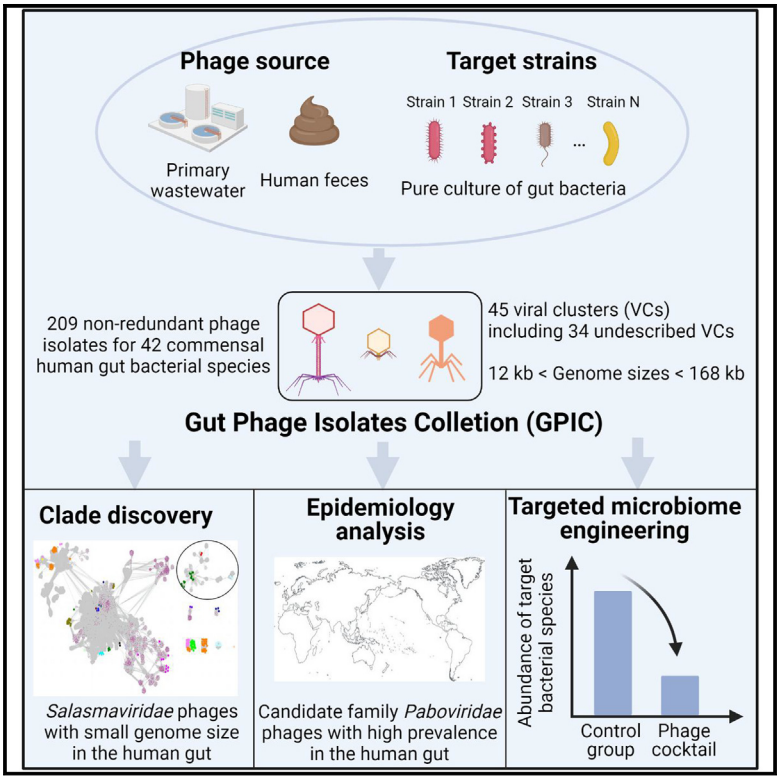
Resource

Cell Host & Microbe

Large-scale phage cultivation for commensal human gut bacteria

Graphical abstract



Highlights

- Large-scale cultivation was used to isolate 209 phages for 42 human gut bacterial species
- Genomic analysis reveals isolated phage diversity and discovers two phage families
- *Bacteroides* and *Parabacteroides* strains show great variations in phage susceptibility
- A phage cocktail reduces abundance of the targeted species in communities *in vitro*

Authors

Juntao Shen, Jieqiong Zhang, Luofei Mo, ..., Bing Li, Lei Dai, Yingfei Ma

Correspondence

lei.dai@siat.ac.cn (L.D.), yingfei.ma@siat.ac.cn (Y.M.)

In brief

Shen et al. isolate and characterize a collection of 209 phages for human gut bacteria. They discover *Salasmaviridae* phages with small genomes and highly prevalent candidate *Paboviridae* phages. A phage cocktail reduced abundance of the targeted species in host-derived communities *in vitro*, suggesting their application in targeting the gut microbiome.

Shen et al., 2023, Cell Host & Microbe *31*, 665–677
April 12, 2023
https://doi.org/10.1016/j.chom.2023.03.013

CellPress

Cell Host & Microbe

CellPress

Resource

Large-scale phage cultivation for commensal human gut bacteria

Juntao Shen,[1] Jieqiong Zhang,[1] Luofei Mo,[1,2] Yanchen Li,[1] Yake Li,[1] Cun Li,[1] Xiaoxian Kuang,[1] Zining Tao,[1] Zepeng Qu,[1] Lu Wu,[1] Junyu Chen,[1] Shiying Liu,[1] Linfang Zeng,[1] Zexi He,[1] Zuohong Chen,[3] Yu Deng,[4] Tong Zhang,[4] Bing Li,[3] Lei Dai,[1,2,*] and Yingfei Ma[1,2,5,*]

[1]CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]Guangdong Provincial Engineering Research Center for Urban Water Recycling and Environmental Safety, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China

[4]Environmental Microbiome Engineering and Biotechnology Laboratory, Department of Civil Engineering, The University of Hong Kong, Hong Kong, China

[5]Lead contact

*Correspondence: lei.dai@siat.ac.cn (L.D.), yingfei.ma@siat.ac.cn (Y.M.)

https://doi.org/10.1016/j.chom.2023.03.013

SUMMARY

Phages are highly abundant in the human gut, yet most of them remain uncultured. Here, we present a gut phage isolate collection (GPIC) containing 209 phages for 42 commensal human gut bacterial species. Genome analysis of the phages identified 34 undescribed genera. We discovered 22 phages from the *Salasmaviridae* family that have small genomes (~10–20 kbp) and infect Gram-positive bacteria. Two phages from a candidate family, *Paboviridae*, with high prevalence in the human gut were also identified. Infection assays showed that *Bacteroides* and *Parabacteroides* phages are specific to a bacterial species, and strains of the same species also exhibit substantial variations in phage susceptibility. A cocktail of 8 phages with a broad host range for *Bacteroides fragilis* strains effectively reduced their abundance in complex host-derived communities *in vitro*. Our study expands the diversity of cultured human gut bacterial phages and provides a valuable resource for human microbiome engineering.

INTRODUCTION

Bacteriophages (phages) are highly abundant in the human gut, with estimates ranging from 10^8 to 10^{10} phage particles per gram of human feces.[1,2] Phages, as bacteria predators, greatly influence the composition and function of the human gut microbiome.[3–5] Many studies have discovered the important relationship between gut phage diversity and human health.[6–8] These studies relied heavily on bioinformatic approaches.[9,10] Because of the lack of universal classification marker genes on phage genomes and the high sequence divergence, these bioinformatic approaches are unable to identify novel phages in the absence of reference genomes, and the vast majority (75%–99%) of putative bacterial phage sequences identified from human gut metagenomic data cannot be classified taxonomically or associated with any microbial hosts.[11–15]

Despite the increasing number of human gut bacterial isolates from recent large-scale cultivation studies,[16–18] efforts to isolate phages for these human gut bacteria are still limited to a few clinically potential pathogenic bacteria species, such as *Escherichia coli*,[19] *Enterococcus faecalis*,[20] and *Clostridium difficile*.[21] Besides these studies, *Bacteroides*-infecting phages are the most studied human gut commensal bacterial phages. In the 1980s, phages targeting *Bacteroidetes fragilis* from sewage[22] and feces[23] were successfully isolated. Recently, 27 phages targeting *Bacteroidetes thetaiotaomicron* were isolated from sewage using 8 lipopolysaccharide genetically modified strains.[24] The crAss-like phages, one of the most abundant phages in the human gut, were isolated using *Bacteroides intestinalis*[25] and *B. thetaiotaomicron*[24] as hosts. These results indicated that the effort to isolate phages for gut bacteria is feasible and can lead to a better understanding of the human gut ecosystem.[26,27] However, large-scale phage cultivation for the majority of commensal bacteria species in the human gut microbiome is still lacking.

In addition, phage isolates for gut bacteria can be an effective tool for manipulating the microbiome.[28,29] In gnotobiotic mice, phages can directly knock down susceptible bacteria and lead to cascading effects on other bacterial species, demonstrating the potential to modulate the microbiome.[3] Phage has also been shown to be effective in several cases in which specific phages were used to remove pathogens, resulting in a credible outcome. For example, phages targeting cytolysin-producing *E. faecalis* can attenuate alcoholic liver disease,[30] phages targeting *Fusobacterium nucleatum* can augment the efficacy of chemotherapy[31] and immunotherapy[32] for colorectal cancer, and phages targeting *Klebsiella pneumoniae* can suppress

intestinal inflammation.[33] These studies demonstrated the feasibility of applying phages to gut microbiome engineering and highlighted the unique value of cultured gut phages in understanding the biological function of specific microorganisms in disease.

In this study, we developed a systematic phage cultivation workflow to isolate phages for human gut bacteria, resulting in a collection of 209 unique phages for 42 common bacterial species in the human gut that we isolated using the plaque assay. Comprehensive analyses of these isolated phages demonstrated their prevalence in the human gut and the host range of representative isolated phages. These results shed light on bacteria-phage interactions in the human gut and lay the groundwork to leverage these phages to manipulate the gut microbiome.

RESULTS

Large-scale phage cultivation for commensal human gut bacteria

We developed a systematic phage cultivation workflow to isolate phages for human gut bacteria. Briefly, phage particles were enriched by both the coculture-based and concentration-based methods from the human feces and wastewater samples. Then we implemented the soft agar overlay method to isolate and purify phages (Figures 1A and S1; see STAR Methods). In total, 411 bacteria strains were used as hosts for phage isolation. These bacterial strains were isolated from the human feces of healthy donors and spanned across 5 phyla (Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, and Fusobacteria), 49 genera, and 109 species (Table S1). Our efforts resulted in the successful isolation of phages for 42 human gut bacterial species, including 15 species of Bacteroidetes, 19 Firmicutes, 4 Actinobacteria, and 4 Proteobacteria (Figures 1B and 1C). These 42 bacterial species represent the common commensal bacteria in the human gut, with a median prevalence of 31.9% and an average relative abundance of 36.5% in metagenomic samples (n = 12,485).[34]

To uncover the genome features of these phage isolates, we sequenced and assembled phage genomes, yielding a gut phage isolate collection (GPIC) of nonredundant 209 high-quality phage genomes (Table S1). GPIC phage genome sizes range from 12,228 to 167,563 bp, with a GC content varying from 0.265 to 0.59 (Figure 1D; Table S1). The majority of GPIC phages (77%, n = 160) have genomes ranging in size from 30 to 60 kbp. All of these GPIC phage genomes were predicted to encode 13,488 genes in total; 9% of the genes have a functional annotation by Prokka,[36] and 26% of the genes can be annotated by mapping to the PFAM database using hmmscan (E value = 1e−5)[37] (Table S1). In particular, we identified 126 tRNA genes in the GPIC phage genomes (Table S1), which may be related to a bias in codon or amino acid usage between phages and their bacterial hosts.[38,39] DNA methylase genes were detected in 62 GPIC phages (Table S1). Nine of the GPIC phages encoded integrase genes (Table S1), suggesting that they were temperate phages. Based on the amino acid similarity of phage genes to the viral subset of the TrEMBL database,[15] all 209 GPIC phages were assigned to the class *Caudoviricetes*, which is dominant in the human gut virome.[9,40] Taken together, this work represents a large-scale gut phage cultivation effort that yielded 209 nonredundant phage isolates against 42 commensal human gut bacterial species.

Genomic diversity of GPIC phage isolates

To assess the GPIC phages, we compared the genomes with the RefSeq phage genomes in NCBI. Sequence alignment to the NCBI RefSeq phage sequences showed that ~86% (n = 179) of the GPIC phages lacked similar sequences (query coverage less than 30%) (Figure 2A), indicating that these phages are unique. In the Gut Phage Database (GPD),[11] the largest collection of putative phage sequences predicted from the human gut metagenomes, sequence alignment showed that ~74% (n = 154) of the GPIC phages had highly similar sequences (query coverage ≥ 70% and identity ≥ 60%) (Figure 2A), thus providing direct validation of the metagenome-assembled gut phages in GPD. Moreover, ~16% (n = 33) of the GPIC phages had no similar sequences (query coverage less than 30%) to the GPD genomes; these genomes expanded the reference catalog for future metagenomic analyses of human gut phages.

To assess the genomic diversity, we performed a vConTACT2-based analysis of the GPIC phages, NCBI RefSeq phages (n = 3,502), and GPD phages with a predicted complete genome (n = 14,936) (see STAR Methods).[42] The resulting gene-sharing network had 18,647 nodes linked by 1,515,942 edges, with each node representing a phage genome from GPIC, NCBI RefSeq, or GPD. The nodes formed 146 discrete components (DCs, i.e., subnetworks of inter-connected nodes) (Figure S2). Here, to focus the analysis on the GPIC phages, we only showed 8 DCs that contained the GPIC phages (*I–VIII*) (Figure 2B). DCs *I–III* include phages from all three different collections, DCs *V–VIII* include phages from both GPD (n = 89) and GPIC (n = 80), and DC *IV* includes phages from NCBI RefSeq (n = 13) and GPIC (n = 3) (Figure 2B). In particular, DCs *V–VIII* lack any NCBI RefSeq phages, implying that these phages were not previously isolated. Bacterial hosts of the phages in DCs *V–VIII* include 5 genera: *Bacteroides*, *Phocaeicola*, *Parabacteroides*, *Bifidobacterium*, and *Eggerthella* (Figure 2B), all of which are highly prevalent and abundant in the human gut.[18,34] Some bacteria species within the same genera had distinct phages with high genomic diversity in different DCs (Figure 2B). For example, phages targeting *Bacteroides* were found in three different DCs (*I*, *V*, and *VI*), phages targeting *Bifidobacterium* in two different DCs (*II* and *VII*), and phages targeting *Clostridium* in two different DCs (*I* and *II*). Despite this, the phage congregation in the network is mostly consistent with the phylum-level taxonomy of their hosts (Figure S2).

Based on the vConTACT2 result, all GPIC phages were grouped and distributed into 45 different viral clusters (VCs, Table S1), which approximates viral genus-level taxonomy.[42] Only 11 VCs of them, including 33 GPIC phages (~16%), contained the sequences from the NCBI RefSeq phages (Table S1). The remaining 176 GPIC phage isolates (~84%) that spread across 34 VCs were not classified using this method, indicating that these phages represent the newly isolated members of as-yet undescribed gut phage genera.

Small-genome phages of the *Salasmaviridae*-family-infecting Gram-positive human gut bacteria

We noticed that all phage genomes in DC *II* (Figure 2B) had small genome sizes (less than 28 kb), with 98.2% (535 out of 545) having genome sizes between 12 and 20 kb (Figure 3A). This DC contained 545 phages, including 22 GPIC phages, 10 NCBI RefSeq phages, and 513 GPD phages, which were distributed

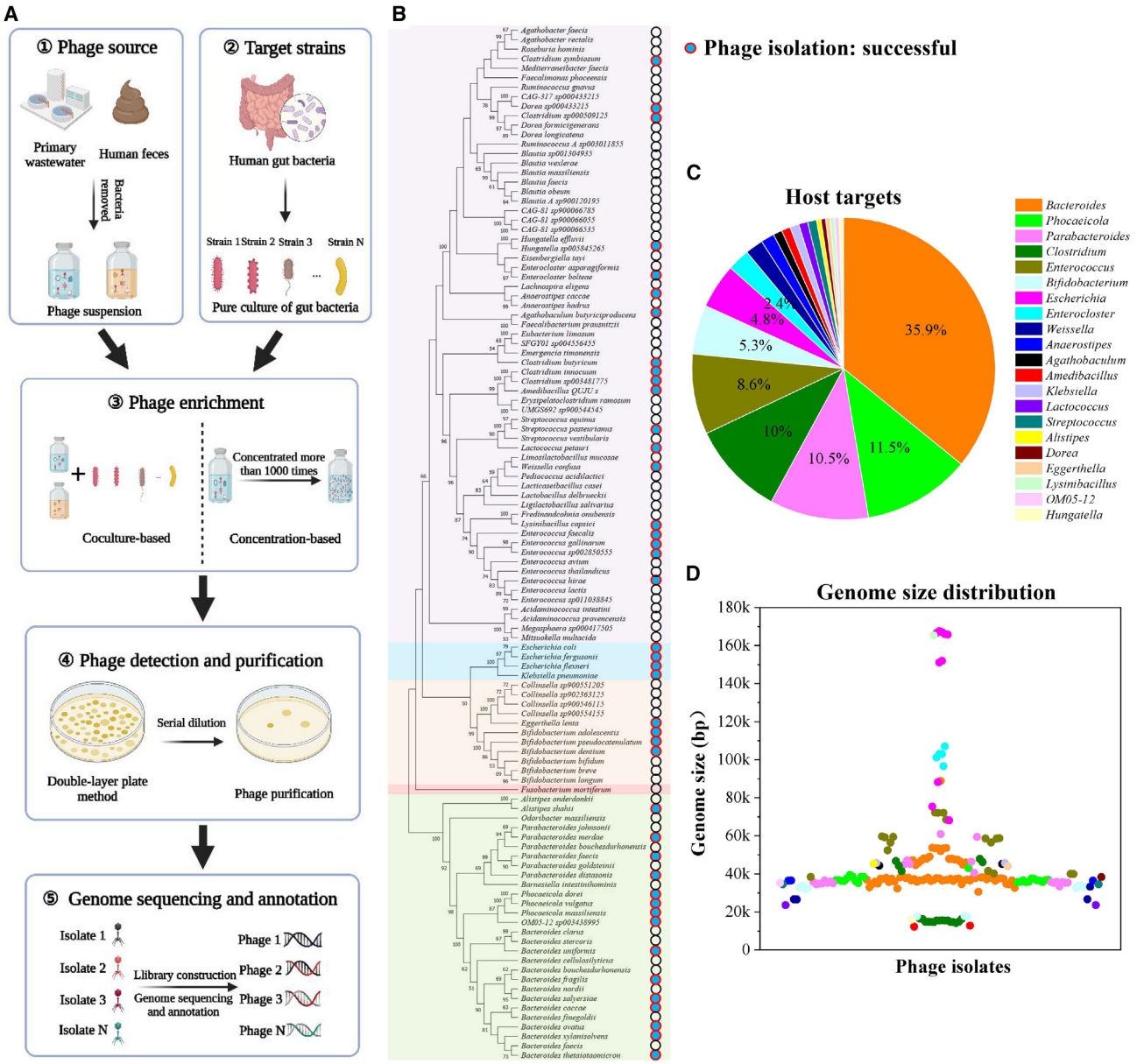


Figure 1. Systematic cultivation of human gut bacteriophages

(A)The workflow for phage isolation, purification, and genomic analysis. Primary wastewater from two municipal sewage treatment plants and healthy human feces were used for phage isolation. All the strains used in this study were isolated from human feces. Two methods for phage enrichment were used, including coculture with target strains and concentration by tangential flow and protein concentrators. The detection and purification of phage from phage-enriched samples used the double-layer plate method. The genomes of all purified phages were extracted, sequenced, and annotated (see STAR Methods, Figure S1, and Table S1).

(B) Phylogenetic tree of 16S rRNA genes from all tested bacterial species. Bacterial species with phage isolates are marked with blue dots on the right, and others are marked with white dots. The phylogenetic tree was constructed using the maximum likelihood method and the Tamura-Nei mode with MEGA X[35] (see STAR Methods).

(C) The host distribution of the 209 GPIC phages.

(D) The genome size distribution of the GPIC phages. A small dot indicates a phage genome, and different colors indicate the phage-targeted bacteria at a genus level.

in 83 VCs (Table S2). The bacterial hosts of the GPIC phages in DC *II* belonged to 4 genera, including *Clostridium*, *Bifidobacterium, Hungatella*, and *Amedibacillus*, and these phages were distributed across 6 different VCs. The hosts of the NCBI RefSeq phages in DC *II* are *Bacillus* (8 phages, including the phage phi29,[43] in 2 VCs) and *Bifidobacterium* (2 phages from honey bees' gut in the same VC).[44] The CRISPR spacer-based host prediction[45] (STAR Methods) showed that the hosts of the

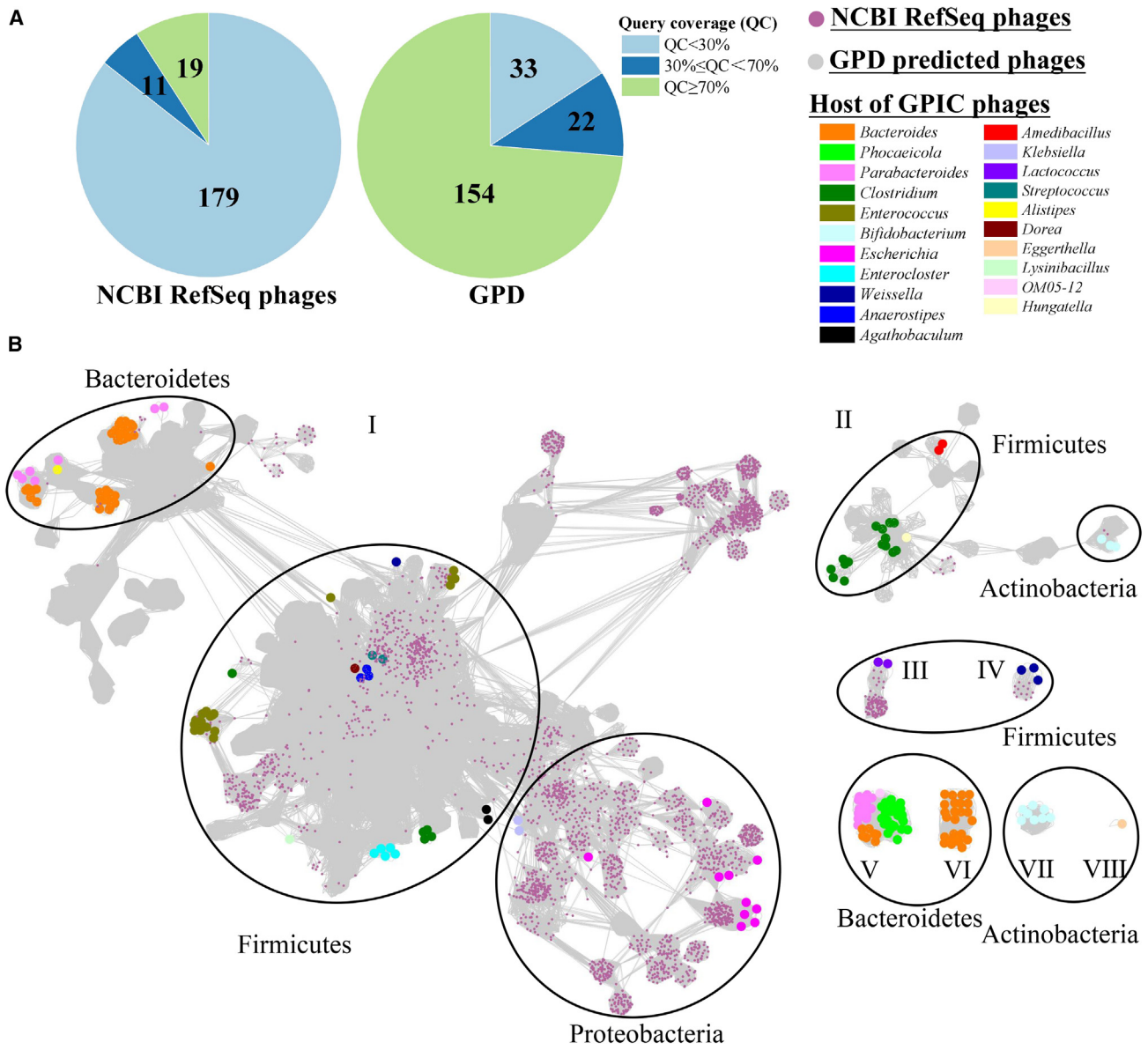


Figure 2. Genomic diversity of the GPIC phages

(A) The proportion of matched GPIC phages in the NCBI nucleotide collection database and the Gut Phage Database (GPD). The genome sequences of the phage isolates were compared with the phage sequences in the NCBI and GPD databases, respectively (identity ≥ 60%). Sequence alignment was performed under the default parameters of Blast+ 2.11.0.[41]

(B) Viral cluster analysis of the GPIC phages based on vConTACT2. A total of 18,653 phage genome sequences were included in the network, of which 3,502 are reference phage genome sequences in the NCBI, 14,936 are complete phage genome sequences predicted in the GPD database, and 209 are GPIC phage isolates. They were clustered together according to their protein content and protein similarity using vConTACT2 with the default sets.[42] Only the nodes related to our phage isolates are kept in the figure. The complete network is shown in Figure S2. Reference phage genomes from the NCBI reference genome database are labeled with lavender blue color. Phage genomes from the GPD database are labeled with light gray color. The 209 GPIC phage genomes are labeled with different colors based on their host taxonomy. See also STAR Methods and Figure S2.

GPD phages in DC *II* were from 31 genera—Firmicutes (25 genera), Actinomyces (5 genera), and Tenericutes (1 genus) (Table S2). Interestingly, all of these bacterial hosts are Gram positive (Table S2).

According to the criteria for the classification of tailed phages by the International Committee for the Taxonomy of Viruses (ICTV),[48,49] all of these phages were categorized as belonging to the *Salasmaviridae* family because they formed a monophyletic group (DC *II* in Figure 2B) and shared a significant number of orthologous genes (n = 2, >10% of genome) with 8 NCBI RefSeq *Salasmaviridae* phages (e.g., *Bacillus* phage phi29). To further reveal the genomic diversity of the *Salasmaviridae* phages, we performed a shared gene-based heatmap analysis on all 545 genomes in DC *II*. The heatmap indicated that all these phages

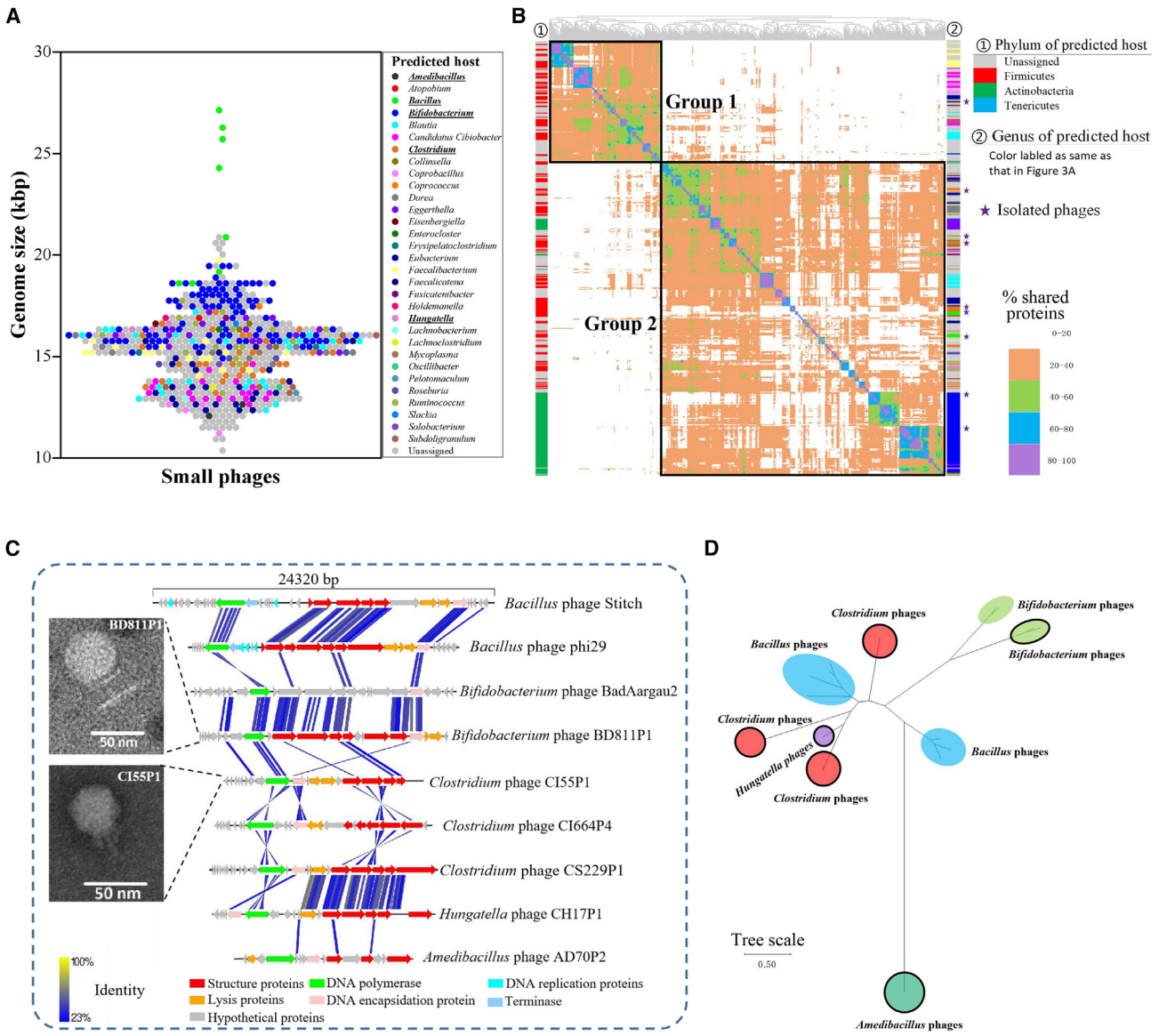


Figure 3. Characterization of the *Salasmaviridae* phages

(A) The genome size range of the *Salasmaviridae* phages. A total of 545 phage genome sequences were included in DC IIt, of which 22 are GPIC phages, 10 are from the NCBI RefSeq, and 513 are complete phage genome sequences predicted in the GPD database.

(B) A heatmap based on shared genes of the *Salasmaviridae* phages. The CRISPR spacer-based host prediction[45] showed that the hosts of this branched phage include bacteria from 31 genera—Firmicutes (25 genera), Actinomyces (5 genera), and Tenericutes (1 genus). The orthologous proteins were identified by OrthoFinder with the default set.[46]

(C) Genome-wide comparison of 9 selected *Salasmaviridae* phage isolates and the phage morphology of two representative GPIC *Salasmaviridae* phages. Genome comparison and visualization were performed using Easyfig[47] in which the alignment was performed using the default parameters of Blastn in the software.

(D) Unrooted maximum-likelihood phylogenetic tree of the *Salasmaviridae* phages. The tree was constructed based on the phage DNA polymerases of the isolated phages, including 22 GPIC phages and 10 NCBI reference phages. The evolutionary history was inferred by using the maximum likelihood method and a JTT matrix-based model with MEGA X.[35] The GPIC phages are marked in a black circle. See also Figures S3A and S3B and Table S2.

formed two subfamilies. The proportion of orthologous proteins shared by these two subfamilies was less than 20% (Figure 3B). The subfamily I (153 genomes, 16 VCs) contains only two isolated *Amedibacillus*-infecting GPIC phages, AM70P1 and AM70P2, and the subfamily II (392 genomes, 67 VCs) contains 30 isolated phages infecting *Clostridium*, *Hungatella*, *Bacillus*, and *Bifidobacterium* (Figure 3B; Table S2). Based on the host information of the phages, of all VCs in DC *II*, 87% (n = 72) had hosts limited to a single bacterial genus.

To characterize the genome organization of the *Salasmaviridae* phages, 6 GPIC phages and 3 NCBI RefSeq phages (representing 9 VCs in DC *II*) were selected (Figure 3C). Annotation revealed that

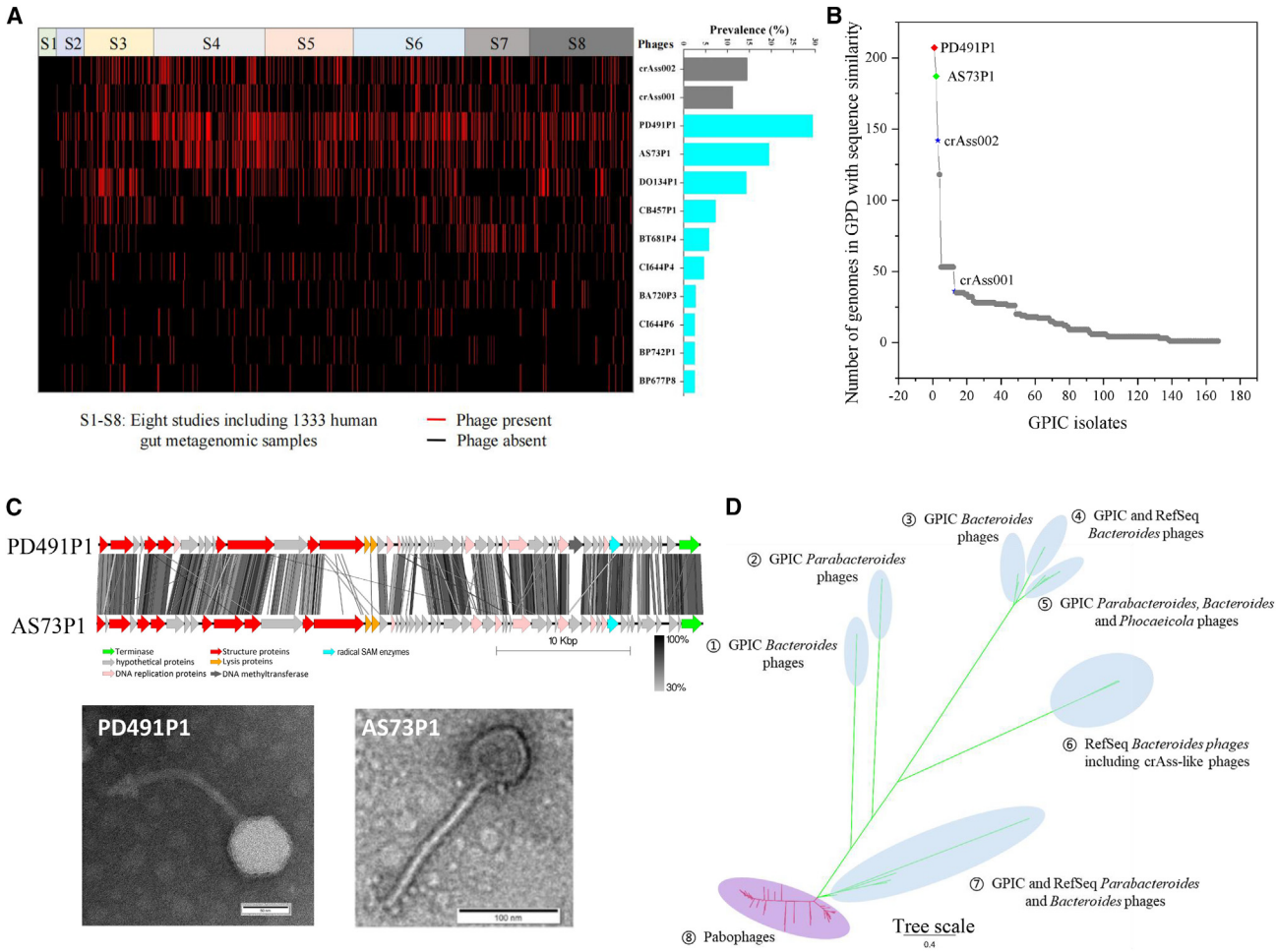


Figure 4. Characterization of the candidate family *Paboviridae* phages

(A) Distribution of the GPIC phage isolates in the human gut. Top 10 highly prevalent GPIC phages, crAss001 and crAss002 in 1,333 human gut samples from 8 studies. A phage with genome breadth coverage by the reads in a sample greater than 30% is considered to be present in the sample. The phage presence is indicated by a red line in the graph (see STAR Methods).

(B) Number of the GPD phage genomes similar to those of the GPIC phages, including PD491P1, AS73P1, crAss001, and crAss002. The phages in the GPD database were selected based on sequence similarity (sequence identity $\geq$ 60% and breadth coverage $\geq$ 70%) (see STAR Methods).

(C) Genome-wide comparison and phage morphology of the GPIC phages PD491P1 and AS73P1. The genome can be divided into four main regions encoding the genes for DNA processing, lysis and structure-related proteins, and hypothetical proteins, respectively. Genome comparison and visualization were performed using Easyfig[47] in which the alignment was performed using the default parameters of Blastn in the software (see STAR Methods).

(D) Phylogenetic tree of the isolated phages infecting the strains of the *Bacteroidetes* phylum and the candidate family *Paboviridae* phages. The tree was constructed based on the large terminase subunits using the MAFFT alignment.[52] The phages include isolated phages infecting the strains of the *Bacteroidetes* phylum and all the candidate family *Paboviridae* phages from the GPD database based on the identity of the large terminase subunit sequences with those of the GPIC phages PD491P1 and AS73P1. The genus-level taxonomy of bacterial hosts of these phages is marked. See also Figures S3C–S3F and Table S3.

the *Salasmaviridae* phage genomes had a similar gene organization with 4 distinct regions (Figure 3C), encoding genes for structural proteins, lysis-related proteins, proteins involved in DNA processing, and hypothetical proteins, respectively. Two GPIC phages (*Bifidobacterium* phage BD811P1 and *Clostridium* phage CI55P1) with small genomes (17,657 and 14,205 bp) were selected for transmission electron microscopy (TEM) and dynamics of infection, showing that both were short-tailed phages with head diameters of 50 and 36 nm, respectively, and exhibiting significant inhibitory effects on their bacterial hosts (Figures 3C and S3A). Based on the accurate host information of the isolated *Salasmaviridae* phages (22 GPIC phages and 10 NCBI RefSeq phages), we constructed phylogenetic trees for these phages using four conserved orthologous proteins, including DNA polymerase B, DNA encapsidation protein, major capsid protein, and portal proteins (Figures 3D and S3B). All of these trees clustered the 32 phages into 9 groups in agreement with the vConTACT2-based result; all phages in each group had hosts limited to a single bacterial genus.

An undescribed and highly prevalent phage family in the human gut

We determined the prevalence of 209 GPIC phages and 2 isolated crAss-like phages (crAss001[25] and crAss002[50]) in 1,333

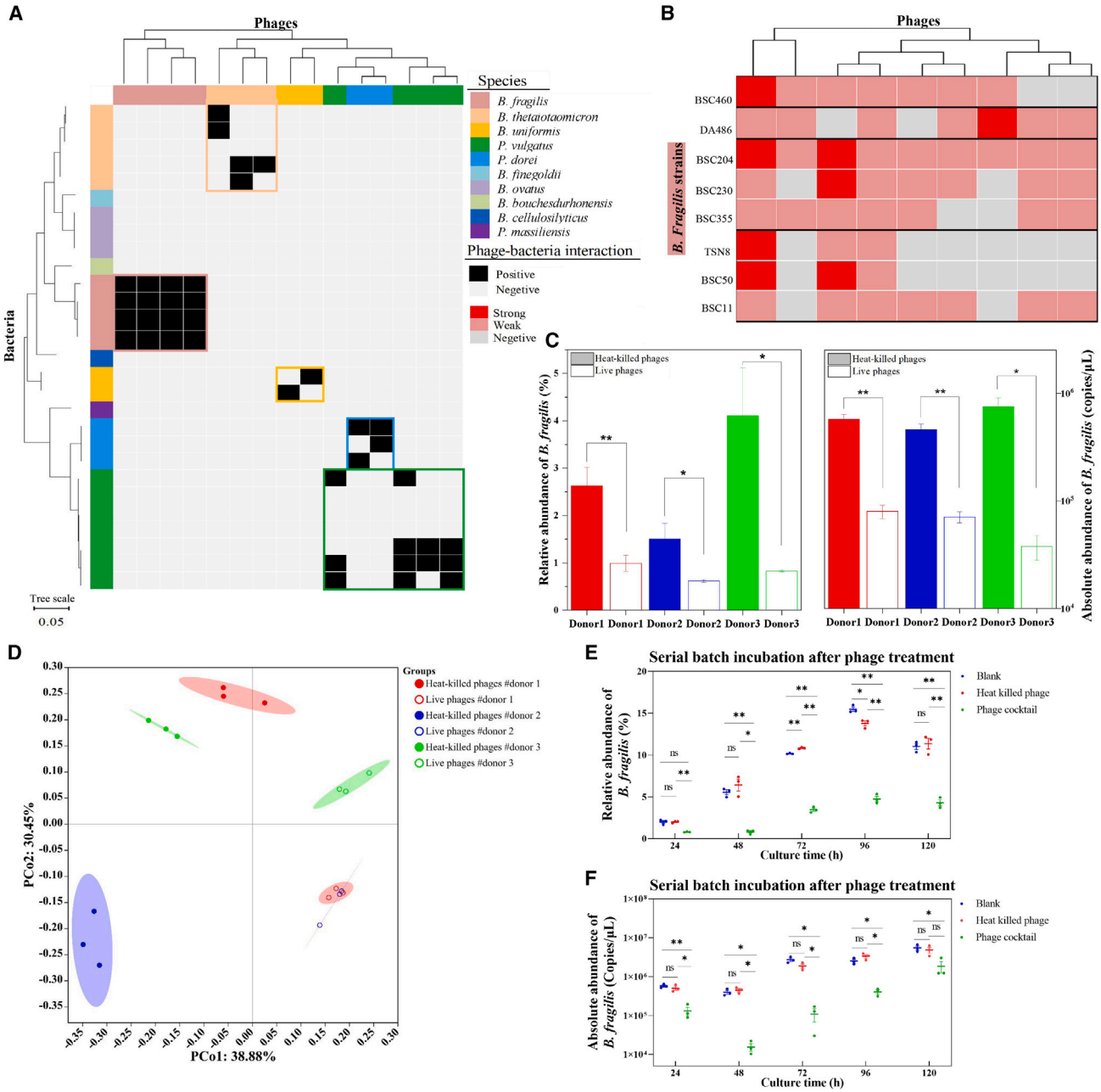


Figure 5. Host specificity of the human gut phages

(A) Cross-species infection assay. Rows represent *Bacteroides* strains; columns represent *Bacteroides* phages. Left: phylogenetic tree of *Bacteroides* strains based on the 16S rRNA gene. The phylogenetic tree was constructed by using the maximum likelihood method and the Tamura-Nei mode with MEGA X.[35] Top: cluster of *Bacteroides* phages based on shared genes (see Figure S4). The black square indicates that the corresponding strain can be infected by the phage, and the gray square indicates that it cannot be infected.

(B) Phage infection assay on *B. fragilis* strains isolated from different individuals. Rows represent *Bacteroides* strains from four donors; columns (top) represent *B. fragilis* phages, including phages BF503P1, BF494P1, BF344P1, BF766P4, BF766P1, BF702P1, BF486P1, BF695P2, and BF698P1. The light red square indicates that the phage has a strong or weak effect on the corresponding strain, and the gray square indicates that the phage does not affect the corresponding strain. Strong: the phage can form a plaque with a clear and large zone on the plate. Weak: the phage can form a plaque with a vague zone. Negative: no plaque formed. (See STAR Methods and Figure S4.)

(C) Alteration of relative abundance and absolute abundance of *B. fragilis* in the *in vitro* culture of human gut microbiome. Fecal samples from 3 donors were cultured with the phage cocktail or the heat-killed phages *in vitro* and were sampled at 24 h for metagenomic sequencing.

(D) Alteration of gut microbiome in response to phage treatment *in vitro*. PCoA (principal coordinate analysis) based on Jaccard dissimilarity matrices showed significant differences between the cultures with the heat-killed phage treatment and the phage cocktail treatment.

(E and F) The relative abundance (E) and the absolute abundance (F) of *B. fragilis* in the cultures of donor #1 for 120 h by subcultivation every 24 h.

(legend continued on next page)

publicly available human gut metagenomes from eight studies (Figure 4A; see STAR Methods). By mapping metagenomic reads to the phage genomes (coverage ≥ 30%), we discovered that the prevalence of 2 GPIC phages (*Parabacteroide* phage PD491P1 29.5%, and *Alistipes* phage AS73P1 19.5%) was higher than that of the two isolated crAss-like phages (crAss001, 11.3% and crAss002, 14.5%) but lower than that of the prototype crAss-phage (41%) (NC_024711). We did not detect any temperate phage elements (e.g., *integrase*) in the phage genomes, nor did we detect the phage genomes in the genomes of their hosts, implying that they were lytic phages. The phage PD491P1 was detected in the samples in all countries except Tanzania (Figure 4A), with the highest prevalence reaching 43.0% in Denmark-Spain samples (Figure S3C). The top 10 GPIC phages with a high prevalence in the human gut were differently distributed between the Tanzanian and the populations from other countries (ANOSIM, $p < 0.0001$), as we can see that 9 of 10 of these phage isolates were not detected in the metagenomic sequencing datasets from the Tanzanian (Figure S3C).[51] The Tanzanian dataset was sampled from Hadza people in Africa, who typically live in rural settings distant from urban centers. This implied that the living habitat could affect the human gut microbiome. We also searched in the GPD database for the highly similar sequences (coverage ≥ 70% and identity ≥ 60%) to the two GPIC phages, yielding a total of 207 (PD491P1) and 187 (AS73P1), respectively, when compared with the crAss-like phage isolates (36 for crAss001 and 142 for crAss002) (Figure 4B). This suggested that *Parabacteroide* phage PD491P1 and *Alistipes* phage AS73P1 represent members of the most prevalent isolated phages in the human gut so far.

The genome sequences of two GPIC phages, PD491P1 and AS73P1, infecting distinct hosts were highly similar, with a similar genome size (~45 kb) and 42 functional genes shared, accounting for 61% of their total number of genes (Figure 4C). The genomes are organized into 4 distinct regions encoding for structural proteins, lysis-related proteins, proteins involved in DNA replication processing, and hypothetical proteins. TEM showed that the morphology of *Parabacteroides* phage PD491P1 and *Alistipes* phage AS73P1 are both long-tailed phages with head diameters of 63 and 52 nm, respectively (Figure 4C). These two phages both exhibited modest inhibitory effects on their bacterial hosts (Figure S3D). In addition, we discovered genes for radical S-adenosylmethionine (SAM) enzymes in the genomes of these two phage genomes (Figure 4C), which are thought to be involved in promoting anaerobic bacterial metabolism during infection.[53] It is worth noting that a *Parabacteroides* phage (phage PDS1, MN929097) has been independently isolated and sequenced and has high sequence identity with phages PD491P1 (identity = 97.37%) and AS73P1 (identity = 82.82%) (Figure S3E).

We discovered 508 GPD genomes encoding the large terminase subunits with significant similarity to those of the phages PD491P1 and AS73P1 (E value < 1e−6) (Table S3). According to CRISPR spacer-based bacterial host prediction,[45] 82.2% of these phages (419 out of 510) were assigned to hosts belonging to 8 genera in the *Bacteroidales* order, namely, *Bacteroides*, *Alistipes*, *Odoribacter*, *Parabacteroide*, *Prevotella*, *Barnesiella*, *Butyricimonas*, and *Dysgonomonas* (Table S3). We constructed a phylogenetic tree using the large terminase subunit to investigate the evolutionary relationship between these phages and the currently isolated phages that infect the phylum of *Bacteroidetes* (Figure 4D). These phages significantly diverged from other *Bacteroidetes* phages and formed a distinct clade.

To classify this clade of phages according to the criteria established by the ICTV, we performed a shared gene-based heatmap analysis of all complete phage genomes (n = 258), resulting in 4 distinct groups. These 258 phages were assigned to 4 VCs by vConTACT2, have an average genome size of 44.27 kbp, and encode 56 genes on average. These groups shared more than 28% (n = 16) but less than 70% of homologous proteins (Figure S3C). The members of G1 (VC_117_0 and VC_118_0) had the hosts of *Bacteroides*, *Odoribacter*, *Parabacteroides*, and *Alistipes*; G2 (VC_617_0) had the hosts of *Bacteroides*, *Prevotella*, and *Dysgonomonas*; G3 (VC_120_0) had the hosts of *Bacteroides*, *Prevotella*, and *Odoribacter*; and G4 (VC_120_0) had the hosts of *Odoribacter*. In 95% of these 258 complete phages, 30 orthologous proteins were identified, 24 of which were annotated as unknown function proteins, indicating that the function of these conserved proteins was specific to the phages of this clade but needs further investigation. Thus, accordingly, we tentatively classify these phages as an undescribed family in the class *Caudoviricetes*, named as the candidate *Paboviridae* family, which infects *Parabacteroides*, *Alistipes*, and other bacterial genera within the *Bacteroidales* order in the human gut.

The host specificity of *Bacteroides* and *Parabacteroides* phages

Bioinformatic analysis indicated that one-third of human gut phages likely have a broad host range and are not limited to a single host species.[11] To explore the extent to which the human gut phages can infect hosts across distinct species, we conducted experiments to determine the host range of a subset of GPIC phages. First, to test if the phages could infect their hosts across species, we selected 15 GPIC *Bacteroides* phages and 28 strains from 9 *Bacteroides* species to perform a cross-infection experiment (Figures 5A and S4A). However, in all of these tests, we observed that none of the phages could infect cross-species bacterial strains (Figure 5A). Then, to test if the phages could infect cross-species bacterial strains derived from the same individual donor, we determined the host range of 9 *B. fragilis* phages (Figures S4B and S4C) on all 58 *Bacteroides* isolates from two individual donors, 17 of which were *B. fragilis* strains, and 41 of which belonged to the other 5 *Bacteroides* species, showing that none of the phages can infect their hosts of different species (Table S4). A similar result was obtained for the *Parabacteroides* phages (Figures S4E and S4F; Table S4).

Blank, the culture was not treated. Heat-killed phages: the culture was treated with heat-killed phages. The phage cocktail consists of 8 phages targeting *B. fragilis* strains, including phages BF344P1, BF702P1, BF766P4, BF766P1, BF695P2, BF698P1, BF486P1, and BF494P1. Relative abundance and absolute abundance of *B. fragilis* were estimated using Kraken2[54] and the qPCR assay, respectively. Scale bars represent ±SEM. p values between the cultures with heat-killed phages and the living phage cocktail were determined by t test with a two-tailed test. *$p < 0.05$, **$p < 0.01$, ns: no significance. Bonferroni correction was used for multiple hypotheses testing.[55] See STAR Methods and Figure S5.

Human gut bacteria exhibit strain-level diversity in gene content and function.[56–59] In this study, we discovered strain-level diversity in phage susceptibility for all tested bacterial species, including *B. fragilis*, *B. thetaiotaomicron*, *Phocaeicola dorei*, and *P. vulgatus* (Figure 5A; Table S4). In particular, 8 *B. fragilis* strains from four individuals showed different patterns in phage susceptibility (Figure 5B), even between the strains with a high ANI (average nucleotide identity) > 99.99% (Figure S4D). For example, strains BSC50, TSN8, and BSC11 were isolated from the same individual donor; the former two were infected by 3 phages with varying degrees of infection efficiency, whereas strain BSC11 was infected by 7 phages. A similar result was observed in the case of *P. merdae* (Figures S4G–S4I). This can be explained by phase variation of capsular polysaccharide operons and/or bacterial receptor gene regulation for these bacterial strains.[60,61]

The high host specificity of phages allows for precise microbiota modulation.[62,63] To demonstrate the feasibility of using GPIC phages to modulate human gut microbiota, we selected 8 *B. fragilis* phages with distinct genome sequences (Figure S4C) and varying host specificity (Figures 5B and S5A) to prepare an 8-phage cocktail with a broad host range for *B. fragilis* strains. This phage cocktail was added to *in vitro* culture[64] of gut microbial communities derived from stool samples of three donors (see STAR Methods). The result indicated that both the absolute abundance and the relative abundance of *B. fragilis* in the community were significantly reduced under 24 h treatment of the phage cocktail, in comparison with the control group ($p < 0.05$) (Figure 5C). Interestingly, we found that the composition of the gut microbiome altered substantially in response to the phage cocktail treatment (Figure 5D; Table S5), but no significant difference in alpha-diversity was observed for all three samples. Furthermore, using the *in vitro* culture of the gut microbial community derived from the stool sample of donor#1, we compared the effects of the phage cocktail and the individual phages under longer incubation time (up to 72 h). We found that the phage cocktail significantly reduced the abundance of *B. fragilis* compared with the control group (blank or heat-killed phages) ($p < 0.05$) (Figure S5B), and the knockdown effect of the phage cocktail was significantly stronger than individual phages (Figure S5C). Finally, we performed serial cultivation of stool (donor #1)-derived *in vitro* communities for 120 h (subcultivation every 24 h) after phage treatment. The phage cocktail significantly reduced the abundance of *B. fragilis* throughout the 120 h experiment (Figures 5E and 5F), showing a long-lasting knockdown effect on the target species. It is worth noting that the abundance of *B. fragilis* slightly increased during the late stages of incubation cultivation (Figures 5E and 5F), hinting at potential phage-sensitivity of strain-level variations in phage susceptibility[65] and/or rapid bacteria-phage co-evolution.[66]

DISCUSSION

A massive number of putative phage genomes have been identified using bioinformatics approaches from human gut metagenomic sequencing data.[11,13,14] The culture and isolation of human gut phages, especially for commensal anaerobic gut bacteria,[26] are thus of considerable interest. In this study, we obtained 209 isolated phages for 42 gut bacterial species that represent a large phylogenetic diversity of human gut bacteria across 4 phyla. This effort expands the existing collections of human gut phage isolates with high-quality reference genomes, and comprehensive analyses provide additional findings for the human gut virome when compared with the gut phage sequence databases predicted from the metagenomic data. For example, this study defines an undescribed family named as candidate family *Paboviridae* with a high prevalence in the human gut, showing that phage cultivation can be essential to unraveling the dark matter of gut phages. The experiment in which we successfully knocked down the targeted species in the microbiome *in vitro* using a phage cocktail highlights the value of the isolated gut phages in the microbiome study.

In this study, we demonstrated the feasibility of isolating specific phages for a variety of commensal gut bacteria derived from the human gut. Our study demonstrated that a large number of gut bacterial phages can be isolated and cultured using the double agar plate method. The failure of phage isolation for some bacterial species indicated that the workflow could be improved in the future for phages that do not form plaques (e.g., crAss002 phages[50]). Interestingly, the majority (199/209) of the GPIC phages were isolated from the sewage wastewater samples (Table S1). We conducted a metagenomic analysis of the samples for phage isolation, and the result indicated that the diversity of the phage population (VCs, n = 100) in the sewage wastewater samples was greater than that in the human fecal samples (VCs, average n = 68.6) (Figure S1C). This discrepancy may not be sufficient to explain the bias in isolating phages from these two types of samples. It has been observed that bacteria and their phages co-exist in the human gut.[15] A study has shown that phase-variable capsular polysaccharides and lipoproteins can modify gut phage susceptibility in *B. thetaiotaomicron*.[60] Moreover, in our study, the host range determination of the phages against the strains of a specific species of *Bacteroides* and *Parabacteroides* demonstrated that bacterial strains sharing a high identity (ANI > 99.9%) genome in an individual human gut have variable susceptibility to phages. Thus, this evidence suggests that many gut bacteria have evolved to acquire resistance to co-existing phages. This could also explain why we failed to isolate phages for the gut bacteria from autologous feces despite comprehensive enrichment of the phage particles (Table S1).

Phage therapy has shown great potential for precise modulation of the human gut microbiota,[33,67] where phages were used to remove specific pathogens or pathobionts. The GPIC phage isolates provide a valuable resource for modulating the human gut microbiota. In this study, we showed that the relative abundance of *B. fragilis* in the *in vitro* cultured human gut microbiome was successfully knocked down by a phage cocktail. Enterotoxigenic *B. fragilis* is a candidate pathogen in inflammatory bowel disease and colorectal cancer.[68–70] The results presented here demonstrate that the GPIC phage isolates have potential applications for the removal of targeted pathobionts in the human gut microbiome. In addition, phage-based interventions will provide a unique tool to perturb the human gut microbiome and advance our understanding of the complex ecological network. However, we note that the targeted species may co-exist with phages after the initial decline in abundance (Figures 5E and 5F). This phenomenon has also been observed in animal experiments.[3,71] Potential mechanisms[72,73] include the regulation of

bacterial gene expression,[61] phase variation of capsular polysaccharide operons,[60] and strain-level diversity of the human gut species.[74] The introduction of additional phages in the cocktail and the development of engineered phages[75,76] may overcome this limitation in future applications.

In conclusion, GPIC is a unique resource that offers access to hundreds of phage isolates targeting human gut strains. We envision that GPIC will help elucidate the role of phages in the gut microbiome and support the application of phage-based microbiota modulation in scientific and therapeutic investigations.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacterial strains
  - Bacterial culture conditions
  - Ethics statement
- METHOD DETAILS
  - Phage isolation from sewage wastewater samples and human feces
  - Phage genome sequencing, assembly and annotation
  - Phage host range assays
  - Transmission Electron Microscopy
  - Phage enrichment from sewage wastewater samples and human feces for metagenomic sequencing
  - Comparative genomic analysis of phage genomes
  - Identification of phage orthologous proteins and classification of phage genomes
  - Distribution of phage isolates in the human gut
  - CRISPR spacer-based prediction of bacterial hosts
  - Design of a phage cocktail to reduce *B. fragilis* species in the human gut microbiome *in vitro*
  - Metagenomic sequencing and estimation of the relative abundance of *B. fragilis* in the microbiome
  - Estimation of the absolute abundance of *B. fragilis* by the qPCR assay
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chom.2023.03.013.

ACKNOWLEDGMENTS

This work received support from the National Key R&D Program of China (no. 2019YFA0906700), the Strategic Priority Research Program of the Chinese Academy of Sciences (no. XDB29050501), the Guangdong Provincial Key Laboratory of Synthetic Genomics (2019B030301006), the Shenzhen Institute of Synthetic Biology Scientific Research Program (no. JCHZ20200001), the National Natural Science Foundation of China (no. 31971513, no. 32061143023), and the Guangdong basic and applied basic research foundation (no. 2020A1515110184). We thank Chaobi Lei and the Shenzhen Infrastructure for Synthetic Biology for the assistance in the DNA library construction and extraction processes.

AUTHOR CONTRIBUTIONS

Conceptualization, J.S., L.D., and Y.M.; methodology, J.S., J.Z., L.M., Yanchen Li, Yake Li, C.L., X.K., Z.T., Z.Q., J.C., S.L., and L.Z.; investigation, J.S., J.Z., L.D., and Y.M.; writing – original draft, J.S., J.Z., L.W., L.D., and Y.M.; writing – review & editing, J.S., L.D., and Y.M.; funding acquisition, L.D. and Y.M.; resources, Z.C., Y.D., T.Z., and B.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 7, 2022
Revised: January 26, 2023
Accepted: March 8, 2023
Published: April 12, 2023

REFERENCES

1. Lepage, P., Colombet, J., Marteau, P., Sime-Ngando, T., Doré, J., and Leclerc, M. (2008). Dysbiosis in inflammatory bowel disease: a role for bacteriophages? Gut *57*, 424–425. https://doi.org/10.1136/gut.2007.134668.
2. Hoyles, L., McCartney, A.L., Neve, H., Gibson, G.R., Sanderson, J.D., Heller, K.J., and van Sinderen, D. (2014). Characterization of virus-like particles associated with the human faecal and caecal microbiota. Res. Microbiol. *165*, 803–812. https://doi.org/10.1016/j.resmic.2014.10.006.
3. Hsu, B.B., Gibson, T.E., Yeliseyev, V., Liu, Q., Lyon, L., 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.e5. https://doi.org/10.1016/j.chom.2019.05.001.
4. 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. USA *110*, 20236–20241. https://doi.org/10.1073/pnas.1319470110.
5. Rasmussen, T.S., Koefoed, A.K., Jakobsen, R.R., Deng, L., Castro-Mejía, J.L., Brunse, A., Neve, H., Vogensen, F.K., and Nielsen, D.S. (2020). Bacteriophage-mediated manipulation of the gut microbiome - promises and presents limitations. FEMS Microbiol. Rev. *44*, 507–521. https://doi.org/10.1093/femsre/fuaa020.
6. Clooney, A.G., Sutton, T.D.S., Shkoporov, A.N., Holohan, R.K., Daly, K.M., O'Regan, O., Ryan, F.J., Draper, L.A., Plevy, S.E., Ross, R.P., et al. (2019). Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. Cell Host Microbe *26*, 764–778.e5. https://doi.org/10.1016/j.chom.2019.10.009.
7. Gogokhia, L., Buhrke, K., Bell, R., Hoffman, B., Brown, D.G., Hanke-Gogokhia, C., Ajami, N.J., Wong, M.C., Ghazaryan, A., Valentine, J.F., et al. (2019). Expansion of bacteriophages is linked to aggravated intestinal inflammation and colitis. Cell Host Microbe *25*, 285–299.e8. https://doi.org/10.1016/j.chom.2019.01.008.
8. Norman, J.M., Handley, S.A., Baldridge, 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 *160*, 447–460. https://doi.org/10.1016/j.cell.2015.01.002.
9. Liang, G., and Bushman, F.D. (2021). The human virome: assembly, composition and host interactions. Nat. Rev. Microbiol. *19*, 514–527. https://doi.org/10.1038/s41579-021-00536-5.
10. Garmaeva, S., Sinha, T., Kurilshikov, A., Fu, J., Wijmenga, C., and Zhernakova, A. (2019). Studying the gut virome in the metagenomic era: challenges and perspectives. BMC Biol. *17*, 84. https://doi.org/10.1186/s12915-019-0704-y.
11. Camarillo-Guerrero, L.F., Almeida, A., Rangel-Pineros, G., Finn, R.D., and Lawley, T.D. (2021). Massive expansion of human gut bacteriophage diversity. Cell *184*, 1098–1109.e9. https://doi.org/10.1016/j.cell.2021.01.029.

12. Gregory, A.C., Zablocki, O., Zayed, A.A., Howell, A., Bolduc, B., and Sullivan, M.B. (2020). The gut virome database reveals age-dependent patterns of virome diversity in the human gut. Cell Host Microbe *28*, 724–740.e8. https://doi.org/10.1016/j.chom.2020.08.003.

13. Roux, S., Páez-Espino, D., Chen, I.A., Palaniappan, K., Ratner, A., Chu, K., Reddy, T.B.K., Nayfach, S., Schulz, F., Call, L., et al. (2021). IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. Nucleic Acids Res. *49*, D764–D775. https://doi.org/10.1093/nar/gkaa946.

14. Nayfach, S., Páez-Espino, D., Call, L., Low, S.J., Sberro, H., Ivanova, N.N., Proal, A.D., Fischbach, M.A., Bhatt, A.S., Hugenholtz, P., et al. (2021). Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. Nat. Microbiol. *6*, 960–970. https://doi.org/10.1038/s41564-021-00928-6.

15. 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.

16. Lagier, J.C., Dubourg, G., Million, M., Cadoret, F., Bilen, M., Fenollar, F., Levasseur, A., Rolain, J.M., Fournier, P.E., and Raoult, D. (2018). Culturing the human microbiota and culturomics. Nat. Rev. Microbiol. *16*, 540–550. https://doi.org/10.1038/s41579-018-0041-0.

17. Groussin, M., Poyet, M., Sistiaga, A., Kearney, S.M., Moniz, K., Noel, M., Hooker, J., Gibbons, S.M., Segurel, L., Froment, A., et al. (2021). Elevated rates of horizontal gene transfer in the industrialized human microbiome. Cell *184*, 2053–2067.e18. https://doi.org/10.1016/j.cell.2021.02.052.

18. Forster, S.C., Kumar, N., Anonye, B.O., Almeida, A., Viciani, E., Stares, M.D., Dunn, M., Mkandawire, T.T., Zhu, A., Shao, Y., et al. (2019). A human gut bacterial genome and culture collection for improved metagenomic analyses. Nat. Biotechnol. *37*, 186–192. https://doi.org/10.1038/s41587-018-0009-7.

19. Pinto, G., Almeida, C., and Azeredo, J. (2020). Bacteriophages to control Shiga toxin-producing E. coli - safety and regulatory challenges. Crit. Rev. Biotechnol. *40*, 1081–1097. https://doi.org/10.1080/07388551.2020.1805719.

20. Canfield, G.S., and Duerkop, B.A. (2020). Molecular mechanisms of enterococcal-bacteriophage interactions and implications for human health. Curr. Opin. Microbiol. *56*, 38–44. https://doi.org/10.1016/j.mib.2020.06.003.

21. Nale, J.Y., Spencer, J., Hargreaves, K.R., Buckley, A.M., Trzepiński, P., Douce, G.R., and Clokie, M.R. (2016). Bacteriophage combinations significantly reduce Clostridium difficile growth in vitro and proliferation in vivo. Antimicrob. Agents Chemother. *60*, 968–981. https://doi.org/10.1128/AAC.01774-15.

22. Tartera, C., Lucena, F., and Jofre, J. (1989). Human origin of Bacteroides fragilis bacteriophages present in the environment. Appl. Environ. Microbiol. *55*, 2696–2701. https://doi.org/10.1128/aem.55.10.2696-2701.1989.

23. Kai, M., Watanabe, S., Furuse, K., and Ozawa, A. (1985). Bacteroides bacteriophages isolated from human feces. Microbiol. Immunol. *29*, 895–899. https://doi.org/10.1111/j.1348-0421.1985.tb00891.x.

24. Hryckowian, A.J., Merrill, B.D., Porter, N.T., Van Treuren, W., Nelson, E.J., Garlena, R.A., Russell, D.A., Martens, E.C., and Sonnenburg, J.L. (2020). Bacteroides thetaiotaomicron-infecting bacteriophage isolates inform sequence-based host Range predictions. Cell Host Microbe *28*, 371–379.e5. https://doi.org/10.1016/j.chom.2020.06.011.

25. Shkoporov, A.N., Khokhlova, E.V., Fitzgerald, C.B., Stockdale, S.R., Draper, L.A., Ross, R.P., and Hill, C. (2018). ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. Phi. Nat. Commun. *9*, 4781. https://doi.org/10.1038/s41467-018-07225-7.

26. Shareefdeen, H., and Hynes, A.P. (2021). Does over a century of aerobic phage work provide a solid framework for the study of phages in the gut? Anaerobe *68*, 102319. https://doi.org/10.1016/j.anaerobe.2021.102319.

27. Fitzgerald, C.B., Shkoporov, A.N., Upadrasta, A., Khokhlova, E.V., Ross, R.P., and Hill, C. (2021). Probing the "dark matter" of the human gut Phageome: culture assisted metagenomics enables rapid discovery and host-linking for novel bacteriophages. Front. Cell. Infect. Microbiol. *11*, 616918. https://doi.org/10.3389/fcimb.2021.616918.

28. Sabino, J., Hirten, R.P., and Colombel, J.F. (2020). Review article: bacteriophages in gastroenterology-from biology to clinical applications. Aliment. Pharmacol. Ther. *51*, 53–63. https://doi.org/10.1111/apt.15557.

29. Duan, Y., Young, R., and Schnabl, B. (2022). Bacteriophages and their potential for treatment of gastrointestinal diseases. Nat. Rev. Gastroenterol. Hepatol. *19*, 135–144. https://doi.org/10.1038/s41575-021-00536-z.

30. Duan, Y., Llorente, C., Lang, S., Brandl, K., Chu, H., Jiang, L., White, R.C., Clarke, T.H., Nguyen, K., Torralba, M., et al. (2019). Bacteriophage targeting of gut bacterium attenuates alcoholic liver disease. Nature *575*, 505–511. https://doi.org/10.1038/s41586-019-1742-x.

31. Zheng, D.W., Dong, X., Pan, P., Chen, K.W., Fan, J.X., Cheng, S.X., and Zhang, X.Z. (2019). Phage-guided modulation of the gut microbiota of mouse models of colorectal cancer augments their responses to chemotherapy. Nat. Biomed. Eng. *3*, 717–728. https://doi.org/10.1038/s41551-019-0423-2.

32. Dong, X., Pan, P., Zheng, D.W., Bao, P., Zeng, X., and Zhang, X.Z. (2020). Bioinorganic hybrid bacteriophage for modulation of intestinal microbiota to remodel tumor-immune microenvironment against colorectal cancer. Sci. Adv. *6*, eaba1590. https://doi.org/10.1126/sciadv.aba1590.

33. Federici, S., Kredo-Russo, S., Valdés-Mas, R., Kviatcovsky, D., Weinstock, E., Matiuhin, Y., Silberberg, Y., Atarashi, K., Furuichi, M., Oka, A., et al. (2022). Targeted suppression of human IBD-associated gut microbiota commensals by phage consortia for treatment of intestinal inflammation. Cell *185*, 2879–2898.e24. https://doi.org/10.1016/j.cell.2022.07.003.

34. Wu, S., Sun, C., Li, Y., Wang, T., Jia, L., Lai, S., Yang, Y., Luo, P., Dai, D., Yang, Y.Q., et al. (2020). GMrepo: a database of curated and consistently annotated human gut metagenomes. Nucleic Acids Res. *48*, D545–D553. https://doi.org/10.1093/nar/gkz764.

35. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. *35*, 1547–1549. https://doi.org/10.1093/molbev/msy096.

36. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

37. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: the protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. https://doi.org/10.1093/nar/gkaa913.

38. Bailly-Bechet, M., Vergassola, M., and Rocha, E. (2007). Causes for the intriguing presence of tRNAs in phages. Genome Res. *17*, 1486–1495.

39. Delesalle, V.A., Tanke, N.T., Vill, A.C., and Krukonis, G.P. (2016). Testing hypotheses for the presence of tRNA genes in mycobacteriophage genomes. Bacteriophage *6*, e1219441.

40. Mayneris-Perxachs, J., Castells-Nobau, A., Arnoriaga-Rodríguez, M., Garre-Olmo, J., Puig, J., Ramos, R., Martínez-Hernández, F., Burokas, A., Coll, C., Moreno-Navarrete, J.M., et al. (2022). Caudovirales bacteriophages are associated with improved executive function and memory in flies, mice, and humans. Cell Host Microbe *30*, 340–356.e8. https://doi.org/10.1016/j.chom.2022.01.013.

41. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). Blast+: architecture and applications. BMC Bioinformatics *10*, 421. https://doi.org/10.1186/1471-2105-10-421.

42. Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., Brister, J.R., Kropinski, A.M., Krupovic, M., Lavigne, R., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat. Biotechnol. *37*, 632–639. https://doi.org/10.1038/s41587-019-0100-8.

43. Meijer, W.J., Horcajadas, J.A., and Salas, M. (2001). Phi29 family of phages. Microbiol. Mol. Biol. Rev.: MMBR *65*, 261–287. https://doi.org/10.1128/MMBR.65.2.261-287.2001.

44. Bonilla-Rosso, G., Steiner, T., Wichmann, F., Bexkens, E., and Engel, P. (2020). Honey bees harbor a diverse gut virome engaging in nested strain-level interactions with the microbiota. Proc. Natl. Acad. Sci. USA *117*, 7355–7362. https://doi.org/10.1073/pnas.2000228117.

45. Dion, M.B., Plante, P.L., Zufferey, E., Shah, S.A., Corbeil, J., and Moineau, S. (2021). Streamlining CRISPR spacer-based bacterial host predictions to decipher the viral dark matter. Nucleic Acids Res. *49*, 3127–3138. https://doi.org/10.1093/nar/gkab133.

46. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

47. Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011). Easyfig: a genome comparison visualizer. Bioinformatics *27*, 1009–1010. https://doi.org/10.1093/bioinformatics/btr039.

48. Turner, D., Kropinski, A.M., and Adriaenssens, E.M. (2021). A roadmap for genome-based phage taxonomy. Viruses *13*, 506. https://doi.org/10.3390/v13030506.

49. Adriaenssens, E., and Brister, J.R. (2017). How to name and classify your phage: an informal guide. Viruses *9*, 70. https://doi.org/10.3390/v9040070.

50. Guerin, E., Shkoporov, A.N., Stockdale, S.R., Comas, J.C., Khokhlova, E.V., Clooney, A.G., Daly, K.M., Draper, L.A., Stephens, N., Scholz, D., et al. (2021). Isolation and characterisation of PhicrAss002, a crAss-like phage from the human gut that infects Bacteroides xylanisolvens. Microbiome *9*, 89. https://doi.org/10.1186/s40168-021-01036-7.

51. Smits, S.A., Leach, J., Sonnenburg, E.D., Gonzalez, C.G., Lichtman, J.S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J.E., et al. (2017). Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. Science *357*, 802–806. https://doi.org/10.1126/science.aan4834.

52. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

53. Timms, A.R., Cambray-Young, J., Scott, A.E., Petty, N.K., Connerton, P.L., Clarke, L., Seeger, K., Quail, M., Cummings, N., Maskell, D.J., et al. (2010). Evidence for a lineage of virulent bacteriophages that target Campylobacter. BMC Genomics *11*, 214. https://doi.org/10.1186/1471-2164-11-214.

54. Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. *20*, 257. https://doi.org/10.1186/s13059-019-1891-0.

55. Ranstam, J. (2016). Multiple P-values and Bonferroni correction. Osteoarthr. Cartil. *24*, 763–764. https://doi.org/10.1016/j.joca.2016.01.008.

56. Park, S.Y., Rao, C., Coyte, K.Z., Kuziel, G.A., Zhang, Y., Huang, W., Franzosa, E.A., Weng, J.K., Huttenhower, C., and Rakoff-Nahoum, S. (2022). Strain-level fitness in the gut microbiome is an emergent property of glycans and a single metabolite. Cell *185*, 513–529.e21. https://doi.org/10.1016/j.cell.2022.01.002.

57. De Filippis, F., Pasolli, E., Tett, A., Tarallo, S., Naccarati, A., De Angelis, M., Neviani, E., Cocolin, L., Gobbetti, M., Segata, N., et al. (2019). Distinct genetic and functional traits of human intestinal Prevotella copri Strains Are associated with different habitual diets. Cell Host Microbe *25*, 444–453.e3. https://doi.org/10.1016/j.chom.2019.01.004.

58. Vatanen, T., Plichta, D.R., Somani, J., Münch, P.C., Arthur, T.D., Hall, A.B., Rudolf, S., Oakeley, E.J., Ke, X., Young, R.A., et al. (2019). Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life. Nat. Microbiol. *4*, 470–479. https://doi.org/10.1038/s41564-018-0321-5.

59. Carrow, H.C., Batachari, L.E., and Chu, H. (2020). Strain diversity in the microbiome: lessons from Bacteroides fragilis. PLOS Pathog. *16*, e1009056. https://doi.org/10.1371/journal.ppat.1009056.

60. Porter, N.T., Hryckowian, A.J., Merrill, B.D., Fuentes, J.J., Gardner, J.O., Glowacki, R.W.P., Singh, S., Crawford, R.D., Snitkin, E.S., Sonnenburg, J.L., et al. (2020). Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in Bacteroides thetaiotaomicron. Nat. Microbiol. *5*, 1170–1181. https://doi.org/10.1038/s41564-020-0746-5.

61. Lourenço, M., Chaffringeon, L., Lamy-Besnier, Q., Titécat, M., Pédron, T., Sismeiro, O., Legendre, R., Varet, H., Coppée, J.Y., Bérard, M., et al. (2022). The gut environment regulates bacterial gene expression which modulates susceptibility to bacteriophage infection. Cell Host Microbe *30*, 556–569.e5. https://doi.org/10.1016/j.chom.2022.03.014.

62. Sheth, R.U., Cabral, V., Chen, S.P., and Wang, H.H. (2016). Manipulating Bacterial Communities by in situ microbiome Engineering. Trends Genet. *32*, 189–200. https://doi.org/10.1016/j.tig.2016.01.005.

63. Voorhees, P.J., Cruz-Teran, C., Edelstein, J., and Lai, S.K. (2020). Challenges & opportunities for phage-based in situ microbiome engineering in the gut. J. Control. Release *326*, 106–119. https://doi.org/10.1016/j.jconrel.2020.06.016.

64. Aranda-Díaz, A., Ng, K.M., Thomsen, T., Real-Ramírez, I., Dahan, D., Dittmar, S., Gonzalez, C.G., Chavez, T., Vasquez, K.S., Nguyen, T.H., et al. (2022). Establishment and characterization of stable, diverse, fecal-derived in vitro microbial communities that model the intestinal microbiota. Cell Host Microbe *30*, 260–272.e265. https://doi.org/10.1016/j.chom.2021.12.008.

65. Afrizal, A., Jennings, S.A.V., Hitch, T.C.A., Riedel, T., Basic, M., Panyot, A., Treichel, N., Hager, F.T., Wong, E.O.-Y., Wolter, B., et al. (2022). Enhanced cultured diversity of the mouse gut microbiota enables custom-made synthetic communities. Cell Host Microbe *30*, 1630–1645.e25. https://doi.org/10.1016/j.chom.2022.09.011.

66. Shaer Tamar, E., and Kishony, R. (2022). Multistep diversification in spatiotemporal bacterial-phage coevolution. Nat. Commun. *13*, 7971. https://doi.org/10.1038/s41467-022-35351-w.

67. Federici, S., Nobs, S.P., and Elinav, E. (2021). Phages and their potential to modulate the microbiome and immunity. Cell. Mol. Immunol. *18*, 889–904. https://doi.org/10.1038/s41423-020-00532-4.

68. Cao, Y., Wang, Z., Yan, Y., Ji, L., He, J., Xuan, B., Shen, C., Ma, Y., Jiang, S., Ma, D., et al. (2021). Enterotoxigenic Bacteroidesfragilis promotes intestinal inflammation and malignancy by inhibiting exosome-packaged miR-149-3p. Gastroenterology *161*, 1552–1566.e12. https://doi.org/10.1053/j.gastro.2021.08.003.

69. Sears, C.L. (2009). Enterotoxigenic Bacteroides fragilis: a rogue among Symbiotes. Clin. Microbiol. Rev. *22*, 349–369. https://doi.org/10.1128/CMR.00053-08.

70. Boleij, A., Hechenbleikner, E.M., Goodwin, A.C., Badani, R., Stein, E.M., Lazarev, M.G., Ellis, B., Carroll, K.C., Albesiano, E., Wick, E.C., et al. (2015). The Bacteroides fragilis toxin gene is prevalent in the colon mucosa of colorectal cancer patients. Clin. Infect. Dis. *60*, 208–215. https://doi.org/10.1093/cid/ciu787.

71. Maura, D., Morello, E., du Merle, L., Bomme, P., Le Bouguénec, C., and Debarbieux, L. (2012). Intestinal colonization by enteroaggregative Escherichia coli supports long-term bacteriophage replication in mice. Environ. Microbiol. *14*, 1844–1854. https://doi.org/10.1111/j.1462-2920.2011.02644.x.

72. Lourenço, M., De Sordi, L., and Debarbieux, L. (2018). The diversity of bacterial lifestyles hampers bacteriophage tenacity. Viruses *10*, 327.

73. Lourenço, M., Chaffringeon, L., Lamy-Besnier, Q., Pédron, T., Campagne, P., Eberl, C., Bérard, M., Stecher, B., Debarbieux, L., and De Sordi, L. (2020). The spatial heterogeneity of the gut limits predation and fosters coexistence of bacteria and bacteriophages. Cell Host Microbe *28*, 390–401.e5. https://doi.org/10.1016/j.chom.2020.06.002.

74. Patnode, M.L., Guruge, J.L., Castillo, J.J., Couture, G.A., Lombard, V., Terrapon, N., Henrissat, B., Lebrilla, C.B., and Gordon, J.I. (2021). Strain-level functional variation in the human gut microbiota based on bacterial binding to artificial food particles. Cell Host Microbe *29*, 664–673.e5. https://doi.org/10.1016/j.chom.2021.01.007.

75. Bikard, D., Euler, C.W., Jiang, W., Nussenzweig, P.M., Goldberg, G.W., Duportet, X., Fischetti, V.A., and Marraffini, L.A. (2014). Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nat. Biotechnol. *32*, 1146–1150. https://doi.org/10.1038/nbt.3043.

76. Selle, K., Fletcher, J.R., Tuson, H., Schmitt, D.S., McMillan, L., Vridhambal, G.S., Rivera, A.J., Montgomery, S.A., Fortier, L.-C., Barrangou, R., et al. (2020). In vivo targeting of Clostridioides difficile using phage-delivered CRISPR-Cas3 antimicrobials. mBio *11*, e00019–e00020. https://doi.org/10.1128/mBio.00019-20.

77. Buchfink, B., Reuter, K., and Drost, H.G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods *18*, 366–368. https://doi.org/10.1038/s41592-021-01101-x.

78. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.

79. Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 1–11.

80. Bushnell, B. (2014). BBMap: a Fast, Accurate, Splice-Aware Aligner (Lawrence Berkeley National Lab.).

81. Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. *9*, 5114. https://doi.org/10.1038/s41467-018-07641-9.

82. Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics *36*, 1925–1927. https://doi.org/10.1093/bioinformatics/btz848.

83. Nayfach, S., Camargo, A.P., Schulz, F., Eloe-Fadrosh, E., Roux, S., and Kyrpides, N.C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat. Biotechnol. *39*, 578–585. https://doi.org/10.1038/s41587-020-00774-7.

84. 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., and Roux, S. (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.

85. Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Res. *27*, 824–834. https://doi.org/10.1101/gr.213959.116.

86. Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., and Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big Biological Data. Mol. Plant *13*, 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009.

87. Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. Genome Biol. *12*, R60. https://doi.org/10.1186/gb-2011-12-6-r60.

88. Dixon, P. (2003). VEGAN, a package of R functions for community ecology. J. Veg. Sci. *14*, 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

89. Rampelli, S., Soverini, M., D'Amico, F., Barone, M., Tavella, T., Monti, D., Capri, M., Astolfi, A., Brigidi, P., Biagi, E., et al. (2020). Shotgun metagenomics of gut microbiota in humans with up to extreme longevity and the increasing role of xenobiotic degradation. mSystems *5*. e00124–e00120. https://doi.org/10.1128/mSystems.00124-20.

90. Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., Zhang, D., Xia, H., Xu, X., Jie, Z., et al. (2015). Gut microbiome development along the colorectal adenoma-carcinoma sequence. Nat. Commun. *6*, 6528. https://doi.org/10.1038/ncomms7528.

91. Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J.R., Prifti, E., Nielsen, T., et al. (2014). An integrated catalog of reference genes in the human gut microbiome. Nat. Biotechnol. *32*, 834–841. https://doi.org/10.1038/nbt.2942.

92. Zeller, G., Tap, J., Voigt, A.Y., Sunagawa, S., Kultima, J.R., Costea, P.I., Amiot, A., Böhm, J., Brunetti, F., Habermann, N., et al. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol. Syst. Biol. *10*, 766. https://doi.org/10.15252/msb.20145645.

93. Xie, H., Guo, R., Zhong, H., Feng, Q., Lan, Z., Qin, B., Ward, K.J., Jackson, M.A., Xia, Y., Chen, X., et al. (2016). Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. Cell Syst. *3*, 572–584.e3. https://doi.org/10.1016/j.cels.2016.10.004.

94. Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60. https://doi.org/10.1038/nature11450.

95. Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y., et al. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat. Med. *21*, 895–905. https://doi.org/10.1038/nm.3914.

96. Parks, D.H., Chuvochina, M., Chaumeil, P.A., Rinke, C., Mussig, A.J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. Nat. Biotechnol. *38*, 1079–1086. https://doi.org/10.1038/s41587-020-0501-8.

97. Brister, J.R., Ako-Adjei, D., Bao, Y., and Blinkova, O. (2015). NCBI viral genomes resource. Nucleic Acids Res. *43*, D571–D577.

98. Livingston, S.J., Kominos, S.D., and Yee, R.B. (1978). New medium for selection and presumptive identification of the Bacteroides fragilis group. J. Clin. Microbiol. *7*, 448–453. https://doi.org/10.1128/jcm.7.5.448-453.1978.

99. Bloom, S.M., Bijanki, V.N., Nava, G.M., Sun, L., Malvin, N.P., Donermeyer, D.L., Dunne, W.M., Allen, P.M., and Stappenbeck, T.S. (2011). Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. Cell Host Microbe *9*, 390–403. https://doi.org/10.1016/j.chom.2011.04.009.

100. Roux, S., Enault, F., Hurwitz, B.L., and Sullivan, M.B. (2015). VirSorter: mining viral signal from microbial genomic data. PeerJ *3*, e985.

101. Tong, J., Liu, C., Summanen, P., Xu, H., and Finegold, S.M. (2011). Application of quantitative real-time PCR for rapid identification of Bacteroides fragilis group and related organisms in human wound samples. Anaerobe *17*, 64–68. https://doi.org/10.1016/j.anaerobe.2011.03.004.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
GPIC phages, see Table S2	This study	CGMCC 6.9201 - 6.9409
Bacterial strains	This study	N/A
Chemicals, peptides, and recombinant proteins		
Phosphate-buffered saline (PBS)	Sangon Biotech	Cat#E607008
L-CysteineL-Cysteine	Sangon Biotech	Cat#A600132
PEG8000	Sangon Biotech	Cat#A100159A
Hemin	Sangon Biotech	Cat#A602521
Kanamycin	Sangon Biotech	Cat#A100408
Vancomycin	Sangon Biotech	Cat#A100990
Bacteroides Bile Esculin agar	Solarbio	Cat#LA7310
Brain Heart Infusion (BHI)	Huankai Biology	Cat#28363
Vitamin K3	Sangon Biotech	Cat#A502486
SM buffer	Sangon Biotech	Cat#B548130
Ribonuclease A	Takara	Cat#2158
Recombinant DNase I	Takara	Cat#2270A
Premix Taq™	Takara	Cat#RR901A
Critical commercial assays		
Takara MiniBEST Viral RNA/DNA Extraction Kit	Takara	Cat#9766
Nextera DNAIllumina Nextera DNA Flex Library Prep kit	Illumina	Cat#FC-121-1031
QIAamp MinElute Virus Spin Kit	Qiagen	Cat#57704
DNeasy UltraClean 96 Microbial Kit	Qiagen	Cat#10196-4
Genomic DNA Extraction Kit	Sangon Biotech	Cat#B518225
E.Z.N.A.® Gel Extraction Kit	Omega Bio-Tek	Cat#D2500-01
SYBRGreen PCR Master Mix	Yeasen Biotech	Cat#11201ES08
Oligonucleotides		
Bacterial 16S rRNA gene PCR, (27F: 5'-AGAGTTTGATCMTGGCTCAG-3', 1492R: 5'-GGTTACCTTGTTACGACTT-3',)	Sangon Biotech	N/A
qPCR, *B. fragilis* 16S rRNA gene, (Forward: 5'-TGATTCCGCATGGTTTCATT-3', Reverse: 5'-CGACCCATAGAGCCTTCATC3')	Sangon Biotech	N/A
Deposited data		
GPIC phage genomes, see Table S2	GenBank	Accession #: OP172633 - OP172841
Metavirome data of five human feces and one wastewater samples	NCBI	Accession #: PRJNA861902
Metagenomic data of 18 samples of human feces with a 24-hours batch culture in vitro	NCBI	Accession #: PRJNA866369
Metagenomic data of 45 samples of human feces with a 120-hours serial batch culture in vitro	NCBI	Accession #: PRJNA924508
Metagenomic data of 99 samples of human feces with a 72-hours batch culture in vitro	NCBI	Accession #: PRJNA924526

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sequences of 16S rRNA genes and bacterial genomes used in this study	Zenodo	https://doi.org/10.5281/zenodo.7050205
Software and algorithms		
BLAST+	Camacho et al.[41]	v2.11.0
vConTACTv2	Jang et al.[42]	v0.9.19
DIAMOND	Buchfink et al.[77]	v0.9.24
Cytoscape	Shannon et al.[78]	v3.9.1
MEGA	Kumar et al.[35]	X
OrthoFinder	Emms et al.[46]	v2.5.2
Prodigal	Hyatt et al.[79]	v2.6.1
Graphpad Prism	http://www.graphpad.com/scientific-software/	v9.0.0
EasyFig	Sullivan et al.[47]	v2.2.5
CrisprOpenDB	Dion et al.[45]	https://github.com/edzuf/CrisprOpenDB
BBmap	Bushnell et al.[80]	v39.01
MAFFT	Katoh et al.[52]	v7.490
FastANI	Jain et al.[81]	v1.32
Kraken2	Wood et al.[54]	v2.0.7-beta
GTDB-Tk	Chaumeil et al.[82]	v2.1.0
CheckV	Nayfach et al.[83]	v1.0.1
Prokka	Seemann[36]	v1.14.6
VirSorter2	Guo et al.[84]	v1.0.6
SPAdes	Nurk et al.[85]	v3.14.0
Tbtools	Chen et al.[86]	v1.108
qPCRsoft	Analytik Jena	v4.1
LEfSe	Segata et al.[87]	http://huttenhower.sph.harvard.edu/lefse/
Vegan	Dixon[88]	v2.6.4
RStudio	https://www.rstudio.com	v4.1.2
R	https://www.rproject.org/	v4.1.2
Others: Publicly available metagenomic data from eight studies used in this paper		
Study 1: 40 samples from Tanzania	Smits et al.[51]	NCBI: PRJNA392180
Study 2: 62 samples from Italy	Rampelli et al.[89]	NCBI: PRJNA553191
Study 3: 156 samples from Austria	Feng et al.[90]	EBI: PRJEB7774
Study 4: 249 samples from Denmark and Spain	Li et al.[91]	EBI: PRJEB5224
Study 5: 199 samples from France and Germany	Zeller et al.[92]	ENA: PRJEB6070
Study 6: 250 samples from UK	Xie et al.[93]	EBI: PRJEB9584
Study 7: 145 samples from China	Qin et al.[94]	NCBI: PRJNA422434
Study 8: 232 samples from China	Zhang et al.[95]	EBI: PRJEB6997

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and materials should be directed to and will be fulfilled by the lead contact, Yingfei Ma (yingfei.ma@siat.ac.cn).

Materials availability

All the phages isolated in this study have been deposited in CGMCC (China General Microbiological Culture Collection Center, https://cgmcc.net/english/) with the accession numbers CGMCC 6.9201 - CGMCC 6.9409. CGMCC will make these phage isolates available to the international scientific community (https://cgmcc.net/english/deposit). This study did not generate new unique

reagents. This study was approved by the Ethics Committee of the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences with the approval number SIAT-IRB-191015-H0388. Meanwhile, all these phage isolates can be available upon request by contacting the lead contact, Yingfei Ma (yingfei.ma@siat.ac.cn).

Data and code availability

The 209 phage genomes isolated in this study have been uploaded to the NCBI database with the accession numbers OP172633 - OP172841 (Details in Table S1). The metavirome data of five human fecal samples and one wastewater sample have been uploaded to the NCBI database with the BioProject ID PRJNA861902. The metagenomic data of 18 samples of human feces with a 24-hour batch culture in vitro under phage cocktail treatment or heat-killed phage treatment has been uploaded to NCBI, BioProject ID PRJNA866369. The metagenomic data of 45 samples of human feces with a 120-hours serial batch culture (subculture every 24h) in vitro under different treatments has been uploaded to NCBI, BioProject ID PRJNA924508. The metagenomic data of 99 samples of human feces with a 72-hours batch culture (sampling every 24h) in vitro under different treatments have been uploaded to NCBI, BioProject ID PRJNA924526. All the sequences of 16S rRNA genes and bacterial genomes used in this study were deposited at Mendeley Data (https://doi.org/10.17632/ybfcgstgyw.1) or at Zenodo Data (https://doi.org/10.5281/zenodo.7050205).

The present study did not generate code, and mentioned tools used for the data analysis were applied with default parameters unless specified otherwise in the section of STAR Methods.

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

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial strains

Table S1 lists all of the strains used for phage isolation in this study. All the strains were isolated from the human feces of healthy Chinese donors for this study. The main procedures are as follows: Fecal samples were collected from healthy human donors not taking any drugs during the last three months before sampling. The samples were immediately transferred to an anaerobic chamber (Vinyl Anaerobic Chambers, Coylab, USA), homogenized in phosphate-buffered saline (PBS) supplemented with 0.1% cysteine, then diluted and spread on agar plates containing growth medium (YCFA plates). Plates were incubated anaerobically at 37 °C for 2-3 days in an atmosphere of 85% N_2, 5% CO_2, and 10% H_2. A single colony was picked, streaked on a new plate, and incubated anaerobically at 37 °C for another 2-3 days. The purification steps are repeated once to obtain pure strains. The isolates were then transferred to a liquid medium and incubated at 37 °C for another 2 days. Finally, we amplified and sequenced the full 16S rRNA genes obtained using the PCR primer pair (27F 5'-AGAGTTTGATCMTGGCTCAG-3'; 1492R 5'GGTTACCTTGTTACGACTT-3'). Strains with double peaks in the sequences were discarded or entered into a new round of purification process. All purified strains were stored at -80 °C in glycerol suspension (25%, v/v) containing 0.1% cysteine. The taxonomy for the isolates was identified using the full-length 16S rRNA gene sequences by comparing the sequences of 16S rRNA genes against those in the database of GTDB version 202 (Table S1). For the isolates with whole genome sequences, the taxonomy was also analyzed by GTDB-Tk v2.1.0 with the database version of release-202 (Table S1), a toolkit for classifying prokaryotic genomes using the Genomic Taxonomy Database (GTDB).[96] The phylogenetic tree of these strains based on 16S rRNA genes was constructed by using the Maximum Likelihood method and the Tamura-Nei mode with MEGA X.[35]

Bacterial culture conditions

All these bacterial isolates were grown in Brain Heart Infusion (BHI, Huankai Biology) liquid medium containing hemin and vitamin K3. Bacterial isolates were cultured in an anaerobic chamber (Bactron VI, Shellab, Sheldon Manufacturing, Incorporation using a mixed gas of 5% H_2, 5% CO_2, and 90% N_2) at 37 °C. For BHI agar plates, 15 g/L agar was added. For BHI top agar used in soft agar overlays, 4 g/L agar was added before autoclaving. The strains were first re-streaked from the glycerol stock onto BHI agar plates and grown anaerobically for up to 1–2 days. A single colony was subcultured into a pre-reduced liquid BHI broth, and grown anaerobically overnight to provide the starting culture for experiments.

Ethics statement

The use of human fecal samples was explained in detail to the volunteers and approved by the Ethics Committee of the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, with the approval number SIAT-IRB-191015-H0388.

METHOD DETAILS

Phage isolation from sewage wastewater samples and human feces

In this study, phages were isolated from municipal sewage wastewater samples collected from the Nanshan Wastewater Treatment Plant and Futian Wastewater Treatment Plant (Shenzhen, China) or human feces mixtures of 20 individual Chinese donors in Shenzhen, China. Wastewater samples were centrifuged at 9000 rcf for 10 min at 4 °C to remove any remaining solids. The supernatant was then sequentially filtered through 0.8 μm and 0.22 μm polyvinylidene fluoride (PVDF) filters to further remove bacterial cells. Human fecal samples were added SM buffer pre-cooled at 4 °C at a ratio of 10 times the mass of the stool immediately after the samples were collected, and the samples were shaken and mixed. The samples were then sequentially centrifuged at 5000 rcf, 8000 rcf and 13000

rcf for 10 min at 4 °C. Finally, the supernatant was filtered through 0.8 μm and 0.22 μm PVDF filters in turn, and the filtrate was stored at 4 °C.

To ensure the success of phage isolation, phages in these processed samples were enriched in two ways.

1) To enrich the phages specific to a given host, the processed samples were cocultured with the host bacterial cells for 3 consecutive rounds of co-culturing enrichment. Briefly, in the first round of enrichment, we added 12 mL of 3-fold concentrated BHI medium, 600 μL of bacterial cells (OD_{600}~0.3) and 24 mL of the processed wastewater samples or fecal samples to a 50 mL centrifuge tube, and placed the tube in an anaerobic box for overnight incubation at 37 °C. In the second round of enrichment: The culture of the first round was centrifuged for 5 min at 4 °C, 8000 rcf, and the supernatant (6 mL) was passed through a 0.22 μm membrane, and inoculated with 200 μL of fresh bacterial cells and 3 mL of 3-fold concentrated BHI medium in a 15 mL tube overnight in an anaerobic incubator at 37°C. After three rounds of enrichment, the culture was centrifuged at 4 °C, 8000 rcf and filtered with 0.22 μm PVDF membrane, and the filtered culture was detected by spotting method to detect the presence of the phages infecting the target strain in the final coculture. Specifically, 2.5 μL of enriched droplets were placed on the solidified top agar layer containing the target strain, and incubated anaerobically at 37 °C overnight.
2) To enrich all phages, the processed samples were further processed using the tangential flow device and protein concentrators. The membrane pore sizes used for bacterial removal and phage concentration were 0.22 μm and 100 KD, respectively. The resulting phage concentrates were further concentrated by protein concentrators with a pore size of 100 KD. The phages in the processed samples were concentrated up to 2000-fold in this process.

These phage-enriched samples were used to isolate phages and further were purified by the double-layer plate method. To facilitate a diverse collection of phages, no more than 6 plaques from the same plate were selected according to the plaque morphology. The procedure is as follows:

1) Determination of phage titer in the phage-enriched samples: 500 uL of the corresponding bacterial cells (OD_{600}~0.6) plus 4.5 mL of semi-solid medium were used for plating. After solidification, 2.5 μL of each phage-enriched sample was used for plating, and the phage titer was calculated after culturing for 24h.
2) Selection of single phage plaque: 10 μL of the dilution for each phage-enriched sample, 500 μL of bacterial culture solution, and 4.5 mL of semi-solid medium were mixed for plating, and incubated at 37 °C for 24h to observe the plaque morphology on the plate. For each sample, at least 3 dilutions were selected for plating to form plaques clearly on the plate and to see the diversity of plaque morphologies.
3) Purification of the phage isolates derived from a single plaque: according to the plaque morphologies, plaques were picked using a sterilized toothpick and transferred into 1 mL SM buffer. The supernatant was filtered through a 0.22 μm filter membrane for plating and purification again. This procedure was repeated three times consecutively to ensure successful purification for each phage. The well-purified phage isolates were selected and placed in 1 mL SM buffer, filtered with 0.22 μm PVDF membrane, and stored at 4°C.

In summary, two types of samples including human feces and municipal sewage wastewater were used to isolate phages for the human gut bacteria. However, most (199/209) of the GPIC phages were isolated in this study from the sewage wastewater samples (Table S1)). In particular, no phage plaques were observed for all 94 strains from 5 donors when the same fecal samples were used for phage isolation (Table S1). In addition, a step-by-step workflow for phage isolation from primary wastewater is shown in Figure S1A.

Phage genome sequencing, assembly and annotation

High-titer phage stocks were prepared by liquid culture or solid spotting for genome sequencing. 0.2 mL of the purified phage was added to the culture medium (10 mL) of the target strain in the early logarithmic growth phase, and the mixture was cocultured overnight to obtain a high titer of the phage. Alternatively, 20 μL purified phage can also be directly dropped on the soft agar overlay to form large plaques. After overnight incubation, the plaque was resuspended in 4 mL of SM buffer and incubated at room temperature for about 1 h. Finally, the lysate was centrifuged at 5500 rcf for 10 min and filter-sterilized through a 0.22 μm PVDF filter to obtain a high-titer phage stock.

Phage genome DNA was extracted from the high-titer phage stock with Takara MiniBEST Viral RNA/DNA Extraction Kit. The sequencing library was prepared using the Illumina Nextera DNA Flex Library Prep kit and sequenced by Illumina MiSeq. Reads were assembled with SPAdes 3.14.0 with default options after trimming reads with an error probability limit of 0.05.[85] The median positional coverage for these assemblies was 1329× (Figure S1B), and no contamination was detected in the sequencing reads.

The overall quality of the genome assembly was evaluated using CheckV.[83] The completeness level of ~88% of GPIC phages was larger than 90% based on AAI (average amino acid identity) -based approach in CheckV. Details about the predicted genome completeness (AAI-based) and closed genomes of GPIC phages by checkV can be found in Table S1. Closed genomes were also predicted by checking proviruses, direct terminal repeats (DTRs) and inverted terminal repeats (ITRs).[83] Here, a total of 142 genomes were identified as closed genomes, of which, 140 possessed DTRs and 2 possessed ITRs.

Assembled genomes were annotated with Prokka 1.14.6[36] using a custom database constructed from NCBI RefSeq viral genomes[97] as well as using VOGDB's hmm profile (http://vogdb.org/).

Phage host range assays

Host range analysis was carried out using high-titer phage stocks. 2.5 μL high titer phage stocks were plated onto solidified top agar overlays containing target strains. After the spots dried, plates were incubated anaerobically for about 16h before counting plaques. Host susceptibility was established when a lysis plaque was observed. To obtain as many *Bacteroides* strains as possible from a single individual donor, we used two *Bacteroides*-selective media to screen and cultivate *Bacteroides* strains from human feces.[98,99] The two *Bacteroides*-selective media are *Bacteroides* Bile Esculin (BBE) agar supplemented with gentamicin (100 ng/μL) and BBE agar supplemented with kanamycin (10 ng/μL) and vancomycin (0.5 ng/μL), respectively. Ultimately, we screened 58 *Bacteroides* isolates from 200 clones from two individual donors.

Transmission Electron Microscopy

High titer phage stock for each phage isolate was used for Transmission Electron Microscopy (TEM). Five microliters of phage stock were adsorbed onto carbon-coated copper grids and negatively stained with phosphotungstic acid (2% w/v, pH 4.5). After drying, the preparation was examined on a JEM-2100 electron microscope (JEOL, Tokyo, Japan). Bacteriophage size was determined from the average of three independent pictures.

Phage enrichment from sewage wastewater samples and human feces for metagenomic sequencing

The enrichment of viral particles from the sewage wastewater samples used the tangential flow and protein concentrators described above. Viral particles were concentrated from stool samples using polyethylene glycol (PEG) precipitation protocol. Briefly, NaCl was added to 10 mL of fecal phage suspension to a final concentration of 0.5 M (0.29 g/10 mL) and solid PEG8000 was added to a final concentration of 10% (w/v, 1.0 g/10 mL). The mixture was fully dissolved and kept at 4 °C overnight (more than 16h). The liquid from the previous step was centrifuged at 12,000 g for 20 min at 4 °C, the supernatant was removed, and the pellet was collected and resuspended in 400 μL of SM buffer. 360 μL of the supernatant was taken from the previous step, added 40 μl of 10X DNase buffer, and then added 8U TURBO DNase (4 μL) and 20U RNAse (0.2 μL) at 37 °C for 1h. The digested solution was inactivated in a water bath at 70 °C for 10 min and proceeded to the next step of kit purification. Phage genome extraction was performed using the QIAamp MinElute Virus Spin Kit. Reads were assembled using SPAdes[85] and high-quality, non-redundant phage contigs were selected based on the identification of phage sequence, contig length, and read coverage using VirSorter2.[100]

Comparative genomic analysis of phage genomes

All the genomes of the GPIC phage isolates and all the complete phage genomes predicted in the GPD database were pooled and processed with vConTACT2,[42] which compares both the presence and absence of homologous proteins between phage genomes and the sequence identity between homologous proteins and then generates probabilistic clusters of contigs with high similarity. Protein similarities were calculated using DIAMOND[77] (version 0.9.24, running in BLASTp mode with default sets) to compare proteins against the Viral RefSeq v201 database provided by vContACTv2. Protein clusters were generated with the Markov clustering algorithm, and viral clusters (VCs) were generated with the clusterONE algorithm. Benchmarking against the RefSeq phages revealed that the boundaries of GPD VCs were equivalent to a genus level, with 96% of the tested 2,304 phages' VCs agreeing with the ICTV's (International Committee on Taxonomy of Viruses) genus-level viral taxonomy assignments for NCBI RefSeq viruses.[42] The obtained phage similarity network was visualized in Cytoscape v3.9.1.[78] The network was filtered and manually sorted by removing duplicate edges and applying an edge-weighted Spring Embedded Layout. In the network, a discrete component (DC) is made up of the phage genomes that share homologous genes and these phage genomes have no genes shared with the phage genomes of other discrete components.

Identification of phage orthologous proteins and classification of phage genomes

According to the rank-specific demarcation criteria for tailed phages (Class *Caudoviricetes*) by the ICTV, the classification for tailed phages is conducted based on the phage genomes.[48] The viral family is represented by a cohesive and monophyletic group in the main predicted proteome-based clustering tools (*e.g.* vConTACT2 network) and members of the family share a significant number of orthologous genes. Accordingly, in this study, two groups of phages were classified as a viral family, respectively. Briefly, these two groups of phage genomes formed two monophyletic groups in the predicted network by vConTACT2.

The encoded proteins of phages were predicted using Prodigal v2.6.1.[79] Orthologous proteins shared between phages were detected by OrthoFinder with default parameters.[46] The presence/absence of orthologous proteins between phages was initially converted into a binary count matrix where the percentage of shared orthologous proteins was calculated. Heatmaps were made based on the percentage of shared orthologous proteins and clustering of phages was performed. These phages were clustered based on the percentage of shared orthologous proteins and visualized using TBtools.[86] The orthologous proteins were separately aligned using the MAFFT v7.490 with the default sets. The phylogenetic trees were inferred by using the Maximum Likelihood method and JTT matrix-based model with MEGA X.[35]

Distribution of phage isolates in the human gut

Metagenomic reads from 1,333 human gut samples were mapped to our phage isolates using BBmap with default settings.[80] These samples were drawn from eight different studies.[51,89–95] Read coverage of targeted phage isolates in different metagenomic samples

is presented as a heatmap. In this study, a phage genome with a coverage of more than 30% was considered to be present in this metagenomic sample. The source of these 1,333 human gut metagenomic samples used in this study was listed in key resources table.

CRISPR spacer-based prediction of bacterial hosts

The prediction of host bacteria for the phages in this study used a CRISPR spacer-based approach as described in the previous study.[45] CrisprOpenDB is a command-line host prediction tool,[45] and its default settings were used in this study. The host prediction tool provides four levels of prediction (level 1–4). In this study, we only considered the first two levels of predictions (levels 1 and 2).

Design of a phage cocktail to reduce *B. fragilis* species in the human gut microbiome *in vitro*

To knock down *B. fragilis* from the human gut microbiome, according to their host ranges (see phage host range assays), 8 phages (F766P4, BF486P1, BF494P1, BF766P1, BF702P1, BF695P2, BF698P1, and BF344P1) with distinct genome sequences (Figure S4C) were selected to make up the cocktail with a host range for all the tested *B. fragilis* strains. Each phage was adjusted to $5*10^7$ PFU/mL and equally mixed for the cocktail. The human fecal samples were collected from healthy Chinese donors (donors #1, #2, and #3). The samples were mixed with 25% glycerol and 0.1% cysteine and filtered using gauze to remove solid residues, then stored at -80 °C for use. In the *in vitro* culture experiment, 20 μL of the processed human fecal sample for each donor was added to 2 mL fresh YCFA liquid medium (the modified version of DSMZ media 1611, https://mediadive.dsmz.de/medium/1611?mobile=1) and cultured in 96 deep-well plates (Sangon Biotechnology, Shanghai, China) in an anaerobic chamber (Bactron VI Sheldon Manufacturing, using a mixed gas of 5% H_2, 5% CO_2, and 90% N_2) at 37 °C for 24h. The phage cocktail (20 μL) was added to the wells at the initial timepoint. The heat-killed phage cocktail was obtained by incubating the phage cocktail in a FALCON polypropylene round-bottom tube at 98 °C for 30 min. After 24h of incubation, samples were collected for DNA extraction and subsequent analysis.

For the experiment with long-time incubation, 20 μL of the processed human fecal sample of donor #1 was used. The sample and phages (20 μL, $5*10^7$ PFU/mL) were added into 2 mL of the YCFA liquid medium in 96 deep-well plates and cultured in an anaerobic chamber at 37 °C for 72h. In total, 11 experimental groups were carried out, including 1 treatment group using the phage cocktail, 8 treatment groups using 8 individual phages, 1 control group using the heat-killed phage cocktail, and 1 control group without any treatment (i.e. blank). Each experimental group included three replicates. At the incubation time of 24, 48 and 72h, 0.5 mL samples were collected for DNA extraction and subsequent analysis.

For the subcultivation experiment, 20 uL of the processed human fecal sample of donor #1 was used. The sample was grown in 2 mL of the YCFA liquid medium in 96 deep-well plates and cultured in an anaerobic chamber at 37 °C with 4 consecutive transfers. The phage cocktail and the heat-killed phage cocktail were added at the initial time. Every 24h, 20 μL culture was transferred to 1.98 mL fresh YCFA medium (*i.e.* 100 fold dilution). One experimental group without treatment was set as a control. Each experimental group included three replicates. The samples (0.5 mL) were collected every 24h for DNA extraction and subsequent analysis.

Metagenomic sequencing and estimation of the relative abundance of *B. fragilis* in the microbiome

The genomic DNA was extracted with the DNeasy UltraClean 96 Microbial Kit (Qiagen) using a Tecan Freedom EVO liquid handler. The sequencing library was prepared using the Illumina Nextera DNA Flex Library Prep kit and sequenced by the Illumina Novaseq PE150 Platform. Reads were assembled with SPAdes 3.14.0 with default options after trimming reads with an error probability limit of 0.05.[85] The Kraken2 v2.0.7-beta (with default parameters and minikraken2_v1_8GB database) was used to obtain the taxonomic abundance in metagenomic data.[54] Alpha-diversity based on the Shannon diversity index and Beta-diversity based on Jaccard dissimilarity matrices was calculated using Vegan package in R.[88] LEfSe was used to identify microbial species with significant changes in abundance between groups.[87]

Estimation of the absolute abundance of *B. fragilis* by the qPCR assay

The total DNA of *B. fragilis* DA486 was extracted using the Genomic DNA Extraction Kit for Bacteria (Sangon Biotechnology, Shanghai, China) as described in the manufacturer's instructions. The primers (F, 5'-TGATTCCGCATGGTTTCATT-3'; R, 5'-CGACCCATAGAGCCTTCATC-3') were designed for specific detection of *B. fragilis*.[101] The PCR products were confirmed on 1.5% agarose gel. The PCR products were purified using the omega E.Z.N.A.® Gel Extraction Kit (Bio-tek, Norcross, USA). The concentration of the purified target DNA was 175.9 ng/μL determined by Nanodrop, and then the purified target DNA was serially diluted with ddH_2O from 175.9 ng/μL to $175.9*10^{-10}$ ng/μL. These tenfold dilutions (from $175.9*10^{-3}$ to $175.9*10^{-10}$ ng/μL) were used as standards in this study.

The primer pair was used to determine the absolute abundance of *B. fragilis* in the microbiome derived from human fecal samples *in vitro*. Quantitative PCR (qPCR) was carried out using a qTOWER3 quantitative PCR System with software version 4.1 (Analytik Jena, Germany). Amplification and detection were carried out with SYBRGreen PCR Master Mix (Yeasen Biotechnology, Shanghai, China) in a final volume of 20 μL containing a final concentration of 0.2 μM primers and 1 μL DNA templates. Amplifications were carried out using the following ramping profile: 1 cycle at 95 °C for 5 min, followed by 40 cycles of 95 °C for 10 s, 60 °C for 20 s, and 72 °C for 20 s. A melting step was added using the default sets of the qTOWER3 system. Standard DNA of the tenfold dilution series was included in each qPCR plate to form a standard curve for target DNA quantification. The

concentration of target DNA was calculated from the standard curves of standard DNA using the qPCRsoft v4.1 of qTOWER3 with default sets.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed using Graphpad Prism 9.0.0. Details of specific analyses and statistic tests are described in applicable figure legends.