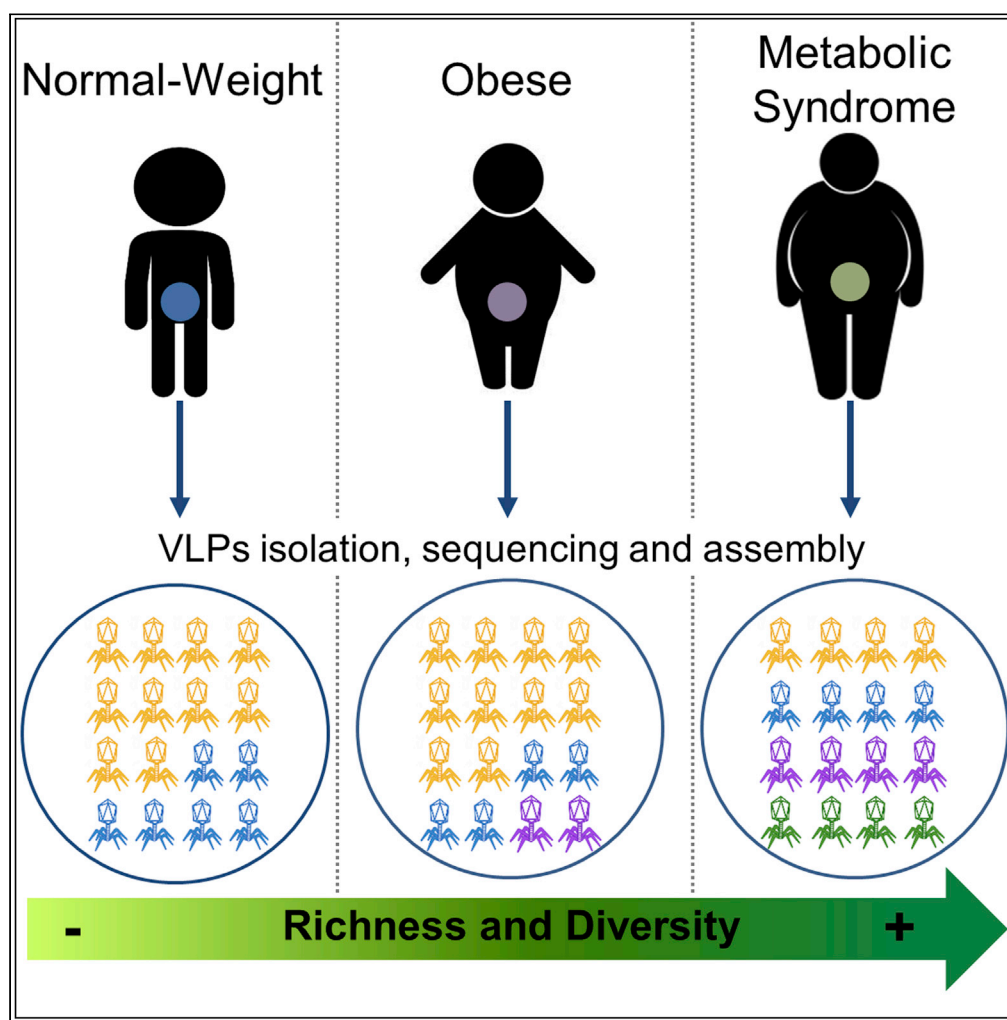


Article

Gut dsDNA virome shows diversity and richness alterations associated with childhood obesity and metabolic syndrome



Shirley Bikel,
Gamaliel López-
Leal, Fernanda
Cornejo-
Granados, ...,
Abigail
Hernández-Reyna,
Alfredo Mendoza-
Vargas, Adrian
Ochoa-Leyva

adrian.ochoa@ibt.unam.mx

Highlights

The VLP abundance remained the same between normal weight and obesity groups

Increased phage richness and diversity is linked to the disease

Bacteriophages dominated the gut virome with high inter-individual diversity

Differential phage abundances and prevalences are associated with the disease

Bikel et al., iScience 24,
102900
August 20, 2021 © 2021 The
Author(s).
[https://doi.org/10.1016/
j.isci.2021.102900](https://doi.org/10.1016/j.isci.2021.102900)

Article

Gut dsDNA virome shows diversity and richness alterations associated with childhood obesity and metabolic syndrome

Shirley Bikel,^{1,4} Gamaliel López-Leal,^{1,4} Fernanda Cornejo-Granados,¹ Luigui Gallardo-Becerra,¹ Rodrigo García-López,¹ Filiberto Sánchez,¹ Edgar Equihua-Medina,¹ Juan Pablo Ochoa-Romo,¹ Blanca Estela López-Contreras,² Samuel Canizales-Quinteros,² Abigail Hernández-Reyna,¹ Alfredo Mendoza-Vargas,³ and Adrian Ochoa-Leyva^{1,5,*}

SUMMARY

Changes in the human gut microbiome are associated with obesity and metabolic syndrome, but the role of the gut virome in both diseases remains largely unknown. We characterized the gut dsDNA virome of 28 school-aged children with healthy normal-weight (NW, n = 10), obesity (O, n = 10), and obesity with metabolic syndrome (OMS, n = 8), using metagenomic sequencing of virus-like particles (VLPs) from fecal samples. The virome classification confirmed the bacteriophages' dominance, mainly composed of Caudovirales. Notably, phage richness and diversity of individuals with O and OMS tended to increase, while the VLP abundance remained the same among all groups. Of the 4,611 phage contigs composing the phageome, 48 contigs were highly prevalent in $\geq 80\%$ of individuals, suggesting high inter-individual phage diversity. The abundance of several contigs correlated with gut bacterial taxa; and with anthropometric and biochemical parameters altered in O and OMS. To our knowledge, this gut phageome represents one of the largest datasets and suggests disease-specific phage alterations.

INTRODUCTION

Childhood obesity (O) is one of the most relevant and severe health problems worldwide; it is a significant risk factor for severe infections, diabetes, and cardiovascular problems later in life. It is the leading cause of adult O, representing a substantial risk for premature death (Biro and Wien, 2010). It is considered a complex disease characterized by abnormal fat accumulation due to an imbalance between energy intake and expenditure (Romieu et al., 2017) that involves genetic, environmental, and lifestyle factors. In Mexico, 17.5% of school-aged children suffer from O (Romero-Martínez et al., 2019), placing it as the country with the world's second highest O rate (WHO, 2017). O is more than an accumulation of fat tissue, it involves chronic low-grade inflammation (Mattos et al., 2016), and it may be associated with metabolic disorders. These include high levels of blood glucose during fasting (hyperglycemia), elevated triglycerides (hypertriglyceridemia), low levels of the beneficial high-density lipoproteins (dyslipidemia), and high blood pressure (hypertension) (Evia-Viscarra et al., 2013). The metabolic syndrome diagnosis requires at least three of these conditions (De Ferranti et al., 2004), unfortunately highly prevalent in children with O (Perichart-Perera et al., 2007; Elizondo-Montemayor et al., 2013). Mexican children are considered a high-risk group (Romero-Martínez et al., 2019) for metabolic syndrome. Its prevalence has to lead to an increased incidence of type 2 diabetes mellitus (T2D) (DiBonaventura et al., 2018) and the development of cardiovascular disease (Franks et al., 2010) in adults.

The human gut microbiome is composed of a vast diversity of bacteria, archaea, and eukaryotic cells that together with viruses (mainly bacteriophages) comprise a diverse and complex ecosystem (Shkoporov and Hill, 2019). Some alterations in the gut microbiota are associated with an increased energy harvest from diet, low-grade inflammation, and altered adipose tissue composition (Pihl et al., 2016). These processes are considered the link between gut microbiota, O, and metabolic syndrome (Bouter et al., 2017). Several studies demonstrated microbiota alterations in O and obesity with metabolic syndrome (OMS) using 16S profiling (Bai et al., 2019; Chen et al., 2020; Gallardo-Becerra et al., 2020; Kim et al., 2020) and

¹Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Cuernavaca, Morelos 62210, México

²Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, UNAM/Instituto Nacional de Medicina Genómica (INMEGEN), México City, México

³Instituto Nacional de Medicina Genómica (INMEGEN), México City, México

⁴These authors contributed equally

⁵Lead contact

*Correspondence: adrian.ochoa@ibt.unam.mx
<https://doi.org/10.1016/j.isci.2021.102900>



metatranscriptomic approaches (Gallardo-Becerra et al., 2020). It has been observed that dietary intervention was associated with changes in the virome community, in which individuals on the same diet converged (Minot et al., 2011). Changes in the virome structure due to a high-fat diet have been observed in mice, suggesting phage-host connections due to the microbial changes induced by diet (Kim and Bae, 2016).

Viral metagenomics is a relatively new and growing research field that studies the complete collection of viruses as part of the microbiota in any given niche (García-López et al., 2019). The gut virome is mainly dominated by bacteriophages (Shkoporov et al., 2019). It regulates the microbial ecosystem and host physiology (Virgin, 2014) through multiple interactions and the co-evolution with the host immune system (Barr et al., 2013), the bacteriome, and horizontal gene transfer events (Maiques et al., 2006). The viromic studies commonly use whole genome amplification (WGA) (Edwards and Rohwer, 2005; Brum and Sullivan, 2015) techniques to obtain a sufficient DNA amount to be sequenced. However, the multiple displacement amplification (MDA) has been associated with quantitative biases (Abulencia et al., 2006; Zhang et al., 2006; Yilmaz et al., 2010) and the preferential amplification of ssDNA viruses (Kim et al., 2008). The MDA-associated artifacts skew the community's taxonomic representation in non-repeatable ways and preclude quantitative analysis of viromes (Yilmaz et al., 2010). To overcome these biases, we used the tagmentation (TAG) method that supports a quantitative method for dsDNA with an ultra-low DNA input (Duhaima et al., 2012). Although the TAG method strongly selects against ssDNA templates (Edwards and Rohwer, 2005; Brum and Sullivan, 2015), 95% of the known human gut phageome is non-enveloped tailed dsDNA phages (Ofir and Sorek, 2018; Sausset et al., 2020).

Human disease-specific changes of the gut virome and phageome (the bacteriophage component of the virome) have been mainly reported in inflammatory bowel disease (Norman et al., 2015; Zuo et al., 2019), AIDS (Monaco et al., 2016), diabetes (Ma et al., 2018), and malnutrition (Reyes et al., 2015). However, studies addressing the role of the virome in O and metabolic syndrome have been limited to animal models (Kim and Bae, 2016; Rasmussen et al., 2020; Schulfer et al., 2020) and human adults (Minot et al., 2011; Manrique et al., 2021), overlooking children cohorts. Here, we characterized the dsDNA virome in ten healthy normal weight (NW) children, ten children with O, and eight children with OMS using metagenomic sequencing of virus-like particles (VLPs) from fecal samples. Our data show that alterations in the gut phageome are present in both O and OMS groups, providing the basis for diagnostic and therapeutic strategies based on phages for managing and preventing these conditions.

RESULTS

The number of VLPs is similar between normal weight, obesity and obesity with metabolic syndrome groups

We used 28 fecal samples previously collected from 7- to 10-year-old 10 healthy NW children, 10 children with O, and eight children with OMS, paired by gender and age (Table S1) and similar middle socioeconomic background (Gallardo-Becerra et al., 2020) (Table S2). The epifluorescence microscopy (Figure 1A) and the transmission electron microscopy (TEM) (Figure 1B) suggested the presence of VLPs in all samples. There was no significant difference in the number of VLPs among the three groups, obtaining an average of 2.56×10^9 , 2.85×10^9 , and 2.70×10^9 VLPs for NW, O, and OMS groups per gram of feces, respectively (Figure 1C and Table S3).

Functional profiles of the reads suggest viral presence in VLP samples

The DNA extracted from VLPs was treated with a TAG method for sequence library construction to avoid the bias of the WGA methods typically used due to the lower amount of DNA. After quality sequencing filters, we obtained an average of 4,871,075 paired-end sequences per sample, producing 11.23 Gb of data (Table S4). To only consider sequences potentially derived from VLPs, the reads mapped to bacterial (~28%) and human (~15%) genomes were discarded from further analyses (Table S4). The removal of bacterial sequences may lead to ignoring potential viral reads from prophages; nonetheless, we preferred to eliminate all potential bacterial DNA for this study. After that, we obtained 74,859,356 reads, an average of 2,673,548 reads per sample, with no significant differences in the sequence depth among the three groups (Figure S1 and Table S4). To obtain a first approach to the functional content of the VLP-derived reads, we annotated the reads using the KEGG database. As previously reported in virome studies (Breitbart et al., 2008; Reyes et al., 2010; Minot et al., 2011; Moreno-Gallego et al., 2019), most reads ($96.2 \pm 1.89\%$) mapped to genes with unknown functions (Figure S2). We also performed a search of the VLP-derived reads against the Prokaryotic Virus Orthologous Groups (pVOGs) and found that a small number of reads ($1.93 \pm 0.62\%$) matched with pVOGs (Figure S3), suggesting that sequences are of unknown viral origin.

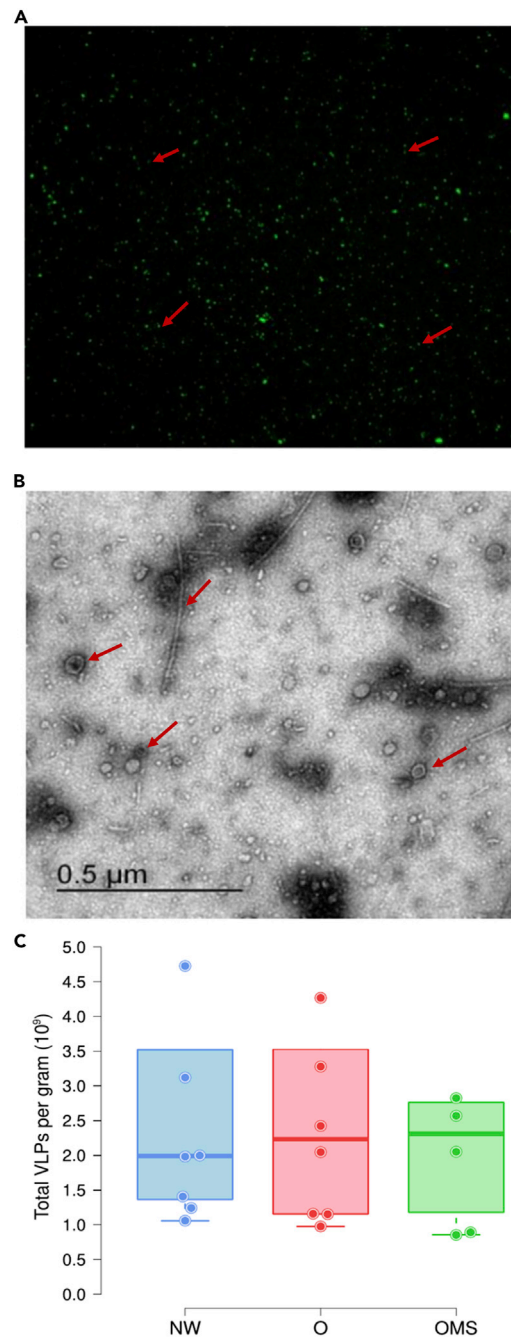


Figure 1. Microscopy visualization and counting of VLPs

(A) SYBR Green I-stained virus-like particles (VLPs) assessed by epifluorescence microscopy. Red rows show an example of the VLPs.

(B) TEM microscopy of VLPs. Red rows show an example of the VLP morphology.

(C) The number of VLPs per gram of fecal matter for each group. Points represent the average number of VLPs for each sample. Error bars indicate the median and interquartile range, and differences were not significant (See [Table S3](#)).

VLP-derived reads show an increased richness in the disease

We conducted 1,000 exercises of 149,000 randomly subsampled viral reads and clustered them at 95% identity to generate unique sequence clusters for each sample. In this way, we could analyze the sequence richness independent of the taxonomic classification and at the same sequence depth for all the samples.

After that, we found an increase of unique sequences in OMS and O groups compared to the NW group (Figure S4), although this was not significant. This result suggests that the obesity groups (O and OMS) had increased viral reads richness compared to the NW.

Bacteriophages dominated the viral reads of the gut virome

After clustering all the viral reads at 95% identity for each sample, there was a reduction of $68 \pm 8\%$ of sequences, resulting in an average of 856,825 unique sequences per sample (Table S4). As a first approach to obtain the potential viral content from the VLP-derived reads, we matched them against the viral non-redundant (NR) RefSeq and found that only $2.95 \pm 0.95\%$ had a match (Figure S5A). From these, $66.95 \pm 6.95\%$ of the reads matched to prokaryotic viruses, $6.79 \pm 3.91\%$ to eukaryotic viruses, and $26.26 \pm 5.79\%$ to an undefined classification including unknown viruses or multiple hits between eukaryotic and prokaryotic viruses (Figure S6).

Virome assembly confirms the dominance of bacteriophages

Considering that the length of a phage genomes could affect the abundance obtained from the sequencing reads (Moreno-Gallego et al., 2019), we performed a *de novo* assembly using the 74,859,356 viral reads from all samples. To avoid chimeric contigs, we collected those contigs covering $\geq 80\%$ of their total size by the viral reads in at least one sample, resulting in 18,602 contigs (≥ 500 nt; largest = 176,210 nt; N50 = 7,480 nt, Table S5). On average, $58.69 \pm 12.79\%$ of the viral reads from each sample mapped back to these contigs, showing a homogeneous contribution of all samples to the assembly (Figure S7 and Table S6). Next, we removed all short contigs (<4 kb) to remove potential fragmented viral genomes. After that, 12,287 contigs (N50 = 9,097 nt, Table S5) were obtained and used for further analyses. Importantly, this reduction in the number of contigs did not lead to a drastic decrease in the read recruitment, remaining on average $54.85 \pm 14.60\%$ of the viral reads from each sample (Figure S7 and Table S6). We found, on average, 7.20 genes per contig (0.87 genes per kb of contig length). We classified these contigs using the DNA and their encoding proteins, obtaining 4,611 contigs as potential prokaryotic viruses, 1,540 as potential eukaryotic viruses, 2,696 contigs with multiple hits to prokaryotic and eukaryotic viruses, and 3,440 contigs with an unknown origin. From the 4,611 prokaryotic contigs (≥ 4 Kb) representing complete or partial genomes, 1,307 were classified using the DNA sequences (dc_megablast) and 3,304 using their encoded proteins (BLASTx). On the other hand, we classified the 12,287 contigs using VirSorter, obtaining a viral classification for 1,542 of them. Contrary, 5,949 contigs had a match with a pVOG. Interestingly, most classified contigs using VirSorter (69.58%) and pVOG (77.16%) coincide with our prokaryotic virus classification (Figure S8), reinforcing our phage classification strategy using both DNA and proteins. Previous virome studies reported that between 29.35% and 48.5% of the viral contigs mapped to a pVOG (Coutinho et al., 2019; Deboutte et al., 2020).

On average, the contig classification revealed $56.30 \pm 5.50\%$ of prokaryotic viruses and $13.94 \pm 2.80\%$ of eukaryotic viruses per sample. These were similar to the taxonomy classification obtained from reads (Figure S6), suggesting that the sequence diversity observed with reads was also captured in the assembled contigs. As expected by using a TAG method, most of the prokaryotic contigs were dsDNA viruses (96%), whereas only 0.1% of them were annotated as ssDNA viruses, and the remaining 4% were unclassified bacterial viruses. We did not observe any influence in the number of reads versus the number of viral contigs obtained per sample (Figure S9).

The children's gut phageome was mainly composed of Caudovirales

The 4,611 contigs (N50 = 10,370 nt, max 176,210 nt; mean 9,347 nt, Table S5) classified as potential prokaryotic viruses were selected as the phageome and represented 37.53% of the total contigs of the virome assembly, hereafter mentioned as phage contigs. The phageome recruited more than half of the viral reads (50.27%) mapped to the original virome assembly and were distributed in an average of 707,571 reads per sample (Table S6).

Caudovirales composed the majority of phage contigs ($91.28 \pm 0.10\%$), followed by crAss-like viruses ($0.64 \pm 0.73\%$) (Figure S10). The high number of contigs classified as Caudovirales coincides with the high number of viral reads also classified as Caudovirales (Figure S5B), suggesting that our genome assembly also reflects the taxonomy of reads by themselves. Within Caudovirales contigs, the most abundant families were Siphoviridae ($35.28 \pm 0.02\%$), Myoviridae ($31.33 \pm 0.03\%$), and Podoviridae ($6.50 \pm 0.01\%$).

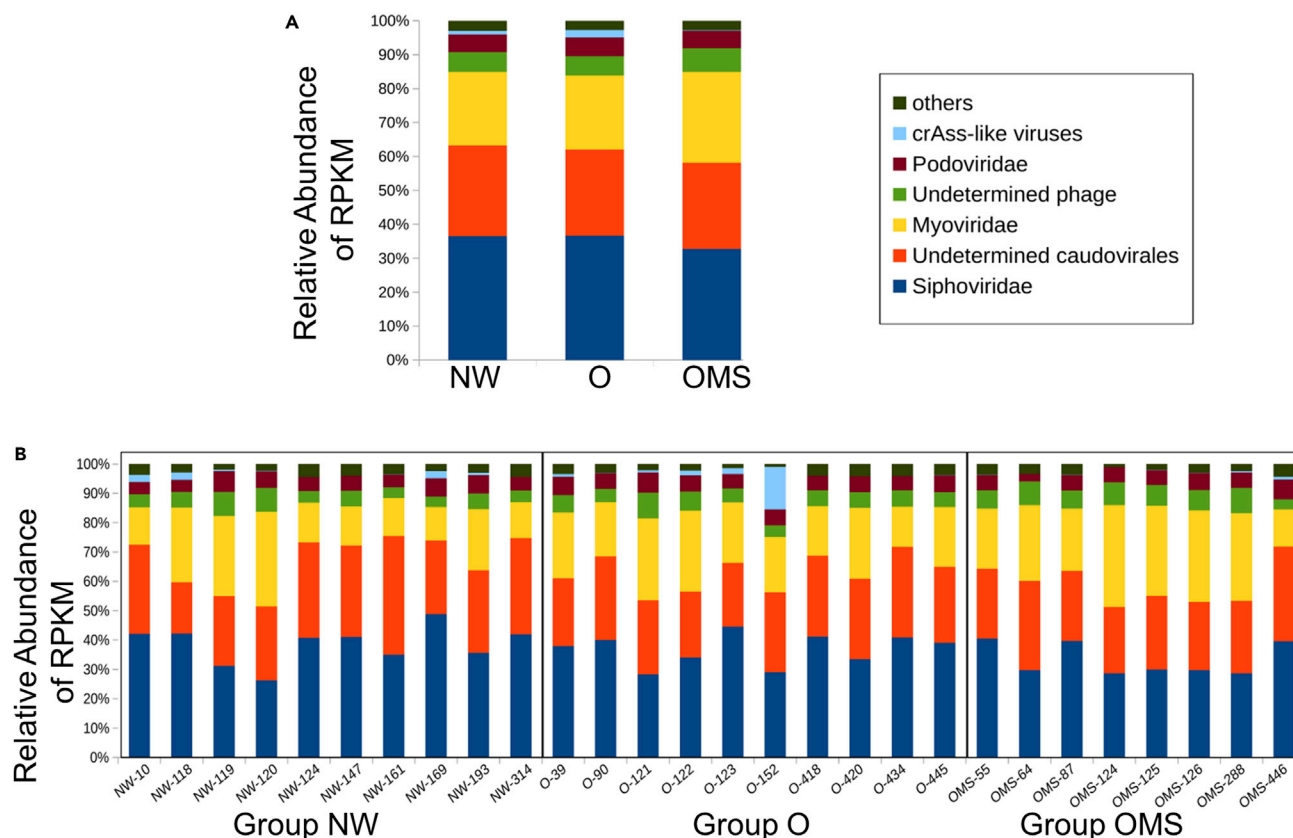


Figure 2. Relative abundance of normalized RPKM for the 4,611 phage contigs

(A and B) (A) Average per group and (B) average per sample. “Others” represents the relative abundance of the less abundant phage contigs.

Given the importance of crAss-like phages in adult viromes (Dutilh et al., 2014; Shkoporov et al., 2018b), we analyzed their presence in our phageome, and we found that an average of $0.64 \pm 0.01\%$ of the contigs could be classified as crAss-like phages (Table S7). On the other hand, we found that the Mexican crAss-phage (Cervantes-Echeverría et al., 2018) was present in 25 out of 28 samples.

Increased richness, diversity, and dominance of phage contigs is linked to the shift from normal weight to obese

We normalized abundance by reads per kilobase per million (RPKM)-sequenced reads per sample to compare groups’ phage abundance and diversity metrics (Reyes et al., 2015). After that, we compared the taxonomy abundances of classified phages among our groups (Figure 2). We found a decreased abundance of Siphoviridae and crAss-like phages (32.72 and 0.18%, respectively) in the OMS group as compared to O (36.66 and 2.2%, respectively) and NW (36.50 and 1.0%, respectively) groups. We also observed an increased abundance for Myoviridae in the OMS group (26.72%) as compared to O (21.77%) and NW (21.70%) groups. However, these abundances were not significantly different among the three groups.

The α -diversities of the phageome showed that phage richness and Shannon diversity increased in O and OMS groups compared to the NW group (Figures 3A and 3B), although the differences among groups were not significant. Notably, these changes in diversity happened maintaining a similar number of VLP counts in the three groups (Figure 1C). The O group exhibited the highest richness, followed by OMS and NW groups (Figure 3B), whereas the OMS group exhibited the highest diversity, followed by O and NW groups (Figure 3A). Notably, this increased richness in obese groups was also supported by our initial viral reads clustering analysis (Figure S4). We also observed that 488 (10.58%) phage contigs accounted for 70% of the normalized reads in the NW samples, whereas 679 (14.73%) and 831 (18.02%) phage contigs accounted the

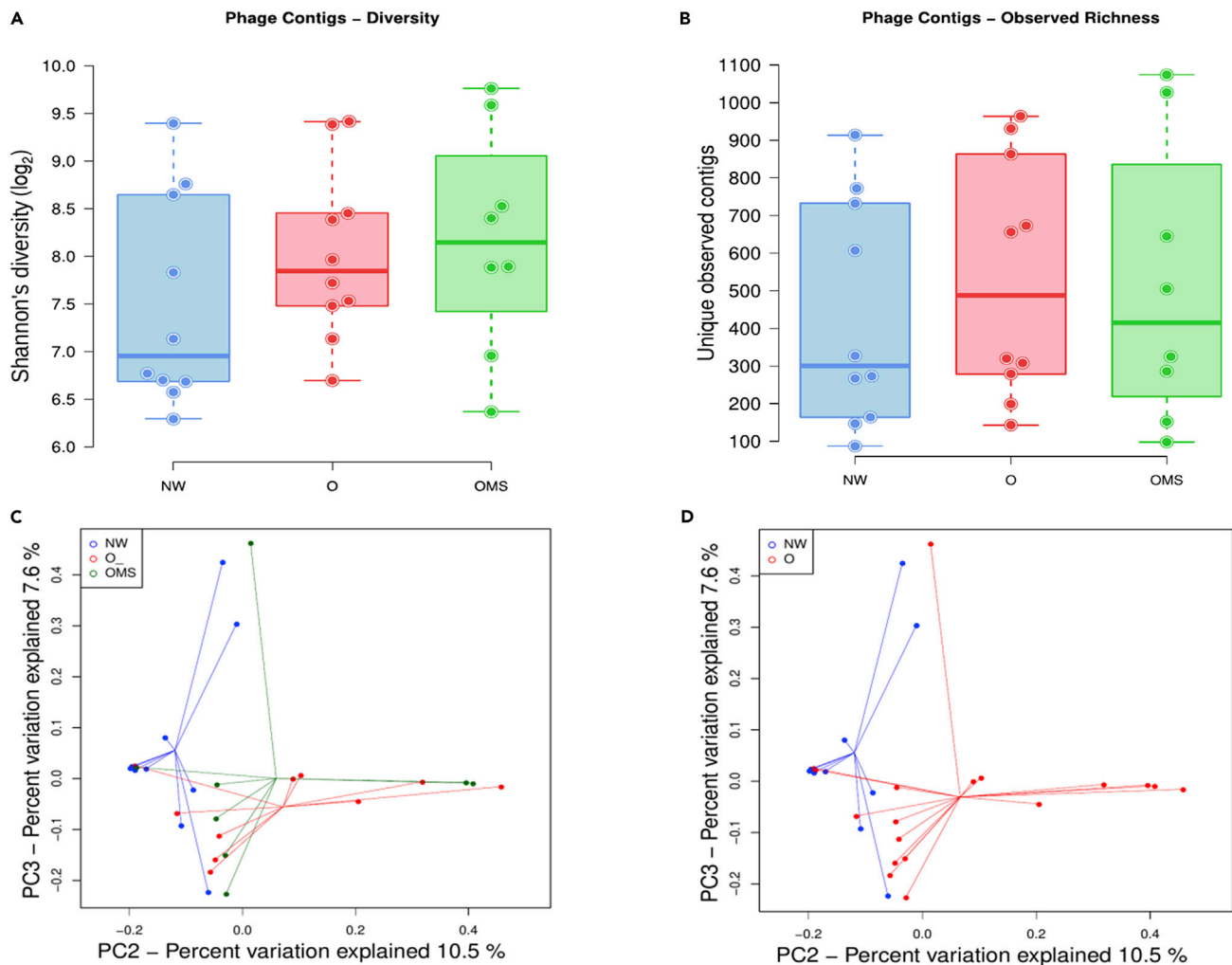


Figure 3. Alpha and beta diversity of the phageome

(A) Phage contigs diversity. Each point shows the median of 10,000 Shannon diversity calculations at an even depth for a single sample based on even resampling, with boxes showing the group's distribution group.

(B) Phage contigs richness. Each point shows the median of 10,000 observed contigs calculations at an even depth for a single sample based on even resampling, with boxes showing the group's distribution. Differences in (A) and (B) were not significant. Error bars represent the mean \pm SD.

(C) Principal coordinates analysis (PCoA) based on Bray-Curtis dissimilarity with samples tagged as NW, O, and OMS.

(D) PCoA based on Bray-Curtis dissimilarity with samples tagged by all obese (O + OMS) in red circles and NW (See [Figures S11](#) and [S12](#)).

70% of the normalized reads in O and OMS groups. These results suggest a considerable increase in the number of dominant phage contigs due to the disease.

On the other hand, the principal coordinates analysis based on Bray-Curtis distances showed no self-consistent clusters by the study group ([Figures 3C](#) and [S11](#)). However, when all the obese samples were tagged together (O + OMS), they cluster separately from NW samples ([Figures 3D](#) and [S11](#)). Although, this difference was not significant ($p = 0.568$). We also performed the Aitchison distance analysis subjected to dimensional reduction with a principal components analysis and showed non-significant separated clusters ([Figure S12](#)).

Several phage contigs were significantly over-abundant in obese and metabolic syndrome groups

We identified the significantly over-abundant phage contigs in O and OMS groups compared to the NW group using all the phages in the normalized abundance matrix. This procedure is analogous to using tables

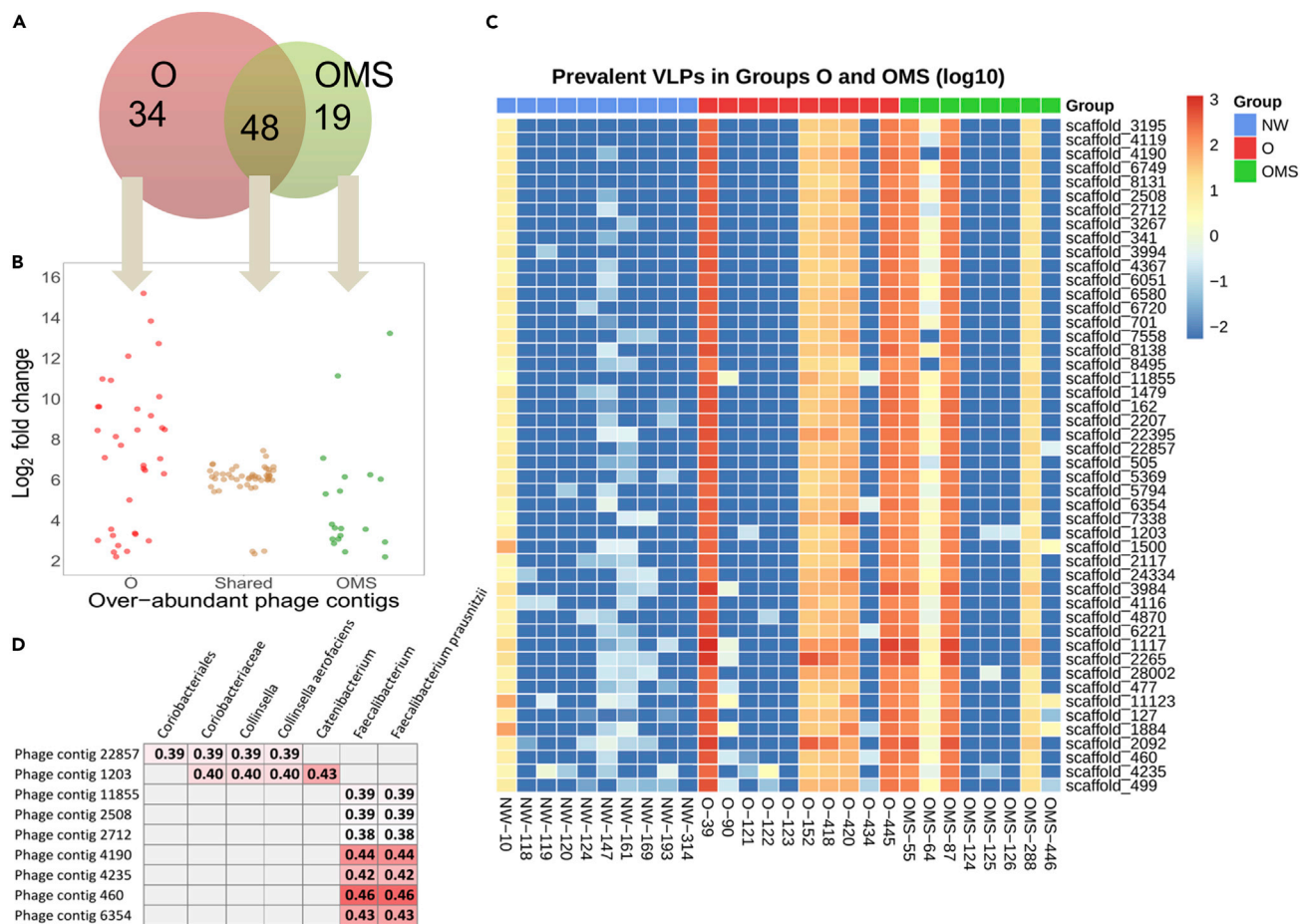


Figure 4. Analysis of over-abundant phage contigs in the obesity (O) and obesity with metabolic syndrome (OMS)

(A) Venn diagram of over-abundant phage contigs in the O (red circle) and OMS (green circle) groups, as compared with the NW group. (B) Expression levels (>2 of log₂ fold change) of over-abundant phage contigs from the Venn diagram. The 34 (red points), 19 (green points), and 48 (brown points) phage contigs of O, OMS, and shared among two groups, respectively. (C) Heatmap of the normalized RPKM abundances of the 48 phage contigs over-abundant in O and OMS groups. We show the abundance distribution for each contig among all samples. (D) Spearman correlation plots of the phage abundances (RPKM) of the 48 phage contigs over-abundant in the disease and the relative abundance of 16S bacterial taxa identified to be significantly associated with obesity. Only significant correlations with unadjusted p values (≤0.05) were displayed. See also [Figure S13](#).

of RPKM for the measurement of differentially expressed genes in RNA-seq experiments. After that, we obtained 111 and 107 phage contigs significantly over-abundant in O and OMS groups, respectively, and the differences were statistically significant at an alpha = 0.05 compared to the NW group. From those, we only selected the phage contigs shared in at least 30% of the O or OMS groups to eliminate the individual phages. This resulted in 82 and 67 phage contigs significantly over-abundant in O and OMS groups, respectively, compared to the NW group, with 48 phage contigs shared between O and OMS groups (Figures 4A and 4B).

We next assessed whether the abundance of the 48 phage contigs (Figure 4C) was associated with a parallel change in bacterial populations on all the samples. To answer this question, we selected the 16S rRNA gene sequencing data of 41 bacterial taxa significantly associated with O and metabolic syndrome (Gallardo-Becerra et al., 2020) and calculated the Spearman correlation between them and the abundance of the 48 phage contigs in all the samples. We found that the abundance of 9 phage contigs correlated with bacterial abundances associated with O and metabolic syndrome (Figure 4D). The samples that led to the correlation between phage contigs and bacteria were the O and OMS groups (Figure S13). We conducted the same correlation analysis for the 34 and 19 over-abundant phages of OMS and O groups (Figure 4A). However, the correlations were mainly caused by outlier samples and zero abundance values for most phage contigs

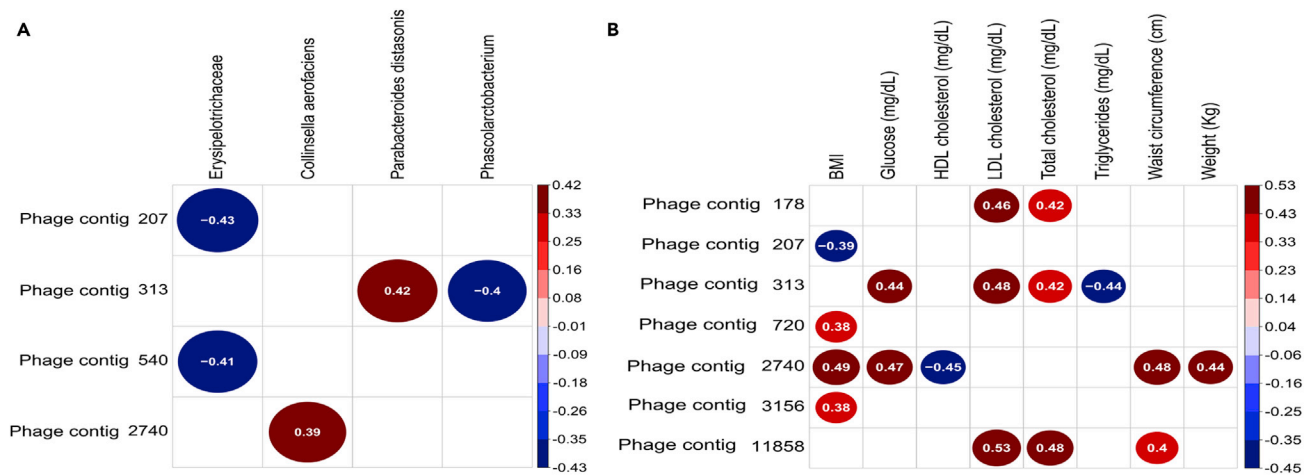


Figure 6. Disease-specific bacteria-phageome patterns in obesity and metabolic syndrome

(A) Spearman correlation plots of the phage abundances (RPKM) of the 48 phage contigs with a higher prevalence ($\geq 80\%$ of all the samples) and the relative abundance of 16S bacterial taxa identified to be significantly associated with the disease.

(B) Spearman correlation plots of the phage abundances (RPKM) of the 48 phage contigs with a higher prevalence ($\geq 80\%$ of all the samples) and the clinical and anthropometrical parameters altered in obesity and metabolic syndrome. Only significant correlations with unadjusted p values ≤ 0.05 were displayed. See also [Figures S15](#) and [S18](#).

The abundance of “core phages” correlated with bacterial taxa associated with obesity and metabolic syndrome

We assessed whether the 48 core phage contigs were associated with parallel changes in bacterial populations. To this end, we calculated the Spearman correlation between the 16s rRNA gene sequencing data of 41 bacterial taxa and the abundance of 48 core phage contigs in all the samples. After that, we only selected the Spearman correlations with $r^2 > 0.3$ and unadjusted p value ≤ 0.05 .

The abundance of four phage contigs was correlated with bacterial abundances ([Figures 6A](#) and [S15](#)). Phage contig 2740, which was more prevalent in O than NW group ([Figure S16A](#)), was positively correlated with the abundance of *Collinsella aerofaciens* ([Figures 6A](#) and [S15](#)), a prevalent bacteria in the OMS group. Phage contig 313, more abundant in O vs. OMS group ([Figure S16A](#)), showed a positive correlation with *Parabacteroides distasonis*, also prevalent in O compared to OMS group, and a negative correlation with an undetermined species from genus *Phascolarctobacterium* (more abundant in NW vs. O) ([Figures 6A](#) and [S15](#)). Furthermore, the phage contig 313 showed high similarity (99.4% nt identity) with previously reported *Bacteroides* plasmids (sequence ID: CP059857.1 and AP019726.1). Phage contigs 207 and 540 had a lower prevalence in OMS and O groups than the NW group ([Figure 6A](#)), and both contigs were negatively correlated with *Erysipelotrichaceae*, an over-abundant family in O and OMS groups ([Figures 6A](#) and [S15](#)). These data might imply that some of the changes in the abundances of specific phage-bacteria interactions may be partly associated with the bacterial changes found in O and metabolic syndrome.

We also assessed whether these core phage contigs were associated with a parallel change in all bacterial populations, independently of the bacterial association to O and metabolic syndrome. To this end, we selected all the taxa of the 16s rRNA gene sequencing and calculated the Spearman correlation between them and the abundance of these core phage contigs in all the samples. Only the abundance of 30 phage contigs correlated with bacterial abundances associated with O and metabolic syndrome ([Figure S17](#)).

The abundances of core phage contigs correlated with anthropometric and biochemical parameters altered by obesity and metabolic syndrome

We also analyzed whether the 48 core phage contigs correlated with the anthropometric and clinical parameters typically altered in obesity and metabolic syndrome, such as high body mass index (BMI), low levels of high-density lipoprotein (HDL), high levels of triglycerides, high glucose level, high waist circumference, and high weight ([Gallardo-Becerra et al., 2020](#)). The Spearman correlation between the 48 phage

contigs and the anthropometrical and biochemical parameters, including all the samples of our study, showed positive correlations with several phage contigs (Figures 6B, S18, and S16B), suggesting an association between these phage contigs and anthropometric and biochemical parameters. We also found a negative correlation between the abundance of several phage contigs (Figure S18) and BMI, HDL levels, and triglycerides (Figures 6B and S18).

DISCUSSION

Metagenomic analysis of viruses, one of the most poorly understood components of the human gut microbiome, has recently revolutionized our view of the gut microbiome highlighting the critical role of the interactions between phages and bacteria in health and disease (Breitbart et al., 2003; Bikel et al., 2015; Garmaeva et al., 2019).

Here, we report a large-scale sequencing viral metagenome where instead of solely using the sequencing reads, we assembled the phageome community, obtaining 4,611 phage contigs ≥ 4 Kb representing complete or partial genomes. We used this size threshold to avoid having partial viral genomes and decrease the selection of eukaryotic viruses, mainly composed of the Anelloviridae family, with a reported genome size between 3 and 4 Kb (Reyes et al., 2015) and considering that shortest phage genomes are ≥ 4 Kb (Hatfull, 2008). Most virome studies use MDA phi29 WGA (Sutton and Hill, 2019) which unevenly amplifies linear genome fragments and preferentially amplifies small circular ssDNA viruses, e.g., the Microviridae family (Roux et al., 2016). This methodology probably limits statistical analyses of the viral community composition to presence-absence observation because relative abundances might be biased toward specific viruses. In contrast, we used a TAG method free of WGA. Although this method may now select against ssDNA (Yilmaz et al., 2010; Roux et al., 2016), it allowed us to perform analyses associated with the abundance of dsDNA phages, the major components of the human gut virome (Kim et al., 2011). Probably because of the TAG method, the abundance of Microviridae in our samples was much smaller than that previously found in most virome studies (Norman et al., 2015; Shkoporov et al., 2018b), highlighting the need for quantitative analysis protocols in virome studies, as was recently described (Roux et al., 2016).

The TAG method impacted the proportion of both Microviridae (0.031%) and of Inoviridae (0.137%) to a lesser extent (Table S7) since both families are ssDNA viruses. We do not know why the method affected the abundance of Microviridae more than Inoviridae. However, given that we used the same protocols for all samples, we expected the same bias for both families. The selection against ssDNA templates had been observed before in mock community experiments (Kim and Bae, 2011; Roux et al., 2016), favoring ssDNA viruses to be systematically under-represented (>10 -fold) in TAG and over-represented (<10 -fold) in WGA-obtained viromes (Roux et al., 2016). To avoid this bias, we decided to eliminate all ssDNA viruses from our analysis.

The Caudovirales were the most abundant phage order in our three groups. This predominance is consistent with that observed in previous human gut viromes (Norman et al., 2015; Reyes et al., 2015; Manrique et al., 2016). We showed that children with O and OMS had specific changes in their gut phageomes, particularly in the abundance of specific phage contigs. Importantly, the recruited children had similar lifestyles, the same ethnographic region, and relatively homogeneous environments, making the effects of socioeconomic, cultural, and nationality not confounding factors. We consider this feature to minimize bias in our results, as reported in previous studies (He et al., 2018; Stagaman et al., 2018; Zhong et al., 2019). Notably, we found a similar number of VLPs among the groups, independently of the disease status. Contrary, previous reports on Crohn disease have observed that sick patients harbored significantly more VLPs than healthy individuals (Norman et al., 2015).

We observed an increased diversity and richness in OMS and O groups than NW group, although it was not significant, probably to the limited sample size. This suggests that the expansion of specific phages in O and metabolic syndrome could decrease the presence of others, maintaining similar VLP amounts independently of the disease status. Interestingly, both sequencing reads and metagenomic viral assembly supported the increased richness observed in O and OMS groups. We previously observed a significant increase in bacterial richness and diversity in O and metabolic syndrome groups than the healthy NW group using the same set of samples used for the virome (Gallardo-Becerra et al., 2020). The increase of richness and diversity of phages and the bacteria associated with O and metabolic syndrome agrees with the recent proposal that the virome diversity is associated with adults' intestinal bacterial diversity (Moreno-Gallego et al., 2019).

An increase of viral richness and diversity has also been reported in Crohn disease, ulcerative colitis, and murine colitis (Duerkop et al., 2018). Contrary, low viral richness and diversity were detected in individuals with type 1 diabetes mellitus (Zhao et al., 2017). The phage contigs found at high prevalence in NW individuals were significantly lower in individuals with O and OMS. This would agree with a recent study in ulcerative colitis and Crohn disease that also found a reduction of phage contigs typically abundant in healthy individuals (Manrique et al., 2016). These results suggest that the loss of some phages with high prevalence in the NW group could be associated with O and metabolic syndrome.

We found a higher diversity and richness associated with OMS than in obese individuals, indicating the importance of studying the Phageome in these pathologies as separated diseases. Furthermore, we also found specific over-abundant phages for O group different from OMS group. There is a wide-ranging heterogeneity among individuals with O regarding their risk for developing metabolic dysfunction and its attendant complications (Viveiros and Oudit, 2020). The gut microbiota is involved in the etiology of O and O-related complications such as nonalcoholic fatty liver disease, insulin resistance, and T2D (Pedersen et al., 2016; Canfora et al., 2019), suggesting the importance of studying O separately from metabolic syndrome.

In this regard, recent studies suggest that the microbiota from metabolically healthy individuals with O could transition to OMS (Yuan et al., 2021), and a metatranscriptome and 16S profiling demonstrated significant differences between O and metabolic syndrome (Gallardo-Becerra et al., 2020). The abundance of some metabolism-related bacteria was associated with circulating inflammatory compounds in individuals with O without metabolic syndrome, suggesting that gut microbiota changes in metabolically healthy children with O conceivably serve as a compensatory response to a surfeit of nutrients (Yuan et al., 2021). The 19 phage contigs that were only over-abundant in the OMS group may be further studied as biomarkers linked to the development of metabolic syndrome in children with O. It would be interesting to validate this assumption by analyzing phage contigs at the functional level (Turnbaugh et al., 2009) and analyzing the over-abundant unique phages' functional profile in O and OMS groups.

We detected a high inter-individual variability of the phageome, with 75.41% of phage contigs detected in less than 50% of the individuals. In contrast, only 48 (1.04%) phage contigs were present in >80% of the individuals (core phages) independently of the disease. This observation agrees with previous evidence of a reduced healthy core in the human gut phageome, composed only of 23 phages (Manrique et al., 2016). The concept of a reduced core virome is still controversial. It has received support from a recent study with adult monozygotic twins, in which only 18 contigs were found in all individuals (Moreno-Gallego et al., 2019). In contrast, the compilation of a large-scale gut virome database called into question the existence of a human core gut virome (Gregory et al., 2019; Shkoporov et al., 2019). Our findings on a reduced number of shared phages support the idea of a highly individual-specific gut virome (Shkoporov et al., 2018b; Moreno-Gallego et al., 2019; Shkoporov et al., 2019) and a high inter-individual viral diversity (Reyes et al., 2010; Minot et al., 2013; Shkoporov et al., 2019) because only 1% of our viral assembly (4611 phage contigs) belongs to core phage contigs. This percentage is similar to previously reported core contigs, where 0.6% of the viral assembly (3639 viral clusters) (Shkoporov et al., 2019) and 1.3% of the viral assembly (1703 viral contigs) (Manrique et al., 2016) were shared among the majority of individuals.

According to previous reports, contamination with some bacteria from different extraction kits (including ours) was reported (Laurence et al., 2014; Salter et al., 2014), mostly are soil- or water-dwelling bacteria frequently related with nitrogen fixation, probably associated with nitrogen that is often used instead of air in ultrapure water storage tanks (Kulakov et al., 2002). However, we did not find reports about viral contamination. Unlike high-biomass samples (such as from feces), kit contamination is a particular challenge for low-biomass studies (such as from blood or the lung), which may provide little template DNA to compete with that in the reagents for amplification (Tanner et al., 1998; Willerslev et al., 2004), suggesting that if there was contamination in our samples, it would be minimal and this should not be of significance on our results.

The criteria used to define the presence of a viral sequence in a sample (Gregory et al., 2019) are still questionable. According to the previous reports (He et al., 2018; Zhong et al., 2019) and our results, we suggest that the highly homogeneous geographic and ethnic representation across our dataset of samples was an essential factor that allowed us to establish a human core phageome, being this highly reduced with only 48 phage

contigs. We followed similar sequence thresholds recently proposed for accurate estimation of viral community composition and diversity (Reyes et al., 2015), such as (i) contig length ≥ 4 kb, (ii) coverage determined from reads mapped at $\geq 90\%$ identity, and (iii) $\geq 80\%$ of contig length with $\geq 1\times$ coverage. However, we should note that our assemblies may represent fragments of the same phage genome or family. More extensive studies using stringent standard bioinformatics parameters, such as previously suggested (Roux et al., 2017), are necessary to generate more accurate estimates about the core phageome and its prevalence in the human population. We suggest that the use of only contigs covering $\geq 80\%$ of their total size by the viral reads in at least one sample decreased the rate of reads aligned ($\sim 58\%$) to the initial virome assembly. Removing all shorter contigs (< 4 kb) did not lead to a drastic reduction in reads recruitment, remaining $\sim 54\%$ of the viral reads aligned. Similar reads recruitment percentage (49.66%) was obtained in a recent human gut virome using an extensive post-assembly decontamination process (Shkoporov et al., 2018b). Compelling evidence supports the importance of studying the virome concurrently with the bacteriome to obtain a holistic picture of the gut ecosystem changes in a disease such as inflammatory bowel disease (IBD) (Norman et al., 2015; Cornuault et al., 2018), where phage abundance was associated with changes in the abundance of specific gut bacterial species (Reyes et al., 2013). Also, fecal metabolomics in mice revealed that phage predation in the mouse gut microbiota could potentially impact the mammalian host by changing the levels of key metabolites involved in essential functions for the host (Hsu et al., 2019). According to these, we assessed whether the highly abundant phage contigs were associated with a parallel change in bacterial populations and the clinical parameters significantly altered in O and metabolic syndrome. Although the biochemical and anthropometric parameters reflected the difference between groups according to the disease, this group separation also correlated with several abundances of phages. In this regard, we observed that the abundance of phage contig 2740 positively correlated with *Collinsella aerofaciens*, low-density lipoprotein (LDL), BMI, waist circumference, glucose, weight, and lower HDL levels.

Further, *Collinsella aerofaciens* was significantly over-abundant in the OMS group (Gallardo-Becerra et al., 2020) and showed a positive correlation with triglycerides and a negative correlation with HDL in the same sample set. However, *Collinsella* is a highly abundant taxon in 12-month-old breastfed infants. It was previously reported as a signature of the developing anaerobic infant microbiome and could be involved in acquiring crAss-like phages (Bäckhed et al., 2015; Siranosian et al., 2020). These results suggest that the gut virome is also altered along with the microbiome in children with O. The phage contig 313 positively correlated with *Parabacteroides distasonis*, high LDL, glucose, and total cholesterol levels. This bacteria was enriched in O vs. OMS groups, and it was associated with increased LDL levels in the cohort of our study (Gallardo-Becerra et al., 2020) and weight gain and hyperglycemia in other studies (Wang et al., 2019).

The phage contig 313 showed high similarity with previously reported *Bacteroides* plasmids (sequence ID: CP059857.1 and AP019726.1). Because *Bacteroides* and *Parabacteroides* are closely related bacteria, these results suggest these groups of bacteria as putative hosts. Moreover, phage contig 313 was positively correlated with *Parabacteroides distasonis*. Together, these results suggest that this phage could be opting for a pseudo-lysogenic cycle and multiply with their hosts (Shkoporov et al., 2018a) since the replication strategy of crAss-like bacteriophages is that it co-replicates with its host in a way that does not disrupt the host proliferation (Shkoporov et al., 2018a). On the other hand, the high similarity of contig 313 with previously reported *Bacteroides* plasmids follows this group of bacteria as putative hosts. Unfortunately, little is known about the diversity of plasmids and pseudo-lysogenic bacteriophages in these groups of bacteria. However, more studies are needed to understand better viral replication that would help interpret these relationships. In contrast, the phage contig 313 showed a negative correlation with the genus *Phascolarctobacterium*, enriched in the NW vs. O groups (Gallardo-Becerra et al., 2020). This suggests that the increased abundance of this phage contig in O could be inhibiting the abundance of *Phascolarctobacterium*, potential protective bacteria against O (Gallardo-Becerra et al., 2020). Also, the increased abundance of phage contigs 207 and 540 in the NW group correlated with a decreased abundance of *Erysipelotrichaceae*, a bacterial family significantly increased in the OMS group and positively correlated with waist circumference (Gallardo-Becerra et al., 2020). A high abundance of *Erysipelotrichaceae* has been associated with host dyslipidemia in O, metabolic syndrome, and hypercholesterolemia (Spencer et al., 2011). It suggests that phage contigs 207 and 540 could be used to diminish the abundance of *Erysipelotrichaceae* in O and OMS groups (Gallardo-Becerra et al., 2020).

These examples open the possibility of using phages as a therapeutic option against the bacterial changes typically associated with O. Indeed, it has been suggested that gut phageome represents a source of

individual phages with potential therapeutic applications (Ott et al., 2017). Even successful treatments against *Clostridium difficile* using bacteria-free fecal filtrate provided the first evidence that phageome manipulation may be an effective therapeutic strategy to stabilize the bacterial eubiosis in the microbiome (Ott et al., 2017). Compelling evidence supports the idea that shifts in the microbial system during infancy may increase the risk of O later in life (Scheepers et al., 2015). Therefore, manipulating the gut microbiota using phages at an early stage might offer the prevention and treatment of the bacterial changes associated with O. Fecal microbiota transplantation revealed that phages could be co-transferred with bacteria (Draper et al., 2018).

We believe that our study provides a better knowledge of the phage-bacteria interactions in the gut microbiome. The development of *in vivo* models to test the phage-bacteria dynamics in O will undoubtedly be an essential area that could help complement the understanding of phage's role in microbiota changes associated with O and metabolic syndrome.

Limitations of the study

This study could be limited by the number of studied samples, which probably accounted for the lack of statistical significance obtained in some of the analyses. We performed an accurate estimation of viral community composition and diversity. However, we should note that our assemblies may represent fragments of the same phage genome that could affect accurate estimates about our phageome and its prevalence in the human population.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCE TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - Experimental model and subject details
 - Viral-like particles (VLPs) isolation
 - Microscopy visualization and VLPs counts
 - Viral DNA shotgun sequencing
 - Cleaning and clustering of sequenced reads
 - Analysis of viral reads richness
 - Functional profiles and pVOGs analysis
 - Classification of viral reads
 - De Novo contig assembly
 - Taxonomic classification of *de novo* assembly
 - Differential abundance of phage contigs
 - Richness and diversity of phage contigs
 - Bacteria and biochemical parameters correlations
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.102900>.

ACKNOWLEDGMENTS

We thank Juan Manuel Hurtado Ramírez for the informatics technical support. We thank the National Laboratory of Advanced Microscopy for help with epifluorescence microscopy and Dr. Guadalupe Trinidad Zavala Padilla for her help with TEM microscopy, both in the Biotechnology Institute at UNAM. SB thanks to the Master's and Doctoral Program in Medical, Dental and Health Sciences, in the field of knowledge: Experimental Clinical Research in Health, disciplinary field: Clinical Biochemistry, at UNAM. Doctoral fellowship: 405490260, CVU: 481385. F.C.G. and L.G.B. acknowledge the support of CONACyT as post-graduate fellows. G.L.-L. was supported by CONACyT postdoctoral fellowship CB-2013/223279. We also

thank the Unidad de Secuenciación Masiva from INMEGEN for sequencing technical support. There is no conflict of interest for authors. This research was funded by the DGAPA PAPIIT UNAM (IN215520) and CONACYT grant Ciencia-Frontera-2019-263986 both from Mexico to A.O.-L. We also acknowledge the support of program Actividades de Intercambio Académico 2019 CIC-UNAM-CIAD.

AUTHOR CONTRIBUTIONS

Conceptualization, S.B., G.L.-L., F.C.-G., and A.O.-L.; data curation, S.B., G.L.-L., F.C.-G., L.G.-B., R.G.-L., F.S., E.E.-M., J.P.O.-R., B.E.L.-C., S.C.-Q., and A.O.-L.; formal analysis, S.B., G.L.-L., F.C.-G., L.G.-B., R.G.-L., F.S., E.E.-M., J.P.O.-R., B.E.L.-C., S.C.-Q., and A.O.-L.; funding acquisition, S.C.-Q. and A.O.-L.; investigation and methodology, S.B., G.L.-L., F.C.-G., L.G.-B., F.S., E.E.-M., J.P.O.-R., A.H.-R., A.M.-V., and A.O.-L.; project administration, S.B., G.L.-L., F.C.-G., and A.O.-L.; writing original draft, S.B., G.L.-L., F.C.-G., and A.O.-L.; writing review & editing, S.B., G.L.-L., F.C.-G., L.G.-B., R.G.-L., F.S., E.E.-M., J.P.O.-R., A.H.-R., A.M.-V., B.E.L.-C., S.C.-Q., and A.O.-L.. All authors read and approved the final manuscript.

DECLARATIONS OF INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

We used the previously published samples (16) under the approval of the Ethics Committee of the National Institute of Genomic Medicine (INMEGEN) in Mexico City to approve the study. The parents or guardians of donors signed the informed consent form for participation, and the donors assented to participate.

Received: January 11, 2021

Revised: May 1, 2021

Accepted: July 21, 2021

Published: August 20, 2021

REFERENCES

- Abulencia, C.B., Wyborski, D.L., Garcia, J.A., Podar, M., Chen, W., Chang, S.H., Chang, H.W., Watson, D., Brodie, E.L., Hazen, T.C., et al. (2006). Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl. Environ. Microbiol.* 72, 3291–3301. <https://doi.org/10.1128/AEM.72.5.3291-3301.2006>.
- Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., et al. (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* 17, 690–703. <https://doi.org/10.1016/j.chom.2015.04.004>.
- Bai, J., Hu, Y., and Bruner, D.W. (2019). 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. <https://doi.org/10.1111/ijpo.12480>.
- Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A.S., Doran, K.S., et al. (2013). Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl. Acad. Sci. U S A* 110, 10771–10776. <https://doi.org/10.1073/pnas.1305923110>.
- Bikel, S., Valdez-Lara, A., Cornejo-Granados, F., Rico, K., Canizales-Quinteros, S., Soberón, X., Del Pozo-Yauner, L., and Ochoa-Leyva, A. (2015). Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: Towards a systems-level understanding of human microbiome. *Comput. Struct. Biotechnol. J.* 390–401. Elsevier. <https://doi.org/10.1016/j.csbj.2015.06.001>.
- Biro, F.M., and Wien, M. (2010). Childhood obesity and adult morbidities. *Am. J. Clin. Nutr.* 1499S. American Society for Nutrition. <https://doi.org/10.3945/ajcn.2010.28701B>.
- Bouter, K.E., van Raalte, D.H., Groen, A.K., and Nieuwdorp, M. (2017). Role of the gut microbiome in the Pathogenesis of obesity and obesity-Related metabolic Dysfunction. *Gastroenterology* 152, 1671–1678. <https://doi.org/10.1053/j.gastro.2016.12.048>.
- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., and Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* 185, 6220–6223. <https://doi.org/10.1128/JB.185.20.6220-6223.2003>.
- Breitbart, M., Haynes, M., Kelley, S., Angly, F., Edwards, R.A., Felts, B., Mahaffy, J.M., Mueller, J., Nulton, J., Rayhawk, S., et al. (2008). Viral diversity and dynamics in an infant gut. *Res. Microbiol.* 159, 367–373. <https://doi.org/10.1016/j.resmic.2008.04.006>.
- Brum, J.R., and Sullivan, M.B. (2015). Rising to the challenge: Accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* 147–159. Nature Publishing Group. <https://doi.org/10.1038/nrmicro3404>.
- Canfora, E.E., Meex, R.C.R., Venema, K., and Blaak, E.E. (2019). Gut microbial metabolites in obesity, NAFLD and T2DM. *Nat. Rev. Endocrinol.* 261–273. Nature Publishing Group. <https://doi.org/10.1038/s41574-019-0156-z>.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. <https://doi.org/10.1038/nmeth.f.303>.
- Cervantes-Echeverría, M., Equihua-Medina, E., Cornejo-Granados, F., Hernández-Reyna, A., Sánchez, F., López-Contreras, B.E., Canizales-Quinteros, S., and Ochoa-Leyva, A. (2018). Whole-genome of Mexican-crAssphage isolated from the human gut microbiome. *BMC Res. Notes* 11, 902. <https://doi.org/10.1186/s13104-018-4010-5>.
- Chen, X., Sun, H., Jiang, F., Shen, Y., Li, X., Hu, X., Shen, X., and Wei, P. (2020). Alteration of the gut microbiota associated with childhood obesity by 16S rRNA gene sequencing. *PeerJ*, 2020. <https://doi.org/10.7717/peerj.8317>.
- Cornuault, J.K., Petit, M.A., Mariadassou, M., Benevides, L., Moncaut, E., Langella, P., Sokol, H., and De Paepe, M. (2018). Phages infecting *Faecalibacterium prausnitzii* belong to novel viral genera that help to decipher intestinal viromes. *Microbiome* 6, 65. <https://doi.org/10.1186/s40168-018-0452-1>.

- Coutinho, F.H., Edwards, R.A., and Rodríguez-Valera, F. (2019). Charting the diversity of uncultured viruses of Archaea and Bacteria. *BMC Biol.* 17, 1–16. <https://doi.org/10.1186/s12915-019-0723-8>.
- Deboutte, W., Beller, L., Yinda, C.K., Maes, P., de Graaf, D.C., and Matthijssens, J. (2020). Honeybee-associated prokaryotic viral communities reveal wide viral diversity and a profound metabolic coding potential. *Proc. Natl. Acad. Sci. U S A* 117, 10511–10519. <https://doi.org/10.1073/pnas.1921859117>.
- De Ferranti, S.D., Gauvreau, K., Ludwig, D.S., Neufeld, E.J., Newburger, J.W., and Rifai, N. (2004). Prevalence of the metabolic syndrome in American adolescents: Findings from the Third National Health and Nutrition Examination Survey. *Circulation* 110, 2494–2497. <https://doi.org/10.1161/01.CIR.0000145117.40114.C7>.
- DiBonaventura, M.D., Meincke, H., Le Lay, A., Fournier, J., Bakker, E., and Ehrenreich, A. (2018). Obesity in Mexico: prevalence, comorbidities, associations with patient outcomes, and treatment experiences. *Diabetes Metab. Syndr. Obes.* 11, 1–10. <https://doi.org/10.2147/DMSO.S129247>.
- Draper, L.A., Ryan, F.J., Smith, M.K., Jalanka, J., Mattila, E., Arkkila, P.A., Ross, R.P., Satokari, R., and Hill, C. (2018). Long-term colonisation with donor bacteriophages following successful faecal microbial transplantation. *Microbiome* 6, 220. <https://doi.org/10.1186/s40168-018-0598-x>.
- Duerkop, B.A., Kleiner, M., Paez-Espino, D., Zhu, W., Bushnell, B., Hassell, B., Winter, S.E., Kyrpides, N.C., and Hooper, L.V. (2018). Murine colitis reveals a disease-associated bacteriophage community. *Nat. Microbiol.* 3, 1023–1031. <https://doi.org/10.1038/s41564-018-0210-y>.
- Duhaime, M.B., Deng, L., Poulos, B.T., and Sullivan, M.B. (2012). Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: A rigorous assessment and optimization of the linker amplification method. *Environ. Microbiol.* 14, 2526–2537. <https://doi.org/10.1111/j.1462-2920.2012.02791.x>.
- Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* 5, 1–11. <https://doi.org/10.1038/ncomms5498>.
- Edwards, R.A., and Rohwer, F. (2005). Viral metagenomics. *Nat. Rev. Microbiol.* 3, 504–510. <https://doi.org/10.1038/nrmicro1163>.
- Elizondo-Montemayor, L., Gutierrez, N.G., Moreno, D.M., Martínez, U., Tamargo, D., and Treviño, M. (2013). School-based individualised lifestyle intervention decreases obesity and the metabolic syndrome in Mexican children. *J. Hum. Nutr. Diet.* 26, 82–89. <https://doi.org/10.1111/jhn.12070>.
- Evía-Viscarra, M.L., Rodea-Montero, E.R., Apolinar-Jiménez, E., and Quintana-Vargas, S. (2013). 'Metabolic syndrome and its components among obese (BMI \geq 95th) Mexican adolescents'. *Endocr. Connections* 2, 208–215. <https://doi.org/10.1530/ec-13-0057>.
- Franks, P.W., Hanson, R.L., Knowler, W.C., Sievers, M.L., Bennett, P.H., and Looker, H.C. (2010). Childhood obesity, other cardiovascular risk factors, and premature death. *New Engl. J. Med.* 362, 485–493. <https://doi.org/10.1056/NEJMoa0904130>.
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>.
- Gallardo-Becerra, L., Cornejo-Granados, F., García-López, R., Valdez-Lara, A., Bikel, S., Canizales-Quinteros, S., López-Contreras, B.E., Mendoza-Vargas, A., Nielsen, H., and Ochoa-Leyva, A. (2020). Metatranscriptomic analysis to define the Secrebiome, and 16S rRNA profiling of the gut microbiome in obesity and metabolic syndrome of Mexican children. *Microb. Cell Factories* 19. <https://doi.org/10.1186/s12934-020-01319-y>.
- García-López, R., Pérez-Brocá, V., and Moya, A. (2019). Beyond cells – the virome in the human holobiont. *Microb. Cell.* 373–396. Shared Science Publishers OG. <https://doi.org/10.15698/mic2019.09.689>.
- Garmaeva, S., Sinha, T., Kurilshikov, A., Fu, J., Wijmenga, C., and Zernakova, A. (2019). Studying the gut virome in the metagenomic era: Challenges and perspectives. *BMC Biol.* 1–14. BioMed Central Ltd. <https://doi.org/10.1186/s12915-019-0704-y>.
- Gregory, A.C., Zablocki, O., Howell, A., Bolduc, B., and Sullivan, M.B. (2019). The human gut virome database. *bioRxiv*, 655910. <https://doi.org/10.1101/655910>.
- Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O., Pratama, A.A., Gazitúa, M.C., Vik, D., Sullivan, M.B., et al. (2021). VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9, 37. <https://doi.org/10.1186/s40168-020-00990-y>.
- Hatfull, G.F. (2008). Bacteriophage genomics. *Curr. Opin. Microbiol.* 447–453. NIH Public Access. <https://doi.org/10.1016/j.mib.2008.09.004>.
- He, Y., Wu, W., Zheng, H.M., Li, P., McDonald, D., Sheng, H.F., Chen, M.X., Chen, Z.H., Ji, G.Y., Zheng, Z.D.X., et al. (2018). Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat. Med.* 24, 1532–1535. <https://doi.org/10.1038/s41591-018-0164-x>.
- Hsu, B.B., Gibson, T.E., Yeliseyev, V., Bry, L., Silver, P.A., and Gerber, G.K. (2019). Dynamic Modulation of the gut microbiota and metabolome by bacteriophages in a mouse model. *Cell Host & Microbe* 25, 803–814. <https://doi.org/10.1016/j.chom.2019.05.001>.
- Huson, D.H., Albrecht, B., Bağcı, C., Bessarab, I., Górská, A., Jolic, D., and Williams, R.B.H. (2018). MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. *Biol. Direct* 13, 6. <https://doi.org/10.1186/s13062-018-0208-7>.
- Kim, K.H., Chang, H.W., Nam, Y., Roh, S.W., Kim, M.S., Sung, Y., Jeon, C.O., Oh, H.M., and Bae, J.W. (2008). Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl. Environ. Microbiol.* 74, 5975–5985. <https://doi.org/10.1128/AEM.01275-08>.
- Kim, K.H., and Bae, J.W. (2011). Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl. Environ. Microbiol.* 77, 7663–7668. <https://doi.org/10.1128/AEM.00289-11>.
- Kim, M.-S., and Bae, J.-W. (2016). Spatial disturbances in altered mucosal and luminal gut viromes of diet-induced obese mice. *Environ. Microbiol.* 18, 1498–1510. <https://doi.org/10.1111/1462-2920.13182>.
- Kim, M.H., Yun, K.E., Kim, J., Park, E., Chang, Y., Ryu, S., Kim, H.L., and Kim, H.N. (2020). Gut microbiota and metabolic health among overweight and obese individuals. *Scientific Rep.* 10, 1–11. <https://doi.org/10.1038/s41598-020-76474-8>.
- Kim, M.S., Park, E.J., Roh, S.W., and Bae, J.W. (2011). Diversity and abundance of single-stranded DNA viruses in human feces. *Appl. Environ. Microbiol.* 77, 8062–8070. <https://doi.org/10.1128/AEM.06331-11>.
- Kulakov, L.A., McAlister, M.B., Ogdén, K.L., Larkin, M.J., and O'Hanlon, J.F. (2002). Analysis of bacteria contaminating ultrapure water in industrial systems. *Appl. Environ. Microbiol.* 68, 1548–1555. <https://doi.org/10.1128/AEM.68.4.1548-1555.2002>.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
- Laurence, M., Hatzis, C., and Brash, D.E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS One* 9. <https://doi.org/10.1371/journal.pone.0097876>.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Ma, Y., You, X., Mai, G., Tokuyasu, T., and Liu, C. (2018). A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* 6, 24. <https://doi.org/10.1186/s40168-018-0410-y>.
- Maiques, E., Úbeda, C., Campoy, S., Salvador, N., Lasa, I., Novick, R.P., Barbé, J., and Penadés, J.R. (2006). β -lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *J. Bacteriol.* 188, 2726–2729. <https://doi.org/10.1128/JB.188.7.2726-2729.2006>.
- Manrique, P., Bolduc, B., Walk, S.T., Van Oost, J. Der, De Vos, W.M., and Young, M.J. (2016). Healthy human gut phageome. *Proc. Natl. Acad. Sci. U S A* 113, 10400–10405. <https://doi.org/10.1073/pnas.1601060113>.

- Manrique, P., Zhu, Y., van der Oost, J., Herrema, H., Nieuwdorp, M., de Vos, W.M., and Young, M. (2021). Gut bacteriophage dynamics during fecal microbial transplantation in subjects with metabolic syndrome. *Gut Microbes* 13, 1–15. <https://doi.org/10.1080/19490976.2021.1897217>.
- Mattos, R.T., Medeiros, N.I., Menezes, C.A., Fares, R.C.G., Franco, E.P., Dutra, W.O., Rios-Santos, F., Correa-Oliveira, R., and Gomes, J.A.S. (2016). Chronic low-grade inflammation in childhood obesity is associated with decreased il-10 expression by monocyte subsets. *PLoS One* 11. <https://doi.org/10.1371/journal.pone.0168610>.
- Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* 21, 1616–1625. <https://doi.org/10.1101/gr.122705.111>.
- Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013). Rapid evolution of the human gut virome. *Proc. Natl. Acad. Sci. U S A* 110, 12450–12455. <https://doi.org/10.1073/pnas.1300833110>.
- Monaco, C.L., Gootenberg, D.B., Zhao, G., Handley, S.A., Ghebremichael, M.S., Lim, E.S., Lankowski, A., Baldrige, M.T., Wilen, C.B., Flagg, M., et al. (2016). Altered virome and bacterial microbiome in human Immunodeficiency Virus-associated Acquired Immunodeficiency syndrome. *Cell Host and Microbe* 19, 311–322. <https://doi.org/10.1016/j.chom.2016.02.011>.
- Moreno-Gallego, J.L., Chou, S.P., Di Rienzi, S.C., Goodrich, J.K., Spector, T.D., Bell, J.T., Youngblut, N.D., Hewson, I., Reyes, A., and Ley, R.E. (2019). Virome diversity correlates with intestinal microbiome diversity in adult monozygotic twins. *Cell Host and Microbe* 25, 261–272. e5. <https://doi.org/10.1016/j.chom.2019.01.019>.
- Norman, J.M., Handley, S.A., Baldrige, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., Fleshner, P., et al. (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell*. <https://doi.org/10.1016/j.cell.2015.01.002>.
- Ofir, G., and Sorek, R. (2018). Contemporary phage Biology: From Classic models to New insights. *Cell*, 1260–1270. <https://doi.org/10.1016/j.cell.2017.10.045>.
- Ott, S.J., Waetzig, G.H., Rehman, A., Moltzau-Anderson, J., Bharti, R., Grasis, J.A., Cassidy, L., Tholey, A., Fickenscher, H., Seeger, D., et al. (2017). Efficacy of Sterile fecal Filtrate transfer for Treating Patients with Clostridium difficile Infection. *Gastroenterology* 152, 799–811. e7. <https://doi.org/10.1053/j.gastro.2016.11.010>.
- Pedersen, H.K., Gudmundsdottir, V., Nielsen, H.B., Hyotylainen, T., Nielsen, T., Jensen, B.A.H., Forslund, K., Hildebrand, F., Pridi, E., Falony, G., et al. (2016). Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* 535, 376–381. <https://doi.org/10.1038/nature18646>.
- Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. <https://doi.org/10.1093/bioinformatics/bts174>.
- Perichart-Perera, O., Balas-Nakash, M., Schiffman-Selechnik, E., Barbato-Dosal, A., and Vadillo-Ortega, F. (2007). Obesity Increases metabolic syndrome risk factors in school-Aged children from an Urban school in Mexico City. *J. Am. Diet. Assoc.* 107, 81–91. <https://doi.org/10.1016/j.jada.2006.10.011>.
- Pihl, A.F., Fonvig, C.E., Stjernholm, T., Hansen, T., Pedersen, O., and Holm, J.C. (2016). The role of the gut microbiota in childhood obesity. *Child. Obes.* 292–299. Mary Ann Liebert Inc. <https://doi.org/10.1089/chi.2015.0220>.
- Rasmussen, T.S., Mentzel, C.M.J., Kot, W., Castro-Mejia, J.L., Zuffa, S., Swann, J.R., Hansen, L.H., Vogensen, F.K., Hansen, A.K., and Nielsen, D.S. (2020). Faecal virome transplantation decreases symptoms of type 2 diabetes and obesity in a murine model. *Gut* 69, 2122–2130. <https://doi.org/10.1136/gutjnl-2019-320005>.
- Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466, 334–338. <https://doi.org/10.1038/nature09199>.
- Reyes, A., Wu, M., McNulty, N.P., Rohwer, F.L., and Gordon, J.I. (2013). Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proc. Natl. Acad. Sci. U S A* 110, 20236–20241. <https://doi.org/10.1073/pnas.1319470110>.
- Reyes, A., Blanton, L.V., Cao, S., Zhao, G., Manary, M., Trehan, I., Smith, M.I., Wang, D., Virgin, H.W., Rohwer, F., et al. (2015). Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl. Acad. Sci. U S A* 112, 11941–11946. <https://doi.org/10.1073/pnas.1514285112>.
- Rho, M., Tang, H., and Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* 38, e191. <https://doi.org/10.1093/nar/gkq747>.
- Romero-Martínez, M., Shamah-Levy, T., Vielma-Orozco, E., Heredia-Hernández, O., Mojica-Cuevas, J., Cuevas-Nasu, L., Rivera-Dommarco, J., Gómez-Humarán, M., Gaona-Pineda, E.B., Gómez-Acosta, L.M., et al. (2019). National Health and Nutrition Survey 2018–19: Methodology and perspectives. *Salud Publica Mex.* 61, 917–923. <https://doi.org/10.21149/11095>.
- Romieu, I., Dossus, L., Barquera, S., Blottière, H.M., Franks, P.W., Gunter, M., Hwalla, N., Hursting, S.D., Leitzmann, M., Margetts, B., et al. (2017). Energy balance and obesity: what are the main drivers? *Cancer Causes Control* 28, 247–258. <https://doi.org/10.1007/s10552-017-0869-z>.
- Roux, S., Solonenko, N.E., Dang, V.T., Poulos, B.T., Schwenck, S.M., Goldsmith, D.B., Coleman, M.L., Breitbart, M., and Sullivan, M.B. (2016). Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ*, 2016. <https://doi.org/10.7717/peerj.2777>.
- Roux, S., Sullivan, M.B., Emerson, J.B., and Eloe-Fadrosh, E.A. (2017). Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ*, e3817. <https://doi.org/10.7717/peerj.3817>.
- Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., and Walker, A.W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12, 87. <https://doi.org/10.1186/s12915-014-0087-z>.
- Sausset, R., Petit, M.A., Gaboriau-Routhiau, V., and De Paepe, M. (2020). New insights into intestinal phages. *Mucosal Immunol.* 205–215. Springer Nature. <https://doi.org/10.1038/s41385-019-0250-5>.
- Scheepers, L.E.J.M., Penders, J., Mbakwa, C.A., Thijs, C., Mommers, M., and Arts, I.C.W. (2015). The intestinal microbiota composition and weight development in children: The KOALA Birth Cohort Study. *Int. J. Obes.* 39, 16–25. <https://doi.org/10.1038/ijo.2014.178>.
- Schulfer, A., Santiago-Rodríguez, T.M., Ly, M., Borin, J.M., Chopyk, J., Blaser, M.J., and Pride, D.T. (2020). Fecal viral community Responses to High-Fat diet in mice. *mSphere* 5. <https://doi.org/10.1128/msphere.00833-19>.
- Shkoporov, A.N., Khokhlova, E.V., Fitzgerald, C.B., Stockdale, S.R., Draper, L.A., Ross, R.P., and Hill, C. (2018a). ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. *Nat. Commun.* 9, 1–8. <https://doi.org/10.1038/s41467-018-07225-7>.
- Shkoporov, A.N., Ryan, F.J., Draper, L.A., Forde, A., Stockdale, S.R., Daly, K.M., McDonnell, S.A., Nolan, J.A., Sutton, T.D.S., Dalmasso, M., et al. (2018b). Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* 6, 68. <https://doi.org/10.1186/s40168-018-0446-z>.
- Shkoporov, A.N., Clooney, A.G., Sutton, T.D.S., Ryan, F.J., Daly, K.M., Nolan, J.A., McDonnell, S.A., Khokhlova, E.V., Draper, L.A., Forde, A., et al. (2019). The human gut virome is highly diverse, Stable, and individual specific. *Cell Host Microbe* 26, 527–541. e5. <https://doi.org/10.1016/j.chom.2019.09.009>.
- Shkoporov, A.N., and Hill, C. (2019). Bacteriophages of the human gut: The “Known unknown” of the microbiome. *Cell Host Microbe*, 195–209. <https://doi.org/10.1016/j.chom.2019.01.017>.
- Siranosian, B.A., Tamburini, F.B., Sherlock, G., and Bhatt, A.S. (2020). Acquisition, transmission and strain diversity of human gut-colonizing crAss-like phages. *Nat. Commun.* 11, 1–11. <https://doi.org/10.1038/s41467-019-14103-3>.
- Spencer, M.D., Hamp, T.J., Reid, R.W., Fischer, L.M., Zeisel, S.H., and Fodor, A.A. (2011). Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. *Gastroenterology* 140, 976–986. <https://doi.org/10.1053/j.gastro.2010.11.049>.
- Stagaman, K., Cepon-Robins, T.J., Liebert, M.A., Gildner, T.E., Urlacher, S.S., Madimenos, F.C., Guillemin, K., Snodgrass, J.J., Sugiyama, L.S., and Bohannan, B.J.M. (2018). Market Integration Predicts human gut microbiome Attributes across

a Gradient of Economic development. *mSystems* 3. <https://doi.org/10.1128/msystems.00122-17>.

Sutton, T.D.S., and Hill, C. (2019). Gut bacteriophage: Current understanding and challenges. *Front. Endocrinol.* 784. *Frontiers Media S.A.* <https://doi.org/10.3389/fendo.2019.00784>

Tanner, M.A., Goebel, B.M., Dojka, M.A., and Pace, N.R. (1998). Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl. Environ. Microbiol.* 64, 3110–3113. <https://doi.org/10.1128/aem.64.8.3110-3113.1998>.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., et al. (2009). A core gut microbiome in obese and lean twins. *Nature* 457, 480–484. <https://doi.org/10.1038/nature07540>.

Virgin, H.W. (2014). The virome in mammalian physiology and disease. *Cell*, 142–150. *Cell Press*. <https://doi.org/10.1016/j.cell.2014.02.032>.

Viveiros, A., and Oudit, G.Y. (2020). The dual nature of obesity in metabolic programming: Quantity versus quality of adipose tissue. *Clin. Sci.* 2447–2451. *Portland Press Ltd.* <https://doi.org/10.1042/CS20201028>.

Wang, K., Liao, M., Zhou, N., Bao, L., Ma, K., Zheng, Z., Wang, Y., Liu, C., Wang, W., Wang, J., et al. (2019). *Parabacteroides distasonis* Alleviates obesity and metabolic Dysfunctions via Production of Succinate and Secondary bile acids. *Cell Rep.* 26, 222–235.e5. <https://doi.org/10.1016/j.celrep.2018.12.028>.

WHO (2017). *Childhood Overweight and Obesity (WHO)*.

Willerslev, E., Hansen, A.J., and Poinar, H.N. (2004). Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol. Evol.* 141–147. *Elsevier Ltd.* <https://doi.org/10.1016/j.tree.2003.11.010>.

Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15, R46. <https://doi.org/10.1186/gb-2014-15-3-r46>.

Yilmaz, S., Allgaier, M., and Hugenholtz, P. (2010). Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat. Methods*, 943–944. <https://doi.org/10.1038/nmeth1210-943>.

Yuan, X., Chen, R., McCormick, K.L., Zhang, Y., Lin, X., and Yang, X. (2021). The role of the gut microbiota on the metabolic status of obese

children. *Microb. Cell Factories* 20, 53. <https://doi.org/10.1186/s12934-021-01548-9>.

Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W., and Church, G.M. (2006). Sequencing genomes from single cells by polymerase cloning. *Nat. Biotechnol.* 24, 680–686. <https://doi.org/10.1038/nbt1214>.

Zhao, G., Vatanen, T., Droit, L., Park, A., Kostic, A.D., Poon, T.W., Vlamakis, H., Siljander, H., Härkönen, T., Hämäläinen, A.M., et al. (2017). Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children. *Proc. Natl. Acad. Sci. U S A* 114, E6166–E6175. <https://doi.org/10.1073/pnas.1706359114>.

Zhong, H., Penders, J., Shi, Z., Ren, H., Cai, K., Fang, C., Ding, Q., Thijs, C., Blaak, E.E., Stehouwer, C.D.A., et al. (2019). Impact of early events and lifestyle on the gut microbiota and metabolic phenotypes in young school-age children. *Microbiome* 7, 2. <https://doi.org/10.1186/s40168-018-0608-z>.

Zuo, T., Lu, X.J., Zhang, Y., Cheung, C.P., Lam, S., Zhang, F., Tang, W., Ching, J.Y.L., Zhao, R., Chan, P.K.S., et al. (2019). Gut mucosal virome alterations in ulcerative colitis. *Gut* 68, 1169–1179. <https://doi.org/10.1136/gutjnl-2018-318131>.

STAR★METHODS

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
RNA later	Thermo Fisher Scientific	Cat. AM7020
SM Buffer	Nalgene	Cat. 7252520
DNase I	Thermo Fisher Scientific	Cat. 18047019
SYBR Green	Thermo Fisher Scientific	Cat. K0221
Critical commercial assays		
QIAamp MinElute Virus Spin Kit	Qiagen	Cat. 57704
Qubit	Thermo Fisher Scientific	Cat. Q32851
Illumina Nextera XT DNA Library	Illumina	Cat. FC-131-1024
Deposited data		
Assembled contigs	This paper	BioProject PRJNA646512 BioSample SAMN15545081
Software and algorithms		
Trim Galore 1.12	The Babraham Institute	https://github.com/FelixKrueger/TrimGalore
Fastx Toolkit 0.7	Hannon Lab	http://hannonlab.cshl.edu/fastx_toolkit/index.html
CD-HIT 4.6	(Fu et al., 2012)	http://cd-hit.org/
R 3.6.2	N/A	https://www.r-project.org/
IDBA-UD assembler 1.1.1	(Peng et al., 2012)	https://i.cs.hku.hk/~alse/hkubrg/projects/idba_ud/
Bowtie 2.3.5	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
VirSorter 2.2.1	(Guo et al., 2021)	https://github.com/jiarong/VirSorter2
QIIME 1.9	(Caporaso et al., 2010)	http://qiime.org/
DESeq2 3.13	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Frag Gene Scan 1.31	(Rho et al., 2010)	https://sourceforge.net/projects/fraggenescan/
Meta Genome Analyzer 6.18.3	(Huson et al., 2018)	https://software-ab.informatik.uni-tuebingen.de/download/megan6/welcome.html
Other		
In-house scripts describing the data analysis process	This paper	https://github.com/lab8a/2021-iScience-Phageome

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Adrian Ochoa-Leyva (adrian.ochoa@ibt.unam.mx).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All original code describing the data analysis process are available on GitHub at <https://github.com/lab8a/2021-iScience-Phageome>. The sequence data have been deposited in the NCBI under the NCBI BioProject accession number: PRJNA646512. Accession numbers are also listed in the [key resources table](#). Any

additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Experimental model and subject details

The information on the fecal samples and the biochemical parameters information used in this study are fully described in a previous study of our group (Gallardo-Becerra et al., 2020). Briefly, we analyzed the stools from 10 normal weight (NW), 10 obese (O), and 8 obese with metabolic syndrome (OMS) children, aged 7–10 years old. All children came from households with a middle economic class income and belonged to a similar socio-cultural status. All of them lived in Mexico City at the time of collection and did not practice any sport regularly.

The study groups were paired by gender and age (Table S1). Feces samples were collected and refrigerated at 4°C and transported to the research facility within the following 12 h in a portable cooler with ice packs to preserve the temperature. At the research facility samples were aliquoted into 200 mg portions in sterile plastic containers with RNA later and stored at –70°C. For the biochemical parameters, 5 mL of blood samples were drawn after 8–12 h fasting on the same day of the feces collection. Also, anthropometric parameters, blood pressure, and body mass index were measured following standardized procedures, as previously described (Gallardo-Becerra et al., 2020).

Obesity was defined by a body mass index (BMI) \geq 95th percentile. In contrast, NW was defined as a BMI between the 15th and 75th percentiles considering age and gender, based on the Centers for Disease Control and Prevention (CDC). Metabolic syndrome parameters were determined according to previous reports (De Ferranti et al., 2004), and OMS were defined by the presence of waist circumference $>$ 75th percentile considering age and gender, and at least two of the following metabolic traits: (1) triglycerides $>$ 1.1 mmol/L (100 mg/dL); (2) HDL cholesterol $<$ 1.3 mmol/L (50 mg/dL), (3) glucose $>$ 6.1 mmol/L (110 mg/dL) and (4) systolic blood pressure $>$ 90th percentile considering gender, age, and height. Children in the O group were selected with no more than one metabolic syndrome trait.

Exclusion criteria for all samples included recent bodyweight loss $>$ 10%, antibiotic intake 3 months before sample collection, and the occurrence of diarrhea or acute gastrointestinal illness during the same period. The Ethics Committee of the Instituto Nacional de Medicina Genómica (INMEGEN) in Mexico City, Mexico, approved the study. Each child's parents or legal guardians signed the informed consent form for participation, and all children assented to participate.

Viral-like particles (VLPs) isolation

Viral-like particles (VLPs) were isolated from \sim 250 mg of fecal sample suspended by vortexing in 1 mL of sterile SM Buffer (pH 7.5), (Cat. 725-2520, Nalgene, NY, USA) to stabilize phage particles. The homogenates were centrifuged at $4,700 \times g$ for 30 min at room temperature, and the supernatant was filtered through 0.45 μ m (Cat. 725-2545, Nalgene) and 0.22 μ m PES filters (Cat. 725-2520, Nalgene NY, USA) to remove cell debris and bacterial-sized particles. The filtrate was then re-suspended in 15 mL of SM buffer and concentrated to 200 μ l at 4°C with an Amicon Ultra 15 filter unit, 100 KDa (Cat. UFC910024, Millipore, MA, USA) to remove cellular debris. The concentrate was transferred to a 1.5 mL microfuge tube and incubated with 40 μ l chloroform for 10 min at room temperature to degrade any remaining bacterial and human cell debris. Non-virus protected DNA was eliminated with 2.5 units per milliliter of DNase I following the manufacture's procedure (Cat 18047-019, Invitrogen, MA, USA). After incubation, the DNase was inactivated at 65°C for 10 min. The samples were stored at -80°C until further processing.

Microscopy visualization and VLPs counts

We used epifluorescence microscopy to quantify the isolated VLPs. We stained 10 μ l of the concentrated VLPs samples with a mix of SYBR Green (Cat. K0221, Thermo Fisher Scientific, MA, USA) and 10 μ l of paraformaldehyde previously filtered through 0.22 μ m PES membranes (Cat. 725-2520, Nalgene NY, USA). Five fields per sample were observed and quantified with an Olympus FV1000 Multiphoton Confocal Microscope, and each field was quantified in triplicate using the free image processing software Fiji. The average number of VLPs from the five fields by triplicate was used per sample to calculate the amount per gram of feces per sample (Table S3). Additionally, an aliquot of 8 μ l from the concentrated VLPs samples was

observed in the transmission electron microscope to corroborate the phage nature (morphology and structure) in VLPs (Figure 1B).

Viral DNA shotgun sequencing

The DNA of VLPs was extracted following the manufacturer's protocol for the QIAamp MinElute Virus Spin Kit (Cat. 57704, Qiagen, Hilden, Alemania). The resulting DNA for each sample was quantified with a Qubit fluorometer (Cat. Q32851, Thermo Fisher Scientific, CA, USA), and diluted with ribonuclease-free water to a concentration of 0.3 ng/μl. From this DNA, we prepared independent sequencing libraries following Illumina Nextera XT DNA Library Preparation protocol (Cat. FC-131-1024, Illumina, CA, USA) that supports ultra-low DNA input with unique barcodes for multiplexing. For DNA tagmentation, we mixed 5 μl of DNA at 0.3 ng/μl with the tagmentation reaction mix. Next, we added the indexed oligos and amplified the library for 12 cycles. Each library was purified with 30 μl of AMPure XP beads (Cat. A63881, Beckman Coulter, CA, USA) to obtain ~600 bp DNA fragments.

The size and quality of each library were assessed with a DNA bioanalyzer 2100 (Cat. 5067-4626, Agilent Technologies, CA, USA). All barcoded libraries were pooled together and sequenced using the Illumina NextSeq500 platform in the 2x150 pair-end mode at the Sequencing Unit Facility of the National Institute for Genomic Medicine, México.

Cleaning and clustering of sequenced reads

Total reads were dereplicated, adapters and low-quality bases (PHRED Q30) were trimmed using Trim_Galore (<https://github.com/FelixKrueger/TrimGalore>), and the first 20 nucleotides were removed with Fastx Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Human and bacterial reads by read mapping using BWA (against the Homo sapiens GRCh38.p13 reference genome GenBank GCA_000001405.28) and Kraken (Wood and Salzberg, 2014) against bacteria NR database, with default parameters. All reads mapped to those genomes were removed, and the remaining reads were named quality-filtered reads. Quality-filtered reads were clustered at a 95% identity using CD-HIT (Fu et al., 2012) to remove redundancy and generate a unique sequence dataset.

Analysis of viral reads richness

The viral richness between groups (NW, O, OMS) was determined, collecting 1,000 random subsamples of 149,000 single-end quality-filtered reads using seqtk subseq according to the smallest sample (NW_169: 149,775), and later each sub-sample was clustered at a 95% identity level using CD-HIT to identify the unique groups of reads.

Functional profiles and pVOGs analysis

The quality-filtered reads were mapped onto the viral NR RefSeq and POGs databases using BLASTX with a maximum e-value cutoff of 0.001 and a maximum of 50 reported target sequences. After mapping, an abundance matrix was generated using an in-house bash script. The matrix was then annotated according to the KEGG annotation of each protein using the UniProtKB online database and an in-house bash script. The KEGG functional profile was then generated using the relative abundance for each protein and function for each sample. The quality-reads were mapped against the Prokaryotic Virus Orthologous Groups (pVOGs) database using BLASTX with a maximum e-value cutoff of 0.001, and a maximum of 50 reported target sequences. The final results were filtered with an in-house bash script to get the final results with the pVOGS classification for each sample.

Classification of viral reads

The quality-filtered unique sequences were taxonomically classified into orders and families, according to the International Committee on Taxonomy of Viruses (ICTV), using BLASTX with maximum e-value cutoff 0.001 against the NR RefSeq viral database and considering the lowest-common ancestor algorithm in MetaGenomeAnalyzer (MEGAN6) (Huson et al., 2018) using the following parameters: Min Support: 1, Min Score: 40.0, Max Expected: 0.01, Top Percent: 10.0, Min-Complexity filter: 0.44. Absolute read counts for selected viral taxa were normalized using all the reads from each sample, obtaining the relative abundance for each sample using R scripts.

De Novo contig assembly

The total quality-filtered reads from all samples were pooled for *de novo* assembly using IDBA-UD assembler (Peng et al., 2012) with a k-mer length of 20-125 with scaffolding rounds. Each sample's reads were mapped separately with Bowtie2 (v2.3.5) (Langmead and Salzberg, 2012) against the viral assembly using the end-to-end mode with the default parameters. Viral scaffolds covered $\geq 80\%$ in length by the reads by at least one sample were used as a cut-off to discard chimeras. Next, we kept the scaffolds ≥ 4 Kb for downstream analysis to reduce the probability of selecting viral genome fragments.

To eliminate the redundant contigs, we used CD-HIT using a 95% clustering identity. To know the number of genes per contig length, we conducted the gene prediction of each contig using FragGeneScan (Rho et al., 2010) with the following parameters -complete=0 -train=illumina_5.

Taxonomic classification of *de novo* assembly

The taxonomy classification of each contig was obtained using dc_megablast against the NT NCBI viral genomes database with maximum e-value cutoff 0.001 and the maximum number of target sequences to report set to 50 hits (Moreno-Gallego et al., 2019). The taxonomy of each contig was assigned by the lowest-common-ancestor algorithm in MEtaGenomeANalyzer (MEGAN6) using the discontinuous megablast results with the following parameters: Min Support: 1, Min Score: 40.0, Max Expected: 0.01, Top Percent: 10.0, Min-Complexity filter: 0.44. After that, all contigs without taxonomic classification with dc_megablast were searched against the NR NCBI viral protein database using BLASTX with the maximum e-value cutoff 0.001 and a maximum number of reported target sequences set to 50. The taxonomy of each contig was assigned by the lowest-common-ancestor algorithm in MetaGenomeANalyzer (MEGAN6), using the BLASTX results with the following parameters: Min Support: 1, Min Score: 40.0, Max Expected: 0.01, Top Percent: 10.0, Min-Complexity filter: 0.44. The assembly of 12,287 contigs was classified using VirSorter2 (v2.2.1) (Guo et al., 2021) with the default parameters following the tutorial provided by the authors. After that, we contrasted the VirSorter classification against the previous one using the NT NCBI viral genomes and NR NCBI viral protein databases (Guo et al., 2021). The assembly of 12,287 contigs was mapped against the Prokaryotic Virus Orthologous Groups (pVOGs) database using BLASTX with a maximum e-value cutoff 0.001, and maximum target sequences to report set to 50.

Differential abundance of phage contigs

The recruitment of reads to the contigs assembly was used to construct an abundance matrix, applying the filter of coverage and length as previously recommended (Roux et al., 2017). The coverage was defined from reads mapping (Bowtie2) at $\geq 90\%$ identity and $\geq 80\%$ length. Mapping outputs were converted into an abundance matrix using an in-house R script normalized by Reads Per Kilobase per Million sequenced reads per sample (RPKM) (Reyes et al., 2015).

Richness and diversity of phage contigs

Richness and diversity of the contig assembly were evaluated based on the median of 10,000 rarefactions at a depth of the smallest sample based on the RPKM matrix using QIIME 1.9 (Caporaso et al., 2010). Based on the presence of phage contigs in the samples, each phage contig was classified as either core phages: detected in $>80\%$ of the samples; common phages: in $>50\%$ and $<80\%$; and individual phages: appearing in $<50\%$ of the population.

Bacteria and biochemical parameters correlations

All correlations were calculated using the Spearman coefficient with rcorr function in R considering all samples. We considered the RPKM matrix for the contigs phage abundance and the relative frequency of the significant over-abundant taxa previously reported between NW, O, and OMS for the microbiota abundance (Gallardo-Becerra et al., 2020).

QUANTIFICATION AND STATISTICAL ANALYSIS

The differences in the number of VLPs between the three groups (Figure 1C) were evaluated with pairwise Mann-Whitney-Wilcoxon non-parametric tests in R.

For each sample of the analysis of viral reads richness (Figure S4), the median of all unique observations was calculated. The resulting groups were tested for normality with Shapiro-Wilk tests in R. Differences were evaluated with pairwise Mann–Whitney–Wilcoxon non-parametric tests in R.

The RPKM matrix was used to determine statistical differential viral taxonomic abundances between groups using a Kruskal Wallis test using the Holm–Sidak method, with $\alpha = 0.01$ for multiple test correction (Figure 2). This matrix was also used to determine statistical differential individual contig-abundance (Figures 4A and 4B) using DESeq2 (Love et al., 2014) with log fold change ≥ 2 and FDR adjustment (p -value ≤ 0.05) using Benjamini–Hochberg correction.

Medians for richness and diversity of phage contigs were obtained per sample and groups were evaluated for normality, following pairwise Mann–Whitney–Wilcoxon non-parametric tests in R (Figures 3A and 3B).

The prevalence of highly abundant contigs found in $>80\%$ of samples in the NW group was compared across the three groups using pairwise Mann–Whitney–Wilcoxon non-parametric tests using R (Figure S14).

For the bacteria and biochemical and anthropometric correlations, we selected the Spearman correlations with $R^2 > 0.3$ and p -value ≤ 0.05 (Figures 6, S13, S15, S17, and S18). However, after applying an FDR correction for the p -values, we did not get a significant correlation below the 0.05 cutoff. Thus, to still attend to the tendencies among groups, we referred to the unadjusted p -values in the correlation analysis.

For the Euclidean distances comparison the relative abundance phage tables were subjected to a center log-ratio transformation with the mixOmics library v6.10.9 in R using an offset of $\min_value = 1e-7$ to deal with zero logs. The resulting Euclidean matrix was then subjected to dimensional reduction with a principal components analysis (with vegan 2.5-6 in R). Euclidean distances were then calculated to create an adjacency matrix for group testing with ANOSIM and adonis (carried out with vegan) and posthoc pairwise group testing (Figure S12).