Supplementary Methods

Bacterial DNA Sequencing

DNA libraries were prepared using a KAPA HyperPlus kit (KAPA Biosystems, Indianapolis, IN) following the manufacturer's instructions, except that NEBNext multiplex oligos were used for Illumina (New England BioLabs, Ipswich, MA) during the adapter ligation and barcoding steps. The prepared target library size was 450 to 550 nucleotides. The concentration of the library was quantified using a KAPA Illumina Library Quantification kit (KAPA Biosystems) and adjusted to the mean library size measured by MCE-202 MultiNA (Shimadzu, Kyoto, Japan). The libraries were pooled and sequenced on a HiSeq2500 sequencer (2 × 250 paired-end reads, HiSeq Rapid SBS Kit v2, Illumina, San Diego, CA). For each run, 10 libraries from the bacterial fractions were pooled at equimolar concentrations.

Viral DNA Sequencing

Extracted viral DNAs were fragmented identically using the Covaris M220 system (Covaris, Woburn, MA). The prepared fragment size was about 350 nucleotides. After fragmentation, DNA libraries were prepared using a Swift 1S Plus DNA Library Kit for Illumina. The concentration of the library was quantified using an Illumina Library Quantification kit and adjusted to the mean library size measured by MCE-202 MultiNA. The libraries were pooled and sequenced on a HiSeq2500 sequencer (2 × 250 paired-end reads, HiSeq Rapid SBS Kit v2). For each run on the HiSeq2500 sequencer, 40 libraries from the viral fractions were pooled at equimolar concentration.

Processing the Sequencing Data and Quality Assessment

Sequencing reads were demultiplexed using Illumina CASAVA software and processed using the following 3 steps:

(1) Adaptor sequence trimming. The adaptor sequences were removed using Cutadapt software v1.2.1 (http://cutadapt.readthedocs.io/en/stable/index.html).

(2) Nucleotide trimming for filtering low-quality sequenced nucleotides and removing duplicates. The first and last 10 nucleotides of each read were removed. Next, low-quality nucleotides within 20 nucleotides of both ends with Phred quality scores <20 were trimmed, as were the polynucleotides at the end of the sequences. After trimming such low-quality nucleotides, reads with lengths shorter than 75 nucleotides, reads with low-complexity (DUST score >7), exact duplicates, and sequences containing Ns and singletons were removed. This step was performed using PRINSEQ software lite v0.20.4 (http://prinseq.sourceforge.net/; -trim_right 10 -trim_left 10 -trim_qual_right 20 -trim_qual_left 20 -trim_qual_window 20 -trim_ns_right 1 -min_len 75 -lc_method dust -lc_threshold 7 -ns_max_n 0 -derep 1). Samples containing large portions of low-quality reads were excluded from further processing and analysis.

(3) Error base correction. Correction of sequencing errors based on the Hamming graph and Bayesian subclustering was performed using BayesHammer software[1] (SPAdes v3.11.0) (spades.py –only-error-correction).

Metagenome Assembly

The quality-filtered, error-corrected reads from each sample were assembled using MetaSPAdes[2,3] v3.11.0 with default k-mer lengths (options: –meta –only-assembler) (Supplementary Figure 1*A*). To compare the contig abundances across the samples, the assembled contigs (with lengths ≥1 kb for the viral fraction and ≥5 kb for the bacterial fraction) from each sample were pooled. CD-HIT-EST[4] (v4.6) was used to cluster the pooled contigs at 95% global average nucleotide identity (-c 0.95 -G 1 -n 10 -mask NX, Supplementary Figure 1*B*). The contigs from the viral and bacterial fractions were treated separately. From the nonredundant pooled contigs, circular contigs were identified by detecting overlaps of the 5′- and 3′-end sequences (> 50 nucleotides overlap at 100% identity) of the contigs using MegaBLAST (BLAST+ v2.7.1).[5] Each detected circular contig was trimmed to remove the redundant parts. Circular contigs longer than 1.5 kb and linear contigs longer than 5 kb were used for the analyses.

Viral Contig Extraction

The contigs constructed from the viral fraction were screened using the gene enrichment-based VirSorter[6] (v1.0.3) method and the k-mer frequency-based VirFinder[7] (v1.1) method to identify and remove bacteria-like contigs. VirSorter was performed using both the RefSeqABVir (–db 1) and Viromes (–db 2) databases in the "virome decontamination" mode (–virome 1) to extract the viral contigs (category 1, 2, or 3). VirFinder was performed using a default prediction model and $P < .05$ as the threshold. The viral contigs detected by either or both methods were used for further analyses. After this process, the remaining contigs were classed as viral contigs and used for the following analyses.

Viral Nucleotide and Protein Database

To classify the viral contigs, we prepared viral genome and protein databases. The viral RefSeq sequences (v84, containing 9497 genomes and 231,157 proteins) were downloaded from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239). Taxonomic lineage information was assigned to each viral sequence using NCBI taxonomy data (ftp://ftp.ncbi.nih.gov/pub/taxonomy/, downloaded on November 30, 2017).

Viral Classification by Nucleotide Alignment

We initially classified the viral contigs using viral RefSeq genomes (first part of Figure 1*A* shown in red). The viral contig sequences were searched against viral RefSeq genomes (v84) using MegaBLAST (BLAST+ v.2.7.1) with an *E* value of $<10^{-10}$. Because some viral contig sequences contained multiple fragments mapping to different viral genomes, significant alignments up to the top 5 were considered when determining viral taxonomy. If the top 5 alignments covered more than 50% of a contig sequence, the lowest common ancestor (LCA) of the top 5 hits was determined using blast2lca (https://github.com/emepyc/Blast2lca) (modified to use accession.version identifiers) with NCBI taxonomy (downloaded on November 30, 2017) and was assigned to the contig. Because p-crAssphage classification is conducted together with the detection of crAss-like marker genes in a later step of the pipeline, contigs classified as p-crAssphage at this step were set as unclassified.

Gene Prediction

The contigs that were not classified in the previous step (MegaBLAST with viral RefSeq genomes) were analyzed according to their open-reading frames (ORFs). The ORFs on the contigs were predicted using MetaProdigal[8] (v2.6.3) with the metagenomics procedure (-p meta). To predict genes spanning the 3′ to 5′ ends of a circular contig, a temporary version of the circular contig was used in the ORF prediction, where the first 1500 nucleotides were duplicated and added to the end of the contig.

crAss-like Phage Detection

Using the ORF annotation on the contigs, we initially detected crAss-like phage contigs using the crAss-like phage detection method reported previously.[9] The amino acid sequences of the prototypical crAssphage (p-crAssphage, NC_024711.1) genetic signatures, namely the polymerase (UGP_018) and terminase (UGP_092), were queried using BLASTP (BLAST+ v2.7.1) against the ORFs on the viral contigs with an *E* value $< 1 \times 10^{-5}$ and an alignment length $\geq$ 350. The viral contigs containing a BLASTP hit of either the p-crAssphage polymerase or terminase with a minimum contig length of 70 kb were classified as crAss-like phages. The similarities between the crAss-like phage contigs and the p-crAssphage genome (NC_024711.1) were assessed using MegaBLAST (BLAST+ v2.7.1). When the crAss-like phage contigs exhibited $\geq$95% sequence identity across 80% of the contig (the criteria previously used for genome-based species separation[10,11]) against the p-crAssphage genome, they were classified as p-crAssphage (the part with the gray background in Supplementary Figure 2*B*).

Tentative Viral Contig Classification

To annotate the predicted ORFs, the putative amino acid sequences encoded by the ORFs were queried using GHOST-MP[12] (v1.3.4) against the viral RefSeq protein (v84) with an *E* value of $< 1 \times 10^{-5}$ and a bitscore of $>$ 50. The predicted ORFs were also queried by hmmscan in HMMER3[13] (v3.1b2) against the PfamA[14] (v31.0) database with an *E* value $< 1 \times 10^{-5}$.

The viral RefSeq proteins with the top 3 closest homologies (*E* value $< 1 \times 10^{-5}$ and bitscore $>$ 50) were considered for each ORF. The taxonomic lineages of the 3 viral proteins were compared from the species level to the order level. For each level, if 2 or more of the hits shared the same taxon, the ORF was assigned to that taxon. To analyze the taxonomy of the entire contig, the taxonomic lineages of the classified ORFs within the contig were compared. From the species level to the order level, a taxon that was common in more than 50% of the classified ORFs was assigned to the contig, a process analogous to a previously reported method.[15]

Viral Classification With Pfam Structural Proteins

To improve the tentative classification of the viral contigs, we additionally used the phage structural proteins from the PfamA annotation (*E* value $< 1 \times 10^{-5}$). The contigs were classified as *Caudovirales*, *Myoviridae*, or *Microviridae* when the contigs possessed a phage tail protein, phage tail sheath protein, or Microviridae capsid protein gene, respectively. The Pfam entries for the phage structural proteins used in this classification are listed in Supplementary Table 2.

When a taxon from a certain rank was undetermined but instead determined as a higher rank, the upper-level taxon name with the prefix “uc_” (unclassified_) was used for its classification at the lower rank (eg, a contig with no family assignment but classified as *Caudovirales* at the order level was labeled as uc_Caudovirales at the family level). Contigs that could not be classified were labeled “Unclassified.” Contigs containing homologous proteins from different orders of viruses were labeled “Unclassified (Multiple).”

No viral contigs classified as eukaryotic viruses were detected in this study.

Bacterial Taxonomic Assignment

We assigned bacterial taxonomy to the contigs using PhyloPythiaS+.[16] The whole PhyloPythiaS+ pipeline was run (options: -n -g -o s16 mg -t -p c -r -s) using the reference database “NCBI201502” with the following configuration parameters: maxLeafClades = 500 and minPercentInLeaf = 0.05 (the others parameters were set as default). To obtain the taxonomic profile of each sample, the quality-filtered, error-corrected reads were mapped to microbial taxonomy-specific marker genes with MetaPhlAn2.0.[17]

Calculating the Read Coverage on Contigs

The quality-filtered, error-corrected reads were mapped to the nonredundant pooled contigs using the bbmap tool from BBtools with $\geq$95% identity and the ambiguous mapping option (ambiguous = random). A contig was considered “detected” in a sample when more than 75% of the contig’s length was covered by mapped reads, as recommended in a previous study.[18] The abundance of a contig was calculated as the average contig coverage (number of nucleotides mapped to the contig divided by the contig length), in which the abundance of a "nondetected" contig was set to 0 and normalized by the total number of nucleotides of the mapped reads in a sample, to have a total number of nucleotides equal to 1 × 10^9. In the viral library preparation, the double-stranded DNA (dsDNA) genomes provide twice as many templates as the ssDNA genomes per single virus through the dsDNA denaturation step. Therefore, in the viral fraction from a sample, the read coverage of the contigs that were not classified as ssDNA viruses was divided by 2, as reported in a previous study.[19]

Analysis of Diversity

Alpha diversity (Shannon index), beta diversity (Bray–Curtis dissimilarity), and richness measurements were performed using the *vegan* R package (https://cran.r-project.org/web/packa).

Analysis of Prophages

Prophage sequences in the bacterial contigs were detected by combining the following 2 approaches. In the first approach, prophage sequences were predicted according to known viral signatures with VirSorter[6] (v1.0.3). The bacterial contigs (≥5 kb) from individual samples were analyzed by VirSorter using both RefSeqABVir (–db 1) and Viromes (–db 2). The predicted prophage sequences from VirSorter categories 4 or 5 (presence of viral hallmark genes or enrichment of viral-like genes in a prophage region) were extracted. The positions of the predicted prophage sequences on the nonredundant pooled bacterial contigs were obtained through MegaBLAST (BLAST+ v2.7.1) searches (E value $<1 \times 10^{-100}$ and ≥95% identity), and the prophage sequences were merged when their positions overlapped. Prophage sequences longer than 3 kb were extracted and listed as candidate prophage sequences. In the second approach, we searched for prophages using the viral contigs generated in this study. The viral contigs were queried against the bacterial contigs using MegaBLAST (BLAST+ v2.7.1) with 90% of the viral contig length aligned at a minimum identity of 95%. We considered the aligned sequence to be a prophage when it satisfied the following criteria: (1) the bacterial contig sequence was 5 kb longer than the aligned viral contig sequence, and (2) the aligned viral contig was detected as circular or the bacterial contig contained a MetaPhlAn2.0[17] bacterial taxonomy-specific marker gene (MegaBLAST; E value $< 1 \times 10^{-10}$). Because the detected prophage sequences could include partial sequences from a single prophage region, the detected prophage sequences located within 1 kb on a contig were merged. Finally, the detected prophage sequences longer than 10 kb or the prophage sequences detected on the ssDNA phage contigs (known ssDNA phage genomes are shorter than 10 kb) were extracted and listed as candidate prophage sequences.

The candidate prophage sequences from the first and second approaches were compared to define the final prophage sequences. We first considered the prophage sequences detected by the circular viral contigs using the second approach to be highly reliable. Then, the prophage sequences from the first approach that overlapped with the circular viral contig-derived prophages (>75% overlap in either sequence) were removed from the candidate list. We next examined the remaining candidate prophage sequences. If the prophage sequences from the 2 approaches overlapped (>75% overlap in either sequence), the overlapping prophage sequences were merged. The merged and non-overlapping candidate prophage sequences, as well as the circular viral contig-derived prophage sequences, were defined as the final prophage sequences.

"Active prophages" are prophages with sequences that were detected in the viral contigs. We believe that such phages exist both as prophages and viral particles in the gastrointestinal tract. "Not activated prophages" are those with sequences that were not detected in the viral contigs. These are likely to be dormant phages. To assess the "activated" prophages that were induced and released from host bacteria, the detected prophage sequences were compared with the viral contig sequences. A prophage was considered "activated" when the prophage sequence aligned (MegaBLAST; ≥95% identity) with a viral contig sequence (>75% overlap in either sequence; all of the prophages detected by the second approach may be "activated" prophages).

Analysis of CRISPR Spacers

CRISPR regions are diverse across individuals, and therefore the bacterial contigs from individual samples were analyzed to determine the CRISPR regions. CRISPR repeats and spacers on bacterial contigs (≥5 kb) from individual samples were predicted using the CRISPR array identification program CRISPRDetect[20] (-array_quality_score_cutoff 3). The identified CRISPR spacers were linked to the representative (pooled) bacterial contigs using the clustering information from CD-HIT-EST, which was performed during contig pooling and the redundancy removal process.

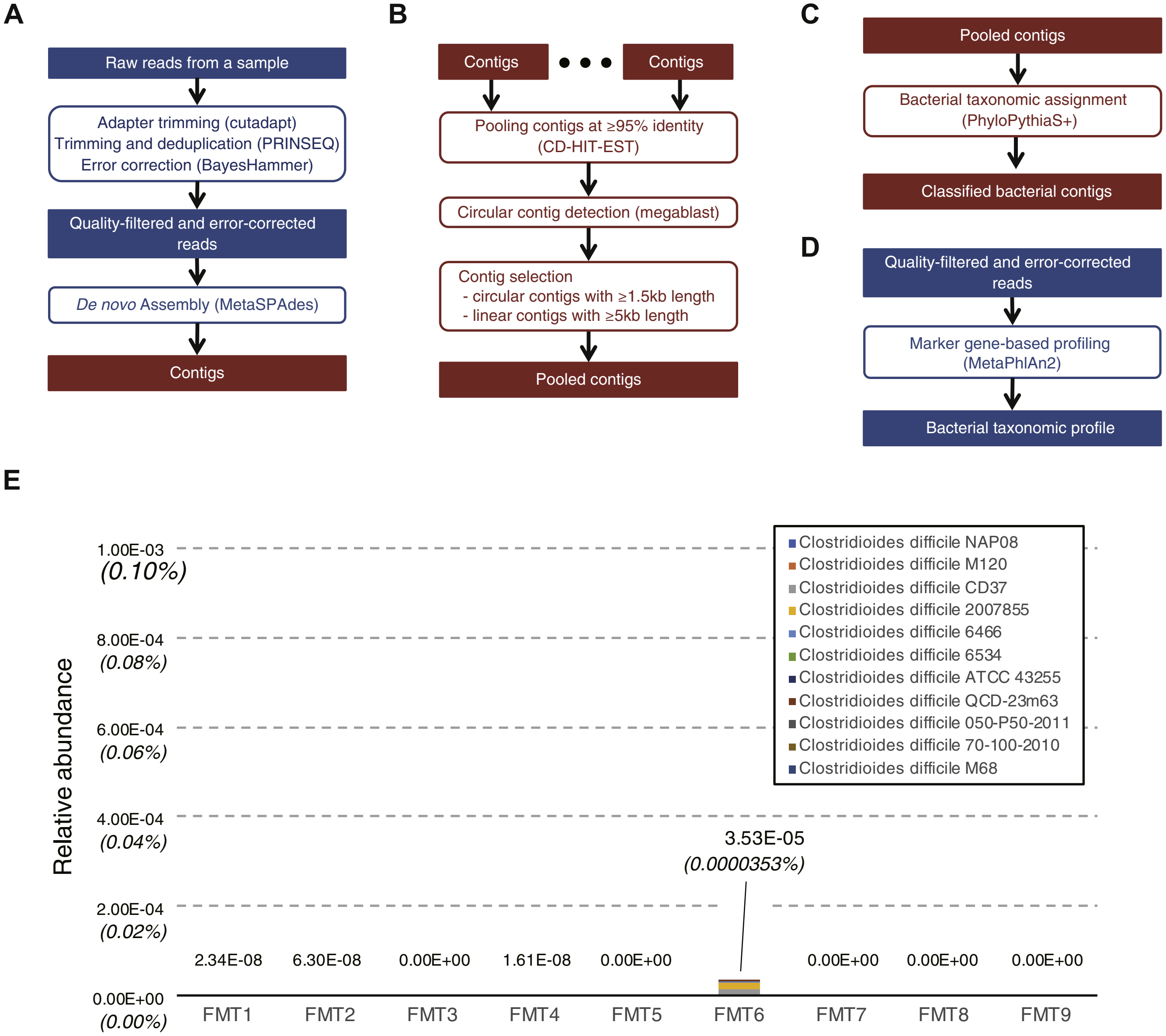
To identify the target phages of the CRISPR spacers, we queried the predicted spacers using BLASTN (BLAST+ v2.7.1)[5] against the viral contigs and extracted the aligned spacers when >90% of their lengths aligned with a minimum identity level of 95% and a maximum E value of 5×10^{-3}. The former 2 criteria were the most important; however, the E value cutoff permitted the removal of short spacers that may be false-positives. The viral contigs constructed by our virome pipeline were used for the target search.

Statistical Analysis of Gene Functions

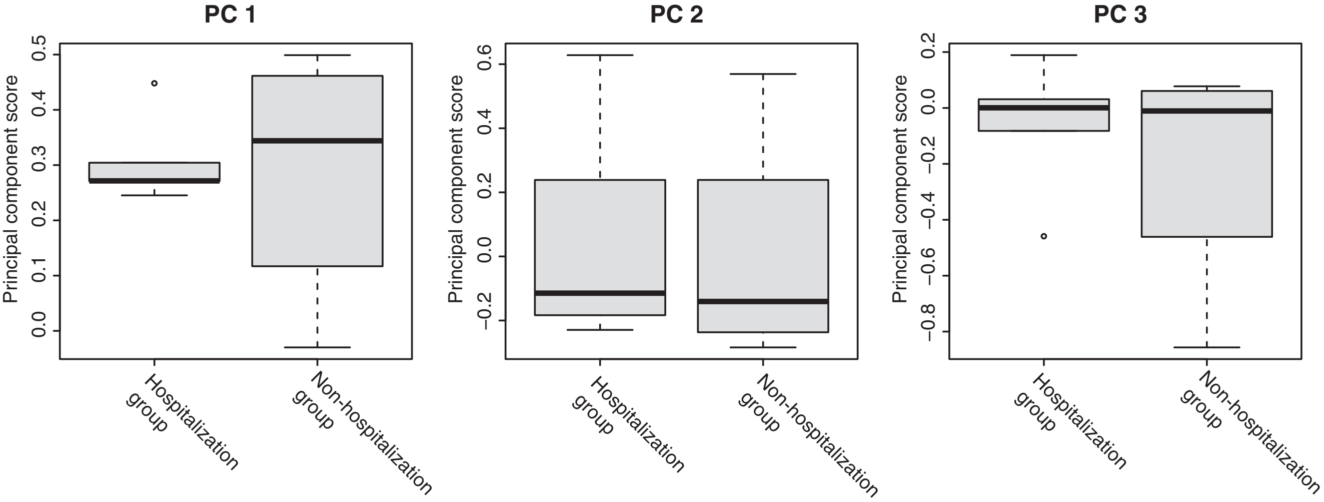
The detected ORFs in each sample were annotated according to KEGG prokaryote genes and corresponding KOs using the most significant hit (E value $< 1 \times 10^{-5}$ and bitscore > 50). For the bacteriome, the functional KEGG pathway was assessed using the ratio of the observed KOs divided by the total KOs constituting a particular KEGG pathway. KEGG pathways with observed KO ratios < 0.1 on average were removed from the analysis. The difference in the observed KO ratios between the pre- and post-FMT samples was evaluated using a paired Student *t* test. Compared with the bacteriome sequences, the virome sequences contained a larger number of unknown sequences that were not classified as KEGG genes. In fact, we found that a small number of KEGG pathways had KO ratios ≥ 0.1. Therefore, for the virome, the abundance of each KO term was used for comparisons, and its related pathways were analyzed. The number of detected ORFs annotated with a KO term was compared using a paired Student *t* test between the pre- and post-FMT samples. KO terms with average ORF counts of less than 0.5 were removed from the analysis. *P* values were corrected using the false discovery rate q-values estimated by the positive false discovery rate method using the *qvalue* R package. The heatmap was generated by R with the heatmap.2 function in the gplots library. Hierarchical clustering was also performed by the heatmap.2 function with default parameters. We used the prcomp function of R with default parameters for the principal component analysis.

Supplementary References

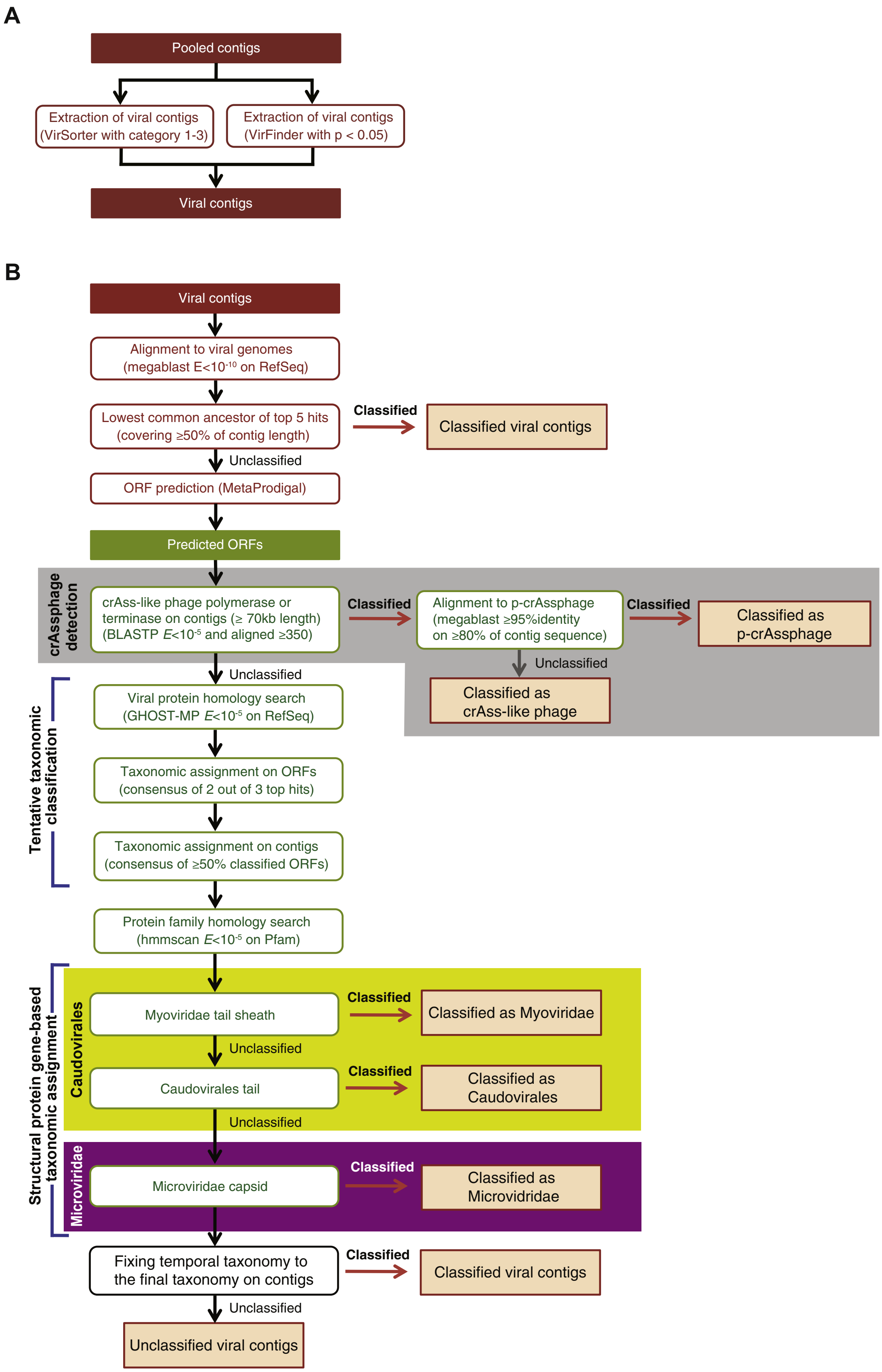
1. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–477.
2. **Nurk S, Meleshko D**, Korobeynikov A, et al. metaSPAdes: a new versatile metagenomic assembler. Genome Res 2017;27:824–834.
3. Pasolli E, Asnicar F, Manara S, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell 2019;176:649–662.e20.
4. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006;22:1658–1659.
5. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009; 10:421.
6. Roux S, Hallam SJ, Woyke T, et al. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. Elife 2015;4.
7. **Ren J, Ahlgren NA**, Lu YY, et al. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome 2017;5:69.
8. Hyatt D, LoCascio PF, Hauser LJ, et al. Gene and translation initiation site prediction in metagenomic sequences. Bioinformatics 2012;28:2223–2230.
9. Guerin E, Shkoporov A, Stockdale SR, et al. Biology and taxonomy of crAss-like bacteriophages, the most abundant virus in the human gut. Cell Host Microbe 2018;24:653–664.e6.
10. Deng L, Ignacio-Espinoza JC, Gregory AC, et al. Viral tagging reveals discrete populations in Synechococcus viral genome sequence space. Nature 2014;513:242–245.
11. Brum JR, Ignacio-Espinoza JC, Roux S, et al. Ocean plankton. Patterns and ecological drivers of ocean viral communities. Science 2015;348:1261498.
12. Kakuta M, Suzuki S, Izawa K, et al. A massively parallel sequence similarity search for metagenomic sequencing data. Int J Mol Sci 2017;18.
13. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 2011;39:W29–37.
14. Finn RD, Coggill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 2016;44:D279–D285.
15. Kang DW, Adams JB, Gregory AC, et al. Microbiota Transfer Therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study. Microbiome 2017;5:10.
16. Gregor I, Droge J, Schirmer M, et al. PhyloPythiaS+: a self-training method for the rapid reconstruction of low-ranking taxonomic bins from metagenomes. PeerJ 2016;4:e1603.
17. Truong DT, Franzosa EA, Tickle TL, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods 2015;12:902–903.
18. Roux S, Emerson JB, Eloe-Fadrosh EA, et al. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ 2017;5:e3817.
19. Roux S, Solonenko NE, Dang VT, et al. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. PeerJ 2016;4: e2777.
20. Biswas A, Staals RH, Morales SE, et al. CRISPRDetect: A flexible algorithm to define CRISPR arrays. BMC Genomics 2016;17:356.



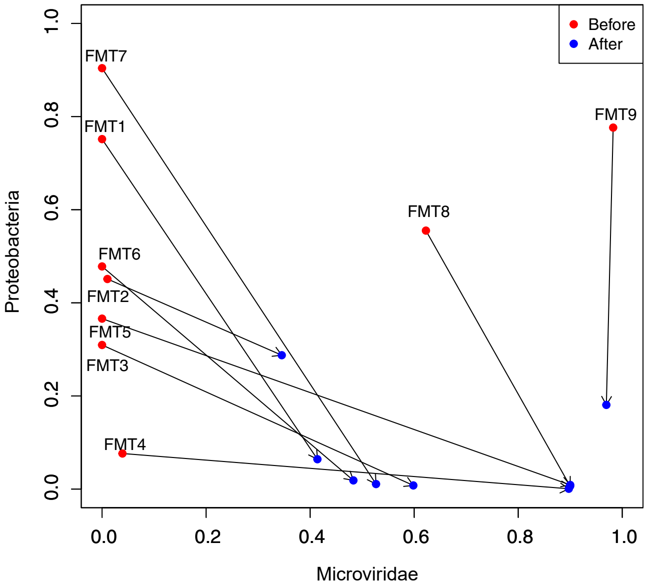
Supplementary Figure 1. Bacteriome analysis pipeline and the proportion of *C difficile* in the pre-FMT samples. (*A*) Analysis pipeline for the construction of contigs from bacterial sequence data. (*B*) Analysis pipeline for the construction of pooled contigs from the sample-derived contigs. (*C*) Analysis pipeline for the construction of classified bacterial contigs from the pooled contigs. (*D*) Analysis pipeline for the bacterial taxonomic profiles. (E) Relative abundance of *C difficile* in the pre-FMT samples.



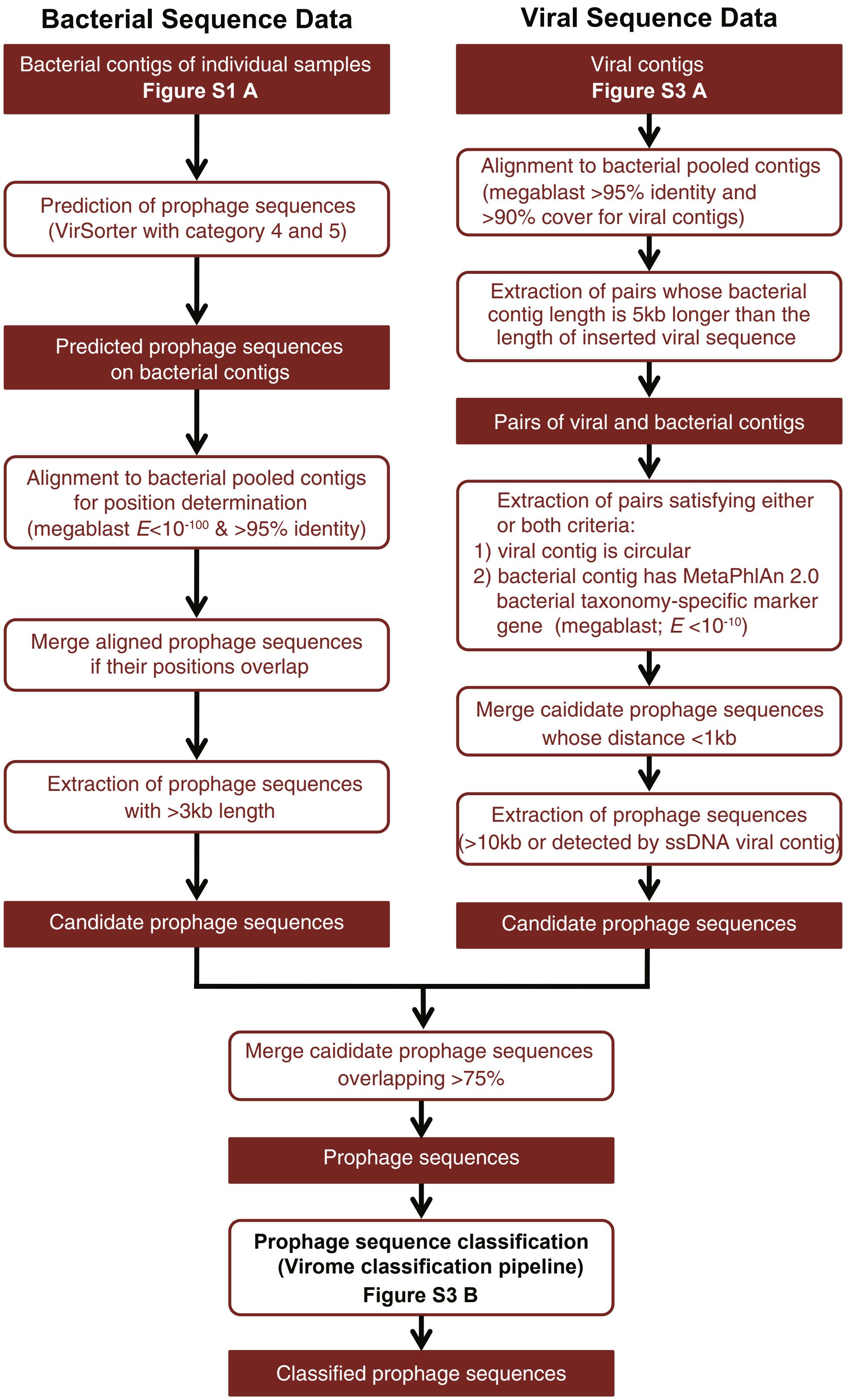
Supplementary Figure 2. Principal component (PC) scores. PC1, PC2, and PC3 scores for bacterial relative abundance between the hospitalization group (FMT2, FMT3, FMT6, FMT8, and FMT9) and the nonhospitalization group (FMT1, FMT4, FMT5, and FMT7).


A
Pooled contigs
Extraction of viral contigs (VirSorter with category 1-3)
Extraction of viral contigs (VirFinder with p < 0.05)
Viral contigs
B
Viral contigs
Alignment to viral genomes (megablast E<10^{-10} on RefSeq)
Lowest common ancestor of top 5 hits (covering ≥50% of contig length)
Classified
Classified viral contigs
Unclassified
ORF prediction (MetaProdigal)
Predicted ORFs
crAssphage detection
crAss-like phage polymerase or terminase on contigs (≥ 70kb length) (BLASTP E<10^{-5} and aligned ≥350)
Classified
Alignment to p-crAssphage (megablast ≥95%identity on ≥80% of contig sequence)
Classified
Classified as p-crAssphage
Unclassified
Classified as crAss-like phage
Unclassified
Tentative taxonomic classification
Viral protein homology search (GHOST-MP E<10^{-5} on RefSeq)
Taxonomic assignment on ORFs (consensus of 2 out of 3 top hits)
Taxonomic assignment on contigs (consensus of ≥50% classified ORFs)
Protein family homology search (hmmscan E<10^{-5} on Pfam)
Structural protein gene-based taxonomic assignment
Caudovirales
Myoviridae tail sheath
Classified
Classified as Myoviridae
Unclassified
Caudovirales tail
Classified
Classified as Caudovirales
Unclassified
Microviridae
Microviridae capsid
Classified
Classified as Microvidridae
Fixing temporal taxonomy to the final taxonomy on contigs
Classified
Classified viral contigs
Unclassified
Unclassified viral contigs

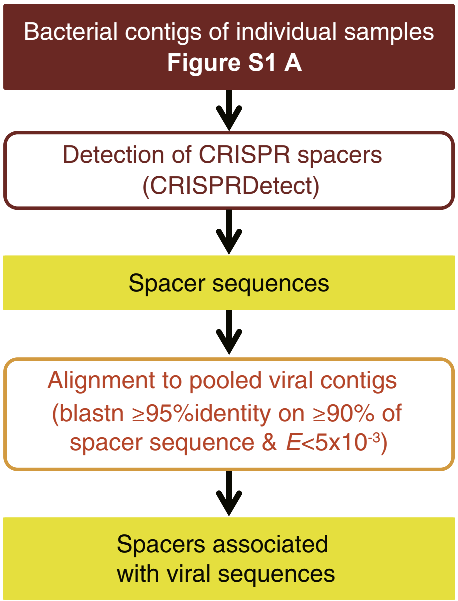
Supplementary Figure 3. Virome analysis pipeline. (*A*) Extraction of viral contigs using VirSorter and VirFinder. (*B*) Viral classification pipeline.



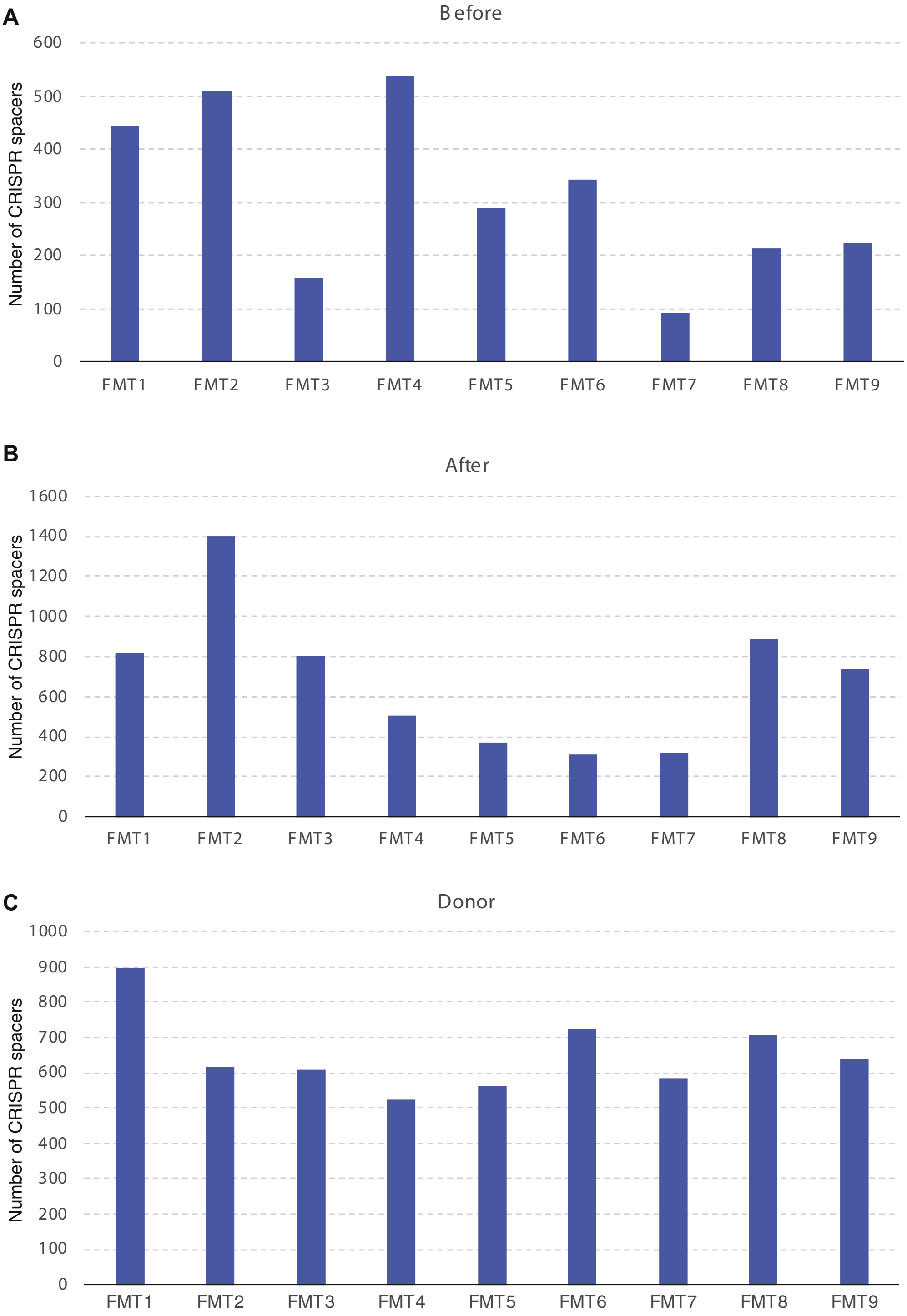
Supplementary Figure 4. Negative correlation between *Microviridae* abundance and Proteobacteria abundance. Comparison of the relative abundances of *Microviridae**** and Proteobacteria*** in fecal samples from patients with rCDI before and after FMT. ****P* < .001, paired Student *t* test (before FMT vs after FMT).


Bacterial Sequence Data
Viral Sequence Data
Bacterial contigs of individual samples
Figure S1 A
Prediction of prophage sequences (VirSorter with category 4 and 5)
Predicted prophage sequences on bacterial contigs
Alignment to bacterial pooled contigs for position determination (megablast $E<10^{-100}$ & >95% identity)
Merge aligned prophage sequences if their positions overlap
Extraction of prophage sequences with >3kb length
Candidate prophage sequences
Viral contigs
Figure S3 A
Alignment to bacterial pooled contigs (megablast >95% identity and >90% cover for viral contigs)
Extraction of pairs whose bacterial contig length is 5kb longer than the length of inserted viral sequence
Pairs of viral and bacterial contigs
Extraction of pairs satisfying either or both criteria:
1) viral contig is circular
2) bacterial contig has MetaPhlAn 2.0 bacterial taxonomy-specific marker gene (megablast; $E<10^{-10}$)
Merge caididate prophage sequences whose distance <1kb
Extraction of prophage sequences (>10kb or detected by ssDNA viral contig)
Candidate prophage sequences
Merge caididate prophage sequences overlapping >75%
Prophage sequences
Prophage sequence classification (Virome classification pipeline)
Figure S3 B
Classified prophage sequences

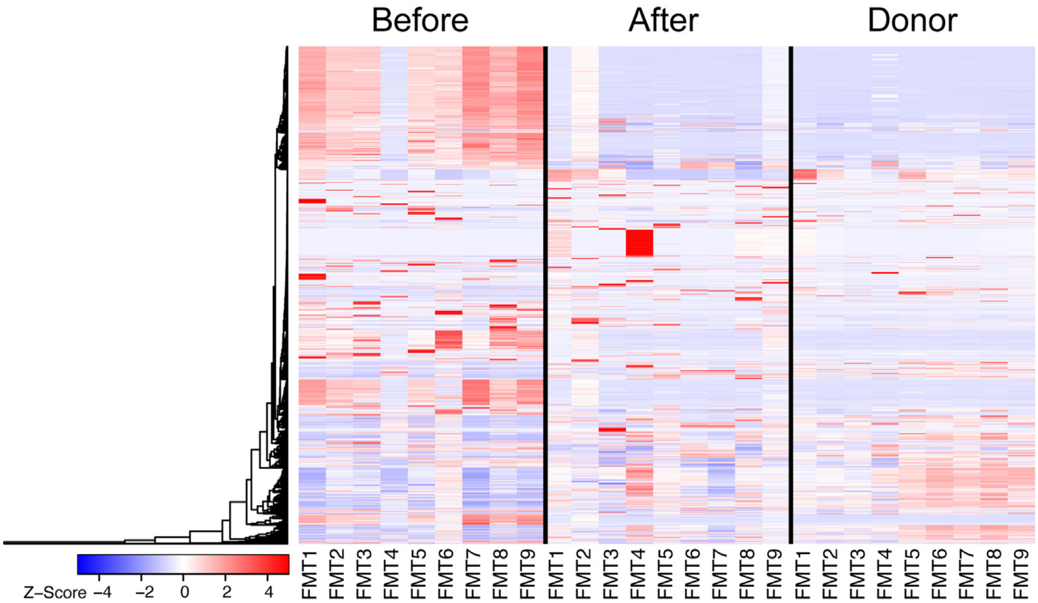
Supplementary Figure 5. Prophage detection pipeline. Prophage analysis pipeline based on bacterial and viral contigs.


Bacterial contigs of individual samples
Figure S1 A
Detection of CRISPR spacers
(CRISPRDetect)
Spacer sequences
Alignment to pooled viral contigs
(blastn ≥95%identity on ≥90% of spacer sequence & E<5x10-3)
Spacers associated with viral sequences

Supplementary Figure 6. CRISPR spacer detection pipeline. Pipeline for the detection of CRISPR spacer and CRISPR-based host–parasite associations based on bacterial and viral contigs.



Supplementary Figure 7. CRISPR spacer numbers. The number of identified CRISPR spacers. (*A*) Pre-FMT samples. (*B*) Post-FMT samples. (*C*) Donor samples.



Supplementary Figure 8. Bacterial functional profile distribution in recipients before and after FMT and in donors. Gene-level functional profiles across samples before FMT, after FMT, and in donors, for 6082 KEGG Orthology genes (E value $< 1 \times 10-5$ and bitscore > 50).