



Article

Ecological Structuring of Temperate Bacteriophages in the Inflammatory Bowel Disease-Affected Gut

Hiroki Nishiyama ¹, Hisashi Endo ¹, Romain Blanc-Mathieu ² and Hiroyuki Ogata ^{1,*}

¹ Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji 611-0011, Japan; hiroki@kuicr.kyoto-u.ac.jp (H.N.); endo@scl.kyoto-u.ac.jp (H.E.)

² Laboratoire de Physiologie Cellulaire & Végétale, CEA, CNRS, INRA, IRIG, Université Grenoble Alpes, 38000 Grenoble, France; romain.blancmathieu@cea.fr

* Correspondence: ogata@kuicr.kyoto-u.ac.jp; Tel.: +81-774-38-3270

Received: 30 September 2020; Accepted: 23 October 2020; Published: 27 October 2020



Abstract: The aim of this study was to elucidate the ecological structure of the human gut temperate bacteriophage community and its role in inflammatory bowel disease (IBD). Temperate bacteriophages make up a large proportion of the human gut microbiota and are likely to play a role in IBD pathogenesis. However, many of these bacteriophages await characterization in reference databases. Therefore, we conducted a large-scale reconstruction of temperate bacteriophage and bacterial genomes from the whole-metagenome sequence data generated by the IBD Multi'omics Database project. By associating phages with their hosts via genome comparisons, we found that temperate bacteriophages infect a phylogenetically wide range of bacteria. The majority of variance in bacteriophage community composition was explained by variation among individuals, but differences in the abundance of temperate bacteriophages were identified between IBD and non-IBD patients. Of note, in active ulcerative colitis patients, temperate bacteriophages infecting *Bacteroides uniformis* and *Bacteroides thetaiotaomicron*—two species experimentally proven to be beneficial to gut homeostasis—were over-represented, whereas their hosts were under-represented in comparison with non-IBD patients. Supporting the mounting evidence that gut viral community plays a vital role in IBD, our results show potential association between temperate bacteriophages and IBD pathogenesis.

Keywords: temperate bacteriophage; inflammatory bowel disease; ulcerative colitis; Crohn's disease; gut microbiota

1. Introduction

Inflammatory bowel disease (IBD) encompasses a group of intestinal disorders, the most common of which are Crohn's disease (CD) and ulcerative colitis (UC). These disorders are characterized by chronic inflammation that occurs throughout multiple layers of the gastrointestinal tract in CD and in the inner layer of the colon in UC. The exact cause of IBD is not clear; however, the gut microbiota, together with host genetics and environmental factors (e.g., diet, antibiotics, and drugs), plays an important role in the pathogenesis of IBD [1,2]. Although bacteria have been the focus of most IBD-related gut microbiota studies, recent studies have shown that bacteriophages in the gut may also play an important role in the pathogenesis of IBD [3–5].

The human gut bacteriophage community is thought to be largely composed of temperate bacteriophages [6], which are characterized by their ability to replicate through two different life cycles: the lysogenic cycle and the lytic cycle [7]. In the lysogenic cycle, temperate bacteriophages are integrated into their host genome in the form of prophages. During this process, viral genes, including those associated with anti-microbial activity, virulence, and toxin production, are horizontally transferred to the host. Upon induction, temperate bacteriophages enter the lytic cycle, in which

virions are produced from prophages and released into the outer environment following lysis of the host bacterium.

Recent studies on gut bacteriophage communities in IBD-affected individuals suggest that *Caudovirales* are associated with the pathogenesis of IBD [3,4]. However, insights based on viral taxonomy require careful consideration because the majority of viruses in the human gut are yet to be taxonomically classified [8]. Furthermore, viral taxonomy is under constant review by the International Committee on the Taxonomy of Viruses (ICTV) to accommodate newly sequenced viral genomes [9]. For example, the definition of the *Caudovirales* order was recently changed to include two new families: *Ackermannviridae* and *Herelleviridae* [6]. To address these problems, we aimed to characterize the ecological structure of functional temperate bacteriophage communities in the guts of IBD patients by identifying associations between these viruses and their bacterial hosts via reference-independent methods using a large publicly available microbiome dataset.

2. Materials and Methods

2.1. IBDMDB Data Depository

This study was conducted using publicly available whole-metagenome shotgun sequencing (WMGS) data generated by the IBD Multi-omics Database (IBDMDB) research team as part of the Human Microbiome Project 2 [1,10]. Specifically, 1338 sets of quality-filtered WMGS data were used. Sequences (i.e., reads) were generated via DNA sequencing using the Illumina HiSeq (2 × 101 bp) system (Illumina, San Diego, CA, USA) from longitudinal stool samples from 65 CD patients, 38 UC patients, and 27 non-IBD patients. Non-IBD patients had gastrointestinal symptoms but were not diagnosed with IBD [1]. Quality control of WMGS data was performed by the IBDMDB research team [1]. Briefly, the AnADAMA2 pipeline (<http://huttenhower.sph.harvard.edu/anadama2>) was used to remove low quality regions from reads, and any reads matching the human genome were discarded. All sequence data, including the WMGS metadata, were downloaded from the IBDMDB data depository (<http://ibdmdb.org>).

2.2. Assessment of Disease Activity Among CD and UC Patients

CD patients with a Harvey–Bradshaw index (HBI) score ≥ 5 were identified as having active CD [11]. CD patients with an HBI score < 5 were classified as having inactive CD. Among UC patients, those with a simple clinical colitis activity index (SCCAI) score ≥ 5 were identified as having active UC [12], whereas those with a SCCAI score < 5 were identified as having inactive UC.

2.3. De Novo Contig Assembly

Quality-filtered WMGS data from the 1327 stool samples with sufficient sequencing depth (i.e., minimum of 10,000 reads) were selected for *de novo* assembly. Reads from each stool sample were assembled into contiguous sequences (i.e., contigs) using MEGAHIT (version 1.1.4) ($-kmin\ 21, -kmax\ 251, -k-step\ 10$) [13]. To assemble the genomes from the microbes present at low abundance, we also co-assembled the reads from the stool samples taken from the same individual using the same method.

2.4. Generation of Viral Operational Taxonomic Units (OTUs) for Temperate Bacteriophages

Viral regions within contigs were predicted using VirSorter (version 1.0.5) ($-db\ 2, -diamond$) [14]. To reduce false identification, predicted viral regions with at least one hallmark viral gene and/or enrichment of either viral-like genes or non-*Caudovirales* viral genes (i.e., VirSorter categories 1, 2, 4, and 5) were selected for further analyses. The selected viral regions were grouped into viral OTUs at 95% sequence similarity using CD-HIT-EST (version 4.6.8) ($-c\ 0.95, -aS\ 0.85, -n\ 10, -g\ 1$) [15]. To identify viral OTUs representing temperate bacteriophages, MetaGeneMark (version 4.30) was first used to derive translated peptide sequences from the viral regions [16]. Subsequently, the peptide sequences were searched against the Pfam-A database (version 33.0) using hmmscan (version 3.3) ($-cut_ga$) [17,18]. Viral OTUs encoding one or more of the following viral integrase protein domains were defined

as temperate bacteriophage and were selected for the subsequent analyses: PF00589, PF02899, PF09003, PF14659, PF13356, PF13495, PF12834, PF12835, PF13009, PF14657, PF02914, PF02316, PF02920, PF09299. Taxonomic classification of the viral OTUs was performed by classifying representative sequences (the longest) from each OTU using vConTACT v2.0 (version 0.9.12) (`-rel-mode Diamond, -db ProkaryoticViralRefSeq94-Merged`) [19].

2.5. Reconstruction of Bacterial Genomes via Metagenomic Binning

Contigs were binned into metagenome-assembled genomes (MAGs) using METAWRAP (version 1.1.8) (`binning, -metabat2, -maxbin2, -concoct`) [20]. Briefly, reads from the same individual were mapped on the contigs using Burrows–Wheeler aligner [21]. Contigs ≥ 1000 bp in length (1500 bp for MetaBAT2) were then binned into MAGs based on their read coverage using three binning tools: MaxBin2 (version 2.2.5), MetaBAT2 (version 2.12.1), and CONCOCT (version 1.0.0) [22–24]. Subsequently, METAWRAP `binning_refiner` was used to hybridize MAGs generated via different binning tools and the best set of MAGs was chosen based on completion and contamination statistics calculated by CheckM (version 1.0.13) (`lineage_wf`) [20,25]. Of these, MAGs with a minimum completion of 70% and maximum contamination of 10% were dereplicated (i.e., grouped) into MAG clusters with a minimum genome-wide average nucleotide identity (gANI) threshold of 99% and a minimum aligned fraction threshold of 60% using dRep (version 2.2.2) [26]. Representative MAGs were chosen for each cluster based on completeness, contamination, strain heterogeneity, N50, and size [26]. NCBI taxonomy was determined for each representative MAG by placing them into a protein reference tree using GTDB-Tk (version 1.0.2) (`classify_wf`) [27]. Six MAG clusters assigned to archaea were removed. The resulting collection of 3133 bacterial MAG clusters were selected for subsequent analyses.

2.6. Construction of a Bacterial Phylogenetic Tree

A bacterial phylogenetic tree was constructed by first extracting translated peptide sequences from the representative draft genome of each MAG cluster using MetaGeneMark (version 4.30) [16]. The translated peptide sequences were then used to generate a phylogenetic tree using PhyloPhlAn (version 0.99) (`-user_tree`) [28]. Briefly, input peptides sequences matching to PhyloPhlAn's collection of 400 marker proteins (i.e., proteins broadly conserved in bacteria and archaea) were identified using USEARCH (version 5.2.236) [29]. Subsequently, these proteins are aligned with MUSCLE (version 3.8.31) [30] and reconstructed into a phylogenetic tree with FastTree (version 2.1.11) [31]. The visual representation of the phylogenetic tree was generated using iTOL (version 5.6) [32].

2.7. Assignment of Hosts to Viral OTUs

Two strategies based on previous studies were used for host assignment of viral OTUs: binning-and basic local alignment search tool (BLAST)-based methods [33–35]. In the binning-based method, a MAG cluster was assigned as the host of a viral OTU if the contigs from the MAG included a region corresponding to the viral OTU. In the BLAST-based method, bacterial sequences adjacent to prophage sequences (i.e., VirSorter categories 4 and 5) were aligned against MAGs using BLASTn in the BLAST+ package (version 2.8.1) (`-perc_identity 95, -evalue 1e-50, -word_size 11`) [36]. If both the upstream and downstream regions of a prophage sequence in a viral OTU aligned with contigs from a MAG cluster with sequence similarity $>95\%$ and query coverage $>80\%$, the MAG cluster was considered to be the host. Long queries were trimmed to 1000 bp starting from the edge of the selected prophage region. Prophage sequences that were not flanked on both sides by sequences ≥ 250 bp were excluded from the host assignment process.

2.8. Mapping Reads to Viral OTUs and MAGs

Quality-filtered reads from 1338 gut metagenomes were mapped to the representative sequences of each viral OTU and MAG cluster using bowtie2 (version 2.3.4.3) (`-maxins 1000`) [37]. Resulting bam files were processed using CoverM (version 0.3.0, <https://github.com/wwood/CoverM/>) (`-proper-pairs-only`) so that only reads that aligned across $\geq 80\%$ of their length with a minimum of 95% sequence identity

were retained. The bam files were then further processed using samtools (version 1.9) (-f 3, -F 2304) [38], and only the primary alignments from the properly paired reads were retained. Counts of mapped read pairs were retrieved using idxstat in samtools (version 1.9) and placed into count tables for both viral OTUs and MAG clusters [38].

2.9. Calculation of Phage-to-Host Ratios

For each viral OTU, the phage-to-host ratio within each sample was calculated by dividing viral fragments per kilobase of transcript per million mapped reads (FPKM) by the sum of the FPKMs of the viral hosts. Phage-to-host ratios were not calculated in cases where none of the hosts received at least 100 mapped reads to avoid falsely-high phage-to-host ratios.

2.10. Analysis of Ecological Diversity

For all subsequent analyses, a set of 1290 samples with more than 1,000,000 reads mapped to MAG contigs and viral OTUs was used. Alpha diversity within each stool sample was assessed using Shannon's diversity index analysis of genome-length-normalized count tables after rarefying at the minimum sample depth using phyloseq R package (version 1.30.0) [39]. This process was repeated 10 times and the average value for each sample was reported. The following linear mixed-effects model was used to assess the statistical significance of the effect of diagnosis factor on the alpha diversity of gut microbial communities using the nlme R package (version 3.1-147) [40]:

$$\text{Shannon's diversity index} \sim \text{disease status} + \text{sex} + \text{consent age} + \text{antibiotics} + \text{chemotherapy} + \text{immunosuppressants} + \text{site name} + (1|\text{participant ID}).$$

Significance (p value) was determined using the Wald test. Richness was also quantified using the approach described above.

Beta diversity among gut microbial communities was assessed using Bray–Curtis dissimilarity analysis of genome-length-normalized count tables after rarefying using phyloseq R package (version 1.30.0) [39]. The rarefying depth was set at the depth of the smallest sample. The effect of individuality on variance in the gut microbial communities was quantified using PERMANOVA implemented in phyloseq R package's Adonis function [39,41]. The same method was used to quantify the effect of diagnosis on gut microbiota variance; however, rarefied counts were first averaged for each individual prior to genome-length-normalization. Dissimilarity matrices were visualized in two dimensions using uniform manifold approximation and projection for dimension reduction (UMAP) (version 0.4.2) ($n_neighbors = 30$, $min_dist = 0.3$, $n_components = 2$, $metric = braycurtis$) [42]. Parallel computations were conducted using GNU parallel (version 20180322) [43].

2.11. Analysis of Differential Abundance

Differences in the abundance of gut microorganisms between different diagnosis groups were assessed using the fitZig function in metagenomeSeq R package (version 1.30.0) [44]. First, viral OTUs or MAG clusters that were not present in more than 25% of samples in one of the diagnosis groups were removed. Subsequently, read counts were normalized using the cumulative sum scaling method in metagenomeSeq [44]. A zero-inflated Gaussian mixture model fitted to the normalized read counts was then used to test for differential abundance between the diagnosis groups at a false discovery rate (FDR) cutoff of 0.1 and a \log_2 fold-change cutoff of 1. The FDR was calculated using the Benjamini–Hochberg method. The model used patient ID as a blocking factor along with the following covariates: disease status, sex, consent age (i.e., age of the patient at consent), antibiotics, immunosuppressants, site name, and normalization factor. As advised by Paulson et al. (2013), the normalization factor was set using the median scaling factor [44]. Maximum iteration was set to 100 (default 10). In addition, significant features were restricted to those with an effective sample size greater than or equal to the average as recommended by the authors [44]. To further minimize the possibility of false positives, only those

significant features for which the median abundance in the higher diagnosis group was greater than that in the lower diagnosis group were retained. Ten samples belonging to a CD patient (participant ID C3031) were excluded from this analysis because consent age, metadata required in the current zero-inflated Gaussian mixture model, were unavailable.

2.12. Data Availability

Count tables, metadata, and the sequences of MAGs and viral regions can be obtained from the following site: <ftp://ftp.genome.jp/pub/db/community/ibd-phage/>.

3. Results

3.1. Taxonomic Classification of Temperate Bacteriophages in the Human Gut

A total of 17,536,516 genome fragments were assembled from the 1327 sets of WMGS data generated by the IBDMDB team from stool samples from IBD patients and other patients diagnosed with non-IBD gastrointestinal disorders. To characterize the temperate bacteriophage community in the gut, 50,624 regions predicted to be viral based on the presence of hallmark viral genes were clustered into 17,331 species-rank virus groups based on sequence similarity. Of these viral OTUs, 3843 were associated with putative host sequences and found to encode integrase genes. A total of 1797 viral OTUs with a phage-to-host ratio >10 in at least one sample were conservatively classified as functional (i.e., non-defunct) temperate bacteriophages. The representative viral regions of these viral OTUs ranged in size from 4331 to 313,294 bp, with a median size of 33,609 bp. Only 28 (1.56%) of the temperate bacteriophage OTUs were taxonomically annotated. These belonged to one of the following three *Caudovirales* families: *Myoviridae*, *Podoviridae*, and *Siphoviridae* (Figure 1A).

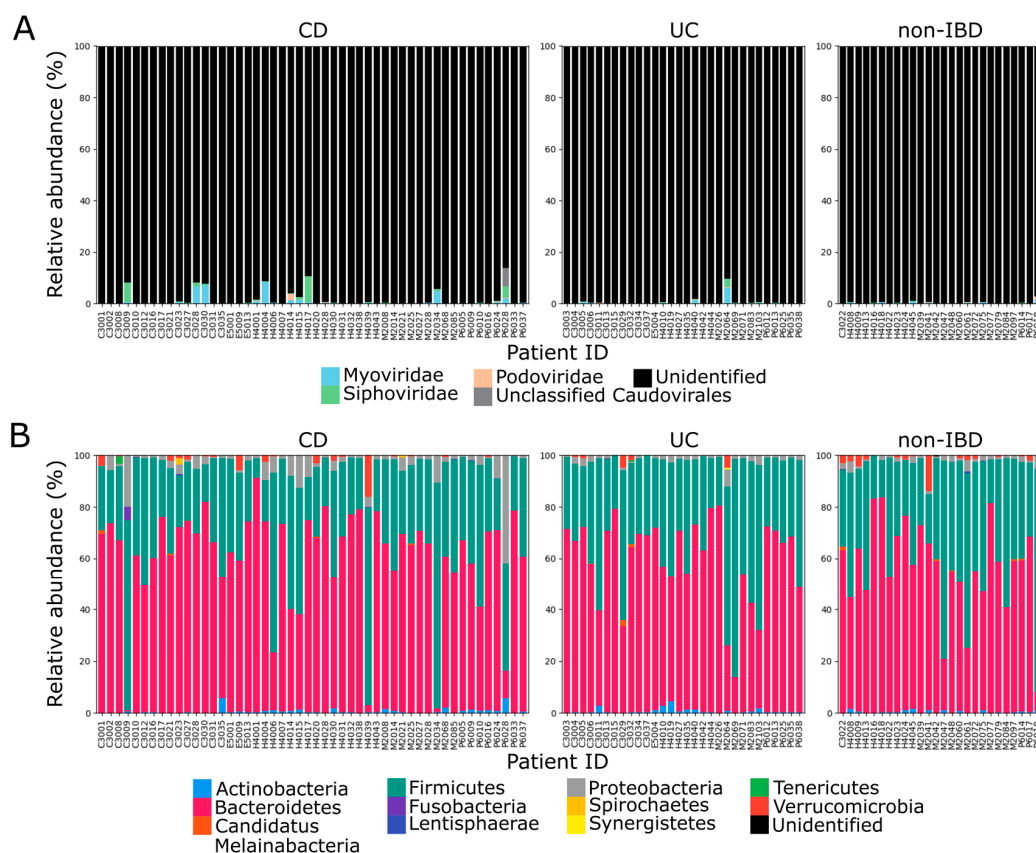


Figure 1. Relative abundance of gut microbes and viruses in inflammatory bowel disease (IBD) and non-IBD patients. Mean relative abundance within each patient is plotted at the family level for temperate bacteriophages (A) and at the phylum level for bacteria (B). Relative abundance was calculated based on genome length normalized read counts.

3.2. Infection of a Phylogenetically Wide Range of Human Gut Bacteria by Temperate Bacteriophages

To determine the host distribution of temperate bacteriophages in the human gut, bacterial genomes were reconstructed from metagenome data. By binning contigs based on read coverage and k-mer occurrence, 15,906 MAGs with completion >70% and contamination <10% were reconstructed. These MAGs were then dereplicated into 3133 bacterial MAG clusters based on gANI values. The representative MAGs for each of the MAG clusters ranged in size from 685,598 to 6,988,682 bp, with a median size of 2,635,319 bp. Most (91.3%) of the MAGs were identified as belonging to the phyla *Bacteroidetes* and *Firmicutes*, which are the two most dominant phyla in the human gut (Figure 1B) [45].

A total of 1153 MAG clusters (36.8%) were found to be infected with temperate bacteriophages based on sequence comparisons (Figure 2). Taxonomic classification of the MAG clusters showed that temperate bacteriophages infect a wide range of bacteria distributed across the *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, and *Verrucomicrobia* phyla (corresponding to 12 classes, 13 orders, 28 families, 63 genera, and 101 species) (Figure 3A, Table S1). To understand the host range of temperate bacteriophages, the last common ancestors of host taxonomies were also calculated (Figure 3B). Based on these results, 94.8% of viral OTUs were found to have a narrow host range at the species and genus levels.

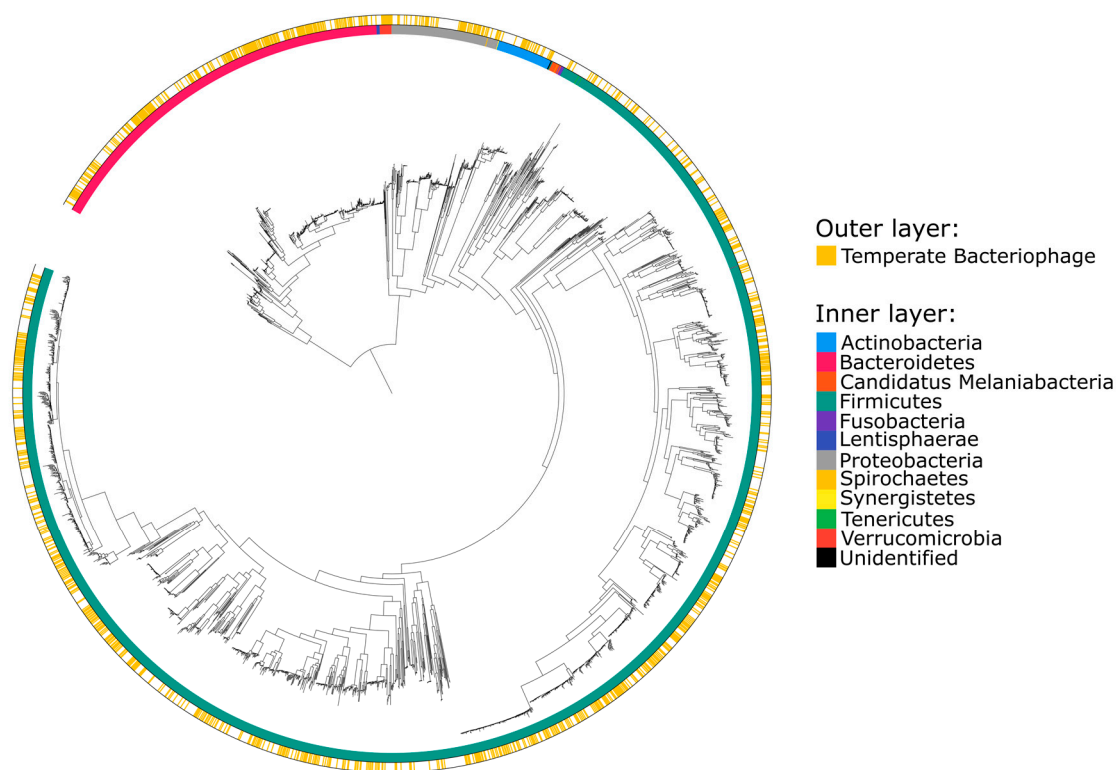


Figure 2. Phylogenetic tree based on a comprehensive collection of reconstructed bacterial genomes. The inner layer colors correspond to the bacterial phyla represented at the tip of each branch. The outer layer indicates bacteria associated with functional temperate bacteriophages.

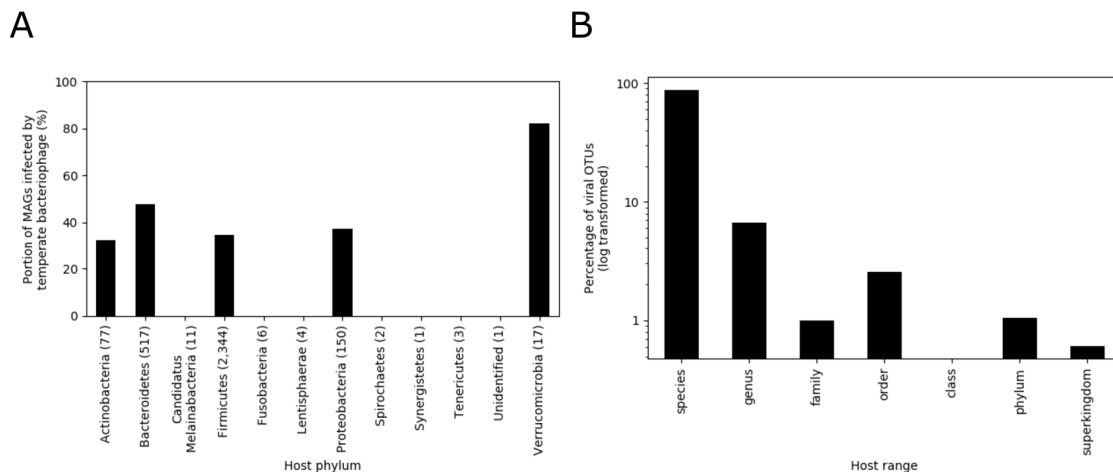


Figure 3. Characteristics of phage-host associations. **(A)** Distribution of temperate bacteriophage host phyla. The numbers following the phylum names indicate the number of metagenome-assembled genomes (MAG) clusters obtained for each phylum. **(B)** Host range of temperate bacteriophages.

3.3. Alpha Diversity of Gut Microbes in IBD and Non-IBD Patients

Dysbiosis is a state of microbiota imbalance characterized by factors such as decreased diversity and altered composition [46]. The gut microbiota from IBD patients has been shown to be dysbiotic [46]. To investigate differences in alpha diversity between IBD patients and patients with non-IBD gastrointestinal disorders, Shannon's diversity and richness of bacterial and temperate bacteriophage communities were investigated (Figures S1B,C and S2B,C). The same measures were also used to assess the alpha diversity of induced temperate bacteriophage (i.e., temperate bacteriophages in the lytic cycle) communities in each sample. Temperate bacteriophages were conservatively considered to be induced (i.e., active) only when the phage-to-host ratio was >10 , based on the threshold utilized in a previous study [35] (Figures S1A and S2A). Consistent with the results from the IBDMDB research group [1], most of the comparisons displayed lower alpha diversity in the IBD patients than in the non-IBD patients, but the comparisons lacked statistical significance ($p > 0.05$ with Holm's correction).

3.4. Variation of Gut Microbial Community Composition among Stool Samples

To assess variation in microbial community composition among samples, Bray–Curtis dissimilarity between all samples was assessed and results were plotted in two dimensions using UMAP (Figure 4). Consistent with previous reports, the temperate bacteriophage and bacterial community compositions were highly specific to each individual (effect size = 71.2% and 69.7% for temperate bacteriophage and bacterial communities, respectively; $p < 0.0001$, PERMANOVA test) [1,47]. The effect of individuality on the composition of induced temperate bacteriophage communities could not be estimated because of a lack of induced temperate bacteriophages in many samples. The effect of disease status on microbial community composition was also examined but differences were not significant ($p > 0.05$, PERMANOVA test).

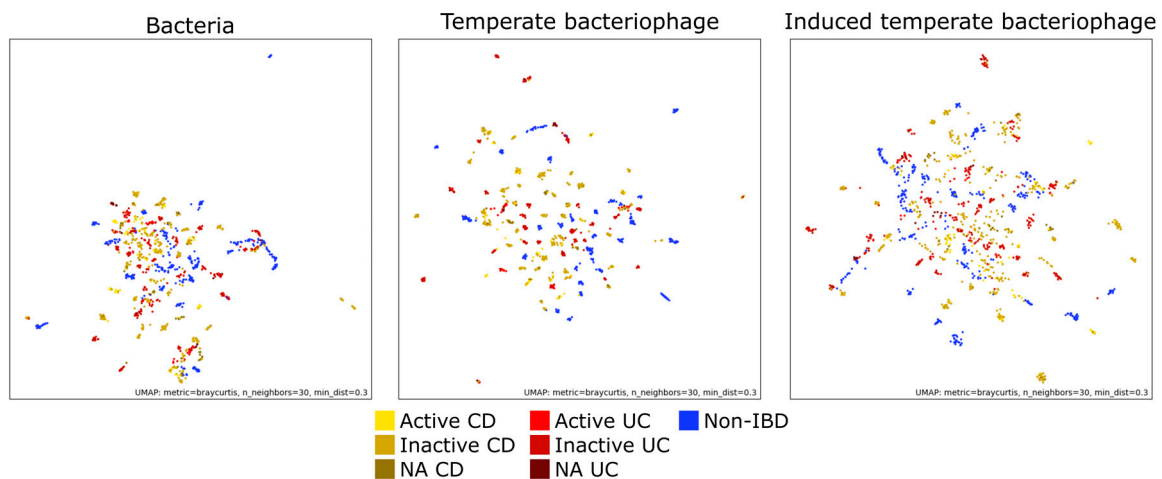


Figure 4. Ordination of gut microbial communities. Dissimilarity among gut microbial communities was assessed using the Bray–Curtis dissimilarity measure and ordinated in two dimensions using uniform manifold approximation and projection for dimension reduction (UMAP). The following samples were used: 87 active Crohn’s disease (CD) samples from 24 patients, 448 inactive CD samples from 48 patients, 45 NA CD samples from 17 patients, 29 active ulcerative colitis (UC) samples from 9 patients, 298 inactive UC samples from 29 patients, 21 NA UC samples from 7 patients, and 362 non-IBD samples from 26 patients. NA CD and NA UC samples refers to those lacking Harvey–Bradshaw index (HBI) and simple clinical colitis activity index (SCCAI) scores, respectively.

3.5. Differential Abundance of Temperate Bacteriophages and Their Hosts

To gain a better understanding of the role played by the microbial community in IBD at a finer scale, differences in the abundances of the temperate bacteriophages and bacteria between the IBD and non-IBD patients were assessed at the level of viral OTUs and MAG clusters, respectively (Figures 5 and 6). In comparisons between CD and non-IBD patients, 182 and 105 viral OTUs were found to be differentially abundant between active CD and non-IBD patients and between inactive CD and non-IBD patients, respectively (FDR < 0.1) (Tables S2 and S4). Additionally, 54 and 102 MAG clusters were shown to be differentially abundant between active CD and non-IBD patients and between inactive CD and non-IBD patients, respectively (FDR < 0.1) (Tables S3 and S5). Over-represented MAG clusters in CD patient samples predominantly consisted of *Dialister invisus* (active CD: 24 MAG clusters, inactive CD: 24 MAG clusters), whereas under-represented MAG clusters in these patients were largely attributed to *Faecalibacterium prausnitzii* (active CD: 7 MAG clusters, inactive CD: 20 MAG clusters). Of note, similar differences in abundance (i.e., the same trend of over-/under-representation) were not observed for bacterial hosts of the differentially-abundant viral OTUs (except in two cases) in any of the comparisons (active CD vs non-IBD and inactive CD vs non-IBD).

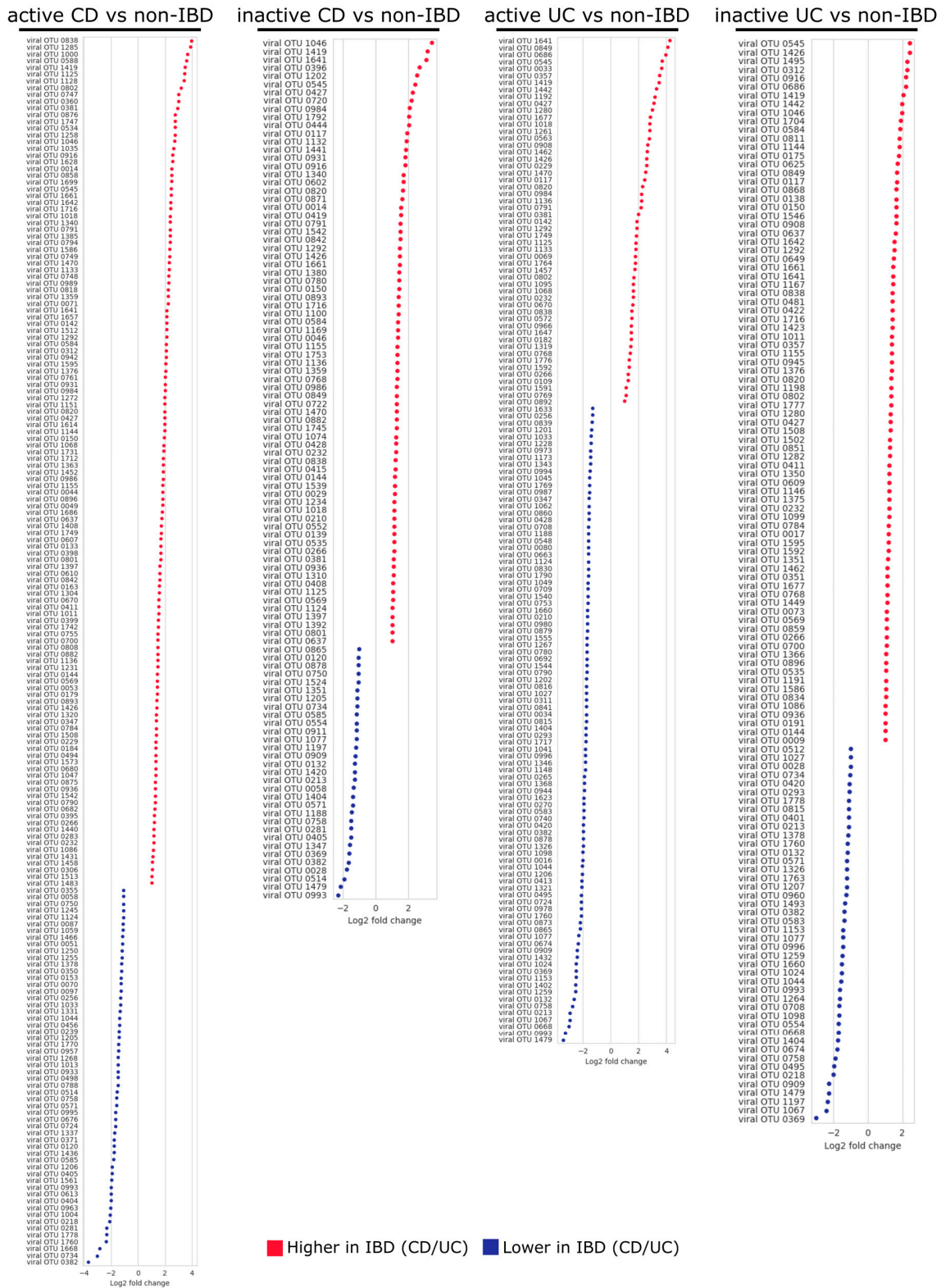


Figure 5. Log₂ fold-change values for temperate bacteriophages displaying statistically significant differences in abundance between IBD and non-IBD patients (FDR < 0.1). Positive fold-change values indicate a higher abundance in IBD patients, whereas a negative fold-change indicates higher abundance in non-IBD patients. The following samples were used: 87 active CD samples from 24 patients, 448 inactive CD samples from 48 patients, 29 active UC samples from 9 patients, 298 inactive UC samples from 29 patients, and 362 non-IBD samples from 26 patients.



Figure 6. Log₂ fold-change values for bacteria displaying statistically significant differences in abundance between IBD and non-IBD patients (FDR < 0.1). Positive fold-change indicates a higher abundance in IBD patients, whereas a negative fold-change indicates higher abundance in non-IBD patients. MAG cluster is abbreviated to “M.” The following samples were used: 87 active CD samples from 24 patients, 448 inactive CD samples from 48 patients, 29 active UC samples from 9 patients, 298 inactive UC samples from 29 patients, and 362 non-IBD samples from 26 patients.

Comparisons between UC and non-IBD patients yielded 145 and 126 viral OTUs that were differentially abundant between active UC and non-IBD patients and between inactive UC and non-IBD patients, respectively (FDR < 0.1) (Tables S6 and S8). In addition, 160 and 89 MAG clusters were differentially abundant between active UC and non-IBD patients and between inactive UC and non-IBD patients, respectively (FDR < 0.1) (Tables S7 and S9). In the active UC vs. non-IBD comparison, notable differences included an increased abundance of *Clostridium* spp. (21 MAG clusters) and decreased abundance of *Alistipes* spp. (28 MAG clusters) and *Bacteroides* spp. (58 MAG clusters) in the

active UC patients compared with the non-IBD patients. The inactive UC vs. non-IBD comparison revealed an increased abundance of *Bacteroides* spp. (14 MAG clusters) and *Clostridium* spp. (25 MAG clusters) in the inactive UC patients compared with the non-IBD patients. As observed in the comparisons between CD and non-IBD patients, differences in the abundance of host MAG clusters were not identified for the majority of the differentially-abundant viral OTUs. With the exception of 49 differentially-abundant viral OTUs identified in the active UC vs. non-IBD comparison and 15 differentially-abundant viral OTUs identified in the inactive UC vs. non-IBD comparison, most of the differences in viral OTU abundance were not reflected in host abundance. Of note, viral OTUs 0791, 0838, 0802, and 1592 were over-represented in active UC patients, whereas their hosts, MAG clusters 0358, 0655, 0655, and 2980, respectively, were under-represented in the same patients. Taxonomical classification analysis identified MAG clusters 0358 and 0655 as *Bacteroides uniformis* and MAG cluster 2980 as *Bacteroides thetaiotaomicron*.

4. Discussion

In the current study, we characterized the ecological structure of the temperate bacteriophage community in the guts of IBD-affected individuals by identification of bacterial hosts. Previous studies have reported that the human gut viral community largely consists of unclassified members [3,8,48]. In line with these observations, we were only able to taxonomically assign 28 (1.56%) of the viral OTUs corresponding to the temperate bacteriophages that we identified in the current study. To address this problem, we characterized the temperate bacteriophage community by host assignment based on sequence comparisons between reconstructed microbial and viral genomes. Our findings revealed that a wide range of bacteria across the *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, and *Verrucomicrobia* phyla are infected with functional temperate bacteriophage in the human gut (Figures 2 and 3A). In addition, the majority (94.8%) of the temperate bacteriophage OTUs were found to possess a narrow host range at the level of species or genus. This finding is similar to the results from a previous study showing that 99% of the phage OTUs from various environments possess a narrow host range at the species or genus level [34].

Analysis of beta diversity in gut microbial communities revealed that community composition was highly specific to each individual, and was not associated with disease status (Figure 4). This result is consistent with a previous report by the IBDMDB research team that suggested that individuality is the most important factor in explaining variance in gut microbial community structure [1]. In addition, another study on gut bacteriophage communities found that community composition was specific to each individual [47]. In the current study, we also observed the same trend for the temperate bacteriophage community in IBD-affected gut samples.

Among the CD vs. non-IBD comparisons (i.e., active CD vs. non-IBD and inactive CD vs. non-IBD), with the exception of two viral OTUs, differential abundance was not observed among the bacterial hosts of viral OTUs that were differentially abundant (Figures 5 and 6). For UC and non-IBD patients, the hosts of 96 (66.2%) and 111 (88.1%) differentially-abundant viral OTUs did not show the same trend in abundance in the active UC vs. non-IBD and inactive UC vs. non-IBD comparisons, respectively. This suggests that the differential abundance displayed by these temperate bacteriophages is not merely a reflection of shifts in host abundance. However, techniques such as the single-cell viral tagging used by Džunková et al. will be needed in order to validate the current results [49].

In the active UC vs. non-IBD comparison, four temperate bacteriophages (viral OTUs 0791, 0838, 0802, and 1592) that were over-represented in active UC patients were accompanied by under-representation of their hosts (*B. uniformis*: MAG clusters 0358 and 0655, *B. thetaiotaomicron*: MAG cluster 2980). Interestingly, both *B. uniformis* and *B. thetaiotaomicron* have been experimentally proven to be beneficial to gut homeostasis [50–52]. *B. uniformis* isolated from a healthy fecal donor decreased the expression of pro-inflammatory cytokine interleukin-8 in lipopolysaccharide-induced human colorectal adenocarcinoma cells [50]. In another study, *B. thetaiotaomicron* significantly increased the concentration of interleukin-6 (IL-6) secreted by intraepithelial lymphocytes in the mouse colon, whereas IL-6 deficiency significantly decreased the integrity of the epithelial barrier [51]. Furthermore,

administration of *B. thetaiotaomicron* was found to ameliorate colon damage in IBD model mice [52]. Therefore, increased lysis of these beneficial bacteria by temperate bacteriophages may contribute to disease pathogenicity in patients with active UC.

In contrast to our results, a study conducted by the IBDMDB research group reported no differences in microbial abundance between IBD and non-IBD patients [1]. We suspect that this is largely the result of differences in how microbial abundance was assessed. In the current study, microbial abundance was assessed by read mapping to a collection of genome fragments reconstructed from the actual gut microbiota under study. In the previous study, microbial abundance was assessed via read mapping to clade-specific gene markers using MetaPhlan2 [53]. Although both methods have pros and cons, the current approach was selected for the current study because a large proportion of temperate bacteriophages are yet to be characterized in reference databases.

The current study only explored ecological traits (i.e., alpha diversity, beta diversity, and differential abundances) in temperate bacteriophage communities; thus, no assessments of the functional traits of temperate bacteriophages were conducted. The reason for limiting the design of our study in this way is related to the difficulty involved in recovering full-length temperate bacteriophage genomes from whole metagenome sequence data. Future developments in the metagenome contig binning technique for the viral genomes currently un-represented in reference databases may promote deeper understanding of the relationships between gene functional traits in temperate bacteriophages and IBD. We also acknowledge that our study is limited by the modest number of patients and lack of healthy subjects. Future studies on a larger population of patients and inclusion of healthy subjects will be needed for further confirmation of our results.

In conclusion, our results support growing evidence that the gut temperate bacteriophage community does indeed have a relationship with the pathogenesis of IBD. Notably, when compared with non-IBD patients, the temperate bacteriophages infecting two species that were experimentally shown to be beneficial to gut homeostasis, *B. uniformis* and *B. thetaiotaomicron*, were over-represented in patients with active UC. Worth mentioning is that, our study has reconstructed a large collection of functional temperate bacteriophage and bacterial genomes that may be of benefit to future studies. Further understanding of the role played by temperate bacteriophages in the human gut microbiota has potential to lead to the development of new phage therapy for IBD.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2076-2607/8/11/1663/s1>: Figure S1: Comparisons of Shannon's diversity index values for bacteria and temperate bacteriophages between IBD and non-IBD patients; Figure S2: Comparisons of bacterial and temperate bacteriophage community richness between IBD and non-IBD patients; Table S1: Temperate bacteriophages and their assigned hosts; Table S2: Differentially-abundant viral OTUs between active CD and non-IBD patients with FDR values < 1.0; Table S3: Differentially-abundant MAG clusters between active CD and non-IBD patients with FDR values < 1.0; Table S4: Differentially-abundant viral OTUs between inactive CD and non-IBD patients with FDR values < 1.0; Table S5: Differentially-abundant MAG clusters between inactive CD and non-IBD patients with FDR values < 1.0; Table S6: Differentially-abundant viral OTUs between active UC and non-IBD patients with FDR values < 1.0; Table S7: Differentially-abundant MAG clusters between active UC and non-IBD patients with FDR values < 1.0; Table S8: Differentially-abundant viral OTUs between inactive UC and non-IBD patients with FDR values < 1.0; Table S9: Differentially-abundant MAG clusters between inactive UC and non-IBD patients with FDR values < 1.0.

Author Contributions: Conceptualization, H.N.; Data curation, H.N.; Formal analysis, H.N.; Funding acquisition, H.N. and H.O.; Investigation, H.N.; Methodology, H.N. and R.B.-M.; Project administration, H.N.; Resources, H.O.; Software, H.N.; Supervision, H.E., R.B.-M. and H.O.; Validation, H.N.; Writing—original draft, H.N.; Writing—review & editing, H.N., H.E., R.B.-M. and H.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the JSPS KAKENHI (grant number 19J14228) and Scientific Research on Innovative Areas from the Ministry of Education, Culture, Science, Sports and Technology (MEXT) of Japan (grant numbers 16H06429, 16K21723, and 16H06437).

Acknowledgments: The authors acknowledge the IBDMDB research team, led by Ramnik Xavier and Curtis Huttenhower for generating the phenomenal metagenome dataset and making it available to the public along with its high-quality metadata. We also express gratitude to the NIH Common Fund for supporting Ramnik Xavier and Curtis Huttenhower (grant number 8U54DK102557). Computational time was provided by the SuperComputer System, Institute for Chemical Research, Kyoto University. We thank Tamsin Sheen and Sandra Cheesman for editing a draft of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Lloyd-Price, J.; Arze, C.; Ananthakrishnan, A.N.; Schirmer, M.; Avila-Pacheco, J.; Poon, T.W.; Andrews, E.; Ajami, N.J.; Bonham, K.S.; Brislawn, C.J.; et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **2019**, *569*, 655–662. [[CrossRef](#)] [[PubMed](#)]
2. Zuo, T.; Ng, S.C. The Gut Microbiota in the Pathogenesis and Therapeutics of Inflammatory Bowel Disease. *Front. Microbiol.* **2018**, *9*, 2247. [[CrossRef](#)]
3. 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. Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* **2015**, *160*, 447–460. [[CrossRef](#)] [[PubMed](#)]
4. Zuo, T.; Lu, X.-J.; Zhang, Y.; Cheung, C.P.; Lam, S.; Zhang, F.; Tang, W.; Ching, J.Y.L.; Zhao, R.; Chan, P.K.S.; et al. Gut mucosal virome alterations in ulcerative colitis. *Gut* **2019**, *68*, 1169–1179. [[CrossRef](#)] [[PubMed](#)]
5. Maronek, M.; Link, R.; Ambro, L.; Gardlik, R. Phages and Their Role in Gastrointestinal Disease: Focus on Inflammatory Bowel Disease. *Cells* **2020**, *9*, 1013. [[CrossRef](#)] [[PubMed](#)]
6. Garmaeva, S.; Sinha, T.; Kurilshikov, A.; Fu, J.; Wijmenga, C.; Zhernakova, A. Studying the gut virome in the metagenomic era: Challenges and perspectives. *BMC Biol.* **2019**, *17*, 1–14. [[CrossRef](#)]
7. Howard-Varona, C.; Hargreaves, K.R.; Abedon, S.T.; Sullivan, M.B. Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. *ISME J.* **2017**, *11*, 1511–1520. [[CrossRef](#)]
8. Shkoporov, A.N.; Hill, C. Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. *Cell Host Microbe* **2019**, *25*, 195–209. [[CrossRef](#)] [[PubMed](#)]
9. Simmonds, P.; Adams, M.J.; Benkő, M.; Breitbart, M.; Brister, J.R.; Carstens, E.B.; Davison, A.J.; Delwart, E.L.; Gorbalenya, A.E.; Harrach, B.; et al. Virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* **2017**, *15*, 161–168. [[CrossRef](#)] [[PubMed](#)]
10. Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project: Dynamic Analysis of Microbiome-Host Omics Profiles during Periods of Human Health and Disease. *Cell Host Microbe* **2014**, *16*, 276–289. [[CrossRef](#)]
11. Harvey, R.F.; Bradshaw, J.M. A simple index of Crohn’s-disease activity. *Lancet* **1980**, *1*, 514. [[CrossRef](#)]
12. Walmsley, R.S.; Ayres, R.C.S.; Pounder, R.E.; Allan, R.N. A simple clinical colitis activity index. *Gut* **1998**, *43*, 29–32. [[CrossRef](#)]
13. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015**, *31*, 1674–1676. [[CrossRef](#)] [[PubMed](#)]
14. Roux, S.; Enault, F.; Hurwitz, B.L.; Sullivan, M.B. VirSorter: Mining viral signal from microbial genomic data. *PeerJ* **2015**, *3*, e985. [[CrossRef](#)]
15. 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. [[CrossRef](#)]
16. Zhu, W.; Lomsadze, A.; Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **2010**, *38*, e132. [[CrossRef](#)] [[PubMed](#)]
17. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2018**, *47*, D427–D432. [[CrossRef](#)] [[PubMed](#)]
18. Eddy, S.R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **2011**, *7*, e1002195. [[CrossRef](#)]
19. 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. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **2019**, *37*, 632–639. [[CrossRef](#)]
20. Uritskiy, G.V.; DiRuggiero, J.; Taylor, J. MetaWRAP—A flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **2018**, *6*, 1–13. [[CrossRef](#)]
21. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]

22. Wu, Y.-W.; Simmons, B.A.; Singer, S.W. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **2015**, *32*, 605–607. [[CrossRef](#)] [[PubMed](#)]
23. Kang, D.D.; Li, F.; Kirton, E.; Thomas, A.; Egan, R.; An, H.; Wang, Z. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **2019**, *7*, e7359. [[CrossRef](#)]
24. Alneberg, J.; Bjarnason, B.S.; De Bruijn, I.; Schirmer, M.; Quick, J.; Ijaz, U.Z.; Lahti, L.; Loman, N.J.; Andersson, A.F.; Quince, C. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **2014**, *11*, 1144–1146. [[CrossRef](#)] [[PubMed](#)]
25. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [[CrossRef](#)]
26. Olm, M.R.; Brown, C.T.; Brooks, B.; Banfield, J.F. dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **2017**, *11*, 2864–2868. [[CrossRef](#)] [[PubMed](#)]
27. Chaumeil, P.-A.; Mussig, A.J.; Hugenholtz, P.; Parks, D.H. GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **2019**, *36*, 1925–1927. [[CrossRef](#)]
28. Segata, N.; Börnigen, D.; Morgan, X.C.; Huttenhower, C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* **2013**, *4*, 1–11. [[CrossRef](#)]
29. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [[CrossRef](#)]
30. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
31. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)] [[PubMed](#)]
32. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [[CrossRef](#)] [[PubMed](#)]
33. Kim, M.-S.; Bae, J.-W. Lysogeny is prevalent and widely distributed in the murine gut microbiota. *ISME J.* **2018**, *12*, 1127–1141. [[CrossRef](#)] [[PubMed](#)]
34. Gao, N.L.; Zhang, C.; Zhang, Z.; Hu, S.; Lercher, M.J.; Zhao, X.-M.; Bork, P.; Liu, Z.; Chen, W.-H. MVP: A microbe-phage interaction database. *Nucleic Acids Res.* **2018**, *46*, D700–D707. [[CrossRef](#)]
35. Waller, A.S.; Yamada, T.; Kristensen, D.M.; Kultima, J.R.; Sunagawa, S.; Koonin, E.V.; Bork, P. Classification and quantification of bacteriophage taxa in human gut metagenomes. *ISME J.* **2014**, *8*, 1391–1402. [[CrossRef](#)]
36. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.S.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
37. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)]
38. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
39. McMurdie, P.J.; Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **2013**, *8*, e61217. [[CrossRef](#)]
40. Pinheiro, J.; Bates, D.; DebRoy, S.; Sarkar, D.; Heisterkamp, S.; Willigen, B.V. nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-147. 2020. Available online: <https://CRAN.R-project.org/package=nlme> (accessed on 7 May 2020).
41. Anderson, M.J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **2001**, *26*, 32–46.
42. McInnes, L.; Healy, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *J. Open Source Softw.* **2018**, *3*, 861. [[CrossRef](#)]
43. Tange, O. GNU Parallel—The Command-Line Power Tool. *USENIX Mag.* **2011**, *36*, 42–47.
44. Paulson, J.N.; Stine, O.C.; Bravo, H.C.; Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* **2013**, *10*, 1200–1202. [[CrossRef](#)] [[PubMed](#)]
45. Cresci, G.A.; Bawden, E. Gut Microbiome: What We Do and Don't Know. *Nutr. Clin. Pract.* **2015**, *30*, 734–746. [[CrossRef](#)] [[PubMed](#)]

46. Kostic, A.D.; Xavier, R.J.; Gevers, D. The Microbiome in Inflammatory Bowel Disease: Current Status and the Future Ahead. *Gastroenterology* **2014**, *146*, 1489–1499. [[CrossRef](#)] [[PubMed](#)]
47. Shkoporov, A.N.; Clooney, A.G.; Sutton, T.D.; Ryan, F.J.; Daly, K.M.; Nolan, J.A.; McDonnell, S.A.; Khokhlova, E.V.; Draper, L.A.; Forde, A.; et al. The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell Host Microbe* **2019**, *26*, 527–541. [[CrossRef](#)] [[PubMed](#)]
48. Shkoporov, A.N.; Ryan, F.J.; Draper, L.A.; Forde, A.; Stockdale, S.R.; Daly, K.M.; McDonnell, S.A.; Nolan, J.A.; Sutton, T.D.; Dalmasso, M.; et al. Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* **2018**, *6*, 1–17. [[CrossRef](#)] [[PubMed](#)]
49. Džunková, M.; Low, S.J.; Daly, J.N.; Deng, L.; Rinke, C.; Hugenholtz, P. Defining the human gut host–phage network through single-cell viral tagging. *Nat. Microbiol.* **2019**, *4*, 2192–2203. [[CrossRef](#)] [[PubMed](#)]
50. Hiippala, K.; Kainulainen, V.; Suutarinen, M.; Heini, T.; Bowers, J.R.; Jasso-Selles, D.; Lemmer, D.; Valentine, M.; Barnes, R.; Engelthaler, D.M.; et al. Isolation of Anti-Inflammatory and Epithelium Reinforcing Bacteroides and Parabacteroides Spp. from a Healthy Fecal Donor. *Nutrients* **2020**, *12*, 935. [[CrossRef](#)]
51. Kuhn, K.A.; Schulz, H.M.; Regner, E.H.; Severs, E.L.; Hendrickson, J.D.; Mehta, G.; Whitney, A.K.; Ir, D.; Ohri, N.; Robertson, C.E.; et al. Bacteroidales recruit IL-6-producing intraepithelial lymphocytes in the colon to promote barrier integrity. *Mucosal Immunol.* **2017**, *11*, 357–368. [[CrossRef](#)]
52. Delday, M.; Mulder, I.; Logan, E.T.; Grant, G. Bacteroides thetaiotaomicron Ameliorates Colon Inflammation in Preclinical Models of Crohn’s Disease. *Inflamm. Bowel Dis.* **2018**, *25*, 85–96. [[CrossRef](#)]
53. Truong, D.T.; Franzosa, E.A.; Tickle, T.L.; Scholz, M.; Weingart, G.; Pasolli, E.; Tett, A.; Huttenhower, C.; Segata, N. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **2015**, *12*, 902–903. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).