



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**Programa de Maestría y Doctorado en Ciencias Médicas,**  
**Odontológicas y de la Salud**

**Investigación Clínica Experimental en Salud**

**FACULTAD DE QUÍMICA**

**EL PAPEL DEL VIROMA EN EL MICROBIOMA  
INTESTINAL ASOCIADO A LA OBESIDAD CON  
COMPLICACIONES METABÓLICAS EN POBLACIÓN  
INFANTIL MEXICANA**

**TESIS**

**QUE PARA OPTAR POR EL GRADO DE:**

**DOCTORA EN CIENCIAS**

**PRESENTA**

**M.C. SHIRLEY BIKEL**

**Director de tesis:**

**Dr. Adrián Ochoa Leyva**

**Instituto de Biotecnología (IBt), UNAM**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

# ÍNDICE

---

<b>1. INDICE DE FIGURAS, TABLAS Y ANEXOS</b>	<b>3</b>
<b>2. ABSTRACT</b>	<b>5</b>
<b>3. RESUMEN</b>	<b>6</b>
<b>4. HOJA DE ABREVIATURAS</b>	<b>7</b>
<b>5. INTRODUCCIÓN</b>	<b>8</b>
<b>6. ANTECEDENTES</b>	<b>14</b>
<b>7. PLANTEAMIENTO DEL PROBLEMA</b>	<b>16</b>
<b>8. JUSTIFICACIÓN</b>	<b>16</b>
<b>9. HIPÓTESIS</b>	<b>17</b>
<b>10. OBJETIVO GENERAL</b>	<b>17</b>
<b>11. OBJETIVOS ESPECÍFICOS</b>	<b>17</b>
<b>12. MATERIALES Y MÉTODOS</b>	<b>18</b>
12.1. Población de estudio	18
12.2. Aislado de partículas virales (VLPs)	19
12.3. Visualización microscópica y conteo de VLPs	19
12.4. Extracción y secuenciación masiva del ADN viral	20
12.5. Limpieza y agrupación de las lecturas secuenciadas	20
12.6. Análisis de la riqueza de lecturas virales	21
12.7. Perfiles funcionales y análisis de pVOGs de las lecturas derivadas de VLPs	21
12.8. Clasificación de las lecturas virales	21
12.9. Ensamble De Novo	22
12.10. Clasificación taxonómica de los contigs del ensamble de novo	22
12.11. Análisis de las abundancias diferenciales de los “contigs de fagos”	23
12.12. Riqueza y diversidad de los “contigs de fagos”	23
12.13. Prevalencia de los “contigs de fagos” altamente abundantes en NW	23
12.14. Correlaciones de los fagos con bacterias y parámetros bioquímicos	23
12.15. Comparación de las distancias Euclidianas	24
12.16. Disponibilidad de datos y software	24
<b>13. RESULTADOS</b>	<b>25</b>
13.1. Caracterización de la población de estudio	25
13.2. Selección de la población de estudio	26
13.3. El número de partículas virales (VLPs) es similar entre NW, O y OMS	29
13.4. Perfiles funcionales de las lecturas sugieren presencia viral en las muestras de VLPs	30
13.5. El número de lecturas únicas, derivadas de los VLPs, muestran una riqueza incrementada en la enfermedad	32
13.6. Los bacteriófagos dominaron las lecturas virales del viroma intestinal	33
13.7. El ensamble viral confirma el dominio de los bacteriófagos	35
13.8. El fagoma intestinal de niños estuvo compuesto principalmente de Caudovirales	39
13.9. El aumento de la riqueza, diversidad, y dominancia de los contigs de fagos está vinculado al cambio de peso normal saludable a obesidad	40
13.10. Varios contigs de fagos estuvieron significativamente sobreabundantes en la obesidad y en el síndrome metabólico	43
13.11. El fagoma fue principalmente específico para cada individuo	45
13.12. La enfermedad alteró la prevalencia de los contigs de fagos altamente abundantes en NW	46

13.13. Las abundancias de los “fagos core” correlacionaron con taxones bacterianos asociados a obesidad y síndrome metabólico	47
13.14. Las abundancias de los “contigs de fagos core” correlacionaron con los parámetros bioquímicos y antropométricos alterados en la obesidad y síndrome metabólico	50
<b>14. DISCUSIÓN</b>	<b>51</b>
<b>15. CONCLUSIONES</b>	<b>57</b>
<b>16. PERSPECTIVAS</b>	<b>58</b>
<b>17. BIBLIOGRAFÍA</b>	<b>59</b>
<b>18. ANEXO</b>	<b>67</b>
<b>19. ARTÍCULOS DERIVADOS DE LA PRESENTE TESIS</b>	<b>95</b>

# 1. ÍNDICE DE FIGURAS, TABLAS y ANEXOS

---

## FIGURAS

<b>Figura 1.</b> Tasas de obesidad en niños en edad escolar a nivel mundial	<b>8</b>
<b>Figura 2.</b> Anomalías metabólicas que integran el diagnóstico de síndrome metabólico	<b>9</b>
<b>Figura 3.</b> Grupos taxonómicos principales del microbioma intestinal humano y de los niveles de Dominio/Reino	<b>10</b>
<b>Figura 4.</b> Dinámica fago-bacteria en la microbiota intestinal	<b>12</b>
<b>Figura 5.</b> Componentes del síndrome metabólico en la población de niños con peso normal saludable (NW), obesos (O) y obesos con síndrome metabólico (OMS).	<b>26</b>
<b>Figura 6.</b> Visualización de la cantidad y tipo de componentes del síndrome metabólico para cada una de las muestras seleccionadas para el estudio	<b>28</b>
<b>Figura 7.</b> Visualización microscópica y conteo de VLPs	<b>29</b>
<b>Figura 8.</b> Secuencias de VLPs detectadas en las muestras NW, O, y OMS	<b>31</b>
<b>Figura 9.</b> Asignación funcional de VLPs	<b>32</b>
<b>Figura 10.</b> Agrupaciones de secuencias únicas (eje y) por grupo	<b>33</b>
<b>Figura 11.</b> Asignación taxonómica viral de las lecturas derivadas de los VLPs	<b>34</b>
<b>Figura 12.</b> Clasificación taxonómica de las lecturas y de los contigs virales	<b>35</b>
<b>Figura 13.</b> Porcentaje de las lecturas de calidad que mapearon contra los diferentes ensambles de contigs	<b>36</b>
<b>Figura 14.</b> Diagrama de Venn mostrando la superposición entre las tres clasificaciones obtenidas de la NR y NT y de los contigs clasificados usando VirSorter.	<b>38</b>
<b>Figura 15.</b> Correlación entre las lecturas de calidad y el número de contigs por muestra	<b>39</b>
<b>Figura 16:</b> Abundancia relativa de los contigs de fagos clasificados taxonómicamente por muestra y por grupo	<b>40</b>
<b>Figura 17.</b> Abundancia relativa del RPKM normalizado para los 4,611 contigs de fagos	<b>41</b>
<b>Figura 18.</b> Diversidad Alpha y Beta del fagoma	<b>42</b>
<b>Figure 19.</b> Análisis de los contigs de fagos sobreabundantes en la obesidad (O) y en la obesidad con síndrome metabólico (OMS).	<b>44</b>
<b>Figura 20.</b> “Heat-maps” de presencia-ausencia de la distribución de los contigs de fagos entre todas las muestras	<b>46</b>
<b>Figura 21.</b> Prevalencia del Fagoma core en NW.	<b>47</b>
<b>Figura 22.</b> Patrones de interacción del Fagoma específicos de la enfermedad en la obesidad y el síndrome metabólico	<b>48</b>

<b>Figura 23.</b> Gráficas de cajas de las abundancias de los contigs de fagos en NW, O y OMS que significativamente correlacionaron	<b>49</b>
--	-----------

---

## **TABLAS**

<b>Tabla 1.</b> Características de las muestras seleccionadas.	<b>27</b>
<b>Tabla 2.</b> Muestras pareadas por edad y género.	<b>28</b>
<b>Tabla 3.</b> Resultado de todos los análisis “Quast” de los ensamblajes virales.	<b>36</b>
<b>Tabla 4.</b> Número de lecturas mapeadas a los ensamblajes de contigs.	<b>37</b>

---

## **ANEXO**

<b>Anexo 1.</b> Tabla con la lista de los estudios del viroma humano publicados hasta la actualidad	<b>68</b>
<b>Anexo 2.</b> Tabla con la lista de los estudios de trasplantes de fagos realizados hasta la actualidad	<b>81</b>
<b>Anexo 3.</b> Conteo de VLPs por muestra	<b>64</b>
<b>Anexo 4.</b> Resultados de la secuenciación de VLPs	<b>65</b>
<b>Anexo 5.</b> Abundancia relativa de la clasificación taxonómica de los 4,611 contigs de fagos por muestra	<b>68</b>
<b>Anexo 6.</b> Análisis de Coordinadas Principales (PCoA) basado en la disimilitudde Bray-Curtis.	<b>69</b>
<b>Anexo 7.</b> PCoA basado en coordenadas CLR	<b>70</b>
<b>Anexo 8.</b> Gráficas de correlación de Spearman	<b>71</b>
<b>Anexo 9.</b> Regresión lineal y correlación de Spearman de los contigs de fagos que correlacionaron significativamente con taxones bacterianos específicos de la enfermedad	<b>73</b>
<b>Anexo 10.</b> Gráfica de correlación “contigs de fagos core”-bacteria	<b>74</b>
<b>Anexo 11.</b> Regresión lineal y correlación de Spearman de los contigs de fagos que significativamente correlacionaron con parámetros bioquímicos y antropométricos específicos	<b>75</b>

## 2. ABSTRACT

*Changes in the human gut microbiome are associated with obesity and metabolic syndrome. However, the role of gut virome in both diseases remains largely unknown. Therefore, we characterized the gut dsDNA virome of 28 school-age children with healthy normal-weight (NW, n=10), obese (O, n=10), and obese with metabolic syndrome (OMS, n= 8), using metagenomic sequencing of virus-like particles (VLPs) isolated from fecal samples. The classification of all virome assembly confirmed the bacteriophages dominance, mainly composed of Caudovirales. Notably, we observed that phage richness and diversity in O and OMS individuals tended to increase, while the VLP abundance remained the same among the three groups. In addition, the prevalence of the highly abundant NW phage contigs was altered in O and OMS individuals. Of the 4,611 phage contigs composing the Phageome, only 48 were highly prevalent in  $\geq 80\%$  of all individuals, suggesting high inter-individual phage diversity. Several of those contigs correlated with gut bacterial taxa, anthropometric and biochemical parameters typically altered by obesity, suggesting an association between specific phages and these variables. We also found phage contigs that were over-abundant in the diseases. To our knowledge, this human gut Phageome represents one of the largest-sequencing datasets to analyze disease-specific alterations in childhood obesity and obesity with metabolic syndrome.*

### 3. RESUMEN

*Cambios en el microbioma intestinal humano han sido asociados con obesidad y síndrome metabólico. Sin embargo, el papel del viroma intestinal en ambas enfermedades sigue siendo en gran parte desconocido. Por lo tanto, caracterizamos por primera vez el viroma dsDNA intestinal de 28 niños en edad escolar con peso normal (NW, n=10), con obesidad sin síndrome metabólico (O, n=10), y con obesidad con síndrome metabólico (OMS, n=8), mediante secuenciación metagenómica de partículas tipo virales (VLPs\_ por sus siglas en inglés “viral like particles”) aisladas de muestras fecales. La clasificación de todo el ensamble del viroma confirmó el dominio de bacteriófagos (o fagos), principalmente compuestos de Caudovirales. Notablemente, observamos que la riqueza y diversidad de fagos en individuos O y OMS tendieron a incrementar con respecto a los individuos NW, mientras que la abundancia de VLPs se mantuvo igual entre los tres grupos. Además, la prevalencia de los contigs de fagos altamente abundantes en el grupo NW, estuvo alterada en los individuos O y OMS. De los 4611 contigs de fagos que componen el Fagoma, solo 48 estuvieron altamente prevalentes en  $\geq 80\%$  de todos los individuos, sugiriendo que la diversidad de fagos es alta entre los individuos. Varios contigs de fagos correlacionaron con taxones bacterianos intestinales, con parámetros bioquímicos y antropométricos típicamente alterados en la obesidad, sugiriendo una asociación entre fagos específicos y estas variables. También, encontramos un conjunto de contigs de fagos que estuvo sobreabundante en la enfermedad. A nuestro conocimiento, este Fagoma intestinal humano representa a uno de los conjuntos de datos de secuenciación más grandes para analizar alteraciones específicas de la enfermedad en la obesidad infantil y en la obesidad con síndrome metabólico.*



#### 4. HOJA DE ABREVIATURAS

**NW:** Peso normal saludable (Normal Weight)

**O:** Obesidad

**OMS:** Obesidad con Síndrome Metabólico

**VLPs:** Partículas tipo virales (virus-like particles)

**TA:** Tensión arterial

**TG:** Triglicéridos

**CC:** Circunferencia de cintura

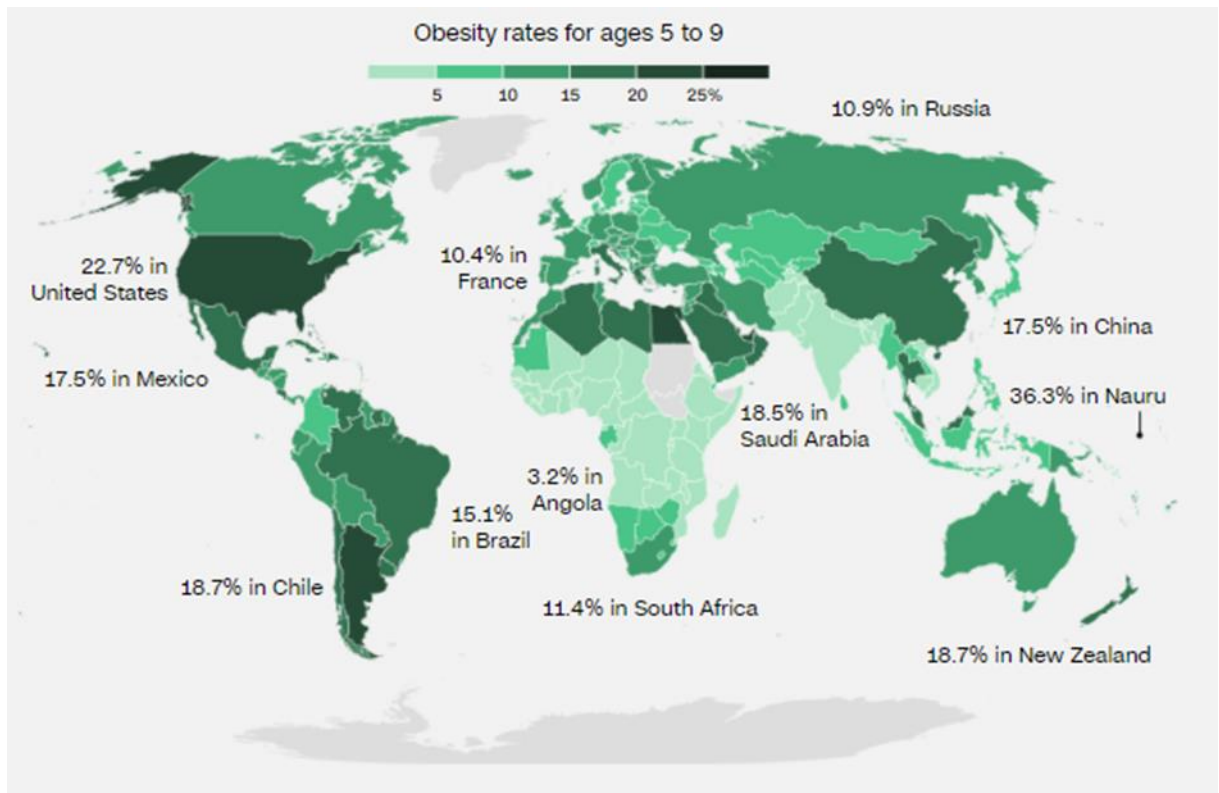
**IMC:** Índice de masa corporal

**HDL:** colesterol de lipoproteínas de alta densidad

**LDL:** Colesterol de lipoproteínas de baja densidad

## 5. INTRODUCCIÓN

La obesidad infantil es uno de los problemas de salud más relevantes y severos alrededor del mundo. La obesidad es una enfermedad multifactorial definida como la acumulación excesiva de grasa corporal, en donde existe un desequilibrio entre el consumo y gasto de energía, y que involucra factores genéticos, ambientales y de estilo de vida. Ante el incremento de la obesidad en los países desarrollados y en vías de desarrollo, la Organización Mundial de la Salud (OMS) ha declarado a esta enfermedad como una epidemia mundial, la cual contribuye a incrementar la morbilidad y mortalidad de diversas enfermedades. México, ocupa el segundo lugar en el mundo con una de las tasas de obesidad infantil más altas<sup>1</sup>, en donde el 17.5% de los niños en edad escolar sufren obesidad<sup>2</sup> (Figura 1).

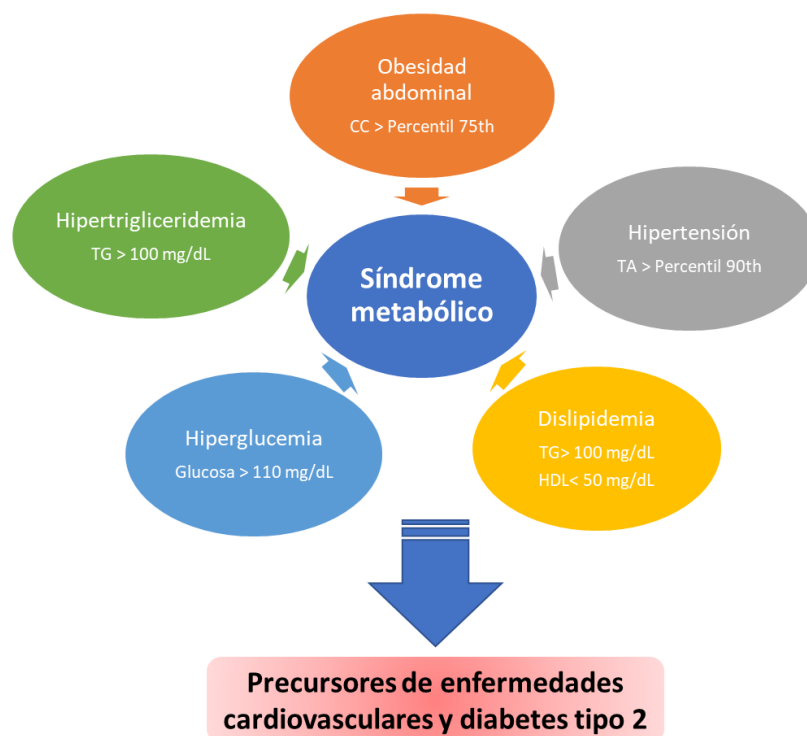


**Figura 1. Tasas de obesidad en niños en edad escolar a nivel mundial.** Prevalencia de la obesidad en niños de 5-9 años de edad en el año 2016. Fuente: Organización Mundial de la Salud (OMS). <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>.

La obesidad, considerada también como un estado de inflamación crónica de bajo grado<sup>3</sup>, se asocia con el desarrollo de distintas comorbilidades que incluyen altos niveles de azúcar en

sangre durante el ayuno (hiperglicemia), altos niveles de triglicéridos (hipertrigliceridemia), bajos niveles de lipoproteínas de alta densidad-HDL (dislipidemia) y presión sanguínea elevada (hipertensión) <sup>4</sup> (Figura 2). Al menos tres de estos criterios constituyen el diagnóstico de síndrome metabólico <sup>5</sup>.

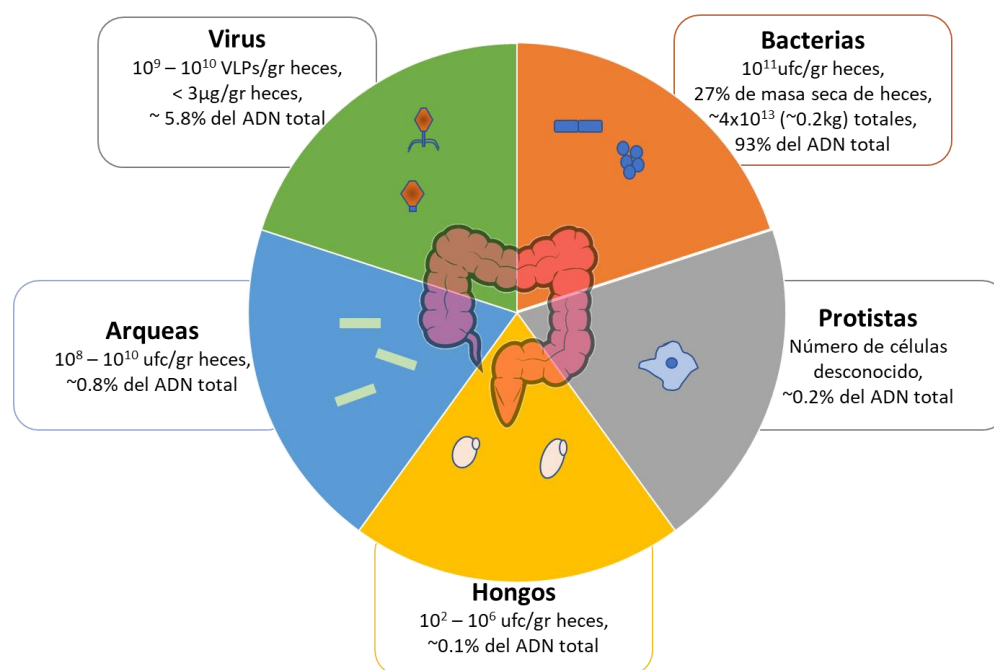
Notablemente, los niños mexicanos son considerados un grupo de alto riesgo para desarrollar síndrome metabólico <sup>2</sup>, en quienes se ha reportado una prevalencia del 53% al 62% entre población obesa <sup>4,6</sup>, constituyendo un factor de riesgo incrementado para el desarrollo de Diabetes Mellitus tipo 2 y de enfermedades cardiovasculares en la edad adulta <sup>7,8</sup> (Figura 2). Además, la obesidad en edades tempranas es un factor importante para el pronóstico tanto de obesidad como de sus complicaciones en el adulto, y representa un factor de riesgo sustancial de muerte prematura <sup>9</sup>.



**Figura 2. Anomalías metabólicas que integran el diagnóstico de síndrome metabólico y que en conjunto constituyen un factor de riesgo para el desarrollo de enfermedades cardiovasculares y de diabetes tipo 2.** CC: Circunferencia de cintura, TG: Triglicéridos, TA: tensión arterial.

En los últimos años, el estudio del microbioma intestinal ha cobrado cada vez más importancia debido a que se ha encontrado altamente relacionado con la salud humana y con diversas enfermedades. El microbioma intestinal humano está compuesto por una gran diversidad

de bacterias, arqueas, y células eucariotas, que junto con los virus (principalmente bacteriófagos) componen un ecosistema diverso y complejo <sup>10</sup> (Figura 3).

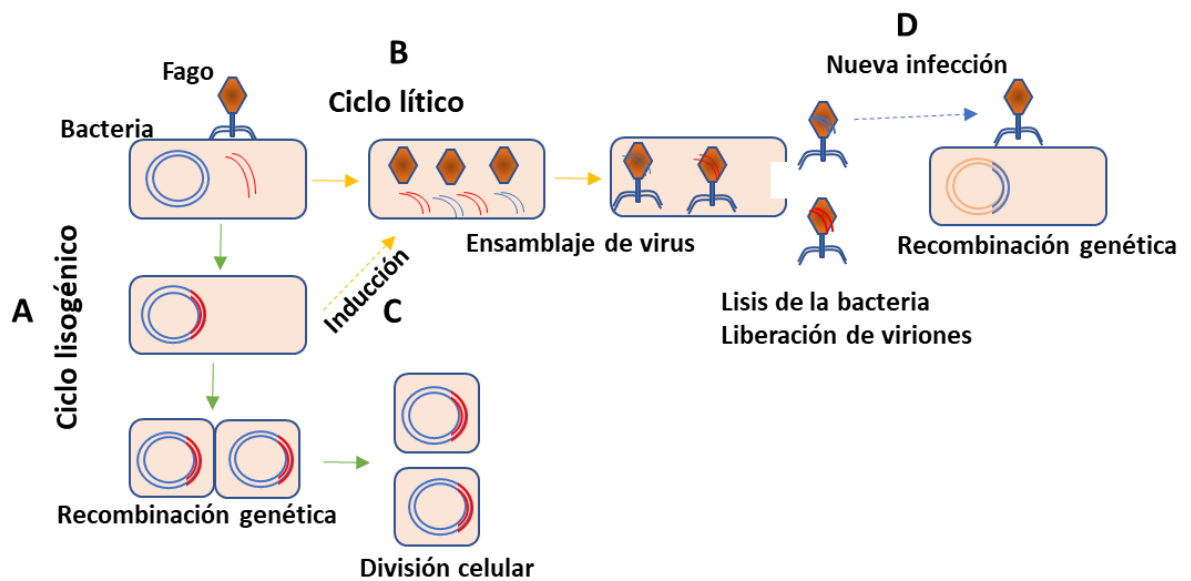


**Figura 3. Grupos taxonómicos principales del microbioma intestinal humano y de los niveles de Dominio/Reino.** <sup>11-17</sup>

Diferentes estudios han demostrado que existen alteraciones de la microbiota intestinal en la obesidad y en la obesidad con síndrome metabólico usando perfiles del 16S rRNA <sup>6,18-20</sup> y aproximaciones de metatranscriptómica <sup>6</sup>. Interesantemente, se ha demostrado que cambios en la composición de la microbiota intestinal están potencialmente asociados a un aumento en la incorporación de energía de la dieta, a un grado bajo de inflamación crónica, y a una composición alterada del tejido adiposo <sup>21</sup>. Estos procesos han sido considerados el vínculo entre el microbioma intestinal, la obesidad y el síndrome metabólico <sup>22</sup>. Al igual que en el bacterioma intestinal, ha sido observado que la intervención dietética también se asocia a cambios en la comunidad del viroma intestinal, en donde los individuos con la misma dieta convergieron <sup>23</sup>. Incluso, cambios en la estructura del viroma debido a dietas altas en grasa también han sido observados en ratones, sugiriendo conexiones entre fagos y sus huéspedes bacterianos a partir de los cambios microbianos inducidos por la dieta <sup>24</sup>.

La metagenómica viral es un campo relativamente nuevo y en crecimiento que estudia la colección completa de virus que forman parte de la microbiota en un nicho dado <sup>25</sup>. El viroma

intestinal está dominado, principalmente, por bacteriófagos (Fagoma) <sup>26</sup>, que tienen la capacidad de regular el ecosistema microbiano y la fisiología del hospedero <sup>27</sup> a través de múltiples interacciones con el sistema inmune <sup>28</sup> y con el bacterioma, incluyendo eventos de transferencia génica horizontal <sup>29</sup> (Figura 4). Como parásitos obligados de bacteria, la persistencia de fagos en un ecosistema microbiano es dependiente de la presencia de sus huéspedes bacterianos. A su vez, el Fagoma intestinal desempeña un papel clave en la conformación de la composición y función de la microbiota intestinal humana <sup>30-32</sup>, mediante uno de sus dos principales ciclos de replicación, lítico o lisogénico <sup>33</sup> (Figura 4). En ambos casos, un virión de fago se une a la superficie de la célula huésped utilizando una proteína de unión a receptores que desencadena la inserción de su genoma en la célula bacteriana. Para el caso de los fagos líticos, la traducción del material genético del fago por la célula huésped resulta en la replicación del genoma del fago, el montaje de partículas de fago y la lisis del huésped. Esto resulta en la liberación de nuevos viriones de fago hacia el entorno, en donde pueden infectar a los huéspedes cercanos <sup>34</sup>. De esta manera, la cantidad y tipo de bacterias presentes en la microbiota puede modificarse <sup>34</sup>. Alternativamente, la infección lisogénica resulta en la replicación del genoma de fago dentro de la célula huésped sin la síntesis inmediata de viriones de fago. El genoma del fago se integra en el genoma del huésped en donde existe como profago, y se replica junto con el genoma bacteriano, persistiendo así en su descendencia bacteriana <sup>34</sup>. De esta manera, nuevas funciones pueden ser adoptadas por las células bacterianas, las cuales se transmitirán a toda su progenie. Después de un evento de inducción, el fago lisogénico iniciará la traducción de su genoma y la producción subsiguiente de viriones del fago que conducen a la lisis del hospedador <sup>34</sup>. En el caso de la pseudolisogenia, el genoma del fago persiste como un episoma dentro de la célula huésped, separado del genoma del huésped. Para garantizar que las células hijas posteriores contengan genomas de fagos, los fagos pseudolisogénicos pueden usar sistemas de mantenimiento como la toxina-antitoxina <sup>35,36</sup>. Sin embargo, también se han informado casos de células hijas que carecen de fagos pseudolisogénicos <sup>36</sup>. Es así como los fagos han sido íntimamente asociados con la virulencia y evolución de bacterias patógenas, así como con la resistencia a antibióticos y nuevas capacidades metabólicas <sup>34</sup>.



**Figura 4. Dinámica fago-bacteria en la microbiota intestinal.** Esquema de los ciclos (A) lisogénico y (B) lítico. (C) Inducción del ciclo lítico por separación del fago del cromosoma bacteriano. (D) Transferencia génica horizontal mediada por bacteriófagos.

Se ha demostrado la transmisión de genes resistentes a antibióticos (ARG) por fagos en un gran número de bacterias, incluidas *Salmonella*, *Clostridium*, *Streptococcus*, *Staphylococcus* y *Bacillus*<sup>37-39</sup>. Un estudio reciente describió la mayor incidencia de ARGs en secuencias de viromas de pacientes con fibrosis quística en comparación a los viromas de individuos sin la enfermedad<sup>40</sup>. Este estudio destacó el papel potencial de los fagos en la evolución de la resistencia a los antibióticos, describiendo la presencia de genes relacionados con bombas de eflujo, genes de resistencia a fluoroquinolonas y genes de  $\beta$ -lactamasa en la fracción viral<sup>40</sup>. En un segundo estudio, los autores trataron ratones con ampicilina o ciprofloxacino y secuenciaron los viromas de las muestras fecales colectadas<sup>41</sup>. Con respecto a las muestras control de los ratones no tratados, la exposición a antibióticos causó incrementos de 2-3 razones de cambio en las lecturas virales anotadas como ARGs. Además, las transducciones con mezclas de fagos en los ratones tratados con antibióticos resultaron en incrementos de 2 razones de cambio en ARGs en comparación a una microbiota sin tratamiento previo<sup>41</sup>. Un tercer estudio comparó los genomas fecales y orales de comunidades virales y bacterianas provenientes de humanos sometidos a terapia con antibióticos<sup>42</sup>. Mientras este estudio reportó un cambio significativo en las comunidades bacterianas después del tratamiento con antibióticos, no se observó ningún aumento significativo en las lecturas de ARGs entre las muestras virales de los individuos tratados con antibióticos. Sin embargo, un nuevo análisis de estos datos<sup>43</sup> sugiere que los estudios mencionados sufrieron de falsos positivos inflados,

debido a un contenido excesivo de ADN bacteriano y a la utilización de umbrales relajados para la detección in silico de ARG. Por lo tanto, los autores propusieron un algoritmo automatizado específico para detectar o descartar contaminación bacteriana (VirSorter)<sup>44</sup>, el cual utiliza un umbral conservador para cuantificar ARG con mejor precisión y eliminar el tipo de falsos positivos identificados en estos estudios. Sugiriendo, además, que el ensamblaje de contigs debe usarse siempre que sea posible para confirmar que los nuevos ARG están realmente presentes en los contigs virales.

A pesar de la gran cantidad de fagos que hay en el intestino humano ( $>10^{10}$  g<sup>-1</sup>)<sup>11,12</sup> y del impacto considerable sobre los ecosistemas microbianos, siguen siendo uno de los miembros menos comprendidos del microbioma intestinal, debido a que la mayor proporción del viroma es desconocida, siendo la fracción identificable de tan solo 1-14%<sup>45</sup>. Hasta ahora, cambios específicos del viroma y del Fagoma intestinal humanos han sido reportados, principalmente, en la enfermedad inflamatoria intestinal<sup>30,32</sup>, SIDA<sup>46</sup>, diabetes<sup>47</sup>, y malnutrición<sup>48</sup> (ver Anexo 1 para más ejemplos). Sin embargo, los estudios que han abordado el papel del viroma en la obesidad y en el síndrome metabólico han estado limitados a modelos animales<sup>24,49,50</sup> y a humanos adultos<sup>23,31</sup>, pasando por alto las cohortes de niños.

Para conocer si el Fagoma intestinal juega un papel esencial en la obesidad y en el síndrome metabólico, caracterizamos el Fagoma intestinal en niños con peso normal saludable (NW), con obesidad (O), y con obesidad con síndrome metabólico (OMS), mediante la secuenciación metagenómica de partículas virales (VLPs) provenientes de muestras fecales.

## 6. ANTECEDENTES

Desde que los fagos fueron identificados por primera vez por Frederick Twort en 1915 <sup>51</sup>, se han utilizado métodos basados en cultivos, como los ensayos de placa, para seleccionar y cuantificar los títulos de fagos de muchos entornos. Hoy en día, estos métodos todavía juegan un papel central en la identificación de fagos que se dirigen a bacterias específicas y han contribuido a la comprensión de la mecánica de las interacciones entre los fagos y los ciclos de replicación. Sin embargo, como se desconoce la gran mayoría de los fagos y de sus hospedadores que participan en la interacción fago-huésped en el intestino, estos métodos no son adecuados para la caracterización a gran escala de un ecosistema complejo como el intestino humano. Hasta la fecha, la mayoría de la investigación del metagenoma viral (viroma) se ha centrado en comunidades ambientales como las del océano <sup>52,53</sup>. Actualmente, los estudios de viromas se apoyan en gran medida en enfoques metagenómicos basados en la secuenciación para investigar las comunidades de fagos intestinales y para tratar de comprender su papel en la configuración del microbioma intestinal <sup>45</sup>. Esto implica secuenciar el ADN y ARN virales totales de una comunidad después de la separación física del componente bacteriano, utilizando software de ensamblaje para recrear los genomas virales dentro de esa comunidad y caracterizar la abundancia y función de esos genomas. El problema es que con este tipo de secuenciación no se puede confirmar que realmente sean fagos activos como partículas virales.

Dada la amplia evidencia de que los fagos pueden dar forma a la composición y función de las comunidades bacterianas, el viroma del intestino humano se ha estudiado en varias enfermedades, como en la enfermedad inflamatoria intestinal <sup>30,32</sup>, SIDA <sup>46</sup>, diabetes <sup>47</sup>, y malnutrición <sup>48</sup> (Anexo 1). Sin embargo, los hallazgos han sido algo contradictorios y cualquier papel potencial del viroma en la configuración del microbioma en la enfermedad sigue siendo difícil de alcanzar. Estudios recientes han informado que las poblaciones de fagos intestinales no se alteraron significativamente en enfermedades como el cáncer colorrectal y el SIDA asociado al VIH <sup>46,54</sup>, a pesar de las asociaciones establecidas entre el microbioma intestinal y estas enfermedades <sup>13,55</sup>. Sin embargo, muchos estudios de viromas basados en secuencias excluyen la materia oscura viral de los análisis, y trabajan en gran medida con una pequeña fracción de secuencias de fagos conocidas (por lo general, del 1 al 14% del conjunto de datos) <sup>45</sup>. Esto puede tener profundas implicaciones para las conclusiones extraídas de estos estudios, ya que los cambios en la fracción conocida pueden no reflejar cambios en el viroma en su conjunto. Como resultado, se utilizan cada vez más métodos de análisis independientes de la base de datos que incluyen fracciones conocidas y desconocidas del viroma <sup>11</sup>. Sin embargo, los altos niveles de inter-individualidad hacen que las señales biológicas en los estudios de viromas sean difíciles de detectar <sup>56-</sup>



<sup>58</sup>. Además, los estudios de viomas son particularmente susceptibles a sesgos metodológicos debido a las dificultades en la evaluación comparativa de las herramientas bioinformáticas *de novo* y al predominio de secuencias desconocidas en conjuntos de datos de viomas <sup>59,60</sup>. Estudios recientes han propuesto la utilización de ciertos parámetros bioinformáticos astringentes para los análisis de secuencias con el objetivo de lograr obtener estimaciones más precisas de la composición de las comunidades virales <sup>48</sup>, sin embargo, todavía se corre el riesgo de que los ensamblajes virales estén compuestos, en parte, por fragmentos de genomas virales parciales, requiriendo en la actualidad de estudios más extensos como los que se han sugerido previamente <sup>61</sup>.

A pesar de las dificultades para los análisis de los viomas intestinales, la primera evidencia de que el vioma tiene potencial como herramienta para dar forma al microbioma y puede desempeñar un papel en la eficacia del tratamiento con trasplante de microbiota fecal (FMT) fue informada por Ott et al. <sup>62</sup>. En este estudio, los pacientes con enfermedad de Crohn recibieron filtrados fecales de donantes sanos que resultaron en la eliminación de los síntomas hasta por 6 meses. Por otro lado, se están realizando ensayos de terapia con fagos existentes <sup>63-66</sup> con resultados prometedores, siendo en su mayoría estudios de intervención con fagos basados en pares únicos fago-hospedador (o cocteles que contienen un número limitado de fagos) *in vitro* (para mayores detalles ver Anexo 2). Por lo tanto, superar los desafíos que implica el estudio del vioma es de vital importancia.

## **7. PLANTEAMIENTO DEL PROBLEMA**

La población infantil mexicana presenta una de las prevalencias más altas de obesidad y comorbilidades asociadas siendo a su vez un factor importante para el pronóstico de obesidad y sus complicaciones en el adulto. Sin embargo, la dieta y la actividad física, ampliamente reconocidas como los factores de riesgo modificables más importantes para el desarrollo de la obesidad y sus complicaciones metabólicas, no explican en su totalidad la alta prevalencia de dicha enfermedad, y su manejo no ha sido llevado con éxito, hasta hoy en día. Esto demanda la atención del problema por parte de la comunidad científica para identificar aquellos factores que pueden predisponer, retardar o proteger a los niños obesos de desarrollar alteraciones metabólicas a corto o largo plazo. Recientemente, cambios en la estructura de la microbiota intestinal han sido propuestos como uno de los principales factores que se asocian al desarrollo de la obesidad<sup>6,18-22</sup>, sin embargo, se ha observado que alteraciones en la composición y función de la microbiota intestinal podrían estar asociadas al tipo de Fagoma con el que interactúan<sup>23,57,67</sup>. Por lo tanto, proponemos analizar el Fagoma intestinal en niños obesos, y su posible asociación con el bacterioma intestinal y el desarrollo de obesidad y síndrome metabólico.

## **8. JUSTIFICACIÓN**

Uno de los grandes retos que enfrentan los estudios de la microbiota humana es poder relacionar el conocimiento actual de la ecología microbiana con el de las actividades moleculares desarrolladas por cada uno de los microorganismos presentes en su conjunto y su asociación a enfermedades tales como la obesidad y el síndrome metabólico asociado a la obesidad. Sin embargo, los factores que dan forma a la composición de estas comunidades bacterianas siguen siendo poco conocidos y esto ha ralentizado el desarrollo de terapias y biomarcadores basados en el estudio del microbioma. Debido a que los fagos son los componentes que mayor proporción ocupan dentro de la microbiota intestinal y cuentan con la capacidad de regular la composición y función bacterianas, es necesario generar estrategias que nos permitan comprender a mayor profundidad y desde una perspectiva integradora, la dinámica de la microbiota intestinal humana, no solo a través del estudio del bacterioma sino también a través del Fagoma con el que interactúa. A la fecha, no existen reportes de estudios de metagenómica de partículas virales del intestino en la obesidad y el síndrome metabólico.

## **9. HIPÓTESIS**

Cambios de la diversidad y riqueza del Fagoma de la microbiota intestinal se asocia con la obesidad y la obesidad con síndrome metabólico en la población infantil mexicana.

## **10. OBJETIVO GENERAL**

Determinar la diversidad y riqueza del Fagoma intestinal y su asociación con la obesidad y obesidad con síndrome metabólico en población infantil mexicana mediante un análisis integrador de metagenómica de partículas virales y secuenciación del gen ribosomal 16S.

## **11. OBJETIVOS ESPECÍFICOS**

- Determinar la composición del Fagoma intestinal en niños con peso normal saludable, obeso y obeso con síndrome metabólico.
- Determinar la asociación de la diversidad y riqueza del Fagoma con la obesidad y con la obesidad con síndrome metabólico.
- Analizar la relación entre el bacterioma y el Fagoma con la obesidad y con la obesidad con síndrome metabólico.
- Evaluar la relación del Fagoma con los parámetros clínicos típicamente alterados en la obesidad y en el síndrome metabólico.
- Evaluar si existen cambios en la prevalencia del fagoma asociados a la enfermedad, además de si existe un nicho de fagos “core” que sean compartidos entre la mayoría de las muestras.

## 12. MATERIALES Y MÉTODOS

### 12.1. Población de estudio

Analizamos las muestras fecales de 10 niños con peso normal saludable (NW), 10 niños obesos (O), y 8 niños obesos con síndrome metabólico (OMS), con edades entre 7 a 10 años, que asisten a las Convivencias de Verano de la Secretaría de Salud. Dichas muestras son de una cohorte colectada y estudiada por el Dr. Samuel Canizales del Instituto Nacional de Medicina Genómica. Todos los niños provienen de hogares de clase económica media y pertenecen a un estatus socio-cultural similar. Todos ellos vivían en la Ciudad de México en el momento de la recolección de datos y no practicaban ningún deporte con regularidad. Los grupos de estudio fueron pareados por edad y género (Tablas 1 y 2). Las muestras fecales fueron recolectadas y refrigeradas en sus casas a 4 °C y transportadas a las instalaciones de investigación dentro de las siguientes 12 hrs en una heladera portátil con paquetes de hielo para preservar la temperatura. Las muestras fueron recibidas en las instalaciones de investigación temprano por la mañana. Se realizaron alícuotas de 200 mg en contenedores plásticos esterilizados y con RNA later. Las muestras fueron almacenadas en el banco de muestras fecales del Dr. Samuel Canizales de la Unidad de Genómica de Poblaciones Aplicada a la Salud UNAM-INMEGEN en contenedores plásticos con RNA later a -70°C. A cada participante se le tomó una bioquímica sanguínea. Las muestras sanguíneas de 5 mL fueron extraídas después de 8-12 hrs de ayuno el mismo día de la recolección de heces. A cada uno de los donantes se les realizó una historia clínica, así como también se les midió la presión arterial, y se obtuvieron parámetros antropométricos, siguiendo procedimientos estandarizados, como ha sido descrito previamente.

La obesidad fue definida con un índice de masa corporal (IMC)  $\geq$  percentil 95th. Por el contrario, NW fue definido con un IMC entre los percentiles 15th y 75th, considerando edad y género; basado en el Centro para el Control y Prevención de Enfermedades (CDC – por sus siglas en inglés: “Centers for Disease Control and Prevention”). Los parámetros de síndrome metabólico en niños fueron determinados acorde a reportes previos <sup>5</sup>, y los individuos OMS fueron definidos por la presencia de una circunferencia de cintura  $>$  percentil 75th considerando edad y género, junto con al menos dos de los siguientes rasgos metabólicos: (1) triglicéridos  $>$  1.1 mmol/L (100 mg/dL); (2) colesterol HDL  $<$  1.3 mmol/L (50 mg/dL), (3) glucosa  $>$  6.1 mmol/L (110 mg/dL) y (4) presión sanguínea sistólica  $>$  percentil 90th considerando género, edad y altura. Los niños del grupo O fueron seleccionados con no más de uno de los rasgos para síndrome metabólico (Figura 2). Los criterios de exclusión para las muestras incluyeron pérdida de peso reciente  $>$  10%, ingesta de antibióticos en los últimos 3 meses previos a la toma de las muestras, y la aparición de diarrea o alguna enfermedad

gastrointestinal aguda durante el mismo período.

El Comité de Ética del Instituto Nacional de Medicina Genómica (INMEGEN) en la Ciudad de México, aprobó el estudio. Los padres o tutores legales de cada niño firmaron el formato de consentimiento informado para su participación, y todos los niños asintieron participar.

### **12.2. Aislado de partículas virales (VLPs)**

Las partículas virales (VLPs) fueron aisladas de ~250 mg de muestra fecal suspendida por vórtice en 1 mL de Buffer SM (pH 7.5) (Cat. N° 725-2520, Nalgene), para estabilizar las partículas virales durante la manipulación, el cual fue previamente esterilizado por temperatura y filtrado a través de membranas PES de 0.22  $\mu\text{m}$  (Cat. N° 725-2520, Nalgene). El homogeneizado fue centrifugado a 4,700 x g por 30 minutos a temperatura ambiente con la finalidad de pasar las VLPs a la capa del sobrenadante, el cual, posteriormente, fue filtrado a través de membranas PES de 0.45  $\mu\text{m}$  (Cat. N° 725-2545, Nalgene) y de 0.22  $\mu\text{m}$  (Cat. N° 725-2520, Nalgene), para eliminar restos celulares y partículas de tamaño bacteriano. El filtrado fue re-suspendido en 15 mL de Buffer SM y concentrado a 200  $\mu\text{l}$  a 4°C con unidades de filtración de Amicon Ultra 15, 100 Kda (Cat. No. UFC910024, Millipore) durante varias veces, los cuales ayudaron a remover proteínas y células residuales. El concentrado fue transferido a un tubo de microcentrífuga de 1.5 mL e incubado con 40  $\mu\text{l}$  de cloroformo por 10 minutos a temperatura ambiente para degradar cualquier resto celular humano y bacteriano. El ADN no protegido por virus fue eliminado con 2.5 unidades por mililitro de Dnasa I siguiendo los procedimientos de manufactura (Cat No. 18047-019, Invitrogen). Después de la incubación, la Dnasa fue inactivada a 65°C por 10 minutos. Finalmente, las muestras fueron almacenadas a -80°C hasta su posterior procesamiento.

### **12.3. Visualización microscópica y conteo de VLPs**

Usamos microscopía de epifluorescencia para cuantificar los VLPs aislados. Para ello, teñimos 10  $\mu\text{l}$  del concentrado de cada muestra de VLPs con un mix de 2  $\mu\text{l}$  de SYBR Green (Invitrogen) (a partir de un concentrado de 5  $\mu\text{l}$  de SYBR Green en 45  $\mu\text{l}$  de agua filtrada) y 10  $\mu\text{l}$  de paraformaldehído, previamente filtrado a través de membranas PES de 0.22  $\mu\text{m}$  (Millipore). Cinco campos por muestra fueron observados y cuantificados con el Microscopio Confocal de Multifotones Olympus FV1000 y cada campo fue cuantificado en triplicado utilizando el software de procesamiento de imágenes Fiji. El promedio del número de VLPs proveniente de los cinco campos por triplicado fue obtenido por muestra para calcular la cantidad de VLPs por gramo de heces fecales por muestra (Anexo

3). Las diferencias en el número de VLPs entre los tres grupos fueron evaluadas por pares con pruebas no paramétricas de Mann–Whitney–Wilcoxon en R. Adicionalmente, una alícuota de 8  $\mu$ l del concentrado de las muestras de VLPs fue observada por microscopía electrónica de transmisión para corroborar la naturaleza del fago (morfología y estructura) en los VLPs. Las imágenes fueron observadas con una resolución de entre 200 nm - 500 nm, y obtenidas a 10kX - 80kX.

#### **12.4. Extracción y secuenciación masiva del ADN viral**

El ADN de las VLPs fue extraído siguiendo el protocolo de manufactura para el kit QIAampMinElute Virus Spin (QIAGEN, Cat. 57704). El ADN resultante para cada muestra fue cuantificado con un fluorómetro de Qubit (Invitrogen, Cat. Q32851) y diluido en agua libre de ribonucleasas a una concentración de 0.3 ng/ $\mu$ L. A partir de este concentrado de ADN, preparamos librerías de secuenciación independientes siguiendo el protocolo de Illumina Nextera XT DNA Library Preparation (Illumina, Cat. FC-131-1024), que permite utilizar cantidades ultra-pequeñas de ADN, con códigos de barra únicos para el multiplexado de muestras. Para esto, mezclamos 5  $\mu$ l del concentrado de ADN con el mix de reactivos para la tagmentación. A continuación, agregamos los oligos indexados y amplificamos la librería por 12 ciclos. Cada librería fue purificada con 30  $\mu$ l de perlas AMPure XP para obtener fragmentos de ADN de ~600 pb. El tamaño y la calidad de cada librería fueron analizados con el DNA bioanalyzer 2100 (Agilent Technologies, Cat. 5067-4626). Todas las librerías con sus códigos de barra fueron agrupadas juntas y secuenciadas usando la plataforma de Illumina NextSeq500 en el modo “pair-end” 2x150 en las instalaciones de la Unidad de Secuenciación del Instituto Nacional de Medicina Genómica, México.

#### **12.5. Limpieza y agrupación de las lecturas secuenciadas**

Las lecturas totales fueron dereplicadas. Los adaptadores y las bases de baja calidad (PHRED Q30) fueron recortados usando Trim\_Galore (<https://github.com/FelixKrueger/TrimGalore>), y los primeros 20 nucleótidos fueron removidos con Fastq ([http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)). Las lecturas de humano y de bacteria fueron removidas por mapeo de lecturas usando BWA (Burrows-WheelerAligner) (contra el genoma de referencia del Homo Sapiens v38) y contra la base de datos NR de bacterias usando Kraken, con parámetros predeterminados. Todas las lecturas mapeadas a estos genomas fueron eliminadas, y las lecturas restantes fueron nombradas “lecturas de calidad filtradas”. Las “lecturas de calidad filtradas”

fueron agrupadas a un 95% de identidad usando CD-HIT (<http://weizhongli-lab.org/cd-hit/>) para remover redundancias y generar secuencias únicas.

### **12.6. Análisis de la riqueza de lecturas virales**

La riqueza viral entre grupos (NW, O, OMS) fue determinada mediante la generación al azar de 1000 sub-muestras de 149,000 “lecturas de calidad filtradas” “single-end”, de acuerdo a la muestra más pequeña (NW\_8: 149775) y, posteriormente, cada sub-muestra fue agrupada a un 95% de identidad usando CD-HIT para identificar grupos únicos de lecturas. Para cada muestra, la mediana de todas las observaciones únicas fue calculada. Los grupos resultantes fueron probados para la normalidad mediante la prueba de Shapiro-Wilk en R. Las diferencias fueron evaluadas por pares con la prueba no paramétrica de Mann–Whitney–Wilcoxon en R.

### **12.7. Perfiles funcionales y análisis de pVOGs de las lecturas derivadas de VLPs**

Las “lecturas de calidad filtradas” fueron mapeadas sobre las bases de datos NRRefSeq y POGs usando BLASTX con un e-value máximo de 0.001 y un máximo de 50 secuencias “target” reportadas. Después del mapeo, una matriz de abundancias fue generada usando un script “in-house”. La matriz fue entonces anotada acorde a las anotaciones de KEGG para cada proteína usando la base de datos online de UniProtKB y un script “in-house”. Fue generado un perfil funcional de KEGG usando la abundancia relativa de cada proteína y función para cada muestra. Las “lecturas de calidad filtradas” fueron mapeadas también contra la base de datos “Prokaryotic Virus Orthologous Groups” (pVOGs) usando BLASTX con un e-value máximo de 0.001 y un máximo de 50 secuencias “target” reportadas. Los resultados finales fueron filtrados con un script “in-house” para obtener la clasificación de pVOGs para cada muestra.

### **12.8. Clasificación de las lecturas virales**

Las secuencias únicas fueron clasificadas taxonómicamente en órdenes y familias, acorde al Comité Internacional de la Taxonomía de los Virus (ICTV, por sus siglas en inglés), usando BLASTX con un e-value de corte máximo de 0.001 contra la base de datos de la NR RefSeq viral y considerando el algoritmo basado en el ancestro común más bajo en MetaGenomeANalyzer (MEGAN6) usando los siguientes parámetros: Min Support: 1, Min Score: 40.0, Max Expected: 0.01, Top Percent: 10.0, Min-Complexity filter: 0.44. Los conteos de lecturas absolutas para los taxones virales seleccionados fueron normalizados usando todas las lecturas provenientes de cada muestra, obteniendo la abundancia relativa para cada muestra usando scripts en R.

## **12.9. Ensamble *De Novo***

Las “lecturas de calidad filtradas” provenientes de todas las muestras se unieron para construir un ensamble *de novo* usando el ensamblador IDBA-UD con una longitud de k-meros de 20-125 con rondas de andamio. Las lecturas de cada muestra fueron mapeadas en forma separada con Bowtie2 contra el ensamble viral usando el modo “end-to-end” con los parámetros preestablecidos. Los contigs virales que fueron cubiertos por las lecturas en  $\geq 80\%$  de su longitud en al menos una muestra, fueron usados como evidencia para descartar quimeras de contigs. Posteriormente, nos quedamos con los scaffolds  $\geq 4$  Kb para los siguientes análisis, con la finalidad de disminuir la probabilidad de seleccionar fragmentos de genomas virales. Para eliminar los contigs redundantes, usamos CD-HIT a un 95% de identidad. Para conocer el número de genes por longitud del contig, realizamos una predicción de los genes de cada contig usando FragGeneScan con los siguientes parámetros: -complete=0 -train=illumina\_5.

### **12.10. Clasificación taxonómica de los contigs del ensamble *de novo***

La clasificación taxonómica de cada contig fue obtenida usando dc-megablast contra la base de datos de genomas virales del NCBI NT con un e-value de 0.001 y un número máximo de secuencias “target” para reportar sets a 50 hits <sup>57</sup>. La taxonomía de cada contig fue asignada usando el algoritmo del ancestro común más bajo en MetaGenomeANalyzer (MEGAN6) usando los resultados de megablast discontinuos con los siguientes parámetros: Min Support: 1, Min Score: 40.0, Max Expected: 0.01, Top Percent: 10.0, Min-Complexity filter: 0.44. Después de eso, todos los contigs sin clasificación taxonómica con dc\_megablast fueron usados para buscar contra la base de datos de proteínas virales del NCBI NR usando BLASTX con el máximo e-value de 0.001 y un número máximo de secuencias “target” para reportar un set de 50. La taxonomía de cada contig fue asignada por el algoritmo del ancestro común más bajo en MetaGenomeANalyzer (MEGAN6), usando los resultados de BLASTX con los siguientes parámetros: Min Support: 1, Min Score: 40.0, Max Expected: 0.01, Top Percent: 10.0, Min-Complexity filter: 0.44. El ensamble de 12,287 contigs fue clasificado usando VirSorter2 (versión 2.2.1) con los parámetros preestablecidos siguiendo el tutorial proporcionado por los autores. Posterior a esto, contrastamos la clasificación de VirSorter contra la clasificación previa usando las bases de datos de los genomas virales de NT NCBI y de las proteínas virales NR NCBI <sup>68</sup>. El ensamble de 12,287 contigs fue mapeado contra la base de datos “Prokaryotic Virus Orthologous Groups” (pVOGs) usando BLASTX con un e-value máximo de 0.001, y un máximo de secuencias “target” para reportar un set de 50.



### **12.11. Análisis de las abundancias diferenciales de los “contigs de fagos”**

El reclutamiento de las lecturas al ensamble de contigs fue usado para construir una matriz de abundancias, aplicando el filtro de cobertura y longitud como ha sido previamente recomendado <sup>61</sup>. La cobertura fue definida a partir del mapeo de los reads (Bowtie2) a  $\geq 90\%$  de identidad y a  $\geq 80\%$  de la longitud del contig. Los mapeos de salida fueron convertidos, usando un script “in-house” en R, a una matriz de abundancias normalizadas del número de lecturas por kilobase por millón de lecturas secuenciadas por muestra (RPKM- por sus siglas en inglés: Reads Per Kilobase per Million) <sup>48</sup>. Esta matriz fue usada para determinar abundancias taxonómicas virales estadísticamente diferenciales entre los grupos usando la prueba de Kruskal Wallis y usando el método de Holm-Sidak, con un  $\alpha = 0.01$  para correcciones de pruebas múltiples. Esta matriz fue utilizada también para determinar abundancias de contigs individuales estadísticamente diferenciales usando DESeq2 con una razón de cambio ( $\log_2$ )  $\geq 2$  y un ajuste FDR (valor  $p \leq 0.05$ ) usando la corrección de Benjamini-Hochberg.

### **12.12. Riqueza y diversidad de los “contigs de fagos”**

La riqueza y diversidad de los contigs del ensamble fueron evaluadas en base a la mediana de 10,000 rarefacciones a una profundidad de la muestra más pequeña con base en la matriz de RPKM usando QIIME 1.9. Las medianas fueron obtenidas por muestra y por grupo, cuyos valores fueron evaluados para la normalidad, seguido de pruebas por pares no paramétricas de Mann–Whitney–Wilcoxon en R. Según el nivel de presencia de los “contigs de fagos” en las muestras, cada “contig de fago” fue clasificado como fagos “core”: detectado en  $>80\%$  de las muestras; fagos comunes: en  $>50\%$  y  $< 80\%$ ; y fagos individuales: apareciendo en  $< 50\%$  de la población total.

### **12.13. Prevalencia de los “contigs de fagos” altamente abundantes en NW**

La prevalencia de los contigs altamente abundantes que fueron encontrados en  $>80\%$  de las muestras dentro del grupo NW fue comparada a través de los tres grupos usando, por pares, pruebas no paramétricas de Mann–Whitney–Wilcoxon, en R.

### **12.14. Correlaciones de los fagos con bacterias y parámetros bioquímicos**

Todas las correlaciones fueron realizadas usando el coeficiente de correlación de Spearman en R, considerando todas las muestras. Utilizamos la abundancia en RPKM para los “contigs de fagos”, y la frecuencia relativa de los taxones bacterianos significativamente sobreabundantes previamente reportados entre NW, O y OMS, para la abundancia de la microbiota <sup>6</sup>. Solo seleccionamos las

correlaciones de Spearman con  $r^2 > 0.3$  y un valor de  $p \leq 0.05$ . Sin embargo, después de aplicar correcciones FDR para los valores de  $p$ , no obtuvimos correlaciones significativas por debajo de 0.05. Por lo tanto, para seguir atendiendo a las tendencias entre los grupos, nos hemos referido a valores de  $p$  no ajustados en los análisis de correlación.

### **12.15. Comparación de las distancias Euclidianas**

Las tablas de abundancias relativas de fagos fueron sometidas a una transformación logarítmica central con la librería `mixOmics v6.10.9` en R usando un valor mínimo de  $1e^{-7}$  para poder lidiar con registros de cero. La matriz Euclideana resultante fue entonces sometida a una reducción dimensional con un análisis de componentes principales (PcoA) (con `vegan 2.5-6` en R). Las distancias Euclidianas fueron calculadas para crear una matriz de adyacencia para las pruebas grupales con ANOSIM y `adonis` (realizado con `vegan`) y pruebas posthoc por pares de los grupos.

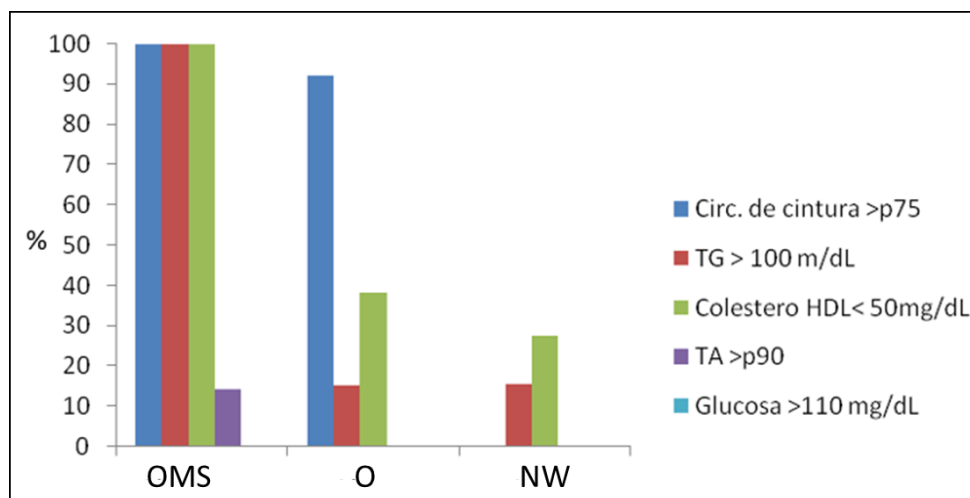
### **12.16. Disponibilidad de datos y software**

Los scripts que describen los procesos de análisis de los datos están disponibles en GitHub en <https://github.com/lab8a/2021-iScience-Phageome>. Los datos de secuenciación han sido depositados en NCBI bajo el número de acceso del “BioProject”: PRJNA646512.

## 13. RESULTADOS

### 13.1. Caracterización de la población de estudio

Este estudio se realizó a partir de una población de niños caracterizada por pertenecer a un estado socio-económico medio, compartir la misma ubicación geográfica, pertenecer al mismo grupo étnico y contar con estilos de vida similares (no practican ejercicio físico con regularidad). A partir de una cohorte de 750 muestras fecales previamente colectadas por el laboratorio del Dr. Samuel Canizales Quinteros de la *Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, UNAM/Instituto Nacional de Medicina Genómica (INMEGEN)*, analizamos los expedientes clínicos de las 333 muestras que fueron conservadas en “RNA later”, para estudios complementarios del metatranscriptoma. De ellas, 221 fueron eliminadas por criterios de exclusión, y otras 34 fueron eliminadas por presentar expediente incompleto. De las 78 muestras restantes, 51 (65.4%) provenían de niños sanos con peso normal saludable (NW), 13(16.7%) de niños obesos, y 14 (17.9%) de niños obesos con síndrome metabólico. El 51.8% de los niños obesos entre 7 y 10 años de edad presentaron síndrome metabólico, similar a resultados previos obtenidos de niños mexicanos obesos <sup>4,6</sup>. Dentro de la población de niños obesos con síndrome metabólico, los siguientes componentes fueron observados: obesidad abdominal (14 niños, 100%), hipertrigliceridemia (14 niños, 100%), colesterol HDL bajo (14 niños, 100%) e hipertensión arterial (2 niños, 14.3%) (Figura 5). Dentro de las poblaciones de niños obesos y de niños sanos también encontramos algunos de los componentes del síndrome metabólico (Figura 5). Dentro de la población de niños obesos, 12 presentaron obesidad abdominal, 2 tuvieron hipertrigliceridemia y 5 niños tuvieron colesterol HDL bajo. En la población de niños sanos, 8 tuvieron hipertrigliceridemia y 14 tuvieron colesterol HDL bajo. No observamos niños con hiperglucemia en ninguno de los tres grupos (figura 5).



**Figura 5. Componentes del síndrome metabólico en la población de niños con peso normal saludable (NW), obesos (O) y obesos con síndrome metabólico (OMS).** Se representa el porcentaje de niños que tienen componentes del síndrome metabólico dentro de cada población de estudio. TG: triglicéridos, TA: tensión arterial.

### 13.2. Selección de la población de estudio

Para caracterizar la composición del Fagoma seleccionamos 28 muestras conformadas por 10 niños de peso normal saludable (NW), 10 obesos (O) y 8 obesos con síndrome metabólico (OMS), de 7 a 10 años de edad, pareados por edad y género (Tablas 1 y 2). Para esta selección consideramos que los componentes del síndrome metabólico fueran los mismos entre los niños del grupo OMS para evitar factores de confusión en los resultados (Figura 6) y que los niños del grupo NW no tuvieran ningún componente del síndrome metabólico alterado.

Muestra ID	Género	Edad	Peso (Kg)	Altura (cm)	IMC (percentil)	CC (cm, percentil)	TA (mmHg, percentil)	Glucosa (mg/dL)	TG (mg/dL)	Colesterol HDL (mg/dL)
NW_118	F	9	34	142	56th	58.5, 25th	95/70, 23th/79th	81	79	50
NW_119	M	7	26	128.6	58th	56.3, 25th	98/65, 40th/69th	92	35	61
NW_120	F	8	29	134	56th	59.4, 25th	91/66, 19th/72th	91	44	71
NW_010	M	7	22	121.3	31th	61.4, 25th	89/67, 26th/80th	81	95	50
NW_124	M	10	32	141.2	30th	60.9, 25th	100/75, 38th/88th	86	57	54
NW_147	F	9	30	132.3	65th	63.1, 25th	89/60, 14th/51th	79	66	61
NW_161	M	9	28	134	36th	60.4, 25th	97/64, 38th/64th	94	42	78
NW_169	M	8	27	134	27th	58.5, 25th	99/71, 45th/82th	92	71	50
NW_193	M	9	22	122.2	18th	55.1, 25th	92/62, 38th/64th	85	42	75
NW_314	F	7	22	124.5	15th	49.6, 25th	86/61, 14th/61th	81	35	75
O_121	M	10	43	137.6	95th	80.4, >75th	100/68, 44th/74th	90	55	52
O_122	M	9	44	138.5	96th	87, >90th	99/62, 37th/52th	90	56	59
O_123	F	9	62	139.9	99th	100.4, >90th	100/72, 42th/85th	90	86	51
O_152	M	9	63	153.5	98th	96.5, >90th	113/68, 71th/64th	90	73	39
O_039	M	8	40	137	95th	76.7, >90th	86/54, 7th/27th	97	90	30
O_418	M	10	51	148.6	96th	80, >75 th	117/57, 84th/32th	91	64	41
O_420	F	7	28	119.9	95th	71, >90th	82/61, 10th/63th	91	93	52
O_434	F	8	38	135.7	95th	70.7, >75 th	105/59, 64th/46th	88	151	54
O_445	M	8	43	137.3	98th	77.3, >90th	120/79, 95th/94th	93	60	52
O_090	M	9	45	143.2	95th	84.4, >90 th	91/64, 11th/57th	96	72	40
OMS_124	M	9	59	146.1	98th	91.9, >90th	124/80, 96th/93th	81	135	47
OMS_125	M	7	52	136.4	99th	90.2, >90th	98/62, 32th/54th	89	152	33
OMS_126	F	9	44	131.2	98th	89.4, >90th	88/60, 14th/53th	79	126	42
OMS_288	M	7	48	136.5	99th	88.2, >90th	104/72, 55th/83	91	128	39
OMS_446	M	9	45	133.2	98th	83.3, >90th	119/78, 96th/94th	90	306	33
OMS_055	M	10	55	142.5	98th	95.6, >90th	88/60, 7th/46th	91	276	24
OMS_064	F	10	42	136	95th	77.6, >75th	99/68, 41th/76th	84	331	31
OMS_087	M	10	68	149.2	99th	103.7, >90th	100/61, 28th/45th	91	121	28

**Tabla 1. Características de las muestras seleccionadas.** M: masculino, F: femenino, CC: circunferencia de cintura, TA: tensión arterial, percentil Sistólica/Diastólica: percentil sistólica/percentil diastólica, TG: triglicéridos.

Muestra ID	Género	Edad		Muestra ID	Género	Edad
NW_119	Masculino	7	VS	OMS_125	Masculino	7
NW_010	Masculino	7	VS	OMS_288	Masculino	7
NW_314	Femenino	7	VS	O_420	Femenino	7
NW_120	Femenino	8	VS	O_434	Femenino	8
NW_169	Masculino	8	VS	O_418	Masculino	8
NW_118	Femenino	9	VS	O_123	Femenino	9
NW_147	Femenino	9	VS	OMS_126	Femenino	9
NW_161	Masculino	9	VS	O_122	Masculino	9
NW_193	Masculino	9	VS	O_152	Masculino	9
NW_124	Masculino	10	VS	OMS_055	Masculino	10
O_418	Masculino	10	VS	OMS_087	Masculino	10
O_090	Masculino	9	VS	OMS_446	Masculino	9
O_445	Masculino	8	VS	OMS_124	Masculino	9
OMS_064	Femenino	10	VS	O_121	Masculino	10

Tabla 2. Muestras pareadas por edad y género. VS: versus.

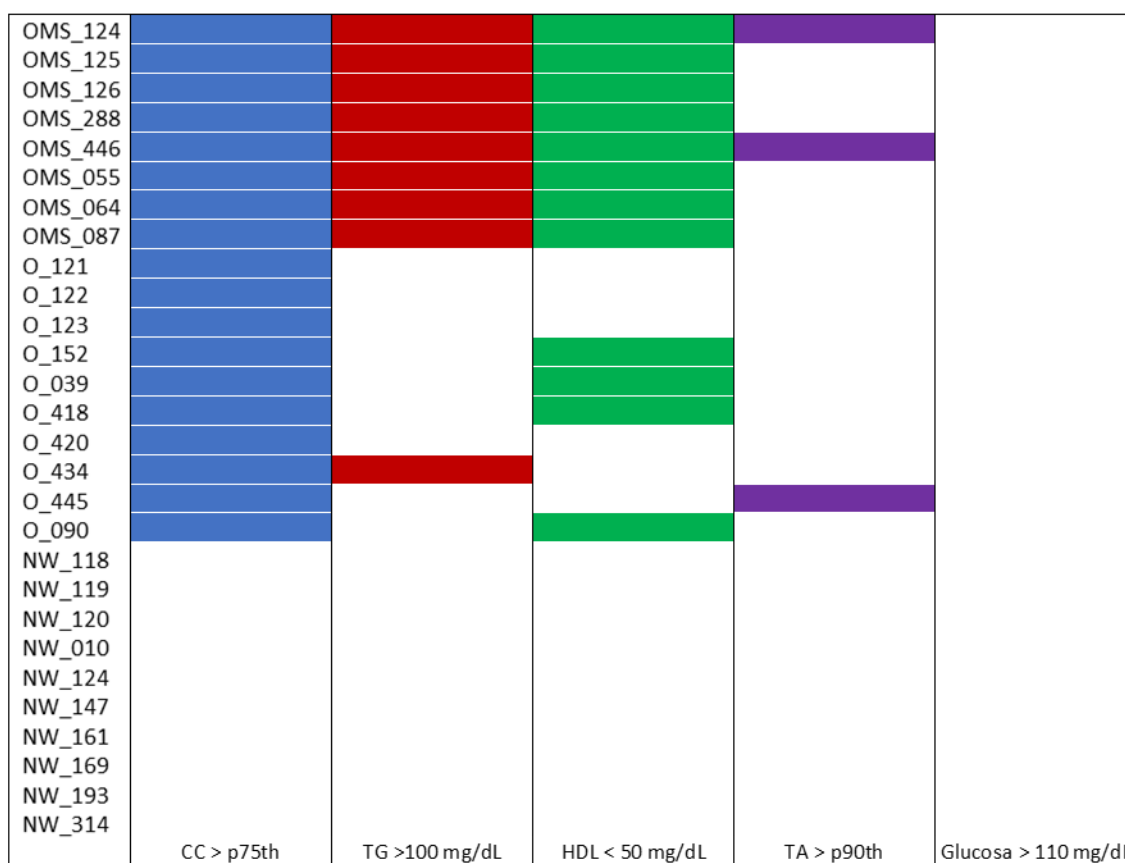
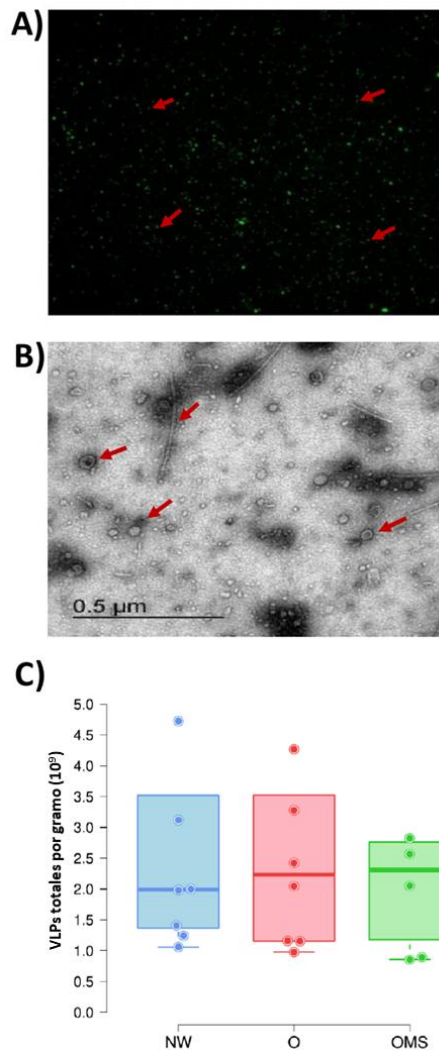


Figura 6. Visualización de la cantidad y tipo de componentes del síndrome metabólico para cada una de las muestras seleccionadas para el estudio. CC: Circunferencia de cintura, TG: Triglicéridos, TA: Tensión arterial.

### 13.3. El número de partículas virales (VLPs) es similar entre NW, O y OMS.

Con la finalidad de caracterizar el viroma intestinal de los niños sanos, obesos y obesos con síndrome metabólico, primero extrajimos las partículas virales derivadas de las muestras fecales de los tres grupos (ver Anexo 3). La microscopía de epifluorescencia (Figura 7A) y la microscopía electrónica de transmisión (TEM) (figura 7B) sugirieron la presencia de VLPs en todas las muestras. No hubo diferencia significativa en el número de VLPs entre los tres grupos, obteniendo un promedio de  $2.56 \times 10^9$ ,  $2.85 \times 10^9$ , y  $2.70 \times 10^9$  VLPs para NW, O y OMS, respectivamente (figura 7C y Anexo 3).

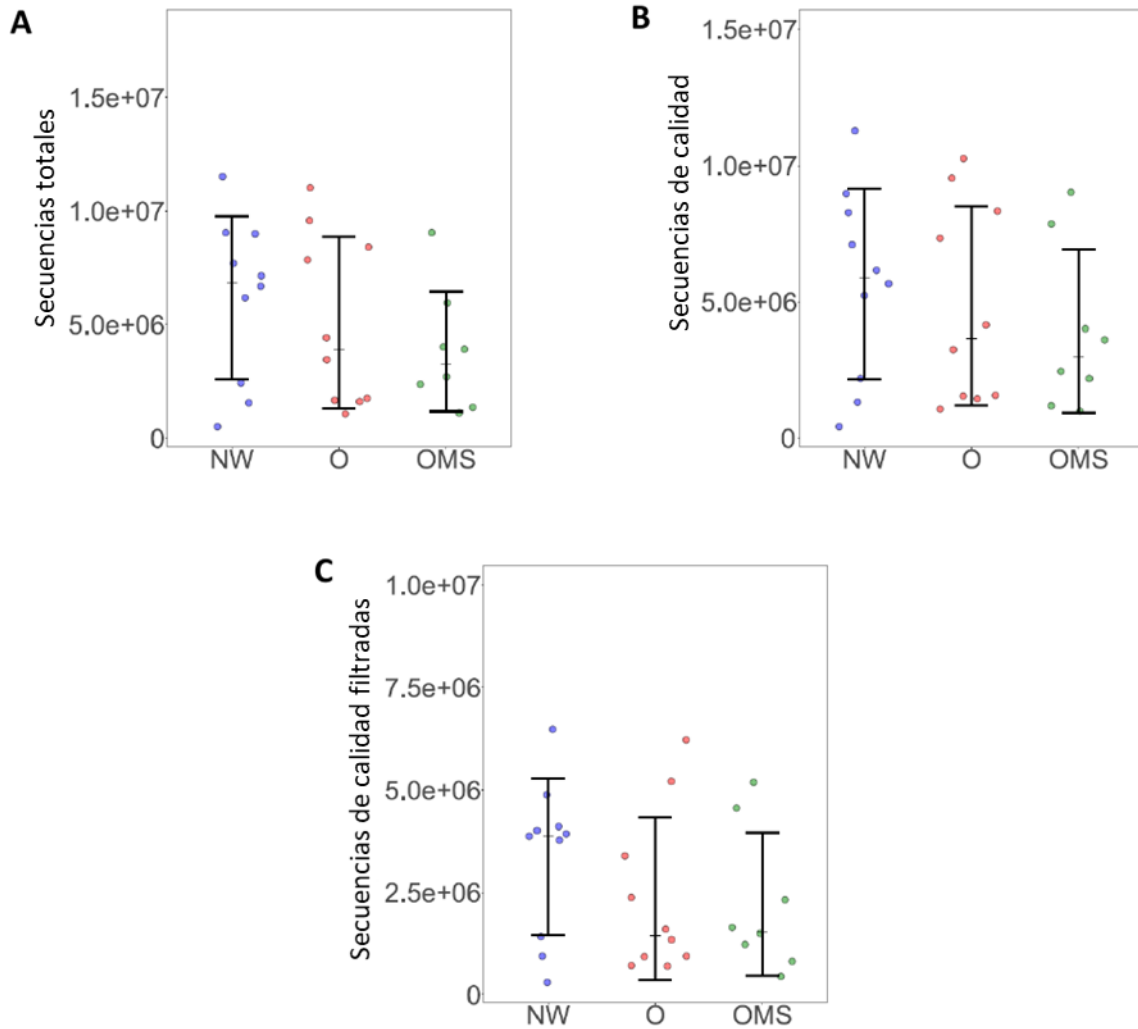


**Figura 7. Visualización microscópica y conteo de VLPs.** A) VLPs teñidas con SYBR Green I y visualizadas mediante microscopía de epifluorescencia. Las flechas rojas muestran ejemplos de VLPs. B) Microscopía TEM de VLPs. Las flechas rojas muestran un ejemplo de la morfología de las VLPs. C) Número de VLPs por gramo de materia fecal para cada grupo. Los puntos representan el número promedio de VLPs por muestra. Las barras de error indican la mediana y el rango intercuartílico. Las diferencias no fueron significativas (Ver Anexo 3).

#### **13.4. Perfiles funcionales de las lecturas sugieren presencia viral en las muestras de VLPs**

El ADN extraído de los VLPs fue tratado con un método de tagmentación (TAG) para la construcción de las librerías con la finalidad de evitar los sesgos de los métodos de amplificación de genomas completos típicamente utilizados debido a la baja cantidad de ADN. Después de aplicar los filtros de secuenciación de calidad, obtuvimos un promedio de 4,871,075 lecturas “paired-end” por muestra, produciendo 11.23 Gb de datos (Anexo 4). Para considerar solo a las secuencias potencialmente derivadas de los VLPs, las lecturas mapeadas a genomas de bacteria (~28%) y humano (~15%) fueron descartadas para los siguientes análisis (Anexo 4). La eliminación de las secuencias bacterianas puede llegar a ignorar lecturas virales potenciales a partir de un profago; sin embargo, preferimos eliminar todo el ADN potencialmente bacteriano para este estudio. Posteriormente, obtuvimos 74,859,356 lecturas, un promedio de 2,673,548 secuencias filtradas por calidad por muestra, sin diferencias significativas en la profundidad de secuencia entre los tres grupos (Figura 8 y Anexo 4), lo que sugiere una profundidad de secuenciación uniforme entre las muestras.

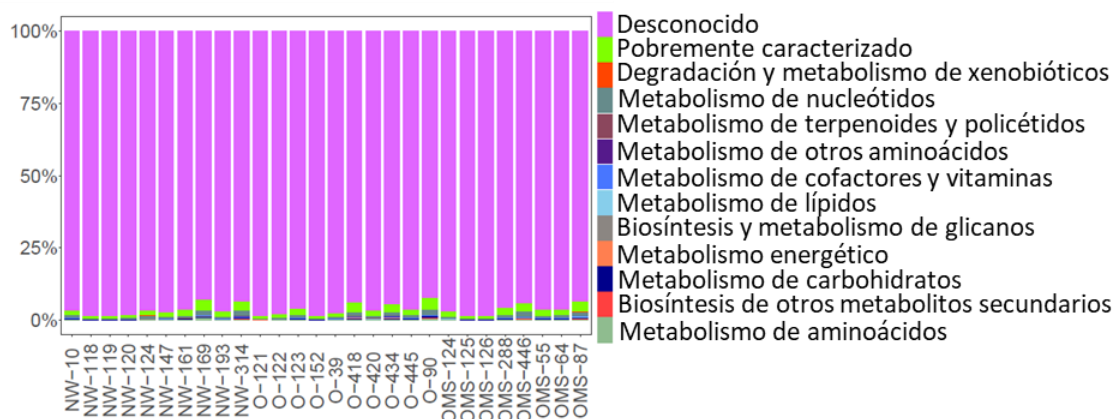




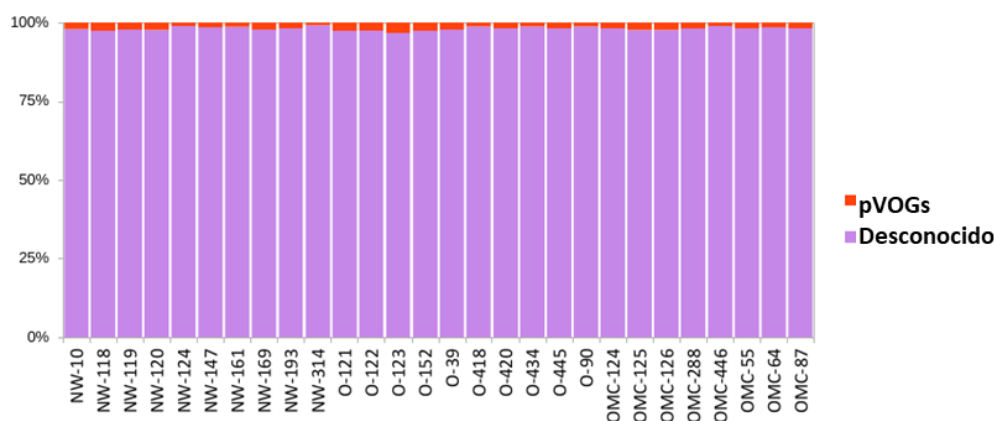
**Figura 8. Secuencias de VLPs detectadas en las muestras NW, O, y OMS.** Se presentan gráficas que contienen el número total de secuencias (A), secuencias filtradas por calidad (B), y secuencias de calidad filtradas (las secuencias de humano y bacteria fueron removidas) (C). Las barras de error indican la mediana y el rango intercuartílico. El número de secuencias por muestra se indica con puntos. No hubieron diferencias significativas entre los grupos.

Para obtener una primera aproximación del contenido funcional de las lecturas derivadas de los VLPs, anotamos las lecturas usando la base de datos de KEGG. Tal como ha sido reportado en estudios del viroma <sup>23,57,69,70</sup>, la mayoría de las lecturas ( $96.2 \pm 1.89\%$ ) mapearon a genes con función desconocida (Figura 9A). También realizamos una búsqueda de las lecturas derivadas de los VLPs contra Grupos Ortólogos de Virus Procariontas (pVOGs, por sus siglas en inglés, y encontramos que un pequeño número de lecturas ( $1.93 \pm 0.62\%$ ) se emparejaron con pVOGs (figura 9B), sugiriendo que las secuencias son de origen viral desconocido.

**A**



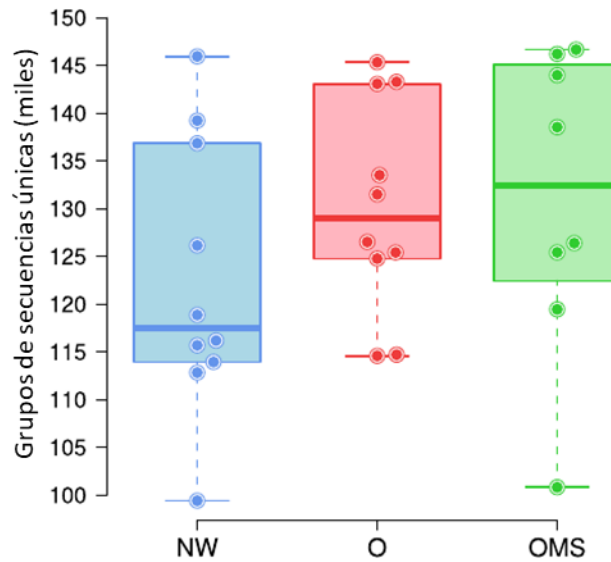
**B**



**Figura 9. Asignación funcional de VLPs.** Se presenta la abundancia relativa de las categorías de KEGG (A) y pVOGs (B) en las lecturas provenientes de los VLPs. “Desconocido” (color morado) indica la proporción de lecturas que no pudieron ser clasificadas funcionalmente en A o que no pudieron ser asignadas a las categorías de pVOGs en B.

### 13.5. El número de lecturas únicas, derivadas de los VLPs, muestra una riqueza incrementada en la enfermedad.

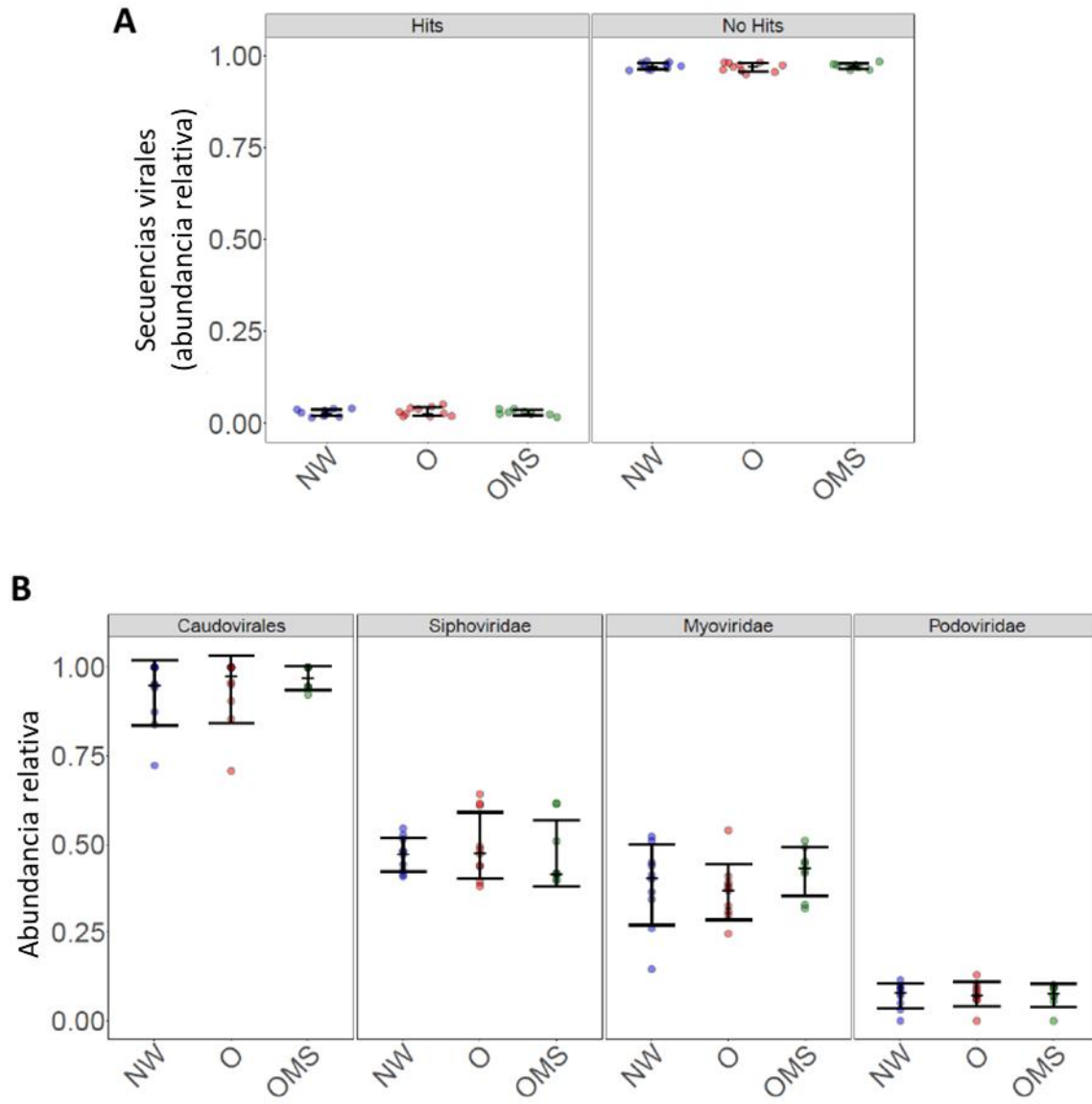
Realizamos 1,000 ejercicios al azar de 149,000 lecturas virales sub-muestreadas al azar y agrupadas a un 95% de identidad para generar grupos de secuencias únicas para cada muestra. De esta manera, pudimos analizar la riqueza de lecturas independientemente de la clasificación taxonómica y a la misma profundidad de secuencia para todas las muestras. Como resultado, encontramos un incremento de secuencias únicas en OMS y O en comparación a NW (Figura 10), aunque no fue significativo. Notablemente, esto sugiere que ambos grupos de obesidad incrementaron la riqueza de lecturas virales en comparación con el grupo de peso normal saludable.



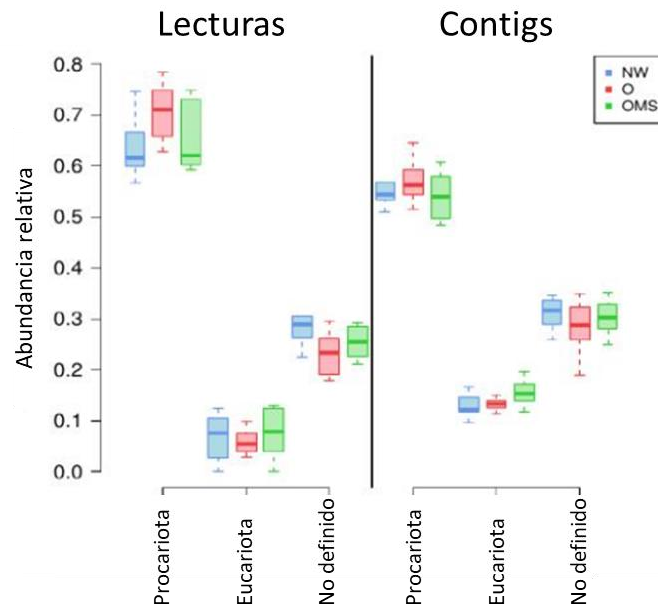
**Figura 10. Agrupaciones de secuencias únicas (eje y) por grupo.** Cada punto muestra la mediana de 1,000 iteraciones a una profundidad de secuencia de 149,000 lecturas por muestra. Las cajas muestran la distribución de datos para cada grupo. Las barras de error indican la mediana y el rango intercuartílico. No hubo diferencia significativa entre grupos.

### 13.6. Los bacteriófagos dominaron las lecturas virales del viroma intestinal

Después de agrupar todas las lecturas virales a un 95% de identidad en cada muestra, hubo una reducción de  $68 \pm 8\%$  de secuencias, resultando en un promedio de 856,825 secuencias únicas por muestra (Anexo 4). Como una primera aproximación para obtener el contenido viral potencial de las lecturas derivadas de los VLPs, éstas fueron emparejadas contra la base de datos de proteínas virales de RefSeq NR y encontramos que tan solo el  $2,95 \pm 0,95\%$  se emparejó (Figura 11A). De las lecturas emparejadas, el  $66,95 \pm 6,95\%$  clasificaron para virus procariotas,  $6,79 \pm 3,91\%$  para virus eucariotas, y  $26,26 \pm 5,79\%$  clasificaron como no definidas que incluyen virus desconocidos o “hits” múltiples entre virus eucariotas y procariotas (figura 12).



**Figura 11. Asignación taxonómica viral de las lecturas derivadas de los VLPs.** A. La abundancia relativa de las secuencias únicas que emparejaron contra una secuencia de proteína viral (hits), o que no emparejaron (no hits), se muestra en cada grupo. B. La abundancia relativa de las secuencias únicas asignadas a Caudovirales y sus miembros familiares taxonómicos en los grupos de NW, O y OMS. El número de secuencias únicas por muestra se muestra con puntos. Las barras de error indican la mediana y el rango intercuartílico.



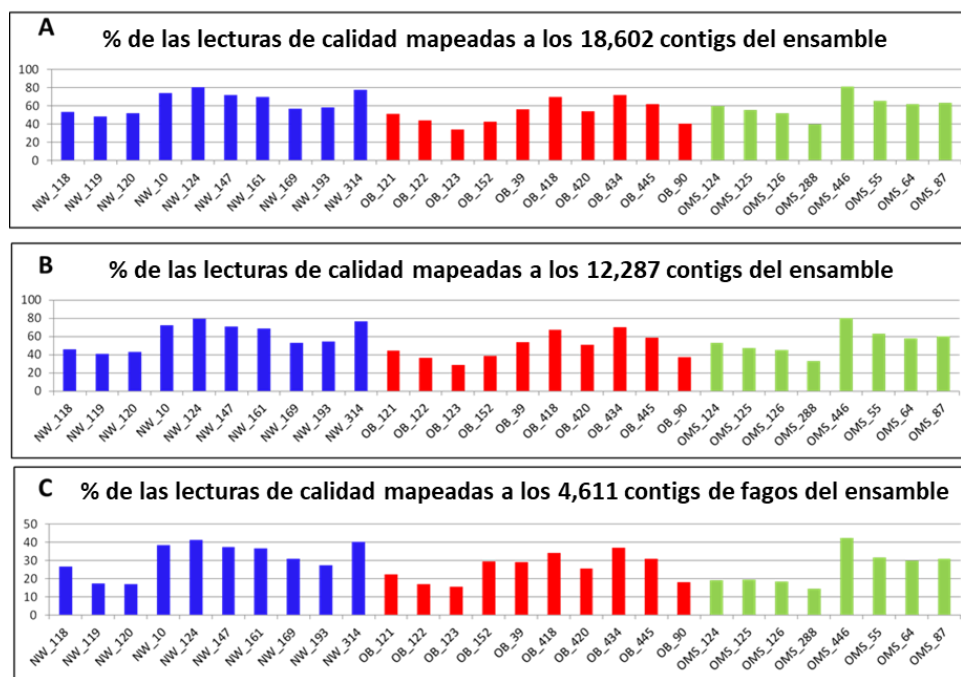
**Figura 12. Clasificación taxonómica de las lecturas y de los contigs virales.** Las abundancias relativas de las lecturas de secuenciación y de los contigs fueron asignadas a la clasificación viral indicada. Las barras de error indican la mediana y el rango intercuartílico.

### 13.7. El ensamble viral confirma el dominio de los bacteriófagos

Teniendo en cuenta que la longitud de los genomas de fagos podría afectar la abundancia obtenida de las lecturas de secuenciación<sup>57</sup>, realizamos un ensamble *de novo* usando los 74,859,356 lecturas virales provenientes de todas las muestras. Para evitar contigs quiméricos, seleccionamos aquellos contigs que tuvieron una cobertura  $\geq 80\%$  de su tamaño total por las lecturas virales en al menos una muestra, resultando en 18,602 contigs ( $\geq 500$  nt; el más largo: 176,210 nt; N50 = 7,480 nt) (Tabla 3). En promedio,  $58.69 \pm 12.79\%$  de las lecturas virales de cada muestra mapearon contra estos contigs, demostrando una contribución homogénea de todas las muestras al ensamble (Figura 13, Tabla 4). Posteriormente, eliminamos los contigs más pequeños ( $< 4$ Kb) para remover genomas virales potencialmente fragmentados. Después de esto, 12,887 contigs (N50 = 9,097 nt) (Tabla 3) fueron obtenidos y usados para los siguientes análisis. Importantly, esta reducción en el número de contigs no conllevó a una disminución drástica en el reclutamiento de lecturas, quedando en promedio  $54.85 \pm 14.60\%$  de las lecturas virales de cada muestra (Figura 13, Tabla 4).

Ensamble	Ensamble de 18,602 contigs	Ensamble de 12,287 contigs	Ensamble de 4,611 contigs de fagos
# contigs ( $\geq 0$ bp)	18602	12287	4611
# contigs ( $\geq 1000$ bp)	18602	12287	4611
# contigs ( $\geq 5000$ bp)	8760	8760	3445
# contigs ( $\geq 10000$ bp)	2621	2621	1158
# contigs ( $\geq 25000$ bp)	364	364	184
# contigs ( $\geq 50000$ bp)	69	69	40
Longitud total ( $\geq 0$ bp)	127030619	105263498	43100311
Longitud total ( $\geq 1000$ bp)	127030619	105263498	43100311
Longitud total ( $\geq 5000$ bp)	89542009	89542009	37896750
Longitud total ( $\geq 10000$ bp)	47836069	47836069	22217717
Longitud total ( $\geq 25000$ bp)	14977436	14977436	7975753
Longitud total ( $\geq 50000$ bp)	5068871	5068871	3137340
# contigs	18602	12287	4611
Contig más largo	176210	176210	176210
Longitud total	127030619	105263498	43100311
GC (%)	49.69	49.7	49.71
N50	7480	9097	10370
N75	4615	5841	6332
L50	4450	3125	1093
L75	9954	6797	2452
# N's per 100 kbp	0.05	0.05	0

**Tabla 3. Resultado de todos los análisis “Quast” de los ensamblajes virales.** Segunda columna: ensamblaje del viroma completo, Tercera columna: ensamblaje de los contigs >4Kb, Cuarta columna: ensamblaje de los 4,611 contigs clasificados como fagos.

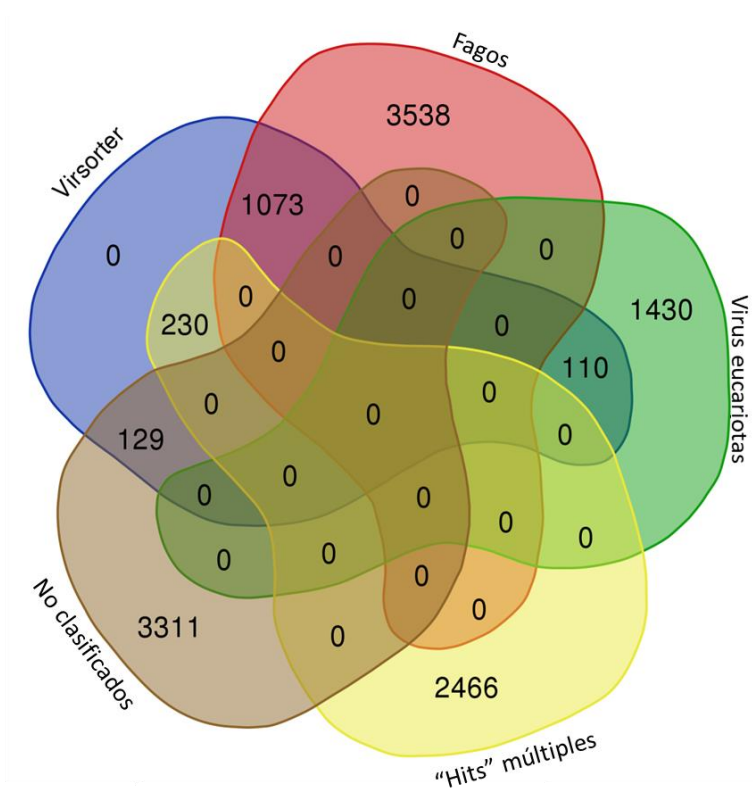


**Figura 13. Porcentaje de las lecturas de calidad que mapearon contra los diferentes ensamblajes de contigs.** A) Ensamblaje de los 18,602 contigs (ensamblaje del viroma completo), B) Ensamblaje de los 12,287 contigs >4Kb, C) Ensamblaje de los 4,611 contigs clasificados como fagos.

Muestra ID	Tipo de enfermedad	Lecturas de calidad filtradas (paired_seq)	Lecturas de calidad mapeadas a los 18,602 contigs del ensamble	% de lecturas mapeadas a los 18,602 contigs del ensamble	Lecturas de calidad mapeadas a los 12,287 contigs del ensamble	% de lecturas mapeadas a los 12,287 contigs del ensamble	Lecturas de calidad mapeadas a los 4,611 contigs de fagos del ensamble	% de lecturas mapeadas a los 4,611 contigs de fagos del ensamble
NW_118	Peso normal saludable	1415490	1043862	73.75	1021513	72.17	544750	53.33
NW_119	Peso normal saludable	6483370	3446174	53.15	3000980	46.29	1723270	57.42
NW_120	Peso normal saludable	4105736	1964917	47.86	1680101	40.92	703738	41.89
NW_10	Peso normal saludable	3930198	2020079	51.40	1689329	42.98	663764	39.29
NW_124	Peso normal saludable	3873168	3127816	80.76	3084258	79.63	1600132	51.88
NW_147	Peso normal saludable	3777848	2720684	72.02	2674804	70.80	1407664	52.63
NW_161	Peso normal saludable	4883020	3413536	69.91	3350099	68.61	1780477	53.15
NW_169	Peso normal saludable	2995550	170192	56.82	159755	53.33	92395	57.84
NW_193	Peso normal saludable	4007412	2338006	58.34	2174355	54.26	1093916	50.31
NW_314	Peso normal saludable	945188	733140	77.57	728349	77.06	379352	52.08
O_121	Obeso	6222168	3160537	50.79	2777058	44.63	1379468	49.67
O_122	Obeso	3390676	1478086	43.59	1234603	36.41	575184	46.59
O_123	Obeso	714182	242360	33.94	203736	28.53	110118	54.05
O_152	Obeso	5213758	2212411	42.43	2032799	38.99	1536439	75.58
O_39	Obeso	2377282	1336861	56.23	1273595	53.57	691467	54.29
O_418	Obeso	933454	647537	69.37	631942	67.70	318605	50.42
O_420	Obeso	1345222	726624	54.02	686337	51.02	343508	50.05
O_434	Obeso	945650	679514	71.86	667387	70.57	350182	52.47
O_445	Obeso	1599244	982795	61.45	946904	59.21	495001	52.28
O_90	Obeso	699572	282616	40.40	263250	37.63	126100	47.90
OMS_124	Obeso con Síndrome Metabólico	2319728	1376347	59.33	1240376	53.47	442514	35.68
OMS_125	Obeso con Síndrome Metabólico	5190602	2889218	55.66	2471914	47.62	1004408	40.63
OMS_126	Obeso con Síndrome Metabólico	4555974	2352597	51.64	2066334	45.35	828660	40.10
OMS_288	Obeso con Síndrome Metabólico	1229670	481883	39.19	410001	33.34	175817	42.88
OMS_446	Obeso con Síndrome Metabólico	817376	661852	80.97	656773	80.35	345680	52.63
OMS_55	Obeso con Síndrome Metabólico	1641244	1077006	65.62	1032205	62.89	515729	49.96
OMS_64	Obeso con Síndrome Metabólico	1497064	929203	62.07	866113	57.85	445543	51.44
OMS_87	Obeso con Síndrome Metabólico	445510	281515	63.19	270091	60.63	138128	51.14

**Tabla 4. Número de lecturas mapeadas a los ensamblados de contigs.**

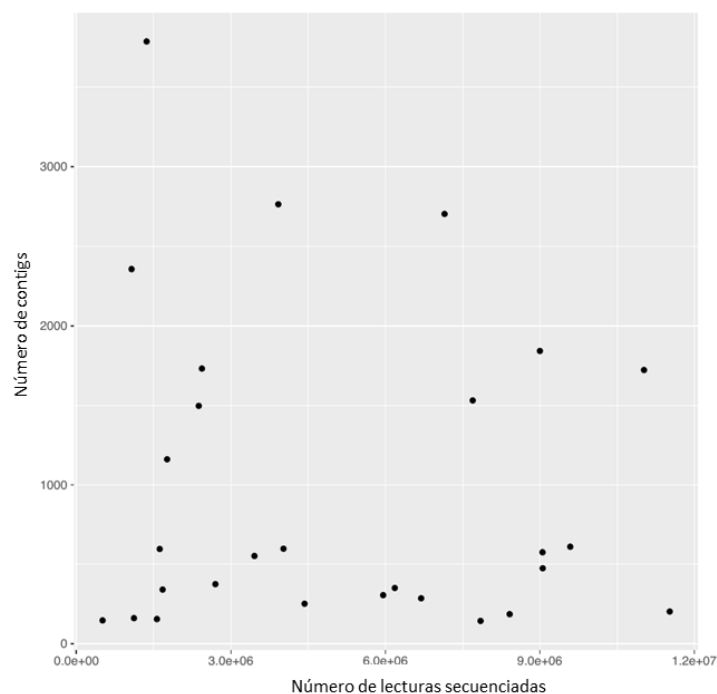
Encontramos, en promedio, 7.20 genes por contig (0.87 genes por kb de la longitud del contig). Clasificamos a estos contigs usando el ADN y la codificación de proteínas, obteniendo 4,611 contigs como virus procariotas potenciales, 1,540 como virus eucariotas potenciales, 2,696 contigs con “hits” múltiples a virus procariotas y eucariotas, y 3,440 contigs con un origen desconocido. De los 4,611 contigs procariotas ( $\geq 4\text{Kb}$ ), que pudieran estar representando genomas completos o parciales, 1,307 fueron clasificados usando las secuencias de ADN (dc-megablast) y 3,304 usando sus proteínas codificadas (BLASTx). Por otro lado, clasificamos los 12,287 contigs usando VirSorter, obteniendo una clasificación viral para 1,542 de ellos. Contrariamente, 5,949 contigs emparejaron con un pVOG. De forma interesante, la mayoría de los contigs clasificados usando VirSorter (69.58%) y pVOG (77.16%) coincide con nuestra clasificación de virus procariotas (Figura 14), reforzando nuestra estrategia de clasificación de fagos usando tanto ADN como proteínas. Estudios previos del viroma han reportado que entre el 29.35% y el 48.5% de los contigs virales mapearon a pVOG <sup>71,72</sup>.



**Figura 14. Diagrama de Venn mostrando la superposición entre las tres clasificaciones obtenidas de la NR y NT y de los contigs clasificados usando VirSorter.**

En promedio, la clasificación de contigs reveló,  $56.30 \pm 5.50\%$  de virus procariotas y el  $13.94 \pm 2.80\%$  de virus eucariotas por muestra. Notablemente, esto fue similar a la clasificación taxonómica obtenida de las lecturas (Figura 12), sugiriendo que la diversidad de secuencias observada en las lecturas fue también capturada en los contigs ensamblados. Como se esperaba debido al uso del método de tagmentación, la mayoría de los contigs procariotas fueron virus dsDNA (96%), mientras que solo el 0.1% fueron anotados como virus ssDNA, y el 4% restante fueron virus bacterianos no clasificados. No observamos ninguna influencia en el número de lecturas versus el número de contigs virales obtenidos por muestra (Figura 15).





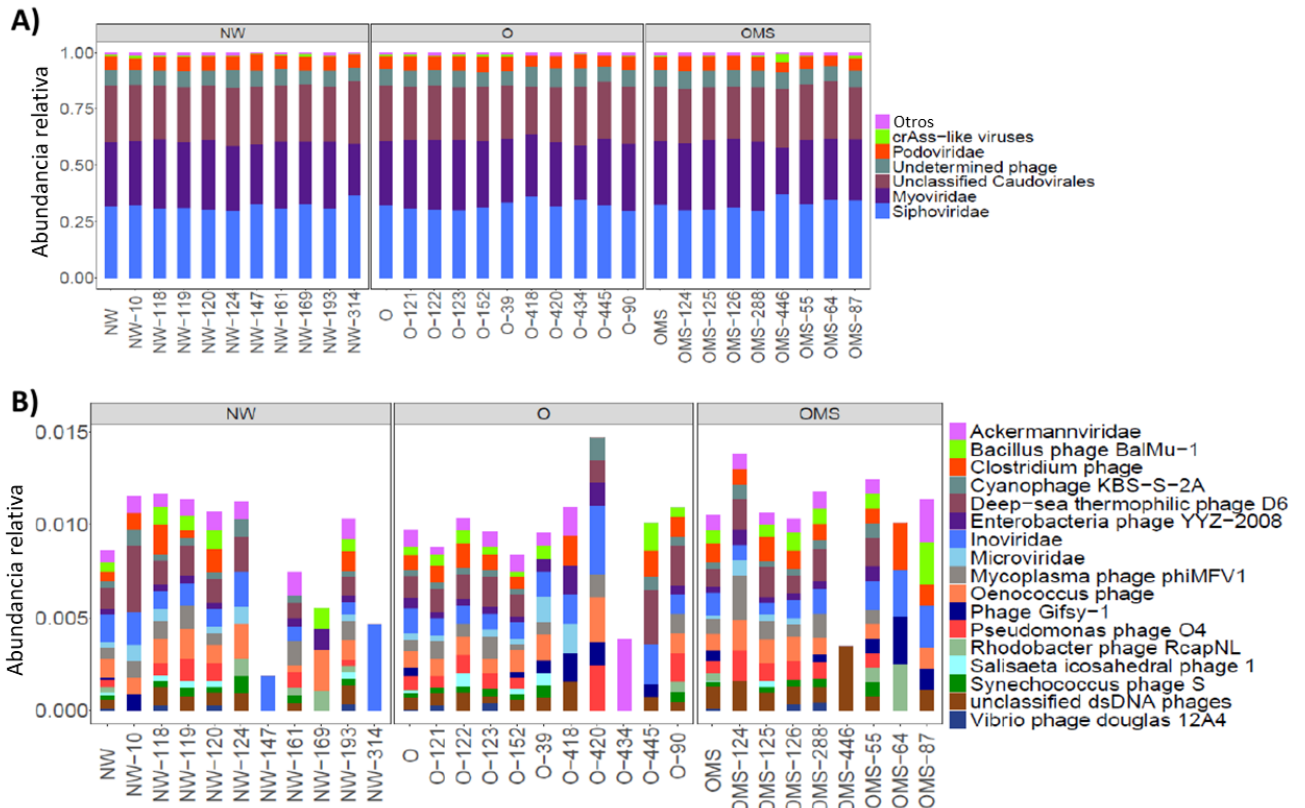
**Figura 15.** Correlación entre las lecturas de calidad y el número de contigs por muestra.  $R^2 = 0.0131$ ,  $p = 1$ .

### **13.8. El fagoma intestinal de niños estuvo compuesto principalmente de Caudovirales**

Los 4,611 contigs (N50 = 10,370 nt, max 176,210 nt; promedio 9,347 nt) (Tabla 3) clasificados como virus procariotas potenciales fueron seleccionados como el Fagoma y representó 37.53% de los contigs totales del ensamble del viroma, a partir de aquí mencionados como contigs de fagos. El Fagoma reclutó más de la mitad de las lecturas virales (50.27%) que fueron mapeadas al ensamble del viroma original y estuvieron distribuidas en un promedio de 707,571 lecturas por muestra (Tabla 4).

Los Caudovirales fueron los contigs de fagos más abundantes (91.28±0.10%), seguidos por los virus crAss-like (0.64±0.73%) (figura 16). El alto número de contigs clasificados como Caudovirales (Figura 11B), sugiere que nuestro ensamble de genomas también refleja la taxonomía de las lecturas por sí mismas. Dentro de los contigs pertenecientes a la orden de los Caudovirales, las familias más abundantes fueron Siphoviridae (35.28±0.02%), Myoviridae (31.33±0.03%), y Podoviridae (6.50±0.01%).

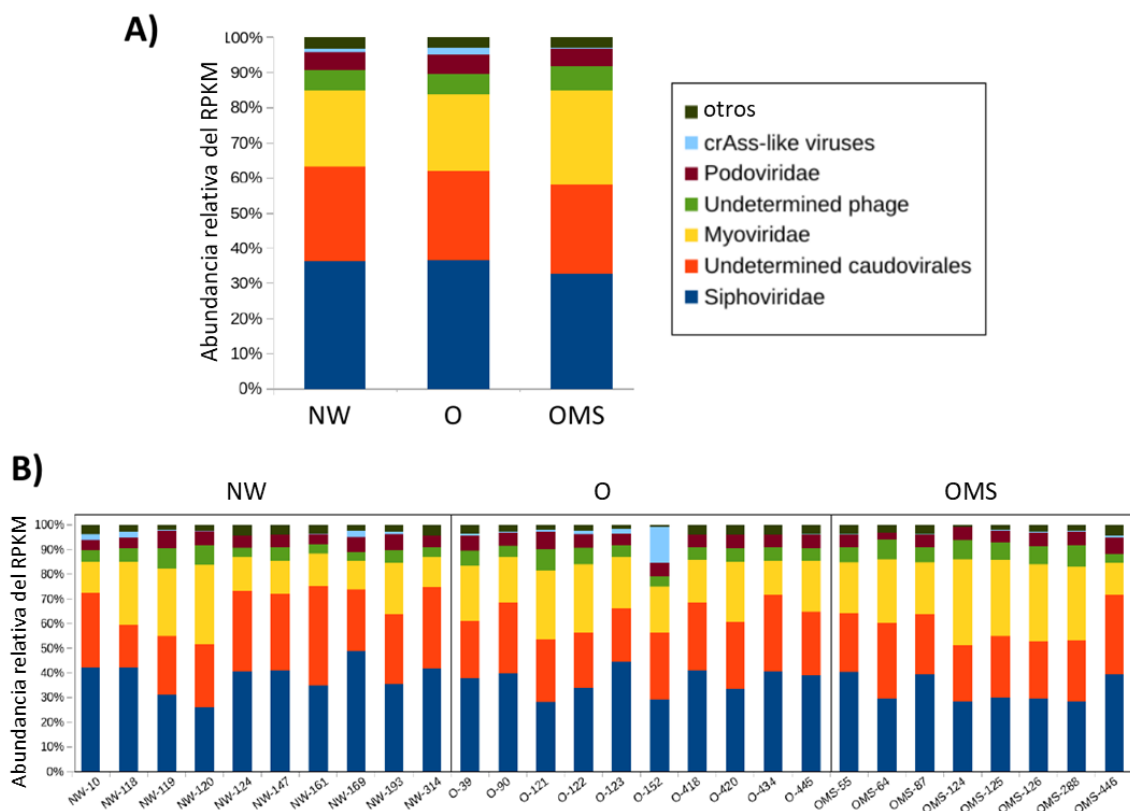
Dada la importancia de los fagos crAss-like en los viromas adultos <sup>73,74</sup>, analizamos su presencia en nuestro Fagoma, y encontramos que un promedio de 0.64±0.01% de los contigs pudieron ser clasificados como fagos crAss-like (Anexo 5). Por otro lado, encontramos que el crAssphage Mexicano <sup>75</sup> estuvo presente en 25 de las 28 muestras.



**Figura 16: Abundancia relativa de los contigs de fagos clasificados taxonómicamente por muestra y por grupo. A.** Abundancia relativa de los contigs de fagos más abundantes. “Otros” (color morado) indica la suma de la abundancia relativa de los contigs de fagos mostrados en B. **B.** Abundancia relativa de los contigs de fagos menos abundantes.

### 13.9. El aumento de la riqueza, diversidad, y dominancia de los contigs de fagos está vinculado al cambio de peso normal saludable a obesidad

Normalizamos la abundancia de los contigs de fagos mediante “el número de Lecturas por Kilobase por Millon de lecturas secuenciadas por muestra” (RPKM- por sus siglas en inglés: “Reads Per Kilobase per Million”) (ver métodos) por muestra para comparar las métricas de abundancia y diversidad de los fagos entre los grupos<sup>48</sup>. Posteriormente, comparamos las abundancias taxonómicas de los fagos clasificados a través de nuestros grupos (Figura 17). Encontramos una disminución en la abundancia de Siphoviridae y crAss-like (32.72 y 0.18%, respectivamente) en OMS en comparación con O (36.66 y 2.2%, respectivamente) y NW (36.50 y 1.0%, respectivamente). También observamos un aumento en la abundancia de Myoviridae en OMS (26.72%) en relación a O (21.77%) y NW (21.70%). Sin embargo, estas abundancias no fueron significativamente diferentes entre los tres grupos.

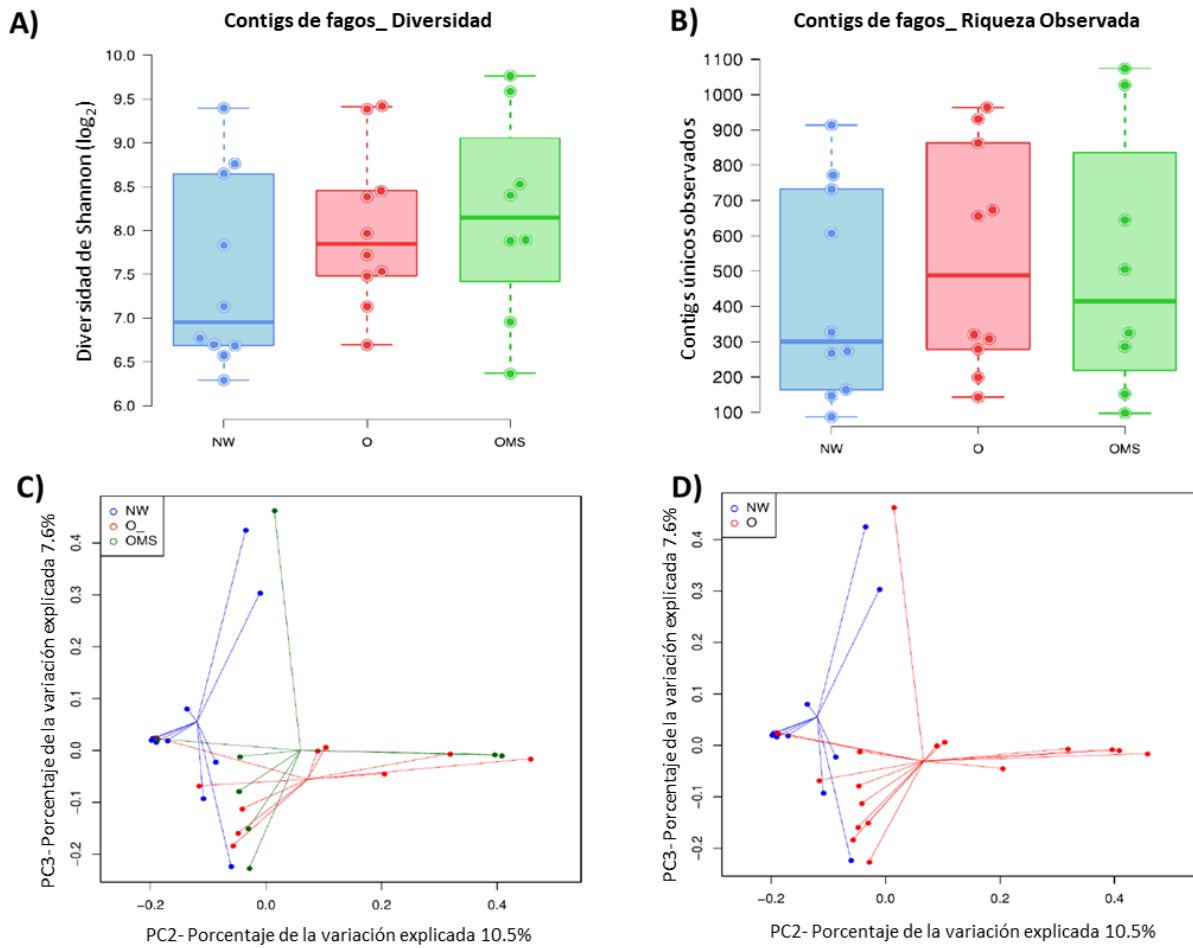


**Figura 17. Abundancia relativa del RPKM normalizado para los 4,611 contigs de fagos.** A) Promedio por grupo. B) Promedio por muestra. “otros” en la leyenda representa la abundancia relativa de los contigs de fagos menos abundantes.

Las diversidades alpha del Fagoma mostraron que la riqueza y la diversidad de Shannon incrementaron en O y OMS en comparación a NW (Figura 18A-B), aunque las diferencias entre grupos no fueron significativas. Notablemente, estos cambios en la diversidad sucedieron manteniendo números similares de conteos de VLPs en los tres grupos (Figura 7C). el grupo O tuvo la riqueza más alta, seguido por OMS y NW (Figura 18B), mientras el grupo OMS exhibió la diversidad más alta, seguida por O y NW (Figura 18A). Notablemente, esta riqueza incrementada en los grupos de obesidad estuvo también respaldada por nuestro análisis inicial de agrupamientos de lecturas virales (Figura 10). También observamos que 488 (10.58%) contigs de fagos representaron el 70% de las lecturas normalizadas en las muestras NW, mientras 679 (14.73%) y 831 (18.02%) contigs de fagos representaron el 70% de las lecturas normalizadas en los grupos O y OMS. Estos resultados sugieren un incremento considerable en el número de contigs de fagos dominantes debido a la enfermedad.

Por otro lado, el PCoA basado en las distancias de Bray-Curtis mostraron agrupaciones que no fueron auto-consistentes con el grupo de estudio (Figura 18C y Anexo 6). Sin embargo, cuando todas

las muestras de obesos fueron unidas (O + OMS), éstas formaron un grupo separado de las muestras NW (Figura 18D y Anexo 6), aunque esta diferencia no fue significativa ( $p= 0.568$ ). También realizamos un análisis de distancias de Aitchison sujeto a una reducción dimensional con un análisis de componentes principales que mostró agrupaciones separadas no significativas (Anexo 7).

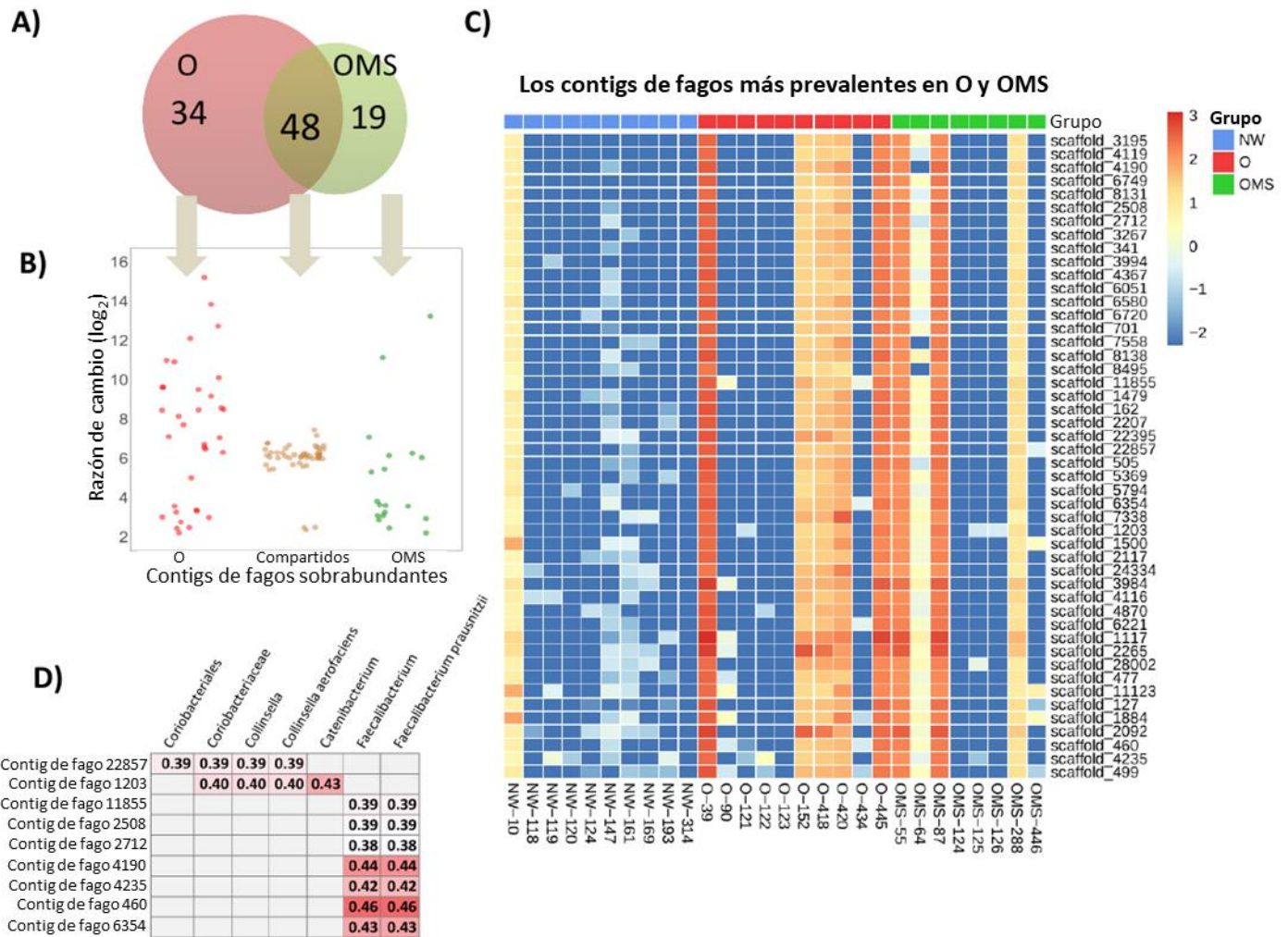


**Figura 18. Diversidad Alpha y Beta del fagoma.** A) Diversidad de los contigs de fagos. Cada punto representa la mediana de 10,000 cálculos de la diversidad de Shannon a una profundidad uniforme para una muestra. Las cajas muestran la distribución de datos en cada grupo. B) Riqueza de los contigs de fagos. Cada punto representa la mediana de 10,000 cálculos de contigs de fagos observados a una profundidad uniforme para una muestra. Las cajas muestran la distribución de datos dentro de cada grupo. Las diferencias en A y B no fueron significativas. C) Análisis de Coordenadas Principales (PCoA) basado en la disimilitud de Bray-Curtis para las muestras NW (azul), O (rojo) y OMS (verde). D) PCoA basado en la disimilitud de Bray-Curtis para todas las muestras de obesos (O + OMS) (rojo) y NW (azul). Ver anexos 4 y 5.

### **13.10. Varios contigs de fagos estuvieron significativamente sobreabundantes en la obesidad y en el síndrome metabólico**

Identificamos los contigs de fagos que resultaron estar significativamente sobreabundantes en los grupos O y OMS en comparación con el de NW usando todos los fagos con la matriz de abundancias normalizadas. Este procedimiento es análogo al uso de tablas de RPKM para identificar a los genes expresados diferencialmente en experimentos de RNA-seq. Después de esto, obtuvimos 111 y 107 contigs de fagos significativamente sobreabundantes en O y OMS, respectivamente; y las diferencias fueron estadísticamente significativas con un  $\text{Alpha} = 0.05$  en relación a NW. De éstos, solo seleccionamos los contigs de fagos compartidos entre al menos el 30% de las muestras del grupo de O o del grupo de OMS para eliminar fagos individuales. Esto resultó en 82 y 67 contigs de fagos significativamente sobreabundantes en O y OMS, respectivamente, en comparación con NW, y con 48 contigs de fagos compartidos entre O y OMS (Figura 19A-B).

Posteriormente, evaluamos si las abundancias de los 48 contigs de fagos (Figura 19C) estuvieron asociadas a cambios paralelos en la población bacteriana, sobre todas las muestras. Para responder esta pregunta, seleccionamos los datos de secuenciación del gen 16S rRNA de los 41 taxones bacterianos significativamente asociados a obesidad y síndrome metabólico <sup>6</sup> y calculamos la correlación de Spearman entre ellos y la abundancia de los 48 contigs de fagos en todas las muestras. Encontramos que la abundancia de 9 contigs de fagos correlacionaron con las abundancias bacterianas asociadas a obesidad y síndrome metabólico (Figura 19D). Las muestras que conllevaron a la correlación entre contigs de fagos y bacterias fueron las pertenecientes a los grupos de O y OMS (Anexo 8). Realizamos el mismo análisis de correlación para los 34 y 19 contigs de fagos específicamente sobreabundantes en OMS y O (Figura 19A). Sin embargo, las correlaciones fueron principalmente causadas por muestras aisladas (“outliers”) y valores de abundancia iguales a cero para la mayoría de los contigs de fagos. Los 19 contigs de fagos que estuvieron solamente sobreabundantes en OMS podría ser estudiado más a fondo como biomarcadores vinculados al desarrollo de síndrome metabólico en niños obesos.

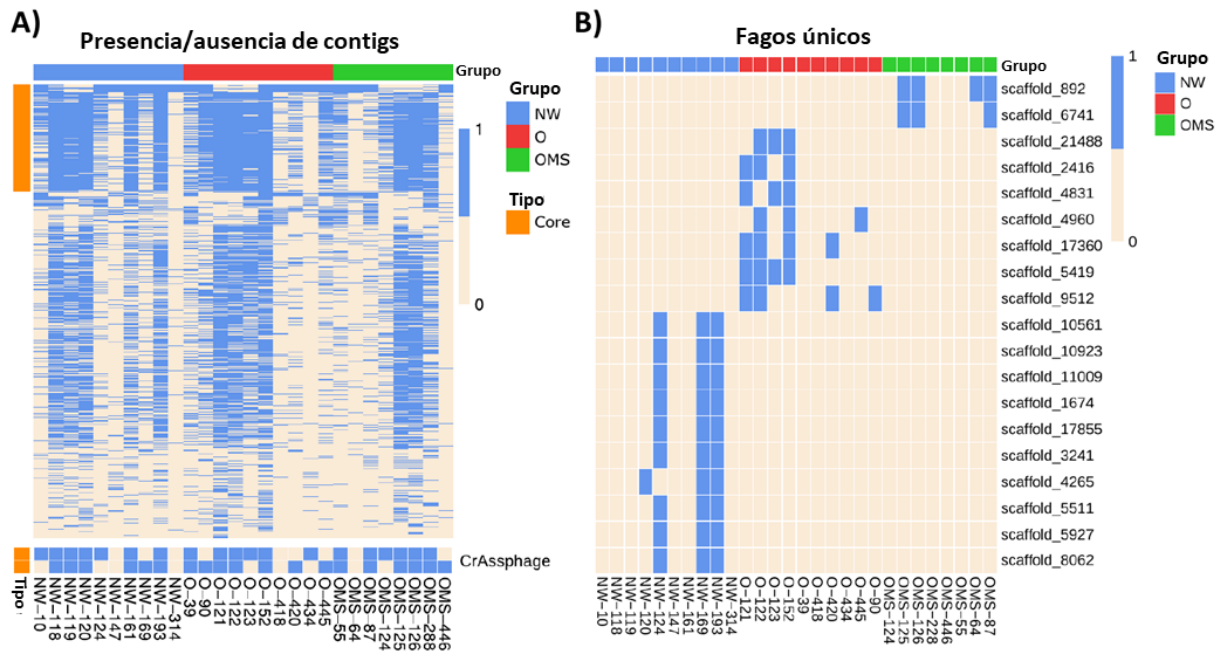


**Figura 19. Análisis de los contigs de fagos sobreabundantes en la obesidad (O) y en la obesidad con síndrome metabólico (OMS).** A) Diagrama de Venn del número de contigs de fagos sobreabundantes en O (círculo rojo) y en OMS (círculo verde), en comparación con NW (círculo azul). El número de contigs exclusivos de O es 34, exclusivos de OMS es 19, y compartidos es 48. B) Los niveles de expresión (razón de cambio en  $\log_2$ ) de los contigs de fagos sobreabundantes del diagrama de Venn. Los 34 (puntos rojos), 19 (puntos verdes), y 48 (puntos marrones) contigs de fagos de O, OMS y los compartidos entre los dos grupos, respectivamente. C) "Heat-map" de las abundancias normalizadas en RPKM de los 48 contigs de fagos sobreabundantes en O y OMS. Se muestra la distribución de las abundancias para cada contig entre todas las muestras. D) Correlación de Spearman de la abundancia (RPKM) de los 48 contigs de fagos sobreabundantes en la enfermedad y la abundancia relativa de los taxones bacterianos 16S identificados por estar significativamente asociados con la obesidad. Solo las correlaciones significativas con un valor de p no ajustado ( $\leq 0.05$ ) fueron presentadas. Ver también Anexo 8.

### **13.11. El Fagoma fue principalmente específico para cada individuo**

La matriz de abundancias normalizadas fue analizada para evaluar la composición del Fagoma en todas las muestras, independientemente del estado de salud de los individuos. De los 4,611 contigs de fagos, solo dos estuvieron presentes en todos los 28 individuos y 48 en más de 23 individuos (>80%), los cuales fueron nombrados los “fagos core”. La mayoría de los contigs de fagos (3,477) estuvieron compartidos en menos de 14 individuos (50% de la población) (Figura 20A). Estos resultados sugieren que la mayor parte del Fagoma fue específico para cada individuo, siendo solo 48 fagos (1.04% del Fagoma) considerados como los “fagos core”. Además, dos de los “fagos core” fueron identificados como fagos “crAss-like” putativos. La presencia de fagos “crAss-like” como parte del “core” da sustento a nuestra definición para “fagos core” debido a que éstos son los fagos más abundantes reportados en el intestino humano de adultos <sup>73</sup>.

No encontramos contigs de fagos únicos y específicos para cada grupo (que estuvieran presentes en 100% de las muestras de un grupo y ausentes en los otros grupos). Esto coincide con la alta presencia de fagos inter-individuales que observamos. Para encontrar posibles fagos únicos, usamos diferentes “cut-offs” de presencia de fagos entre la población, de  $\geq 20$  a  $\geq 80\%$  de las muestras en un grupo y que estuvieran ausentes en todas las muestras de los otros grupos. Encontramos que solo usando un “cut-off” de contigs con  $\geq 30\%$  de presencia, obtuvimos la mayor cantidad de contigs únicos, obteniendo dos contigs de fagos únicos en el grupo OMS, 7 contigs de fagos únicos en el grupo O, y 10 contigs de fagos únicos en el grupo NW (Figura 20B).

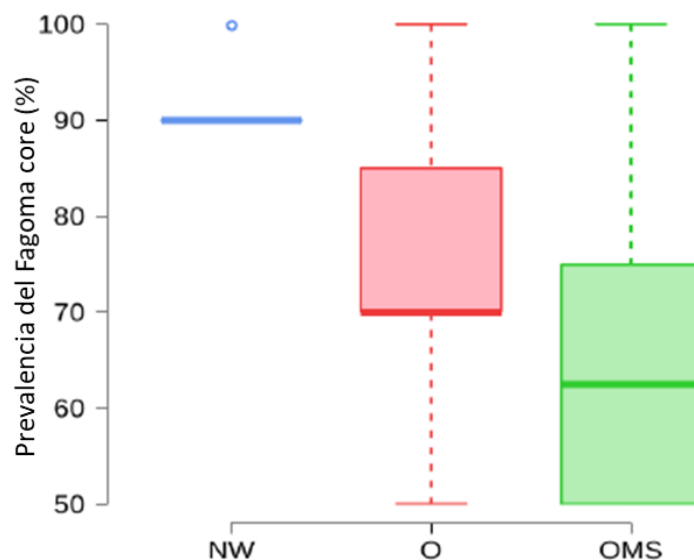


**Figura 20. “Heat-maps” de presencia-ausencia de la distribución de los contigs de fagos entre todas las muestras.** A) Distribución de los contigs de fagos “core” a través de todas las muestras. Los contigs de “crAssphage-like” provenientes del grupo “core” son presentados de forma separada en la base. B) Fagos únicos para cada grupo.

### 13.12. La enfermedad alteró la prevalencia de los contigs de fagos altamente abundantes en NW

Comparamos la prevalencia de los contigs de fagos con una alta presencia (>80% de las muestras) en NW con respecto a las muestras de O y OMS. Encontramos que 52 contigs de fagos estuvieron presentes en >80% de las muestras NW con un promedio de 91.54%. Por el contrario, su prevalencia estuvo significativamente reducida a 76.35% y 68.27% en los individuos O y OMS, respectivamente ( $p = <0.0001$ ) (Figura 21). Estos resultados mostraron que la prevalencia de los contigs de fagos con la presencia más alta en NW estuvo significativamente alterada en la obesidad y en la obesidad con síndrome metabólico.





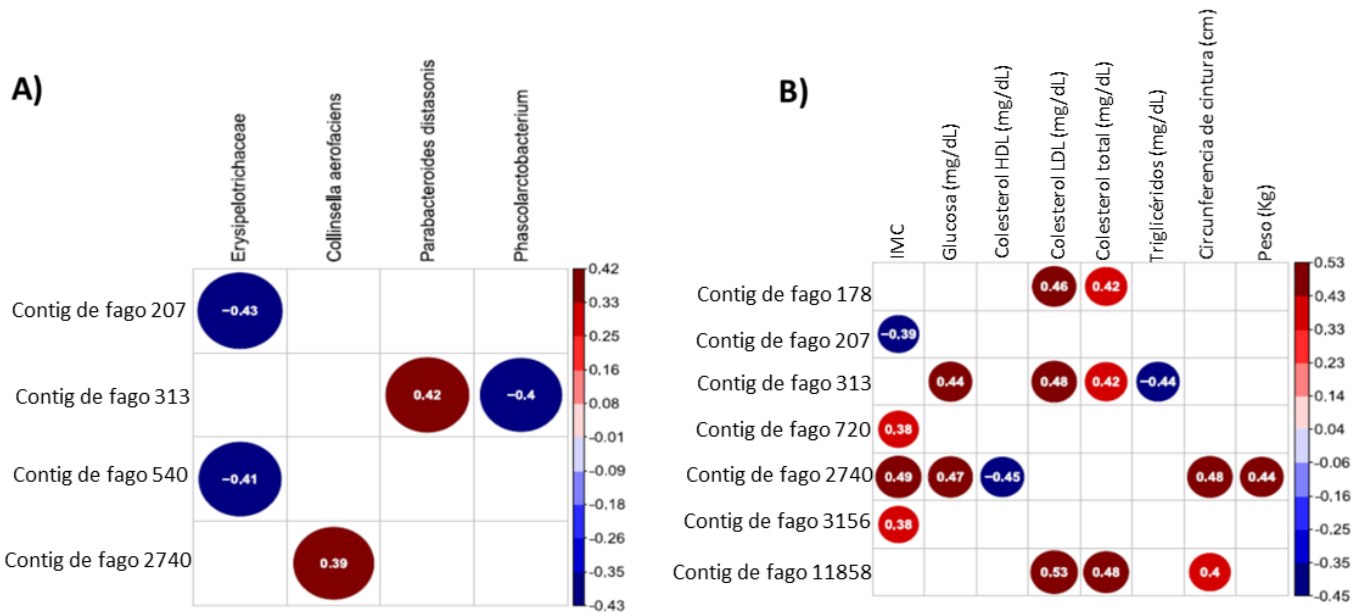
**Figura 21. Prevalencia del Fagoma core en NW.** Prevalencia de los contigs de fagos con la presencia más alta (>80% de las muestras) en NW con respecto a las muestras de O (valor  $p = <0.0001$ ) y OMS (valor  $p = <0.0001$ ). Las barras de error indican la mediana y el rango intercuartílico.

### 13.13. Las abundancias de los “fagos core” correlacionaron con taxones bacterianos asociados a obesidad y síndrome metabólico

Evaluamos si los 48 “contigs de fagos core” estuvieron asociados a cambios paralelos en la población bacteriana. Para este fin, calculamos la correlación de Spearman de los datos de secuenciación del gen 16s rRNA de los 41 taxones bacterianos y la abundancia de los 48 “contigs de fagos core” en todas las muestras. Posterior a esto, solo seleccionamos las correlaciones de Spearman con  $r^2 > 0.3$  y con un valor de  $p$  no ajustada  $\leq 0.05$ .

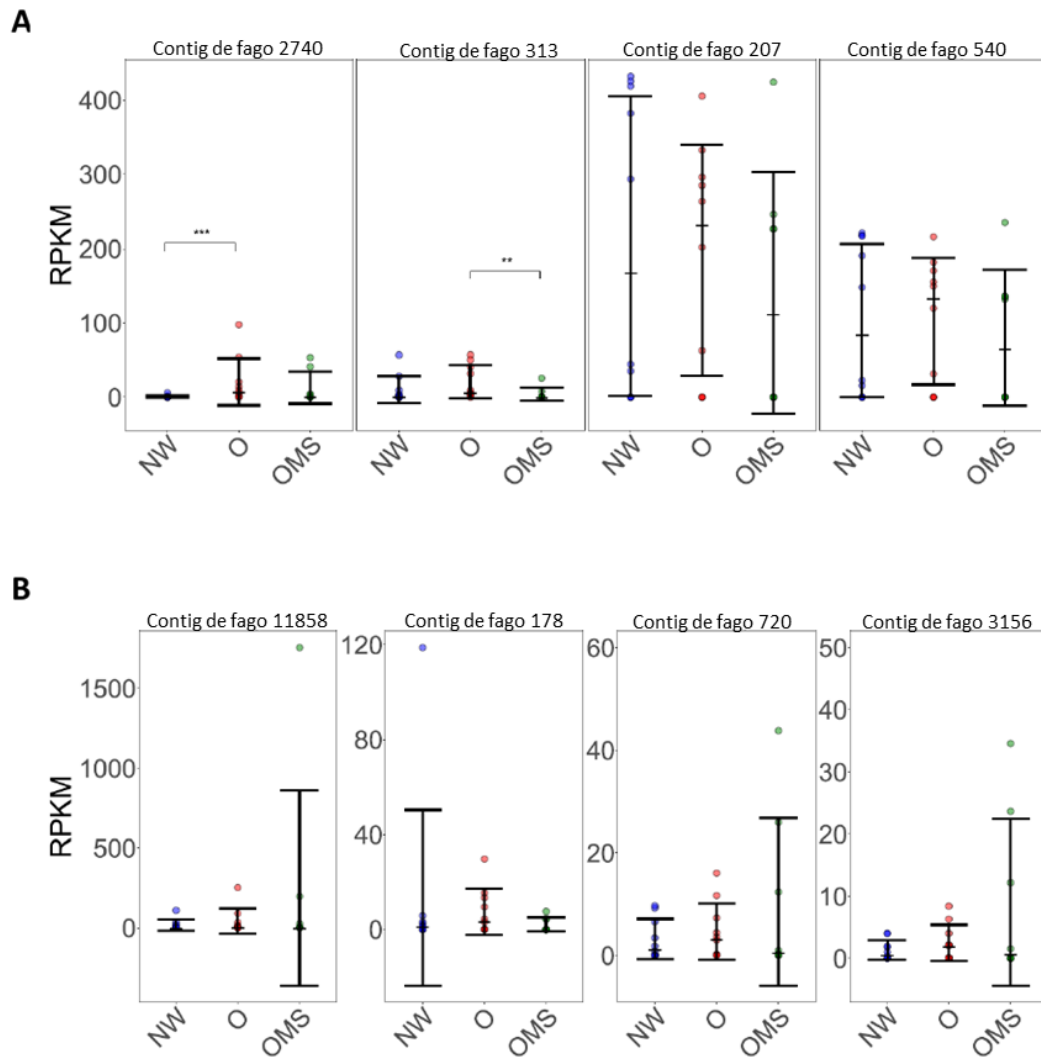
Las abundancias de cuatro contigs de fagos correlacionaron con las abundancias bacterianas (Figura 22A y Anexo 9). El contig de fago 2740, el cual fue más prevalente en O que en NW (Figura 23A), correlacionó positivamente con la abundancia de *Collinsella aerofaciens* (Figura 22A y Anexo 9), una bacteria prevalente en OMS. El contig de fago 313, más abundante en O versus OMS (Figura 23A), mostró una correlación positiva ( $p = 0.03$ ) con *Parabacteroides distasonis*, también más prevalente en O en comparación con OMS, y una correlación negativa con especies no determinadas del género *Phascolarctobacterium* (más abundante en NW versus O) (Figura 22A y Anexo 9). Además, el contig de fago 313 mostró una similitud alta (99.4% de identidad- nt) con plásmidos de *Bacteroides* previamente reportados (ID de secuencia: CP059857.1 and AP019726.1). Los contigs de fagos 207 y 540 tuvieron una prevalencia más baja en OMS y O en comparación con NW (Figura 23A). Éstos correlacionaron negativamente con *Erysipelotrichaceae*, una familia sobreabundante en los grupos de

O y OMS (Figura 22A y Anexo 9). Estos datos pueden sugerir que cambios en las abundancias de fagos específicos podrían estar parcialmente asociados a los cambios bacterianos encontrados en la obesidad y en el síndrome metabólico.



**Figura 22. Patrones de interacción del Fagoma específicos de la enfermedad en la obesidad y el síndrome metabólico.**

A) Correlaciones de Spearman de las abundancias de los 48 contigs de fagos (RPKM) con la prevalencia más alta ( $\geq 80\%$  de todas las muestras) y la abundancia relativa de los taxones bacterianos 16S que están significativamente asociados con la enfermedad. B) Correlaciones de Spearman de las abundancias de los 48 contigs de fagos (RPKM) con la prevalencia más alta ( $\geq 80\%$  de todas las muestras) y los parámetros bioquímicos y antropométricos alterados en la obesidad y en el síndrome metabólico. Solo se presentan las correlaciones significativas con un valor de p no ajustado  $\leq 0.05$ . Ver también Anexo 9.



**Figura 23.** Gráficas de cajas de las abundancias de los contigs de fagos en NW, O y OMS que **significativamente correlacionaron con** A) taxones bacterianos alterados en la obesidad y en la obesidad con síndrome metabólico y B) parámetros bioquímicos y antropométricos típicamente alterados en la obesidad y en el síndrome metabólico. Los contigs de fagos 2740, 313, y 207 mostrados en “A” también correlacionaron con los parámetros bioquímicos y antropométricos. Los puntos representan las abundancias de los contigs de fagos (RPKM) por muestra. Las barras de error indican la mediana y el rango intercuartílico por grupo.

También evaluamos si los “contigs de fagos core” estuvieron asociados con cambios paralelos en todas las poblaciones bacterianas, independientemente de la asociación bacteriana a la obesidad y al síndrome metabólico. Para este fin, seleccionamos todos los taxones derivados de la secuenciación del gen 16S rRNA y calculamos la correlación de Spearman entre ellos y las abundancias de los “contigs de fagos core” en todas las muestras. Las abundancias de 30 contigs de fagos solo correlacionaron con las abundancias de las bacterias asociadas a obesidad y síndrome metabólico (Anexo 10).

#### **13.14. Las abundancias de los “contigs de fagos core” correlacionaron con los parámetros bioquímicos y antropométricos alterados en la obesidad y síndrome metabólico**

Analizamos, también, si los 48 “contigs de fagos core” correlacionan con cambios en los parámetros bioquímicos y antropométricos típicamente alterados en la obesidad y en el síndrome metabólico, tales como IMC, colesterol HDL, triglicéridos, glucosa, circunferencia de cintura, y peso. Las correlaciones de Spearman entre los 48 contigs de fagos y los parámetros bioquímicos y antropométricos, en todas las muestras de este estudio, mostraron correlaciones positivas (Figura 22B, Figura 23 y Anexo 11), sugiriendo una asociación entre estos contigs de fagos y los parámetros bioquímicos y antropométricos. También encontramos correlaciones negativas entre la abundancia de varios contigs de fagos (Anexo 11) y IMC, colesterol HDL, y triglicéridos (Figura 22B y Anexo 11).

## 14. DISCUSIÓN

Los análisis metagenómicos de virus, uno de los componentes menos estudiados del microbioma intestinal humano, ha revolucionado recientemente nuestra visión del microbioma intestinal, destacando el papel crítico de las interacciones entre fagos y bacterias en la salud y en la enfermedad<sup>76-78</sup>.

En este estudio, reportamos un metagenoma viral secuenciado a gran escala donde en lugar de utilizar únicamente lecturas de secuenciación, realizamos un ensamble de la comunidad del Fagoma, obteniendo 4,611 contigs de fagos  $\geq 4\text{Kb}$  que representan genomas completos o parciales. Usamos este umbral de tamaño para evitar quedarnos con genomas virales parciales y para disminuir la selección de virus eucariotas, principalmente compuestos por la familia Anelloviridae, que poseen un tamaño de genoma reportado de entre 3-4 Kb<sup>48</sup>, además de considerar que los genomas de los fagos más pequeños son  $\geq 4\text{Kb}$ <sup>79</sup>. La mayoría de los estudios del viroma usan el método de amplificación del genoma completo (WGA por sus siglas en inglés: Whole Genome Amplification) MDA phi29<sup>34</sup> el cual amplifica de manera irregular fragmentos de genomas lineales y, preferencialmente, amplifica virus con ADN de cadena simple (ssADN) circular, por ejemplo, la familia de Microviridae<sup>59</sup>. Esta metodología, probablemente, limita los análisis estadísticos de la composición de las comunidades virales a observaciones de presencia-ausencia debido a que las abundancias relativas podrían estar sesgadas hacia virus específicos. Por el contrario, nosotros utilizamos un método de tagmentación (TAG) libre de la amplificación del genoma completo. Aunque este método selecciona en contra de ssADN<sup>59,80</sup>, nos ha permitido realizar análisis asociados a la abundancia de fagos con ADN de doble cadena (dsADN), los mayores componentes del viroma intestinal humano<sup>81</sup>. Probablemente, debido al método TAG, la abundancia de Microviridae en nuestras muestras fue mucho menor que la encontrada previamente en la mayor parte de los estudios del viroma<sup>11,30</sup>, destacando la necesidad de protocolos basados en los análisis cuantitativos para los estudios del viroma, como ha sido descrito recientemente<sup>59</sup>.

El método TAG impactó la proporción tanto de Microviridae (0.031%) como de Inoviridae (0.137%) en menor medida (Anexo 5) ya que ambas familias son virus ssADN. No sabemos porqué el método afectó la abundancia de Microviridae más que de Inoviridae. Sin embargo, dado que nosotros usamos los mismos protocolos para todas las muestras, nosotros esperábamos obtener el mismo sesgo para ambas familias. La selección en contra de los templados ssADN han sido observados antes en experimentos de comunidades simuladas<sup>59,82</sup>, favoreciendo la sub-representación sistemática de los virus ssADN (>10 razones de cambio) en TAG y la sobre-representación (>10 razones de cambio) en

WGA<sup>59</sup>. Para evitar este sesgo, decidimos eliminar todos los virus ssADN de nuestro análisis.

Los Caudovirales fueron la orden de fagos más abundantes en nuestros tres grupos. Este predominio es consistente con lo observado previamente en los viomas intestinales humanos<sup>30,31,48</sup>. Por primera vez, nosotros mostramos que los niños con obesidad y con obesidad con síndrome metabólico tuvieron cambios específicos en sus Fagomas intestinales, particularmente, en la abundancia de contigs de fagos específicos. Importantly, los niños reclutados para este estudio provienen de la misma región etnográfica y ambientes relativamente homogéneos, con estilos de vida similares, haciendo que los efectos socioeconómicos, culturales, y nacionales no sean factores confusos. Nosotros consideramos que esta característica poblacional nos ha permitido minimizar sesgos en nuestros resultados, tal como ha sido reportado en estudios previos<sup>83-85</sup>. Notablemente, encontramos números similares de VLPs entre los grupos, independientemente del estado de salud. Por el contrario, pacientes con la Enfermedad de Crohn albergaron de forma significativa más VLPs que en los individuos saludables<sup>30</sup>.

Observamos una diversidad y riqueza incrementadas en OMS y O en comparación con los NW, aunque este incremento no fue significativo, probablemente por el limitado tamaño de muestra. Esto sugiere que la expansión de fagos específicos en la obesidad y en el síndrome metabólico podrían disminuir la presencia de otros, manteniendo cantidades similares de VLPs independientemente del estado de salud. Interesantemente, tanto las lecturas de secuenciación como el ensamble metagenómico viral sustentaron los incrementos en la riqueza observados en O y OMS. Nosotros observamos previamente un incremento significativo en la riqueza y diversidad bacteriana en la obesidad y en el síndrome metabólico en comparación con los individuos de peso normal saludable utilizando el mismo set de muestras que fueron usadas para el viroma<sup>6</sup>. El incremento de la riqueza y diversidad de fagos y de bacterias asociadas con la obesidad y el síndrome metabólico está acorde con la propuesta reciente acerca de que la diversidad del viroma está asociada con la diversidad bacteriana intestinal en adultos<sup>57</sup>.

Aumentos de la riqueza y diversidad viral han sido también reportados en la enfermedad de Crohn<sup>30</sup>, en la colitis ulcerativa<sup>30</sup>, y en la colitis murina<sup>86</sup>. Por el contrario, disminuciones en la riqueza y diversidad viral fueron detectadas en individuos con diabetes mellitus tipo I<sup>87</sup> y en la mucosa intestinal de pacientes con colitis ulcerativa<sup>32</sup>. Por otro lado, en estudios de individuos con HIV<sup>46</sup> y Diabetes Mellitus tipo 2<sup>47</sup> no se encontraron diferencias significativas en la riqueza y diversidad viral. Los contigs de fagos encontrados con una prevalencia alta en los individuos NW estuvieron significativamente disminuidos en individuos O y OMS. Esto estaría acorde a un estudio reciente de colitis ulcerativa y enfermedad de Crohn en donde también se encontró una reducción de los contigs de

fagos típicamente abundantes en los individuos saludables <sup>31</sup>. Estos resultados sugieren que la pérdida de algunos fagos que tienen alta prevalencia en NW podría estar asociada con la obesidad y el síndrome metabólico.

Encontramos cambios sustanciales en la diversidad y riqueza asociados mayormente a la obesidad con síndrome metabólico con respecto a los individuos obesos, indicando la importancia del estudio del Fagoma en estas patologías como enfermedades separadas. Además, también encontramos fagos sobreabundantes específicos para la obesidad y diferentes a los fagos sobreabundantes en la obesidad con síndrome metabólico. Existe una amplia heterogeneidad entre las personas obesas con respecto a su riesgo de desarrollar disfunción metabólica y las complicaciones concomitantes <sup>88</sup>. La microbiota intestinal está involucrada en la etiología de la obesidad y en las complicaciones asociadas a la obesidad tales como la enfermedad del hígado graso no alcohólico, la resistencia a la insulina, y la diabetes mellitus tipo 2 <sup>89,90</sup>, sugiriendo la importancia del estudio de la obesidad de manera separada al síndrome metabólico.

Respecto a esto, estudios recientes sugieren que la microbiota de individuos obesos metabólicamente saludables pudieron transitar hacia la obesidad con síndrome metabólico <sup>91</sup>, además de que perfiles del metatranscriptoma y del 16S han demostrado diferencias significativas entre la obesidad y el síndrome metabólico <sup>6</sup>. La abundancia de algunas bacterias asociadas al metabolismo estuvo asociada a la circulación sanguínea de compuestos inflamatorios en obesos sin síndrome metabólico, sugiriendo que cambios en la microbiota intestinal en niños obesos metabólicamente saludables posiblemente sirva como una respuesta compensatoria a un exceso de nutrientes <sup>91</sup>. Los 19 contigs de fagos que estuvieron sobreabundantes solo en OMS podrían ser estudiados más a fondo como biomarcadores vinculados al desarrollo de síndrome metabólico en niños obesos. Sería interesante validar esta suposición analizando los contigs de fagos a nivel funcional <sup>92</sup> y analizando el perfil funcional de los fagos únicos de los grupos O y en OMS.

Detectamos una alta variabilidad interindividual del Fagoma, siendo 75.41% de los contigs de fagos detectados en menos del 50% de los individuos. Por el contrario, solo 48 (1.04%) de los contigs de fagos estuvieron presentes en >80% de los individuos (fagos “core”) independientemente de la enfermedad. Esta observación está acorde a evidencias previas de un “core” saludable reducido en el Fagoma intestinal humano, compuesto por solo 23 fagos <sup>31</sup>. El concepto de un viroma “core” reducido es todavía controversial. Éste ha sido sustentado por estudios recientes con gemelos monocigotos adultos, en los cuales solo 18 contigs fueron encontrados en todos los individuos <sup>57</sup>. Por el contrario, la compilación de las bases de datos del viroma intestinal a gran escala han puesto en duda la existencia

de un viroma intestinal humano “core”<sup>26,56</sup>. Nuestros hallazgos acerca de un número reducido de fagos compartidos apoyan la idea de un viroma intestinal altamente específico para cada individuo<sup>11,26,57</sup> y una diversidad viral altamente inter-individual<sup>26,70,93</sup> debido a que solo 1% de nuestro ensamble viral (4611 contigs de fagos) pertenece a contigs de fagos “core”. Este porcentaje es similar a contigs “core” reportados previamente, en donde 0.6% del ensamble viral (3639 agrupaciones de virus)<sup>26</sup> y 1.3% del ensamble viral (1703 contigs virales)<sup>31</sup>; estuvieron compartidos entre la mayoría de los individuos.

Acorde a estudios previos, se ha reportado contaminación con alguna bacteria a partir del uso de diferentes kits de extracción (incluyendo los nuestros)<sup>94,95</sup>, siendo en su mayoría bacterias que habitan en el suelo o el agua y que se relacionan frecuentemente con la fijación de nitrógeno, probablemente asociado al nitrógeno que se usa a menudo en lugar de aire en tanques de almacenamiento de agua ultrapura<sup>96</sup>. Sin embargo, no encontramos reportes acerca de contaminación viral. A diferencia de las muestras de biomasa alta (tales como las provenientes de heces fecales), la contaminación del kit es un desafío particular para los estudios de biomasa baja (tal como la proveniente de sangre o del pulmón), que puede proporcionar templados pequeños de ADN para competir con los de los reactivos para la amplificación<sup>97,98</sup>, sugiriendo que si hubo contaminación en nuestras muestras, ésta podría ser mínima y no debería influir significativamente sobre nuestros resultados.

Los criterios utilizados para definir la presencia de una secuencia viral en una muestra<sup>56</sup> todavía es cuestionable. Acorde a reportes previos<sup>83,85</sup> y a nuestros resultados, sugerimos que la representación geográfica y étnica altamente homogénea a través de nuestro set de datos de las muestras fueron factores esenciales que nos permitieron establecer un Fagoma “core” humano, siendo este altamente reducido y compuesto por tan solo 48 contigs de fagos. Nosotros aplicamos “thresholds” de secuencia similares a los recientemente propuestos para estimaciones precisas de la composición y diversidad de las comunidades virales<sup>48</sup>, tales como (i) un tamaño de contig  $\geq 4$  kb, (ii) una cobertura determinada a partir de las lecturas mapeadas con una identidad  $\geq 90\%$ , y (iii)  $\geq 80\%$  de la longitud del contig con una cobertura  $\geq 1\times$ . Sin embargo, debemos hacer notar que nuestro ensamble podría representar fragmentos del mismo genoma de fago o familia. Estudios más extensos usando parámetros bioinformáticos astringentes, tales como los que se han sugerido previamente<sup>61</sup>, son necesarios para generar estimados más precisos sobre el Fagoma “core” y su prevalencia en la población humana. Sugerimos que el uso de únicamente contigs cubiertos por  $\geq 80\%$  de su tamaño total por las lecturas virales en al menos una muestra disminuyó la tasa de lecturas alineadas ( $\sim 58\%$ ) al ensamble del viroma inicial. La eliminación de todos los contigs pequeños ( $< 4$ kb) no conllevó a una reducción drástica del reclutamiento de las lecturas, quedando  $\sim 54\%$  de las lecturas virales alineadas. Un porcentaje similar (49.66%) de



reclutamiento de lecturas fue obtenido en un estudio reciente del viroma intestinal usando un proceso extenso de descontaminación posterior al ensamblaje <sup>11</sup>. Evidencias recientes sustentan la importancia del estudio del viroma al mismo tiempo que el del bacterioma para obtener una imagen holística de los cambios en el ecosistema intestinal en una enfermedad tal como la enfermedad del intestino irritable<sup>30,99</sup>, en donde las abundancias de fagos estuvieron asociadas a cambios en las abundancias de especies bacterianas intestinales específicas <sup>100</sup>. También, la metabolómica fecal en ratones ha revelado que la depredación de fagos en la microbiota intestinal del ratón podría impactar potencialmente al mamífero huésped mediante cambios en los niveles de los principales metabolitos involucrados en funciones esenciales para el huésped <sup>101</sup>. Acorde a esto, evaluamos si los contigs de fagos altamente abundantes estuvieron asociados con un cambio paralelo en las poblaciones bacterianas y con los parámetros clínicos típicamente alterados en la obesidad y en el síndrome metabólico. Aunque los parámetros bioquímicos y antropométricos reflejaron diferencias entre grupos según la enfermedad, esta separación de grupos también correlacionó con varias abundancias de fagos. Con respecto a esto, observamos que la abundancia del contig de fago 2740 correlacionó positivamente con *Collinsella aerofaciens*, LDL, IMC, circunferencia de cintura, glucosa, peso, y bajos niveles del colesterol HDL.

Además, *Collinsella aerofaciens* estuvo significativamente sobreabundante en el grupo OMS <sup>6</sup> y mostró una correlación positiva con triglicéridos, y una correlación negativa con HDL en el mismo set de muestras. Sin embargo, *collinsella* es un taxón altamente abundante en lactantes amamantados de 12 meses de edad. Esto ha sido reportado como una firma del microbioma infantil anaeróbico en desarrollo y podría estar involucrado en la adquisición de fagos crAss-like <sup>102,103</sup>. Estos resultados sugieren que el viroma intestinal también se altera junto con el microbioma en niños obesos.

El contig de fago 313 correlacionó positivamente con *Parabacteroides distasonis*, altos niveles de colesterol LDL, glucosa, y colesterol total. Esta bacteria estuvo enriquecida en O en relación a OMS, y estuvo asociada a incrementos en los niveles del colesterol LDL en nuestra población de estudio <sup>6</sup> y ganancia de peso e hiperglicemia en otros estudios <sup>104</sup>.

El contig de fago 313 mostró una similitud alta con plásmidos de *Bacteroides* reportados (Secuencia ID: CP059857.1 y AP019726.1). Debido a que *Bacteroides* y *Parabacteroides* son bacterias estrechamente relacionadas, estos resultados sugieren que estos grupos de bacterias son hospederos putativos. Además, el contig de fago 313 correlacionó positivamente con *Parabacteroides distasonis*. Juntos, estos resultados sugieren que este fago podría estar optando por un ciclo pseudo-lisogénico y multiplicarse con sus hospederos <sup>105</sup>, dado que la estrategia de replicación de bacteriófagos similares a crAss se co-replican con su hospedero de una manera que no interrumpe la proliferación de éste <sup>105</sup>. Por

otro lado, la alta similitud del contig 313 con plásmidos de *Bacteroides* previamente reportados sugiere a este grupo de bacteria como hospedero putativo. Desafortunadamente, es poco lo que se sabe acerca de la diversidad de plásmidos y bacteriófagos pseudo-lisogénicos en estos grupos de bacterias. La realización de más estudios es necesaria para entender mejor la replicación viral que podría ayudar a interpretar estas relaciones. Por el contrario, el contig de fago 313 mostró una correlación negativa con el género *Phascolarctobacterium*, enriquecido en los NW en relación a O<sup>6</sup>. Esto sugiere que la abundancia incrementada de este contig de fago en la obesidad podría estar inhibiendo la abundancia de *Phascolarctobacterium*, una bacteria potencialmente protectora en contra de la obesidad<sup>6</sup>. También, la abundancia incrementada de los contigs de fagos 207 y 540 en NW correlacionaron con una disminución en la abundancia de *Erysipelotrichaceae*, una familia bacteriana significativamente aumentada en OMS y que correlacionó positivamente con la circunferencia de cintura<sup>6</sup>. Una alta abundancia de *Erysipelotrichaceae* ha sido asociada con dislipidemia en la obesidad, síndrome metabólico, e hipercolesterolemia<sup>106</sup>. Esto sugiere que los contigs de fagos 207 y 540 podrían ser usados para disminuir la abundancia de *Erysipelotrichaceae* en los grupos O y OMS<sup>6</sup>.

Estos ejemplos abren la posibilidad de la utilización de fagos como opciones terapéuticas en contra de los cambios bacterianos típicamente asociados con obesidad. Incluso, se ha sugerido que el Fagoma intestinal representa una fuente de fagos individuales con aplicaciones terapéuticas potenciales<sup>62</sup>. Además, tratamientos exitosos contra *Clostridium difficile* usando filtrados fecales libres de bacterias proporcionaron la primera evidencia de que la manipulación del Fagoma podría ser una estrategia terapéutica efectiva para estabilizar la eubiosis bacteriana en el microbioma<sup>62</sup>. Evidencia contundente sustenta la idea de que cambios en el sistema microbiano durante la infancia puede incrementar el riesgo de obesidad más adelante en la vida<sup>107</sup>. Por lo tanto, la manipulación de la microbiota intestinal usando fagos en etapas tempranas de la vida podría ofrecer la prevención y tratamiento de los cambios bacterianos asociados a obesidad. El trasplante de microbiota ha revelado que los fagos pueden ser co-transferidos junto con las bacterias<sup>108</sup>.

Nosotros creemos que nuestro estudio provee un mejor entendimiento acerca de las interacciones entre fago-bacteria en el microbioma intestinal. El desarrollo de modelos *in vivo* para evaluar las dinámicas fago-bacteria en la obesidad será, sin duda, un área imprescindible que podría ayudar a complementar el entendimiento del rol de los fagos en los cambios de la microbiota asociados a la obesidad y al síndrome metabólico.

## 15. CONCLUSIONES

Uno de los papeles más interesantes del Fagoma intestinal, es su capacidad para interactuar con su hospedero bacteriano e influir sobre el estado de salud del humano, desempeñando un papel clave en la conformación de la composición y función del microbioma intestinal con el que convive. Tanto las lecturas de secuenciación como el ensamble metagenómico viral sustentaron incrementos en la riqueza y diversidad de fagos observados en O y OMS, mientras que cantidades similares de VLPs fueron observadas en los tres grupos, sugiriendo que la expansión de fagos específicos en la obesidad y en el síndrome metabólico podrían disminuir la abundancia de otros manteniendo el mismo número de partículas virales. Los Fagomas de todas las muestras estuvieron conformados, principalmente, por Caudovirales. Sin embargo, encontramos una diversidad viral altamente inter-individual y específica para cada individuo debido a que solo el 1% de nuestro ensamble viral fueron fagos compartidos entre la mayor parte de la población (fagos “core”). A pesar de la alta inter-individualidad del Fagoma, encontramos un grupo de contigs de fagos con una alta prevalencia en los individuos NW la cual disminuyó significativamente en individuos O y OMS, sugiriendo que la pérdida de algunos fagos que tienen alta prevalencia en NW podría estar asociada con la obesidad y el síndrome metabólico. Por otro lado, cambios sustanciales en la diversidad y riqueza estuvieron mayormente asociados a la obesidad con síndrome metabólico en comparación a los individuos obesos, indicando la importancia del estudio del Fagoma en estas patologías como enfermedades separadas. Además, encontramos fagos sobreabundantes específicos para la obesidad y diferentes a los fagos sobreabundantes en la obesidad con síndrome metabólico, sugiriendo, nuevamente, la importancia del estudio de la obesidad de manera separada al síndrome metabólico. El análisis de los contigs de fagos altamente abundantes entre todos los individuos (fagos “core”) mostró que cambios en sus abundancias estuvieron asociados con cambios paralelos en las abundancias de bacterias intestinales específicas y con los parámetros clínicos típicamente alterados en la obesidad y en el síndrome metabólico. Estas asociaciones reflejaron diferencias entre los grupos según la enfermedad, y sugieren que el viroma intestinal también se altera junto con el microbioma en niños obesos. Esto abre la posibilidad de la utilización de fagos como opciones terapéuticas en contra de los cambios bacterianos típicamente asociados con obesidad y síndrome metabólico.

## 16. PERSPECTIVAS

Conocer hasta qué punto el fago del intestino humano le da forma al microbioma determinará si será posible utilizar el fago como herramienta terapéutica en el futuro. Sin embargo, todavía existen lagunas significativas en nuestra comprensión de las interacciones fago-huésped que deben abordarse antes de poder llegar a conclusiones sobre la utilidad de los fagos como bioterapéuticos. En este contexto, profundizar en el entendimiento de la interacción fago-bacteria mediante modelos *in vivo* que permitan interpretar los mecanismos de asociación con la obesidad y el síndrome metabólico es de suma importancia. La evidencia preliminar de que el viroma podría desempeñar un papel en el éxito de la transferencia de la microbiota fecal (FMT) sugiere que puede haber un futuro en el uso del viroma para dar forma al microbioma tanto en la obesidad como en otras enfermedades. Ahondar en el estudio de los fagos que encontramos sobreabundantes y específicos en niños OMS abre la puerta para la posibilidad de desarrollar, no solo estrategias terapéuticas, sino también de prevención mediante la identificación de posibles biomarcadores relacionados con el desarrollo de síndrome metabólico en niños obesos. Finalmente, la realización de estudios que nos permitan obtener un entendimiento integral acerca de la composición del Fagoma intestinal y su interacción con la microbiota, el sistema inmune del hospedero y su entorno ambiental nos permitirá crear más herramientas para disminuir la prevalencia de obesidad infantil y del síndrome metabólico en México y en el mundo.

## 17. BIBLIOGRAFÍA

1. WHO | Childhood overweight and obesity. *WHO* (2017).
2. Romero-Martínez, M. *et al.* National Health and Nutrition Survey 2018-19: Methodology and perspectives. *Salud Publica Mex.* **61**, 917–923 (2019).
3. Mattos, R. T. *et al.* Chronic low-grade inflammation in childhood obesity is associated with decreased il-10 expression by monocyte subsets. *PLoS One* **11**, (2016).
4. Evia-Viscarra, M. L., Rodea-Montero, E. R., Apolinar-Jiménez, E. & Quintana-Vargas, S. Metabolic syndrome and its components among obese (BMI  $\geq$ 95th) Mexican adolescents. *Endocr. Connect.* **2**, 208–215 (2013).
5. De Ferranti, S. D. *et al.* Prevalence of the metabolic syndrome in American adolescents: Findings from the Third National Health and Nutrition Examination Survey. *Circulation* **110**, 2494–2497 (2004).
6. Gallardo-Becerra, L. *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.* **19**, (2020).
7. DiBonaventura, M. D. *et al.* Obesity in Mexico: prevalence, comorbidities, associations with patient outcomes, and treatment experiences. *Diabetes. Metab. Syndr. Obes.* **11**, 1–10 (2018).
8. Franks, P. W. *et al.* Childhood obesity, other cardiovascular risk factors, and premature death. *N. Engl. J. Med.* **362**, 485–493 (2010).
9. Biro, F. M. & Wien, M. Childhood obesity and adult morbidities. *American Journal of Clinical Nutrition* vol. 91 1499S (2010).
10. Shkoporov, A. N. & Hill, C. Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. *Cell Host and Microbe* vol. 25 195–209 (2019).
11. Shkoporov, A. N. *et al.* Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* **6**, 68 (2018).
12. L, H. *et al.* Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res. Microbiol.* **165**, 803–812 (2014).
13. J, W. *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* **25**, 679–689 (2019).
14. M, A. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
15. Huseyin, C. E., Rubio, R. C., O’Sullivan, O., Cotter, P. D. & Scanlan, P. D. The Fungal Frontier: A Comparative Analysis of Methods Used in the Study of the Human Gut Mycobiome. *Front.*

- Microbiol.* **8**, (2017).
16. L, W. *et al.* Colonization and Succession within the Human Gut Microbiome by Archaea, Bacteria, and Microeukaryotes during the First Year of Life. *Front. Microbiol.* **8**, (2017).
  17. R, S., S, F. & R, M. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* **14**, (2016).
  18. Bai, J., Hu, Y. & Bruner, D. W. Composition of gut microbiota and its association with body mass index and lifestyle factors in a cohort of 7-18 years old children from the American Gut Project. *Pediatr. Obes.* **14**, e12480 (2019).
  19. Chen, X. *et al.* Alteration of the gut microbiota associated with childhood obesity by 16S rRNA gene sequencing. *PeerJ* **2020**, (2020).
  20. Kim, M. H. *et al.* Gut microbiota and metabolic health among overweight and obese individuals. *Sci. Rep.* **10**, 1–11 (2020).
  21. Pihl, A. F. *et al.* The role of the gut microbiota in childhood obesity. *Childhood Obesity* vol. 12 292–299 (2016).
  22. Bouter, K. E., van Raalte, D. H., Groen, A. K. & Nieuwdorp, M. Role of the Gut Microbiome in the Pathogenesis of Obesity and Obesity-Related Metabolic Dysfunction. *Gastroenterology* **152**, 1671–1678 (2017).
  23. Minot, S. *et al.* The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* **21**, 1616–1625 (2011).
  24. Kim, M.-S. & Bae, J.-W. Spatial disturbances in altered mucosal and luminal gut viromes of diet-induced obese mice. *Environ. Microbiol.* **18**, 1498–1510 (2016).
  25. García-López, R., Pérez-Brocal, V. & Moya, A. Beyond cells – the virome in the human holobiont. *Microbial Cell* vol. 6 373–396 (2019).
  26. Shkoporov, A. N. *et al.* The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell Host Microbe* **26**, 527-541.e5 (2019).
  27. Virgin, H. W. The virome in mammalian physiology and disease. *Cell* vol. 157 142–150 (2014).
  28. Barr, J. J. *et al.* Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 10771–10776 (2013).
  29. Maiques, E. *et al.*  $\beta$ -lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *J. Bacteriol.* **188**, 2726–2729 (2006).
  30. Norman, J. M. *et al.* Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* (2015) doi:10.1016/j.cell.2015.01.002.

31. Manrique, P. *et al.* Healthy human gut phageome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10400–10405 (2016).
32. Zuo, T. *et al.* Gut mucosal virome alterations in ulcerative colitis. *Gut* **68**, 1169–1179 (2019).
33. Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* vol. 28 127–181 (2004).
34. Sutton, T. D. S. & Hill, C. Gut Bacteriophage: Current Understanding and Challenges. *Frontiers in Endocrinology* vol. 10 784 (2019).
35. NV, R. Replication and Maintenance of Linear Phage-Plasmid N15. *Microbiol. Spectr.* **3**, (2015).
36. W, C., A, M., SK, G., R, L. & A, A. Viral Transmission Dynamics at Single-Cell Resolution Reveal Transiently Immune Subpopulations Caused by a Carrier State Association. *PLoS Genet.* **11**, (2015).
37. K, U., M, K. & R, F. Transduction of drug resistance to tetracycline, chloramphenicol, macrolides, lincomycin and clindamycin with phages induced from *Streptococcus pyogenes*. *J. Antibiot. (Tokyo)*. **28**, 681–688 (1975).
38. Fard, R. M. N., Barton, M. D. & Heuzenroeder, M. W. Bacteriophage-mediated transduction of antibiotic resistance in enterococci. *Lett. Appl. Microbiol.* **52**, 559–564 (2011).
39. M, V. *et al.* Efficient transfer of antibiotic resistance plasmids by transduction within methicillin-resistant *Staphylococcus aureus* USA300 clone. *FEMS Microbiol. Lett.* **332**, 146–152 (2012).
40. Rolain, J. M., Fancello, L., Desnues, C. & Raoult, D. Bacteriophages as vehicles of the resistome in cystic fibrosis. *J. Antimicrob. Chemother.* **66**, 2444–2447 (2011).
41. SR, M., HH, L., CS, S. & JJ, C. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**, 219–222 (2013).
42. SR, A., M, L., TM, S.-R. & DT, P. Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes. *PLoS One* **10**, (2015).
43. Enault, F. *et al.* Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J. 2017 111* **11**, 237–247 (2016).
44. Guo, J. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, (2021).
45. V, A., G, L. & FD, B. Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mob. DNA* **8**, (2017).

46. Monaco, C. L. *et al.* Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host Microbe* **19**, 311–322 (2016).
47. Ma, Y., You, X., Mai, G., Tokuyasu, T. & Liu, C. A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* **6**, 24 (2018).
48. Reyes, A. *et al.* Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11941–11946 (2015).
49. Rasmussen, T. S. *et al.* Faecal virome transplantation decreases symptoms of type 2 diabetes and obesity in a murine model. *Gut* **69**, 2122–2130 (2020).
50. Schulfer, A. *et al.* Fecal Viral Community Responses to High-Fat Diet in Mice. *mSphere* **5**, (2020).
51. Twort, F. W. AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES. *Lancet* **186**, 1241–1243 (1915).
52. BL, H. & MB, S. The Pacific Ocean virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS One* **8**, (2013).
53. BL, H., JR, B. & MB, S. Depth-stratified functional and taxonomic niche specialization in the ‘core’ and ‘flexible’ Pacific Ocean Virome. *ISME J.* **9**, 472–484 (2015).
54. GD, H., MB, D., MT, R., CC, K. & PD, S. Diagnostic Potential and Interactive Dynamics of the Colorectal Cancer Virome. *MBio* **9**, (2018).
55. DM, D. *et al.* Intestinal microbiota, microbial translocation, and systemic inflammation in chronic HIV infection. *J. Infect. Dis.* **211**, 19–27 (2015).
56. Gregory, A. C., Zablocki, O., Howell, A., Bolduc, B. & Sullivan, M. B. The human gut virome database. *bioRxiv* 655910 (2019) doi:10.1101/655910.
57. Moreno-Gallego, J. L. *et al.* Virome Diversity Correlates with Intestinal Microbiome Diversity in Adult Monozygotic Twins. *Cell Host Microbe* **25**, 261-272.e5 (2019).
58. Clooney, A. G. *et al.* Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease. *Cell Host Microbe* **26**, 764-778.e5 (2019).
59. Roux, S. *et al.* Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* **2016**, (2016).
60. TDS, S., AG, C., FJ, R., RP, R. & C, H. Choice of assembly software has a critical impact on virome characterisation. *Microbiome* **7**, (2019).
61. Roux, S., Sullivan, M. B., Emerson, J. B. & Eloë-Fadrosh, E. A. Benchmarking viromics: an in



- silico evaluation of metagenome-enabled estimates of viral community composition and diversity. (2017) doi:10.7717/peerj.3817.
62. Ott, S. J. *et al.* Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients With *Clostridium difficile* Infection. *Gastroenterology* **152**, 799-811.e7 (2017).
  63. A, W., CH, H., EE, A. & DR, H. A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; a preliminary report of efficacy. *Clin. Otolaryngol.* **34**, 349–357 (2009).
  64. BK, C. *et al.* Phage treatment of an aortic graft infected with *Pseudomonas aeruginosa*. *Evol. Med. public Heal.* **2018**, 60–66 (2018).
  65. S, L. *et al.* Phage Therapy for a Multidrug-Resistant *Acinetobacter baumannii* Craniectomy Site Infection. *Open forum Infect. Dis.* **5**, (2018).
  66. Garrett, L. Seven circles of antimicrobial hell. *Lancet* **393**, 865–867 (2019).
  67. Norman, J. M. *et al.* Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **160**, 447–460 (2015).
  68. Guo, J. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).
  69. Breitbart, M. *et al.* Viral diversity and dynamics in an infant gut. *Res. Microbiol.* **159**, 367–373 (2008).
  70. Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
  71. Coutinho, F. H., Edwards, R. A. & Rodríguez-Valera, F. Charting the diversity of uncultured viruses of Archaea and Bacteria. *BMC Biol.* **17**, 1–16 (2019).
  72. Deboutte, W. *et al.* Honey-bee-associated prokaryotic viral communities reveal wide viral diversity and a profound metabolic coding potential. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 10511–10519 (2020).
  73. Dutilh, B. E. *et al.* A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **5**, 1–11 (2014).
  74. Guerin, E. *et al.* Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut. *Cell Host Microbe* **24**, 653-664.e6 (2018).
  75. Cervantes-Echeverría, M. *et al.* Whole-genome of Mexican-crAssphage isolated from the human gut microbiome. *BMC Res. Notes* **11**, 902 (2018).
  76. Breitbart, M. *et al.* Metagenomic analyses of an uncultured viral community from human feces.

- J. Bacteriol.* **185**, 6220–6223 (2003).
77. Bikel, S. *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. 13 390–401 (2015).
  78. Garmaeva, S. *et al.* Studying the gut virome in the metagenomic era: Challenges and perspectives. *BMC Biology* vol. 17 1–14 (2019).
  79. Hatfull, G. F. Bacteriophage genomics. *Current Opinion in Microbiology* vol. 11 447–453 (2008).
  80. Yilmaz, S., Allgaier, M. & Hugenholtz, P. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nature Methods* vol. 7 943–944 (2010).
  81. Kim, M. S., Park, E. J., Roh, S. W. & Bae, J. W. Diversity and abundance of single-stranded DNA viruses in human feces. *Appl. Environ. Microbiol.* **77**, 8062–8070 (2011).
  82. Kim, K. H. & Bae, J. W. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl. Environ. Microbiol.* **77**, 7663–7668 (2011).
  83. He, Y. *et al.* Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat. Med.* **24**, 1532–1535 (2018).
  84. Stagaman, K. *et al.* Market Integration Predicts Human Gut Microbiome Attributes across a Gradient of Economic Development. *mSystems* **3**, (2018).
  85. Zhong, H. *et al.* Impact of early events and lifestyle on the gut microbiota and metabolic phenotypes in young school-age children. *Microbiome* **7**, (2019).
  86. Duerkop, B. A. *et al.* Murine colitis reveals a disease-associated bacteriophage community. *Nat. Microbiol.* **3**, 1023–1031 (2018).
  87. Zhao, G. *et al.* Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E6166–E6175 (2017).
  88. Viveiros, A. & Oudit, G. Y. The dual nature of obesity in metabolic programming: Quantity versus quality of adipose tissue. *Clinical Science* vol. 134 2447–2451 (2020).
  89. Pedersen, H. K. *et al.* Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **535**, 376–381 (2016).
  90. Canfora, E. E., Meex, R. C. R., Venema, K. & Blaak, E. E. Gut microbial metabolites in obesity, NAFLD and T2DM. *Nature Reviews Endocrinology* vol. 15 261–273 (2019).
  91. Yuan, X. *et al.* The role of the gut microbiota on the metabolic status of obese children. *Microb.*

- Cell Fact.* **20**, 53–53 (2021).
92. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
  93. Minot, S. *et al.* Rapid evolution of the human gut virome. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12450–12455 (2013).
  94. Laurence, M., Hatzis, C. & Brash, D. E. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS One* **9**, (2014).
  95. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
  96. Kulakov, L. A., McAlister, M. B., Ogden, K. L., Larkin, M. J. & O’Hanlon, J. F. Analysis of bacteria contaminating ultrapure water in industrial systems. *Appl. Environ. Microbiol.* **68**, 1548–1555 (2002).
  97. Tanner, M. A., Goebel, B. M., Dojka, M. A. & Pace, N. R. Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl. Environ. Microbiol.* **64**, 3110–3113 (1998).
  98. Willerslev, E., Hansen, A. J. & Poinar, H. N. Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology and Evolution* vol. 19 141–147 (2004).
  99. Cornuault, J. K. *et al.* Phages infecting *Faecalibacterium prausnitzii* belong to novel viral genera that help to decipher intestinal viromes. *Microbiome* **6**, 65 (2018).
  100. Reyes, A., Wu, M., McNulty, N. P., Rohwer, F. L. & Gordon, J. I. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20236–20241 (2013).
  101. Hsu, B. B. *et al.* Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. *Cell Host Microbe* **25**, 803–814 (2019).
  102. Bäckhed, F. *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**, 690–703 (2015).
  103. Siranosian, B. A., Tamburini, F. B., Sherlock, G. & Bhatt, A. S. Acquisition, transmission and strain diversity of human gut-colonizing crAss-like phages. *Nat. Commun.* **11**, 1–11 (2020).
  104. Wang, K. *et al.* *Parabacteroides distasonis* Alleviates Obesity and Metabolic Dysfunctions via Production of Succinate and Secondary Bile Acids. *Cell Rep.* **26**, 222-235.e5 (2019).
  105. Shkoporov, A. N. *et al.*  $\Phi$ CrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat. Commun.* **9**, 1–8 (2018).

106. Spencer, M. D. *et al.* Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. *Gastroenterology* **140**, 976–986 (2011).
107. Scheepers, L. E. J. M. *et al.* The intestinal microbiota composition and weight development in children: The KOALA Birth Cohort Study. *Int. J. Obes.* **39**, 16–25 (2015).
108. Draper, L. A. *et al.* Long-term colonisation with donor bacteriophages following successful faecal microbial transplantation. *Microbiome* **6**, 220 (2018).

# 18. ANEXO

**Anexo 1. Tabla con la lista de los estudios del viroma humano publicados hasta la actualidad.** El orden de aparición inicia a partir del estudio más reciente.

PMID	Título	Primer autor	Revista/libro	Año de publicación	PMCID	DOI
34290346	Minimal-moderate variation of human oral virome and microbiome in IgA deficiency	de la Cruz Peña MJ	Sci Rep	2021	PMC8295364	10.1038/s41598-021-94507-8
34238738	[Characteristics of gut virome and microbiome in patients with stroke]	Wang Z	Nan Fang Yi Ke Da Xue Xue Bao	2021	PMC8267978	10.12122/j.issn.1673-4254.2021.06.08
34150671	The Human Gut Phageome: Origins and Roles in the Human Gut Microbiome	Townsend EM	Front Cell Infect Microbiol	2021	PMC8213399	10.3389/fcimb.2021.643214
34071061	Human Milk Virome Analysis: Changing Pattern Regarding Mode of Delivery, Birth Weight, and Lactational Stage	Dinleyici M	Nutrients	2021	PMC8224552	10.3390/nu13061779
34052095	Cross-correlation of virome-bacteriome-host-metabolome to study respiratory health	Iorio A	Trends Microbiol	2021		10.1016/j.tim.2021.04.011
34010651	Stability of the human gut virome and effect of gluten-free diet	Garmaeva S	Cell Rep	2021		10.1016/j.celrep.2021.109132
33975974	Spinal Cord Injury Changes the Structure and Functional Potential of Gut Bacterial and Viral Communities	Du J	mSystems	2021	PMC8125080	10.1128/mSystems.01356-20
33970533	Dynamic changes occur in the DNA gut virome of female cynomolgus macaques during aging	Tan X	Microbiologyopen	2021	PMC8087919	10.1002/mbo3.1186
33959381	Characterization of the gut DNA and RNA Viromes in a Cohort of Chinese Residents and Visiting Pakistanis	Yan Q	Virus Evol	2021	PMC8087960	10.1093/ve/veab022
33923593	Emerging Roles of Gut Virome in Pediatric Diseases	Fulci V	Int J Mol Sci	2021	PMC8073368	10.3390/ijms22084127
33853691	Temporal landscape of human gut RNA and DNA virome in SARS-CoV-2 infection and severity	Zuo T	Microbiome	2021	PMC8044506	10.1186/s40168-021-01008-x
33852569	The gut virome of healthy children during the first year of life is diverse and dynamic	Taboada B	PLoS One	2021	PMC8046192	10.1371/journal.pone.0240958
33845877	Isolation and characterisation of ΦcrAss002, a crAss-like phage from the human gut that infects Bacteroides xylanisolvens	Guerin E	Microbiome	2021	PMC8042965	10.1186/s40168-021-01036-7

33815397	Impact of Early Life Nutrition on Children's Immune System and Noncommunicable Diseases Through Its Effects on the Bacterial Microbiome, Virome and Mycobiome	Fragkou PC	Front Immunol	2021	PMC8012492	10.3389/fimmu.2021.644269
33813257	Assembly of the virome in newborn human infants	Bushman F	Curr Opin Virol	2021	PMC8187319	10.1016/j.coviro.2021.03.004
33806607	Biases in Viral Metagenomics-Based Detection, Cataloguing and Quantification of Bacteriophage Genomes in Human Faeces, a Review	Callanan J	Microorganisms	2021	PMC8000950	10.3390/microorganisms9030524
33791236	Probing the "Dark Matter" of the Human Gut Phageome: Culture Assisted Metagenomics Enables Rapid Discovery and Host-Linking for Novel Bacteriophages	Fitzgerald CB	Front Cell Infect Microbiol	2021	PMC8005731	10.3389/fcimb.2021.616918
33781338	Thousands of previously unknown phages discovered in whole-community human gut metagenomes	Benler S	Microbiome	2021	PMC8008677	10.1186/s40168-021-01017-w
33727396	Breast Milk Virome and Bacterial Microbiome Resilience in Kenyan Women Living with HIV	Maqsood R	mSystems	2021		10.1128/mSystems.01079-20
33678150	Integrated gut virome and bacteriome dynamics in COVID-19 patients	Cao J	Gut Microbes	2021	PMC7946006	10.1080/19490976.2021.1887722
33606979	Massive expansion of human gut bacteriophage diversity	Camarillo-Guerrero LF	Cell	2021	PMC7895897	10.1016/j.cell.2021.01.029
33594055	Analysis of metagenome-assembled viral genomes from the human gut reveals diverse putative CrAss-like phages with unique genomic features	Yutin N	Nat Commun	2021	PMC7886860	10.1038/s41467-021-21350-w
33552999	Enteric Phageome Alterations in Patients With Type 2 Diabetes	Chen Q	Front Cell Infect Microbiol	2021	PMC7862107	10.3389/fcimb.2020.575084
33397897	Longitudinal dynamics of gut bacteriome, mycobiome and virome after fecal microbiota transplantation in graft-versus-host disease	Zhang F	Nat Commun	2021	PMC7782528	10.1038/s41467-020-20240-x
33347881	Effects of Intestinal Fungi and Viruses on Immune Responses and	Iliev ID	Gastroenterology	2021	PMC7956156	10.1053/j.gastro.2020.06.100

	Inflammatory Bowel Diseases					
33176253	Phages to shape the gut microbiota?	Dahlman S	Curr Opin Biotechnol	2021		10.1016/j.copbio.2020.09.016
33451106	Enteric Viruses and Inflammatory Bowel Disease	Tarris G	Viruses	2021	PMC7828589	10.3390/v13010104
33363706	Gut eukaryotic virome in colorectal carcinogenesis: Is that a trigger?	Massimino L	Comput Struct Biotechnol J	2020	PMC7750180	10.1016/j.csbj.2020.11.055
33194818	Virome Sequencing of the Human Intestinal Mucosal-Luminal Interface	Yan A	Front Cell Infect Microbiol	2020	PMC7642909	10.3389/fcimb.2020.582187
33181076	Putting the Virome on the Map: The Influence of Host Geography and Ethnicity on the Gut Virome	Holtz LR	Cell Host Microbe	2020		10.1016/j.chom.2020.10.007
33167516	Metagenomic Analysis of the Enteric RNA Virome of Infants from the Oukasie Clinic, North West Province, South Africa, Reveals Diverse Eukaryotic Viruses	Mogotsi MT	Viruses	2020	PMC7694487	10.3390/v12111260
33014897	Shining Light on Human Gut Bacteriophages	Guerin E	Front Cell Infect Microbiol	2020	PMC7511551	10.3389/fcimb.2020.00481
32958861	Metagenomic sequencing of stool samples in Bangladeshi infants: virome association with poliovirus shedding after oral poliovirus vaccination	Tan SK	Sci Rep	2020	PMC7506025	10.1038/s41598-020-71791-4
32910902	Human-Gut-DNA Virome Variations across Geography, Ethnicity, and Urbanization	Zuo T	Cell Host Microbe	2020		10.1016/j.chom.2020.08.005
32841606	The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut	Gregory AC	Cell Host Microbe	2020	PMC7443397	10.1016/j.chom.2020.08.003
32751611	The Vaginal Virome-Balancing Female Genital Tract Bacteriome, Mucosal Immunity, and Sexual and Reproductive Health Outcomes?	Happel AU	Viruses	2020	PMC7472209	10.3390/v12080832
32654263	Intestinal Virome in Patients With Alcoholic Hepatitis	Jiang L	Hepatology	2020	PMC8159727	10.1002/hep.31459
32652145	Intestinal Virome Signature Associated With Severity of Nonalcoholic Fatty Liver Disease	Lang S	Gastroenterology	2020		10.1053/j.gastro.2020.07.005
32495834	Bacteriophage-mediated manipulation of the gut microbiome - promises and presents limitations	Rasmussen TS	FEMS Microbiol Rev	2020		10.1093/femsre/fuaa020



32461640	The stepwise assembly of the neonatal virome is modulated by breastfeeding	Liang G	Nature	2020	PMC7263352	10.1038/s41586-020-2192-1
32429898	Viral metagenomic analysis of fecal samples reveals an enteric virome signature in irritable bowel syndrome	Ansari MH	BMC Microbiol	2020	PMC7236503	10.1186/s12866-020-01817-4
32406906	Dynamics of the Stool Virome in Very Early-Onset Inflammatory Bowel Disease	Liang G	J Crohns Colitis	2020	PMC7648169	10.1093/ecco-jcc/jjaa094
32325706	Phages and Their Role in Gastrointestinal Disease: Focus on Inflammatory Bowel Disease	Maronek M	Cells	2020	PMC7226564	10.3390/cells9041013
32298613	The crAss-like Phage Group: How Metagenomics Reshaped the Human Virome	Koonin EV	Trends Microbiol	2020		10.1016/j.tim.2020.01.010
32290414	Gut Microbiota beyond Bacteria- Mycobiome, Virome, Archaeome, and Eukaryotic Parasites in IBD	Matijašić M	Int J Mol Sci	2020	PMC7215374	10.3390/ijms21082668
32200373	Beyond the Bacterial Microbiome: Virome of Human Milk and Effects on the Developing Infant	Mohandas S	Nestle Nutr Inst Workshop Ser	2020		10.1159/000504997
32165408	Faecal virome transplantation decreases symptoms of type 2 diabetes and obesity in a murine model	Rasmussen TS	Gut	2020		10.1136/gutjnl-2019-320005
32060814	Enteric Virome and Carcinogenesis in the Gut	Emler C	Dig Dis Sci	2020		10.1007/s10620-020-06126-4
31800147	Higher frequency of vertebrate-infecting viruses in the gut of infants born to mothers with type 1 diabetes	Kim KW	Pediatr Diabetes	2020		10.1111/pedi.12952
31744911	Metagenomics of the faecal virome indicate a cumulative effect of enterovirus and gluten amount on the risk of coeliac disease autoimmunity in genetically at risk children: the TEDDY study	Lindfors K	Gut	2020	PMC7234892	10.1136/gutjnl-2019-319809
31300514	Authors response: giant oversights in the human gut virome	Zuo T	Gut	2020		10.1136/gutjnl-2019-319357
31186296	Giant oversights in the human gut virome	Sutton TDS	Gut	2020		10.1136/gutjnl-2019-319067
32574158	Growth velocity in children with Environmental Enteric Dysfunction	Desai C	PLoS Negl Trop Dis	2020	PMC7310680	10.1371/journal.pntd.0008387

31849833	is associated with specific bacterial and viral taxa of the gastrointestinal tract in Malawian children Gut Bacteriophage: Current Understanding and Challenges	Sutton TDS	Front Endocrinol (Lausanne)	2019	PMC6895007	10.3389/fendo.2019.00784
31823811	Discordant transmission of bacteria and viruses from mothers to babies at birth	Maqsood R	Microbiome	2019	PMC6902606	10.1186/s40168-019-0766-7
31819139	First identification of mammalian orthoreovirus type 3 by gut virome analysis in diarrheic child in Brazil	Rosa UA	Sci Rep	2019	PMC6901473	10.1038/s41598-019-55216-5
31766550	Evolution of BACON Domain Tandem Repeats in crAssphage and Novel Gut Bacteriophage Lineages	Jonge PA	Viruses	2019	PMC6949934	10.3390/v11121085
31757768	Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease	Clooney AG	Cell Host Microbe	2019		10.1016/j.chom.2019.10.009
31662858	The gut virome in inflammatory bowel disease pathogenesis: From metagenomics to novel therapeutic approaches	Ungaro F	United European Gastroenterol J	2019	PMC6794698	10.1177/2050640619876787
31660953	Studying the gut virome in the metagenomic era: challenges and perspectives	Garmaeva S	BMC Biol	2019	PMC6819614	10.1186/s12915-019-0704-y
31600503	The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific	Shkoporov AN	Cell Host Microbe	2019		10.1016/j.chom.2019.09.009
31600497	Not Just a Passing Phage	Sutcliffe SG	Cell Host Microbe	2019		10.1016/j.chom.2019.09.011
31492655	CRISPR-Cas System of a Prevalent Human Gut Bacterium Reveals Hyper-targeting against Phages in a Human Virome Catalog	Soto-Perez P	Cell Host Microbe	2019	PMC6936622	10.1016/j.chom.2019.08.008
31330855	A Protocol for Extraction of Infective Viromes Suitable for Metagenomics Sequencing from Low Volume Fecal Samples	Deng L	Viruses	2019	PMC6669555	10.3390/v11070667
31323792	Human Virome and Disease: High-Throughput Sequencing for Virus Discovery, Identification of Phage-Bacteria Dysbiosis and Development of Therapeutic Approaches with Emphasis on the Human Gut	Santiago-Rodriguez TM	Viruses	2019	PMC6669467	10.3390/v11070656
31316751	A theoretical model of temperate phages as mediators of gut	Lin DM	F1000Res	2019	PMC6611130	10.12688/f1000research.18480.1

	microbiome dysbiosis					
31314945	Comparative genomics groups phages of Negativicutes and classical Firmicutes despite different Gram-staining properties	Rands CM	Environ Microbiol	2019		10.1111/1462-2920.14746
31285584	Global phylogeography and ancient evolution of the widespread human gut virus crAssphage	Edwards RA	Nat Microbiol	2019	PMC7440971	10.1038/s41564-019-0494-6
31255903	What is (not) known about the dynamics of the human gut virome in health and disease	Beller L	Curr Opin Virol	2019		10.1016/j.coviro.2019.05.013
31252683	Phages and Human Health: More Than Idle Hitchhikers	Lawrence D	Viruses	2019	PMC6669647	10.3390/v11070587
31207872	The Revival of the Battle between David and Goliath in the Enteric Viruses and Microbiota Struggle: Potential Implication for Celiac Disease	Lerner A	Microorganisms	2019	PMC6616392	10.3390/microorganisms7060173
31191981	Gut virome of mammals and birds reveals high genetic diversity of the family Microviridae	Wang H	Virus Evol	2019	PMC6555873	10.1093/ve/vez013
30993039	virMine: automated detection of viral sequences from complex metagenomic samples	Garretto A	PeerJ	2019	PMC6462185	10.7717/peerj.6695
30936943	The gut virome: the 'missing link' between gut bacteria and host immunity?	Mukhopadhyay I	Therap Adv Gastroenterol	2019	PMC6435874	10.1177/1756284819836620
30926942	Gut mucosal virome altered in ulcerative colitis	Ray K	Nat Rev Gastroenterol Hepatol	2019		10.1038/s41575-019-0141-y
30842211	Gut mucosal virome alterations in ulcerative colitis	Zuo T	Gut	2019	PMC6582748	10.1136/gutjnl-2018-318131
30815502	Distinct Gut Virome Profile of Pregnant Women With Type 1 Diabetes in the ENDIA Study	Wook Kim K	Open Forum Infect Dis	2019	PMC6386807	10.1093/ofid/ofz025
30763537	Virome Diversity Correlates with Intestinal Microbiome Diversity in Adult Monozygotic Twins	Moreno-Gallego JL	Cell Host Microbe	2019	PMC6411085	10.1016/j.chom.2019.01.019
30763534	Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome	Shkoporov AN	Cell Host Microbe	2019		10.1016/j.chom.2019.01.017
30741981	Higher abundance of enterovirus A species in the gut of children with islet autoimmunity	Kim KW	Sci Rep	2019	PMC6370883	10.1038/s41598-018-38368-8
30691529	Choice of assembly software has a	Sutton TDS	Microbiome	2019	PMC6350398	10.1186/s40168-019-0626-5

30674646	critical impact on virome characterisation Gut Virome Analysis of Cameroonians Reveals High Diversity of Enteric Viruses, Including Potential Interspecies Transmitted Viruses	Yinda CK	mSphere	2019	PMC6344602	10.1128/mSphere.00585-18
30621339	Bacteriocins and Bacteriophages: Therapeutic Weapons for Gastrointestinal Diseases? Metagenomic analysis of intestinal mucosa revealed a specific	Lopetuso LR	Int J Mol Sci	2019	PMC6337595	10.3390/ijms20010183
30252582	eukaryotic gut virome signature in early-diagnosed inflammatory bowel disease	Ungaro F	Gut Microbes	2019	PMC6546319	10.1080/19490976.2018.1511664
30169455	Enteric Virome and Bacterial Microbiota in Children With Ulcerative Colitis and Crohn Disease	Fernandes MA	J Pediatr Gastroenterol Nutr	2019	PMC6310095	10.1097/MPG.0000000000002140
31383878	A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach	d'Humières C	Sci Rep	2019	PMC6683287	10.1038/s41598-019-47656-w
30543873	The intestinal virome of malabsorption syndrome-affected and unaffected broilers through shotgun metagenomics	Lima DA	Virus Res	2019		10.1016/j.virusres.2018.12.005
30619215	The Human Gut Virome in Hypertension	Han M	Front Microbiol	2018	PMC6305721	10.3389/fmicb.2018.03150
30459201	Diagnostic Potential and Interactive Dynamics of the Colorectal Cancer Virome	Hannigan GD	mBio	2018	PMC6247079	10.1128/mBio.02248-18
30449316	Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut	Guerin E	Cell Host Microbe	2018		10.1016/j.chom.2018.10.002
30429469	ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis	Shkoporov AN	Nat Commun	2018	PMC6235969	10.1038/s41467-018-07225-7
30356736	Immunological Tolerance and Function: Associations Between Intestinal Bacteria, Probiotics, Prebiotics, and Phages	Vitetta L	Front Immunol	2018	PMC6189397	10.3389/fimmu.2018.02240
30205462	Insights into the Human Virome Using CRISPR Spacers from	Hidalgo-Cantabrana C	Viruses	2018	PMC6165519	10.3390/v10090479

	Microbiomes					
30181300	The Intestinal Virome and Immunity	Neil JA	J Immunol	2018	PMC6179364	10.4049/jimmunol.1800631
30128999	Diversity Analysis in Viral Metagenomes	Vázquez-Castellanos JF	Methods Mol Biol	2018		10.1007/978-1-4939-8682-8_15
30128992	Viral Genome Isolation from Human Faeces for Succession Assessment of the Human Gut Virome	Mayo-Muñoz D	Methods Mol Biol	2018		10.1007/978-1-4939-8682-8_8
30128989	Extraction and Purification of Viruses from Fecal Samples for Metagenome and Morphology Analyses	Castro-Mejía JL	Methods Mol Biol	2018		10.1007/978-1-4939-8682-8_5
30126456	Tracing mother-infant transmission of bacteriophages by means of a novel analytical tool for shotgun metagenomic datasets: METAnnotatorX	Milani C	Microbiome	2018	PMC6102903	10.1186/s40168-018-0527-z
29751830	Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community	Lim ES	Microbiome	2018	PMC5946436	10.1186/s40168-018-0475-7
29689266	Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival Outcomes	Nakatsu G	Gastroenterology	2018		10.1053/j.gastro.2018.04.018
29668682	Biogeography and environmental conditions shape bacteriophage-bacteria networks across the human microbiome	Hannigan GD	PLoS Comput Biol	2018	PMC5927471	10.1371/journal.pcbi.1006099
29638163	Mining the Virome for Insights into Type 1 Diabetes	Park A	DNA Cell Biol	2018	PMC5944395	10.1089/dna.2018.4185
29631623	Reproducible protocols for metagenomic analysis of human faecal phageomes	Shkoporov AN	Microbiome	2018	PMC5892011	10.1186/s40168-018-0446-z
29615108	Phages infecting Faecalibacterium prausnitzii belong to novel viral genera that help to decipher intestinal viromes	Cornuault JK	Microbiome	2018	PMC5883640	10.1186/s40168-018-0452-1
29570694	Enteric Virome Sensing-Its Role in Intestinal Homeostasis and Immunity	Metzger RN	Viruses	2018	PMC5923440	10.3390/v10040146
29523751	Does the microbiome and virome contribute to myalgic encephalomyelitis/chronic fatigue syndrome?	Newberry F	Clin Sci (Lond)	2018	PMC5843715	10.1042/CS20171330
29378882	Bacteriophages of the Urinary Microbiome	Miller-Ensminger T	J Bacteriol	2018	PMC5847656	10.1128/JB.00738-17

29297933	Quantitative CrAssphage real-time PCR assay derived from data of multiple geographically distant populations	Cinek O	J Med Virol	2018		10.1002/jmv.25012
29272453	Isolation in small populations of Wayampi Amerindians promotes endemicity and homogenisation of their faecal virome, but its distribution is not entirely random	Pérez-Brocail V	FEMS Microbiol Ecol	2018		10.1093/femsec/fix184
29266753	Specific changes of enteric mycobiota and virome in inflammatory bowel disease	Chu Y	J Dig Dis	2018		10.1111/1751-2980.12570
29133882	Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut	Yutin N	Nat Microbiol	2018	PMC5736458	10.1038/s41564-017-0053-y
28539351	Bacteriophage transfer during faecal microbiota transplantation in Clostridium difficile infection is associated with treatment outcome	Zuo T	Gut	2018	PMC5868238	10.1136/gutjnl-2017-313952
29567068	Dynamics of the human gut phageome during antibiotic treatment	Górska A	Comput Biol Chem	2018		10.1016/j.compbiolchem.2018.03.011
29391057	A human gut phage catalog correlates the gut phageome with type 2 diabetes	Ma Y	Microbiome	2018	PMC5796561	10.1186/s40168-018-0410-y
33525769	The human gut virome: form and function	Ogilvie LA	Emerg Top Life Sci	2017		10.1042/ETLS20170039
29026445	Viral communities of the human gut: metagenomic analysis of composition and dynamics	Aggarwala V	Mob DNA	2017	PMC5627405	10.1186/s13100-017-0095-y
28967228	Characterization of the human DNA gut virome across populations with different subsistence strategies and geographical origin	Rampelli S	Environ Microbiol	2017		10.1111/1462-2920.13938
28952375	The gut virome: a neglected actor in colon cancer pathogenesis	Dahiya DK	Future Microbiol	2017		10.2217/fmb-2017-0159
28869283	Review article: the human intestinal virome in health and disease	Carding SR	Aliment Pharmacol Ther	2017	PMC5656937	10.1111/apt.14280
28759053	The eukaryotic gut virome in hematopoietic stem cell transplantation: new clues in enteric graft-versus-host disease	Legoff J	Nat Med	2017		10.1038/nm.4380

28743462	Bacteriophages in the human gut: Our fellow travelers throughout life and potential biomarkers of health or disease	Bakhshinejad B	Virus Res	2017		10.1016/j.virusres.2017.07.013
28732571	Examining the gut bacteriome, virome, and mycobiome in glucose metabolism disorders: Are we on the right track?	Stefanaki C	Metabolism	2017		10.1016/j.metabol.2017.04.014
28466442	Viral metagenomics analysis of feces from coronary heart disease patients reveals the genetic diversity of the Microviridae	Guo L	Virology	2017	PMC6598973	10.1007/s12250-016-3896-0
28448616	Abundance profiling of specific gene groups using precomputed gut metagenomes yields novel biological hypotheses	Yarygin K	PLoS One	2017	PMC5407692	10.1371/journal.pone.0176154
28232956	Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay	Marbouty M	Sci Adv	2017	PMC5315449	10.1126/sciadv.1602105
27935413	Stable core virome despite variable microbiome after fecal transfer	Broecker F	Gut Microbes	2017	PMC5479397	10.1080/19490976.2016.1265196
27326545	Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses	Enault F	ISME J	2017	PMC5315482	10.1038/ismej.2016.90
28856733	A bacteriophages journey through the human body	Barr JJ	Immunol Rev	2017		10.1111/imr.12565
28886006	The enteric virome in hematopoietic stem cell transplantation: ready for its close-up	Takashima S	Nat Med	2017		10.1038/nm.4403
28617757	Looking into Enteric Virome in Patients with IBD: Defining Guilty or Innocence?	Lopes S	Inflamm Bowel Dis	2017		10.1097/MIB.0000000000001167
27912785	Transmission of viruses via our microbiomes	Ly M	Microbiome	2016	PMC5134127	10.1186/s40168-016-0212-z
27721367	A Pair of Identical Twins Discordant for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Differ in Physiological Parameters and Gut Microbiome Composition	Giloteaux L	Am J Case Rep	2016	PMC5058431	10.12659/ajcr.900314
27353648	The Bacterial Microbiome and Virome Milestones of Infant	Lim ES	Trends Microbiol	2016		10.1016/j.tim.2016.06.001

	Development					
27286042	Long-term microbiota and virome in a Zürich patient after fecal transplantation against <i>Clostridium difficile</i> infection	Broecker F	Ann N Y Acad Sci	2016		10.1111/nyas.13100
27206017	Gut Virome and Inflammatory Bowel Disease	Lopetuso LR	Inflamm Bowel Dis	2016		10.1097/MIB.0000000000000807
27148577	Long-term changes of bacterial and viral compositions in the intestine of a recovered <i>Clostridium difficile</i> patient after fecal microbiota transplantation	Broecker F	Cold Spring Harb Mol Case Stud	2016	PMC4849847	10.1101/mcs.a000448
27096311	Viral Safeguard: The Enteric Virome Protects against Gut Inflammation	Karst SM	Immunity	2016		10.1016/j.immuni.2016.04.004
26962942	Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome	Monaco CL	Cell Host Microbe	2016	PMC4821831	10.1016/j.chom.2016.02.011
26919534	Beyond the gut bacterial microbiota: The gut virome	Columpsi P	J Med Virol	2016	PMC7166815	10.1002/jmv.24508
26780893	The fecal virome of South and Central American children with diarrhea includes small circular DNA viral genomes of unknown origin	Phan TG	Arch Virol	2016	PMC4814309	10.1007/s00705-016-2756-4
26568175	Analysis of dsDNA and RNA viromes in methanogenic digesters reveals novel viral genetic diversity	Calusinska M	Environ Microbiol	2016	PMC7163765	10.1111/1462-2920.13127
26564967	Natural mummification of the human gut preserves bacteriophage DNA	Santiago-Rodriguez TM	FEMS Microbiol Lett	2016	PMC5812488	10.1093/femsle/fnv219
27573828	Healthy human gut phageome	Manrique P	Proc Natl Acad Sci U S A	2016	PMC5027468	10.1073/pnas.1601060113
26559140	Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis	Conceição-Neto N	Sci Rep	2015	PMC4642273	10.1038/srep16532
26441889	Direct sequencing of human gut virome fractions obtained by flow cytometry	Džunková M	Front Microbiol	2015	PMC4568480	10.3389/fmicb.2015.00955
26441861	The human gut virome: a multifaceted majority	Ogilvie LA	Front Microbiol	2015	PMC4566309	10.3389/fmicb.2015.00918
26416338	Gut microbiota: The gut virome and bacterial microbiome-the early years	Ray K	Nat Rev Gastroenterol Hepatol	2015		10.1038/nrgastro.2015.169



26366711	Early life dynamics of the human gut virome and bacterial microbiome in infants	Lim ES	Nat Med	2015	PMC4710368	10.1038/nm.3950
26351661	Gut DNA viromes of Malawian twins discordant for severe acute malnutrition	Reyes A	Proc Natl Acad Sci U S A	2015	PMC4586842	10.1073/pnas.1514285112
26257129	The human gut microbiota and virome: Potential therapeutic implications	Scarpellini E	Dig Liver Dis	2015	PMC7185617	10.1016/j.dld.2015.07.008
26137199	Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome	Bikel S	Comput Struct Biotechnol J	2015	PMC4484546	10.1016/j.csbj.2015.06.001
25944980	Gut inflammation and immunity: what is the role of the human gut virome?	Focà A	Mediators Inflamm	2015	PMC4405218	10.1155/2015/326032
25939040	Metagenomic analysis of microbiome in colon tissue from subjects with inflammatory bowel diseases reveals interplay of viruses and bacteria	Wang W	Inflamm Bowel Dis	2015	PMC4450971	10.1097/MIB.0000000000000344
25852662	Genomic characteristics and environmental distributions of the uncultivated Far-T4 phages	Roux S	Front Microbiol	2015	PMC4360716	10.3389/fmicb.2015.00199
25678103	Gut virome sequencing in children with early islet autoimmunity	Kramná L	Diabetes Care	2015		10.2337/dc14-2490
26319032	The Enteric Virome in Inflammatory Bowel Disease	Brooks J	Gastroenterology	2015	PMC7130344	10.1053/j.gastro.2015.08.022
25619688	Disease-specific alterations in the enteric virome in inflammatory bowel disease	Norman JM	Cell	2015	PMC4312520	10.1016/j.cell.2015.01.002
25381053	Metagenomic analysis of the shrew enteric virome reveals novel viruses related to human stool-associated viruses	Sasaki M	J Gen Virol	2015		10.1099/vir.0.071209-0
25262473	Geographic variation in the eukaryotic virome of human diarrhea	Holtz LR	Virology	2014	PMC4254309	10.1016/j.virol.2014.09.012
24656964	Exploiting gut bacteriophages for human health	Dalmasso M	Trends Microbiol	2014		10.1016/j.tim.2014.02.010
26458273	Metagenomic ventures into outer sequence space	Dutilh BE	Bacteriophage	2014	PMC4588555	10.4161/21597081.2014.979664

24036533	Genome signature-based dissection of human gut metagenomes to extract subliminal viral sequences	Ogilvie LA	Nat Commun	2013	PMC3778543	10.1038/ncomms3420
23872943	Viral evolution: Variation in the gut virome	Hofer U	Nat Rev Microbiol	2013		10.1038/nrmicro3092
23836644	Rapid evolution of the human gut virome	Minot S	Proc Natl Acad Sci U S A	2013	PMC3725073	10.1073/pnas.1300833110
23664726	Evidence of the megavirome in humans	Colson P	J Clin Virol	2013		10.1016/j.jcv.2013.03.018
22808158	Evolution and diversity of the Microviridae viral family through a collection of 81 new complete genomes assembled from virome reads	Roux S	PLoS One	2012	PMC3394797	10.1371/journal.pone.0040418
22558115	Comparative (meta)genomic analysis and ecological profiling of human gut-specific bacteriophage $\phi$ B124-14	Ogilvie LA	PLoS One	2012	PMC3338817	10.1371/journal.pone.0035053
22355105	Hypervariable loci in the human gut virome	Minot S	Proc Natl Acad Sci U S A	2012	PMC3309749	10.1073/pnas.1119061109
21880779	The human gut virome: inter-individual variation and dynamic response to diet	Minot S	Genome Res	2011	PMC3202279	10.1101/gr.122705.111
20631792	Viruses in the faecal microbiota of monozygotic twins and their mothers	Reyes A	Nature	2010	PMC2919852	10.1038/nature09199

**Anexo 2. Tabla con la lista de los estudios de trasplantes de fagos realizados hasta la actualidad.** El orden de aparición inicia a partir del estudio más reciente.

PMID	Título	Primer autor	Revista/libro	Año de publicación	PMCID	DOI
33397897	Longitudinal dynamics of gut bacteriome, mycobiome and virome after fecal microbiota transplantation in graft-versus-host disease	Zhang F	Nat Commun	2021	PMC7782528	10.1038/s41467-020-20240-x
33218339	Autochthonous faecal viral transfer (FVT) impacts the murine microbiome after antibiotic perturbation	Draper LA	BMC Biol	2020	PMC7679995	10.1186/s12915-020-00906-0
32495834	Bacteriophage-mediated manipulation of the gut microbiome - promises and presents limitations	Rasmussen TS	FEMS Microbiol Rev	2020		10.1093/femsre/fuaa020
32165408	Faecal virome transplantation decreases symptoms of type 2 diabetes and obesity in a murine model	Rasmussen TS	Gut	2020		10.1136/gutjnl-2019-320005
31941900	Acquisition, transmission and strain diversity of human gut-colonizing crAss-like phages	Siranosian BA	Nat Commun	2020	PMC6962324	10.1038/s41467-019-14103-3
31750259	Transplanting Fecal Virus-Like Particles Reduces High-Fat Diet-Induced Small Intestinal Bacterial Overgrowth in Mice	Lin DM	Front Cell Infect Microbiol	2019	PMC6843071	10.3389/fcimb.2019.00348
30866714	The success of fecal microbial transplantation in Clostridium difficile infection correlates with bacteriophage relative abundance in the donor: a retrospective cohort study	Park H	Gut Microbes	2019	PMC6867182	10.1080/19490976.2019.1586037
30526683	Long-term colonisation with donor bacteriophages following successful faecal microbial transplantation	Draper LA	Microbiome	2018	PMC6288847	10.1186/s40168-018-0598-x

29670863	Commentary: Bacteriophage transfer during faecal microbiota transplantation in Clostridium difficile infection is associated with treatment outcome	Anonye BO	Front Cell Infect Microbiol	2018	PMC5893759	10.3389/fcimb.2018.00104
28539351	Bacteriophage transfer during faecal microbiota transplantation in Clostridium difficile infection is associated with treatment outcome	Zuo T	Gut	2018	PMC5868238	10.1136/gutjnl-2017-313952
28611478	Therapy: Bacteriophages important for FMT efficacy	Dickson I	Nat Rev Gastroenterol Hepatol	2017		10.1038/nrgastro.2017.83
27935413	Stable core virome despite variable microbiome after fecal transfer	Broecker F	Gut Microbes	2017	PMC5479397	10.1080/19490976.2016.1265196
28090385	Fecal microbiota transplantation to fight Clostridium difficile infections and other intestinal diseases	Moelling K	Bacteriophage	2016	PMC5221744	10.1080/21597081.2016.1251380
27286042	Long-term microbiota and virome in a Zürich patient after fecal transplantation against Clostridium difficile infection	Broecker F	Ann N Y Acad Sci	2016		10.1111/nyas.13100
27148577	Long-term changes of bacterial and viral compositions in the intestine of a recovered Clostridium difficile patient after fecal microbiota transplantation	Broecker F	Cold Spring Harb Mol Case Stud	2016	PMC4849847	10.1101/mcs.a000448
27025251	Transfer of Viral Communities between Human Individuals during Fecal Microbiota Transplantation	Chehoud C	mBio	2016	PMC4817255	10.1128/mBio.00322-16

### Anexo 3. Conteo de VLPs por muestra.

Muestra ID	Conteo por triplicado	Conteo de VLPs por campo					Promedio del conteo de VLPs por muestra	VLPs/ 10µl de muestra	VLPs/ 250 mg de heces	VLPs/ gramo de heces
		1	2	3	4	5				
NW-10	1	1355	964	1003	1496	1147	1211	15536178	310723563	1242894254
	2	1760	2506	920	1146	988				
	3	927	1014	707	1086	1147				
NW-118	1	13959	17957	14118	10872	14623	14811	190003029	3800060579	15200242316
NW-119	2	14892	18204	17755	16629	18793				
NW-120	3	10511	17228	14964	10670	10990				
NW-124	1	4339	2210	3779	3081	3530	3041	39006361	780127216	3120508864
	2	3666	3028	4233	1518	2933				
	3	2210	2799	3442	2579	2262				
NW-147	1	540	802	185	891	8110	1371	17591305	351826102	1407304410
	2	643	1494	640	586	678				
	3	1079	1055	833	1606	1427				
NW-161	1	579	556	1835	1811	378	1033	13252237	265044746	1060178984
	2	1218	345	403	315	111				
	3	579	6479	1835	1811	378				
NW-169	1	1305	3777	2405	3092	2969	4606	59083831	1181676615	4726706459
	2	5708	8132	15074	5982	1813				
	3	5708	8132	1795	1774	1419				
NW-193	1	3485	2630	2048	1389	2978	1947	24980526	499610512	1998442049
	2	2394	1571	1821	1606	1867				
	3	2124	1367	892	1570	1467				
NW-314	1	1315	178	2850	4107	3366	1932	24781256	495625122	1982500490
	2	2978	1230	1886	1612	2083				
	3	253	598	839	1349	4332				
O-121	1	41499	47671	50919	48380	43696	46771	600002138	12000042762	48000171046
	2	49528	51804	51156	51744	42080				
	O-123	3	44412	45300	43321	42641				
O-152	1	5718	5498	5126	3636	7262	4159	53349488	1066989755	4267959020
	2	2806	8352	4261	2232	3539				
	3	2309	3322	2914	1576	3829				
O-39	1	1421	2999	2963	3369	4469	2362	30299225	605984499	2423937996
	2	4013	1710	1362	2581	2224				
	3	2587	1209	1002	1959	1560				
O-418	1	878	769	861	2482	1089	1130	14493648	289872962	1159491849
	2	525	1319	884	747	541				
	3	1213	2275	1806	469	1089				
O-420	1	3314	1740	2308	1864	1866	1995	25586886	511737728	2046950913
	2	5933	1238	1357	1426	1862				
	3	1115	828	1671	1534	1862				
O-434	1	2574	271	830	678	1565	952	12216160	244323207	977292829
	2	716	741	830	678	2563				
	3	448	351	418	1106	515				
O-445	1	296	750	1069	966	1688	1124	14420098	288401960	1153607840
	2	928	1086	895	1225	1174				
	3	1869	1238	1023	966	1688				
O-90	1	178	3620	9833	4006	790	3195	40981096	819621915	3278487661
	2	784	6203	2946	6779	2563				
	3	811	752	1903	939	5811				
OMS-124	1	29725	32245	30969	37031	34731	33519	429998753	8599975056	34399900222
OMS-125	2	37130	31979	35136	32670	36854				
OMS-126	3	38089	29937	30916	35198	30176				
OMS-288	1	3381	4479	2692	4097	5630	2752	35307474	706149488	2824597951
	2	2423	2086	3132	2321	2691				
	3	1476	2086	2053	1416	1321				
OMS-446	1	1080	556	721	1684	662	870	11155670	223113408	892453630
	2	731	363	527	705	558				
	3	279	751	888	579	2960				
OMS-55	1	396	660	1512	568	1101	833	10685292	213705835	854823341
	2	560	2155	920	465	818				
	3	510	1077	610	588	554				
OMS-64	1	1487	1265	1945	874	15131	1999	25646753	512935056	2051740223
	2	352	345	304	874	920				
	3	1487	1265	1945	874	920				
OMS-87	1	3315	3927	2659	3118	3449	2504	32119163	642383252	2569533007
	2	2570	2451	1701	2560	2219				
	3	1120	1427	1963	1628	3449				

## Anexo 4. Resultados de la secuenciación de VLPs

Muestra ID	Estado de salud	Lecturas crudas (paired_seq)	Lecturas de calidad (paired_seq)	Lecturas de calidad (%)	Lecturas asignadas a bacteria	Lecturas a signadas a bacteria (%)
NW_118	Peso normal saludable	11523930	11278512	0.98	732044	0.06
NW_119	Peso normal saludable	9001074	8971662	1.00	3371988	0.38
NW_120	Peso normal saludable	7149441	5234936	0.73	751590	0.14
NW_10	Peso normal saludable	2435784	2178708	0.89	460034	0.21
NW_124	Peso normal saludable	6182666	5671004	0.92	1282500	0.23
NW_147	Peso normal saludable	6693958	6153570	0.92	1725486	0.28
NW_161	Peso normal saludable	9052662	8267708	0.91	2126280	0.26
NW_169	Peso normal saludable	505962	417304	0.82	74688	0.18
NW_193	Peso normal saludable	7695966	7099246	0.92	1816298	0.26
NW_314	Peso normal saludable	1561858	1324090	0.85	251708	0.19
O_121	Obeso	9590944	9543816	1.00	1470352	0.15
O_122	Obeso	8413578	8326110	0.99	2616064	0.31
O_123	Obeso	1069494	1066638	1.00	316432	0.30
O_152	Obeso	11023644	10254262	0.93	1833984	0.18
O_39	Obeso	7848438	7339336	0.94	4441906	0.61
O_418	Obeso	1759626	1573216	0.89	454172	0.29
O_420	Obeso	3456064	3245548	0.94	1748292	0.54
O_434	Obeso	1672776	1544018	0.92	421154	0.27
O_445	Obeso	4430022	4148356	0.94	2178058	0.53
O_90	Obeso	1615954	1444198	0.89	384876	0.27
OMS_124	Obeso con síndrome metabólico	4020298	4011836	1.00	1183746	0.30
OMS_125	Obeso con síndrome metabólico	5956082	7864842	1.32	1235256	0.16
OMS_126	Obeso con síndrome metabólico	9055978	9021688	1.00	2352770	0.26
OMS_288	Obeso con síndrome metabólico	2374056	2180938	0.92	225186	0.10
OMS_446	Obeso con síndrome metabólico	1361720	1189158	0.87	240556	0.20
OMS_55	Obeso con síndrome metabólico	3918888	3607194	0.92	1627092	0.45
OMS_64	Obeso con síndrome metabólico	2697578	2444974	0.91	739352	0.30
OMS_87	Obeso con síndrome metabólico	1116012	987244	0.88	480086	0.49

## Continuación Anexo 4

Muestra ID	Lecturas asignadas a humano	Lecturas asignadas a humano (%)	Lecturas de calidad filtradas (sin bacteria y humano) (paired_seq)	Lecturas de calidad filtradas (Sin bacteria y humano) (%)	Lecturas agrupadas (95%)	Tasa normalizada de lecturas únicas/lecturas de calidad filtradas totales
NW_118	4063098	0.360251246	6483370	0.574842674	1555888	0.239981368
NW_119	1493938	0.166517419	4105736	0.457633825	1704013	0.415032287
NW_120	553148	0.105664711	3930198	0.750763333	1371042	0.348848073
NW_10	285808	0.131182334	1415490	0.649692386	358813	0.253490311
NW_124	515296	0.090865039	3873168	0.682977476	746134	0.192641786
NW_147	650124	0.105649891	3777848	0.61392785	768361	0.203385896
NW_161	1258264	0.152190184	4883020	0.590613505	1074267	0.220000532
NW_169	43046	0.103152618	299550	0.717822019	116590	0.389217159
NW_193	1275480	0.17966415	4007412	0.564484172	1348811	0.336579069
NW_314	127186	0.096055404	945188	0.713839694	211178	0.223424335
O_121	1851296	0.193978593	6222168	0.651958085	2362429	0.379679398
O_122	2319370	0.278565861	3390676	0.407234111	1317865	0.388673232
O_123	36024	0.033773408	714182	0.66956362	332384	0.465405177
O_152	3038368	0.296302942	5213758	0.508447902	1574085	0.30190987
O_39	519204	0.070742639	2377282	0.323909683	647382	0.272320238
O_418	185550	0.117943118	933454	0.593341283	269359	0.288561622
O_420	151982	0.046827839	1345222	0.414482238	470981	0.350113959
O_434	177170	0.114746072	945650	0.612460476	277325	0.293263893
O_445	370950	0.089420966	1599244	0.385512719	464474	0.29043348
O_90	359750	0.249100193	699572	0.484401723	281057	0.401755645
OMS_124	508362	0.126715549	2319728	0.578221044	952354	0.410545547
OMS_125	1438984	0.182964133	5190602	0.659975369	2154505	0.415078058
OMS_126	2112944	0.234207168	4555974	0.50500239	1896348	0.41623328
OMS_288	725410	0.332613765	1229670	0.563826207	491198	0.399455138
OMS_446	131178	0.110311666	817376	0.687356937	198943	0.24339227
OMS_55	338782	0.093918431	1641244	0.454991886	478071	0.291285756
OMS_64	208510	0.08528107	1497064	0.612302626	391700	0.261645461
OMS_87	61440	0.062233855	445510	0.451266354	175536	0.394011358

## Continuación Anexo 4

Muestra ID	% hits al (NR) RefSeq viral	% no hits al (NR) RefSeq viral	Lecturas únicas_Virus procariotas	Lecturas únicas_Virus eucariotas	Lecturas únicas_No definidas
NW_118	0.051406578	1.444202646	1.293494705	0.046308254	0.284675842
NW_119	0.258829391	6.226444122	4.838013357	0.707169477	1.989778975
NW_120	0.263431099	6.839259053	4.623630124	0.782627947	2.31264529
NW_10	0.224166527	5.490629376	3.80816206	0.826722261	2.04162208
NW_124	0.055849894	3.054195988	0.864912734	0.088742113	0.571565625
NW_147	0.073027102	3.129665664	1.4163684	0.07656544	0.630235047
NW_161	0.089729134	4.38804518	1.547359876	0.266041843	0.711043873
NW_169	0.01376761	0.472204414	0.291502159	0.010700712	0.088188628
NW_193	0.153215195	5.468917152	2.731818014	0.467325929	1.207889008
NW_314	0.013734264	0.866500747	0.227851371	0	0.144644109
O_121	0.40083626	9.446272415	7.753034943	1.160289288	2.875908638
O_122	0.211703568	5.281439241	4.088594517	0.608280137	1.54090255
O_123	0.07060537	1.314842137	1.563226449	0.160879672	0.407180547
O_152	0.29149985	6.269622647	6.784251504	0.534666617	1.59403712
O_39	0.082780722	2.615645732	2.031843843	0.07619645	0.482270027
O_418	0.020716022	1.102029824	0.399062765	0.028596731	0.152392901
O_420	0.051377401	1.911772006	1.155123427	0.060883362	0.346666175
O_434	0.022454167	1.133495669	0.422124645	0.028781226	0.180989631
O_445	0.049018192	1.887008649	1.122652301	0.057746947	0.317700454
O_90	0.022070691	1.149434917	0.384856647	0.045385779	0.181912107
OMS_124	0.119152554	3.850462336	2.06652891	0.385225637	1.010479318
OMS_125	0.350467567	8.629969464	6.424670676	1.299952031	2.658573484
OMS_126	0.310644455	7.59373906	5.733367772	1.130954577	2.33865909
OMS_288	0.065420112	1.981998069	1.142577765	0.248514815	0.535589093
OMS_446	0.01362589	0.815611027	0.216043688	0	0.140216228
OMS_55	0.050276992	1.942425049	1.139810339	0.063281798	0.32194384
OMS_64	0.038760218	1.593929047	0.82008044	0.053688056	0.27471311
OMS_87	0.018144234	0.713527308	0.412715398	0.019925464	0.118445814

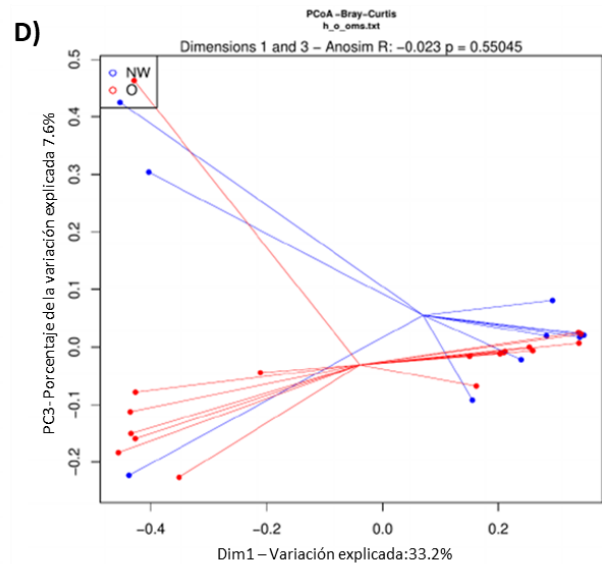
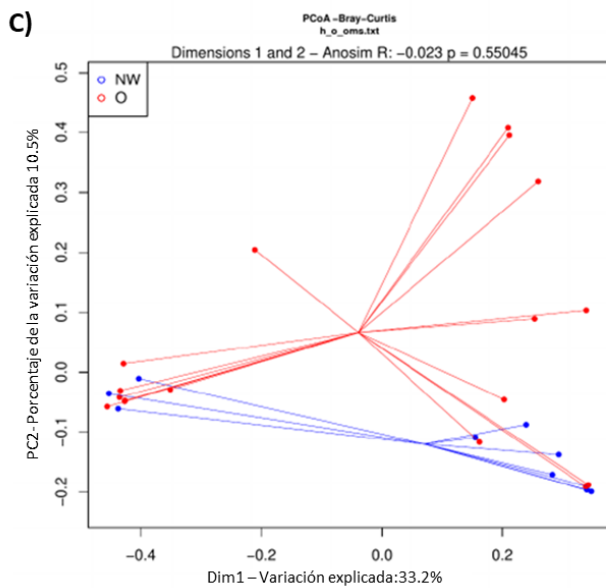
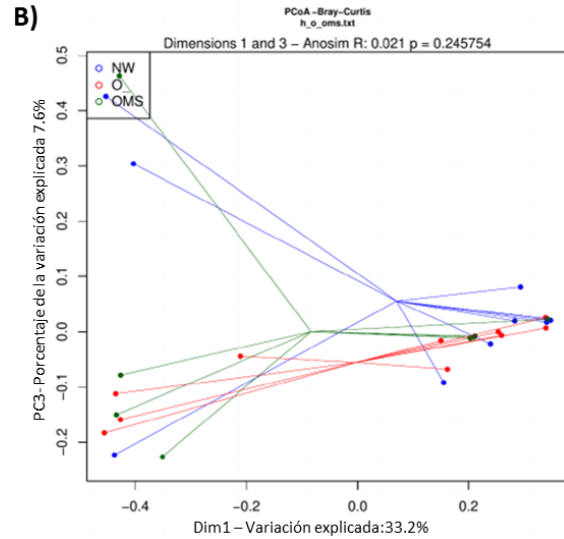
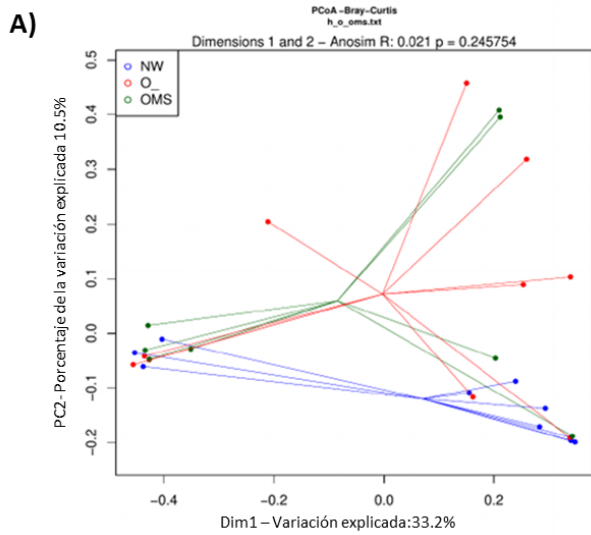
**Continuación Anexo 4**



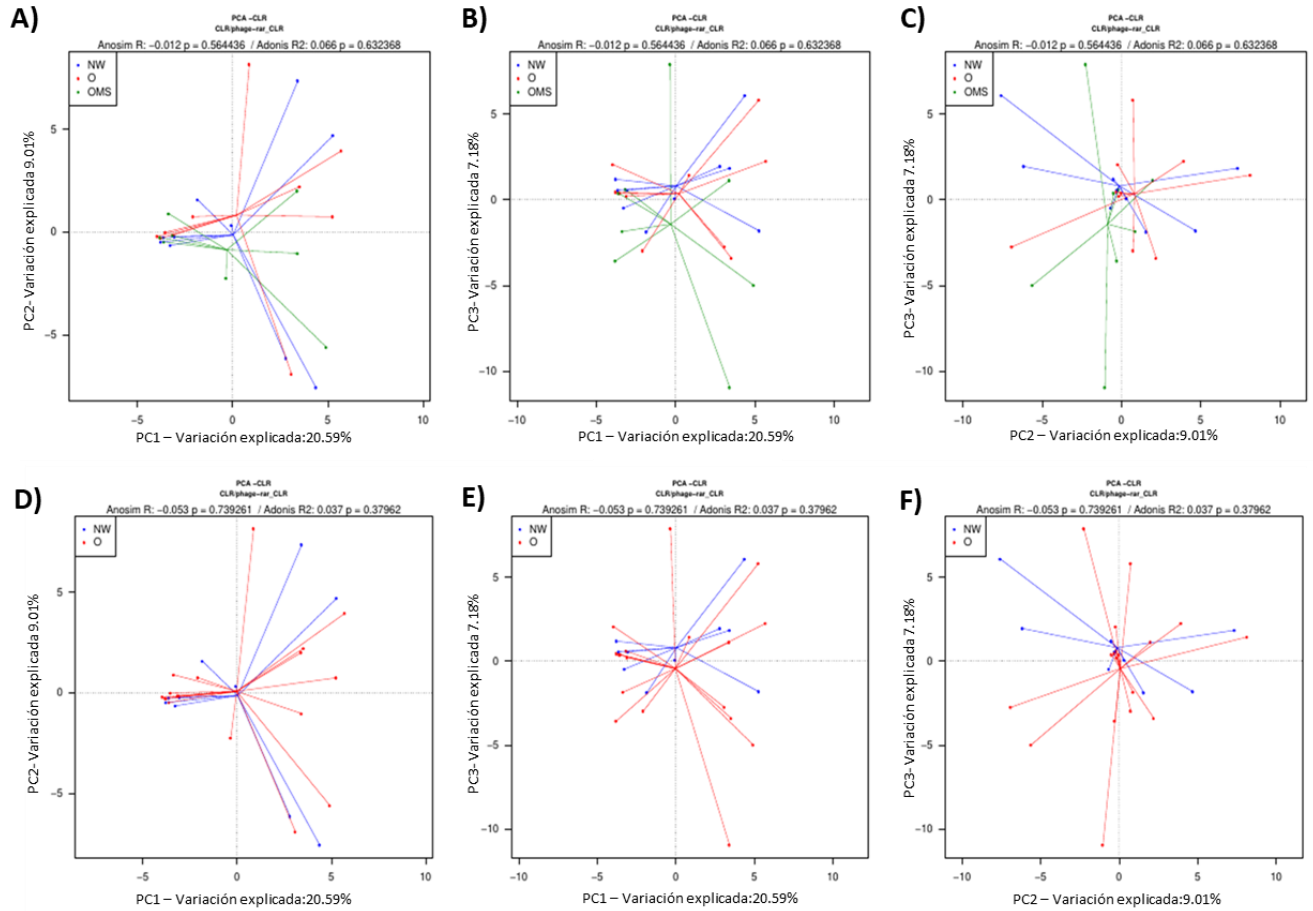
## Anexo 5. Abundancia relativa de la clasificación taxonómica de los 4,611 contigs de fagos por muestra.

Taxonomía de bacteriófagos	NW-10	NW-118	NW-119	NW-120	NW-124	NW-147	NW-161	NW-169	NW-193	NW-314	O-39	O-90	O-121	O-122	O-123	O-152	O-418	O-420	O-434	O-445	OMS-55	OMS-64	OMS-87	OMS-124	OMS-125	OMS-126	OMS-288	OMS-446	Promedio	Promedio (%)	Desviación estándar		
Viruses:dsDNA viruses, no RNA stage;	0.32	0.31	0.31	0.31	0.30	0.33	0.31	0.33	0.31	0.37	0.33	0.30	0.31	0.31	0.30	0.31	0.36	0.32	0.35	0.32	0.33	0.35	0.34	0.30	0.30	0.31	0.30	0.37	0.32	32.21	0.02		
Caudovirales;Siphoviridae;																																	
Viruses:dsDNA viruses, no RNA stage;	0.29	0.31	0.29	0.31	0.29	0.27	0.30	0.28	0.30	0.23	0.28	0.30	0.31	0.31	0.31	0.29	0.27	0.28	0.24	0.30	0.29	0.27	0.27	0.30	0.31	0.31	0.31	0.21	0.29	28.60	0.03		
Caudovirales;Myoviridae;																																	
Viruses:dsDNA viruses, no RNA stage;	0.25	0.24	0.24	0.24	0.26	0.26	0.25	0.25	0.24	0.28	0.23	0.25	0.24	0.24	0.24	0.24	0.21	0.25	0.26	0.25	0.25	0.26	0.23	0.24	0.24	0.23	0.24	0.26	0.25	24.54	0.01		
Caudovirales;																																	
Viruses:undetermined phage;	0.07	0.07	0.07	0.07	0.08	0.07	0.07	0.06	0.07	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.09	0.08	0.08	0.07	0.07	0.07	0.07	0.08	0.07	0.07	0.07	0.07	0.07	7.12	0.01		
Viruses:dsDNA viruses, no RNA stage;																																	
Caudovirales;Podoviridae;	0.05	0.06	0.06	0.06	0.06	0.08	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.07	0.05	0.05	0.06	0.05	0.06	0.05	0.05	0.07	0.06	0.06	0.06	0.04	0.06	5.93	0.01		
Viruses:unclassified bacterial viruses;																																	
crAss-like viruses;	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.04	0.01	0.64	0.01		
Viruses:dsDNA viruses, no RNA stage;																																	
unclassified dsDNA phages;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00		
Deep-sea thermophilic phage D6E;																																	
Viruses:dsDNA viruses;inoviridae;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00		
Viruses:dsDNA viruses, no RNA stage;																																	
unclassified dsDNA phages;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00		
Oenococcus phage;																																	
Viruses:dsDNA viruses, no RNA stage;																																	
unclassified dsDNA phages;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00		
Clostridium phage;																																	
Viruses:unclassified bacterial viruses;																																	
Mycoplasma phage phiMfV1;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00		
Viruses:dsDNA viruses, no RNA stage;																																	
unclassified dsDNA phages;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00		
Viruses:dsDNA viruses, no RNA stage;																																	
Caudovirales;Ackermannviridae;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00		
Viruses:dsDNA viruses, no RNA stage;																																	
unclassified dsDNA phages;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00		
Pseudomonas phage O4;																																	
Viruses:dsDNA viruses, no RNA stage;																																	
unclassified dsDNA phages;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00		
Bacillus phage BaMu-1;																																	
Viruses:dsDNA viruses, no RNA stage;																																	
unclassified dsDNA phages;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00		
Cyanophage KBS-5-24;																																	
Viruses:unclassified bacterial viruses;																																	
Enterobacteria phage YZ-2008;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00		
Viruses:unclassified bacterial viruses;																																	
Synechococcus phage S;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00		
Viruses:dsDNA viruses;Microviridae;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00		
Viruses:unclassified bacterial viruses;																																	
Phage Gifsy-1;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00		
Viruses:dsDNA viruses, no RNA stage;																																	
unclassified dsDNA phages;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00		
Salisaeta icosahedral phage 1;																																	
Viruses:dsDNA viruses, no RNA stage;																																	
unclassified dsDNA phages;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00		
Rhodobacter phage RcapN1;																																	
Viruses:dsDNA viruses, no RNA stage;																																	
unclassified dsDNA phages;	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00		
Vibrio phage douglas 12A4;																																	

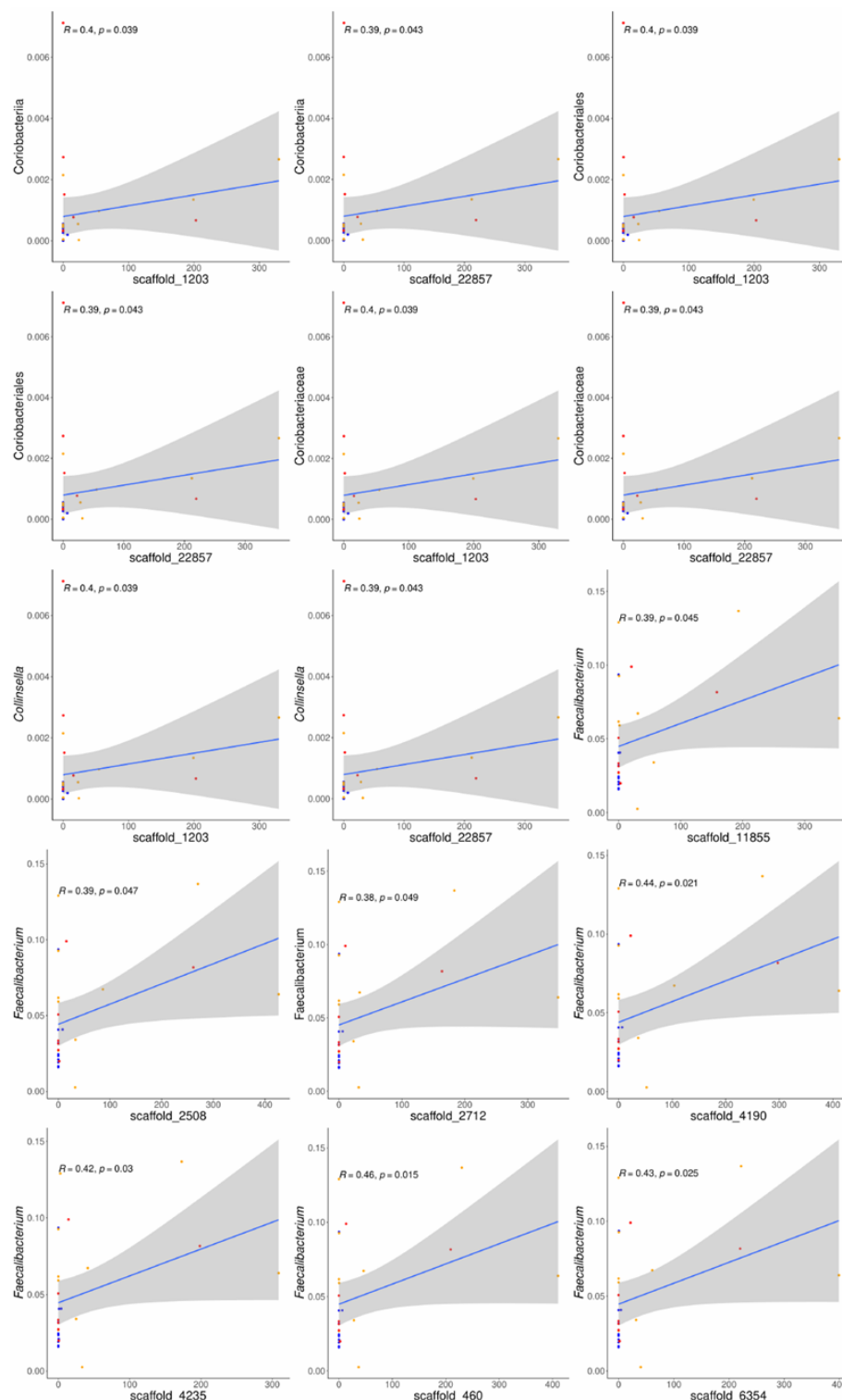
**Anexo 6. Análisis de Coordinadas Principales (PCoA) basado en la disimilitudde Bray-Curtis.**  
 En Muestras NW, O y OMS: A) PC1 vs PC2 y B) PC1 vs PC3. En muestras de todos los obesos (O + OMS) y NW: C) PC1 vs PC2 y D) PC1 vs PC3.

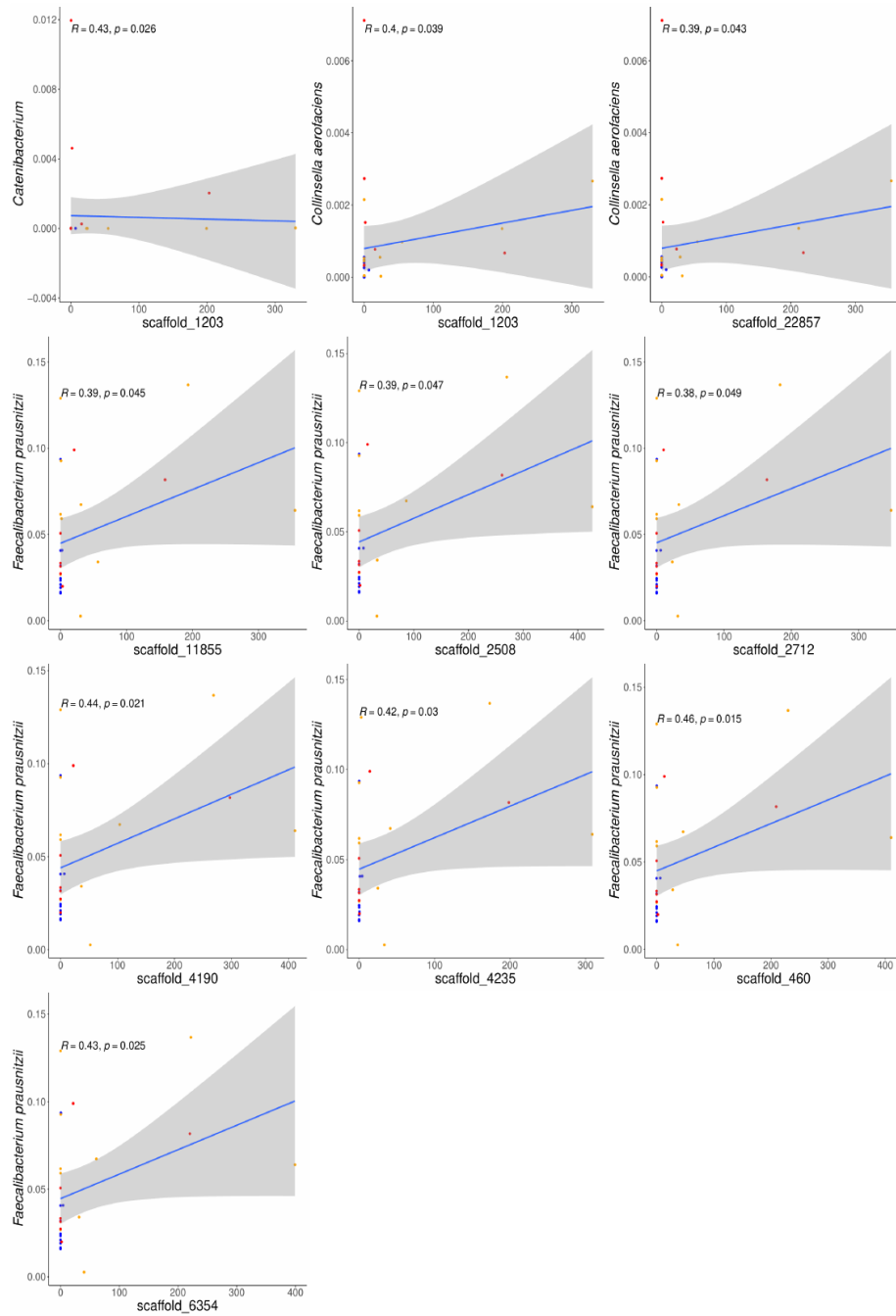


**Anexo 7. PCoA basado en coordenadas CLR.** Muestras NW, O y OMS: A) PC1 vs PC2; B) PC1 vs PC3; y C) PC2 vs PC3. Muestras de todos los obesos (O + OMS) y NW: D) PC1 vs PC2; E) PC1 vs PC3; y F) PC2 vs PC3.

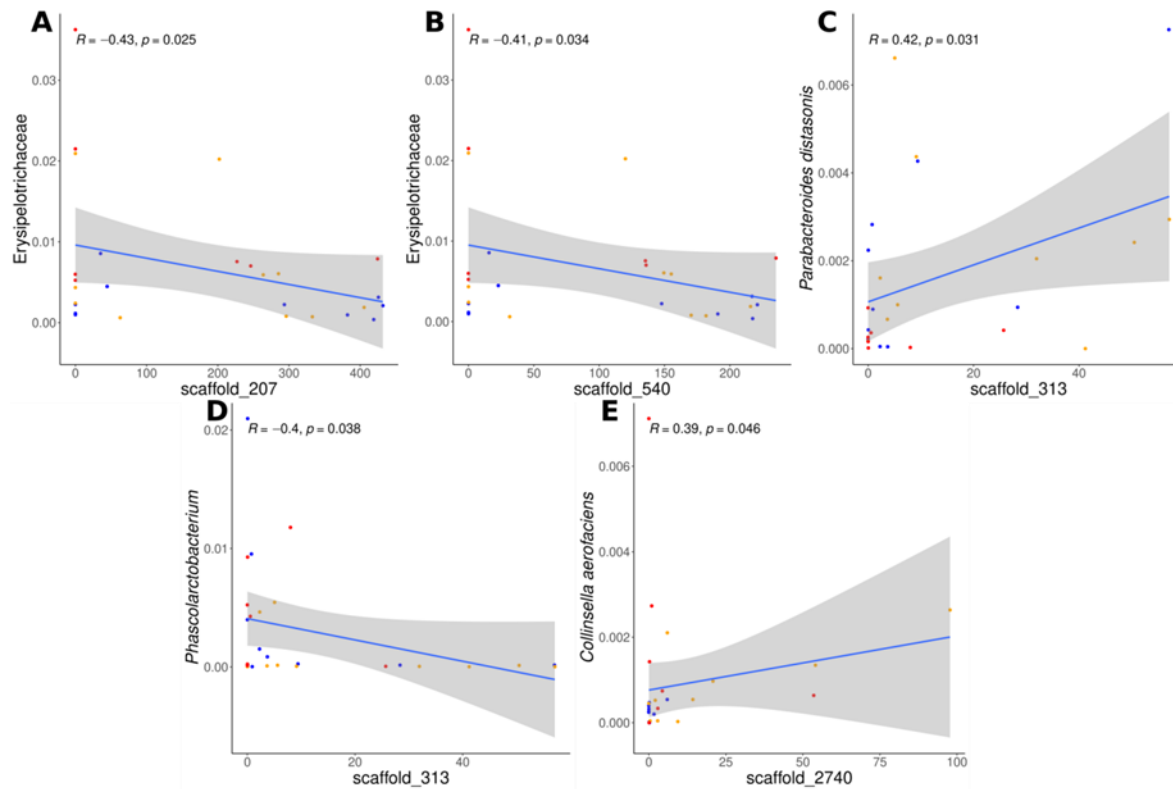


**Anexo 8. Gráficas de correlación de Spearman.** Correlación de Spearman entre los taxones bacterianos alterados en la obesidad y en el síndrome metabólico y la abundancia de los 48 contigs de fagos sobreabundantes y compartidos por las muestras de O y OMS. Círculos azules: muestras NW; círculos naranjas: muestras O; y círculos rojos: muestras OMS.

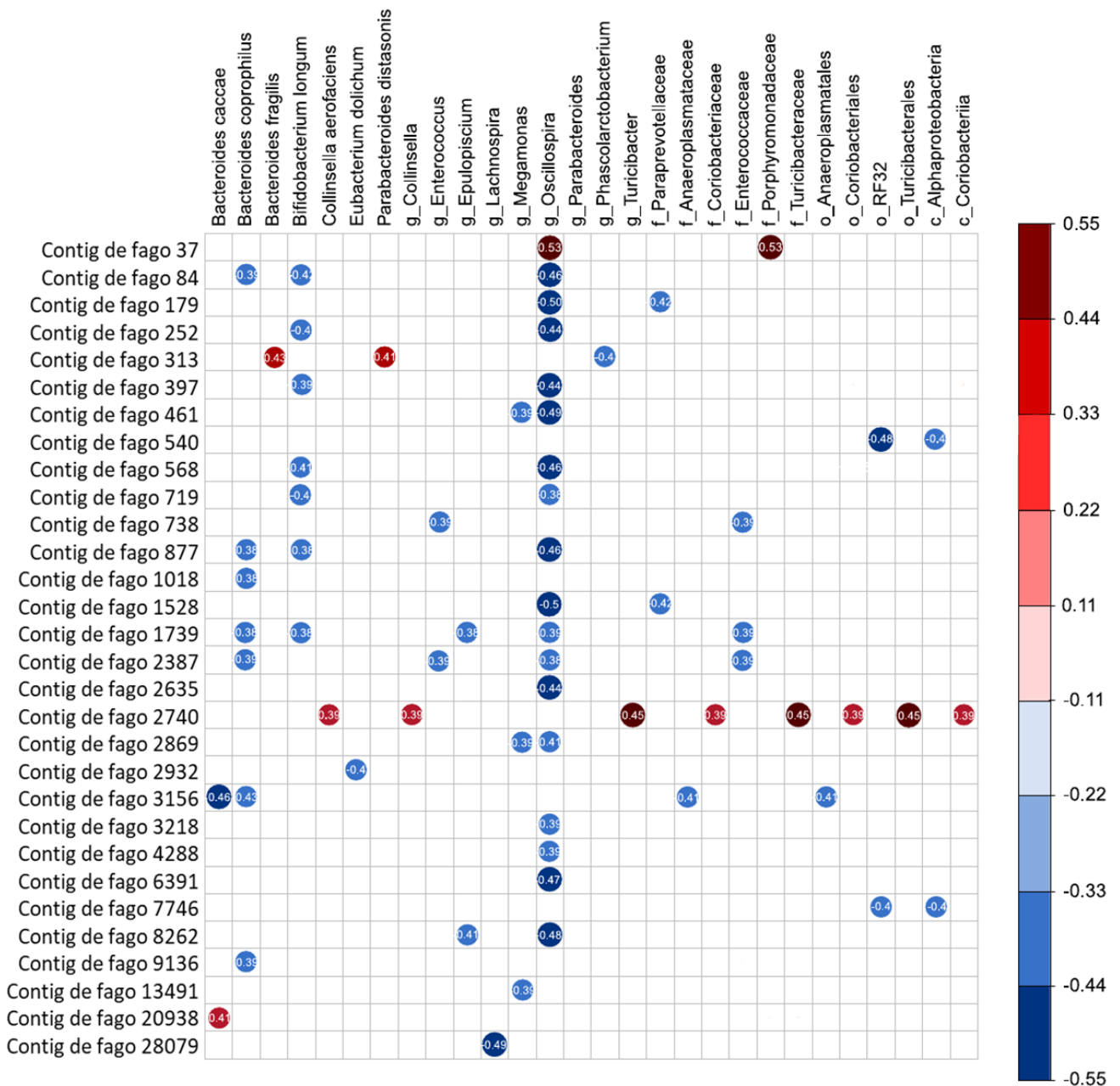




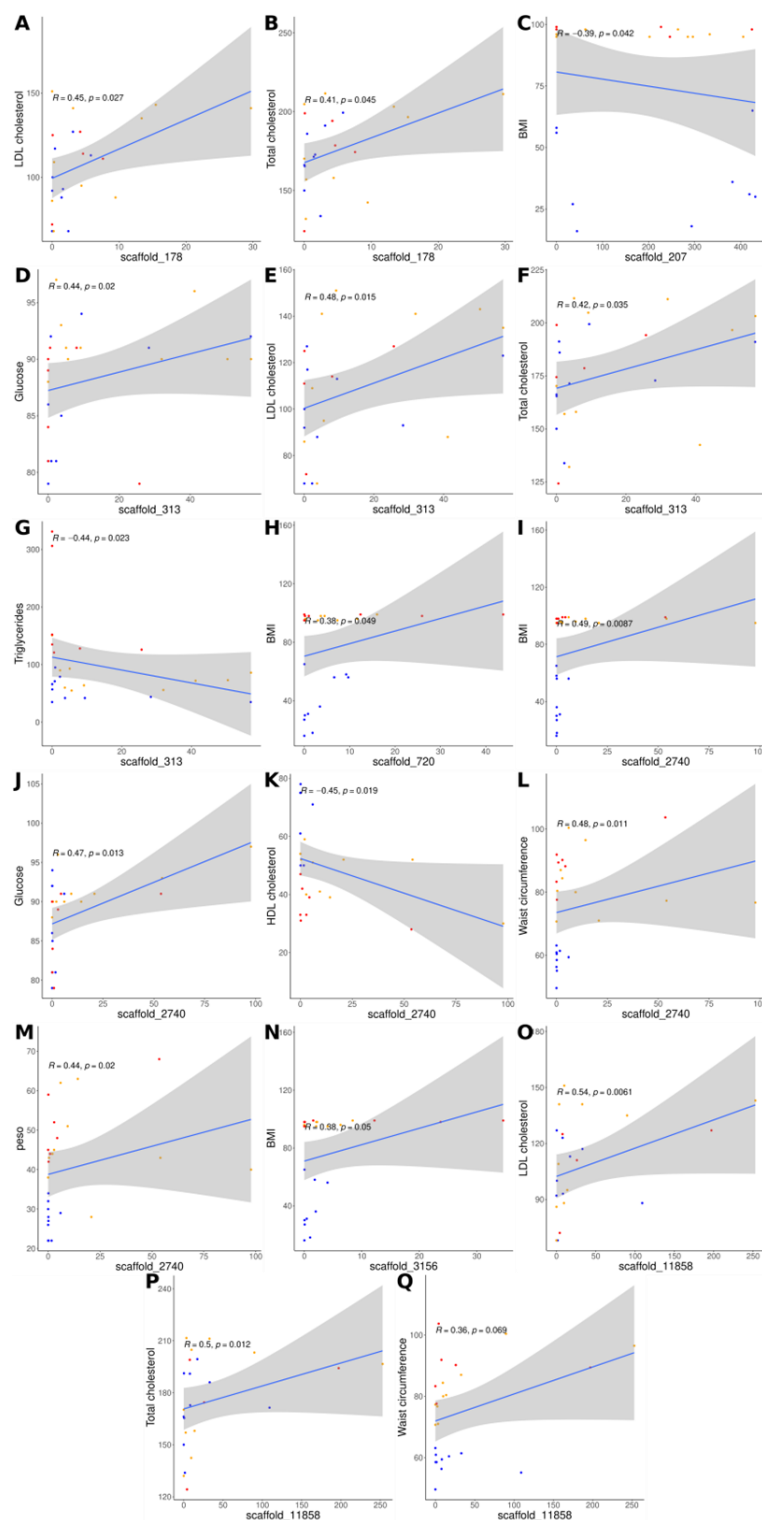
**Anexo 9. Regresión lineal y correlación de Spearman de los contigs de fagos que correlacionaron significativamente con taxones bacterianos específicos de la enfermedad.** El eje Y muestra la abundancia del taxón microbiano, y el eje X muestra la abundancia en RPKM del contig de fago. Círculos azules: muestras NW, círculos naranjas: muestras O; y círculos rojos: muestras OMS.



**Anexo 10. Gráfica de correlación “contigs de fagos core”-bacteria.** Correlaciones de Spearman de las abundancias de los 48 “contigs de fagos core” con la prevalencia más alta ( $\geq 80\%$  de todas las muestras) y la abundancia relativa de todos los taxones bacterianos 16S identificados en las muestras. Solo las correlaciones significativas son mostradas (valor  $p \leq 0.05$ ). La saturación de colores muestra que la correlación numérica del valor de  $p$  fue  $< 0.05$  en todos los casos.



**Anexo 11. Regresión lineal y correlación de Spearman de los contigs de fagos que significativamente correlacionaron con parámetros bioquímicos y antropométricos específicos.** El eje Y muestra los valores de los parámetros bioquímicos y antropométricos, y el eje X muestra la abundancia en RPKM de cada contig de fago. Círculos azules: muestras NW; círculos naranjas: muestras O; y círculos rojos: muestras OMS. BMI= IMC; waist circumference= Circunferencia de cintura.





## 19. Artículos derivados de la presente tesis:

1. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. Bikel S, Valdez-Lara A, Cornejo-Granados F, Rico K, Canizales-Quinteros S, Soberón X, Del Pozo-Yauner L, Ochoa-Leyva A. *Comput Struct Biotechnol J*. 2015 Jun 9;13:390-401. doi: 10.1016/j.csbj.2015.06.001. eCollection 2015.
2. Metatranscriptomic analysis to define the Secrebiome, and 16S rRNA profiling of the gut microbiome in obesity and metabolic syndrome of Mexican children. Gallardo-Becerra L, Cornejo-Granados F, García-López R, Valdez-Lara A, Bikel S, Canizales-Quinteros S, López-Contreras BE, Mendoza-Vargas A, Nielsen H, Ochoa-Leyva A. *Microb Cell Fact*. 2020 Mar 6;19(1):61. doi: 10.1186/s12934-020-01319-y.
3. Gut dsDNA Virome Shows Diversity and Richness Alterations Associated to Childhood Obesity and Metabolic Syndrome. Shirley Bikel, Gamaliel López-Leal, Fernanda Cornejo-Granados, Luigui Gallardo-Becerra, Rodrigo García-López, Filiberto Sánchez, Edgar Equihua-Medina, Juan Pablo Ochoa-Romo, Blanca Estela López-Contreras, Samuel Canizales-Quinteros, Abigail Hernández-Reyna, Alfredo Mendoza-Vargas, Adrian Ochoa-Leyva. *iScience* Available online 24 July 2021, 102900 In Press, Journal Pre-proof.