length. We calculated this value using the Secret-AAR webserver for the different protein data sets [51]. Then, we used a Mann–Whitney test ($p < 0.001$) to establish if there was a significant difference between AAR values of the different data sets.

Data accessibility

The sequencing data have been deposited in the NCBI GEO repository and can be consulted under the

accession number GSE143207. Requests for additional material should be made to the corresponding author.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12934-020-01319-y.

Additional file 1: Table S1. Anthropometric and biochemical characteristics of the analyzed population.

Additional file 2: Figure S1. Alpha diversity per sample at max depth; (a) Shannon index and (b) Observed OTUs.

Additional file 3: Figure S2. Top ten most abundant taxa: Per group a) phyla, b) families and c) genera per group; and per sample d) phyla, e) families and f) genera. NW = Normal Weight, O = Obese, and OMS = Obese with Metabolic Syndrome.

Additional file 4: Figure S3. Principal coordinate analyses of Weighted UniFrac distances. Distances are based on taxa abundance. The two linear combinations explaining the most variation are shown as dimensions 1 and 2. Elipses were calculated based on the most distant samples per group. Samples for which RNA-seq information is available are presented with a larger font.

Additional file 5: Table S2. Pearson correlations between clinical/anthropometric parameters and significantly abuindant taxa among NW, O, and OMS.

Additional file 6: Figure S4. Bioanalyzer profile of each sample used for the metatranscriptome.

Additional file 7: Table S3. Basic statistics of the metatranscriptome assembly.

Additional file 8: Figure S5. GO terms and Enzyme commission class distribution of the total ES proteins. Each pay graph shows the different GO and enzyme terms associated with the complete ES proteins encoded in the metatranscriptome for a) Molecular Function, b) Cellular Component, c) Biological Process, and d) Enzyme categories.

Additional file 9: Figure S6. Recruitment of unique functional categories per sample. Samples are colored by NW, O, and OMS groups. a) Recruitment of unique Gene Ontology (GO) features in progressive randomized no-replacement rarefactions based on Hurlbert calculations. b) Recruitment of unique Enzyme Commission numbers (EC) in progressive randomized no-replacement rarefactions based on Hurlbert calculations.

Additional file 10: Figure S7. Per-sample expression of transcripts associated with the O and OMS groups. The total normalized RSEM abundance of each transcript per sample with a higher abundance in yellow. Only differentially abundant transcripts associated with the case groups are shown. DESeq-based standardization of the expression signal was carried out considering all samples.

Additional file 11: Figure S8. Correlations between differentially abundant transcripts and clinical data. Only significant correlations are shown ($\alpha = 0.05$). Pearson correlation, the p-value is shown. A linear regression model is fitted to the data using the transcript as a predictor; the intercept and slope are shown, and SE is presented as a shadow. Transcripts (x-axis) are shown as relative normalized RSEM abundances. The scale of the clinical data is different for each parameter. **a–d)** Transcripts positively correlating with diastolic blood pressure (percentile). **e–f)** Transcripts negatively correlating with glucose levels (mg/dL). **g–h)** Transcripts negatively correlating with LDL levels (mg/dL). **i–l)** Transcripts positively correlating with systolic blood pressure (percentile). **m)** Transcripts positively correlated to triglyceride levels (mg/dL). **n–o)** Transcripts positively correlated to subject weight (g).

Additional file 12: Figure S9. Carbohydrate-active enzyme distribution. Relative abundance of CAZy enzyme families across the secreted and non-secreted proteins in the NW, O, and OMS groups.

Acknowledgements
We thank Juan Manuel Hurtado Ramírez for the bioinformatic technical support. The authors would like to thank to Filiberto Sánchez, Dr. Ricardo Alfredo Grande Cano and Gloria Tanahiry Vázquez Castro at IBT-UNAM for their helpful support and suggestions on molecular biology and sequencing protocols. L.G.B. and F.C.G. acknowledge the support of CONACyT as a Postgraduate fellow. R.G.L. is part of the DGAPA postdoctoral fellowship program CJIC/CTIC/5750/2018. We also thanks to the Unidad de Secuenciación Masiva from INMEGEN for sequencing technical support. There is not a conflict of interest for authors.

Authors' contributions
Conceptualization: LGB, FCG, RGL and AOL; Data curation: LGB, FCG, RGL, AVL, SB, SCQ, BELC, AMV, HN, and AOL; Formal analysis: LGB, FCG, RGL and AOL; Funding acquisition: SCQ, HN and AOL; Investigation and methodology, LGB, FCG, RGL, AVL, SB, SCQ, BELC, AMV, HN, and AOL; Project administration: LGB, FCG, RGL and AOL; Writing original draft: LGB, FCG, RGL and AOL; Writing review & editing: LGB, FCG, RGL, AVL, SB, SCQ, BELC, AMV, HN, and AOL. All authors read and approved the final manuscript.

Funding
This work was supported by the CONACyT Grant SALUD-2014-C01-234188. This research funded by the DGAPA PAPIIT UNAM (IA203118 and IN215520). We also acknowledge the support of program Actividades de Intercambio Académico 2019 CIC-UNAM-CIAD.

Availability of data and materials
The sequencing data have been deposited in the NCBI GEO repository and can be consulted under the accession number GSE143207. Requests for additional material should be made to the corresponding author.

Ethics approval and consent to participate
The Ethic Committee of the National Institute of Genomic Medicine (INMEGEN) in Mexico City approved the study. The parents or guardians of donors signed the informed consent form for participation, and the donors assented to participate.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, C.P. 62210 Cuernavaca, Morelos, Mexico. [2] Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, UNAM/Instituto Nacional de Medicina Genómica (INMEGEN), México City, Mexico. [3] Instituto Nacional de Medicina Genómica, Secretaría de Salud, México City, Mexico. [4] Section for Bioinformatics, Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark.

Received: 23 December 2019 Accepted: 26 February 2020
Published online: 06 March 2020

References

1. Bikel S, Valdez-Lara A, Cornejo-Granados F, Rico K, Canizales-Quinteros S, Soberón X, et al. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. Comput Struct Biotechnol J. 2015;13:390–401.
2. Deschasaux M, Bouter KE, Prodan A, Levin E, Groen AK, Herrema H, et al. Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. Nat Med. 2018;24:1526–31.
3. Abu-Ali GS, Mehta RS, Lloyd-Price J, Mallick H, Branck T, Ivey KL, et al. Metatranscriptome of human faecal microbial communities in a cohort of adult men. Nat Microbiol. 2018;3:1–14.

4. Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, et al. Relating the metatranscriptome and metagenome of the human gut. Proc Natl Acad Sci USA. 2014;111:E2329–38.
5. Ranjan R, Rani A, Finn PW, Perkins DL. Multiomic strategies reveal diversity and important functional aspects of human gut microbiome. BioMed Res Int. 2018;2018:6074918.
6. McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, et al. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci Transl Med. 2011;3:106ra106–6.
7. Maurice CF, Haiser HJ, Turnbaugh PJ. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell. 2013;152:39–50.
8. Nagao-Kitamoto H, Shreiner AB, Gillilland MG, Kitamoto S, Ishii C, Hirayama A, et al. Functional characterization of inflammatory bowel disease-associated gut dysbiosis in gnotobiotic mice. Cell Mol Gastroenterol Hepatol. 2016;2:468–81.
9. Booijink CCGM, Boekhorst J, Zoetendal EG, Smidt H, Kleerebezem M, de Vos WM. Metatranscriptome analysis of the human fecal microbiota reveals subject-specific expression profiles, with genes encoding proteins involved in carbohydrate metabolism being dominantly expressed. Appl Environ Microbiol. 2010;76:5533–40.
10. Gosalbes MJ, Durbán A, Pignatelli M, Abellan JJ, Jiménez-Hernández N, Pérez-Cobas AE, et al. Metatranscriptomic approach to analyze the functional human gut microbiota. PLoS ONE. 2011;6:e17447.
11. Turnbaugh PJ, Quince C, Faith JJ, McHardy AC, Yatsunenko T, Niazi F, et al. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc Natl Acad Sci USA. 2010;107:7503–8.
12. Ranganathan S, Garg G. Secretome: clues into pathogen infection and clinical applications. Genome Med. 2009;1:113.
13. Tjalsma H, Antelmann H, Jongbloed JDH, Braun PG, Darmon E, Dorenbos R, et al. Proteomics of protein secretion by Bacillus subtilis: separating the "secrets" of the secretome. Microbiol Mol Biol Rev. 2004;68:207–33.
14. Gomez S, Adalid-Peralta L, Palafox-Fonseca H, Cantu-Robles VA, Soberón X, Sciutto E, et al. Genome analysis of excretory/secretory proteins in taenia solium reveals their abundance of antigenic regions (AAR). Sci Rep. 2015;5:9683.
15. Cornejo-Granados F, Zatarain-Barrón ZL, Cantu-Robles VA, Mendoza-Vargas A, Molina-Romero C, Sánchez F, et al. Secretome prediction of two *M. tuberculosis* clinical isolates reveals their high antigenic density and potential drug targets. Front Microbiol. 2017;8:128.
16. Patel AK, Singhania RR, Pandey A, Chincholkar SB. Probiotic bile salt hydrolase: current developments and perspectives. Appl Biochem Biotechnol. 2010;162:166–80.
17. Kumar RS, Suresh CG, Brannigan JA, Dodson GG, Gaikwad SM. Bile salt hydrolase, the member of Ntn-hydrolase family: differential modes of structural and functional transitions during denaturation. IUBMB Life. 2007;59:118–25.
18. Travers MA, Sow C, Zirah S, Deregnaucourt C, Chaouch S, Queiroz RML, et al. Deconjugated bile salts produced by extracellular bile-salt hydrolase-like activities from the probiotic *Lactobacillus johnsonii* La1 Inhibit *Giardia duodenalis* in vitro growth. Front Microbiol. 2016;7:1453.
19. Jiang Q, Chen J, Yang C, Yin Y, Yao K. Quorum sensing: a prospective therapeutic target for bacterial diseases. BioMed Res Int. 2019;2019:2015978.
20. Alessandri G, Ossiprandi MC, MacSharry J, van Sinderen D, Ventura M. Bifidobacterial dialogue with its human host and consequent modulation of the immune system. Front Immunol. 2019;10:2348.
21. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Publ Group. 2008;6:121–31.
22. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009;37:D233–8.
23. Koh A, De Vadder F, Kovatcheva-Datchary P, Bäckhed F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. Cell. 2016;165:1332–45.
24. Evia-Viscarra ML, Rodea-Montero ER, Apolinar-Jiménez E, Quintana-Vargas S. Metabolic syndrome and its components among obese (BMI $\geq$ 95th) Mexican adolescents. Endocr Connect. 2013;2:208–15.
25. Moran MA, Satinsky B, Gifford SM, Luo H, Rivers A, Chan L-K, et al. Sizing up metatranscriptomics. ISME J. 2013;7:237–43.
26. Zeng Q, Li D, He Y, Li Y, Yang Z, Zhao X, et al. Discrepant gut microbiota markers for the classification of obesity-related metabolic abnormalities. Sci Rep: US; 2019. p. 1–10.
27. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature. 2009;457:480–4.
28. Liu R, Hong J, Xu X, Feng Q, Zhang D, Gu Y, et al. Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. Nat Med. 2017;23:859–68.
29. Méndez-Salazar EO, Ortiz-López MG, de LÁ Granados-Silvestre M, Palacios-González B, Menjivar M. Altered gut microbiota and compositional changes in firmicutes and proteobacteria in mexican undernourished and obese children. Front Microbiol. 2018;9:2494.
30. Maya-Lucas O, Murugesan S, Nirmalkar K, Alcaraz LD, Hoyo-Vadillo C, Pizano-Zárate ML, et al. The gut microbiome of Mexican children affected by obesity. Anaerobe. 2019;55:11–23.
31. López-Contreras BE, Morán-Ramos S, Villarruel-Vázquez R, Macías-Kauffer L, Villamil-Ramírez H, León-Mimila P, et al. Composition of gut microbiota in obese and normal-weight Mexican school-age children and its association with metabolic traits. Pediatr Obes. 2018;13:381–8.
32. Silva-Boghossian CM, Cesário PC, Leão ATT, Colombo APV. Subgingival microbial profile of obese women with periodontal disease. J Periodontol. 2018;89:186–94.
33. Lahti L, Salonen A, Kekkonen RA, Salojärvi J, Jalanka-Tuovinen J, Palva A, et al. Associations between the human intestinal microbiota, *Lactobacillus rhamnosus* GG and serum lipids indicated by integrated analysis of high-throughput profiling data. PeerJ. 2013;1:e32.
34. Gomez-Arango LF, Barrett HL, Wilkinson SA, Callaway LK, McIntyre HD, Morrison M, et al. Low dietary fiber intake increases Collinsella abundance in the gut microbiota of overweight and obese pregnant women. Gut Microbes. 2018;9:189–201.
35. Clavel T, Desmarchelier C, Haller D, Gérard P, Rohn S, Lepage P, et al. Intestinal microbiota in metabolic diseases: from bacterial community structure and functions to species of pathophysiological relevance. Gut Microbes. 2014;5:544–51.
36. Wang K, Liao M, Zhou N, Bao L, Ma K, Zheng Z, et al. *Parabacteroides distasonis* alleviates obesity and metabolic dysfunctions via production of succinate and secondary bile acids. Cell Reports. 2019;26:222–5.
37. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med. 2009;1:6ra14.
38. Spencer MD, Hamp TJ, Reid RW, Fischer LM, Zeisel SH, Fodor AA. Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. Gastroenterology. 2011;140:976–86.
39. Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y, et al. Human gut microbiota in obesity and after gastric bypass. Proc Natl Acad Sci USA. 2009;106:2365–70.
40. Zhang C, Zhang M, Wang S, Han R, Cao Y, Hua W, et al. Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. ISME J. 2010;4:232–41.
41. Kaakoush NO. Insights into the role of Erysipelotrichaceae in the human host. Front Cell Infect Microbiol. 2015;5:181.
42. Andoh A, Nishida A, Takahashi K, Inatomi O, Imaeda H, Bamba S, et al. Comparison of the gut microbial community between obese and lean peoples using 16S gene sequencing in a Japanese population. J Clin Biochem Nutr. 2016;59:65–70.
43. Kasai C, Sugimoto K, Moritani I, Tanaka J, Oya Y, Inoue H, et al. Comparison of the gut microbiota composition between obese and non-obese individuals in a Japanese population, as analyzed by terminal restriction fragment length polymorphism and next-generation sequencing. BMC Gastroenterol. 2015;15:100–10.
44. Zacarías MF, Collado MC, Gómez-Gallego C, Flinck H, Aittoniemi J, Isolauri E, et al. Pregestational overweight and obesity are associated with differences in gut microbiota composition and systemic inflammation in the third trimester. PLoS ONE. 2018;13:e0200305.

45. He B, Jin S, Cao J, Mi L, Wang J. Metatranscriptomics of the Hu sheep rumen microbiome reveals novel cellulases. Biotechnol Biofuels. 2019;12:153.
46. Bhattacharya T, Ghosh TS, Mande SS. Global profiling of carbohydrate active enzymes in human gut microbiome. PLoS ONE. 2015;10:e0142038.
47. Ferreira-Halder CV, de Sousa Faria AV, Andrade SS. Action and function of *Faecalibacterium prausnitzii* in health and disease. Best Pract Res Clin Gastroenterol. 2017;31:643–8.
48. Martín R, Miquel S, Benevides L, Bridonneau C, Robert V, Hudault S, et al. Functional characterization of novel *Faecalibacterium prausnitzii* strains isolated from healthy volunteers: a step forward in the use of *F. prausnitzii* as a next-generation probiotic. Front Microbiol. 2017;8:1226.
49. Haro C, Garcia-Carpintero S, Alcala-Diaz JF, Gomez-Delgado F, Delgado-Lista J, Perez-Martinez P, et al. The gut microbial community in metabolic syndrome patients is modified by diet. J Nutr Biochem. 2016;27:27–31.
50. Maffei B, Francetic O, Subtil A. Tracking Proteins secreted by bacteria: what's in the toolbox? Front Cell Infect Microbiol. 2017;7:221.
51. Cornejo-Granados F, Hurtado-Ramírez JM, Hernandez-Pando R, Ochoa-Leyva A. Secret-AAR: a web server to assess the antigenic density of proteins and homology search against bacterial and parasite secretome proteins. Genomics. 2018. https://doi.org/10.1016/j.ygeno.2018.10.007.
52. Ignacio A, Fernandes MR, Rodrigues VAA, Groppo FC, Cardoso AL, Avila-Campos MJ, et al. Correlation between body mass index and faecal microbiota from children. Clin Microbiol Infect. 2016;22(258):e1–8.
53. Wexler AG, Goodman AL. An insider's perspective: bacteroides as a window into the microbiome. Nat Microbiol. 2017;2:17026.
54. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA. 2011;108(Suppl 1):4516–22.

Publisher's Note

Microbial Ecology (2024) 87:8
https://doi.org/10.1007/s00248-023-02313-8

REVIEW

Perspectives in Searching Antimicrobial Peptides (AMPs) Produced by the Microbiota

Luigui Gallardo-Becerra[1] · Melany Cervantes-Echeverría[1] · Fernanda Cornejo-Granados[1] · Luis E. Vazquez-Morado[1] · Adrian Ochoa-Leyva[1]

Received: 26 May 2023 / Accepted: 24 October 2023

Abstract
Changes in the structure and function of the microbiota are associated with various human diseases. These microbial changes can be mediated by antimicrobial peptides (AMPs), small peptides produced by the host and their microbiota, which play a crucial role in host-bacteria co-evolution. Thus, by studying AMPs produced by the microbiota (microbial AMPs), we can better understand the interactions between host and bacteria in microbiome homeostasis. Additionally, microbial AMPs are a new source of compounds against pathogenic and multi-resistant bacteria. Further, the growing accessibility to metagenomic and metatranscriptomic datasets presents an opportunity to discover new microbial AMPs. This review examines the structural properties of microbiota-derived AMPs, their molecular action mechanisms, genomic organization, and strategies for their identification in any microbiome data as well as experimental testing. Overall, we provided a comprehensive overview of this important topic from the microbial perspective.

Keywords Antimicrobial peptides (AMPs) · Microbiota · Microbiome

Introduction

The role of the microbiota is crucial for maintaining good health. It helps to develop the host's physiology, protects against harmful pathogens, and regulates metabolic processes [1]. Moreover, the microbiota also produces metabolites that can affect the host's homeostasis [2]. For instance, short-chain fatty acids (SCFA) produced by some bacterial species play a vital role in cross-communication, mucus barrier [3], gut motility [4], blood pressure regulation [5], bile acids deconjugation [4, 5], amino acid production [6], and vitamin synthesis [7]. Thus, understanding how the microbiome is self-regulated and modulated opens new possibilities for microbiome-based therapies via microbiome engineering [8].

The proteins excreted and secreted by the microbiota, also known as the secrebiome [9], play a crucial role in the communication between the microbiota and their host. These proteins include enzymes, toxins, and antimicrobial peptides (AMPs) [10, 11]. AMPs are an ancestral and effective primary defense mechanism against pathogens such as bacteria, archaea, fungi, and viruses [11]. They do not have enzymatic activities and can act in a monomer or polymer conformation [10, 11]. Also, these peptides autoregulate bacteria, conducting communication with each other through quorum sensing [12], as well as with eukaryotic host cells [13], and regulate virulence systems [14]. In this regard, the microbiota and their host can produce AMPs for microbiota-microbiota or microbiota-host interactions. This review will focus specifically on microbiota-derived AMPs, their general properties, molecular action mechanisms, and strategies for their identification in any microbiome data set, and experimental validation.

General Properties of AMPs

It is well-known that AMPs have low molecular weight and minimal secondary structure compared to regular proteins [15]. These molecules are typically cationic and amphipathic, containing both hydrophobic and hydrophilic regions.

✉ Adrian Ochoa-Leyva
adrian.ochoa@ibt.unam.mx

[1] Departamento de Microbiologia Molecular, Instituto de Biotecnologia, Universidad Nacional Autonoma de Mexico (UNAM), Avenida Universidad 2001, C.P. 62210 Cuernavaca, Morelos, Mexico

Published online: 01 December 2023

They can adopt various conformations, such as α-helical, ß-sheet, or extended (without a specific structural motif). Most AMPs reported in publicly available databases are short peptides, with a typical length ranging from 5 to 50 amino acids [16]. However, some peptides are longer with over 200 aa [17, 18]. To better understand these characteristics, we analyzed the peptide length distributions in three AMP databases: APD3, dbAMP, and DRAMP, which contain both predicted and experimentally tested AMPs. Our analysis found that all AMPs in these databases had a mean of 33–40 amino acids (Supplementary Fig. 1). On the other hand, when we focused only on microbiota-derived AMPs, we observed these peptides had a mean length from 35 to 57 amino acids (Fig. 1).

Based on their structure, microbial AMPs can be categorized into three groups (Fig. 2) [19]. Class I, with peptides of less than 10 kDa with a similar structure to Microcins (Fig. 2A). These peptides are composed of two alpha helices, such as Glycosin F, two beta sheets as in Microcin J25 (Fig. 2A), or a combination of both, which fold the peptide into a horseshoe shape, and at the N-terminal or C-terminal ends, peptide segments can be found without a given conformation, as seen in Ruminococcin C (Fig. 2A)[19–21]. Class II includes peptides larger than 10 kDa and structures like Pediocins (Fig. 2B). These peptides are made up solely of alpha helices that occupy most of the sequence. Lacticin Q and Plantaricin J are among its members (Fig. 2B) [22–25]. Finally, Class III comprises peptides about 30 kDa in size, such as Bacteriolysins (Fig. 2C). These peptides exhibit a structural complexity comparable to native proteins and can be conjugated with other ions (Fig. 2C) [19, 26–28].

Microbial AMPs' Molecular Action Mechanisms

Regarding their biological mechanisms, microbial AMPs can be classified into three groups (Fig. 3). The first includes AMPs that play a regulatory role in the host's immune system. They achieve this by activating and recruiting immunocytes or altering the Toll-like receptor (TLR) recognition of microorganisms (Fig. 3A) [16]. The second group includes AMPs that interact with the membrane or cell wall of the target microorganism, causing lysis (Fig. 3B). The selectivity of these AMPs depends on specific differences in the composition of the membrane or cell wall [11, 16] and can be further classified into barrel-slave [29], carpet-like [30], and toroidal pores [31]. Finally, the third group consists of AMPs that inhibit essential intracellular functions, such as DNA replication (Fig. 3C) [16, 32]. It is worth noting that all three mechanisms are shared between AMPs produced by bacteria, fungi, and protozoans [33, 34].

Microbial AMPs can have additional benefits as antiviral peptides, inhibiting the virus viability [35]. These peptides are effective against DNA and RNA viruses (Fig. 4) and work by disrupting the viral membrane to prevent host cell infection (Fig. 4A) [35]. Other mechanisms include inhibiting the host viral receptor (Fig. 4B) [36], as well as targeting specific intracellular molecules required for viral replication (Fig. 4C) [37]. These findings represent potential alternatives for developing new antiviral treatments based on AMPs.

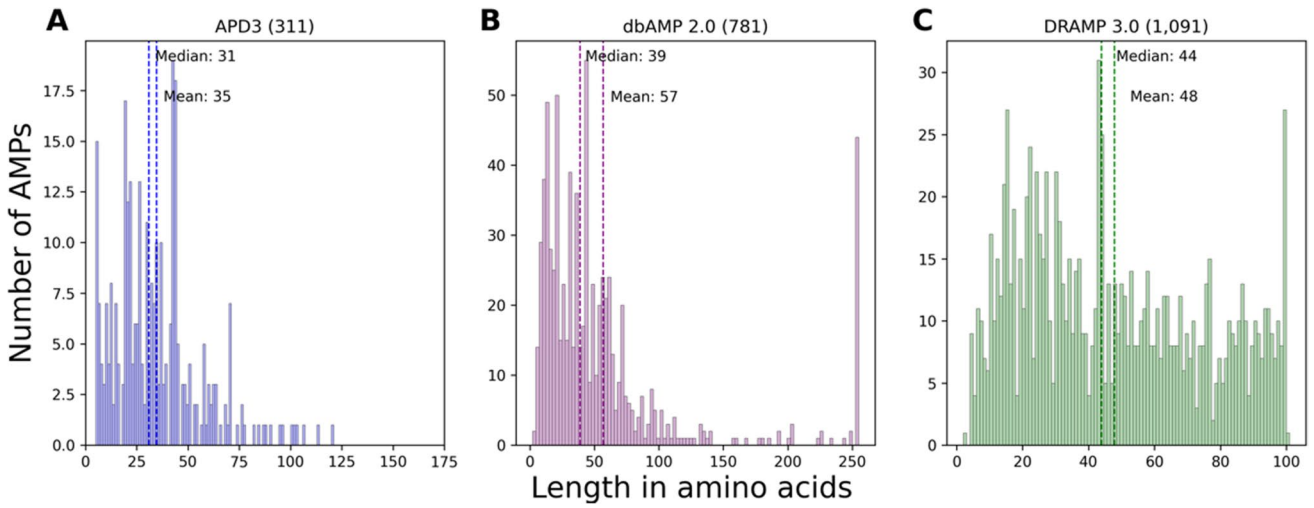


Fig. 1 Length distribution of microbial AMPs deposited in databases. **A** APD3, an experimentally validated database with 311 AMPs produced by prokaryotes, had a median of 31 aa, and a mean of 35 aa; **B** dbAMP 2.0, a collection of experimentally validated and hypothetical AMPs, had 781 peptides produced by microorganisms with a median of 39 aa, and a mean of 57 aa; **C** DRAMP 3.0, also a collection of validated and putative AMPs, contained 1091 peptides from microbial origin, with a median of 44 aa, and a mean of 48 aa

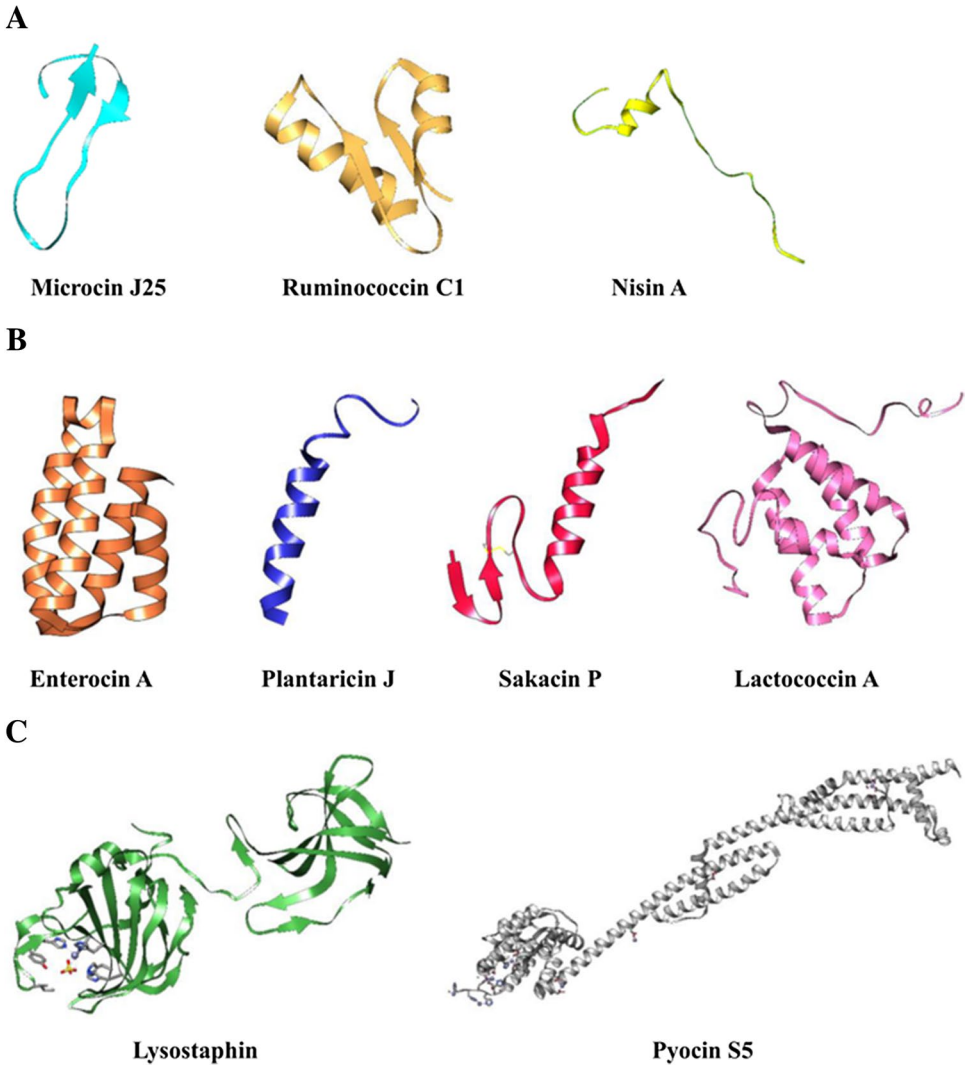


Fig. 2 Three-dimensional structure of the three classes of AMPs. **A** Class I AMPs, which are bacteriocins-like peptides include Microcin J25 (1Q71), Ruminococcin C1 (6T33), and Nisin A (1WCO). **B** Class II AMPs include pediocin-like peptides such as Enterocin A (2BL7), Plantaricin J (2KHG), Sakacin P (1OG7), and Lactococcin A (5LFI). **C** Class III AMPs have bacteriolysin activity, for example, Lysostaphin (4LXC) and Pyocin S5 (6THK). Each class has unique characteristics and plays a crucial role in fighting against harmful microbes. The Protein Data Bank (PDB) identifier of 3D structures is mentioned in parenthesis

Classic Production and Genomic Organization of Microbial AMPs

AMPs are produced through the activation of genes responsible for controlling AMP synthesis. These genes are typically stimulated by infectious or inflammatory processes [38] and are commonly organized in a single or several operons. These operons include one or more structural genes encoding a functional peptide or their inactive precursor and genes for AMP regulation, maturation, export, and self-immunity, typically adjacent in the cluster arrangement [38]. The genetic organization varies among bacterial AMPs. For instance, the cluster of the microcin MccJ25 gene of *Escherichia coli* is partially conserved, including at least one precursor, self-immunity against the AMP, and export genes (Fig. 5A).

On the other hand, enterocin AS-48 produced by *Enterococcus faecalis* may require at least five to seven genes for its production and autoimmunity (Fig. 5B) as well as an accessory operon encoding the ABC transporter protein complex [39]. Complex AMPs, such as lantibiotics produced by Gram-positive bacteria, can undergo post-transcriptional modifications before being secreted. Usually, clusters carry the genes to produce the enzymes responsible for post-transcriptional modification, while other AMPs are translated and exported without modifications [40]. However, there are clusters for some AMPs, such as aureocin A53 of *Staphylococcus aureus*, that do not include genes for post-translational processing (Fig. 5C) [41]. Furthermore, large AMP operons often carry one or more genes of unknown or uncharacterized function [39]. The activation of secretion system genes, typically ABC transporters, is responsible for AMP export from the cytosol to the extracellular environment, completing the production of AMPs [42].

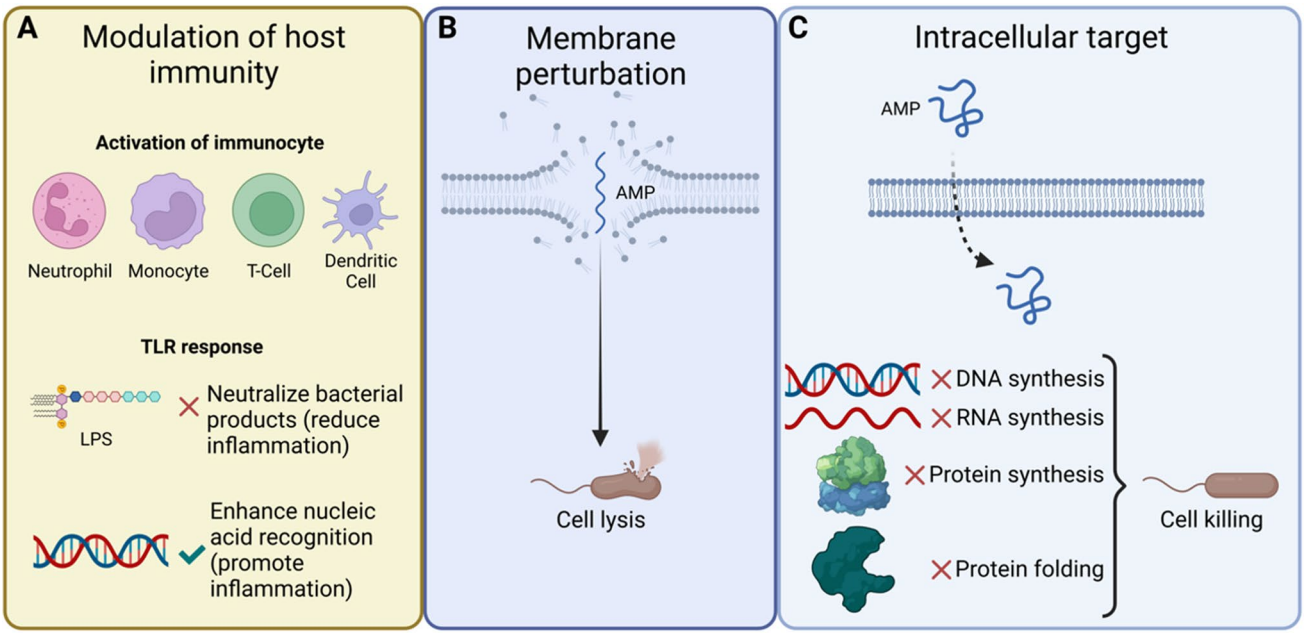


Fig. 3 Microbial AMPs action mechanisms. **A** AMPs significantly impact host immunity by activating the immune cell response or toll-like receptors (TLR). This activity is essential because it can neutralize bacterial products such as lipopolysaccharides (LPS), thus reducing inflammation or enhancing microbial nucleic acid recognition, thereby increasing inflammation. Furthermore, AMPs can directly cause the death of the target microorganisms, either **B** by perturbing its membrane, leading to cell lysis, or **C** by inhibiting vital intracellular functions. These makes AMPs a potent weapon in the fight against harmful bacteria. The image was created using BioRender.com.

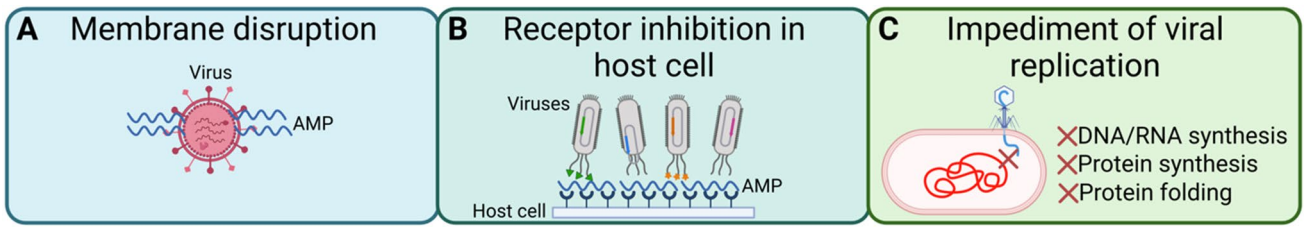


Fig. 4 Action mechanisms of antiviral peptides. Antiviral peptides work by inhibiting the viability of the virus, preventing the infection of target cells, or impeding the virus's ability to replicate. There are several mechanisms by which these peptides work including **A** the direct disruption of the viral membrane, making the virus unable to infect, **B** blocking the viral receptor binding to the host cells, thus preventing the virus from entering the target cell **B**, and **C** blocking intracellular functions necessary for viral replication **C**. Image created using BioRender.com

Microbial AMPs in the Microbiome Regulation

All higher organisms have a close relationship with the microbiota inhabiting them. AMPs, which both the host and the microbiota produce, are essential for crosstalk communication and maintaining the homeostasis of the microbiome. Additionally, AMPs are part of the first line of the host defense by inhibiting the proliferation of potentially harmful pathogens [43, 44]. On the other hand, microorganisms use them to take advantage of an environmental niche by manipulating the host or competing with other microorganisms within the microbiota [16, 44]. AMPs can also regulate species-specific associations and can influence bacterial colonization [45].

The human gut is a fascinating and well-studied example of how microbes interact with their host, and one area that has received much attention is the interaction with microbial AMPs (Fig. 6). Although there is scarce information about the specific bacteria that produce AMPs, we do have some clues. Some of the most common producers of AMPs in the human gut microbiome are *Bacillus* and *Lactobacillus*, transient bacteria colonizing the epithelium [46]. These microbes are known for producing bacteriocins and lipopeptide antibiotics that suppress the growth of potential

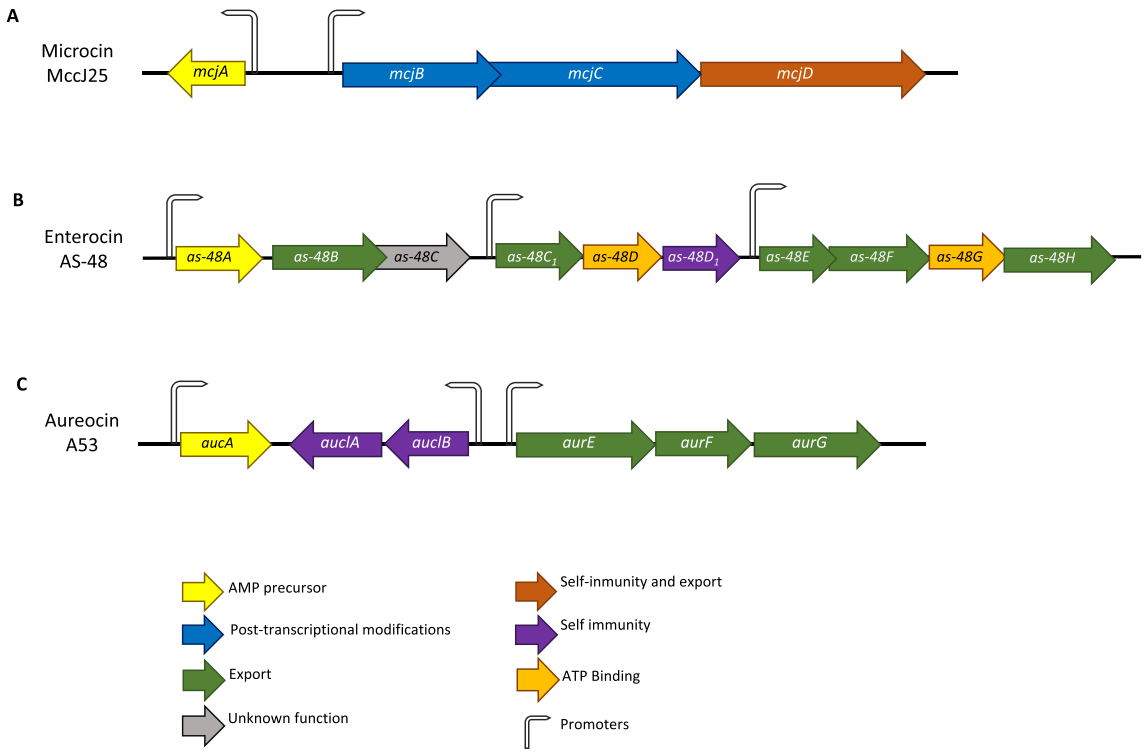


Fig. 5 Examples of the genomic organization of microbial AMPs gene clusters. Arrows indicate the genes colored based on their known or putative functions. A flag indicates promoters and the direction refers to gene transcription. **A** shows the genes required for microcin MccJ25; **B** shows the genes required for Enterocin AS-48, and **C** exhibits genes required for Aureocin A53

pathogens by affecting membrane permeabilization [47–49]. Also, administering probiotics that produce AMPs, such as members of *Lactobacillus* and *Enterococcus*, has improved antimicrobial activity in the intestinal lumen [50]. *Butyrivibrio* is another example of a microbe that produces AMPs and is found in high abundance in the intestine of mice after exercise-induced stress response [51]. Other types of peptides with antimicrobial activities include ribosomally synthesized and post-translationally modified peptides (RiPPs), such as lanthipeptides produced by Firmicutes and Actinobacteria, and sactipeptides, primarily characterized in *Bacillus* species, thiopeptides, reported in *Lactobacillus gasseri*, *Cutibacterium acnes*, *Enterococcus faecalis*, *Streptococcus downei*, and, *S. sobrinus* [2]. It is also worth noting that host intestinal epithelial cells produce AMPs, particularly bacteriocins, lanthipeptides, and sactipeptides, that control the overgrowth of unwanted bacteria in the inner mucus layer [52]. Conversely, the microbiota produces AMPs to compete for gut establishment [16].

The microbiota's mobile genetic elements, including bacteriophages and plasmids are carriers of AMPs enhancing the carrier microorganisms' fitness [53–55]. A significant fraction of AMPs, such as microcins, are usually transported by conjugative plasmids, allowing for their exchange between bacteria. A well-described example is the microcins MccB17 produced by certain strains of *Escherichia coli*, which carry the 70 kb conjugative plasmid pRYC17 with gene clusters to produce precursors, post-transcriptional enzymes, secretion, and autoimmunity genes [56]. Some lytic phages have cell wall hydrolytic AMPs, such as lysins and lysozymes, which form a hole in peptidoglycan structure and release replicated viral particles. Purified phage endolysins have been applied against Gram-positive pathogens such as Streptococcus pyogenes as potential antimicrobial agents [57, 58].

The microbiota is a valuable source of compounds for the industry and a promising source of novel biomolecules like AMPs [14]. However, most microorganisms cannot be cultured, making DNA and RNA sequencing viable alternatives for discovering AMPs in microbial communities. This method also allows us to study AMPs produced by the microbiota without the need to culture the original bacteria. Unfortunately, few bioinformatics protocols are available to identify AMPs in DNA or RNA sequencing datasets. In the next section of this review, we explore the current state of bioinformatic tools for discovering AMPs from the microbiota. Also, we summarize some of the most successful experimental strategies for the functional analysis of AMPs.

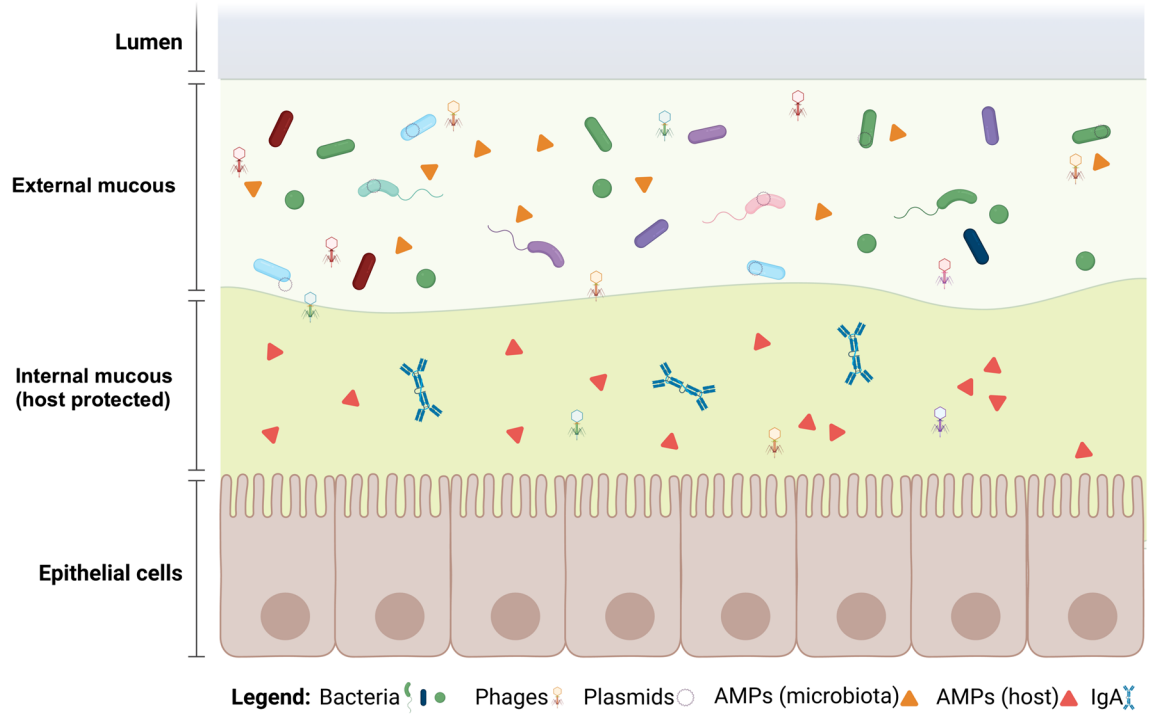


Fig. 6 Illustration of the critical role played by AMPs in the gut microbiome dynamics. AMPs play an essential role in the ecological dynamics of the gut microbiome. The host's AMPs secreted in the gut ensure that the internal mucosa remains uncolonized by potentially harmful bacteria. At the same time, the AMPs produced by the microbiota compete with each other and regulate the host's response to them. Moreover, AMPs can be carried by mobile genetic elements such as bacteriophages and plasmids, which can confer an advantage to their hosts in the microbiome dynamics. The image was created using BioRender.com

Genomic Sciences Applied to Microbiomes to Discover AMPs

The discovery of new microbiome functionalities has increased with the advances in genomic sciences, including metagenomics, metatranscriptomics, viromics, and plasmidomics [8, 59]. Although the methodologies for nucleic acid extraction, library preparation, and subsequent sequencing are beyond the scope of this review, we briefly describe them. For metagenomics, total DNA is extracted, fragmented, and amplified to create the sequencing libraries (Fig. 6). In contrast, metatranscriptomic libraries require total RNA extraction and enrichment of the molecule of interest (mRNA, lincRNA, microRNA, etc.) and subsequently cDNA synthesis following fragmentation and adapter attachment (Fig. 7) [8]. On the other hand, to study the virome, it is necessary to isolate the viral particles (VLPs). To this end, there are several protocols that use particle-selecting filtration or ultracentrifugation with cesium chloride, followed by DNA or RNA extraction of the enriched VLPs [60]. Finally, to analyze the plasmids of the microbiome, it is necessary the depletion of host bacterial DNA before preparing the sequencing libraries. This can be achieved using exonucleases that degrade linear DNA, leaving the plasmid circular DNA intact [60], then the procedure follows the typical sequencing library protocol (Fig. 7).

After conducting the experimental phase, the next step is to analyze the sequencing data using bioinformatics. When writing this review, only a few articles discussed microbial AMPs obtained from the microbiome (Supplementary File 1). For instance, one study discovered five AMPs produced by the gut microbiota of *Ctenopharyngodon idellus* [61]. Another study found new AMPs derived from the *Hirudo medicinalis* microbiome, identifying a new peptide (pept_1545) that could be used for therapeutic purposes due to its widespread antimicrobial activity and lack of toxicity on eukaryotic cells [62].

A method called Metagenomic AMP Classification and Retrieval (Macrel) found 1263 non-redundant AMPs from 182 human gut metagenomes [63]. More recently, a study used metagenomics data and deep learning to identify AMPs from the human gut microbiome, resulting in the discovery of 2389 AMPs, 181 of which were experimentally proven to have antimicrobial activity [64]. Lastly, using a metagenomics approach, two AMPs (HG2 and HG4) were reported from the rumen microbiome. These peptides showed activity against multidrug-resistant bacteria, making them

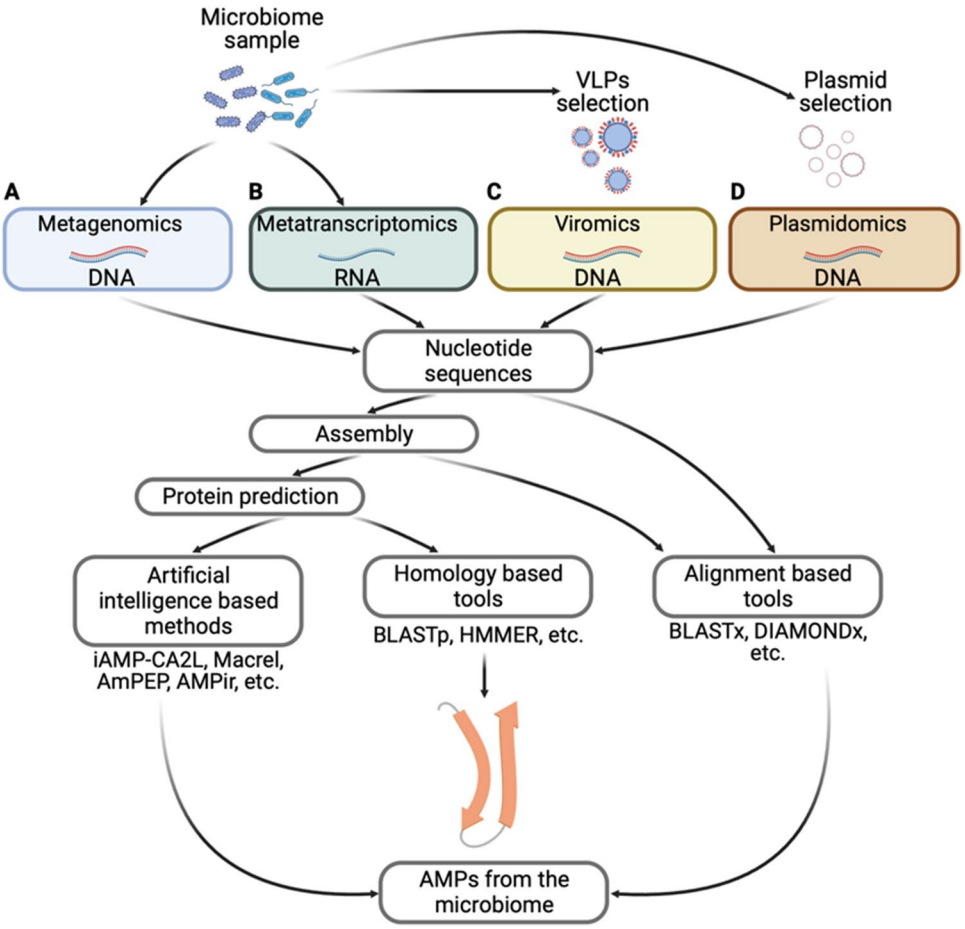


Fig. 7 Experimental and bioinformatic strategies to obtain AMPs from microbiome data. The search for AMPs can be done in four kinds of datasets **A** metagenomics, which involves obtaining all potentially functional AMPs, **B** metatranscriptomics focuses on the expressed AMPs, **C** viromics identifies AMPs encoded in viruses, and lastly, **D** plasmidomics is focused on obtaining AMPs codified in plasmids. Image created using BioRender.com

potentially useful as templates for the treatment of bacterial infections [65].

Metatranscriptomics data led Huang et al. to find microbial AMPs in Taiwanese oolong teas, partially fermented beverages that may impact the microbial communities of the consumer [66]. Another study by Onime et al. found 209 potentially novel AMPs in the rumen of eukaryotic microorganisms using metatranscriptomics. One of these, Lubelisin, was active against methicillin-resistant *Staphylococcus aureus* and maintained low cytotoxicity for humans and sheep [67]. However, as of writing this review, there were no reports of AMPs discovery from viromics or plasmidomics datasets.

Databases, Web Servers, and Bioinformatics Tools to Discover Microbial AMPs

With the discovery of more AMPs, several research groups have developed publicly available databases (Supplementary File 2). A few examples are YADAMP which contains manually curated AMPs that are effective against bacteria [68]; BACTIBASE, which comprises bacteriocins obtained from bacteria [69]; AntiTbPdb, which has experimentally validated AMPs against tubercular or mycobacterial species [70]; ParaPep, containing Anti-parasitical peptides [71]; AVPdb, with curated Antiviral peptides [72]; InverPep, which includes AMPs produced by invertebrates [73]; and PhytAMP, which only has Plant-derived AMPs [74]. Some databases cater to a particular class of AMPs, like Peptaibol [75] and the Defensins Knowledgebase [76]. Finally, more extensive databases such as ADAM [77] or APD3 [78] offer experimentally validated or manually curated AMPs. Other specialized databases provide detailed information on secondary structures, such as DBAASP [79] or CAMPR3 [80]. Others collect data and remove redundancy while unifying classifications facilitate users to find the AMPs they need, such as dbAMP 2.0 [17], LAMP2 [81], and DRAMP 3.0 [18].

Complementing databases, web servers can help predict if a protein sequence has the potential to be an AMP using various algorithms (Supplementary File 3). However, these web servers have limited capacity for uploading and downloading. Therefore, the massive search for AMPs needs to be optimized data. Furthermore, there are tools available to identify AMPs in a local computer, mostly using

a sequence alignment strategy against a protein database, such as BLASTP [82]. Pattern-matching algorithms like Hidden Markov Models (HMM) can also detect remote protein homologs without requiring sequence homology [83]. Although alignment-based methods allow for identifying potential AMPs already reported in databases, they make it challenging to discover new peptides since they depend on already known data. In addition to the sequence alignment methods, other tools consider factors like physic-chemical properties, amino acid composition, and secondary structure to increase the accuracy of the AMPs prediction, such as AMAP [84], AMPir [85], CAMPSign [86], iAMP-2L [87], iAMPred [88], and Macrel [63]. Furthermore, other tools based on artificial intelligence, machine learning, and neural network algorithms can be used to discover AMPs without sequence identity with known AMPs, such as AmPEP [89], amPEPpy 1.0 [90], AMPir [85], APSv2 [91], c_AMP-prediction [64], ClassAMP [92], AMPlify [93], and iAMP-CA2L [94] (Supplementary File 3). These methods allow for finding novel AMPs in genomic data but with the risk of having a higher number of false positives. However, there is no set rule for discovering AMPs from the microbiota, and articles reporting search strategies are scarce. Figure 7 provides a summary of the pipelines for discovering AMPs.

How to Experimentally Test the AMP Function

To analyze the potential function of an AMP, first, the peptide needs to be obtained by direct extraction from the host, chemical synthesis, or heterologous peptide expression and purification from a bacterial system [95]. After acquiring the peptide, the experimental analysis can be divided into three steps [96]. The first step involves testing the peptide in antimicrobial assays, followed by an in vitro cytotoxic assay, and finally, an in vivo test in case of being used for clinical application [97]. This process is illustrated in Fig. 8.

There are various methods available to test the antimicrobial activity of a peptide. The most common ones are broth microdilution, agar diffusion, agar dilution, and the Kirby-Bauer method. In the Kirby-Bauer method, a filter paper disc is impregnated with the AMP and placed on an agar plate inoculated with the target bacteria [95, 97]. Once the plate has been incubated, the diameter of the inhibition zone is measured. Another helpful test is the minimum inhibitory concentration (MIC) assay, which measures the lowest concentration of a peptide required to inhibit the growth of the target microorganism [97]. In addition to these methods, it is also possible to test the efficacy of an antimicrobial peptide against viruses and fungi and its ability to inhibit cancer cell growth [98, 99].

In the in vitro cytotoxic assays, the process involves incubating the AMP with a suspension of mammalian cells and observing the changes in the cells' morphology and viability [52]. Live cells are then stained with a specific dye, and the number is compared to those in a control sample that has not been exposed to the peptide [52]. If a lower percentage of cells is stained in the AMP sample, it indicates a higher level of cytotoxicity [52, 100]. The most common method used to determine the cell viability is the MTT assay, which measures the reduction of a yellow MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) solution to purple formazan by living cells [52]. Another method is the lactate dehydrogenase assay, which measures the release of LDH from damaged cells [52].

When using AMP as a therapy against a pathogen in the cosmetic or pharmaceutical industries, they must pass an in vivo test [101]. The AMP delivery system to kill the pathogen can be topical, intravenous, intraperitoneal,

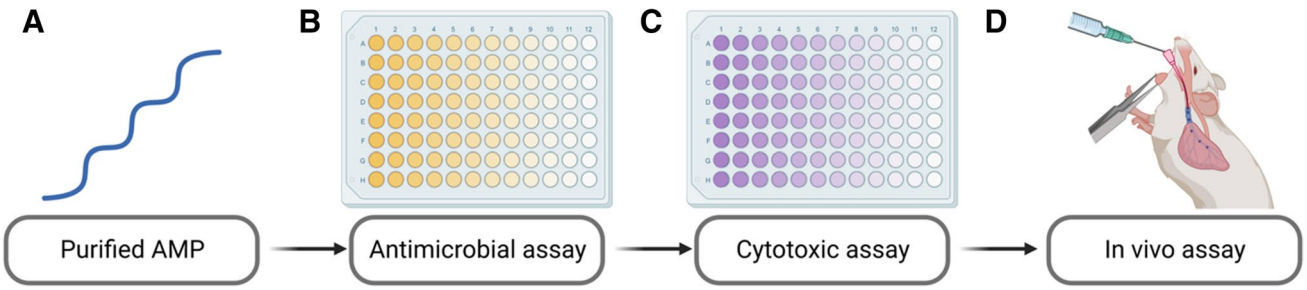


Fig. 8 Typical steps for testing the AMP effect for clinical usage. Firstly, peptide purification is necessary, which can be achieved through natural, chemical, or biological synthesis (**A**). Secondly, selecting an appropriate antimicrobial assay is crucial in determining the efficacy of the AMP against the target microorganism (**B**). This involves incubating the AMP and the microorganism together and observing the effects. Additionally, it may be necessary to conduct cytotoxic assays to determine any potential damage to host cells (**C**). Lastly, the AMP can be tested against the target microorganism using an in vivo model (**C**). This can be achieved through various application methods, including topical, intravenous, intraperitoneal, subcutaneous, oral, intranasal, or inhaled. These steps are vital in ensuring the safety and efficacy of AMPs for clinical usage and should be carefully followed to achieve the best possible results. Image created using BioRender.com

subcutaneous, oral, intranasal, or inhaled [102]. It is crucial to monitor the animal's response to the peptide during the assay and adjust the dose as needed [103].

Future and Perspective of AMPs Derived from Microbiomes

The future of AMPs usage is promising [104], with a wide range of applications that include antibacterial, antifungal, antiviral, and antiparasitic effects [105]. Moreover, researchers are currently exploring their potential use in cancer therapy [106] and as immunomodulators [107]. AMPs are also a safe and well-tolerated natural alternative with few side effects to traditional antibiotics [74], which are becoming less effective due to drug-resistant bacteria [105]. Overall, AMPs represent a hopeful solution to combat infections and diseases with minimal side effects.

The discovery of AMPs traditionally involves screening peptide libraries from the organism of interest to test its antimicrobial activity. This method is largely based on trial-and-error experiments. However, newer approaches use computational methods to predict peptides with antimicrobial activities based on the organism's proteomic or genomic data. The challenge with this approach is selecting the most suitable organism. Fortunately, the microbiota presents a vast reservoir of undiscovered AMPs that could be clinically and industrially valuable, thanks to its large amount of genomic information, long-term co-evolution with the host, and competition between neighboring bacteria [108]. Therefore, integrating experimental and bioinformatics tools focused on discovering AMPs from metagenomes, metatranscriptomes, viromes, and plasmidomes datasets will be of great value.

The AMPs play a crucial role in limiting the growth of unwanted microbiota, particularly pathogens, and shaping the overall microbiome composition [109]. Understanding AMPs could lead to the development of new therapies that can help regulate the microbiome. These can be used to target diseases caused by microbiota dysbiosis, including skin infections [110], eye diseases [111], gastrointestinal diseases [112], bone and joint infections [113], oral diseases [114], and respiratory diseases [115]. Administering the AMPs using mobile elements as carriers, such as bacteriophages or plasmids, could be a great alternative to combat unwanted microorganisms, like multi-resistant pathogens, without promoting antibiotic resistance [116]. Microbiome-derived AMPs have numerous applications in various industries. They can be used as preservatives to control food-borne pathogens [117]. In agriculture, they can act as growth promoters and control plant diseases [118]. In healthcare, they can treat infections as antiseptics, disinfectants, and drugs [119]. They can control microorganism overgrowth in cosmetics and industrial applications [120, 121].

Besides naturally produced AMPs, the design of new peptides with enhanced antimicrobial activity is an active area of research [122, 123]. This includes improving the antimicrobial activity by modifying the peptide sequence and their cationic, hydrophobic, and amphipathic properties [124], where bioinformatic tools and machine learning or deep algorithms play a crucial role in improving antimicrobial peptides AMPs by aiding in their design, prediction, and analysis [125, 126]. Some of these tools include HydrAMP [127], PepGAN [128], AMPAGAN v2 [129], PepCVAE [130], PandoraGan [131], among others [123].

Challenges are still associated with using AMPs as a solution to antimicrobial resistance. These include issues related to their stability, bioavailability, and production cost [132]. Additionally, further research is needed to determine the optimal dosing and delivery strategies to maximize AMP's effectiveness and minimize the risk of side effects [133].

Conclusions

The future of AMP research is promising. There are many undiscovered AMPs produced by the microbiota, that are not harmful to the host, presenting a fantastic research opportunity. There is still much to learn about these AMPs, from their discovery and characterization to understanding how they work. Analyzing the omic data from diverse microbiomes and creating new tools and methods for AMP discovery is essential. With all these efforts, the field of AMPs research will make great strides in the coming years.

In light of the growing concern over antimicrobial resistance, AMPs are an encouraging solution against bacteria resistant to multiple antibiotics and diseases linked to microbiota dysbiosis. However, the rapid degradation of peptides in the body often limits their therapeutic potential. Thus, researchers must develop new techniques to enhance AMP delivery and stability. Therefore, it is critical to use AMPs judiciously and deliberately when deploying them on a large scale to address this issue.

Despite these challenges, ongoing research endeavors will be able to confront these obstacles and refine the utilization of AMPs across several industries. Some strategies under exploration include combining AMPs with conventional antibiotics and bacteriophages, developing advanced delivery systems, and designing AMPs with enhanced properties. These efforts aim to unlock the full potential of AMPs for improved functionality.

It is interesting to consider the role of AMPs in host-microorganism interactions. We often think of them as only being produced by the host to fight off harmful bacteria. Nevertheless, it is essential to remember that bacteria can also produce AMPs to defend against the host and

compete with other bacteria for resources and survival in the ecological niche. It is a complex dynamic that highlights the intricacies of the microbial world.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00248-023-02313-8.

Acknowledgements L.G.B., M.C.E., and L.V.M. thank the Doctoral Biochemical Sciences Program at IBt UNAM and CONACyT by doctoral fellowships C.V.U.: 778192, 887285, and 860364, respectively. Also, F.C.G. would like to thank the Estancias posdoctorales por México 2022 program (C.V.U.: 443238).

Author Contribution Conceptualization: LGB, MCE, LVM, FCG, and AOL; Data curation: LGB, MCE, LVM, FCG, and AOL; Formal analysis: LGB, MCE, LVM, FCG, and AOL; Funding acquisition: FCG, AOL; Investigation and methodology, LGB, MCE, LVM, FCG, and AOL; Writing original draft: LGB, MCE, LVM, FCG, and AOL; Writing review & editing: LGB, MCE, LVM, FCG, and AOL.

Funding This research was funded by CONACyT grant Ciencia de Frontera-2019–263986 and by DGAPA PAPIIT UNAM (IN219723). L.G.B., M.C.E., and L.V.M. were supported by the Doctoral Biochemical Sciences Program at IBt UNAM and CONACyT with the doctoral fellowships C.V.U.: 778192, 887285, and 860364, respectively. F.C.G. was supported by the Estancias posdoctorales por México 2022 program (C.V.U.: 443238).

Data Availability All material relevant to this publication is available in the manuscript and its supplementary information files.

Code Availability Not applicable.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare no competing interests.

References

1. Hooper LV, Gordon JI (2001) Commensal host-bacterial relationships in the gut. Science 292(5519):1115–1118. https://doi.org/10.1126/science.1058709
2. Fobofou SA, Savidge T (2022) Microbial metabolites: cause or consequence in gastrointestinal disease? Am J Physiol-Gastrointest Liver Physiol 322(6):G535–G552. https://doi.org/10.1152/ajpgi.00008.2022
3. Deleu S, Machiels K, Raes J, Verbeke K, Vermeire S (2021) Short chain fatty acids and its producing organisms: an overlooked therapy for IBD? EBioMedicine 66:103293. https://doi.org/10.1016/j.ebiom.2021.103293
4. Krautkramer KA, Fan J, Bäckhed F (2021) Gut microbial metabolites as multi-kingdom intermediates. Nat. Rev. Microbiol. 19(2):2. https://doi.org/10.1038/s41579-020-0438-4
5. Iacob S, Iacob DG and Luminos LM (2019) Intestinal microbiota as a host defense mechanism to infectious threats. Front Microbiol 9:3328. https://doi.org/10.3389/fmicb.2018.03328
6. Collins SM, Surette M, Bercik P (2012) The interplay between the intestinal microbiota and the brain. Nat Rev Microbiol 10(11):11. https://doi.org/10.1038/nrmicro2876
7. Pham VT, Dold S, Rehman A, Bird JK, Steinert RE (2021) Vitamins, the gut microbiome and gastrointestinal health in humans. Nutr Res 95:35–53. https://doi.org/10.1016/j.nutres.2021.09.001
8. Bikel S et al (2015) Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. Comput Struct Biotechnol J 13:390–401. https://doi.org/10.1016/j.csbj.2015.06.001
9. Gallardo-Becerra L et al (2020) Metatranscriptomic analysis to define the Secrebiome, and 16S rRNA profiling of the gut microbiome in obesity and metabolic syndrome of Mexican children. Microb Cell Factories 19(1):61. https://doi.org/10.1186/s12934-020-01319-y
10. Ageitos JM, Sánchez-Pérez A, Calo-Mata P, Villa TG (2017) Antimicrobial peptides (AMPs): ancient compounds that represent novel weapons in the fight against bacteria. Biochem Pharmacol 133:117–138. https://doi.org/10.1016/j.bcp.2016.09.018
11. Epand RM, Vogel HJ (1999) Diversity of antimicrobial peptides and their mechanisms of action. Biochim Biophys Acta BBA - Biomembr 1462(1):11–28. https://doi.org/10.1016/S0005-2736(99)00198-4
12. Monnet V, Juillard V, Gardan R (2016) Peptide conversations in Gram-positive bacteria. Crit Rev Microbiol 42(3):339–351. https://doi.org/10.3109/1040841X.2014.948804
13 Baishya J et al (2021) The impact of intraspecies and interspecies bacterial interactions on disease outcome. Pathogens 10(2):2. https://doi.org/10.3390/pathogens10020096
14. Castillo-Juárez I, Blancas-Luciano BE, García-Contreras R, Fernández-Presas AM (2022) Antimicrobial peptides properties beyond growth inhibition and bacterial killing. PeerJ 10:e12667. https://doi.org/10.7717/peerj.12667
15. Huan Y, Kong Q, Mou H and Yi H (2020) Antimicrobial peptides: classification, design, application and research progress in multiple fields. Front Microbiol 11:582779. https://doi.org/10.3389/fmicb.2020.582779
16. Zhang L, Gallo RL (2016) Antimicrobial peptides. Curr Biol 26(1):R14–R19. https://doi.org/10.1016/j.cub.2015.11.017
17. Jhong J-H et al (2022) dbAMP 2.0: updated resource for antimicrobial peptides with an enhanced scanning method for genomic and proteomic data. Nucleic Acids Res 50(D1):D460–D470. https://doi.org/10.1093/nar/gkab1080
18. Shi G et al (2022) DRAMP 3.0: an enhanced comprehensive data repository of antimicrobial peptides. Nucleic Acids Res 50(D1):D488–D496. https://doi.org/10.1093/nar/gkab651
19. Cardoso MH, Meneguetti BT, Oliveira-Júnior NG, Macedo MLR, Franco OL (2022) Antimicrobial peptide production in response to gut microbiota imbalance. Peptides 157:170865. https://doi.org/10.1016/j.peptides.2022.170865

20. Hols P, Ledesma-García L, Gabant P, Mignolet J (2019) Mobilization of microbiota commensals and their bacteriocins for therapeutics. Trends Microbiol 27(8):690–702. https://doi.org/10.1016/j.tim.2019.03.007
21. Eijsink VGH, Skeie M, Middelhoven PH, Brurberg MB, Nes IF (1998) Comparative studies of Class IIa Bacteriocins of lactic acid bacteria. Appl Environ Microbiol 64(9):3275–3281. https://doi.org/10.1128/AEM.64.9.3275-3281.1998
22. Uteng M et al (2003) Three-dimensional structure in lipid micelles of the pediocin-like antimicrobial peptide Sakacin P and a Sakacin P variant that is structurally stabilized by an inserted C-Terminal disulfide bridge. Biochemistry 42(39):11417–11426. https://doi.org/10.1021/bi034572i
23. Rogne P, Haugen C, Fimland G, Nissen-Meyer J, Kristiansen PE (2009) Three-dimensional structure of the two-peptide bacteriocin plantaricin JK. Peptides 30(9):1613–1621. https://doi.org/10.1016/j.peptides.2009.06.010
24. Acedo JZ, van Belkum MJ, Lohans CT, McKay RT, Miskolzie M, Vederas JC (2015) Solution structure of Acidocin B, a circular bacteriocin produced by Lactobacillus acidophilus M46. Appl Environ Microbiol 81(8):2910–2918. https://doi.org/10.1128/AEM.04265-14
25. Acedo JZ, van Belkum MJ, Lohans CT, Towle KM, Miskolzie M, Vederas JC (2016) Nuclear magnetic resonance solution structures of lacticin Q and Aureocin A53 reveal a structural motif conserved among leaderless bacteriocins with broad-spectrum activity. Biochemistry 55(4):733–742. https://doi.org/10.1021/acs.biochem.5b01306
26. Cascales E et al (2007) Colicin biology. Microbiol Mol Biol Rev 71(1):158–229. https://doi.org/10.1128/MMBR.00036-06
27. Bull M, Plummer S, Marchesi J, Mahenthiralingam E (2013) The life history of Lactobacillus acidophilus as a probiotic: a tale of revisionary taxonomy, misidentification and commercial success. FEMS Microbiol Lett 349(2):77–87. https://doi.org/10.1111/1574-6968.12293
28. Sabala I et al (2014) Crystal structure of the antimicrobial peptidase lysostaphin from Staphylococcus simulans. FEBS J 281(18):4112–4122. https://doi.org/10.1111/febs.12929
29. Wadhwani P, Epand RF, Heidenreich N, Bürck J, Ulrich AS, Epand RM (2012) Membrane-active peptides and the clustering of anionic lipids. Biophys J 103(2):265–274. https://doi.org/10.1016/j.bpj.2012.06.004
30 Brogden KA (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol 3(3):3. https://doi.org/10.1038/nrmicro1098
31. Lohner K, Prossnigg F (2009) Biological activity and structural aspects of PGLa interaction with membrane mimetic systems. Biochim Biophys Acta BBA - Biomembr 1788(8):1656–1666. https://doi.org/10.1016/j.bbamem.2009.05.012
32. Izadpanah A, Gallo RL (2005) Antimicrobial peptides. J Am Acad Dermatol 52(3):381–390. https://doi.org/10.1016/j.jaad.2004.08.026
33. Fernández de Ullivarri M, Arbulu S, Garcia-Gutierrez E and Cotter PD (2020) Antifungal peptides as therapeutic agents. Front Cell Infect Microbiol 10:105. https://doi.org/10.3389/fcimb.2020.00105
34. Giovati L, Ciociola T, Magliani W, Conti S (2018) Antimicrobial peptides with antiprotozoal activity: current state and future perspectives. Future Med Chem 10(22):2569–2572. https://doi.org/10.4155/fmc-2018-0460
35. Bastian A, Schäfer H (2001) Human α-defensin 1 (HNP-1) inhibits adenoviral infection in vitro. Regul Pept 101(1):157–161. https://doi.org/10.1016/S0167-0115(01)00282-8
36. Robinson WE Jr, McDougall B, Tran D, Selsted ME (1998) Anti-HIV-1 activity of indolicidin, an antimicrobial peptide from neutrophils. J Leukoc Biol 63(1):94–100. https://doi.org/10.1002/jlb.63.1.94
37. Sinha S, Cheshenko N, Lehrer RI, Herold BC (2003) NP-1, a Rabbit α-defensin, prevents the entry and intercellular spread of Herpes Simplex Virus Type 2. Antimicrob Agents Chemother 47(2):494–500. https://doi.org/10.1128/AAC.47.2.494-500.2003
38. Duquesne S, Destoumieux-Garzón D, Peduzzi J, Rebuffat S (2007) Microcins, gene-encoded antibacterial peptides from enterobacteria. Nat Prod Rep 24(4):708–734. https://doi.org/10.1039/B516237H
39. Ladjouzi R, Lucau-Danila A, Benachour A and Drider D (2020) A leaderless two-peptide bacteriocin, enterocin DD14, is involved in its own self-immunity: evidence and insights. Front Bioeng Biotechnol 8:644. https://doi.org/10.3389/fbioe.2020.00644
40 Cotter PD, Hill C, Ross RP (2005) Bacteriocins: developing innate immunity for food. Nat Rev Microbiol 3(10):10. https://doi.org/10.1038/nrmicro1273
41. Perez RH, Zendo T, Sonomoto K (2018) Circular and leaderless bacteriocins: biosynthesis, mode of action, applications, and prospects. Front Microbiol 9:2085. https://doi.org/10.3389/fmicb.2018.02085
42. Gebhard S (2012) ABC transporters of antimicrobial peptides in Firmicutes bacteria – phylogeny, function and regulation. Mol Microbiol 86(6):1295–1317. https://doi.org/10.1111/mmi.12078
43. Steinstraesser L, Kraneburg U, Jacobsen F, Al-Benna S (2011) Host defense peptides and their antimicrobial-immunomodulatory duality. Immunobiology 216(3):322–333. https://doi.org/10.1016/j.imbio.2010.07.003
44. Garcia-Gutierrez E, Mayer MJ, Cotter PD, Narbad A (2019) Gut microbiota as a source of novel antimicrobials. Gut Microbes 10(1):1–21. https://doi.org/10.1080/19490976.2018.1455790
45. Gubatan J, Holman DR, Puntasecca CJ, Polevoi D, Rubin SJ, Rogalla S (2021) Antimicrobial peptides and the gut microbiome in inflammatory bowel disease. World J Gastroenterol 27(43):7402–7422. https://doi.org/10.3748/wjg.v27.i43.7402
46. Ilinskaya ON, Ulyanova VV, Yarullina DR and Gataullin IG (2017) Secretome of intestinal bacilli: A natural guard against pathologies. Front Microbiol 8:1666. https://doi.org/10.3389/fmicb.2017.01666
47. Perez KJ, Viana JdS, Lopes FC, Pereira JQ, dos Santos DM, Oliveira JS, Velho RV, Crispim SM, Nicoli JR, Brandelli A and Nardi RMD (2017) Bacillus spp. isolated from Puba as a source of biosurfactants and antimicrobial lipopeptides. Front Microbiol 8:61. https://doi.org/10.3389/fmicb.2017.00061
48. Lim KB, Balolong MP, Kim SH, Oh JK, Lee JY, Kang D-K (2016) Isolation and characterization of a broad spectrum bacteriocin from *Bacillus amyloliquefaciens* RX7. BioMed Res Int 2016:e8521476. https://doi.org/10.1155/2016/8521476
49. Collins FWJ, O'Connor PM, O'Sullivan O, Rea MC, Hill C, Ross RP (2016) Formicin – a novel broad-spectrum two-component lantibiotic produced by Bacillus paralicheniformis APC 1576. Microbiology 162(9):1662–1671. https://doi.org/10.1099/mic.0.000340
50. Mandal SM, Silva ON, Franco OL (2014) Recombinant probiotics with antimicrobial peptides: a dual strategy to improve immune response in immunocompromised patients. Drug Discov Today 19(8):1045–1050. https://doi.org/10.1016/j.drudis.2014.05.019
51. Clark A, Mach N (2016) Exercise-induced stress behavior, gut-microbiota-brain axis and diet: a systematic review for athletes. J Int Soc Sports Nutr 13(1):43. https://doi.org/10.1186/s12970-016-0155-6
52. Maher S, McClean S (2006) Investigation of the cytotoxicity of eukaryotic and prokaryotic antimicrobial peptides in intestinal

epithelial cells in vitro. Biochem Pharmacol 71(9):1289–1298. https://doi.org/10.1016/j.bcp.2006.01.012
53. Schmelcher M, Donovan DM, Loessner MJ (2012) Bacteriophage endolysins as novel antimicrobials. Future Microbiol 7(10):1147–1171. https://doi.org/10.2217/fmb.12.97
54. Fischetti VA (2008) Bacteriophage lysins as effective antibacterials. Curr Opin Microbiol 11(5):393–400. https://doi.org/10.1016/j.mib.2008.09.012
55. O'Flaherty S, Ross RP, Coffey A (2009) Bacteriophage and their lysins for elimination of infectious bacteria. FEMS Microbiol Rev 33(4):801–819. https://doi.org/10.1111/j.1574-6976.2009.00176.x
56. Baquero F, Bouanchaud D, Martinez-Perez MC, Fernandez C (1978) Microcin plasmids: a group of extrachromosomal elements coding for low-molecular-weight antibiotics in Escherichia coli. J Bacteriol 135(2):342–347. https://doi.org/10.1128/jb.135.2.342-347.1978
57 Fischetti VA (2018) Development of phage lysins as novel therapeutics: a historical perspective". Viruses 10(6):6. https://doi.org/10.3390/v10060310
58. Nelson D, Loomis L, Fischetti VA (2001) Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. Proc Natl Acad Sci 98(7):4107–4112. https://doi.org/10.1073/pnas.061038398
59 AmaningDanquah C, Minkah PAB, OseiDuah Junior I, Amankwah KB, Somuah SO (2022) Antimicrobial compounds from microorganisms". Antibiotics 11(3):3. https://doi.org/10.3390/antibiotics11030285
60. Kav AB, Sasson G, Jami E, Doron-Faigenboim A, Benhar I, Mizrahi I (2012) Insights into the bovine rumen plasmidome. Proc Natl Acad Sci 109(14):5452–5457. https://doi.org/10.1073/pnas.1116410109
61 Dong B, Yi Y, Liang L, Shi Q (2017) High throughput identification of antimicrobial peptides from fish gastrointestinal microbiota. Toxins 9(9):9. https://doi.org/10.3390/toxins9090266
62. Grafskaia E et al (2020) The Hirudo Medicinalis microbiome is a source of new antimicrobial peptides. Int J Mol Sci 21:19. https://doi.org/10.3390/ijms21197141
63. Santos-Júnior CD, Pan S, Zhao X-M, Coelho LP (2020) Macrel: antimicrobial peptide screening in genomes and metagenomes. PeerJ 8:e10555. https://doi.org/10.7717/peerj.10555
64 Ma Y et al (2022) Identification of antimicrobial peptides from the human gut microbiome using deep learning. Nat Biotechnol 40(6):6. https://doi.org/10.1038/s41587-022-01226-0
65 Oyama LB et al (2022) In silico identification of two peptides with antibacterial activity against multidrug-resistant Staphylococcus aureus. Npj Biofilms Microbiomes 8(1):1. https://doi.org/10.1038/s41522-022-00320-0
66. Huang K-Y et al (2017) Identification of natural antimicrobial peptides from bacteria through metagenomic and metatranscriptomic analysis of high-throughput transcriptome data of Taiwanese oolong teas. BMC Syst Biol 11(7):131. https://doi.org/10.1186/s12918-017-0503-4
67. Onime LA et al (2021) The rumen eukaryotome is a source of novel antimicrobial peptides with therapeutic potential. BMC Microbiol 21(1):105. https://doi.org/10.1186/s12866-021-02172-8
68. Piotto SP, Sessa L, Concilio S, Iannelli P (2012) YADAMP: yet another database of antimicrobial peptides. Int J Antimicrob Agents 39(4):346–351. https://doi.org/10.1016/j.ijantimicag.2011.12.003
69 Hammami R, Zouhir A, Le Lay C, Ben Hamida J, Fliss I (2010) BACTIBASE second release: a database and tool platform for bacteriocin characterization. BMC Microbiol. 10(1):22. https://doi.org/10.1186/1471-2180-10-22
70 Usmani SS, Kumar R, Kumar V, Singh S, Raghava GPS (2018) AntiTbPdb: a knowledgebase of anti-tubercular peptides. Database 2018:bay025. https://doi.org/10.1093/database/bay025
71 Mehta D et al (2014) ParaPep: a web resource for experimentally validated antiparasitic peptide sequences and their structures. Database 2014:bau051. https://doi.org/10.1093/database/bau051
72. Qureshi A, Thakur N, Tandon H, Kumar M (2014) AVPdb: a database of experimentally validated antiviral peptides targeting medically important viruses. Nucleic Acids Res 42(D1):D1147–D1153. https://doi.org/10.1093/nar/gkt1191
73. Gómez EA, Giraldo P, Orduz S (2017) InverPep: a database of invertebrate antimicrobial peptides. J Glob Antimicrob Resist 8:13–17. https://doi.org/10.1016/j.jgar.2016.10.003
74 Hammami R, Ben Hamida J, Vergoten G, Fliss I (2009) PhytAMP: a database dedicated to antimicrobial plant peptides Nucleic. Acids Res 37(suppl_1):963–968. https://doi.org/10.1093/nar/gkn655
75 Whitmore L, Wallace BA (2004) The Peptaibol Database: a database for sequences and structures of naturally occurring peptaibols. Nucleic Acids Res. 32(suppl_1):D593–D594. https://doi.org/10.1093/nar/gkh077
76. Seebah S et al (2007) Defensins knowledgebase: a manually curated database and information source focused on the defensins family of antimicrobial peptides. Nucleic Acids Res 35(suppl_1):D265–D268. https://doi.org/10.1093/nar/gkl866
77. Lee H-T, Lee C-C, Yang J-R, Lai JZC, Chang KY (2015) A large-scale structural classification of antimicrobial peptides. BioMed Res Int 2015:e475062. https://doi.org/10.1155/2015/475062
78. Wang G, Li X, Wang Z (2016) APD3: the antimicrobial peptide database as a tool for research and education. Nucleic Acids Res 44(D1):D1087–D1093. https://doi.org/10.1093/nar/gkv1278
79. Pirtskhalava M et al (2021) DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. Nucleic Acids Res 49(D1):D288–D297. https://doi.org/10.1093/nar/gkaa991
80. Waghu FH, Barai RS, Gurung P, Idicula-Thomas S (2016) CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. Nucleic Acids Res 44(D1):D1094–D1097. https://doi.org/10.1093/nar/gkv1051
81 Ye G et al (2020) LAMP2: a major update of the database linking antimicrobial peptides. Database 2020:baaa061. https://doi.org/10.1093/database/baaa061
82. Camacho C et al (2009) BLAST+: architecture and applications. BMC Bioinformatics 10(1):421. https://doi.org/10.1186/1471-2105-10-421
83. Walsh CJ, Guinane CM, O'Toole PW, Cotter PD (2017) A profile hidden Markov model to investigate the distribution and frequency of LanB-encoding lantibiotic modification genes in the human oral and gut microbiome. PeerJ 5:e3254. https://doi.org/10.7717/peerj.3254
84. Gull S, Shamim N, Minhas F (2019) AMAP: Hierarchical multilabel prediction of biologically active and antimicrobial peptides. Comput Biol Med 107:172–181. https://doi.org/10.1016/j.compbiomed.2019.02.018
85. Fingerhut LCHW, Miller DJ, Strugnell JM, Daly NL, Cooke IR (2020) ampir: an R package for fast genome-wide prediction of antimicrobial peptides. Bioinformatics 36(21):5262–5263. https://doi.org/10.1093/bioinformatics/btaa653
86. Waghu FH, Idicula-Thomas S (2020) Collection of antimicrobial peptides database and its derivatives: applications and beyond. Protein Sci 29(1):36–42. https://doi.org/10.1002/pro.3714
87. Xiao X, Wang P, Lin W-Z, Jia J-H, Chou K-C (2013) iAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types. Anal Biochem 436(2):168–177. https://doi.org/10.1016/j.ab.2013.01.019

88. Meher PK, Sahu TK, Saini V, Rao AR (2017) Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC. Sci Rep 7(1):1. https://doi.org/10.1038/srep42362
89 Bhadra P, Yan J, Li J, Fong S, Siu SWI (2018) AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest". Sci Rep 8(1):1. https://doi.org/10.1038/s41598-018-19752-w
90. Lawrence TJ et al (2021) amPEPpy 1.0: a portable and accurate antimicrobial peptide prediction tool. Bioinformatics 37(14):2058–2060. https://doi.org/10.1093/bioinformatics/btaa917
91. Veltri D, Kamath U, Shehu A (2018) Deep learning improves antimicrobial peptide recognition. Bioinformatics 34(16):2740–2747. https://doi.org/10.1093/bioinformatics/bty179
92. Joseph S, Karnik S, Nilawe P, Jayaraman VK, Idicula-Thomas S (2012) ClassAMP: a prediction tool for classification of antimicrobial peptides. IEEE/ACM Trans Comput Biol Bioinform 9(5):1535–1538. https://doi.org/10.1109/TCBB.2012.89
93. Li C et al (2022) AMPlify: attentive deep learning model for discovery of novel antimicrobial peptides effective against WHO priority pathogens. BMC Genomics 23(1):77. https://doi.org/10.1186/s12864-022-08310-4
94. Xiao X, Shao Y-T, Cheng X, Stamatovic B (2021) iAMP-CA2L: a new CNN-BiLSTM-SVM classifier based on cellular automata image for identifying antimicrobial peptides and their functional types. Brief Bioinform 22(6):bbab09. https://doi.org/10.1093/bib/bbab209
95. Reller LB, Weinstein M, Jorgensen JH, Ferraro MJ (2009) Antimicrobial susceptibility testing: a review of general principles and contemporary practices. Clin Infect Dis 49(11):1749–1755. https://doi.org/10.1086/647952
96. Malanovic N, Marx L, Blondelle SE, Pabst G, Semeraro EF (2020) Experimental concepts for linking the biological activities of antimicrobial peptides to their molecular modes of action. Biochim Biophys Acta BBA - Biomembr 1862(8):183275. https://doi.org/10.1016/j.bbamem.2020.183275
97. Hoover DG, Harlander SK (1993) Screening methods for detecting bacteriocin activity. In bacteriocins of lactic acid bacteria (pp. 23–39). Academic Press. https://doi.org/10.1016/B978-0-12-355510-6.50010-5
98. Freitas AR, Karpiński TM and Li B (2020) Editorial: antimicrobials and anticancers of bacterial origins. Front Microbiol 11:842. https://doi.org/10.3389/fmicb.2020.00842
99. Hwang JS, Kim SG, Shin TH, Jang YE, Kwon DH, Lee G (2022) Development of anticancer peptides using artificial intelligence and combinational therapy for cancer therapeutics. Pharmaceutics 14(5):5. https://doi.org/10.3390/pharmaceutics14050997
100. Wu X et al (2014) In vitro and in vivo activities of antimicrobial peptides developed using an amino acid-based activity prediction method. Antimicrob Agents Chemother 58(9):5342–5349. https://doi.org/10.1128/AAC.02823-14
101. Arenas I, Ibarra MA, Santana FL, Villegas E, Hancock REW, Corzo G (2020) *In vitro* and *in vivo* antibiotic capacity of two host defense peptides. Antimicrob Agents Chemother 64:7. https://doi.org/10.1128/AAC.00145-20
102. Nordström R, Malmsten M (2017) Delivery systems for antimicrobial peptides. Adv Colloid Interface Sci 242:17–34. https://doi.org/10.1016/j.cis.2017.01.005
103 Jung C-J et al (2021) Identification of potential therapeutic antimicrobial peptides against Acinetobacter baumannii in a mouse model of pneumonia. Sci Rep 11(1):1. https://doi.org/10.1038/s41598-021-86844-5
104. Magana M et al (2020) The value of antimicrobial peptides in the age of resistance. Lancet Infect Dis 20(9):e216–e230. https://doi.org/10.1016/S1473-3099(20)30327-3
105. Luong HX, Thanh TT, Tran TH (2020) Antimicrobial peptides – advances in development of therapeutic applications. Life Sci 260:118407. https://doi.org/10.1016/j.lfs.2020.118407
106. Baxter AA, Lay FT, Poon IKH, Kvansakul M, Hulett MD (2017) Tumor cell membrane-targeting cationic antimicrobial peptides: novel insights into mechanisms of action and therapeutic prospects. Cell Mol Life Sci 74(20):3809–3825. https://doi.org/10.1007/s00018-017-2604-z
107 Hancock REW, Sahl H-G (2006) Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. Nat Biotechnol 24(12):12. https://doi.org/10.1038/nbt1267
108 Boutin S, Dalpke AH (2020) The microbiome: a reservoir to discover new antimicrobials agents. Curr Top Med Chem 20(14):1291–1299
109. Nakatsuji T et al (2017) Antimicrobials from human skin commensal bacteria protect against *Staphylococcus aureus* and are deficient in atopic dermatitis. Sci Transl Med 9(378):eaah4680. https://doi.org/10.1126/scitranslmed.aah4680
110. Mangoni ML, McDermott AM, Zasloff M (2016) Antimicrobial peptides and wound healing: biological and therapeutic considerations. Exp Dermatol 25(3):167–173. https://doi.org/10.1111/exd.12929
111. McDermott AM (2009) The role of antimicrobial peptides at the ocular surface. Ophthalmic Res 41(2):60–75. https://doi.org/10.1159/000187622
112. Kang JK et al (2011) The insect peptide coprisin prevents clostridium difficile-mediated acute inflammation and mucosal damage through selective antimicrobial activity. Antimicrob Agents Chemother 55(10):4850–4857. https://doi.org/10.1128/AAC.00177-11
113. Bormann N et al (2017) A short artificial antimicrobial peptide shows potential to prevent or treat bone infections. Sci Rep 7(1):1506. https://doi.org/10.1038/s41598-017-01698-0
114. Donnelly JP, Bellm LA, Epstein JB, Sonis ST, Symonds RP (2003) Antimicrobial therapy to prevent or treat oral mucositis. Lancet Infect Dis 3(7):405–412. https://doi.org/10.1016/S1473-3099(03)00668-6
115. Majewski K, Kozłowska E, Żelechowska P, Brzezińska-Błaszczyk E (2018) Serum concentrations of antimicrobial peptide cathelicidin LL-37 in patients with bacterial lung infections. Cent Eur J Immunol 43(4):453–457. https://doi.org/10.5114/ceji.2018.81355
116 Mirski T, Lidia M, Nakonieczna A, Gryko R (2019) Bacteriophages, phage endolysins and antimicrobial peptides – the possibilities for their common use to combat infections and in the design of new drugs. Ann Agric Environ Med 26(2):203–209. https://doi.org/10.26444/aaem/105390
117. Wang S, Zeng X, Yang Q, Qiao S (2016) Antimicrobial peptides as potential alternatives to antibiotics in food animal industry. Int J Mol Sci 17(5):603. https://doi.org/10.3390/ijms17050603
118. Li J, Hu S, Jian W, Xie C, Yang X (2021) Plant antimicrobial peptides: structures, functions, and applications. Bot Stud 62(1):5. https://doi.org/10.1186/s40529-021-00312-x
119. Deslouches B, Di YP (2017) Antimicrobial peptides: a potential therapeutic option for surgical site infections. Clin Surg 2:1740
120. Alencar-Silva T et al (2018) Breaking the frontiers of cosmetology with antimicrobial peptides. Biotechnol Adv 36(8):2019–2031. https://doi.org/10.1016/j.biotechadv.2018.08.005
121. Tian T et al (2021) Industrial application of antimicrobial peptides based on their biological activity and structure-activity relationship. Crit Rev Food Sci Nutr 0:1–16. https://doi.org/10.1080/10408398.2021.2019673

122. Yoshida M et al (2018) Using evolutionary algorithms and machine learning to explore sequence space for the discovery of antimicrobial peptides. Chem 4(3):533–543. https://doi.org/10.1016/j.chempr.2018.01.005
123. Aguilera-Puga MDC, Cancelarich NL, Marani MM, de la Fuente-Nunez C, Plisson F (2024) Accelerating the discovery and design of antimicrobial peptides with artificial intelligence. In: Gore M, Jagtap UB (eds), Computational Drug Discovery and Design. Methods in Molecular Biology, vol 2714. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-3441-7_18
124. Han Y, Zhang M, Lai R, Zhang Z (2021) Chemical modifications to increase the therapeutic potential of antimicrobial peptides. Peptides 146:170666. https://doi.org/10.1016/j.peptides.2021.170666
125. Wang Y, Stebe KJ, de la Fuente-Nunez C, Radhakrishnan R (2023) Computational design of peptides for biomaterials applications. ACS Appl Bio Mater. https://doi.org/10.1021/acsabm.2c01023
126. Nagarajan D et al (2018) Computational antimicrobial peptide design and evaluation against multidrug-resistant clinical isolates of bacteria. J Biol Chem 293(10):3492–3509. https://doi.org/10.1074/jbc.M117.805499
127 Szymczak P et al (2023) Discovering highly potent antimicrobial peptides with deep generative model HydrAMP. Nat Commun 14(1):1. https://doi.org/10.1038/s41467-023-36994-z
128. Tucs A, Tran DP, Yumoto A, Ito Y, Uzawa T, Tsuda K (2020) Generating ampicillin-level antimicrobial peptides with activity-aware generative adversarial networks. ACS Omega 5(36):22847–22851. https://doi.org/10.1021/acsomega.0c02088
129. Van Oort CM, Ferrell JB, Remington JM, Wshah S, Li J (2021) AMPGAN v2: machine learning-guided design of antimicrobial peptides. J Chem Inf Model 61(5):2198–2207. https://doi.org/10.1021/acs.jcim.0c01441
130. Dean SN, Alvarez JAE, Zabetakis D, Walper SA and Malanoski AP (2021) PepVAE: variational autoencoder framework for antimicrobial peptide generation and activity prediction. Front Microbiol 12:725727. https://doi.org/10.3389/fmicb.2021.725727
131. Surana S, Arora P, Singh D, Sahasrabuddhe D, Valadi J (2023) PandoraGAN: generating antiviral peptides using generative adversarial network. SN Comput Sci 4(5):607. https://doi.org/10.1007/s42979-023-02203-3
132 Chen CH, Lu TK (2020) Development and challenges of antimicrobial peptides for therapeutic applications. Antibiotics 9(1):1. https://doi.org/10.3390/antibiotics9010024
133. Lau JL, Dunn MK (2018) Therapeutic peptides: historical perspectives, current development trends, and future directions. Bioorg Med Chem 26(10):2700–2707. https://doi.org/10.1016/j.bmc.2017.06.052

Microbial Ecology (2026) 89:15
https://doi.org/10.1007/s00248-025-02620-2

RESEARCH

Bioactive Plasmid- and Phage-Encoded Antimicrobial Peptides (AMPs) in the Human Gut: A Metatranscriptome–Virome Profiling Reveals Exploratory Links to Metabolic Human Diseases

Luigui Gallardo-Becerra[1] · Fernanda Cornejo-Granados[1] · Shirley Bikel[1] · Iván Arenas[2] · Gamaliel López-Leal[3] · Carolina Alvarado-Gonzalez[4] · Filiberto Sánchez-López[1] · Rubiceli Manzo[1] · Gerardo Corzo[2] · Gerardo P. Espino-Solis[4] · Samuel Canizales-Quinteros[5] · Adrian Ochoa-Leyva[1]

Received: 18 July 2025 / Accepted: 28 September 2025

Abstract

Microbe-derived antimicrobial peptides (AMPs) can shape gut community structure; however, their contribution to disease-associated dysbiosis remains poorly understood. We assembled fecal metatranscriptomes from individuals with normal weight (NW), obesity (O), and obesity with metabolic syndrome (OMS), yielding 51,087 non-human transcripts. We screened 1,095 small open reading frames (smORFs) using AMP-prediction algorithms combined with stringent post-hoc bioinformatics filters identifying 51 high-confidence AMP candidates. Most matched bacterial homologs, predominantly *Faecalibacterium prausnitzii*, while eight mapped to plasmids or bacteriophages. Differential expression identified two and four AMPs overexpressed in O and OMS, respectively. Two of them were originated from chromosomes, three from phages, and one from plasmid. Notably, the over-expression of these AMPs was negatively correlated with healthy-associated bacteria and positively correlated with obesity-enriched taxa. Furthermore, these AMPs were broadly detectable across 372 external gut metatranscriptomes (prevalence up to 94% of the samples) indicating conservation within the human gut microbiome and highlighting mobile elements as an overlooked reservoir of transcriptionally active AMPs. Using DNA virome sequencing and prophage analyses, we suggested phage origin of the transcribed AMPs. We further synthesized a phage-encoded AMP (AMP-3020), demonstrating broad-spectrum activity against Gram-positive and Gram-negative bacteria, without detectable cytotoxicity toward human immune T cells. This supports the idea that phages could encode functional AMPs capable of shaping gut community structure by suppressing diverse bacteria without harming host immune cells. Our gut metatranscriptome–virome profiling revealed a conservative core of actively transcribed, plasmid- and phage-encoded AMPs with exploratory associations to obesity/MetS. These findings support mobile-element AMPs as candidate ecological regulators and motivate validation in larger cohorts and mechanistic models.

Keywords Antimicrobial peptides · AMPs · Phages · Metatranscriptome · Virome · Plasmid

✉ Adrian Ochoa-Leyva
adrian.ochoa@ibt.unam.mx

[1] Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, C.P. 62210 Cuernavaca, Morelos, Mexico

[2] Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, 62210 Cuernavaca, Morelos, Mexico

[3] Laboratorio de Biología Computacional y Virómica Integrativa, Centro de Investigación en Dinámica Celular, Universidad Autónoma del Estado de Morelos, 62209 Cuernavaca, Mexico

[4] Facultad de Medicina y Ciencias Biomédicas, Laboratorio Nacional de Citometría de Flujo, Autonomous University of Chihuahua, Circuito Universitario S/N, Campus II, 31125 Chihuahua, Mexico

[5] Facultad de Química, UNAM/Instituto Nacional de Medicina Genómica (INMEGEN), Unidad de Genómica de Poblaciones Aplicada a La Salud, Mexico City, Mexico

Published online: 28 November 2025

Introduction

Childhood obesity has emerged as a critical public health challenge [1], significantly increasing the risk of type 2 diabetes, cardiovascular diseases, and non-alcoholic fatty liver disease in adulthood [2]. It also predisposes to metabolic syndrome (MetS) [3], characterized by high blood glucose levels, elevated triglycerides, low high-density lipoproteins (HDL) levels, and hypertension [4, 5]. Together, these abnormalities diminish quality of life and increase the risk of disability and premature mortality [2, 6]. The situation is particularly alarming in Mexico, where nearly one in five school-aged children is obese and many likely have undiagnosed metabolic syndrome (MetS) [7], underscoring the urgent need for effective preventive and therapeutic strategies.

A growing body of work links obesity to shifts in gut microbial communities, as demonstrated by 16S rRNA profiling, shotgun metagenomics, and metatranscriptomics [8]. Viromics further implicates bacteriophages as key modulators of these communities, which can either maintain a healthy microbe balance (eubiosis) or induce an imbalance (dysbiosis) [9]. Dissecting microbe-phage interactions may enable microbiome-targeted strategies to prevent and treat obesity and metabolic syndrome.

Bidirectional communication between hosts and their resident microbes is mainly mediated by the "secrebiome," a collection of proteins released into the environment [8]. This includes enzymes, toxins, and antimicrobial peptides (AMPs), which are produced by both the host and microbiota as a key innate defense strategy [10, 11]. Despite their small size, AMPs exhibit diverse antimicrobial activities, from membrane disruption to binding essential intracellular targets [12]. In bacteria, AMPs can be synthesized through the proteolytic processing of larger precursor proteins [13, 14], non-ribosomal peptide synthetase (NRPS) pathways [15], and dedicated genomic genes [16]. Recent metagenomic studies have highlighted small open reading frames (smORFs; < 100 codons), as a rich reservoir of novel microbiome-derived AMPs [17–19]. Early large-scale surveys of human microbiomes uncovered thousands of previously unannotated smORFs, some with putative antimicrobial functions [19]. Understanding the diversity and regulation of these smORF-encoded AMPs is crucial for elucidating host-microbiome interactions in health and disease. In the human gut, host-encoded AMPs act as gatekeepers, regulating opportunistic pathogens and the composition of the resident microbiota [20, 21]. They help maintain microbial diversity, which is essential for intestinal health, and they are naturally compatible with the host due to their co-evolution within the intestinal ecosystem [22–24]. However, some AMPs may increase susceptibility to viral infections, underscoring context-dependent effects [25]. At the same time, microbial-derived AMPs foster competition among bacteria, allowing AMP-producing species to secure ecological niches and influence microbial dynamics [26]. Horizontal gene transfer via plasmids and bacteriophages (phages) expands AMP repertoires and disseminates AMP-encoding genes, enhancing bacterial fitness [27–30]. However, no study has established a link between AMPs and the dysbiosis seen in disorders like obesity and metabolic syndrome.

Metatranscriptomics provides insights into which AMP genes are actively transcribed in situ, highlighting those involved in community self-regulation. This study examines gut metatranscriptomes from healthy, obese, and obese with metabolic syndrome individuals to catalog differentially expressed AMPs and explore their roles in microbiota dysbiosis. To our knowledge, this study is among the first to connect the actively transcribed gut AMP repertoire with obesity and metabolic syndrome, laying groundwork for microbiome-targeted interventions.

Materials and Methods

De Novo Assembly and AMP Prediction

We analyzed RNA-seq data from two normal-weight (NW), three obese (O), and three obese children with Metabolic Syndrome (OMS), as detailed in a prior study (BioProject PRJNA600247) [8]. Quality assessment was performed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), followed by filtering with a Q20 Phred score and trimming using Trimmomatic v0.36. We removed sequencing adapters, ambiguous bases and rRNAs using Ribopicker v0.4.3 and the SILVA rRNA database (138 release). The remaining non-rRNA sequences were aligned to the human genome and transcriptome with Kneaddata to eliminate human derived sequences (https://github.com/biobakery/kneaddata). The remaining sequences were then used as input for the de novo transcriptome assembly with the Trinity assembler, adjusting to conserve small transcripts. After that, the original reads were aligned to the transcriptome using Bowtie2. Expression levels were calculated by Expectation (RSEM). After, we removed the transcripts with an expression below 1 in any sample or with less than three cumulative observations across all samples. Then, we employed TransDecoder (https://github.com/TransDecoder/) to identify ORF candidates within protein-coding regions of the transcript sequences, adjusting its parameters to obtain the smORFs (-m 5). The small proteins derived from these smORFs were analyzed for AMP prediction

using Macrel (https://github.com/BigDataBiology/macrel), AxPEP (https://sourceforge.net/projects/axpep/), and AMP Scanner V2 (https://www.dveltri.com/ascan/v2/ascan.html) with default parameters.

Taxonomic Classification, Phylogeny, Genomic Synteny and Prophage Analyses

The transcript sequences encoding AMPs were used for homology search against NCBI's NT database using the BLASTN algorithm with the following parameters: e-value ≤ 0.00001, identity ≥80%, and coverage ≥80%. The associated taxonomy was determined with a Last Common Ancestor (LCA) using MEGAN6 Community Edition (v.6.25.10). The genomic context was obtained for all transcripts, using BLASTN best-hit, and the phylogenic trees were created with iTOL (https://itol.embl.de). Plots for genomic context were created with the AnnotationSketch drawing library (https://genometools.org/annotation sketch.html). The genomic synteny comparisons were conducted using Easyfig (https://mjsull.github.io/Easyfig/) with the BLASTN algorithm. We highlighted genomes containing the AMPs identified in this study in bold in the corresponding figure. Arrows indicate the positions of coding sequences, while shaded lines represent the degree of homology between genomic regions or genomes. We assigned functional categories based on the COG classification system using the COG classifier tool (https://github.com/moshi4/COGclassifier), with each category represented by a distinct color to facilitate interpretation. We screened the bacterial genomes with VirSorter2 to determine whether the AMPs we identified were embedded within prophage regions (https://github.com/jiarong/VirSorter2).

Read Mapping of Viral DNA to Phage Genomes

The quality-filtered phageome reads were obtained from a previously published dataset by our laboratory, which was derived from the same cohort (BioProject: PRJNA646512). We merged all R1 files into a single R1 file and all R2 files into a single R2 file to create a representative population of the entire virome. Next, we created an index for each phage genome using bowtie2-build with default parameters. We aligned the merged reads to the bacteriophage genomes using Bowtie2 with the parameters '—no-unal –end-to-end –very-sensitive.' Finally, we calculated the total genome coverage and the X-fold coverage for each genome.

16S Microbiota Profiling

We obtained the V4 region of 16S rRNA gene for the samples from a previously published dataset by our laboratory (BioProject PRJNA600247) [8]. All quality-filtered sequences were joined and analyzed using QIIME2 (v2024.5). Briefly, raw sequences were imported and dereplicated, followed by de novo clustering at a 97% identity threshold. Taxonomic classification was performed using a Naive Bayes classifier trained on the SILVA 138 reference database. Taxonomy bar plots were generated, and amplicon sequence variants (ASVs) that were present in more than half of the samples (n = 4) were retained for Spearman correlation analysis against AMP abundances.

Peptide Synthesis and Purification

ADR1 and ADR2 were chemically synthesized by a solid-phase method using the Fmoc methodology (GenScript Biotech, Piscataway, NJ) and purified. Briefly, ten milligrams of crude synthetic peptides were dissolved in one milliliter of 20% aqueous acetonitrile solution and were separated each by reverse phase HPLC on an analytic C18 column (Zorbax SB-C18, Agilent, USA). The C18 column was equilibrated in 20% aqueous acetonitrile containing 0.1% TFA, and the synthetic peptides were separated using a linear gradient of acetonitrile/0.1% TFA from 20 to 60% in 40 min at a flow rate of 1 mL/min. The presence of each peptide was monitored at 220 nm. The molecular mass of each peptide was determined by mass spectrometry.

Antibacterial Assays

The assessment of bacterial growth was conducted using a broth microdilution assay according to the Clinical and Laboratory Standards Institute guidelines. The reference strains, *Pseudomonas aeruginosa* (ATCC 27853), *Klebsiella pneumoniae* (ATCC700603), *Staphylococcus aureus* (ATCC 29213) and *Streptococcus pneumoniae* (ATCC46916) were cultured in Mueller–Hinton broth (MHB) at 37ºC during overnight incubation. Following this culture period, the samples were diluted in MHB to achieve an endpoint corresponding to an absorbance between 0.08 and 0.13 units at 625 nm, followed by a further dilution of 1:100 in MHB (approximately 1 × 108 CFU/mL). An aliquot of fifty microliters from each bacterial suspension was introduced into each well of a 96-well microtiter culture plate, which contained 50 µL of MHB supplemented with varying concentrations of the synthetic peptides ADR1 and ADR2 (100, 50, 25, 12.5, 6.2, and 3.13 µg/mL). The growth of each bacterial strain was quantitatively evaluated by measuring absorbance at a wavelength of 630 nm after an incubation period of 18 h at 37 °C.

Flow Cytometry Analysis of T Cell Populations

Human red blood cells were collected from a healthy donor who gave verbal consent for phlebotomy. This related

experiment was approved by the Ethics Committee of the Facultad de Medicina y Ciencias Biomédicas of the Universidad Autónoma de Chihuahua with registration number CI-068–19**.** To assess the potential cytotoxic effects of the synthetic peptides on human immune cells, peripheral blood mononuclear cells (PBMCs) were isolated from three healthy donors. The cells were cultured in RPMI 1640 medium (GIBCO, 11875–085, Paisley, SCT, UK) supplemented with 5% fetal bovine serum (FBS) (GIBCO, 26140–079, Grand Island, NE, USA) and GlutaMAX (GIBCO, 35050–061, Grand Island, NE, USA). PBMCs were seeded in 96-well plates and exposed to each AMP peptide (ADR1 and ADR2) at 20 mg/mL. Cells cultured in medium alone were used as untreated controls. The initial evaluation was performed one hour after peptide exposure to determine the immediate effects on T lymphocyte populations. Subsequently, the PBMCs were incubated overnight at 37 °C in a humidified atmosphere with 5% CO_2 to assess potential delayed or sustained effects. Following incubation, cells were stained with the BD TriTest™ reagent (Becton Dickinson and Company, BD Biosciences, San Jose, CA, USA) to identify $CD3^+$ T cells and their $CD4^+$ and $CD8^+$ subpopulations, the viability of $CD3^+$, $CD4^+$, and $CD8^+$ T cell subpopulations was evaluated using the LIVE/DEAD™ Fixable Near-IR Dead Cell Stain Kit (Invitrogen™, Thermo Fisher Scientific, Eugene, OR, USA). The absolute number of cells was acquired using BDTM Liquid Counting Beads (BD Biosciences). Flow cytometric analysis was performed using a FACS Canto flow cytometer (BD Biosciences), and the data were analyzed with FlowJo software (BD Biosciences). Graphical representations and statistical analyses were performed using GraphPad Prism (GraphPad Software, San Diego, CA, USA). This experimental design allowed for the evaluation of peptide-induced changes in the distribution and viability of key T lymphocyte subsets, providing insight into the potential immunotoxicity or immunomodulatory properties of the peptides in human immune cells.

Results

Identification and Taxonomic Classification of AMP-Encoding smORFs of the Gut Microbial Transcripts

We analyzed fecal metatranscriptomes from a previously characterized pediatric cohort (BioProject PRJNA600247) [8], comprising two normal-weight (NW), three obese (O), and three obese children with metabolic syndrome (OMS) (Fig. 1A). After read-quality control and removal of eukaryotic and prokaryotic ribosomal RNA and human RNA sequences, 42.97 million high-quality paired-end reads remained. De novo assembly with Trinity yielded 51,087 transcripts (N50 = 1,372 bp) and 36.22 million assembled nucleotides (Supplementary Tables 1–2 and Supplementary Fig. 1).Of these, 35,803 transcripts encoded smORFs of less or equal than 100 amino acids. Transcript abundance was recalculated as Fragments per Kilobase of transcript per Million mapped reads (FPKM), and we retained transcripts with an FPKM ≥ 1 present in ≥ 3 samples, resulting in 1,095 reproducibly expressed transcripts (Fig. 1A).

We assessed the antimicrobial potential of 1,095 smORF-encoded peptides using three prediction tools Macrel [31], AxPEP [32], and AMP Scanner V2 [33]. This screening identified 420 potential AMPs supported by a least one tool. We retained only AMPs that met all three criteria: (i) length > 10 amino acids; (ii) encoded by a single, non-overlapping ORF (no overlap with any other coding sequence); and (iii) supported by at least one homolog transcript in NCBI (Fig. 1B). After filtering, 112 AMP candidates remained (Fig. 1B). To mitigate potential overestimation of small ORFs, given their limited prevalence in bacterial genomes [19], we scrutinized our initial set of 112 AMPs and found that several lacked start and/or stop codons in their original transcripts, consistent with possible false positives AMPs. We therefore applied a stringent post hoc filter (Fig. 1C): (i) selected peptides from homologous transcripts with clearly defined start and stop codons; (ii) excluded peptides > 100 aa; (iii) removed sequences no longer predicted as AMPs; and (iv) retained only homolog peptides for which our AMP aligned with > 80% sequence identity and > 40% coverage. After applying these criteria, the AMP set was reduced from 112 to 51. Notably, the retained AMPs align to their reference peptides with a mean coverage of 92% and a mean sequence identity of 97%, indicating that only small N- or C-terminal fragments were typically missing (Supplementary Table 3). All homologous genes in NCBI were annotated as hypothetical or uncharacterized. Consistent with this, our BLAST analyses of the 51 AMP peptide sequences returned only homologs labeled with unknown function. This suggest that these proteins lack prior functional assignments and that our results provide a putative antimicrobial role for this set. From the final 51 AMPs, AxPEP predicted 36 peptides (70.6%), AMP Scanner v2 predicted 23 (45.1%), and Macrel prediected 5 (9.8%). Only three peptides were unanimously predicted by three tools (Supplementary Fig. 2A).

To contextualize predictor performance, we compiled a reference panel of 1,390 experimentally validated microbe-derived peptides from APD3 [34], dbAMP [35], and DRAMP [36], confirming 1,332 (95.8%) as bona-fide AMPs. Benchmarking against this panel highlights substantial inter-tool variability and supports the importance of utilizing multiple predictors for candidate selection.

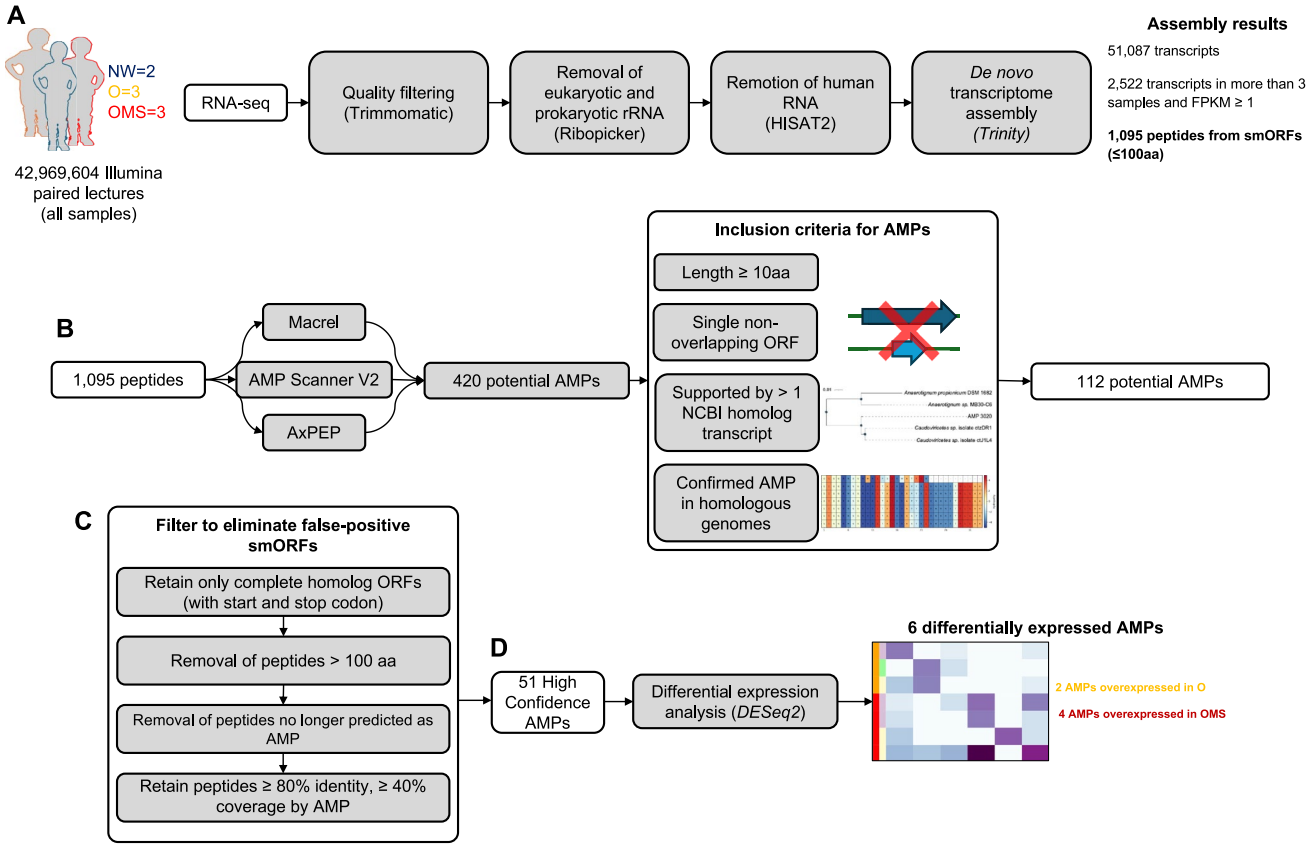


Fig. 1 Workflow for identifying gut-expressed AMPs from fecal metatranscriptomes. **A** Data Processing. Raw Illumina reads were quality filtered and depleted of host and rRNA sequences. The remaining reads were de novo assembled, and smORFs were extracted. **B** AMP discovery. smORFs were screened using three independent predictors Macrel, AMP-Scanner v2, and AxPEP. Candidate AMPs undergo manual curation considering peptide length, genomic context, transcript support, and protein homolog evidence. **C** False-positive filtering of smORFs. Curated AMP candidates were further filtered to remove likely spurious smORFs. **D** Differential Expression. High-confidence AMPs were transcriptionally quantified and compared between obesity (O) and obesity with metabolic syndrome (OMS), yielding the final set of disease-associated, gut expressed AMPs

Taxonomic assignment of the 51 AMP-encoding transcripts indicated that 44 (86.3%) were bacterial, with *Faecalibacterium prausnitzii* accounting for 16 transcripts (31.4%) (Supplementary Fig. 2B). The remaining 7 transcripts (13.7%) were linked to the gut virome, mainly comprising Caudovirales, including Myoviridae and Siphoviridae members. This underscores bacteriophages as an additional source of AMP genes. Notably, one transcript was associated with a plasmid, highlighting the contribution of mobile genetic elements to the gut ecosystem AMP repertoire.

Widespread Transcription of AMPs Differentially Expressed in O and OMS

We compared the expression of the 51 AMPs between the O and OMS groups ($n = 3$ per group). Using DESeq2 with thresholds of fold change > 2 and p-value < 0.05, 6 of 51 AMPs were differentially expressed: two up-regulated in O and four in OMS (Fig. 2A). The taxonomic assignment of these six AMPs revealed a mixture of phage-derived, plasmid-associated, and chromosomally encoded bacterial peptides (Fig. 2A; Supplementary Table 4), suggesting multiple genomic contexts for AMP expression in the gut. The differential expression of these AMPs in O vs. OMS suggests that AMPs dynamics may play a role in the dysbiosis associated with the disease.

To assess biological generalizability, we assessed whether these six AMPs are widely distributed across diverse human gut communities or if they are specific to our cohort. Accordingly, we queried the presence of the six AMPs in an independent dataset of 372 fecal metatranscriptomes from the U.S.-based Health Professionals Follow-Up Study (HPFS; Bioproject PRJNA354235) [37]. This cohort spans a wide range of BMI values, metabolic phenotypes, ages, and dietary backgrounds, thereby offering a robust framework for assessing AMP prevalence across heterogeneous populations. RNA-seq reads were mapped to the six

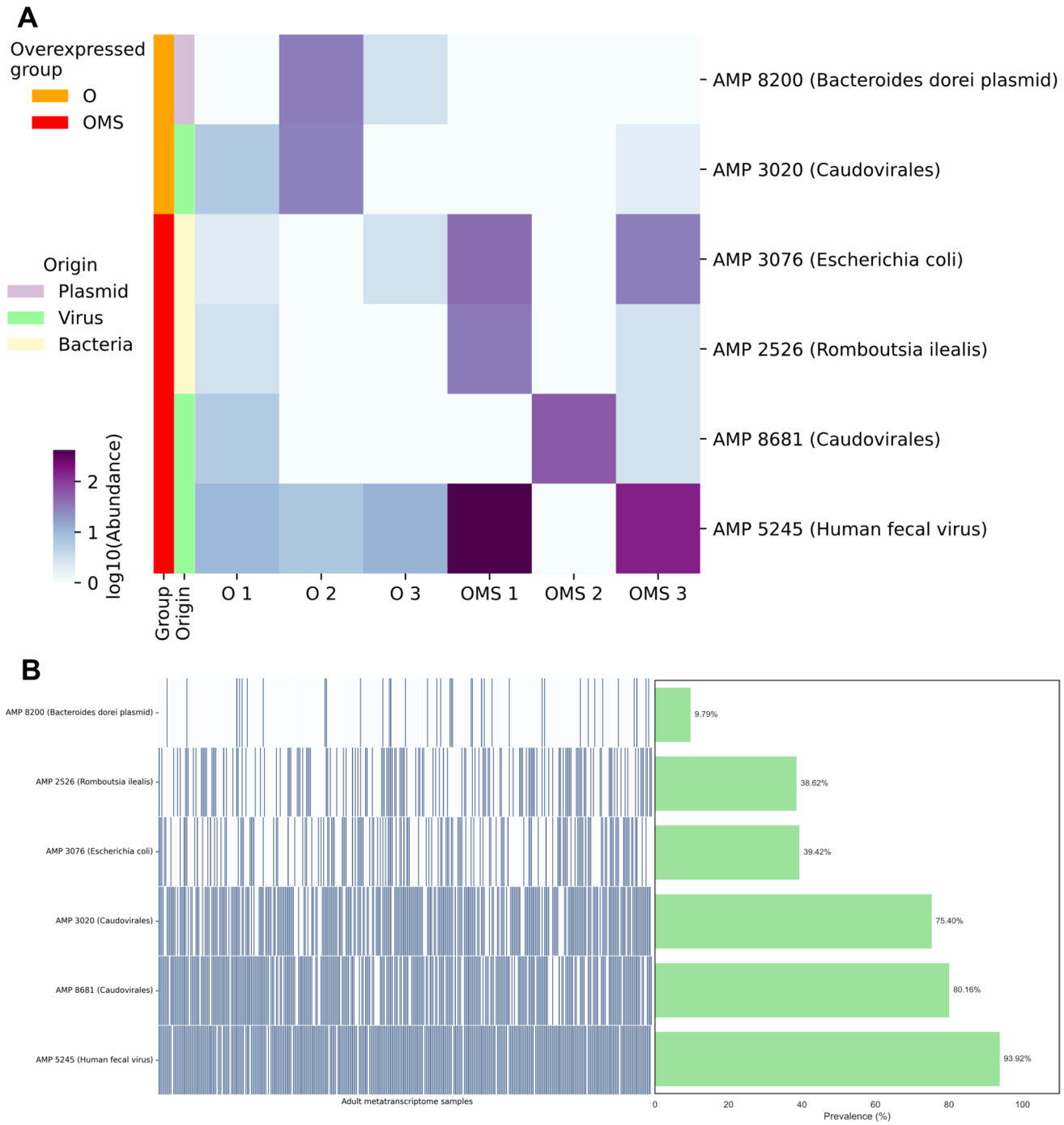


Fig. 2 Overexpression of AMPs associated to obesity and MetS and their prevalence in external datasets. **A** Differential expression heatmap. Rows are AMPs; columns are samples. Row sidebars denote the overexpressed group (O = Obesity or OMS = obesity with MetS), predicted genomic origin (chromosome, plasmid or virus), and the lowest-level taxonomic assignment. **B** External prevalence. Presence of differentially expressed AMPs across 372 independent human gut metatranscriptomes (BioProject PRJNA354235)

AMP-encoding transcripts, and an AMP was considered detected at transcript per million (TPM) > 1. Notably, all six AMPs were observed in a substantial fraction of samples (prevalence 9.8–98.4%; mean 67.93%; median 80.16%; Fig. 2B; Supplementary Table 5), indicating widespread transcription across an independent population. These results point to recurrent AMP expression in the human gut and motivate future work to test their roles in interbacterial antagonism, community structure, and host–microbe signaling.

Chromosomal and Plasmid-Encoded AMPs Exhibit a Conserved Gut AMP Repertoire Across Taxa

Among the six differentially expressed AMPs, two had close protein homologs on bacterial chromosomes with phylogenetically diverse hosts. Specifically, AMP 3076 was identical (100% amino acid sequence identity) to homologs in *Escherichia coli* (Supplementary Fig. 3A and B); AMP 2526 shared 93.75% identity with four *Romboutsia* homologs (Supplementary Fig. 3C and D). The genomic context analyses revealed conserved local synteny around the two AMP surrounding genes, suggesting they are part of conserved genomic modules (Supplementary Fig. 4). The conservation in amino acid sequence and genomic architecture of these AMPs suggests strong purifying selection.

Among chromosomally encoded AMPs, AMP 8200 was notable for its dual genomic occurrence. BLASTP searches identified 15 highly similar protein homologs across *Bacteroides* and *Phocaeicola*, including one copy on a *Phocaeicola dorei* plasmid (Fig. 3A and B). The peptide sequence was 100% conserved in 13 strains, indicating strong purifying selection on this AMP. Additionally, synteny was preserved across host genomes, suggesting a conserved functional genomic region (Fig. 3C). The presence of AMP 8200 in both chromosome and plasmid suggests a potential horizontal transfer event. Thus, we examined whether the 85 kb plasmid carrying the AMP gene was integrated into the chromosome of *P. dorei* strain JR01. Whole-genome alignment indicated that only a 13 kb plasmid fragment (containing the AMP gene) was integrated into the chromosome, rather than the entire plasmid (Supplementary Fig. 5A). To distinguish between plasmid and chromosome origin of transcripts, we re-mapped the RNA-seq reads to the shared 13 kb region. Alignments to the plasmid sequence showed perfect matches, whereas alignments to the chromosomal counterpart contained 17 nucleotide mismatches. This suggests that transcripts mainly originate from the plasmid. Interestingly, when the RNA-seq reads were mapped to the plasmid, we found extensive transcriptional activity across the entire plasmid (95.5% plasmid coverage with mean depth 61.7X) (Supplementary Fig. 5B). These results underscores that mobile genetic elements in the gut can be highly transcribed and may disseminate antimicrobial functions within the community.

AMPs Encoded in Bioactive Phages Suggest the Presence of Phage-Host Dynamics

Three AMPs (AMP 3020; AMP 8681, and AMP 5245) show high sequence similarity to proteins from tailed bacteriophages (Fig. 4), pointing to active phage involvement. For AMP 3020 BLASTP analysis demonstrated a perfect match (100% identity) to proteins from Caudoviricetes phages ctJ1L4 and ctzDR1, and 70.9–74.4% identity to two homologs from *Anaerotignum* strains (Fig. 4A and B). Despite these high sequence similarities, AMP 3020 displayed poor synteny conservation with their surrounding genes in both phages and bacterial genomes (Fig. 4C).

The identical sequence between AMP 3020 and the homolog in phage ctJ1L4 suggests that their overexpression originated from this phage and not from *Anaerotignum* bacteria. To determine the source of expression, we mapped viral-like particle (VLP) DNA-seq reads from the same samples previously published by our laboratory [9] to ctJ1L4 phage genome. After mapping, we found that 70.46% of the genome was covered by virome derived reads, providing strong evidence for the physical presence of this phage as a viral particle, because DNA was isolated from purified VLPs (Supplementary Fig. 6A). Furthermore, the RNA-seq mapping covered 14.06% of the phage genome, indicating transcription of several phage genes (Supplementary Fig. 6B). Together, these data support ctJ1L4 as the likely source of AMP 3020 transcripts rather than *Anaerotignum*.

Given that AMP 3020 also shared 74.36% sequence identity with a protein from *Anaerotignum* sp. MB30-C6 genome, we investigated if the ctJ1L4 phage was integrated into the bacterial chromosome. Only a 372 nucleotides segment of the 32.5 kb ctJ1L4 genome aligned to *Anaerotignum* sp. MB30-C6, corresponding to the AMP 3020 locus (Supplementary Fig. 6C). This suggest that the phage genome was not integrated into the bacterial chromosome. Prophage prediction analysis in the *Anaerotignum* MB30-C6 genome did not identify prophages containing the AMP, further suggesting the bacterial copy is not part of a prophage.

The AMP 8681 was identical (100% sequence identity) to proteins from two Caudoviricetes phages (ctlN07 and ctUyu4) and Clostridiales bacterium KR001 hic 0007, and shows 96.97% identity to proteins from phages ctvmC4 and ctlnE22 (Fig. 4D and E). Despite this high sequence conservation between proteins, synteny around the locus was weakly conserved in both phage and bacterial genomes (Fig. 4F). Given that AMP 8681 had identical homolog proteins in both bacteria and phages, we investigated which of these two genomes could be the potential source of AMP overexpression. Mapping virome-derived reads to ctlN07 phage showed limited coverage (7.18%) (Supplementary Fig. 7A), whereas RNA-seq mapping covered 12.82% of the genome at 52.65X mean depth, indicating transcription despite low DNA abundance(Supplementary Fig. 7B). We evaluated possible phage integration by comparing the 50 kb *ctlN07* genome to the Clostridiales bacterium KR001 hic 0007 genome, and found only a 3-kb shared region (98.10% identity) encompassing the AMP locus, indicating

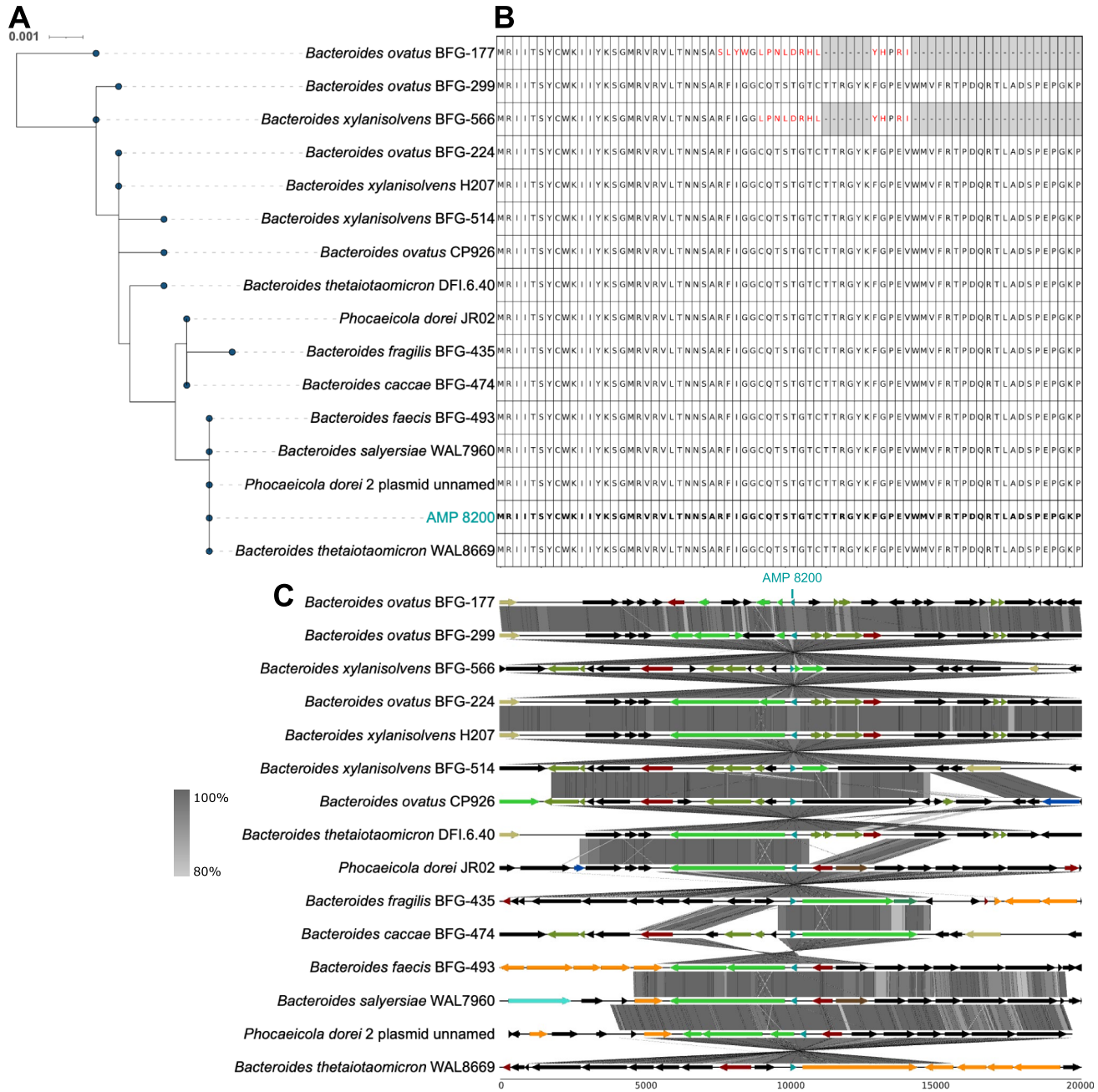


Fig. 3 Molecular phylogeny (transcript), sequence conservation (protein), and conserved synteny (protein) of plasmid-derived AMP 8200. **A** Molecular phylogeny. Neighbor-Joining tree of AMP 8200 homologs. **B** Sequence conservation. Multiple alignment of AMP 8200 homologs; red color indicates the amino acid differences compared to the AMP sequence. **C** Conserved synteny. Gene neighborhoods flanking the AMP 8200 locus across representative genomes. Arrows indicate gene orientation; intergenic distances are to scale. Genes are colored by COG functional category; mobile-element genes are highlighted. The colors for functional categories were as follows: translation, ribosomal structure and biogenesis (blue); RNA processing and modification (forest green); transcription (dark red); replication, recombination and repair (orange); chromatin structure and dynamics (purple); cell cycle control, cell division, and chromosome partitioning (golden yellow); nuclear structure (steel blue); defense mechanisms (firebrick red); signal transduction mechanisms (lime green); cell wall/membrane/envelope biogenesis (dark brown); cell motility (cyan); cytoskeleton (royal blue); extracellular structures (slate gray); intracellular trafficking, secretion, and vesicular transport (magenta); posttranslational modification, protein turnover, and chaperones (dark orange); mobilome elements such as prophages and transposons (olive green); energy production and conversion (tomato red); carbohydrate transport and metabolism (turquoise); amino acid transport and metabolism (deep pink); nucleotide transport and metabolism (violet); coenzyme transport and metabolism (sea green); lipid transport and metabolism (khaki); inorganic ion transport and metabolism (cobalt blue); secondary metabolite biosynthesis, transport, and catabolism (indian red); general function prediction only (dark gray); and unknown function (black)

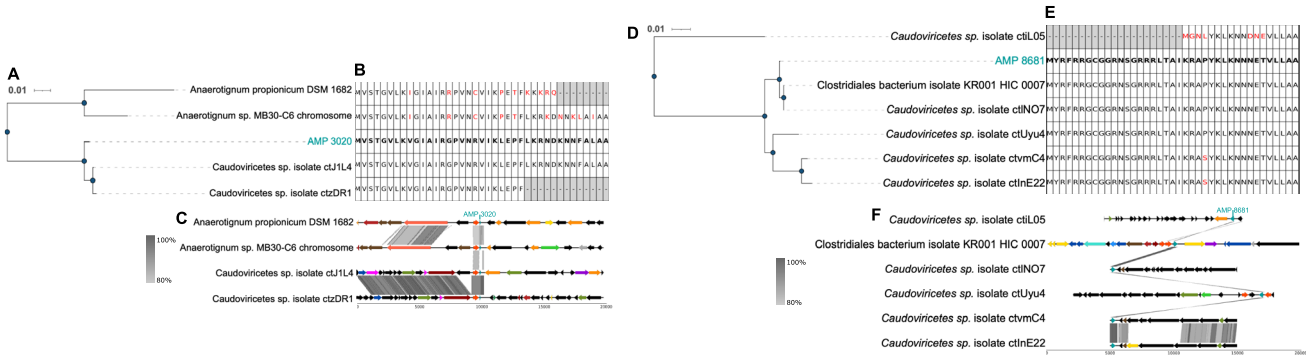


Fig. 4 Molecular phylogeny (transcript), sequence conservation (protein), and conserved synteny (protein) of phage-derived AMP 3020 (**A**, **B** and **C**) and AMP 8681 (**D**, **E** and **F**). Molecular phylogeny. Neighbor-Joining tree of AMP 3020 (**A**) and AMP 8681 (**D**) homologs. Sequence conservation. Multiple alignment of AMP 3020 (**B**) and AMP 8681 (**E**) homologs; red color indicates the amino acid differences compared to the AMP sequence. Conserved synteny. Gene neighborhoods flanking the AMP 3020 (**C**) and AMP 8681 (**F**) locus across representative genomes. Arrows indicate gene orientation; intergenic distances are to scale. Genes are colored by COG functional category; mobile-element genes are highlighted. The colors were represented as detailed in the legend of Fig. 3

no phage integration in the bacterial genome (Supplementary Fig. 7C). Additionally, prophage prediction in the bacterial genome did not identify ctlN07 or other prophages carrying the AMP, suggesting a non-viral origin of this AMP in the bacterial genome. Thus, AMP 8681 overexpression could be originated from either the phage or the bacterium additional functional assays are needed to resolve the primary source.

The AMP 5245 shows 100% identity to multiple phage proteins and to homologs in *Blautia wexlerae* (Fig. 5A and B), with two conserved genomic architectures spanning phages and bacterial contexts (Fig. 5C). We investigated the source of the observed AMP overexpression. VLP DNA-Seq reads covered 88.7% of the cognate phage genome, and RNA-seq reads covered 86.7% at a mean depth 1,777X, indicating an strong transcription of the majority of phage genes (Supplementary Fig. 8A and B). Next, we analyzed if phage genome was integrated into the bacterial genome, and the alignment of the 8-kb human fecal virus clone to *B. wexlerae* DSM 19850 revealed a 3-kb region (98.35% identity), containing the AMP locus (Supplementary Fig. 8C), suggesting a partial integration of phage (39.53% of the phage genome) into the bacterial genome. However, prophage prediction in *B. wexlerae* did not detect any prophage elements matching the human fecal virus clone, arguing against integration. Additionally, prophage prediction did not detect any other prophages containing the AMP. The high percentage of phage genome coverage with VLP DNA and RNA reads therefore implicates the phage as the principal source of AMP 5245 transcripts. Collectively, mapping and synteny analyses support phage origin and active transcription for AMP 3020 and AMP 5245, while AMP 8681 remains ambiguous, highlighting dynamic phage–host interactions and potential bidirectional exchange of AMP genes.

Correlation Analysis links AMPs Expression with Bacterial Taxa

To explore associations between the six overexpressed AMPs and gut microbiota, we correlated the AMP expression with 16S rRNA profiles from the same samples (BioProject PRJNA600247). We observed significant negative correlations in which higher AMP expression coincided with lower taxon abundance (Fig. 6A). Notable inverse correlations included: AMP 5865 with *Anaeroplasma*, *Clostridium*, and *Bacteroides* spp.; AMP 5245 with > 20 taxa, prominently *Akkermansia*, Christensenellaceae, *Moryella*, and *Oscillibacter*; AMP 3076 with Christensenellaceae, *Bilophila*, and multiple Lachnospiraceae lineages; AMP 3096 with Desulfovibrionaceae, *Bilophila wadsworthia*, *Oscillibacter*, and *A. muciniphila*; and AMPs 2526 and AMP 2198 with *A. muciniphila* and *Eubacterium* spp.

Beyond inverse associations, we also identified statistically significant positive correlations between the expression of several AMPs and specific bacterial taxa (Fig. 6B). These likely reflect ecological compatibility, co-expression via shared mobile elements, or shared niche responses. Notable positive correlations included: AMPs 2198 and 2526 with *Colidextribacter*; AMPs 3076, 5245 and 3096 with Candidatus Soleaferrea; AMP 8681 with several *Alistipes* spp. and > 11 taxa; and AMP 3020 with > 21 taxa, including *Adlercreutzia*, *Butyricimonas*, *Dialister invisus*, several *Eubacterium* spp., *Gordonibacter*, *Monoglobus*, Prevotellaceae bacterium, and *Sutterella*.

Taken together, the inverse correlations (often involving gut health–associated taxa (Table 1)) and the positive correlations (frequently implicating obesity-linked lineages (Table 1)) suggest a model in which AMP expression

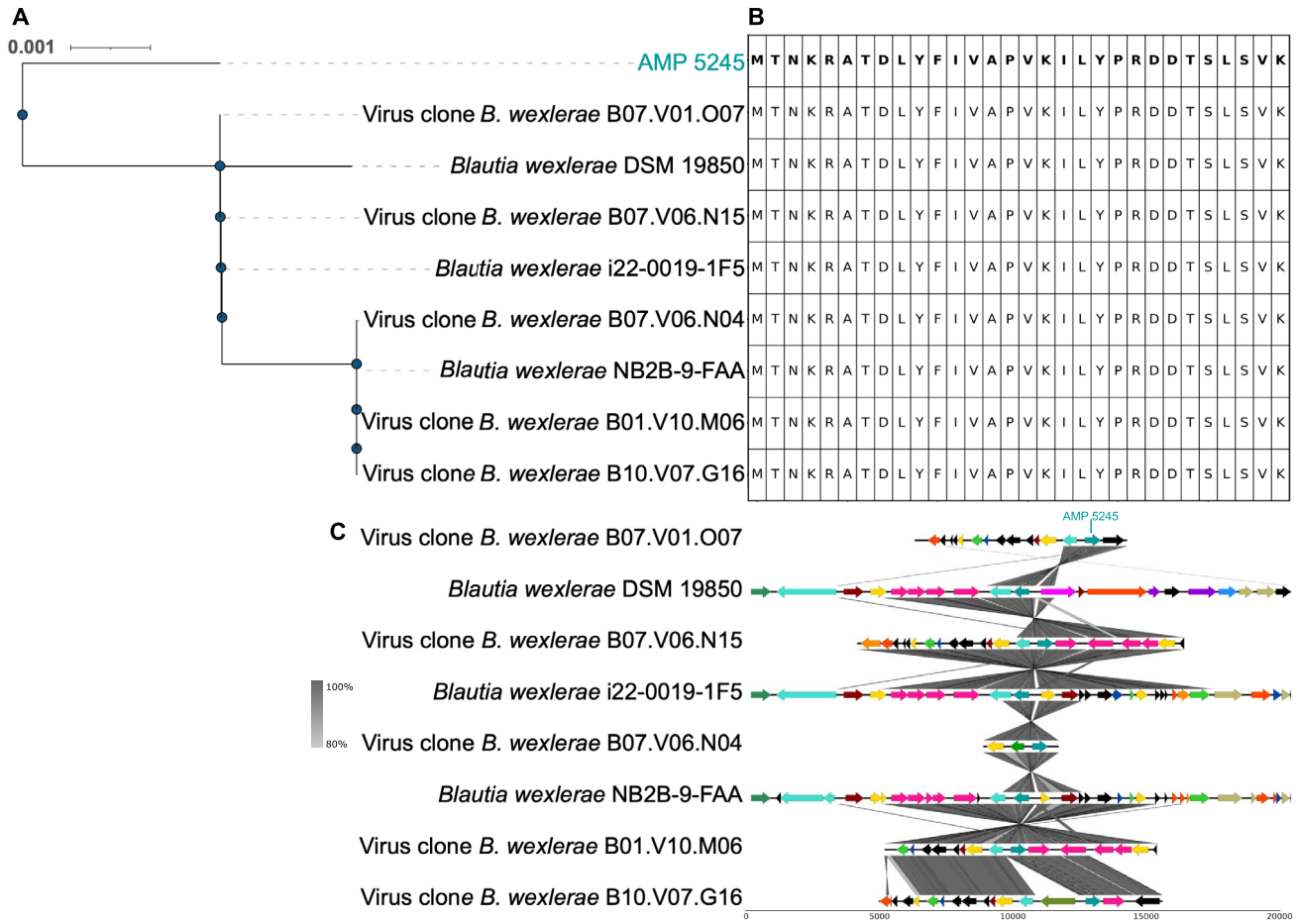


Fig. 5 Molecular phylogeny (transcript), sequence conservation (protein), and conserved synteny (protein) of phage-derived AMP 5245. **A** Molecular phylogeny. Neighbor-Joining tree of AMP 5245 homologs. **B** Sequence conservation. Multiple alignment of AMP 5245 homologs; red color indicates the amino acid differences compared to the AMP sequence. **C** Conserved synteny. Gene neighborhoods flanking the AMP 5245 locus across representative genomes. Arrows indicate gene orientation; intergenic distances are to scale. Genes are colored by COG functional category; mobile-element genes are highlighted. The colors were represented as detailed in the legend of Fig. 3

contributes to microbiome restructuring in obesity/MetS, directly constraining beneficial taxa and indirectly favoring obesity-associated taxa via competitor suppression (Fig. 8). However, these are observational, non-causal associations; given the modest sample size, we consider them exploratory and hypothesis-generating, meriting confirmation in larger cohorts and targeted mechanistic studies.

Experimental Validation of Phage-Encoded AMP 3020: Broad-Spectrum Antibacterial Activity without Detectable T-Cell Toxicity

To test our in-silico predictions, we selected the AMP 3020 for experimental validation based on its confirmed phage origin. We synthesized two variants (Fig. 7A): ADR1 which retains the initiator methionine, and ADR2, which lacks this residue and begins with valine. ADR2 starts with valine, as it has been reported that some peptides undergo post-translational processing when the second residue of a nascent peptide is a short one; here, the second residue was valine [38]. This single amino acid difference also allowed us to assess whether minimal N-terminal changes modulate activity. Both peptides significantly inhibited the growth of Gram-negative bacteria (*Pseudomonas aeruginosa* and *Klebsiella pneumoniae*) and Gram-positive bacteria (*Staphylococcus aureus* and *Streptococcus pneumoniae*) relative to controls (Fig. 7B–E). ADR1 showed greater potency against *P. aeruginosa*, whereas ADR2 was more effictive against *K. pneumoniae*. Both variants exhibited antibacterial activity against *S. pneumoniae*, with only ADR1 inhibited *S. aureus*. Notably, neither peptide impaired primary human T-lymphocyte viability (Fig. 7F), with similar cell death frequencies (0.7–14.4%) to those of untreated controls (Fig. 7F). The absolute cell

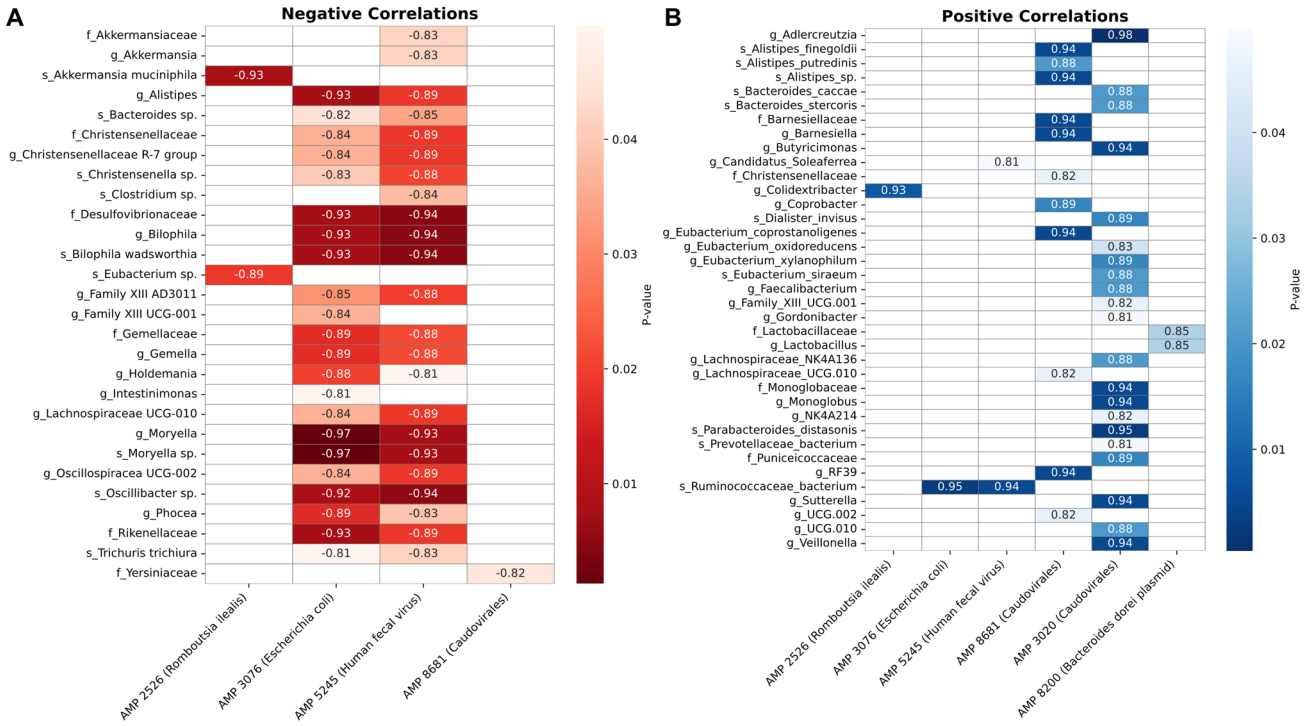


Fig. 6 Heat-map of correlation analysis between bacterial taxa (16S) and AMP expression. **A** Negative correlations. **B** Positive correlations. Columns are AMPs; rows are taxa. Color indicates correlation strength (p-value). Value inner box indicates the correlation value. Only correlations passing significance were displayed ($p < 0.05$)

counts confirmed no significant loss of viable T cells, in contrast to the cytotoxic positive control (PMA/ionomycin) (Supplementary Figs. 9 and 10). Overall, ADR1 and ADR2 exhibited antibacterial activity with no detectable T-cell toxicity under our assay conditions. The absence of the initiator methionine in ADR2 may subtly alter peptide folding or charge distribution and thereby tune target specificity, a hypothesis that merits future mechanistic testing.

Discussion

This study provides a framework for discovering and characterizing AMPs in the human gut microbiome using metatranscriptomics and viromics, representing a promising frontier to understand their role in host-microbiota interactions [10]. Using our methodology described in Fig. 1 we identified 51 high-confidence AMP candidates from 1,095 expressed small open reading frames (smORFs). The limited overlap among the results of the different AMP prediction tools highlights the methodological variability in AMP discovery. Notably, many expressed AMPs were linked to *Faecalibacterium prausnitzii*, a known gut commensal bacterium. Additionally, the identification of AMP-encoding transcripts from plasmids and phages expands the potential origins of AMPs beyond traditional chromosomal sources. One of the key smORF quality criteria proposed by Sberro et al. is to reduce false positives by retaining only smORFs that contain both start and stop codons [19]. However, metatranscriptomic assemblies often yield short contigs representing partial transcripts. To address this, we leveraged homologs transcripts and their predicted proteins to complete the coding sequences of our AMPs. This step was critical for eliminating protein truncation artifacts. Notably, our AMPs aligned to their reference peptides with a mean coverage of 92% and a mean sequence identity of 97%, indicating that only small N- or C-terminal fragments were missing in the RNAseq.

The intestinal epithelial interface contains AMPs of both host and microbial origin, which play a crucial role in shaping the surrounding microbiota [39]. Consistent with emerging evidence that bioactive peptides can remodels gut microbiota [40], our data suggests that AMPs may contribute to the microbiota alterations observed in obesity and metabolic syndrome [8]. Notably, the two phage-derived AMP 3020 variants (ADR1 and ADR2) displayed selective antibacterial activity while remaining non-cytotoxic to primary human T-cell subsets (CD3 +, CD4 +, and CD8 +). This profile aligns them with host defense peptides (HDPs), a broader class that includes bacteriophage-encoded peptides

Table 1 Correlations between gut-expressed antimicrobial peptides (AMPs) and bacterial taxa

AMP (ID)	Correlation (direction)	Correlated taxa (Family/Genus/species)	Healthy or Obesity -related association	References
AMP 2526	↓ Negative	*A. muciniphila*	Healthy linked	[52, 61]
AMP 2526	↓ Negative	*Eubacterium spp.*	Mixed	[63, 64]
AMP 3076 and AMP 5245	↓ Negative	*Bacteroides, Oscillibacter, Alistipes, Holdemia, Christensenella spp, Christensenellaceae R-7 group, Lachnospiraceae UCG-010*	Healthy linked	[54–59, 65–68]
AMP 3076	↓ Negative	*Intestimonas*	Healthy linked	[69]
AMP 5245	↓ Negative	*Clostridium spp.*	Mixed	[70, 71]
AMP 5245	↓ Negative	*Akkermansia*	Healthy linked	[52, 72]
AMP 8681	↓ Negative	*Yersiniaceae*	Obesity linked	[73]
AMP 2526	↑ Positive	*Colidextribacter*	Obesity linked	[53, 74]
AMP 3076 and AMP 5245	↑ Positive	*Ruminococcaceae bacterium*	Obesity linked	[60]
AMP 5245	↑ Positive	*Candidatus Soleaferrea*	Mixed	[61]
AMP 8681	↑ Positive	*Alistipes spp., Barnesiella, Coprobacter, Eubacterium coprostanoligenes, Lachnospiraceae UCG-010, RF39*	Healthy linked	[54–59, 65–68]
AMP 3020	↑ Positive	*Adlercreutzia, Dialister invisus, Eubacterium oxidoreducens, Monoglobus, Family XIII UCG-001, Sutterella, Veillonella*	Obesity linked	[75–81]
AMP 3020	↑ Positive	*Bacteroides caccae, Butyricimonas, Prevotellaceae bacterium, Veillonella*	Mixed	[82–84]
AMP 3020	↑ Positive	*Bacteroides stercoris, Eubacterium coprostanoligenes, Eubacterium xylanophilum, Eubacterium siraeum, Faecalibacterium, Parabacteroides distasonis*	Healthy linked	[63, 66, 85–88]
AMP 8200	↑ Positive	*Lactobacillus*	Healthy linked	[89]

The table lists AMP identifiers (AMP ID), the direction of the association (Direction; ↑ positive, ↓ negative) between AMP expression and the relative abundance of each taxon across samples, the most specific taxonomic label available (Correlated taxa; Family/Genus/species), and a literature-based summary of whether each taxon is typically linked to healthier or obesity-related states (Healthy or Obesity-related association). Mixed indicates heterogeneous or context-dependent evidence acrossstudies. Joint entries (e.g., "AMP 3076 and AMP 5245") indicate that both AMPs share the same direction of association with the listed taxon

capable of modulating inter-microbial competition and host-microbiota interactions [41, 42]. The absence of T cell cytotoxicity emphasizes the immunological neutrality of these peptides in vitro, suggesting some phage encoded AMPs produced within the gut can reshaped bacterial communities without directly harming host. These findings advance our understanding of virome-microbe crosstalk and suggest that phage-derived AMPs may influence ecosystem composition while preserving host immune balance [43, 44]. Moreover, the obesity-associated upregulation of several AMPs and their negative correlations with taxa linked to metabolic health (Table 1) is consistent with AMP-mediated pressures, suggesting their potential role favoring dysbiosis linked to obesity and MetS (Fig. 8). While these associations are not evidence of causality, they motivate mechanistic studies (e.g., gnotobiotic models and in situ peptide perturbations) to test whether phage-encoded AMPs can drive microbiota shifts relevant to metabolic disease.Differential expression analysis comparing obesity with obesity and metabolic syndrome identified six AMPs that were significantly overexpressed, originating from chromosomal, plasmid, and phage genomic contexts. This supports the idea that the host metabolic status influences microbial gene expression for AMP production. Importantly, all sixAMPs were also detected in 372 samples from an independent gut metatranscriptome dataset, suggesting their commonality and ecological significance in the human gut microbiome, rather than being unique to our cohort. Their widespread prevalence argues for ecological relevance and nominates them as candidate biomarkers of microbiome perturbations associated with metabolic disease (Fig. 8).

We identified three AMPs with close homologs in phages, an underexplored reservoir of antimicrobial compounds. AMP 3020 shared 98% identity with a protein from the Caudoviricetes phage ctJ1L4. DNA VLP-virome read mapping suggested the presence of ctJ1L4 virions and metatranscriptome read mapping shows an active transcription of multiple phage genes, including the AMP. Synthetic AMP 3020 inhibited both Gram-positive and Gram-negative bacteria, consistent with broad-spectrum

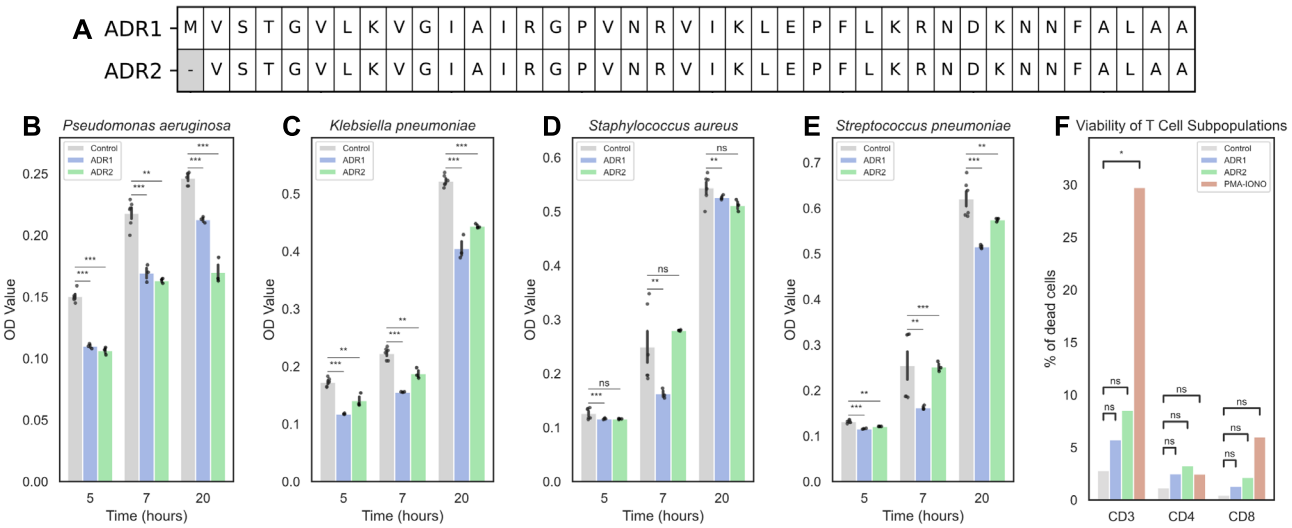


Fig. 7 In-vitro characterization of phage-encoded AMP 3020. **A** Amino acid sequences of AMP 3020 (ADR1) and its variant (ADR2). (B-E) Time-course antimicrobial activity against **B**) *Pseudomonas aeruginosa*, **C**) *Klebsiella pneumoniae*, **D**) *Staphylococcus aureus*, and **E**) *Streptococcus pneumoniae*. **F**) Viability of T cell subpopulations after 24 h exposure to 20 µg of ADR1 or ADR2. Bars show mean ± standard deviation. Statistical significance was assessed using two-sided t-tests: ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$

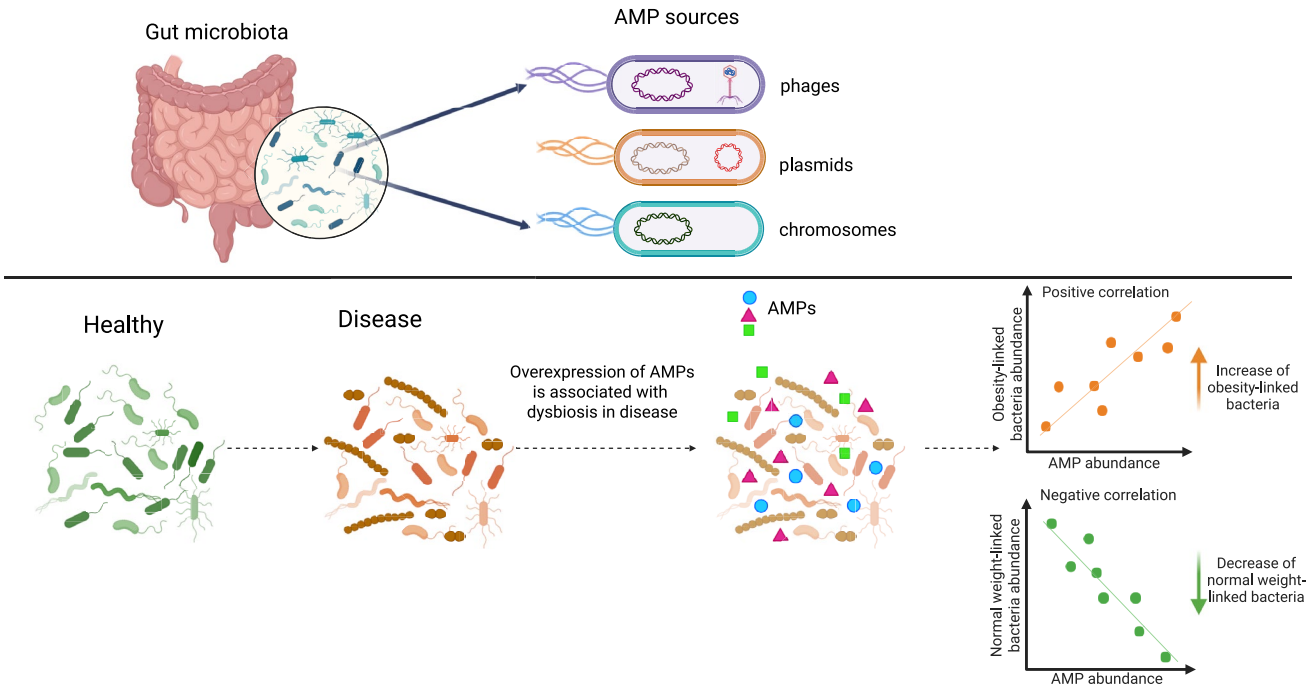


Fig. 8 Gut-expressed antimicrobial peptides (AMPs) originated from bacterial chromosomes, plasmids, and phages. AMP over-expression negatively correlates with healthy-associated taxa and positively correlates with obesity-linked taxa. The microbiota dysbiosis observed in obesity and Metabolic syndrome can be associated with AMP overexpression. Icons indicate AMP source; arrows show correlation direction. Created in BioRender: https://BioRender.com/w7dlcld

activity. Although related proteins exist in two bacterial genomes, we found no evidence of ctJ1L4 integration or prophages carrying the AMP gene, suggesting that in vivo AMP trasncription stemmed primarily from lytic phage activity, which was suggested by DNA VLP-virome read mapping. Removal of the N-terminal Methionine altered

antibacterial activity, reflecting the evolutionary adaptability of phage-derived peptides. Methionine is essential for stabilizing protein structures and may also act as a regulatory switch through reversible redox reactions [45]. Overall, these findings expand the functional repertoire of gut phages, particularly those within the dominant Caudovirales lineage found in the human gut virome [9] and reinforce phages as a rich, underrecognized source of AMPs capable of reshaping microbioal communities (Fig. 8). Future mechanistic studies in phage–host systems will be essential to establish causality and therapeutic potential.

AMPs 8681 and 5245 show 100% amino acid identity between their phage- and bacterium-encoded homologs. DNA VLP virome read mapping suggests the phage presence. Additionally, metatranscriptomic data showed active transcription of several phage genes, suggesting that phages substantially contributed of AMP transcripts in vivo, although bacterial loci may also play a role in the transcription. Prophage scans show that both chromosomal AMP loci were located outside predicted prophage regions, suggesting that the integration of both AMPs into the bacterial genomes was not related to a prophage. Interestingly, the expression of AMP 5245 negatively correlated with Moryella and Eubacterium species, known butyrate producers with anti-inflammatory effects, suggesting that this AMP may antagonize taxa linked to intestinal health [46, 47]. We also detected a plasmid-encoded homolog protein of AMP 8200, a plasmid described in *Phocaeicola dorei* [48]. The peptide exhibits 100% sequence identity and conserved synteny across additional *Phocaeicola* species. RNA-seq read mapping revealed robust transcription across the entire plasmid, suggesting the plasmid as the primary source of AMP transcripts, reinforcing the concept that plasmids serve as reservoirs of AMPs in the gut microbiome [49]. Together, these observations highlight the potential for plasmids and phages to spread competitive traits such as AMPs via horizontal transfer.

Additional AMPs 3076 and 2526, shared homolog proteins across diverse gut bacterial commensals, including *Escherichia coli* and *Romboutsia* spp., suggesting their widespread distribution in the intestinal microbiota. These AMPs demonstrated high sequence and synteny conservation across bacterial genomes. Expression of these AMPs negatively correlated with microbial groups associated with metabolic health, such as *Akkermansia muciniphila and Christensenellaceae,* suggesting a potential link with microbial dysbiosis observed in obesity and metabolic syndrome [50, 51].

Given our modest sample size, the correlations between AMPs and disease-associated taxa should be viewed as exploratory. Interestingly, AMP 2526 (overexpressed in OMS) was negatively correlated with *A. muciniphila*, taxa frequently linked to metabolic health (Table 1) [52]. By contrast, AMP 2526 correlated positively with Colidextribacter, a genus increased in obese mice and associated with higher body/liver weight and circulating lipids (Table 1) [53]. Overall, the correlation pattern links AMP 2526 to an obesogenic microbiome: it is negatively associated with taxa typically tied to metabolic health and positively associated with obesogenic taxa, suggesting AMP 2526 may mark, or contribute to, dysbiosis and adverse metabolic status in OMS, but these are associations only. Validation in larger, independent cohorts and mechanistic models is required before inferring causality. Another example was the AMPs 3076 and 5245 negatively correlated with several healthy linked taxa, such as *Bacteroides* and *Oscillibacter* (Table 1). Furthermore, these AMPs also were negatively *Alistipes* which was enriched in normal-weight individuals compared to subjects with obesity [54, 55]; *Holdemania* which is is more abundant in normal-weight children [56]; and Christensenella spp., that reduce diet-induced weight gain, dyslipidemia, and hepatic steatosis in mice [57]. These AMPs also were negatively correlated with Christensenellaceae R-7 group, Lachnospiraceae UCG-010, and Rikenellaceae which are linked to healthier metabolic profiles (Table 1) [58, 59]. This suggests the possibility that both AMPs may constrain these beneficial taxa. Nonetheless, these observations require validation in larger, independent cohorts and experimental systems before inferring causality. Another example was the AMPs and 5245 negatively correlated with several healthy linked taxa, such as *Bacteroides* and *Oscillibacter* (Table 1). Furthermore, these AMPs also were negatively *Alistipes* which was enriched in normal-weight individuals compared to subjects with obesity [54, 55]; *Holdemania* which is is more abundant in normal-weight children [56]; and Christensenella spp., that reduce diet-induced weight gain, dyslipidemia, and hepatic steatosis in mice [57]. These AMPs also were negatively correlated with Christensenellaceae R-7 group, Lachnospiraceae UCG-010, and Rikenellaceae which are linked to healthier metabolic profiles (Table 1) [58, 59]. This suggest the possibility that both AMPs may constrain these beneficial taxa. Nonetheless, these observations require validation in larger, independent cohorts and experimental systems before inferring causality.

By contrast, AMPs 3076, and 5245, correlated positively with a Ruminococcaceae bacterium; a family shows context-dependent links to obesity, enriched in some cohorts, reduced or unchanged in others [60]. AMP 5245 also positively correlated with Candidatus, whose metabolic associations are mixed—reported as protective in some settings but linked to cardiometabolic risk in others [61]. Taken together, AMPs 3076 and 5245 align with depletion of taxa linked to leanness and some increases in taxa with mixed/obesogenic associations. Unlike the other overexpressed AMPs, AMP 8681 shows the opposite pattern: it associates negatively

with taxa linked to adverse metabolic states and positively with taxa tied to metabolic health. Specifically, it correlates inversely with Yersiniaceae (often elevated in obesity [62] and positively with *Alistipes* spp., *Barnesiella*, *Coprobacter*, *Eubacterium coprostanoligenes*, Lachnospiraceae UCG-010, and the RF39 clade—groups typically more abundant in lean or metabolically healthy cohorts, as early described. These are correlations only and likely reflect context-dependent shifts rather than direct AMP effects.

Overall, these patterns suggest potential coexistence or co-regulation, which may arise from: (i) ecological compatibility—AMP activity suppressing competitors and indirectly favoring non-target taxa; (ii) co-expression driven by mobile elements (plasmids/phages) carrying both AMP genes and host-maintenance functions; and/or (iii) shared responses to niche conditions (e.g., substrates, stress signals) that simultaneously elevate AMP transcription and the abundance of certain lineages. These positive co-occurrence patterns, together with the inverse correlations reported above, point to structured community responses to AMP expression—encompassing both potential constraint of sensitive taxa and concomitant expansion of compatible or co-regulated lineages. Given the modest sample size, we interpret these findings as exploratory and hypothesis-generating; they nominate specific AMP–taxon pairs for targeted validation in larger cohorts and mechanistic experiments.

Metatranscriptomic profiling provides a dynamic map of microbiome activity, revealing which genes are actively expressed under conditions such as obesity and metabolic syndrome. This approach differs from traditional metagenomics, which only lists gene presence. Our analysis identified actively expressed AMPs, helping differentiate between latent genetic potential and those that actively influence host microbiota and disease associations. Our findings highlight that gut-expressed AMPs were derived from diverse genomic sources, including bacteria, plasmids, and phages, indicating their significant ecological roles in gut microbial dynamics. This proof-of-concept study shows that metatranscriptomic data can uncover relevant, expressed AMPs implicated in the gut microbiome regulation. The differential expression of these AMPs associated with disease, the negative and positive correlations with microbial taxa, along with their antimicrobial properties that do not affect host immunity points to their importance in shaping the gut microbiota. Moreover, the presence of mobile genetic elements, such as plasmids and phages, reinforces the need to rethink the role of phages in gut ecology, acting not only as predators but also as regulators. While our study was exploratory, it establishes metatranscriptomics and viromics as a practical framework to uncover relevant, expressed AMPs implicated in microbiome regulation (Fig. 8). This work sets the stage for future studies on AMPs and their therapeutic potential in microbiome-targeted treatments in obesity and MetS-associated dysbiosis.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00248-025-02620-2.

Acknowledgements L.G.B. thanks to the Doctoral Biochemical Sciences Program at IBt UNAM and CONACyT for doctoral fellowship C.V.U.: 887285. Also, F.C.G. would thank the Estancias posdoctorales por México 2022 program (C.V.U.: 443238). This research was funded by CONACyT grant Ciencia de Frontera-2019-263986 and by DGAPA PAPIIT UNAM (IN219723). We thank M.T.I Juan Manuel Hurtado Ramírez for informatics technical support. Also, the authors would like to thank the "Unidad Universitaria de Secuenciación Masiva y Bioinformática" of the "Laboratorio Nacional de Apoyo Tecnológico a las Ciencias Genómicas," UNAM, especially to Ricardo Alfredo Grande Cano and Lizeth A. Matías Valdez for the technical sequencing support.

Authors Contributions Conceived of or designed study: LGB, FCG, SB, and AOL; Analyzed Data: LGB, FCG, AA, GLL, CAG, FS, RM, GC, GPE, SCQ, and AOL; Formal analysis: LGB, FCG, AA, GLL, CAG, FS, GC, GPE, SCQ, and AOL; Contributed new methods or models: LGB, FCG, AA, GLL, CAG, FS, GC, GPE, SCQ, and AOL; Wrote the paper: LGB, FCG, AA, GLL, CAG, FS, GC, GPE, SCQ, and AOL. Funding acquisition: FCG, SCQ, AOL. All authors read and approved the final manuscript.

Funding This research was funded by *CONACyT grant Ciencia de Frontera-2019–263986 and by DGAPA PAPIIT UNAM (IN219723).* L.G.B. was supported by the Doctoral Biochemical Sciences Program at IBt UNAM and CONACyT with the doctoral fellowship C.V.U.: 778192. F.C.G. was supported by the *Estancias posdoctorales por México 2022 program (C.V.U.: 443238).*

Data Availability The data that support the findings of this study are openly available in NCBI GEO repository with accession number GSE143207 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE143207) and the NCBI BioProject: PRJNA600247 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA600247) and PRJNA646512 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA646512). All the code used for this project was deposited in this GitHub repository: (https:/github.com/LuiguiGallardo/amps_microbiome_2025). Requests for additional material should be made to the corresponding author.

Declarations

Competing Interest The authors declare no competing interests.

References

1. Biro FM, Wien M (2010) Childhood obesity and adult morbidities. Am J Clin Nutr 91(5):1499S-1505S. https://doi.org/10.3945/ajcn.2010.28701B
2. Franks PW, Hanson RL, Knowler WC, Sievers ML, Bennett PH, Looker HC (2010) Childhood obesity, other cardiovascular risk factors, and premature death. N Engl J Med 362(6):485–493. https://doi.org/10.1056/NEJMoa0904130
3. Perichart-Perera O, Balas-Nakash M, Schiffman-Selechnik E, Barbato-Dosal A, Vadillo-Ortega F (2007) Obesity increases metabolic syndrome risk factors in school-aged children from an urban school in Mexico City. J Am Diet Assoc 107(1):81–91. https://doi.org/10.1016/j.jada.2006.10.011
4. Al-Hamad D, Raman V (2017) Metabolic syndrome in children and adolescents. Transl Pediatr. https://doi.org/10.21037/tp.2017.10.02
5. de Ferranti SD, Gauvreau K, Ludwig DS, Neufeld EJ, Newburger JW, Rifai N (2004) Prevalence of the metabolic syndrome in American adolescents. Circulation 110(16):2494–2497. https://doi.org/10.1161/01.CIR.0000145117.40114.C7
6. Evia-Viscarra ML, Rodea-Montero ER, Apolinar-Jiménez E, Quintana-Vargas S (2013) Metabolic syndrome and its components among obese (BMI ≥95th) Mexican adolescents. Endocr Connect 2(4):208–215. https://doi.org/10.1530/EC-13-0057
7. Shamah-Levy T et al (2023) Prevalencias de sobrepeso y obesidad en población escolar y adolescente de México. Ensanut Continua 2020–2022. Salud Pública México 65:s218–s224. https://doi.org/10.21149/14762
8. Gallardo-Becerra L et al (2020) Metatranscriptomic analysis to define the Secrebiome, and 16S rRNA profiling of the gut microbiome in obesity and metabolic syndrome of Mexican children. Microb Cell Fact 19(1):61. https://doi.org/10.1186/s12934-020-01319-y
9. Bikel S et al (2021) Gut dsDNA virome shows diversity and richness alterations associated with childhood obesity and metabolic syndrome. iScience 24(8):102900. https://doi.org/10.1016/j.isci.2021.102900
10. Gallardo-Becerra L, Cervantes-Echeverría M, Cornejo-Granados F, Vazquez-Morado LE, Ochoa-Leyva A (2023) Perspectives in searching antimicrobial peptides (AMPs) produced by the microbiota. Microb Ecol 87(1):8. https://doi.org/10.1007/s00248-023-02313-8
11. Izadpanah A, Gallo RL (2005) Antimicrobial peptides. J Am Acad Dermatol 52(3):381–390. https://doi.org/10.1016/j.jaad.2004.08.026
12. Ageitos JM, Sánchez-Pérez A, Calo-Mata P, Villa TG (2017) "Antimicrobial peptides (AMPs): Ancient compounds that represent novel weapons in the fight against bacteria." Biochem Pharmacol 133:117–138. https://doi.org/10.1016/j.bcp.2016.09.018
13. Torres MDT, Melo MCR, Flowers L, Crescenzi O, Notomista E, De La Fuente-Nunez C (2021) Mining for encrypted peptide antibiotics in the human proteome. Nat Biomed Eng 6(1):67–75. https://doi.org/10.1038/s41551-021-00801-1
14. Pizzo E, Cafaro V, Di Donato A, Notomista E (2018) Cryptic antimicrobial peptides: identification methods and current knowledge of their immunomodulatory properties. Curr Pharm Des 24(10):1054–1066. https://doi.org/10.2174/1381612824666180327165012
15. Nolan EM, Walsh CT (2009) How nature morphs peptide scaffolds into antibiotics. ChemBioChem 10(1):34–53. https://doi.org/10.1002/cbic.200800438
16. Singh N, Abraham J (2014) Ribosomally synthesized peptides from natural sources. J Antibiot (Tokyo) 67(4):277–289. https://doi.org/10.1038/ja.2013.138
17. Torres MDT et al (2024) "Mining human microbiomes reveals an untapped source of peptide antibiotics." Cell. https://doi.org/10.1016/j.cell.2024.07.027
18. Santos-Júnior CD et al (2024) Discovery of antimicrobial peptides in the global microbiome with machine learning. Cell. https://doi.org/10.1016/j.cell.2024.05.013
19. Sberro H et al (2019) Large-scale analyses of human microbiomes reveal thousands of small, novel genes. Cell 178(5):1245-1259.e14. https://doi.org/10.1016/j.cell.2019.07.016
20. Nakatsuji T et al (2017) Antimicrobials from human skin commensal bacteria protect against *Staphylococcus aureus* and are deficient in atopic dermatitis. Sci Transl Med 9(378):eaah4680. https://doi.org/10.1126/scitranslmed.aah4680
21. Ma Y et al (2022) Identification of antimicrobial peptides from the human gut microbiome using deep learning. Nat Biotechnol 40(6):6. https://doi.org/10.1038/s41587-022-01226-0
22. Sun D et al (2021) Angiogenin maintains gut microbe homeostasis by balancing α-proteobacteria and lachnospiraceae. Gut 70(4):666–676. https://doi.org/10.1136/gutjnl-2019-320135
23. González-Dávila P et al (2022) Catestatin selects for colonization of antimicrobial-resistant gut bacterial communities. ISME J 16(8):1873–1882. https://doi.org/10.1038/s41396-022-01240-9
24. Magana M et al (2020) The value of antimicrobial peptides in the age of resistance. Lancet Infect Dis 20(9):e216–e230. https://doi.org/10.1016/S1473-3099(20)30327-3
25. Deng Y, Yang S, Zhang L, Chen C, Cheng X, Hou C (2024) Chronic bee paralysis virus exploits host antimicrobial peptides and alters gut microbiota composition to facilitate viral infection. ISME J 18(1):wrae051. https://doi.org/10.1093/ismejo/wrae051
26. Garcia-Gutierrez E, Mayer MJ, Cotter PD, Narbad A (2019) Gut microbiota as a source of novel antimicrobials. Gut Microbes 10(1):1–21. https://doi.org/10.1080/19490976.2018.1455790
27. Kommineni S et al (2015) Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract. Nature 526(7575):719–722. https://doi.org/10.1038/nature15524
28. Nakazono K et al (2022) Complete sequences of epidermin and nukacin encoding plasmids from oral-derived *Staphylococcus epidermidis* and their antibacterial activity. PLoS ONE 17(1):e0258283. https://doi.org/10.1371/journal.pone.0258283
29. Du J et al (2023) Enhancing bacteriophage therapeutics through in situ production and release of heterologous antimicrobial effectors. Nat Commun 14(1):4337. https://doi.org/10.1038/s41467-023-39612-0
30. Tyagi JL, Gupta P, Ghate MM, Kumar D, Poluri KM (2024) Assessing the synergistic potential of bacteriophage endolysins and antimicrobial peptides for eradicating bacterial biofilms. Arch Microbiol 206(6):272. https://doi.org/10.1007/s00203-024-04003-6
31. Santos-Júnior CD, Pan S, Zhao X-M, Coelho LP (2020) Macrel: antimicrobial peptide screening in genomes and metagenomes. PeerJ 8:e10555. https://doi.org/10.7717/peerj.10555
32. P. Bhadra, J. Yan, J. Li, S. Fong, and S. W. I. Siu, (2018) "AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest". *Sci Rep*, 8(1), https://doi.org/10.1038/s41598-018-19752-w
33. Veltri D, Kamath U, Shehu A (2018) Deep learning improves antimicrobial peptide recognition. Bioinformatics 34(16):2740–2747. https://doi.org/10.1093/bioinformatics/bty179
34. Wang G, Li X, Wang Z (2016) APD3: the antimicrobial peptide database as a tool for research and education. Nucleic Acids Res 44(D1):D1087–D1093. https://doi.org/10.1093/nar/gkv1278
35. Yao L et al (2025) dbAMP 3.0: updated resource of antimicrobial activity and structural annotation of peptides in the post-pandemic

era. Nucleic Acids Res 53(D1):D364–D376. https://doi.org/10.1093/nar/gkae1019
36. Ma T et al (2025) DRAMP 4.0: an open-access data repository dedicated to the clinical translation of antimicrobial peptides. Nucleic Acids Res 53(D1):D403–D410. https://doi.org/10.1093/nar/gkae1046
37. Abu-Ali GS et al (2018) Metatranscriptome of human faecal microbial communities in a cohort of adult men. Nat Microbiol 3(3):356–366. https://doi.org/10.1038/s41564-017-0084-4
38. Ben-Bassat A, Bauer K (1987) Amino-terminal processing of proteins. Nature 326(6110):315–315. https://doi.org/10.1038/326315a0
39. Zong X, Fu J, Xu B, Wang Y, Jin M (2020) Interplay between gut microbiota and antimicrobial peptides. Anim Nutr 6(4):389–396. https://doi.org/10.1016/j.aninu.2020.09.002
40. Xu F et al (2016) GLP-1 receptor agonist promotes brown remodelling in mouse white adipose tissue through SIRT1. Diabetologia 59(5):1059–1069. https://doi.org/10.1007/s00125-016-3896-5
41. Haney EF, Straus SK, Hancock REW (2019) Reassessing the host defense peptide landscape. Front Chem. https://doi.org/10.3389/fchem.2019.00043
42. Peng S-Y, You R-I, Lai M-J, Lin N-T, Chen L-K, Chang K-C (2017) Highly potent antimicrobial modified peptides derived from the *Acinetobacter baumannii* phage endolysin LysAB2. Sci Rep 7(1):1. https://doi.org/10.1038/s41598-017-11832-7
43. Ritz NL et al (2024) The gut virome is associated with stress-induced changes in behaviour and immune responses in mice. Nat Microbiol 9(2):359–376. https://doi.org/10.1038/s41564-023-01564-y
44. Wei Y (2024) Bacteriophages: a double-edged sword in the gastrointestinal tract. Front Microbiomes 3:1450523. https://doi.org/10.3389/frmbi.2024.1450523
45. Aledo JC (2019) Methionine in proteins: the cinderella of the proteinogenic amino acids. Protein Sci 28(10):1785–1796. https://doi.org/10.1002/pro.3698
46. Lu H et al (2022) Butyrate-producing *Eubacterium rectale* suppresses lymphomagenesis by alleviating the TNF-induced TLR4/MyD88/NF-κB axis. Cell Host Microbe 30(8):1139-1150.e7. https://doi.org/10.1016/j.chom.2022.07.003
47. Sung MM et al (2016) Improved glucose homeostasis in obese mice treated with resveratrol is associated with alterations in the gut microbiome. Diabetes 66(2):418–425. https://doi.org/10.2337/db16-0680
48. Kijner S, Ennis D, Shmorak S, Florentin A, Yassour M (2024) CRISPR-Cas-based identification of a sialylated human milk oligosaccharides utilization cluster in the infant gut commensal *Bacteroides dorei*. Nat Commun 15(1):105. https://doi.org/10.1038/s41467-023-44437-y
49. Fernández-Pato A et al (2025) Early-life development of the gut virome and plasmidome: a longitudinal study in cesarean-born infants. Cell Rep. https://doi.org/10.1016/j.celrep.2025.115731
50. Waters JL, Ley RE (2019) "The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health." BMC Biol 17:83. https://doi.org/10.1186/s12915-019-0699-4
51. Everard A et al (2013) Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. Proc Natl Acad Sci U S A 110(22):9066–9071. https://doi.org/10.1073/pnas.1219451110
52. Xu Y, Wang N, Tan H-Y, Li S, Zhang C, Feng Y (2020) Function of *Akkermansia muciniphila* in obesity: interactions with lipid metabolism, immune response and gut systems. Front Microbiol. https://doi.org/10.3389/fmicb.2020.00219
53. Yu Z, Yu X-F, Kerem G, Ren P-G (2022) Perturbation on gut microbiota impedes the onset of obesity in high fat diet-induced mice. Front Endocrinol. https://doi.org/10.3389/fendo.2022.795371
54. Thingholm LB et al (2019) Obese individuals with and without type 2 diabetes show different gut microbial functional capacity and composition. Cell Host Microbe 26(2):252-264.e10. https://doi.org/10.1016/j.chom.2019.07.004
55. Duan M, Wang Y, Zhang Q, Zou R, Guo M, Zheng H (2021) Characteristics of gut microbiota in people with obesity. PLoS ONE 16(8):e0255446. https://doi.org/10.1371/journal.pone.0255446
56. Orbe-Orihuela YC et al (2022) Association of gut microbiota with dietary-dependent childhood obesity. Arch Med Res 53(4):407–415. https://doi.org/10.1016/j.arcmed.2022.03.007
57. Mazier W et al (2021) A new strain of *Christensenella minuta* as a potential biotherapy for obesity and associated metabolic diseases. Cells 10(4):823. https://doi.org/10.3390/cells10040823
58. Ahmad MA, Karavetian M, Moubareck CA, Wazz G, Mahdy T, Venema K (2023) Association of the gut microbiota with clinical variables in obese and lean Emirati subjects. Front Microbiol. https://doi.org/10.3389/fmicb.2023.1182460
59. Tavella T et al (2021) Elevated gut microbiome abundance of Christensenellaceae, Porphyromonadaceae and Rikenellaceae is associated with reduced visceral adipose tissue and healthier metabolic profile in Italian elderly. Gut Microbes 13(1):1880221. https://doi.org/10.1080/19490976.2021.1880221
60. Novikova E, Belkova N, Pogodina A, Romanitsa A, Rychkova L (2022) ODP377 adolescents with obesity breastfed until four months age have high abundance of Ruminococcaceae bacteria in gut microbiota. J Endocr Soc 6(Supplement_1):A599. https://doi.org/10.1210/jendso/bvac150.1242
61. X. Lin and K. Xu, "Causal Link Between Gut Microbiota and Arterial Occlusive Diseases: Insights From Mendelian Randomization and Bioinformatics Analysis," *Catheter. Cardiovasc. Interv.*, https://doi.org/10.1002/ccd.70047.
62. Palmas V et al (2021) Gut microbiota markers associated with obesity and overweight in Italian adults. Sci Rep 11(1):5532. https://doi.org/10.1038/s41598-021-84928-w
63. Lai X et al (2024) *Eubacterium siraeum* suppresses fat deposition via decreasing the tyrosine-mediated PI3K/AKT signaling pathway in high-fat diet-induced obesity. Microbiome 12(1):223. https://doi.org/10.1186/s40168-024-01944-4
64. Dong TS et al (2020) A distinct brain-gut-microbiome profile exists for females with obesity and food addiction. Obesity 28(8):1477–1486. https://doi.org/10.1002/oby.22870
65. Wu L, Park S-H, Kim H (2023) "Direct and Indirect Evidence of Effects of Bacteroides spp. on Obesity and Inflammation." Int J Mol Sci 25(1):438. https://doi.org/10.3390/ijms25010438
66. Ryu SW et al (2023) "Anti-obesity activity of human gut microbiota Bacteroides stercoris KGMB02265." Arch Microbiol 206(1):19. https://doi.org/10.1007/s00203-023-03750-2
67. Orsso CE et al (2021) Composition and functions of the gut microbiome in pediatric obesity: relationships with markers of insulin resistance. Microorganisms 9(7):1490. https://doi.org/10.3390/microorganisms9071490
68. Alcazar M et al (2022) Gut microbiota is associated with metabolic health in children with obesity. Clin Nutr 41(8):1680–1688. https://doi.org/10.1016/j.clnu.2022.06.007
69. Rampanelli E et al (2025) Gut bacterium *Intestinimonas butyriciproducens* improves host metabolic health: evidence from cohort and animal intervention studies. Microbiome 13(1):15. https://doi.org/10.1186/s40168-024-02002-9
70 Woting A, Pfeiffer N, Loh G, Klaus S, Blaut M (2014) Clostridium ramosum Promotes High-Fat Diet-Induced Obesity in Gnotobiotic Mouse Models. mBio 5(5):e01530-14. https://doi.org/10.1128/mBio.01530-14

71. Liao J et al (2023) *Clostridium butyricum* strain CCFM1299 reduces obesity via increasing energy expenditure and modulating host bile acid metabolism. Nutrients 15(20):4339. https://doi.org/10.3390/nu15204339
72. Zhou Q et al (2020) Gut bacteria *Akkermansia* is associated with reduced risk of obesity: evidence from the American Gut Project. Nutr Metab 17:90. https://doi.org/10.1186/s12986-020-00516-1
73. Sun J et al (2023) "The visceral adipose tissue bacterial microbiota provides a signature of obesity based on inferred metagenomic functions." Int J Obes 47(10):1008–1022. https://doi.org/10.1038/s41366-023-01341-1
74. Zhu X et al (2023) A high-fat diet increases the characteristics of gut microbial composition and the intestinal damage associated with non-alcoholic fatty liver disease. Int J Mol Sci 24(23):16733. https://doi.org/10.3390/ijms242316733
75. Dekker Nitert M, Mousa A, Barrett HL, Naderpoor N, de Courten B (2020) Altered gut microbiota composition is associated with back pain in overweight and obese individuals. Front Endocrinol 11:605. https://doi.org/10.3389/fendo.2020.00605
76. McLeod A et al (2023) Comparing the gut microbiome of obese, African American, older adults with and without mild cognitive impairment. PLoS ONE 18(2):e0280211. https://doi.org/10.1371/journal.pone.0280211
77. Lu M et al (2024) Causal relationship between gut microbiota and childhood obesity: a Mendelian randomization study and case–control study. Clin Nutr ESPEN 63:197–206. https://doi.org/10.1016/j.clnesp.2024.05.012
78. Murga-Garrido SM et al (2023) Virulence Factors of the Gut Microbiome Are Associated with BMI and Metabolic Blood Parameters in Children with Obesity. Microbiol Spectr 11(2):e0338222. https://doi.org/10.1128/spectrum.03382-22
79. Song X et al (2019) Inulin can alleviate metabolism disorders in ob/ob mice by partially restoring leptin-related pathways mediated by gut microbiota. Genomics Proteomics Bioinformatics 17(1):64–75. https://doi.org/10.1016/j.gpb.2019.03.001
80. Sen T et al (2017) Diet-driven microbiota dysbiosis is associated with vagal remodeling and obesity. Physiol Behav 173:305–317. https://doi.org/10.1016/j.physbeh.2017.02.027
81. Yun Y et al (2017) Comparative analysis of gut microbiota associated with body mass index in a large Korean cohort. BMC Microbiol 17(1):151. https://doi.org/10.1186/s12866-017-1052-0
82. Chakraborti CK (2015) New-found link between microbiota and obesity. World J Gastrointest Pathophysiol 6(4):110–119. https://doi.org/10.4291/wjgp.v6.i4.110
83. Chambers ES et al (2019) Dietary supplementation with inulin-propionate ester or inulin improves insulin sensitivity in adults with overweight and obesity with distinct effects on the gut microbiota, plasma metabolome and systemic inflammatory responses: a randomised cross-over trial. Gut 68(8):1430–1438. https://doi.org/10.1136/gutjnl-2019-318424
84. Lee H et al (2022) A novel bacterium, *Butyricimonas virosa*, preventing HFD-induced diabetes and metabolic disorders in mice via GLP-1 receptor. Front Microbiol 13:858192. https://doi.org/10.3389/fmicb.2022.858192
85. Hu J et al (2022) Gut Microbiota Signature of Obese Adults Across Different Classifications. Diabetes Metab Syndr Obes Targets Ther 15:3933–3947. https://doi.org/10.2147/DMSO.S387523
86. Sabater C et al (2025) Enzymatic synthesis of β-galactosylated xylitol derivatives modulates gut microbiota and improves obesity-related metabolic parameters in mice. Food Funct 16(13):5493–5510. https://doi.org/10.1039/D5FO00978B
87. Maioli TU et al (2021) Possible benefits of *Faecalibacterium prausnitzii* for obesity-associated gut disorders. Front Pharmacol. https://doi.org/10.3389/fphar.2021.740636
88. Wang K et al (2019) *Parabacteroides distasonis* alleviates obesity and metabolic dysfunctions via production of succinate and secondary bile acids. Cell Rep 26(1):222-235.e5. https://doi.org/10.1016/j.celrep.2018.12.028
89. Choi WJ et al (2020) "Lactobacillus plantarum LMT1-48 exerts anti-obesity effect in high-fat diet-induced obese mice by regulating expression of lipogenic genes." Sci Rep 10(1):869. https://doi.org/10.1038/s41598-020-57615-5