

# Diversity and Abundance of Single-Stranded DNA Viruses in Human Feces<sup>∇†</sup>

Min-Soo Kim, Eun-Jin Park, Seong Woon Roh, and Jin-Woo Bae\*

Department of Life and Nanopharmaceutical Sciences and Department of Biology, Kyung Hee University, 1 Hoegi-dong, Dongdaemun-gu, Seoul 130-701, Republic of Korea

Received 26 July 2011/Accepted 9 September 2011

**In this study, we investigated the abundance and diversity of single-stranded DNA (ssDNA) viruses in fecal samples from five healthy individuals through a combination of serial filtration and CsCl gradient ultracentrifugation. Virus abundance ranged from 10<sup>8</sup> to 10<sup>9</sup> per gram of feces, and virus-to-bacterium ratios were much lower (less than 0.1) than those observed in aquatic environments (5 to 10). Viral DNA was extracted and randomly amplified using phi29 polymerase and analyzed through high-throughput 454 pyrosequencing. Among 400,133 sequences, an average of 86.2% viromes were previously uncharacterized in public databases. Among previously known viruses, double-stranded DNA podophages (52 to 74%), siphophages (11 to 30%), myophages (1 to 4%), and ssDNA microphages (3 to 9%) were major constituents of human fecal viromes. A phylogenetic analysis of 24 large contigs of microphages based on conserved capsid protein sequences revealed five distinct newly discovered evolutionary microphage groups that were distantly related to previously known microphages. Moreover, putative capsid protein sequences of five contigs were closely related to prophage-like sequences in the genomes of three *Bacteroides* and three *Prevotella* strains, suggesting that *Bacteroides* and *Prevotella* are the sources of infecting microphages in their hosts.**

Viruses, particularly bacteriophages, are one of the main drivers of mortality and the evolutionary change of microorganisms through horizontal genetic transfer (HGT) (40, 59). This viral activity is believed to play a significant role in nutrient cycling and carbon flow in biogeochemical and ecological processes (47, 61). However, the investigation of the ecological role of viruses has been focused on aquatic systems, especially on the marine environment, even though viruses are believed to be ubiquitous in all ecosystems (40, 59).

The human gastrointestinal tract, considered a forgotten organ, harbors an overwhelmingly large number of unknown microbes, such as bacteria, archaea, microbial eukarya, and viruses, that are armed through direct selective pressures from the immune system (27). The maintenance and compositional changes in the gut microbiota are known to be closely linked to human physiology, nutrition, and the prevalence of disease (56). Recently the connection between an altered gut microbiota and the pathogenesis of metabolic syndromes or inflammations, such as obesity, diabetes, and inflammatory bowel disease, has been increasingly reported (17, 57, 60).

Given the critical contribution of viruses to host mortality and genetic diversity in the ecosystem, the ecological role of viruses also might be of significance to the microbial ecology of the gut. The functional redundancy of the gut microbiota not only confers stability for their fitness (27) but also contributes to the complementation of metabolic features not coded by the

human genome (37). The frequent occurrence of HGT among gut microbes (39, 48) and the observation of a large number of phage-related genes from the human gut microbial metagenome implies a role of viruses in gut homeostasis (37). Still, human gut viruses remain largely unknown and have not received much attention as an important constituent of the gut microbiome. Few studies have examined viral diversity in the human gut, and those studies examined samples from a few individuals using shotgun library construction with limited resolution (4, 5, 67). Recent work characterizing fecal viromes has found that the viral-microbial interactions in the human intestine potentially were different from the predator-prey relationship, known as “kill the winner,” exhibited predominantly in many other environments (38).

With the development of advanced molecular techniques, our knowledge of viral diversity has been widely revolutionized. Increased attention has been paid to the study of viruses in a wider range of environments as the diversity and innovations associated with viruses have been shown to be much higher than previously recognized. The diversity and abundance of single-stranded DNA (ssDNA) viruses, particularly microphages, have been discovered using multiple-displacement amplification (MDA) with phi29 polymerase, which made it possible to study ssDNA viruses by taking advantage of random priming and the preferential amplification of circular genomes (16). As a result, a high genotypic diversity of ssDNA viruses has been found in many studies; environments as diverse as rice paddy soil (23), microbialites (12), seawater (1), reclaimed water (42), and an Antarctic lake (28) have been investigated. However, no evidence for elucidating their host ranges has been reported yet (53).

Based on the recognition of the coexistence of viruses and gut bacteria (5), further research is needed to determine their functional role and the interactions among viruses, bacteria,

\* Corresponding author. Mailing address: Department of Life and Nanopharmaceutical Sciences and Department of Biology, Kyung Hee University, 1 Hoegi-dong, Dongdaemun-gu, Seoul 130-701, Republic of Korea. Phone: 82 2 961 2312. Fax: 82 2 961 0244. E-mail: baejw@khu.ac.kr.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

∇ Published ahead of print on 23 September 2011.

and gut epithelial cells. In this study, we analyzed randomly amplified fecal viromes from five healthy individuals by 454 pyrosequencing to characterize the genetic diversity and composition of DNA viruses, especially ssDNA viruses. We also determined the diversity and structure of bacterial communities based on 16S rRNA genes to compare them to the viral assemblages. We believe that this study should greatly improve our knowledge of the diversity of unknown ssDNA viruses in the human intestine.

## MATERIALS AND METHODS

**Sample preparation.** All procedures were reviewed and approved by the Kyung Hee University Institutional Review Board (KHU IRB 2010-008). Five healthy and similarly aged adults (23, 27, 28, 28, and 29 years old) living in Seoul, South Korea, were selected for the analysis of viral diversity. Participants had no known illnesses related to the gastrointestinal tract and had balanced meals at regular times. The procedures used for the collection of viruses were the same as those described by Thurber et al. (54). About 10 to 15 g of fresh fecal samples from each individual, designated F-A, F-B, F-C, F-D, and F-E (see Table S1 in the supplemental material), was collected and resuspended by stirring them into 10 volumes of SM buffer (0.1 M NaCl, 50 mM Tris-HCl [pH 8.0], 10 mM MgSO<sub>4</sub>). An additional serial filtration of the suspensions using 10- $\mu$ m (Whatman, Maidstone, United Kingdom), 5- $\mu$ m (Advantec, Tokyo, Japan), 1- $\mu$ m (Advantec), 0.45- $\mu$ m (Millipore, Billerica, MA), and 0.22- $\mu$ m nitrocellulose filter papers (Millipore) and a 0.20- $\mu$ m syringe filter (Satorius, Goettingen, Germany) was employed to isolate the viral particles. CsCl layers of 1.7, 1.5, and 1.2 g ml<sup>-1</sup> were used to purify the viral particles using density gradient ultracentrifugation at 88,250  $\times$  g for 2 h at 4°C in a swinging bucket rotor (Optima L-100 K; Beckman Coulter, Fullerton, CA). To harvest viruses, a 1.2- to 1.5 g ml<sup>-1</sup> fraction was collected and concentrated using centrifugal concentration filters (Amicon Ultra-15; 30 kDa; Millipore) before the extraction of the viral nucleic acids.

**Viral and bacterial enumeration.** Viral and bacterial enumeration were performed as previously described (54). Briefly, 0.20- $\mu$ m filtrates for viruses and fecal suspension for bacteria were diluted and attached onto 0.02- $\mu$ m Anodisc filters (Whatman). Filter membranes were stained with 2 $\times$  SYBR gold (Invitrogen, Carlsbad, CA) for 15 min in the dark and rinsed with 1 ml of sterile water. After being air dried, the filters were mounted onto a glass slide with an antifade agent for visualization and quantification by confocal laser-scanning microscopy (CLSM; LSM 510, Carl Zeiss MicroImaging). For the enumeration of viruses and bacteria, more than 10 fields of view and at least 200 viruses and bacteria from each sample were observed. Particles smaller than 0.5  $\mu$ m were considered viruses, whereas those ranging from 0.5 to 10  $\mu$ m were classified as bacteria.

**Morphology of viruses.** For the morphological characterization of viruses, the same density gradient purification procedure as that used by Park et al. (34) was conducted. Briefly, the fraction collected during sample preparation was pelleted by ultracentrifugation at 100,000  $\times$  g for 3 h at 4°C. Pellets were resuspended into SM buffer and purified by a second CsCl density gradient at 88,250  $\times$  g for 2 h. Collected viruses from a 1.2- to 1.5 g ml<sup>-1</sup> fraction were resuspended with 10 volumes of deionized water and concentrated into final volumes of less than 0.1 ml using centrifugal concentration filters (Amicon Ultra-15; 30 kDa). Copper grids with 200-mesh carbon-coated Formvar were floated on a droplet of the concentrated sample, negatively stained with 2% uranyl acetate, washed with deionized water, and air dried. Images that were magnified 50,000 $\times$  to 100,000 $\times$  were taken using an energy-filtering transmission electron microscope (LIBRA 120; Carl Zeiss, Germany).

**Viral DNA extraction and sequencing.** Before the extraction of the viral nucleic acids, viruses stained by SYBR gold were observed to ensure the absence of microbial cells. Subsequently, DNase I (25 U/ml at 37°C for 1 h; Takara Bio, Japan) was added to the viral concentrates of each sample to remove any free DNA. Viral genomic DNA was extracted using the QIAamp MinElute virus spin kit (Qiagen, Valencia, CA). Extracted viral genomic DNA, starting at 40 to 50 ng per sample, was amplified using the GenomiPhi V2 DNA amplification kit (Amersham Biosciences, NJ) in a 1-h reaction to minimize amplification bias toward circular genomes (23). The amplification of viral DNA was performed in duplicate, and samples were pooled prior to pyrosequencing. Amplified viral DNA (approximately 10  $\mu$ g) from each sample was sequenced with the Genome Sequencer FLX titanium (454 Life Sciences, CT). An eighth of a PicoTiterPlate device was used for each sample.

**16S rRNA gene amplification and barcoded pyrosequencing for the investigation of bacterial diversity.** Approximately 1 g of each fecal sample was fully homogenized by a pestle supplemented with 5 volumes of liquid nitrogen.

Genomic DNA then was extracted, precipitated, and purified according to the improved fecal DNA extraction method (66). Highly variable regions, V1 and V2, of the 16S rRNA gene sequences were amplified from the extracted bacterial DNA using barcode primers (8F and 518R) (3). A bacterial forward primer designed according to a previous study by Dethlefsen et al. (13) and reverse primer 518R (5'-WTTACCGCGGCTGCTGG-3') were used for amplification. Specifically, to enable the recognition of sequences from each sample, barcodes (4 to 8 nucleotides; TACG, ATAGC, CATCTG, ATAGTGC, and TAGCA TCG) and linkers (2 nucleotides; AC) were tagged at the 5' ends of the primers. For the libraries of each sample, eight replicated PCR amplicons were mixed to minimize PCR bias (36). A 30- $\mu$ l PCR mix (2 $\times$  solution type *Taq* premix; Solgent, South Korea) was prepared containing each barcode-primer set. To each reaction mixture, 1  $\mu$ l (5 to 10 ng) of the extracted DNA template was added. PCR conditions were 94°C for 2 min; 15 cycles of 94°C for 20 s, 55°C for 10 s, and 72°C for 40 s; and a final extension at 72°C for 5 min. Tubes containing only barcode-primer sets were amplified to serve as a negative control. The eight PCR replicates were pooled and purified with the QIAquick PCR purification kit (Qiagen). Three additional cycles of amplification were employed under the same PCR conditions as those described above (except using an annealing temperature of 60°C) to minimize heteroduplex molecules (52). Products were purified with the QIAquick PCR purification kit. Finally, the pooled DNA (1  $\mu$ g from each sample) was sequenced with the 454 pyrosequencing Genome Sequencer FLX titanium. An eighth of a PicoTiterPlate device was used to sequence the samples.

**Bioinformatics. (i) Preprocessing.** To analyze the viral metagenomic sequences, raw sequence reads of each sample were filtered to minimize the effects of poor sequence quality and sequencing errors. Sequences containing more than one ambiguous base call (Ns), average quality scores lower than 25, and a length shorter than 100 bp were removed using MOTHUR v.1.10.2 (44). Exact duplicates, a known systematic artifact generated by 454-based pyrosequencing, were excluded using the 454 replicate filter using the CD-HIT program (18) provided by the CAMERA 2 server (<https://portal.camera.calit2.net>), using a minimum sequence identity of 100% and 1 bp as a maximum difference in length (33). Bacterial raw sequence reads were filtered in the same way as the viral metagenomic sequences. Sequence reads were sorted into each sample according to barcode-tagged primer sequences and then aligned using Pyrosequencing Aligner in RDP (8). Barcode and primer sequences were manually trimmed, keeping the V1 and V2 regions, using the BioEdit program (19). Reads including either a sequencing error in the barcode and primer regions or reads shorter than 301 bp were discarded.

**(ii) Viral taxonomic identification.** Sequence reads from each virome were compared to the SEED nonredundant (nr) protein database (14), CAMERA's nonidentical (nr) peptide sequences, and viral proteins databases (<http://camera.calit2.net>; v2.0.5.0). The NCBI reference data sets of CAMERA contain the sequence data released from GenBank (release 177) and RefSeq (release 40), and the MEGAN program (v3.9) (22) was used for CAMERA searches based on the lowest-common-ancestor algorithm with two minimal support hits. Sequences were searched against three public databases using the BLASTx algorithm. Matches with expected (E) values of less than 10<sup>-5</sup> for the SEED nr database, 10<sup>-3</sup> for CAMERA nr peptide sequences, and 10<sup>-5</sup> for CAMERA viral proteins were used for identification. The best hit with an annotated protein sequence in the three databases was classified as known, whereas those sequences with no matches at less than the E value were classified as unknown. The relative abundance of viral families within each virome was determined by normalizing hit numbers based on genome size and correcting for the phi29 polymerase bias (preferentially amplified as 100 times more [23]). Sequences classified as RNA viruses (only four hits in all BLAST results) were excluded because only DNA was used for sequencing in this study.

**(iii) Structure and diversity estimation.** The diversity and structure of the viral community from the five viromes were determined using alpha diversity analysis in the CAMERA 2 server as previously described (62). The best-rank abundance model of the five viromes was the lognormal model with the lowest error rate. The diversity and structure of bacterial communities were determined with MOTHUR v.1.10.2 (44). A sequence similarity of 97% was used as a cutoff for the assignment of operational taxonomic units (OTUs) in this study. Species richness estimators (Chao1 and Ace) and a biodiversity index (Shannon-Wiener) were calculated.

**(iv) Phylogenetic diversity of the microphage sequences.** To determine the genetic diversity of ssDNA viruses, a phylogenetic analysis of the sequences assigned to the microphages was carried out as previously described (12). Identical sequences preprocessed from the five viromes were assembled (provided they met the 98% minimum identity and 35-bp minimum overlap criteria) using the 454 *de novo* assembler (Roche Applied Science). From the assembled con-

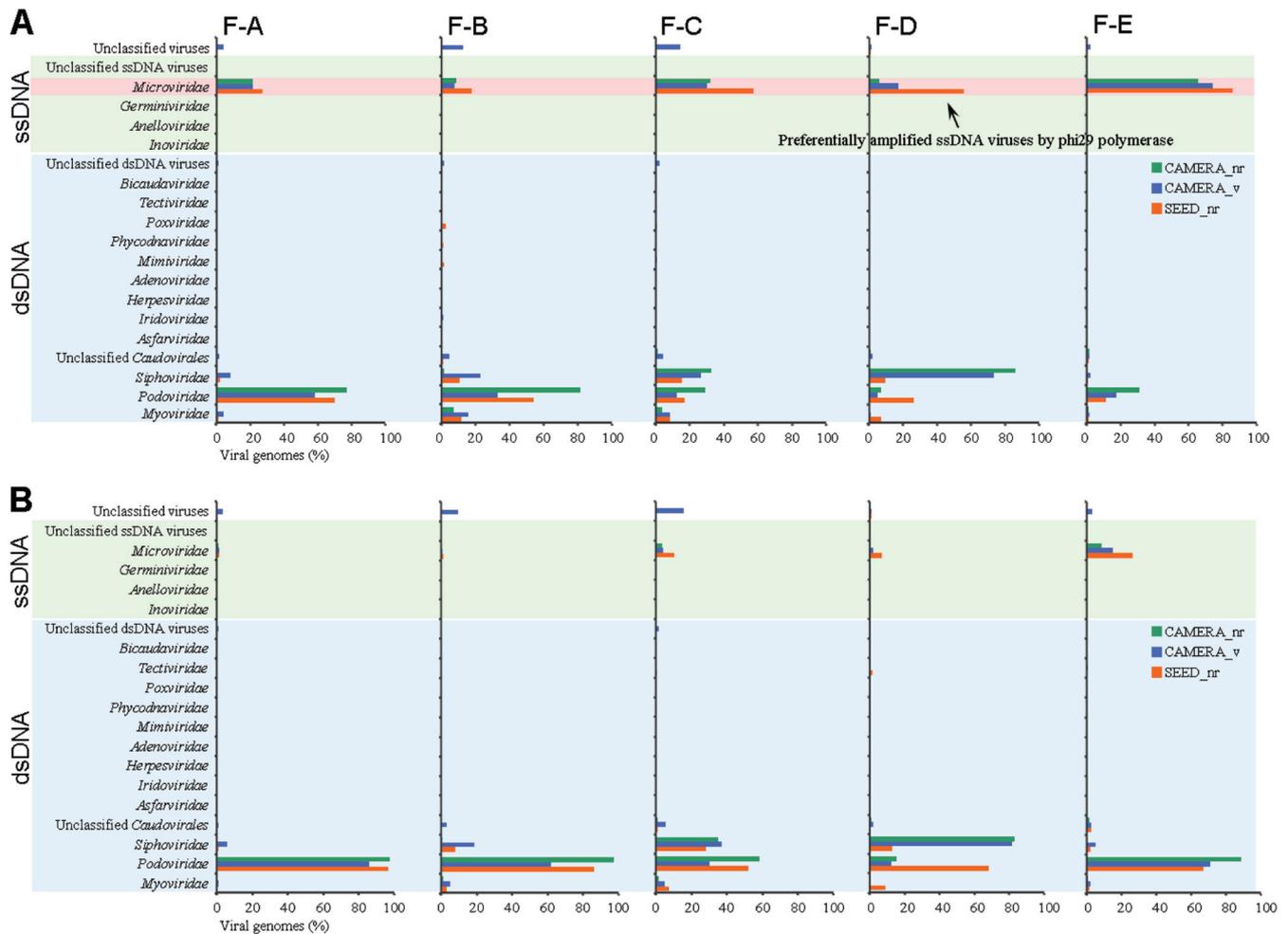


FIG. 1. Relative abundance of viral families in five human gut viromes. Viral sequences were compared to public databases (SEED nr, CAMERA nr, and CAMERA viral protein databases). Viral sequences were classified into viral families based on best matches. (A) Distribution of viral families based on sequence matches. (B) Distribution of viral families corrected by the phi29 polymerase bias and normalized by genome size.

tigs, large contigs (>500 bp) related to the putative major capsid protein gene of microphages (tBLASTx against viral genome database; E value,  $<10^{-5}$ ) were extracted, and predicted open reading frames (ORF) were obtained by using an ORF finder with a six-reading-frame translation (65). Multiple alignments of partial sequences of major capsid protein from large contigs with 13 sequences from the Sargasso Sea (1), 8 PCR products from marine stromatolites (12), 2 reconstructed contigs from the viromes of Antarctic lake (28), 15 sequences of cultured members in the family *Microviridae*, and 6 putative capsid protein sequences in the genomes of *Bacterioides* (*B. plebeius* DSM 17135, *B. eggerthii* DSM 20697, and *Bacterioides* sp. strain 2\_2\_4) and *Prevotella* (*P. bergensis* DSM 17361, *P. buccalis* ATCC 35310, and *Prevotella* sp. strain oral taxon 317) in the data from the human microbiome project (32) were performed using MUSCLE (<http://www.ebi.ac.uk/Tools/muscle>) with default settings. The latter were archived from BLASTp searches against the GenBank nr database. The phylogenetic tree was constructed using MEGA 4.0 (49) and was tested by 1,000 randomly replicated bootstraps in the neighbor-joining algorithm (43).

(v) **Metabolic analysis of viromes.** To determine the metabolic profiles of the five viromes, viral sequence reads and large contigs (>500 bp) were compared to the SEED subsystem data set (14) by BLASTx searches (E value,  $<10^{-2}$ ). Best matches were used for the inference of the functions of the genes from the five viromes.

**Virome accession numbers.** The fecal viromes from five volunteers are accessible in the NCBI Short Read Archive under accession number SRP005097. The viromes also are accessible in the SEED database under accession numbers 4449580.3 to 4449584.3.

## RESULTS AND DISCUSSION

**Identification of DNA viruses in human feces.** We obtained enriched viruses from the fecal samples of five individuals, which were purified with a combination of size-dependent filtration and CsCl gradient ultracentrifugation. Randomly amplified viral genomes were used for high-throughput shotgun sequencing, with an average of 100,929 sequences achieved (see Table S2 in the supplemental material). Following the quality-based screening of the raw data, an average of 81,347 sequences (longer than 100 bp, with an average read length of 448 bp) was used to characterize the fecal viral assemblage. Based on the analyses of viral sequences, 72.8 to 93.7% of the sequence reads (classified as unknown) were not identified in BLASTx searches against three public databases, and a small number of the sequences (classified as known; 6.3 to 27.2%) showed a significant degree of similarity to microbial groups, including viruses (see Table S2). The present results are similar to those of recent viral metagenomic studies showing that approximately 40 to 50%, or occasionally up to 90%, of viral sequences from environmental samples are uncharacterized

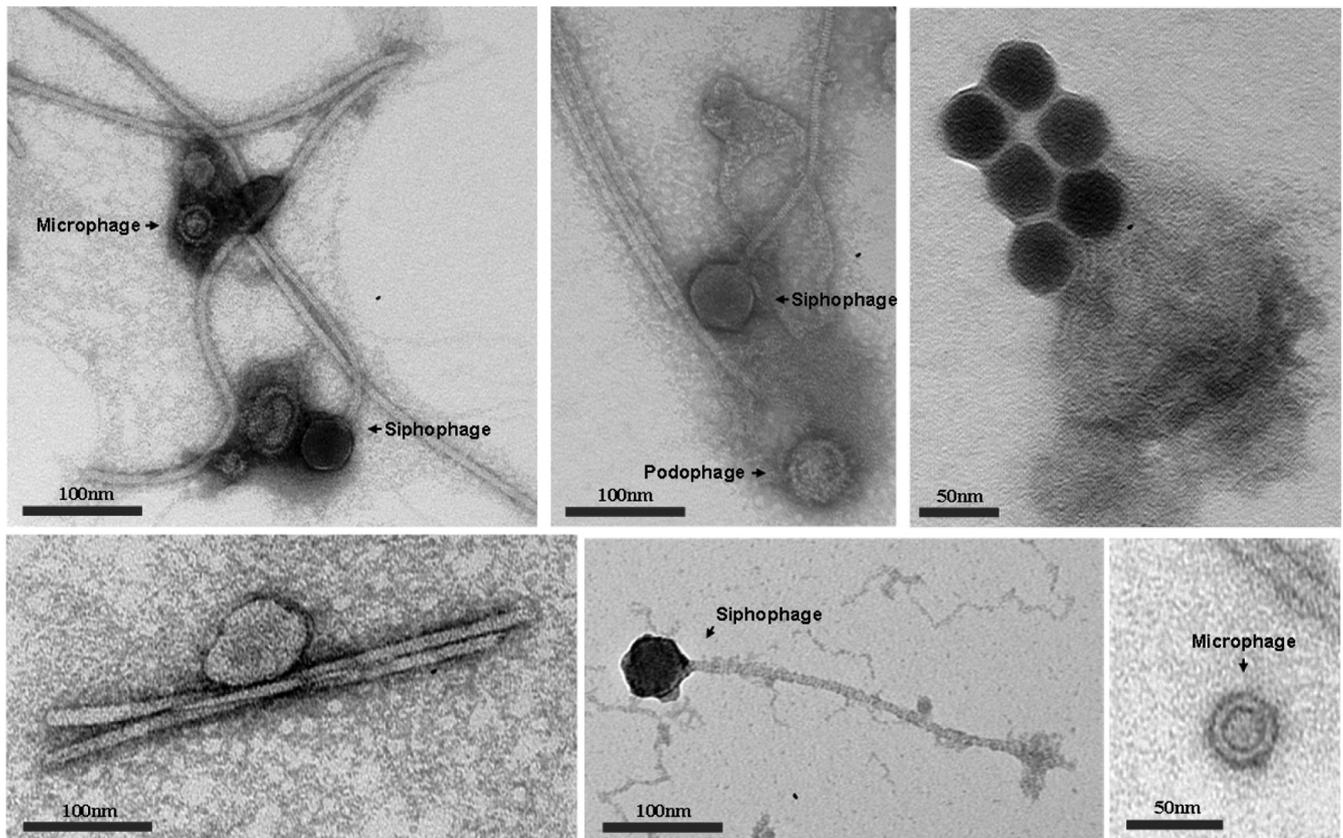


FIG. 2. Transmission electron micrographs showing diverse viral morphologies in human feces. Scale bars are shown on the bottom of each panel.

(24). This indicates that currently known viruses do not adequately represent the viral assemblages, as most viruses are unknown to science (40). Data accumulated so far on bacteriophages are especially sparse compared to those for animal and plant viruses. Indeed, most phages infecting bacteria belong to only three phyla: *Proteobacteria*, *Firmicutes*, and *Actinobacteria* (24). Numerous metagenomic studies on the microbiota of the human gut demonstrated the predominance of core gut bacteria, such as *Faecalibacterium*, *Alistipes*, *Bacteroides*, *Blautia*, *Roseburia*, *Ruminococcus*, *Dorea*, and *Bifidobacterium* (50). Still, except for the only *Bacteroides* phage (B40-8) and six *Clostridium* phages (c-st, phiCD119, phi3626, phiC2, phiCD27, and phiCTP1), no viruses that infect these dominant genera have been isolated. Thus, the isolation and characterization of novel viruses in the human intestine should be a priority to enable the description of its viral assemblage and its role in the gut microbiota.

While previously known viral sequences from five viromes were classified into 16 viral families (Fig. 1A; also see Table S3 in the supplemental material), most sequences were concentrated in only four families: double-stranded DNA (dsDNA) viruses that belong to three families of the *Caudovirales* order, podophages (25 to 45%), siphophages (8 to 27%), and myophages (3 to 6%), and ssDNA viruses that belong to the microphage family (27 to 49%). Due to the preferential amplification by phi29 polymerase (23), a large portion (27 to 49%) of the sequences shared a high degree of similarity

to ssDNA microphages containing small circular genomes. To determine the relative abundance of each viral family, the overestimated abundance of microphages was corrected by considering the phi29 polymerase bias and normalizing it by genomic size (Fig. 2B). The corrected abundances suggest that the fecal viromes were dominated by podophages (52 to 74%), siphophages (11 to 30%), microphages (3 to 9%), and myophages (1 to 4%). The morphologies of dominant phages from unknown viruses are shown in Fig. 2. Eukaryotic DNA viruses, such as large nucleocytoplasmic DNA viruses (NCLDVs), including poxviruses, asfarviruses, iridoviruses, mimiviruses, and phycodnaviruses, as well as herpesviruses, were detected only rarely (less than 1%). No sequence was classified as being an ascovirus, which is a member of the NCLDV families (see Table S3).

**Low abundance and low diversity of viruses in human feces.**

The abundance of viruses and bacteria in the five fecal samples was calculated using CLSM by staining with SYBR Gold. Viral abundance ranged from  $1.1 \times 10^8$  to  $1.5 \times 10^9$  viruses  $g^{-1}$  (wet weight), whereas bacterial abundance ranged from  $3.9 \times 10^9$   $g^{-1}$  to  $7.6 \times 10^9$  bacteria  $g^{-1}$  (wet weight) (Fig. 3). This is the first research to determine accurate total viral abundance in human fecal samples, which is slightly higher than that in colonic biopsy specimens ( $1.2 \times 10^8$  viruses biopsy specimen $^{-1}$ ) (26). Because the direct count of fluorescent particles is a well-defined and reliable technique with high accuracy (21), many studies of viral ecology conducted in various envi-

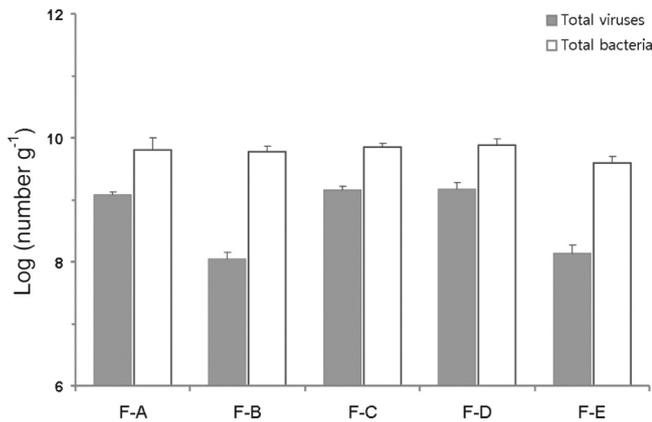


FIG. 3. Abundance of viruses and bacteria in human feces determined by fluorescence staining with confocal laser-scanning microscopy.

ronments employed this method (21, 35, 42). Virus-to-bacterium ratios (VBRs) also were defined, and they were in the range of 0.019 to 0.209 (means  $\pm$  standard deviations [SD],  $0.129 \pm 0.094$ ). These ratios are considerably lower than those reported for other environments, such as seawater and marine sediment. In aquatic environments, VBR have been shown to range from 5 to 10 (63), and in marine sediments a wide range, from 0.11 to 71, has been described (9, 10, 15). In soil, however, a low VBR of 0.04 was reported (2).

With the lower abundance of viruses than bacteria, the lower indexes of species richness and Shannon biodiversity were calculated for the viral assemblages than those obtained for the bacterial communities, as computed according to the *in silico* comparison of viral and bacterial diversity (see Table S4 in the supplemental material). While bacterial species richness was estimated as  $1,372 \pm 305$  OTUs in Chao1 and  $1,907 \pm 490$  OTUs in Ace, species richness was substantially lower for the five viromes examined, ranging from 18 to 401 viral genotypes (means  $\pm$  SD,  $101 \pm 168$ ). Similarly, the diversity of the viral assemblages (Shannon index,  $3.09 \pm 0.64$  nats) was lower than that of the bacterial communities (Shannon index,  $4.54 \pm 0.17$  nats) in the fecal samples. Based on the kill-the-winner model, the ecological role of viruses is as regulators of bacterial populations through viral predation (46). Accordingly, inactive virus forms with diverse genotypes and high abundance are distributed ubiquitously in the environment (6). However, the observed low VBR and reduced diversity of fecal viromes in this study do not support the kill-the-winner model.

In addition, many sequences were categorized prominently as prophage-related genes. The functional profiling of the fecal

viromes showed that many viral sequences and large contigs were identified as lysogeny-related genes of the virulence category and subsequently prokaryotic metabolism-related genes of the cellular category, although only a few were identified (E value of  $<10^{-2}$ ;  $3.3\% \pm 1.3\%$  for viral sequences and  $18.9\% \pm 4.1\%$  for large contigs) (see Fig. S1 in the supplemental material). Based on the recent work of Reyes et al. (38) characterizing fecal viromes, a few temperate phages dominating with a high genetic stability were present in the human intestine; consequently, the association between viruses and microbial hosts in the human intestine cannot be described as a predator-prey relationship driven by the lytic cycle. Considering the gene transfer capacity of viruses via lysogenic conversion, prokaryotic metabolism-related genes in the viromes might originate from gut bacteria as lysogenic conversion genes (7). Currently, the role of prophage elements in their host genomes has been gaining increased attention, with their beneficial effects on their hosts' phenotypes beginning to be understood (7, 41, 58). According to these studies, the bacteriophages in the human intestine might be helpful for gut bacteria to enable their colonization and adaptation in the intestinal community.

**Divergence of ssDNA viruses in viromes.** Based on the composition of the viral assemblage in the human feces, ssDNA viruses were identified, with members of the viral family *Microviridae* being the major constituents of fecal viromes. However, no other ssDNA viruses, such as circovirus, nanovirus, or parvovirus, were detected. Similarly, anellovirus, geminivirus, and inovirus were hardly observed (see Table S3 in the supplemental material). The genetic diversity of the ssDNA microphages in the fecal viromes was determined based on a conserved major capsid protein (pfam02305) of microphages. From the assemblies of viral sequences, we found a total of 45 contigs with significant similarities (tBLASTx; E value of  $<10^{-5}$ ) to *Chlamydia* microphage, *Spiroplasma* microphage, and *Bdellovibrio* microphage (Table 1). Of these, 24 contigs contained nearly full-length capsid protein sequences (F-A, 8 sequences; F-B, 3 sequences; F-C, 3 sequences; F-D, 6 sequences; and F-E, 4 sequences). To characterize the contigs, we first analyzed putative capsid protein sequences of the five viromes by comparing them to the GenBank nr database in BLASTp searches (Table 2). All capsid protein sequences from the five viromes shared a significant degree of similarity to known capsid protein sequences of microphages (33 to 56% identities). Replication initiation protein sequences (Rep; PHA0330) also were observed in the 24 contigs, and 20 contigs showed a similarity to known Rep sequences of microphages (27 to 42% identities) (Table 2). Four contigs did not contain Rep sequences due to their short length. Surprisingly, five

TABLE 1. Overview of the total number of viral sequences and contigs for five fecal samples

Sample	No. of raw reads	High-quality, nonredundant sequences	No. of all assembled contigs (>100 bp)	No. of singletons	No. of large contigs (>500 bp)	No. of known contigs	No. of microphage contigs
F-A	113,054	90,273	634	2,945	465	70	9
F-B	109,569	88,810	810	2,930	571	82	4
F-C	68,391	53,342	737	1,783	549	90	10
F-D	115,121	95,902	367	2,210	246	43	17
F-E	98,511	71,806	296	2,825	181	35	5



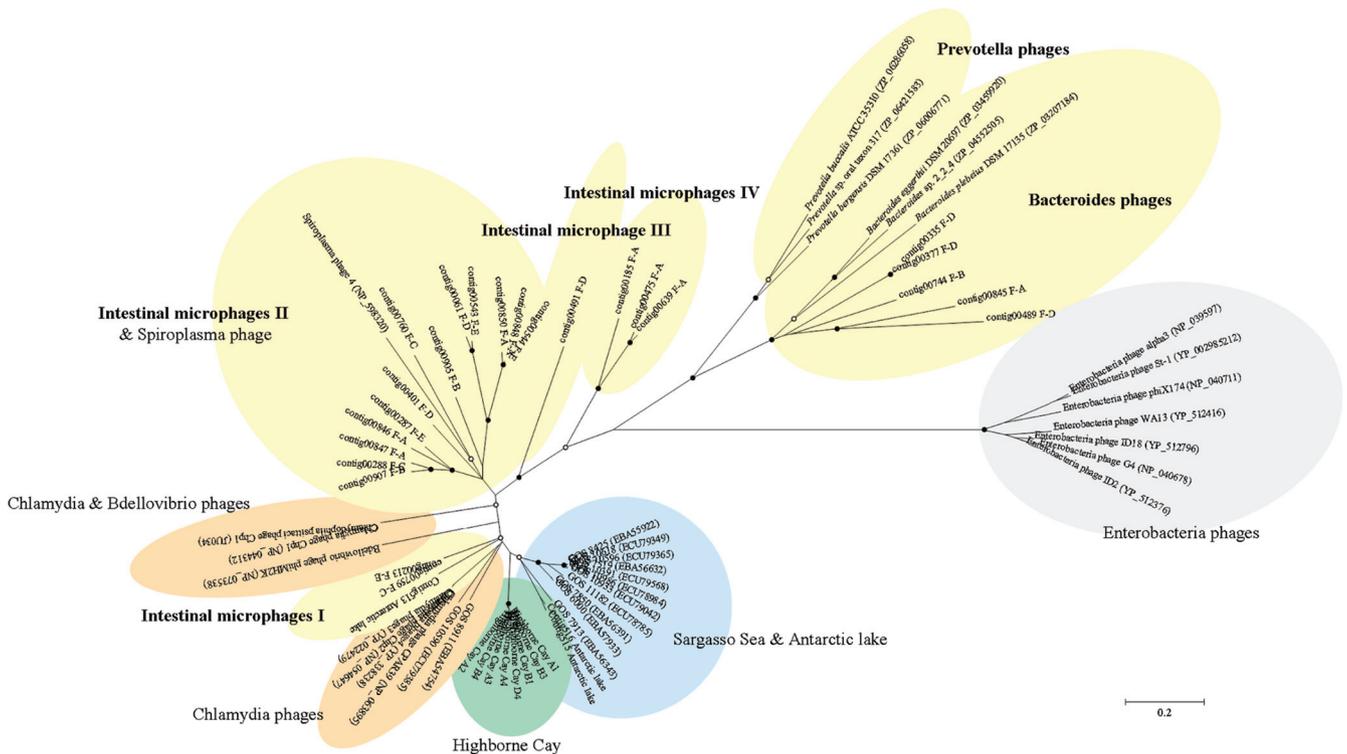


FIG. 4. Phylogenetic tree generated by the neighbor-joining algorithm of capsid protein sequences from fecal samples, cultured isolates, and environmental samples. Environmental sequences from previous studies of the Sargasso Sea (1), Highborne Cay (12), and Antarctic lake (28), 15 known sequences from cultured isolates (*Chlamydia* phages, *Bdellovibrio* phages, and *Spiroplasma* phages [in green] and *Enterobacteria* phages), and the predicted sequences from three *Bacteroides* and three *Prevotella* strains were aligned with the fecal sequences. Partial sequences of capsid protein were used for the phylogenetic analysis. Sequences were enveloped by their origins, and subgroups from fecal samples were intestinal microphages I, II, III, and IV and *Bacteroides* microphages. Filled and empty circles at internal nodes indicate bootstrap values greater than 90 and 50%, respectively. The scale bar represents 0.2 amino acid substitutions per site.

capsid protein sequences (contig845 from F-A; contig744 from F-B; and contig335, contig377, and contig489 from F-D) shared a significant similarity with six prophage-like sequences in the genomes of three *Bacteroides* and three *Prevotella* strains (34 to 39% identity). The 24 contigs were arranged with publicly known sequences from cultured isolates and environmental samples (see Fig. S2 in the supplemental material), and the phylogenetic relationship subsequently was determined (Fig. 4; also see Fig. S3 in the supplemental material). As a result, five different evolutionary groups, named as intestinal microphages I, II, III, and IV and *Bacteroides* phages, were clearly formed, and all were only distantly related to previously known microphages. The *Bacteroides* phages contained five sequences from the F-A, F-B, and F-D viromes and three sequences from three *Bacteroides* spp., whereas the *Prevotella* spp. sequences, named the *Prevotella* phages, were grouped separately from the *Bacteroides* phages.

The phylogenetic analysis showed that the microphages isolated from human feces were distantly related to previously known microphages, indicating that the diversity of ssDNA microphages in the human intestine is much higher than previously foreseen. The existence of a higher genotypic diversity of microphages in the human gut than in other environments also brings expectations of their important ecological role in the human intestine. Due to technical limitations (16), previous studies have been concentrated on dsDNA viruses, espe-

cially ssDNA and ssRNA viruses. Thus, current knowledge about the diversity and distribution of ssDNA microphages still is incomplete. Here, we showed that the diversity of ssDNA microphages may be much greater than previously recognized. Hence, groups defined by their sources of origin are clearly shown not only among gut microphages but also among all other microphages. Indeed, the geographical distribution of ssDNA microphages in the marine environment has gained much attention (55). However, this observation also may be valid for different environments, as shown in Fig. 3. Such biogeographical variability implies a certain connection between the lifestyle of microphages and the distribution of their hosts (55). ssDNA microphages and their host ranges should be further explored to enable the understanding of their ecological role in these environments.

***Bacteroides* and *Prevotella* as infecting hosts of ssDNA microphages in the human intestine.** The close relatedness between five microphage sequences from the fecal viromes and those of three *Bacteroides* and three *Prevotella* strains, based on major capsid protein and Rep sequences, suggests that microphages in the human gut are closely identified as having *Bacteroides* and *Prevotella* as their hosts. Gut bacteria and *Bacteroides* and *Prevotella* species of the phylum *Bacteroidetes* are capable of utilizing diverse types of dietary fibers, thus producing short-chain fatty acids that potentially are effective as an energy supplement and in gut inflammatory responses

(11, 30, 45). For this reason, changes in their relative abundances are closely related to the prevalence of obesity, gut inflammation, and autoimmune disease (17, 25, 57). In many previous studies, exchanges of functional genes, such as carbohydrate-active enzymes (CAZymes), were shown in dominant gut microbes (29, 51), and remarkably, a prevalence of CAZyme-encoding genes rearranged by HGT was found in a *Bacteroides* population (51). In the *Bacteroides* phage cluster, the transfer of CAZymes from the marine bacterium *Zobellia galactanivorans* to the Japanese gut bacterium *Bacteroides plebeius* via diet has been reported recently (20). Due to the possibility of gene acquisition by lateral gene transfer, symbiotic *Bacteroidetes* can be present and adapted to distinct niches (64). The relatedness results showing fecal microphages with prophage-like elements of *Bacteroides* and *Prevotella* indicate that fecal microphages might play an important role in the genotypic diversity of *Bacteroides* and *Prevotella* populations by acting as gene exchangers in the human intestine. A few genes introduced from mobile elements, such as bacteriophages, could make a strain physiologically distinct (31) and contribute to the colonization of the human intestine by *Bacteroides* and *Prevotella*. The ecology of bacteriophages in the human intestine will be further studied. In particular, the influence of lysogenic phages on host gut bacteria needs to be explored. We believe that this study greatly contributes to our knowledge of viral ecology in the human intestine.

#### ACKNOWLEDGMENTS

This work was supported by a grant (09172KFDA996) from the Korea Food & Drug Administration in 2011, as well as by the Basic Science Research Program (no. 2011-0005513) and Midcareer Researcher Program (no. 2011-0028854) through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (MEST). M.-S.K was supported by a Hi Seoul Science (humanities) fellowship funded by the Seoul Scholarship Foundation.

We are grateful to F. E. Angly for bioinformatics advice and Dong Hyun Kim and Young Mi Kim for technological assistance.

#### REFERENCES

- Angly, F. E., et al. 2006. The marine viromes of four oceanic regions. *PLoS Biol.* **4**:e368.
- Ashelford, K. E., M. J. Day, and J. C. Fry. 2003. Elevated abundance of bacteriophage infecting bacteria in soil. *Appl. Environ. Microbiol.* **69**:285–289.
- Baker, G. C., J. J. Smith, and D. A. Cowan. 2003. Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Methods* **55**:541–555.
- Breitbart, M., et al. 2008. Viral diversity and dynamics in an infant gut. *Res. Microbiol.* **159**:367–373.
- Breitbart, M., et al. 2003. Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* **185**:6220–6223.
- Breitbart, M., and F. Rohwer. 2005. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* **13**:278–284.
- Canchaya, C., G. Fournous, S. Chibani-Chennoufi, M. L. Dillmann, and H. Brussow. 2003. Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* **6**:417–424.
- Cole, J. R., et al. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37**:D141–D145.
- Danovaro, R., E. Manini, and A. Dell'Anno. 2002. Higher abundance of bacteria than of viruses in deep Mediterranean sediments. *Appl. Environ. Microbiol.* **68**:1468–1472.
- Danovaro, R., and M. Serresi. 2000. Viral density and virus-to-bacterium ratio in deep-sea sediments of the eastern Mediterranean. *Appl. Environ. Microbiol.* **66**:1857–1861.
- De Filippo, C., et al. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci. U. S. A.* **107**:14691–14696.
- Desnues, C., et al. 2008. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* **452**:340–343.
- Dethlefsen, L., S. Huse, M. L. Sogin, and D. A. Relman. 2008. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* **6**:e280.
- Dinsdale, E. A., et al. 2008. Functional metagenomic profiling of nine biomes. *Nature* **452**:629–632.
- Drake, L. A., K. H. Choi, A. G. Edward Haskell, and F. C. Dobbs. 1998. Vertical profiles of virus-like particles and bacteria in the water column and sediments of Chesapeake Bay, U. S. A. *Aquatic Microb. Ecol.* **16**:17–25.
- Edwards, R. A., and F. Rohwer. 2005. Viral metagenomics. *Nat. Rev. Microbiol.* **3**:504–510.
- Frank, D. N., et al. 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U. S. A.* **104**:13780–13785.
- Gomez-Alvarez, V., T. K. Teal, and T. M. Schmidt. 2009. Systematic artifacts in metagenomes from complex microbial communities. *ISME J.* **3**:1314–1317.
- Hall, T. A. 1999. BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**:95–98.
- Hehemann, J. H., et al. 2010. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**:908–912.
- Helton, R. R., L. Liu, and K. E. Wommack. 2006. Assessment of factors influencing direct enumeration of viruses within estuarine sediments. *Appl. Environ. Microbiol.* **72**:4767–4774.
- Huson, D. H., A. F. Auch, J. Qi, and S. C. Schuster. 2007. MEGAN analysis of metagenomic data. *Genome Res.* **17**:377–386.
- Kim, K. H., et al. 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl. Environ. Microbiol.* **74**:5975–5985.
- Kristensen, D. M., A. R. Mushegian, V. V. Dolja, and E. V. Koonin. 2010. New dimensions of the virus world discovered through metagenomics. *Trends Microbiol.* **18**:11–19.
- Larsen, N., et al. 2010. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One* **5**:e9085.
- Lepage, P., et al. 2008. Dysbiosis in inflammatory bowel disease: a role for bacteriophages? *Gut* **57**:424–425.
- Ley, R. E., D. A. Peterson, and J. I. Gordon. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**:837–848.
- López-Bueno, A., et al. 2009. High diversity of the viral community from an Antarctic lake. *Science* **326**:858–861.
- Lozupone, C. A., et al. 2008. The convergence of carbohydrate active gene repertoires in human gut microbes. *Proc. Natl. Acad. Sci. U. S. A.* **105**:15076–15081.
- Maslowski, K. M., et al. 2009. Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**:1282–1286.
- Morowitz, M. J., et al. 2011. Strain-resolved community genomic analysis of gut microbial colonization in a premature infant. *Proc. Natl. Acad. Sci. U. S. A.* **108**:1128–1133.
- Nelson, K. E., et al. 2010. A catalog of reference genomes from the human microbiome. *Science* **328**:994–999.
- Niu, B., L. Fu, S. Sun, and W. Li. 2010. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics* **11**:187.
- Park, E. J., et al. 2010. Metagenomic analysis of the viral community in fermented Foods. *Appl. Environ. Microbiol.* **77**:1284–1291.
- Patel, A., et al. 2007. Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nat. Protoc.* **2**:269–276.
- Polz, M. F., and C. M. Cavanaugh. 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **64**:3724–3730.
- Qin, J., et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**:59–65.
- Reyes, A., et al. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**:334–338.
- Roberts, A. P., et al. 2008. Revised nomenclature for transposable genetic elements. *Plasmid* **60**:167–173.
- Rohwer, F., D. Prangishvili, and D. Lindell. 2009. Roles of viruses in the environment. *Environ. Microbiol.* **11**:2771–2774.
- Roossinck, M. J. 2011. The good viruses: viral mutualistic symbioses. *Nat. Rev. Microbiol.* **9**:99–108.
- Rosario, K., C. Nilsson, Y. W. Lim, Y. Ruan, and M. Breitbart. 2009. Metagenomic analysis of viruses in reclaimed water. *Environ. Microbiol.* **11**:2806–2820.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- Schloss, P. D., et al. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**:7537–7541.
- Sokol, H., et al. 2008. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U. S. A.* **105**:16731–16736.
- Suttle, C. A. 2007. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**:801–812.

47. **Suttle, C. A.** 1994. The significance of viruses to mortality in aquatic microbial communities. *Microb. Ecol.* **28**:237–243.
48. **Tamames, J., and A. Moya.** 2008. Estimating the extent of horizontal gene transfer in metagenomic sequences. *BMC Genomics* **9**:136.
49. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
50. **Tap, J., et al.** 2009. Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* **11**:2574–2584.
51. **Tasse, L., et al.** 2010. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* **20**:1605–1612.
52. **Thompson, J. R., L. A. Marcelino, and M. F. Polz.** 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR.' *Nucleic Acids Res.* **30**:2083–2088.
53. **Thurber, R. V.** 2009. Current insights into phage biodiversity and biogeography. *Curr. Opin. Microbiol.* **12**:582–587.
54. **Thurber, R. V., M. Haynes, M. Breitbart, L. Wegley, and F. Rohwer.** 2009. Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* **4**:470–483.
55. **Tucker, K. P., R. Parsons, E. M. Symonds, and M. Breitbart.** 2010. Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *ISME J.* **5**:822–830.
56. **Turnbaugh, P. J., et al.** 2007. The human microbiome project. *Nature* **449**:804–810.
57. **Turnbaugh, P. J., et al.** 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**:1027–1031.
58. **Wang, X., et al.** 2010. Cryptic prophages help bacteria cope with adverse environments. *Nat. Commun.* **1**:147.
59. **Weinbauer, M. G.** 2004. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**:127–181.
60. **Wen, L., et al.** 2008. Innate immunity and intestinal microbiota in the development of type 1 diabetes. *Nature* **455**:1109–1113.
61. **Wilhelm, S. W., and C. A. Suttle.** 1999. Viruses and nutrient cycles in the sea. *Bioscience* **49**:781–788.
62. **Willner, D., et al.** 2009. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One* **4**:e7370.
63. **Wommack, K. E., and R. R. Colwell.** 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**:69–114.
64. **Xu, J., et al.** 2007. Evolution of symbiotic bacteria in the distal human intestine. *Plos Biol.* **5**:e156.
65. **Yooseph, S., et al.** 2007. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *Plos Biol.* **5**:e16.
66. **Yu, Z., and M. Morrison.** 2004. Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36**:808–812.
67. **Zhang, T., et al.** 2006. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* **4**:0108–0118.