Microbiome

RESEARCH

Open Access

Gut virome-wide association analysis identifies cross-population viral signatures for inflammatory bowel disease

Xiangge Tian[1,2†], Shenghui Li[3†], Chao Wang[2†], Yanyan Zhang[1,2†], Xiaoying Feng[1], Qiulong Yan[1,2*], Ruochun Guo[3], Fan Wu[1], Chunxue Wu[1], Yan Wang[1], Xiaokui Huo[1,2*] and Xiaochi Ma[1,2*]

Abstract

Background The gut virome has been implicated in inflammatory bowel disease (IBD), yet a full understanding of the gut virome in IBD patients, especially across diverse geographic populations, is lacking.

Results In this study, we conducted a comprehensive gut virome-wide association study in a Chinese cohort of 71 IBD patients (15 with Crohn's disease and 56 with ulcerative colitis) and 77 healthy controls via viral-like particle (VLP) and bulk virome sequencing of their feces. By utilizing an integrated gut virus catalog tailored to the IBD virome, we revealed fundamental alterations in the gut virome in IBD patients. These characterized 139 differentially abundant viral signatures, including elevated phages predicted to infect *Escherichia*, *Klebsiella*, *Enterococcus_B*, *Streptococcus*, and *Veillonella species*, as well as IBD-depleted phages targeting *Prevotella*, *Ruminococcus_E*, *Bifidobacterium*, and *Blautia species*. Remarkably, these viral signatures demonstrated high consistency across diverse populations such as those in Europe and the USA, emphasizing their significance and broad relevance in the disease context. Furthermore, fecal virome transplantation experiments verified that the colonization of these IBD-characterized viruses can modulate experimental colitis in mouse models.

Conclusions Building upon these insights into the IBD gut virome, we identified potential biomarkers for prognosis and therapy in IBD patients, laying the foundation for further exploration of viromes in related conditions.

Keywords Gut virome, Inflammatory bowel disease, Viral-like particle virome, Bulk virome, Viral signatures, IBD-characterized viruses

Background

The human gut virome represents a dynamic community of viruses inhabiting the gastrointestinal tract, primarily comprising bacteriophages with limited eukaryotic and archaeal viral presence. Factors influencing gut virome composition and diversity include genetics, diet, antibiotic use, and environmental exposures [1–4]. This complex virome has been implicated in the pathogenesis of several human disorders. Notably, it was first associated with inflammatory bowel disease (IBD) [5], and subsequent research linked the gut virome to conditions such as obesity, diabetes [6], metabolic syndrome [7], colorectal cancer [8], autoimmune diseases [9–11], and liver

[†] Xiangge Tian, Shenghui Li, Chao Wang, and Yanyan Zhang contributed equally to this work.

*Correspondence:
Qiulong Yan
qiulongy1988@163.com
Xiaokui Huo
huoxiaokui@163.com
Xiaochi Ma
maxc1978@163.com
[1] Second Affiliated Hospital, Dalian Medical University, Dalian 116023, China
[2] Dalian Key Laboratory of Metabolic Target Characterization and Traditional Chinese Medicine Intervention, College of Basic Medical Sciences, Dalian Medical University, Dalian 116044, China
[3] Puensum Genetech Institute, Wuhan 430076, China

disease [12, 13]. These interactions occur within the context of the gut microbiome and the host immune system and are crucial for maintaining gut homeostasis and defending against pathogens [14].

The study of the gut virome predominantly employs two approaches: bulk metagenome sequencing and virus-like particle (VLP) enrichment followed by sequencing. While bulk metagenome analysis has long been used to characterize the entire microbial community (i.e., bacteriome, archaeome, and mycobiome) in the gut [15], its use in predicting viral genomes is relatively recent, with viral reference databases still evolving [16]. In contrast, VLP enrichment-based technology has emerged as a promising tool for studying the gut virome in the past few years [17, 18], yielding a greater proportion of viral sequences and simplifying the identification of viruses in complex microbiota. Nevertheless, VLP enrichment has its challenges, including incomplete databases and potential methodological biases, requiring resolution to enhance the understanding of viral characteristics linked to diseases.

Both VLP and bulk virome approaches have been used to investigate the characteristics of the gut virome in patients with IBD, including Crohn's disease (CD) and ulcerative colitis (UC). Early VLP-based studies revealed a significant expansion of *Caudovirales* bacteriophages in CD and UC patients [5]. A subsequent study expanded on this by identifying IBD-specific virome changes and found increased temperate phages in CD patients [19]. Moreover, bulk virome analysis revealed significant alterations in eukaryotic viruses in IBD [20], consistent with findings in children with IBD [21, 22]. Experimental advances have demonstrated that fecal virus-like particles isolated from UC patients exacerbate colitis severity in mice [23], and viromes from IBD colon tissue promote inflammation in mice [24]. However, further research is needed to obtain detailed information about gut viral structural changes and their impact on IBD. Comprehensive technologies and reference databases are essential for the universal and streamlined identification of these associated changes.

In this study, we designed and carried out a virome-wide association study based on VLP and bulk metagenome sequencing and analysis of fecal samples from an IBD patient and healthy control cohort (total $n = 148$). We established an integrated gut virus catalog tailored to the IBD virome, facilitating the identification of numerous viral signatures. Notably, these signatures displayed remarkable consistency across diverse populations, including those in Europe and the USA. Moreover, we conducted fecal virome transplantation (FVT) experiments using human fecal VLPs in a mouse model, confirming the roles of IBD viral signatures in disease progression. Our findings provide valuable insights into the IBD gut virome, offering a range of reproducible biomarkers with potential implications for prognostic and therapeutic strategies, and establish a framework for exploring virome landscapes in other relevant disorders.

Results

Subject characteristics

Our cohort was composed of 71 patients with IBD (including 15 with CD and 56 with UC) and 77 healthy controls. The phenotypic characteristics of all the subjects are summarized in Supplementary Table 1. There were no significant differences in age, sex, or body mass index (BMI) between IBD patients and healthy controls, and these characteristics did not differ between patients with CD and those with UC. The disease activity of IBD patients was determined by the simplified Crohn's Disease Activity Index (sCDAI) [25], which revealed that 66.7% of CD patients and 82.1% of UC patients experienced moderate or severe activity periods.

Gut virome sequencing and virus catalog construction

To characterize IBD virome, we processed and analyzed fecal samples from 148 subjects using VLP metagenomic and bulk metagenomic sequencing technologies (detailed in "Methods"). This generated a total of 462.0 Gbp (VLP clean data, averaging 3.3 ± 1.9 Gbp per sample) and 1.14 Tbp (bulk clean data, average 7.7 ± 5.2 Gbp per sample) of high-quality nonhuman data (Fig. 1a). The metagenomic data of each sample were individually de novo assembled, and the viral sequences (representing > 64,000 candidates for all samples) were identified from these assembled contigs using an integrated pipeline detailed in the "Methods" section, consistent with our prior studies [18, 26, 27]. As a considerable number of viral sequences in this study were potentially fully assembled in other Chinese metagenomes, we incorporated ~ 17,000 viral genomes from the comprehensive Chinese gut virus catalog (cnGVC) [26], which exhibited high similarity and coverage with our current viral candidates. Next, we combined all viral sequences and removed duplicates with > 95% nucleotide identity across > 70% of the sequences, resulting in a catalog of 10,054 viral operational taxonomic units (vOTUs) for subsequent analyses (Supplementary Table 2). Utilizing CheckV [28], it was estimated that the majority (60.6%) of vOTUs possessed complete or high-quality (≥ 90% completeness) genomes, with over 90% of vOTUs showing low contamination (< 10%) (Fig. 1b). Taxonomically, 39.5% and 4.9% of vOTUs could be assigned to known prokaryotic or eukaryotic viral families, respectively, while the remaining vOTUs (55.6% of all vOTUs) remained unclassified at the family level. Consistent with previous studies [18, 19,

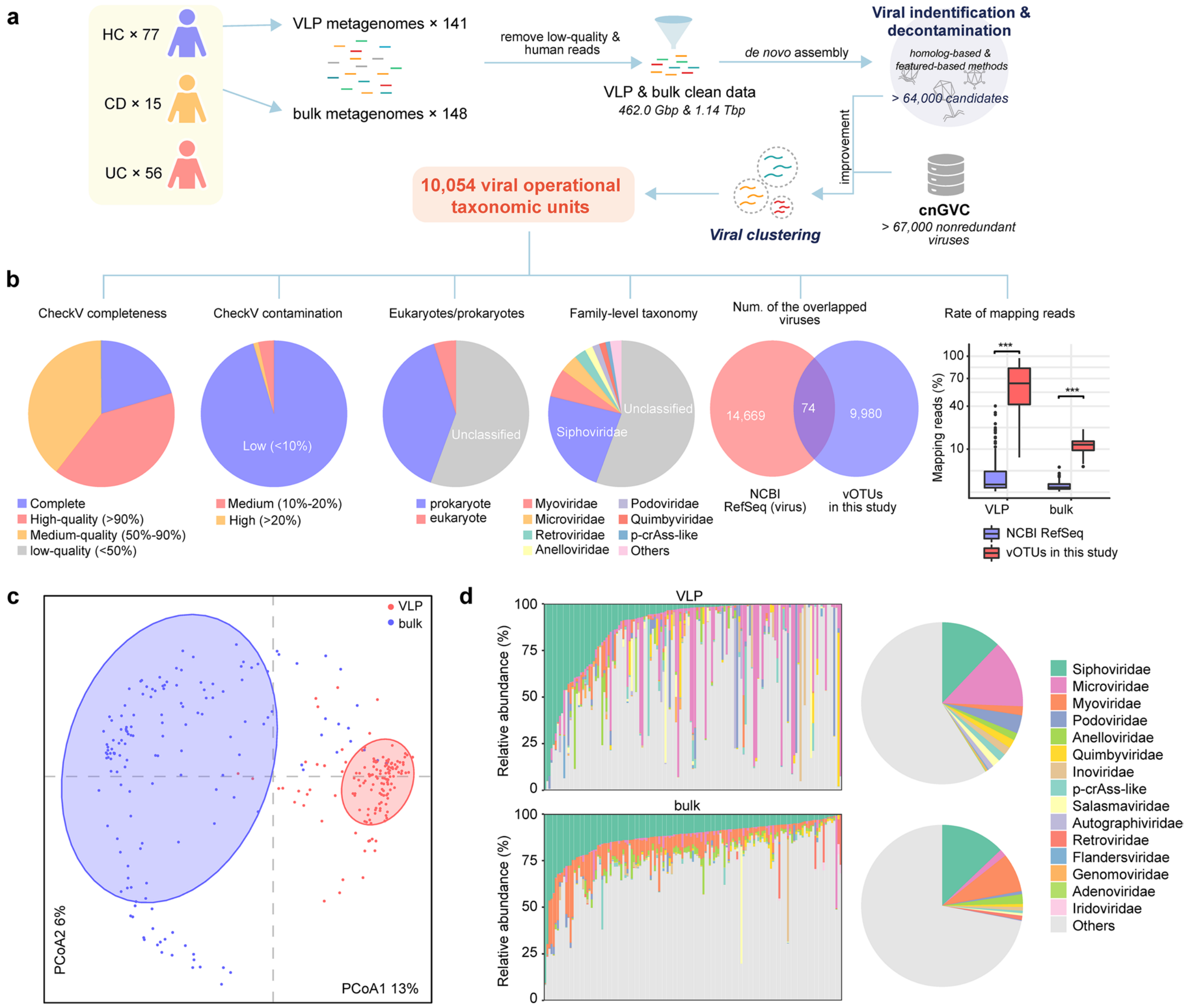


Fig. 1 Overview of the study design and the integrated gut virus catalog. **a** Workflow of the construction of the nonredundant gut virus catalog from IBD patients and healthy subjects. **b** Statistics of the gut virus catalog. Pie plots show the completeness, contamination, proportions of eukaryotic and prokaryotic viruses, and family-level taxonomic classification of the virus catalog. Venn diagram shows the overlap between the integrated gut virus catalog and the viruses from the NCBI RefSeq database. Boxplot shows the fraction of mapped metagenomic reads to the integrated gut virus catalog and the viruses from the NCBI RefSeq database. Wilcoxon rank-sum test: ****p* < 0.001. **c** Principal coordinate analysis (PCoA) reveals the difference between the VLP virome and bulk virome (141 paired samples were included in the analysis). Samples are shown at the first and second principal coordinates (PCoA1 and PCoA2), and the variance explained (%) by these two PCoAs is displayed. Ellipses represent a 95% confidence interval surrounding each group. **d** Comparison of the gut viral structure between VLPs and bulk viromes. Bar plots show the family-level composition of each sample from the VLP and bulk viromes. Pie plots show the overall proportions

29], we found that the most dominant classifiable vOTU families were *Siphoviridae*, *Myoviridae*, *Microviridae*, and *Podoviridae*. Significantly, only 74 vOTUs (< 1% of all vOTUs) exhibited species-level homology with viral genomes from the NCBI RefSeq database (Fig. 1b), highlighting the extraordinary novelty of our catalog. Moreover, an average of 59.8% and 11.7% of the metagenomic reads from the VLP and bulk datasets, respectively, were captured by our vOTU catalog, which was more than an order of magnitude greater than that of the RefSeq viruses (3.1% and 0.3%, respectively; Fig. 1b). This underscores the broader and more comprehensive profiling capability of our viral genome catalog in studying IBD and healthy viromes.

Principal coordinate analysis (PCoA) based on the Bray–Curtis distance at the vOTU level clearly separated the VLP and bulk viromes (permutational multivariate analysis of variance [PERMANOVA] $p < 0.001$;

Fig. 1c). This pattern was consistent with the family-level viral profile (Supplementary Fig. 1a). Almost all families exhibited significantly different relative abundances between various samples obtained from the two technologies, especially in the case of two predominant families, *Microviridae* (average relative abundance 13.5% vs. 1.4% in VLP and bulk viromes, respectively, Wilcoxon rank-sum test $q = 2.5 \times 10^{-11}$) and *Myoviridae* (1.8% vs. 7.8%, $q = 1.6 \times 10^{-16}$) (Fig. 1d). Notably, of the 401 *Microviridae* vOTUs present in the VLP metagenomes, only 61.8% (248/401) had reads captured by the bulk metagenomic samples (Supplementary Fig. 1b). These results concur with prior studies highlighting the substantial differences between VLPs and bulk technologies in virome capture and profiling [26, 30], emphasizing the need to integrate both methods to comprehensively study entire gut viral communities.

Eukaryotic and prokaryotic virome diversity in IBD patients and healthy subjects

To gauge the gut virome diversity concerning IBD, we conducted a comparative analysis of the observed index (reflecting viral richness) and Shannon index (representing viral diversity) within the viromes of IBD patients and healthy controls. The examination was carried out separately for eukaryotic and prokaryotic viromes due to disparities in the results obtained (as elaborated below). In the eukaryotic virome, we observed a significant increase in both the observed and Shannon indices for IBD patients in comparison to healthy controls. This pattern held true for both the VLP and bulk metagenomic datasets (Fig. 2a). The elevation in eukaryotic viruses, which can infect host cells in the gut, aligns with earlier research suggesting that the heightened presence of eukaryotic viruses may serve as a hallmark in IBD patients, particularly those with UC [20]. In contrast, when considering the prokaryotic virome, we observed that diversity indices (especially the Shannon index) were greater in IBD patients than in healthy controls in the VLP dataset. However, a reversal of this trend was observed in the bulk dataset, where the indices were significantly lower in IBD patients than in controls (Fig. 2b). This discrepancy may stem from the preference of the VLP technique for free viral particles, which tends to increase in IBD patients. In contrast, bulk metagenomes predominantly capture actively infecting viruses or integrated prophages [30], and these are diminished in IBD patients along with some bacteria. As a corroborative point, we found that gut bacteriome diversity significantly decreased in IBD patients in comparison to controls. Importantly, this bacteriome diversity exhibited a stronger positive correlation with the prokaryotic virome diversity in the bulk dataset than in the VLP dataset (Supplementary Fig. 2).

Gut virome structure associated with IBD

PCoA analysis of the gut virome structure indicated a clear demarcation between IBD patients and healthy subjects, irrespective of whether VLP or bulk viromes were considered (Fig. 2c), underscoring a profound alteration in the viromes of IBD patients. Intriguingly, no significant difference was noted between the gut viromes of CD and UC patients in the PCoA plot. In contrast to IBD-control differentiation, which accounted for 1.9% (PERMANOVA $p < 0.001$) and 7.2% ($p < 0.001$) of the gut virome variance in the VLP and bulk datasets, respectively, the effect of CD-UC differentiation was substantially lower at 0.8% (PERMANOVA $p = 0.513$) and 1.7% ($p = 0.029$). Furthermore, demographic parameters such as age, sex, BMI, and disease activity had minimal impacts on gut virome variance (effect size < 0.5%, PERMANOVA $p > 0.05$ for all parameters).

Given the similarity in gut virome diversity and structure between CD patients and UC patients, we proceeded with a case–control comparison of the gut viral profiles for all IBD patients against those of healthy controls. This analysis focused on the 20 most abundant viral families, representing 99.7% of the total relative abundance of known viral families in both the VLP and bulk viromes. The results revealed that 11 out of the 20 viral families in the VLP metagenomes displayed trends of change consistent with those observed in the bulk metagenomes (Fig. 2d; Supplementary Table 3). Among these viral families, *Siphoviridae*, the predominant viral family in the human gut, displayed a mean relative abundance that was 250% greater (Wilcoxon rank-sum test $q = 4.0 \times 10^{-5}$) in the VLP dataset and 89% greater ($q = 6.2 \times 10^{-4}$) in the bulk dataset of IBD patients compared to controls. However, in the IBD gut virome, *Myoviridae* and *Salasmaviridae* were significantly enriched in the VLP dataset but not in the bulk dataset, while *Podoviridae* and *Inoviridae* were enriched in the bulk dataset but not in the VLP dataset. In contrast, *crAss-like* and *Quimbyviridae* viruses experienced a significant reduction in relative abundance in both datasets of IBD patients. Reduction in *Microviridae*, *Adenoviridae*, *Gratiaviridae*, and *Autographiviridae* was exclusively observed in the VLP dataset and not in the bulk dataset. *Anelloviridae* exhibited conspicuous enrichment among IBD patients in the VLP dataset but a marked reduction in the bulk dataset, whereas *Inoviridae* exhibited the opposite trend. (Fig. 2d). At the vOTU level, almost all *Anelloviridae* vOTUs demonstrated increased relative abundance in IBD patients, regardless of whether they were detected in the VLP or bulk datasets (Supplementary Fig. 3). Furthermore, consistent with prior studies [5, 21], it is noteworthy that the relative abundance of *Caudovirales* (primarily composed of *Siphoviridae* and *Myoviridae*) significantly increased in IBD patients

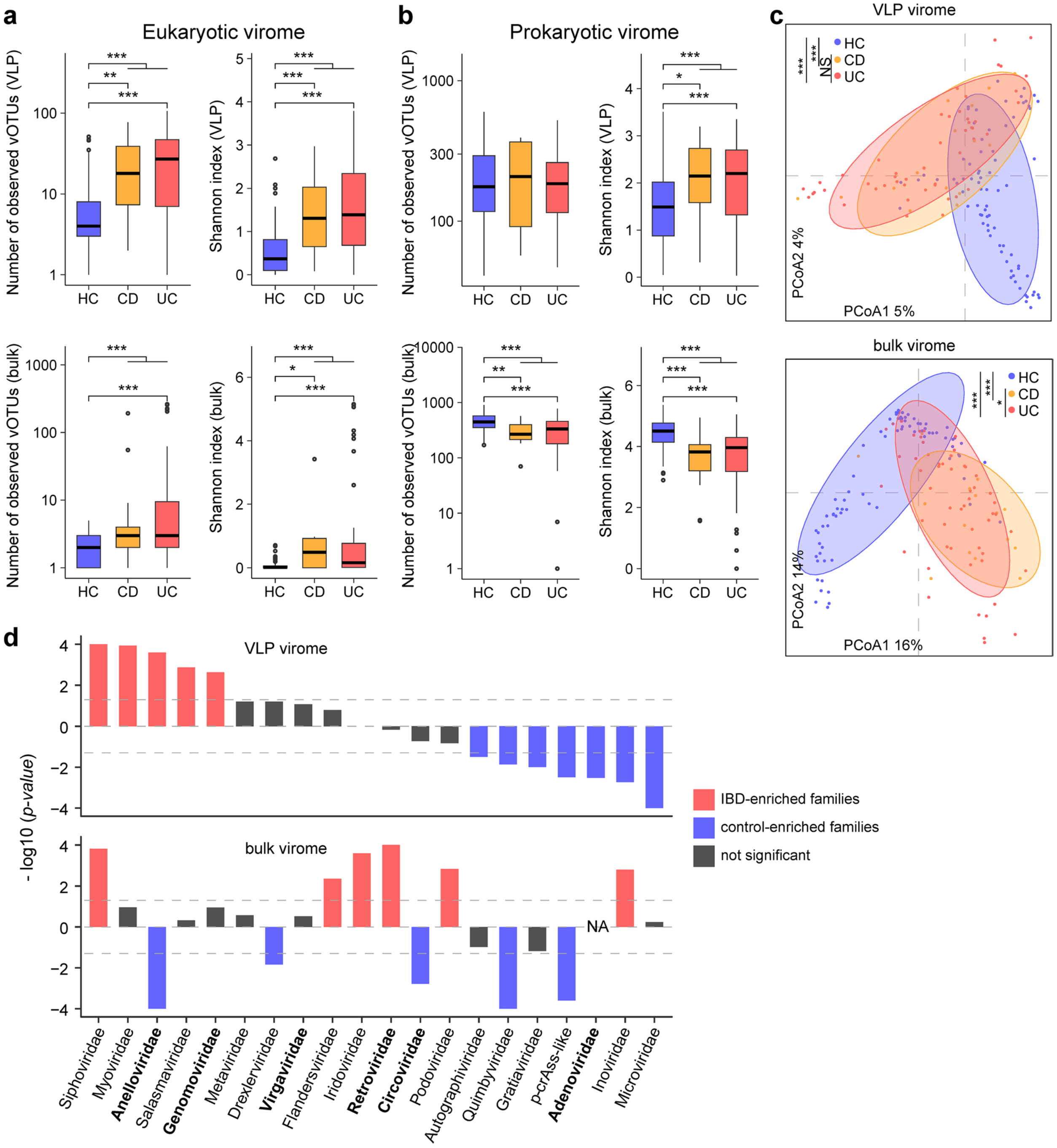


Fig. 2 Diversity and structure of the gut virome associated with IBD. **a**–**b** Boxplot showing the comparison of eukaryotic (**a**) and prokaryotic (**b**) viromes between IBD patients and healthy controls. Boxes represent the interquartile range between the first and third quartiles and the median (internal line). Whiskers denote the lowest and highest values within 1.5 times the range of the first and third quartiles, respectively; dots represent outlier samples beyond the whiskers. Wilcoxon rank-sum test: **p* < 0.05; ***p* < 0.01; ****p* < 0.001. HC, healthy controls. **c** Principal coordinate analysis (PCoA) reveals the differences in the VLP (upper panel) and bulk (bottom panel) viromes of IBD patients and healthy controls. The samples are shown at the first and second principal coordinates (PCoA1 and PCoA2), and the ratio of variance contributed by these two PCs is shown. PERMANOVA *adonis* test: ****p* < 0.001; ns not significant. The ellipses represent a 95% confidence interval surrounding each group. **d** Family-level comparison of the VLP and bulk viromes of patients and controls. The bar plot shows *p*-values from the Wilcoxon rank-sum test for comparison between two groups. The direction (positive or negative) of the bars indicates how the average relative abundance of each family differs between the patient and control groups. The names of the eukaryotic viruses are bolded

compared to controls in both datasets, while that of *Petitvirales* (mainly composed of *Microviridae*) significantly decreased (Supplementary Fig. 4).

Identification of IBD-associated viral signatures

To determine the viral species associated with IBD, we conducted a case–control comparison of the vOTU composition in both the VLP and bulk viromes using the Wilcoxon rank-sum test with Benjamini–Hochberg adjustment. This analysis identified a total of 139 vOTUs that displayed significant differences in relative abundances and consistent trends in case–control comparisons in both datasets (Fig. 3a; Supplementary Table 4); these vOTUs were designated IBD-associated gut viral signatures. Notably, a Fisher's exact test analysis indicated a high level of agreement in the identification of differential vOTUs between the two datasets (odds ratio = 163; $p < 0.001$), signifying that these IBD-associated markers

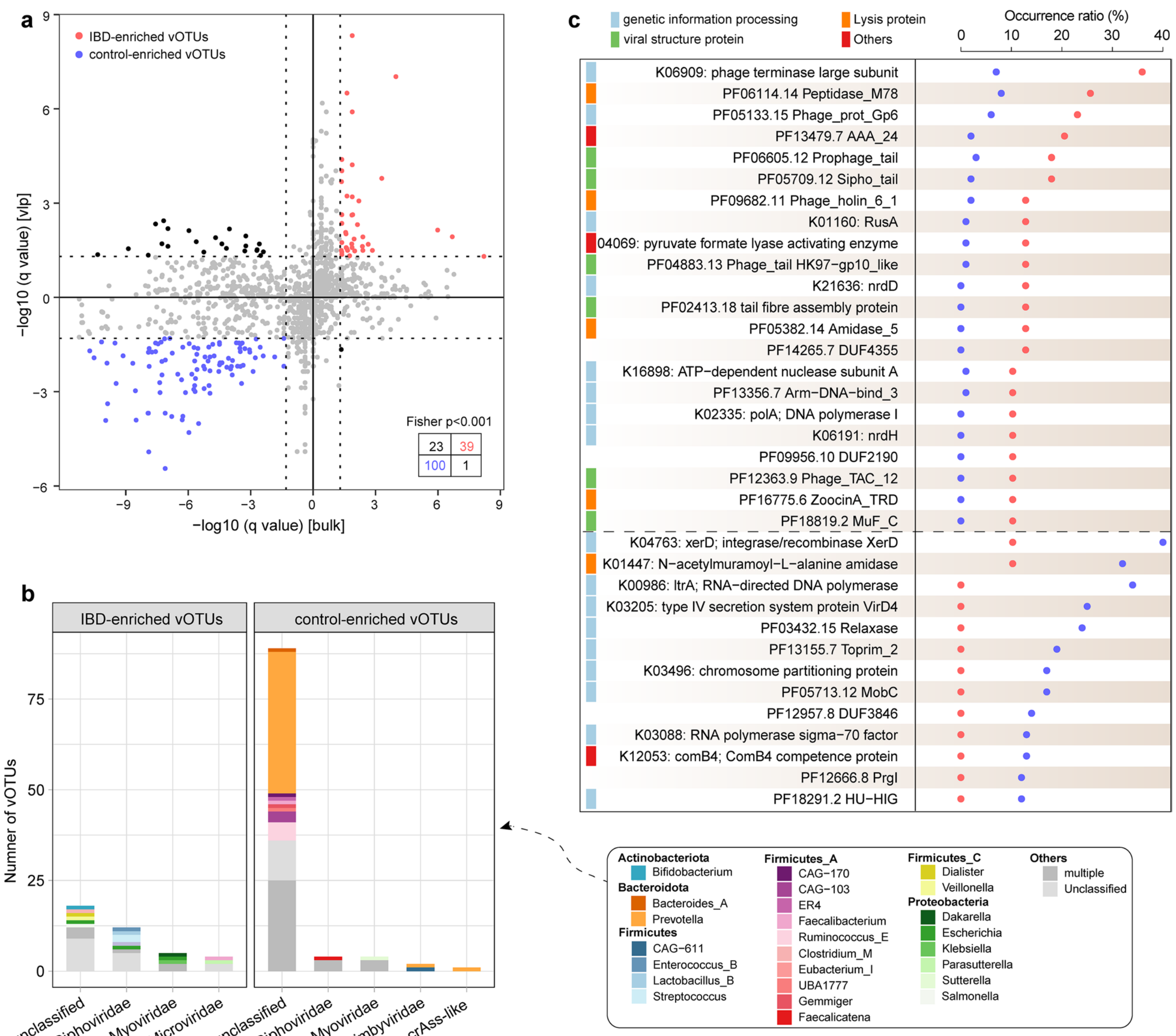


Fig. 3 Identification of IBD-associated vOTUs and exploration of their prokaryotic hosts and potential functions. **a** Scatter plot of *q*-values of vOTUs in the VLP and bulk datasets. Red and blue nodes represent the vOTUs that are significantly enriched in the viromes of IBD patients and healthy controls, respectively, with the same tendency in the VLP and bulk datasets. Statistical significance is determined using the Wilcoxon rank-sum test with Benjamini–Hochberg adjustment. **b** Distribution of the taxonomic annotation and host assignment of the IBD-enriched and control-enriched vOTUs. The vOTUs are grouped at the family level, and the prokaryotic host taxa are also shown at the family level. **c** Occurrence rates of 40 differential functions in IBD-enriched and control-enriched vOTUs. Red and blue nodes represent the IBD-enriched and control-enriched vOTUs, respectively. Statistical significance is determined using Fisher's exact test with Benjamini–Hochberg adjustment, and a $q < 0.05$ was considered statistically significant

were not randomly identified. Among these 139 IBD-associated markers, 39 were IBD-enriched vOTUs, representing members of the *Siphoviridae*, *Myoviridae*, and *Microviridae* families, as well as unclassified viruses (Fig. 3b). In contrast, the majority of control-enriched vOTUs (89 out of 100) were from unknown families, with a few originating from the *Siphoviridae*, *Myoviridae*, *Quimbyviridae*, and *crAss-like* families.

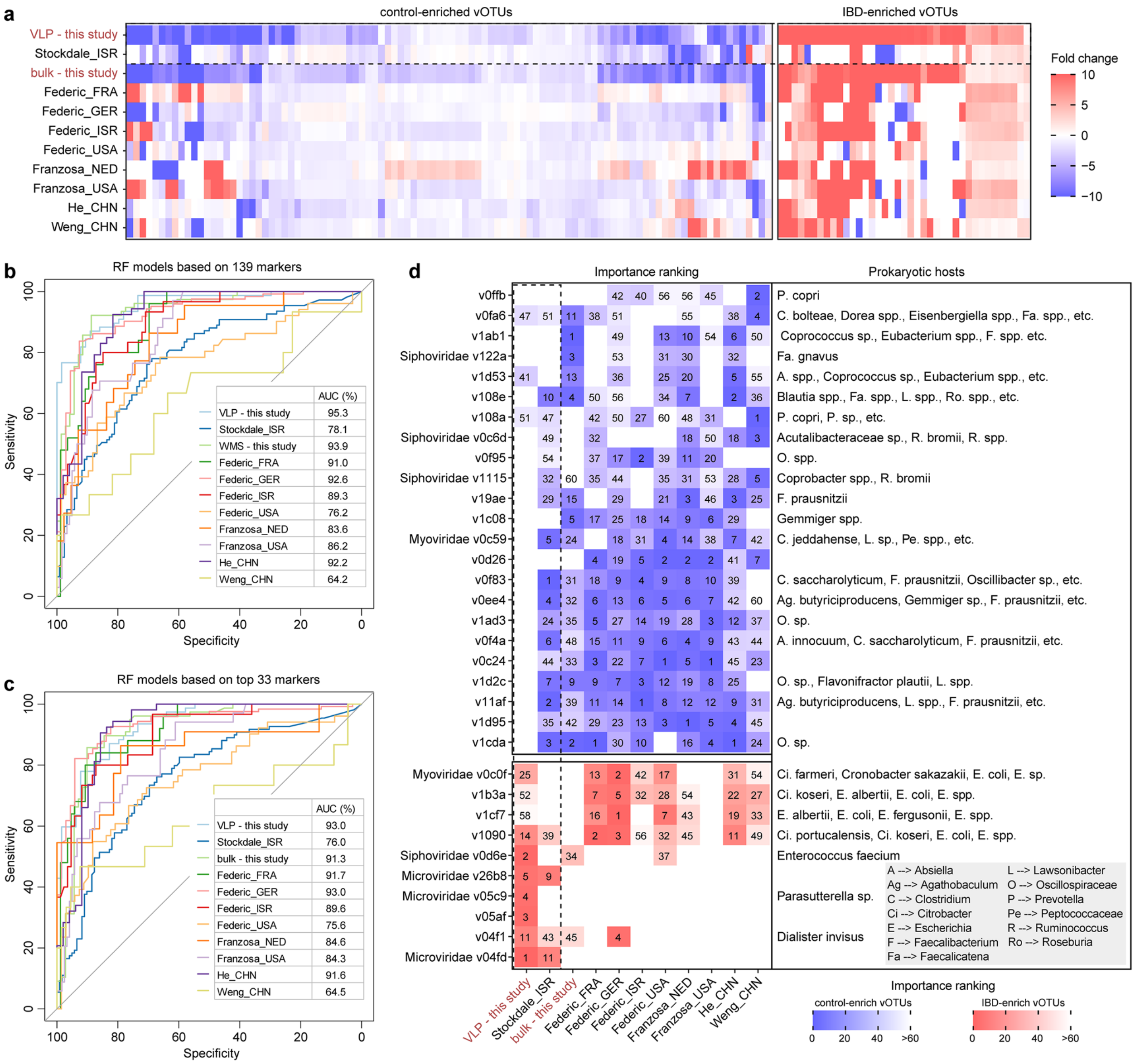
To investigate the prokaryotic hosts infected by these viral markers associated with IBD, we performed virus-host prediction based on the comprehensive Unified Human Gastrointestinal Genome (UHGG) collection containing 4644 gut prokaryotes [31] (see "Methods" for details). Approximately, 80% of the viral markers (112/139) were assigned to at least one prokaryotic host, with these hosts primarily belonging to the Bacteroidota, Firmicutes_A, Firmicutes, and Proteobacteria phyla (Supplementary Fig. 5). A notable difference in host preference was observed between IBD- and control-enriched vOTUs (Fig. 3b). For instance, IBD-enriched vOTUs frequently include genera containing common bacterial pathogens, such as *Escherichia*, *Klebsiella*, and *Streptococcus*. Moreover, the genus *Veillonella*, which was identified as an IBD-enriched bacteria in previous studies [32], was also infected by IBD-enriched vOTUs. In contrast, there was a high percentage of control-enriched vOTUs with predicted hosts (54%) infected with species from Bacteroidota, particularly *Prevotella* species, with most of these vOTUs lacking annotations from known viral families. A few control-enriched vOTUs were predicted to infect species from *Ruminococcus_E* and *Faecalibacterium*.

Furthermore, we conducted a functional comparison between IBD- and control-enriched vOTUs. Gene prediction of these vOTUs resulted in 8466 protein-coding genes, 49.2% of which were annotated as functional orthologs within the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Pfam databases [33, 34], spanning 1747 functional families. PCoA and PERMANOVA revealed a significant difference in the functional profiles between IBD-enriched and control-enriched vOTUs (PERMANOVA $p < 0.001$; Supplementary Fig. 6). Using Fisher's exact test with Benjamini–Hochberg adjustment, we identified 35 functional families that significantly differed in the occurrence rate between IBD-enriched and control-enriched vOTUs ($q < 0.05$; Fig. 3c). Functions related to viral structure proteins, particularly those involved in phage tail assembly ($n = 6$), were notably prevalent in IBD-enriched vOTUs but largely absent in control-enriched vOTUs. This suggests that the control-enriched vOTUs were more likely to belong to non-tailed phages. Moreover, four lysis proteins (PF06114.14, PF09682.11, PF05382.14, and PF16775.6), which interfere with the growth of host bacteria, were also found to be more commonly present in IBD-enriched vOTUs. Conversely, control-enriched vOTUs more frequently encoded genes associated with genetic information processing, indicating a greater focus on genetic replication and expression. Notably, more than 1/3 of the control-enriched vOTUs possessed the reverse transcriptase gene (K00986). These viruses were not annotated as *Retroviridae*, which predominantly infect human and animal vertebrates but were more likely to infect members of Bacteroidota. Numerous studies have demonstrated the presence of reverse transcriptase homologous genes in bacterial genomes, known as retrons, which function as components of the bacterial defense system against phages [35, 36]. Interestingly, we observed that, in addition to carrying the reverse transcriptase gene, several control-enriched vOTUs also contained typical viral protein sequences such as capsid and tape measure proteins (Supplementary Fig. 7). This suggests that these reverse transcriptase genes may not originate from host bacterial genomes but rather from the phages infecting them.

The IBD viral signatures are consistent across global cohorts

To assess the reliability and consistency of the 139 IBD-associated vOTUs identified in this study, we validated them in publicly available IBD cohorts from diverse geographic origins. We collected 1171 gut VLP and bulk metagenomic sequencing samples from 9 case–control datasets, including data from the USA, Europe, and China (Supplementary Table 5). We processed these samples by filtering out low-quality and human reads, mapping them to our integrated gut virus catalog containing 10,054 vOTUs, and calculating the relative abundance of each vOTU. PCoA revealed a substantial difference in viral composition between IBD patients and healthy controls among all samples from the 11 datasets (including the 9 public datasets and the VLP and bulk datasets from this study) (PERMANOVA $p < 0.001$; Supplementary Fig. 8), indicating the presence of common signatures in different populations. We then examined the direction of enrichment in the relative abundance of the 139 IBD-associated vOTUs between IBD patients and healthy controls in each cohort. Impressively, the direction of enrichment of these vOTUs displayed remarkable consistency across cohorts (average consistency rate of 70.0 ± 10.4% among the nine public datasets), despite a few exceptions in individual cohorts (Fig. 4a). This finding underscores the robustness of the IBD-associated vOTUs identified in this study in other public databases.

We built random forest (RF) models using the 139 IBD-associated vOTUs to evaluate their potential for classifying IBD patients from healthy controls in each



b

	AUC (%)
VLP - this study	95.3
Stockdale_ISR	78.1
WMS - this study	93.9
Federic_FRA	91.0
Federic_GER	92.6
Federic_ISR	89.3
Federic_USA	76.2
Franzosa_NED	83.6
Franzosa_USA	86.2
He_CHN	92.2
Weng_CHN	64.2

c

	AUC (%)
VLP - this study	93.0
Stockdale_ISR	76.0
bulk - this study	91.3
Federic_FRA	91.7
Federic_GER	93.0
Federic_ISR	89.6
Federic_USA	75.6
Franzosa_NED	84.6
Franzosa_USA	84.3
He_CHN	91.6
Weng_CHN	64.5

d

	VLP - this study	Stockdale_ISR	bulk - this study	Federic_FRA	Federic_GER	Federic_ISR	Federic_USA	Franzosa_NED	Franzosa_USA	He_CHN	Weng_CHN	Prokaryotic hosts
v0ffb					42	40	56	56	45		2	P. copri
v0fa6	47	51	11	38	51			55		38	4	C. bolteae, Dorea spp., Eisenbergiella spp., Fa. spp., etc.
v1ab1			1		49		13	10	54	6	50	Coprococcus sp., Eubacterium spp., F. spp. etc.
Siphoviridae v122a			3		53		31	30		32		Fa. gnavus
v1d53	41		13		36		25	20		5	55	A. spp., Coprococcus sp., Eubacterium spp., etc.
v108e		10	4	50	56		34	7		2	36	Blautia spp., Fa. spp., L. spp., Ro. spp., etc.
v108a	51	47		42	50	27	60	48	31		1	P. copri, P. sp., etc.
Siphoviridae v0c6d		49		32				18	50	18	3	Acutalibacteraceae sp., R. bromii, R. spp.
v0f95		54		37	17	2	39	11	20			O. spp.
Siphoviridae v1115		32	60	35	44		35	31	53	28	5	Coprobacter spp., R. bromii
v19ae		29	15		29		21	3	46	3	25	F. prausnitzii
v1c08			5	17	25	18	14	9	6	29		Gemmiger spp.
Myoviridae v0c59		5	24		18	31	4	14	38	7	42	C. jeddahense, L. sp., Pe. spp., etc.
v0d26				4	19	5	2	2	2	41	7	
v0f83		1	31	18	9	4	9	8	10	39		C. saccharolyticum, F. prausnitzii, Oscillibacter sp., etc.
v0ee4		4	32	6	13	6	5	6	7	42	60	Ag. butyriciproducens, Gemmiger sp., F. prausnitzii, etc.
v1ad3		24	35	5	27	14	19	28	3	12	37	O. sp.
v0f4a		6	48	15	11	9	6	4	9	43	44	A. innocuum, C. saccharolyticum, F. prausnitzii, etc.
v0c24		44	33	3	22	7	1	5	1	45	23	
v1d2c		7	9	9	7	3	12	19	8	25		O. sp., Flavonifractor plautii, L. spp.
v11af		2	39	11	14	1	8	12	12	9	31	Ag. butyriciproducens, L. spp., F. prausnitzii, etc.
v1d95		35	42	29	23	13	3	1	5	4	45	
v1cda		3	2	1	30	10		16	4	1	24	O. sp.
Myoviridae v0c0f	25			13	2	42	17			31	54	Ci. farmeri, Cronobacter sakazakii, E. coli, E. sp.
v1b3a	52			7	5	32	28	54		22	27	Ci. koseri, E. albertii, E. coli, E. spp.
v1cf7	58			16	1		7	43		19	33	E. albertii, E. coli, E. fergusonii, E. spp.
v1090	14	39		2	3	56	32	45		11	49	Ci. portucalensis, Ci. koseri, E. coli, E. spp.
Siphoviridae v0d6e	2		34				37					Enterococcus faecium
Microviridae v26b8	5	9										
Microviridae v05c9	4											Parasutterella sp.
v05af	3											
v04f1	11	43	45		4							Dialister invisus
Microviridae v04fd	1	11										

Fig. 4 Alterations in IBD-associated vOTUs in the external validation cohort. **a** Heatmap showing the fold changes in IBD-associated vOTUs in the VLP and bulk datasets of this study and external validation cohorts. Fold change > 0, enriched in patients; fold change < 0, enriched in controls. **b–c** Receiver operating characteristic (ROC) analysis of the classification of IBD/control status in each cohort using the random forest model trained by 139 IBD-associated vOTUs (**b**) and 34 top-ranking vOTUs (**c**). The classification performance of the model was assessed by the area under the ROC curve (AUC). **d** Heatmap showing the 34 most important vOTUs in the random forest models and their host information

dataset, including both the VLP and bulk metagenomic datasets from this study and the 9 public datasets. The models achieved an average intra-dataset area under the curve (AUC) of 0.857 (ranging from 0.642 to 0.936) in these datasets, with five datasets exhibiting an AUC of over 0.90 (Fig. 4b), supporting the generalizability and reliability of our markers across different datasets. Additionally, we developed RF models based on a subset of 33 IBD-associated vOTUs that integrated the top 5 most important features in the model for each dataset. The average intra-dataset AUC of these models was found to be high at 0.850 (ranging from 0.645 to 0.930), which was comparable to the performance of the RF models based on all markers (Fig. 4c). The 33 IBD-associated vOTUs consisted of 23 control-enriched vOTUs and 10 IBD-enriched vOTUs (Fig. 4d). Notably, some of the control-enriched vOTUs were predicted to infect butyrate-producing bacteria, such as *Agathobaculum* *butyriciproducens*, and *Faecalibacterium prausnitzii*. However, although three control-enriched vOTUs, v0c24 v0d26, and v1d95, were crucial to the prediction accuracy of most classification models, their taxonomic annotation and host preference remain completely unknown. On the other hand, the ten IBD-enriched vOTUs were found to frequently infect opportunistic pathogens, including *Citrobacter farmeri*, *Escherichia coli*, and *Escherichia fergusonii* (Fig. 4d).

We also performed cross-dataset prediction based on these 33 IBD-associated vOTUs (Supplementary Fig. 9). The cross-dataset RF models achieved an average AUC of 0.758 (ranging from 0.496 to 0.922). Considering the variation in sample sizes across datasets, we performed a leave-one-dataset-out (LODO) analysis, where each dataset was treated as a test dataset in turn. The LODO AUCs ranged from 0.599 to 0.919, with an average AUC of 0.799 (Supplementary Fig. 9). Collectively, these results demonstrate the robustness of the IBD-associated viral signatures identified in this study across diverse datasets, highlighting their potential as reliable targets for future microbiota intervention studies in IBD patients.

Colonization of IBD-associated viruses regulates experimental colitis in mice

Although IBD-associated gut viral signatures have been identified, experimental evidence regarding the influence of these viruses on IBD is limited. To address this, we conducted fecal microbiome transplantation (FMT) from either IBD patients or healthy subjects into broad spectrum antibiotic-treated mice to generate human microbiota-associated (HMA) mice, followed by high-dose FVT from patients and controls into these HMA mice and the administration of 2% DSS in their drinking water to induce experimental colitis (Fig. 5a). The donors for these experiments, including eight IBD patients (three CD patients and five UC patients) and seven healthy individuals, were randomly selected from our original cohort, and their fresh feces and fecal virus-like particles were pooled for the FMT and FVT procedures, respectively (see "Methods" for details). We then assessed the viral and bacterial composition of mouse feces before and after the FVT procedures using shotgun metagenomic sequencing. Low levels of viruses were detectable in the HMA mice before FVT, possibly reflecting the baseline proportion of viruses in their feces (Fig. 5b). Following FVT and DSS-induced colitis, we observed that the mice receiving the fecal virome from healthy donors (referred to as "HC-FVT mice") showed a significant expansion of control-enriched vOTUs compared to their pre-FVT state, with average relative abundances increasing from 2.0% to 52.4% (Student's t-test $p = 2.5 \times 10^{-8}$). Similarly, the mice that received the virome from IBD patients (referred to as "IBD-FVT mice") exhibited a substantial expansion of IBD-enriched vOTUs, with average relative abundances increasing from 2.3% to 44.9% ($p = 9.1 \times 10^{-8}$). We also estimated the virus-to-bacterium ratio of mouse feces and found that this ratio reached 51–1087 (average 366 in HC-FVT mice and 272 in IBD-FVT mice) in post-FVT samples (Supplementary Fig. 10), suggesting that the overgrowth of these viruses may have overwhelmed the existing bacterial content in the mouse gut microbiota. These findings demonstrate that the recipient mice efficiently recapitulated the viral features of the patient or control donors virome.

Next, we compared colitis-related symptoms in mice from the HC-FVT, IBD-FVT, and sham FVT groups. After DSS induction, the mice exhibited a colitis phenotype characterized by body weight loss, diarrhea, rectal bleeding, colon shortening, pathological changes, increased levels of inflammatory factors, and decreased expression of tight junction proteins, indicating that DSS induced colonic inflammation and impaired gut barrier function. Regardless of whether the initial FMT was from IBD patients or healthy donors, we found that compared with the sham group, the HC-FVT group exhibited reduced colitis-related symptoms, including a significant reduction in the disease activity index (DAI), an elongated colon length, and less severe pathological colon injury (Fig. 5c-e). In contrast, the IBD-FVT group displayed exacerbated colitis, with greater DAI scores, more severe pathological injuries, and shorter colons than both the HC-FVT and sham groups (Fig. 5c-e and Supplementary Fig. 11). Moreover, the expression levels of adhesion and tight junction-related proteins, such as E-cadherin, β-catenin, and occludin, were upregulated in the HF-FVT group and downregulated in the IBD-FVT group (Fig. 5f and Supplementary Fig. 12), indicating improved gut permeability in the HF-FVT group and worsened gut permeability in the IBD-FVT group. Additionally, the IBD-FVT group, irrespective of whether the HMA mice received microbiota from IBD patients or healthy donors, exhibited significantly greater levels of inflammatory cytokines, including TNF-α, IL-6, IFN-γ, and IL-1β, than the HC-FVT or sham groups (Fig. 5g). In contrast, the HC-FVT group exhibited reduced levels of certain cytokines such as IL-6, IFN-γ, and IL-1β in IBD HMA mice compared with those in the sham group (Fig. 5g). Taken together, these findings indicate that the colonization of viruses from healthy donors (HC-FVT) has a mitigating effect on colitis in a mouse model, whereas that of viruses from IBD patients (IBD-FVT) exacerbates colitis with a substantial expansion of IBD-enriched vOTUs in the mouse gut microbiota.

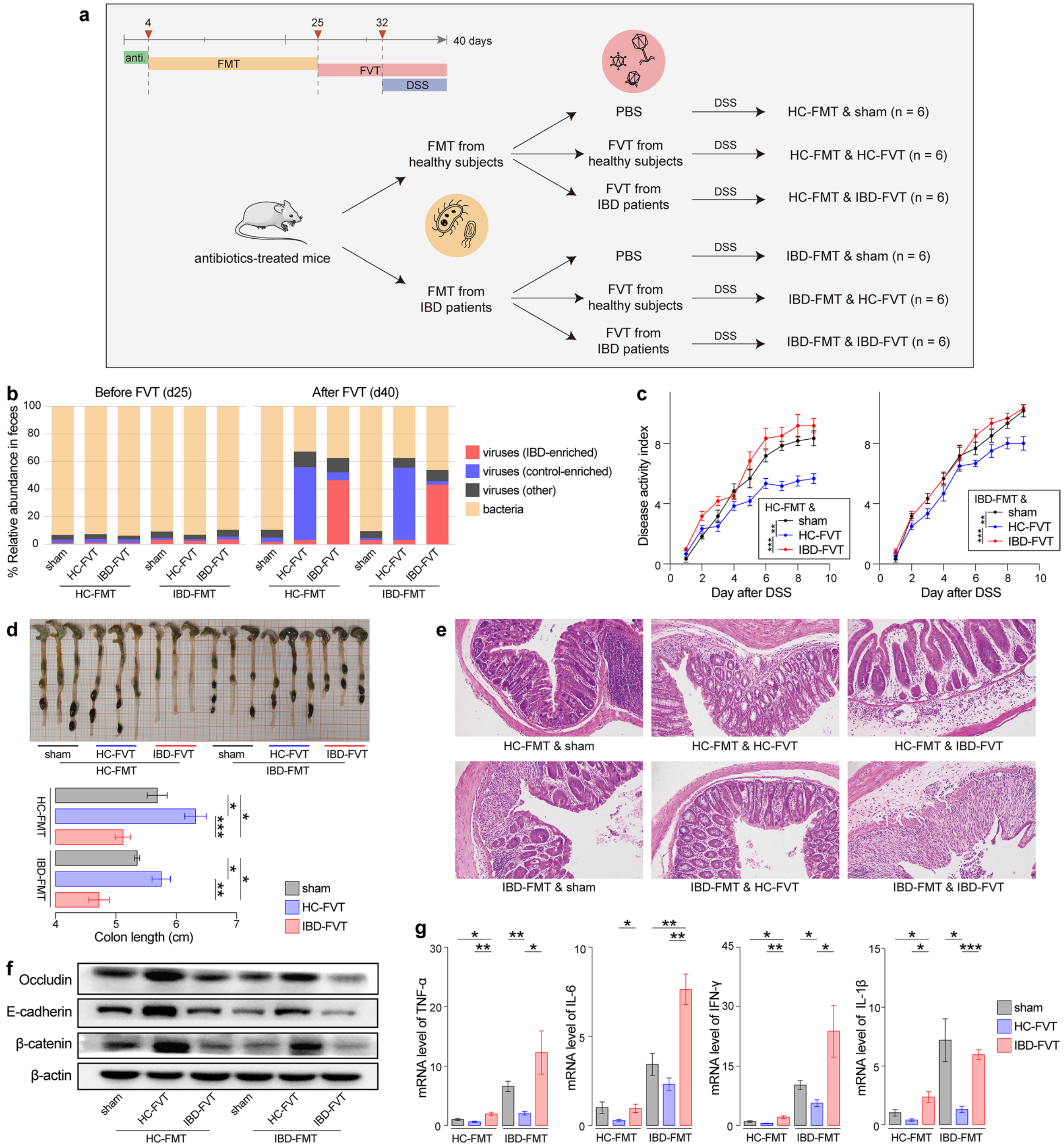


Fig. 5 Biological effect evaluation of the gut virome on DSS-induced colitis in mice. **a** Schematic diagram of the construction of the DSS-induced mouse model of colitis. **b** Bar plot showing the microbial composition of the feces of mice before and after FVT. **c** Evaluation of the disease activity index in colitis mice that received gut viromes from IBD patients and healthy controls. **d** Colon length of mice with colitis. **e** Representative images of hematoxylin and eosin (HE) staining of the mouse colon. **f** The expression levels of adhesion and tight junction-related proteins in the mice with colitis. **g** Levels of inflammatory cytokines in the colon of colitis mice. $N = 6$. Statistical tests were performed using Student's *t*-test: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

Discussion

Improved approaches for investigating the IBD gut virome

IBD is a severe gastrointestinal disorder that is increasingly attributed to dysbiosis of the gut microbiota, leading to immune dysfunction and the onset of the disease [37, 38]. While extensive research has identified the significant roles of gut bacteria and even fungi in IBD pathology, there has been a notable lack of information regarding the involvement of viruses. The limitations in studying the gut virome within the context of specific diseases can be attributed to two primary technical challenges. First, detection methods for viral communities have often been restricted to either VLP or bulk metagenomics, lacking comprehensive cross-validation. Second, despite the construction of several large viral genome reference databases, there remains a substantial reservoir of unexplored viral sequences within the human gut [16, 26]. This reliance on reference databases for gut virome analysis significantly restricts the scope of viral investigations. In this study, we sought to overcome these challenges by employing both VLP and bulk metagenomics sequencing methods in a cohort of individuals to conduct an extensive investigation into the "whole gut virome" associated with IBD. Notably, these two sequencing approaches demonstrated distinctions in their capacity to capture viruses. In alignment with prior assumptions [19, 30], VLP sequencing exhibited a preference for capturing free viral particles, particularly those from the *Microviridae* family, whereas bulk sequencing was more adept at reconstructing viral sequences that had partly integrated into bacterial hosts. Furthermore, we constructed a reference gut viral catalog that represented more than 10,000 nonredundant vOTUs based on our high-depth sequenced samples. The majority (61.2%) of vOTUs in this study were not present in existing gut virus catalogs (i.e., the Gut Virome Database (GVD) [30], Gut Phage Database (GPD) [39], or Metagenomic Gut Virus (MGV) [40]) (Supplementary Fig. 13), and almost all vOTUs had no homology with the viruses in the NCBI RefSeq database. These findings suggest that our catalog substantially supplements the deficiencies of previous viral references, allowing for a more comprehensive examination of the IBD gut virome. Overall, our results underscore the significance of utilizing complementary sequencing techniques and viral reference databases to fully explore the diversity of viruses within the human gut. This study presents a paradigm for future virome-wide studies of other relevant disorders.

Characteristics of the IBD gut virome and viral signatures

The results from both the VLP and bulk datasets consistently indicated an increase in eukaryotic virome richness and evenness in IBD patients compared to healthy controls. In particular, two eukaryotic viral families, *Retroviridae* and *Genomoviridae*, were found to be enriched in IBD patients. Although the overgrowth of the *Retroviridae* family in the gut has been implicated in several diseases including CD [26, 41], its specific influence on IBD remains unclear. Our analysis showed that VLP data supported the enrichment of *Anelloviridae* in IBD patients, whereas bulk data yielded the opposite result. Considering that at the vOTU level, the majority of *Anelloviridae* vOTUs in both VLP and bulk samples were enriched in patients, we believe that the enrichment of *Anelloviridae* in patients appears to be more credible. The enrichment of *Anelloviridae* was also previously observed in a study of onset CD [22], possibly related to long-term immune therapy [42, 43]. Notably, another study focusing on the gut mucosa virome found significantly greater levels of eukaryotic *Hepadnaviridae* transcripts in patients with UC than in both healthy controls and CD patients [20]. However, this phenomenon was not observed in our datasets.

In terms of the prokaryotic virome, the alpha diversity results for the VLP and bulk datasets revealed diametrically opposite findings. In the bulk dataset, we observed a significant decrease in prokaryotic virome diversity in IBD patients compared with the controls, and it exhibited a significant positive correlation with bacteriome diversity. Previous studies have extensively reported a substantial decrease in bacterial diversity in IBD patients [44–46]. These findings demonstrated a strong dependency of the bacterial community on the prokaryotic viral community in the bulk dataset. In contrast, the VLP dataset showed a significant increase in prokaryotic virome diversity in the IBD group, which was entirely unrelated to the bacteriome. Combining these findings with previous research suggesting that VLPs may preferentially capture free-floating viruses outside host cells, while bulk analysis captures a greater proportion of intracellular viral communities. These observations suggest a more frequent occurrence of prokaryotic cell lysis within the gut of IBD patients.

At the family level, we found that *Siphoviridae* and *Myoviridae* were significantly enriched in IBD patients, while *crAss-like* and *Quimbyviridae* were decreased. *Siphoviridae* and *Myoviridae* constitute the major viral communities in the human gut [1]; most members of these two families are temperate viruses, but their functions in the IBD virome remain unknown. Additionally, both *Siphoviridae* and *Myoviridae* belong to the order Caudovirales, and their enrichment was consistent with previous studies of the IBD virome [5, 47]. At the vOTU level, we also identified numerous IBD-enriched *Siphoviridae* and *Myoviridae* vOTUs. These vOTUs contained a large proportion of phages infecting bacteria such as *Escherichia*, *Klebsiella*, *Enterococcus_B*, *Streptococcus*, and *Veillonella*. *Escherichia*, *Klebsiella*, and *Enterococcus* spp. are typical opportunistic pathogens capable of inducing inflammation and triggering various diseases [48–50], and their pathogenicity in IBD has been recently reported [51, 52]. Overproliferation of *Escherichia* phages and other Enterobacteriaceae phages has also been observed in mucosal viromes of UC patients [53]. *Streptococcus* and *Veillonella* are commensal bacteria in the gut that have proinflammatory properties and are associated with diseases such as cirrhosis and obesity [54]. Although *Streptococcus* phages have been reported to be associated with metabolic and autoimmune diseases [55], their role in IBD has not been reported. In addition, our results showed that IBD-enriched vOTUs were more frequently associated with IBD-related pathogenic bacteria as their hosts, suggesting their involvement in the development of IBD through their prokaryotic hosts, as observed in immune diseases [10, 56]. Regarding IBD-depleted viruses, *crAss-like* and *Quimbyviridae* are newly identified viral clades that tend to infect Bacteroidetes, which are major polysaccharide degraders in the gut [57]. Similarly, at the vOTU level, numerous unclassified control-enriched vOTUs are predicted to infect Bacteroidetes, particularly *Prevotella* species, which are known to widely participate in plant polysaccharide utilization [58]. Studies have shown that these two clades of viruses, *crAss-like* and *Quimbyviridae*, have larger genomes and a substantial number of functional auxiliary genes, especially carbohydrate-active enzymes involved in polysaccharide metabolism [59, 60]. Therefore, the decrease in *crAss-like* and *Quimbyviridae* viruses and *Prevotella* phages may imply the loss of certain core viral functions, such as polysaccharide metabolism, in the gut virome of IBD patients.

Universality and disease relevance of the IBD-associated viral signatures

Recent meta-analyses of the gut virome in colorectal cancer patients have revealed distinctive viral signatures among diverse populations and suggested that their virome characteristics may operate independently from the bacteriome [8]. Similarly, by studying fecal samples from populations spanning Europe, the USA, Israel, and China, we found that despite population heterogeneity and differences in sample treatment methods, most of the IBD-associated viral signatures identified in our current study could be validated in other datasets. This indicates that our dataset and analyses effectively captured the universal signal characteristics in the IBD virome. The consistency of gut viral signatures in transcontinental datasets may not only be correlated with common changes in the bacterial microbiome but also suggest that some viruses can independently influence IBD in a consistent manner across geographic populations, potentially through mechanisms such as immunoregulation or viral infection [61]. Furthermore, we discovered that the classifying models with only 33 viral markers could effectively distinguish between the IBD and control groups in nearly all datasets (average AUC > 0.80), suggesting their potential diagnostic value for IBD and relevant diseases.

Previous animal experiments have confirmed that FVT from UC patients can exacerbate the severity of DSS-induced colitis [23]. Here, we performed a similar FVT experiment by transplanting VLPs from IBD patients and healthy controls into an HMA mouse model of experimental colitis. Surprisingly, we observed that IBD-associated viruses (both IBD-enriched and control-enriched viruses) could colonize the mice at very high levels after FVT, with an average relative abundance exceeding 50% and more than 200 times the number compared to bacteria. Accompanying the colonization of IBD-associated viruses, mice exhibited significant changes in colitis, pathological injury, and intestinal permeability. Isolation and direct functional investigation of these IBD-associated viruses may be needed to address causality between the gut virome and IBD progression in future research. Importantly, in contrast to previous studies focusing on the exacerbation of disease by IBD-FVT, our research found that HC-FVT effectively improved intestinal permeability and inflammatory markers in mice with colitis. This finding suggests the potential of HC-FVT as a therapeutic approach for the IBD population.

Furthermore, we also found that compared with mice with IBD-FMT, mice with HC-FMT exhibited an alleviated IBD phenotype. The improvement derived from the HC-FMT was comparable to that of the HC-FVT. It would be interesting to differentiate the contributions of FMT and FVT to IBD treatment. However, this was impossible in the present study because the mice received FMT and FVT simultaneously. The use of an antibiotic cocktail before the establishment of an HMA mouse model is another point that requires attention. Broad-spectrum antibiotic regimens can induce a 10,000-fold reduction in the gut bacterial load, leading to a significant increase in the diversity and relative abundance of the gut fungal microbiota, and a decrease in gut virome diversity [3, 62, 63]. Additional studies using germ-free mice or clinical trials may be needed to further clarify the role of FVT in IBD progression in the future.

In discussing the limitations of our study, it is important to acknowledge that while we made efforts to ensure age, sex, and BMI matching between the control and disease groups during sample collection, and established strict exclusion criteria such as excluding individuals with liver diseases, kidney diseases, cardiometabolic diseases, and cancers, there are potential confounding factors related to participants' income levels, dietary habits, exercise patterns, and lifestyles. These differences among subjects may have influenced the study outcomes. Therefore, we advocate that future research exercise greater caution in selecting control groups, striving to align demographic, socioeconomic, and lifestyle factors across different cohorts to enhance the reliability and applicability of the research findings. Additionally, the presence of potential impurities within VLPs is a limitation of this study. These impurities, including large glycoprotein molecules, lipoprotein molecules, outer membrane vesicles, bacterial fragments, fungal fragments, and even trace amounts of secondary metabolites and free nucleic acid, could influence our research outcomes. Subsequent research efforts aimed at further separating phages or eukaryotic viruses exerting effects will overcome this issue.

Conclusions

Through integrated VLP and bulk virome studies, we demonstrated a significant disruption in the gut virome of IBD patients and identified 139 IBD-associated vOTUs. Furthermore, we substantiated that these IBD viral signatures can be validated across diverse human population datasets and possess potential modulatory features in regulating disease severity in animal models. This study provides a detailed analysis of the gut virome landscape for IBD, and the results and resources provided may promote future mechanistic and therapeutic work.

Methods

Subjects and sample collection

Ethical approval for this study was obtained from the Ethics Committee of Xinhua Hospital affiliated with Dalian University (approval no. XH2020A008), and written informed consent was obtained from all participants. All subjects involved in this study were recruited from Dalian, Liaoning province, China, and were rigorously screened to exclude individuals with liver diseases, kidney diseases, cardiometabolic diseases (e.g., diabetes, moderate to severe hypertension), and cancers. Additionally, none of the participants had taken antibiotics within 1 month or antiviral drugs within 3 months prior to sample collection. The IBD patients were enrolled with inclusion criteria necessitating a clear diagnosis by a licensed physician in strict accordance with the ECCO-ESGAR Guideline for Diagnostic Assessment in Inflammatory Bowel Disease [64]. The Simplified Crohn's Disease Activity Index (sCDAI) integrates clinical symptoms, laboratory parameters, medication utilization, and the presence of perianal disease to offer a comprehensive assessment of disease severity. Healthy participants were matched with IBD patients in terms of age, sex, and BMI to ensure consistency in the results. More than 20 g of fresh fecal specimen per individual were collected from both IBD patients and healthy controls, using sterile fecal collection containers. Fecal specimens were collected and transported to the laboratory within 2 h. One portion of the samples was directly frozen at − 80 °C for subsequent NGS sequencing, while another portion was subjected to processing for VLPs enrichment. For the latter, a parallel set of samples was supplemented with glycerol as a protective agent before being frozen at − 80 °C. All samples for VLPs enrichment were stored frozen for less than 1 month before processing.

VLP enrichment and metagenomic sequencing

The procedures of VLP enrichment, viral DNA extraction, and metagenomic sequencing were conducted following our previously established protocols with minor adjustments [18, 26]. Briefly, 0.17 g of fecal material for each specimen was mixed with 1 mL of Hank's balanced salt solution (HBSS, devoid of phenol red) and vigorously homogenized using a vortex mixer (with pulses lasting a minimum of 15 s). Subsequently, the samples underwent centrifugation at 10,000 × *g* for 10 min at 4 ℃. The resulting supernatant was then successively filtered through 0.45-μm and 0.2-μm filters. Afterward, the samples were subjected to ultracentrifugation at 750,000 × *g* for 60 min at 8 ℃, and the resulting pellet was resuspended in 500 μL of HBSS; 120 μL of this resuspension was transferred and treated with a mixture of nucleases (comprising 2.4 μL of TURBO DNase (4.8 U, Invitrogen), 8 μL of RNase A/T1 Mix (16 μg RNase A, 40 U RNase T1, Thermo Scientific), and 1 μL of Benzonase (5 U, EMD Millipore)) for 120 min at 37 ℃. Following this, nucleic acids were immediately extracted using the TIANamp Viral Genome DNA/RNA extraction kit (TIANGEN, China), following the manufacturer's instructions. Subsequently, a DNA sequencing library was constructed utilizing the NEB Next Ultra DNA Library Prep Kit (NEB, USA) in accordance with the manufacturer's guidelines, with unique index codes assigned to each sample. The library's quality was confirmed using an Agilent 2100 instrument. Index-coded samples were then clustered using the Illumina PE Cluster Kit (Illumina, USA) on a cBot Cluster Generation System, following the manufacturer's protocols. Following cluster generation, the DNA libraries were subjected to metagenome shotgun sequencing on the Illumina NovaSeq platform, resulting in the generation of 150 bp paired-end reads.

Bulk metagenomic sequencing

For bulk metagenomic analyses, each fecal specimen from IBD patients or healthy controls was subjected to microbial DNA extraction using the TIANamp Stool DNA Kit (TIANGEN, China). DNA quality was assessed using the Qubit 2.0 system, and the extracted DNA samples were stored at − 80 ℃ until needed. After that, the DNA library and bulk metagenomic sequencing were consistent with the operations of the VLP samples.

Quality control and assembly of metagenomic datasets

Raw reads from the VLP and bulk metagenomic datasets were preprocessed using fastp v.0.23.2 [65] with adapter and polyG tail trimming, and low-quality reads were removed if the length was less than 60 bp or over more than 30% of them had an average Phred quality score of less than 20. The preprocessed reads were then aligned to the human genome (GRCh38) and the *Escherichia* phage phiX174 genome (NCBI accession NC_001422.1). Any reads that mapped to these genomes were excluded, and the remaining reads were recognized as clean reads for each sample. Next, the clean reads from each sample were subjected to contig assembly using MEGAHIT v1.2.967 [66] with the parameters: "–k-list 21,41,61,81,101,121,141".

Identification and clustering of viral sequences

We identified viral sequences using a well-established method described in our previous studies [18, 26]. Only contigs exceeding a length of 2000 bp were used for this analysis. Firstly, these contigs were assessed using CheckV v0.7.0 [28]. Those containing more than ten host genes, exceeding five times the number of viral genes, were excluded from further consideration. Among the remaining contigs, potential viral sequences were identified if they met any of the following criteria: (1) contigs with a higher count of viral genes in comparison to host genes, as determined by CheckV v0.7.0 [28]; (2) contigs with a score > 0.90 and a p-value < 0.01 in DeepVirFinder v1.0 [67]; and (3) contigs that were identified as viruses using default parameters in VIBRANT v1.2.1 [68]. To minimize non-viral sequence contamination, we conducted a search for bacterial universal single-copy orthologs (BUSCOs) within the potential viral sequences using hmmsearch [69]. We then calculated the BUSCO ratio, which measures the ratio of BUSCO counts to the total gene counts within each viral sequence, to assess potential contamination levels. Any sequence exhibiting ≥ 5% BUSCO ratio was then removed from the analysis. The remaining sequences, which possessed a CheckV-estimated completeness exceeding 50%, were assigned as the final viral genomes. To further improve the quality of the viral genomes and deduplicate genomes, we utilized a large gut viral genome database known as cnGVC [26]. In short, we clustered the viral genomes identified in our study with those in cnGVC based on criteria of 95% identity and 70% coverage. Only clusters containing at least one viral genome identified in our study were retained. Within each cluster, the longest sequence was selected as the reference sequence, which was referred to as the vOTU. In the end, we conducted a quality assessment of the genomes for all vOTUs through CheckV v0.7.0.

Taxonomic and functional annotation and host prediction for viral sequences

The taxonomic annotation of vOTUs was conducted by aligning their protein sequences against a comprehensive database. This database was constructed by integrating proteins from various sources, including Virus-Host DB (acquired in May 2021) [70], crAss-like proteins from Guerin's study [71], as well as viral proteins from Benler's and Ye's studies [71, 72]. To predict protein-coding sequences within vOTUs, we employed Prodigal v2.6.3 with the "meta" parameter [73]. Subsequently, we queried these protein sequences against the combined database using diamond v2.0.13.151 with the following parameters: "–id 30 –query-cover 50 –subject-cover 50 –min-score 50" [74]. For small vOTUs containing fewer than 30 genes, we assigned them to a known viral family if more than one-fifth of their proteins matched to a certain family. Conversely, for large vOTUs with 30 or more genes, we assigned them to a known viral family if at least 10 of their proteins matched to the same family.

To elucidate the functions of putative proteins, we used diamond v2.0.13.151 for searches against the KEGG (Kyoto Encyclopedia of Genes and Genomes) database [75], applying criteria of > 50% query coverage and > 60% score. For proteins that remained unassigned to a specific KEGG ortholog (KO), we conducted additional investigations by utilizing the PfamScan tool against Pfam database version 33.1 [34, 76].

We conducted host matching for vOTUs using two criteria within the pool of 4644 prokaryotic species from the Unified Human Gastrointestinal Genome (UHGG) database [31]. Firstly, for prokaryotic genomes, we predicted CRISPR spacer sequences using MinCED v0.4.2 with the parameter "-minNR 2" [77]. When a CRISPR spacer sequence from the host displayed a BLASTn match to a viral genome with a bit-score of 45 or greater, we assigned that virus to the corresponding host. Secondly, the virus was linked to a host if its sequence demonstrated ≥ 90% nucleotide identity and ≥ 30% viral coverage when compared to the host genome.

Taxonomic composition of metagenomic samples

In order to mitigate the issue of false positives, we have taken three steps to regenerate the relative abundance table of vOTUs. The specific steps for generating relative abundances of vOTUs for a metagenomic sample are as follows: (1) we used Bowtie 2 with the options "–end-to-end –fast" to align all clean reads to the 10,054 vOTUs and then utilized Samtools to obtain the sequencing depth of each base position for every vOTU (including positions with a sequencing depth of 0). For each vOTU, all base positions were sorted based on their sequencing depths from lowest to highest, excluding positions in the lowest and highest 10% of depth ranges. Subsequently, the mean depth of the remaining base positions was calculated as the sequencing depth for this vOTU. (2) Additionally, all clean reads were aligned to the 10,054 vOTUs using Kraken with option "–confidence 0.1". If no reads were specifically assigned to a particular vOTU, its sequencing depth obtained in the first step was set to 0. (3) Finally, the relative abundance of each vOTU was calculated by dividing its sequencing depth by the sum of the sequencing depths of all vOTUs. This calculation determined the relative abundance of each vOTU. For the family-level profiles, we aggregated the relative abundances of vOTUs that shared the same family-level annotation.

Statistical analysis

R language (version 4.2.3) was employed for all statistical tests and data visualization in this study.

Alpha diversity

We quantified the number of observed vOTUs per sample by counting those vOTUs that exhibited relative abundances not equal to 0. We derived the Shannon index from the vOTU-level relative abundance profile by applying the *diversity* function with the "index = shannon" parameter. Notably, in computing alpha diversity indexes for the eukaryotic or prokaryotic virome, we exclusively rely on the relative abundance profile of vOTUs assigned to the eukaryotic or prokaryotic families.

Beta diversity

The vOTU-level relative abundance profile was first subjected to square root transformation. Subsequently, Bray–Curtis distances between samples were computed based on the transformed data using the *vegdist* function. Using this distance matrix, we conducted a PERMANOVA using the *adonis* function. Similarly, based on the same distance matrix, PCoA was carried out using the *pcoa* function from the *ape* package. The PCoA plot was generated using the *ggplot* function. Within the plot, we applied the *stat_ellipse* function to include ellipses around each group's centroid, displaying an 80% confidence interval.

Statistical test

To compare the alpha diversity indexes and relative abundances of the viral community between groups, we conducted a Wilcoxon rank-sum test using the *wilcox.test* function. The resulting *p*-values were adjusted for multiple comparisons using the *p.adjust* function with the "method = B-H" parameter (Benjamini–Hochberg method). Spearman correlation analysis was conducted using the *cor.test* function with the "spearman" parameter to evaluate the association between the alpha diversity index of the gut virome and bacteriome. In the functional comparison between IBD-enriched and control-enriched vOTUs, we first calculated the occurrence ratio of each function within each group of viruses. This occurrence ratio is defined as the number of vOTUs possessing the corresponding functional gene divided by the total number of vOTUs within that specific group. Fisher's exact test was applied to compare the occurrence ratios of each function between IBD-enriched and control-enriched vOTUs using the *fisher.test* function.

Random forest model

We additionally downloaded data from nine cohorts, including eight bulk datasets and one VLP dataset. Random forest model was constructed separately for each of the 11 datasets, including our 2 datasets, to perform classification predictions for IBD patients and healthy controls. The model, along with five repeats of fivefold cross-validation, was built based on IBD-associated vOTUs using the *randomForest* function. The importance of IBD-associated vOTUs was assessed using the mean decrease in accuracy (MDA) index, which was obtained using the *importance* function. Subsequently, the importance rank of IBD-associated vOTUs was determined based on the average of MDA indexes derived from the results of five repeats of fivefold cross-validation. To assess the performance of the model, we calculated the AUC using the *roc* function from the *pROC* package. Additionally, we assessed the performance of the cross-cohort model for disease classification. This involved training the model on one dataset and validating it on another dataset. For the leave-one-dataset-out analysis (LODO), we built the model using a training dataset that consisted of data from ten cohorts and subsequently assessed its performance on the remaining cohort.

Animal experiments

All animal studies were approved by the Ethical Committee of Experimental Animal Care of Dalian Medical University (AEE19044).

Preparation of FVT materials and counting of VLPs

Three grams of pooled fecal samples were placed into a sterile 50-mL centrifuge tube. Thirty milliliters of SM buffer were added to create a fecal suspension. The suspension was left to stand on ice for 2 h with two intermittent shakes during this period. Afterward, centrifugation was performed at 5000 × *g* for 5 min at 4 °C. The supernatant was then transferred to a new sterile 50 mL-centrifuge tube and subjected to centrifugation at 12,000 × *g* for 15 min at 4 °C, and the supernatant was transferred to a new sterile 50-mL centrifuge tube. This step was repeated. The samples were then filtered through 0.45-μm and 0.22-μm filter membranes. The samples were then subjected to ultracentrifugation (750,000 × *g*, 60 min), followed by resuspension in saline solution. For counting the virus component prepared in the previous step, fixation was carried out using formalin. Subsequently, filtration was performed using a 0.02-μm filter (Whatman). Afterward, staining was done using SYBR Gold (Thermo Scientific) dye, and images were captured using an epifluorescence microscope for counting. Based on these results, the virus particle concentration for FVT was adjusted to approximately 1×10^9/mL. For animal experiments, 200 μL was administered by gavage each time per mouse.

DSS-induced colitis model

Healthy male C57BL/6 mice were maintained under SPF conditions for 1 week prior to the experiment. Mice received an antibiotics cocktail in the drinking water for 5 days (neomycin 1 mg/mL, streptomycin 1 mg/mL, and bacitracin 1 mg/mL) to deplete gut microbiota. Then, mice received FMT from either IBD patients or healthy controls for 21 days to generate HMA mice, followed by a high-dose FVT for 5 more days. After FMT and FVT, colitis was induced by 2% (w/v) DSS in the drinking water for 7 days, followed by 2 days of regular drinking water before sacrifice. Throughout the experiment, consistent daily dosing was maintained, and changes in the mice's body weight, diarrhea, and rectal bleeding were recorded. At the end of the experiment, the mice were anesthetized, and a colonoscopy was performed. Subsequently, the mice were euthanized, and the colon were collected and the colon length was measured. Approximately 1 cm segments of the colon were placed in formaldehyde fixation solution for histological analysis, while the remaining portions were placed in centrifuge tubes and stored at − 80 °C for subsequent experiments.

Microbial compositional analysis

Fecal specimens of mice were collected before and after the FVT procedures. Each fecal samples underwent for microbial DNA extraction, DNA library preparation, shotgun metagenomic sequencing, and data quality control using the aforementioned methods employed for bulk metagenomic sequencing of human feces. To determine the viral composition of the mouse fecal samples, sequencing reads were mapped to the vOTU catalog of this study using Bowtie 2, and the read count for each vOTU were generated. For each sample, the relative abundances of IBD-enriched and control-enriched vOTUs were calculated by summing the read count of each type of vOTUs and then dividing by the total number of reads of that sample. Similarly, the bacterial composition was determined by mapping the sequencing reads to the UHGG database. Due to the substantial difference in genome size between viruses and bacteria, directly comparing their relative abundances in samples cannot adequately reflect their differences. Therefore, we used the virus-to-bacterium ratio to reflect the difference of "cell count" between viruses and bacteria in each fecal sample. For each virus or bacterium, the number of mapped reads was divided by its genome length to obtain an estimate of "cell count". The sum of "cell counts" for all viruses was divided by the sum of "cell counts" for all bacteria to obtain the virus-to-bacterium ratio.

DAI scores

The disease activity index (DAI), including body weight loss, diarrhea, and rectal bleeding, was assigned scores on a 0–4 scale. Body weight: 0, body weight loss ≤ 0; 1, body weight loss ≤ 5; 2, body weight loss ≤ 10; 3, body weight loss ≤ 15; and 4, body weight loss > 15. Diarrhea: 0, normal stool; 1, soft but still formed stool; 2, soft stool; 3, very soft and wet stool; and 4, watery stool. Rectal bleeding: 0, no bleeding; 1, positive hemoccult; 2, visible blood traces in stool; 3, visible blood traces that adhered to the anus; and 4, gross bleeding.

Hematoxylin and eosin (HE) staining

Fixed colon segments were fixed in 10% buffered formaldehyde solution, embedded in paraffin, and sectioned. Sections were stained with HE for pathological lesion assessment. Pathological injures (crypt damage, inflammation, and ulceration) were assigned scores on a 0–3 scale by experienced and blinded researchers used a pathology scoring method as follows [78]. Inflammation was scored as follows: 0 = none; 1 = slight; 2 = moderate; and 3 = severe. Crypt damage was assigned scores as follows: 0, normal; 1, mild to moderate crypt loss (basal 1/3 damage); 2, severe crypt loss (basal 2/3 damage); and 3, complete crypt loss. Ulceration was assigned scores as follows: 0, normal; 1, destruction of mucosa; 2, destruction of musularis mucosa layer; and 3, destruction of submucosa.

qPCR

The levels of inflammatory cytokines (TNFα, IL1β, IL6, IFNγ) in the colon were determined by qPCR assays. Approximately 25 mg of colon tissue was placed in 500 μL of Trizol to obtain total RNA. A total of 1000 ng of total RNA was reverse-transcribed into cDNA using a reverse transcription kit according to the manufacturer's instructions. The primers used for qPCR were designed as follows: IL1β: Forward 5'-TGCCACCTTTTGACAGTG ATG-3', Reverse 5'-TGATGTGCTGCTGCGAGATT-3'; IL6: Forward 5'-GGGACTGATGCTGGTGACAA-3', Reverse

5’-ACAGGTCTGTTGGGAGTGGT-3’; TNFα: Forward 5’-TAGCCCACGTCGTAGCAAAC-3’, Reverse 5’-TGTCTTTGAGATCCATGCCGT-3’; INFγ: Forward 5’-GGAGGAACTGGCAAAAGGATG-3’, Reverse 5’-GTTGCTGATGGCCTGATTGT-3’; β-actin: Forward 5’-CACCATGTACCCAGGCATTG-3’, Reverse 5’-CCTGCTTGCTGATCCACATC-3’. Cycle threshold (Ct) values were recorded and normalized to the levels of internal reference β-actin.

Western blot

The protein levels of adhesion and tight junction-related proteins (occludin, E-cadherin, β-catenin) were determined by Western blot. Approximately 25 mg of colon tissue was mixed with cell lysis buffer containing PMSF and protease cocktail to obtain the total proteins. The protein concentration was determined using a BCA kit. Subsequently, protein samples were loaded onto the sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gel to undergo electrophoresis. Then, the interesting gel was transferred into PVDF membranes and incubated with primary antibody and secondary antibody sequentially. Finally, the bands were visualized using the Tanon 5200 ECL detection system (Tanon, China) and then the semiquantitative analyses were carried out using the Image J software.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-024-01832-x.

Additional File 1: Supplementary Figure 1. Comparison of gut viromes between VLP and bulk datasets. (a) Principal coordinate analysis (PCoA) reveals the difference between the VLP virome and bulk virome at the family level. Samples are shown at the first and second principal coordinates (PCoA1 and PCoA2), and the ratio of variance contributed by these two PCs is shown. (b) Observation of Microviridae vOTUs in the VLP and bulk datasets. Supplementary Figure 2. Correlation analysis (Spearman correlation test) of the gut bacteriome diversity and VLP/bulk virome diversity. For each panel, scatter plot shows the samples and the fitting line are formed based on the diversity indexes in the bacteriome and virome for all samples. Supplementary Figure 3. Comparison of Anelloviridae vOTUs between IBD patients and healthy controls. Barplot showing the number of vOTUs enriched in IBD patients or healthy controls. Dark red or blue indicate vOTUs with a p-value of less than 0.05 from the Wilcoxon rank-sum test comparing the two groups. Supplementary Figure 4. Comparison of gut viral composition between IBD patients and healthy controls at the order level. Wilcoxon rank-sum test: *, p<0.05; **, p<0.01; ***, p<0.001. HC, healthy controls. Supplementary Figure 5. Distribution of prokaryotic hosts of the IBD-associated vOTUs. Supplementary Figure 6. Comparison of functions between IBD-enriched and control-enriched vOTUs. Principal coordinate analysis (PCoA) reveals the differences in functional profiles between IBD-enriched and controls vOTUs. vOTUs are shown at the first and second principal coordinates (PCoA1 and PCoA2), and the ratio of variance contributed by these two PCs is shown. Statistical significance was obtained by PERMANOVA analysis. Supplementary Figure 7. Partial gene structures of several control-enriched vOTUs with the RNA-dependent DNA polymerase (RdDp) gene. The left-hand side text provides the family-level taxonomic annotation and vOTU ID number corresponding to each vOTU, along with the predicted prokaryotic host for each one. Supplementary Figure 8. Viral community variation among 9 public datasets and the VLP and bulk datasets in this study. Principal coordinates analysis (PCoA) based on the Bray-Curtis distance at the vOTU level. Effect size (R2) and statistical significance were obtained by PERMANOVA analysis. Cross-dataset IBD prediction based on 34 IBD-associated vOTUs. Supplementary Figure 9. Heatmap shows the performance assessed as AUC scores of intra-dataset and cross-dataset IBD predictions using random forest models trained based on 34 IBD-associated vOTUs. The models of intra-dataset prediction (diagonal) are validated using five repeats of five-fold cross-validations. The models of cross-dataset prediction (off-diagonal) are built on the dataset corresponding to each row and validated on the dataset corresponding to each column. The LOCO row refers to leave-one-cohort-out analysis in which models are built on ten datasets combined and validated on the remaining one corresponding to each column. Supplementary Figure 10. Bar plot showing the virus-to-bacterium ratio of feces of mouse before and after FVT. Supplementary Figure 11. Colitis scores of HE staining of the colon in Figure 5e. Supplementary Figure 12. Quantitative analysis of the western blot of Figure 5f. Supplementary Figure 13. Venn plot showing the overlap between our gut virus catalog and the existing gut virus catalogs.

Additional File 2: Table S1. Phenotypic characteristics of 71 IBD patients and 77 healthy controls recruited in this study. Table S2. Detailed information of 10,054 viral operational taxonomic units (vOTUs). Table S3. Comparison of the gut virome between IBD patients and healthy controls at the family level. Table S4. Detailed information of 139 IBD-associated vOTUs. Table S5. External validation cohorts for this study.

Acknowledgements

Not applicable.

Authors’ contributions

X.M., Q.Y., and S.L. conceived the study. X.T., C.W., Q.Y. Y.Z., X.F., F.W., C.W., X.F., and X.H. recruited the subjects and collected the fecal samples. S.L. and R.G. developed the analytical methods and performed the gut virome analyses. Y.Z., F.W., C.W., Q.Y., and X.H. performed animal experiments. S.L., X.M., X.H., and Q.Y. drafted the manuscript. All authors revised the article and approved the final version for publication.

Funding

The authors thank the National Natural Science Foundation of China (No. 82225048 and 82370563), distinguished professor of Liaoning Province (XLYC2002008) and Dalian Science and Technology Leading Talents Project (2019RD15 and 2022RY18), Dalian Science and Technology benefit people project (2023JJ13SN053).

Availability of data and materials

The metagenomic sequencing data of human feces have been deposited in the European Nucleotide Archive (ENA) under the accession IDs PRJEB67456 (https://www.ebi.ac.uk/ena/browser/view/PRJEB67456). The analysis and visualization codes used in this study have been uploaded into the GitHub repository, accessible at: https://github.com/lish2/ibd_virome.

Declarations

Ethics approval and consent to participate

Ethical approval for this study was obtained from the Ethics Committee of Xinhua Hospital affiliated with Dalian University (approval No. XH2020A008), and informed consent was obtained from all participants.
All animal studies were approved by the Ethical Committee of Experimental Animal Care of Dalian Medical University (AEE19044).

Consent for publication

All authors have reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Received: 31 October 2023 Accepted: 8 May 2024
Published online: 18 July 2024

References

1. Zuo T, Sun Y, Wan Y, Yeoh YK, Zhang F, Cheung CP, Chen N, Luo J, Wang W, Sung JJY, Chan PKS, Wang K, Chan FKL, Miao Y, Ng SC. Human-gut-DNA virome variations across geography, ethnicity, and urbanization. Cell Host Microbe. 2020;28:741–51.
2. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD. The human gut virome: inter-individual variation and dynamic response to diet. Genome Res. 2011;21:1616–25.
3. Wang L, Yao H, Morgan DC, Lau KS, Leung SY, Ho JWK, Leung WK. Altered human gut virome in patients undergoing antibiotics therapy for Helicobacter pylori. Nat Commun. 2023;14:2196.
4. Cao Z, Sugimura N, Burgermeister E, Ebert MP, Zuo T, Lan P. The gut virome: a new microbiome component in health and disease. EBioMedicine. 2022;81: 104113.
5. Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, Kambal A, Monaco CL, Zhao G, Fleshner P, Stappenbeck TS, McGovern DPB, Keshavarzian A, Mutlu EA, Sauk J, Gevers D, Xavier RJ, Wang D, Parkes M, Virgin HW. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell. 2015;160:447–60.
6. Yang K, Niu J, Zuo T, Sun Y, Xu Z, Tang W, Liu Q, Zhang J, Ng EKW, Wong SKH, Yeoh YK, Chan PKS, Chan FKL, Miao Y, Ng SC. Alterations in the gut virome in obesity and type 2 diabetes Mellitus. Gastroenterology. 2021;161:1257–69.
7. de Jonge PA, Wortelboer K, Scheithauer TPM, van den Born B-JH, Zwinderman AH, Nobrega FL, Dutilh BE, Nieuwdorp M, Herrema H. Gut virome profiling identifies a widespread bacteriophage family associated with metabolic syndrome. Nat Commun. 2022;13:3594.
8. Chen F, Li S, Guo R, Song F, Zhang Y, Wang X, Huo X, Lv Q, Ullah H, Wang G, Ma Y, Yan Q, Ma X. Meta-analysis of fecal viromes demonstrates high diagnostic potential of the gut viral signatures for colorectal cancer and adenoma risk assessment. J Adv Res. 2023;49:103–14.
9. Tomofuji Y, Kishikawa T, Maeda Y, Ogawa K, Nii T, Okuno T, Oguro-Igashira E, Kinoshita M, Yamamoto K, Sonehara K, Yagita M, Hosokawa A, Motooka D, Matsumoto Y, Matsuoka H, Yoshimura M, Ohshima S, Nakamura S, Inohara H, Mochizuki H, Takeda K, Kumanogoh A, Okada Y. Whole gut virome analysis of 476 Japanese revealed a link between phage and autoimmune disease. Ann Rheum Dis. 2022;81:278–88.
10. Li C, Zhang Y, Yan Q, Guo R, Chen C, Li S, Zhang Y, Meng J, Ma J, You W, Wu Z, Sun W. Alterations in the gut virome in patients with ankylosing spondylitis. Front Immunol. 2023;14:1154380.
11. Guo R, Li S, Zhang Y, Zhang Y, Wang G, Ullah H, Ma Y, Yan Q. Dysbiotic oral and gut viromes in untreated and treated rheumatoid arthritis patients. Microbiology Spectrum. 2022;10: e0034822.
12. Hsu CL, Duan Y, Fouts DE, Schnabl B. Intestinal virome and therapeutic potential of bacteriophages in liver disease. J Hepatol. 2021;75:1465–75.
13. Lang S, Demir M, Martin A, Jiang L, Zhang X, Duan Y, Gao B, Wisplinghoff H, Kasper P, Roderburg C, Tacke F, Steffen H-M, Goeser T, Abraldes JG, Tu XM, Loomba R, Starkel P, Pride D, Fouts DE, Schnabl B. Intestinal virome signature associated with severity of nonalcoholic fatty liver disease. Gastroenterology. 2020;159:1839–52.
14. Foca A, Liberto MC, Quirino A, Marascio N, Zicca E, Pavia G. Gut inflammation and immunity: what is the role of the human gut virome? Mediators Inflamm. 2015;2015: 326032.
15. Wang W-L, Xu S-Y, Ren Z-G, Tao L, Jiang J-W, Zheng S-S. Application of metagenomics in the human gut microbiome. World J Gastroenterol. 2015;21:803–14.
16. Li J, Yang F, Xiao M, Li A. Advances and challenges in cataloging the human gut virome. Cell Host Microbe. 2022;30:908–16.
17. Kleiner M, Hooper LV, Duerkop BA. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. BMC Genomics. 2015;16:7.
18. Yan Q, Wang Y, Chen X, Jin H, Wang G, Guan K, Zhang Y, Zhang P, Ayaz T, Liang Y, Wang J, Cui G, Sun Y, Xiao M, Kang J, Zhang W, Zhang A, Li P, Liu X, Ulllah H, Ma Y, Li S, Ma T. Characterization of the gut DNA and RNA viromes in a cohort of Chinese residents and visiting Pakistanis. Virus Evolution. 2021;7:veab022.
19. Clooney AG, Sutton TDS, Shkoporov AN, Holohan RK, Daly KM, O'Regan O, Ryan FJ, Draper LA, Plevy SE, Ross RP, Hill C. Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. Cell Host Microbe. 2019;26:764–78.
20. Ungaro F, Massimino L, Furfaro F, Rimoldi V, Peyrin-Biroulet L, D'Alessio S, Danese S. Metagenomic analysis of intestinal mucosa revealed a specific eukaryotic gut virome signature in early-diagnosed inflammatory bowel disease. Gut Microbes. 2019;10:149–58.
21. Fernandes MA, Verstraete SG, Phan TG, Deng X, Stekol E, LaMere B, Lynch SV, Heyman MB, Delwart E. Enteric virome and bacterial microbiota in children with ulcerative colitis and Crohn disease. J Pediatr Gastroenterol Nutr. 2019;68:30–6.
22. Tokarz R, Hyams JS, Mack DR, Boyle B, Griffiths AM, LeLeiko NS, Sauer CG, Shah S, Markowitz J, Baker SS, Rosh J, Baldassano RN, Kugathasan S, Walters T, Tagliafierro T, Sameroff S, Lee B, Che X, Oleynik A, Denson LA, Lipkin WI. Characterization of stool virome in children newly diagnosed with moderate to severe ulcerative colitis. Inflamm Bowel Dis. 2019;25:1656–62.
23. Sinha A, Li Y, Mirzaei MK, Shamash M, Samadfam R, King IL, Maurice CF. Transplantation of bacteriophages from ulcerative colitis patients shifts the gut bacteriome and exacerbates the severity of DSS colitis. Microbiome. 2022;10:105.
24. Adiliaghdam F, Amatullah H, Digumarthi S, Saunders TL, Rahman R-U, Wong LP, Sadreyev R, Droit L, Paquette J, Goyette P, Rioux JD, Hodin R, Mihindukulasuriya KA, Handley SA, Jeffrey KL. Human enteric viruses autonomously shape inflammatory bowel disease phenotype through divergent innate immunomodulation. Science Immunology. 2022;7:eabn6660.
25. Thia K, Faubion WA Jr, Loftus EV Jr, Persson T, Persson A, Sandborn WJ. Short CDAI: development and validation of a shortened and simplified Crohn's disease activity index. Inflamm Bowel Dis. 2011;17:105–11.
26. Li S, Yan Q, Zhang Y, Guo R, Zhang P, Lv Q, Chen F, Li Z, Meng J, Li J, Wang G, Chen C, Ullah H, Cheng L, Fan S, Li R, You W, Zhang Y, Ma J, Sun W, Ma X. Cataloguing and profiling of the gut virome in Chinese populations uncover extensive viral signatures across common diseases. bioRxiv. 2022;2012:2027–522048.
27. Li s, Guo R, Zhang Y, Li P, Chen F, Wang X, Li J, Jie Z, Lv Q, Jin H, Wang G, Yan Q. A catalog of 48,425 nonredundant viruses from oral metagenomes expands the horizon of the human oral virome. iScience. 2022;25:104418.
28. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021;39:578–85.
29. Chen C, Yan Q, Yao X, Li S, Lv Q, Wang G, Zhong Q, Tang F, Liu Z, Huang Y, An Y, Zhou J, Zhang Q, Zhang A, Ullah H, Zhang Y, Liu C, Zhu D, Li H, Sun W, Ma W. Alterations of the gut virome in patients with systemic lupus erythematosus. Front Immunol. 2023;13:1050895.
30. Gregory AC, Zablocki O, Zayed AA, Howell A, Bolduc B, Sullivan MB. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. Cell Host Microbe. 2020;28:724–40.
31. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, Pollard KS, Sakharova E, Parks DH, Hugenholtz P, Segata N, Kyrpides NC, Finn RD. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat Biotechnol. 2021;39:105–14.
32. Ruigrok RAAA, Collij V, Sureda P, Klaassen MAY, Bolte LA, Jansen BH, Voskuil MD, Fu J, Wijmenga C, Zhernakova A, Weersma RK, Vila AV. The composition and metabolic potential of the human small intestinal microbiota within the context of inflammatory bowel disease. J Crohns & Colitis. 2021;15:1326–38.
33. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45:D353–61.
34. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. Pfam: the protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9.
35. Millman A, Bernheim A, Stokar-Avihail A, Fedorenko T, Voichek M, Leavitt A, Oppenheimer-Shaanan Y, Sorek R. Bacterial retrons function in anti-phage defense. Cell. 2020;183:1551–61.
36. Toro N, Nisa-Martinez R. Comprehensive phylogenetic analysis of bacterial reverse transcriptases. PLoS ONE. 2014;9: e114083.
37. Ni J, Wu GD, Albenberg L, Tomov VT. Gut microbiota and IBD: causation or correlation? Nat Rev Gastroenterol Hepatol. 2017;14:573–84.
38. Sultan S, El-Mowafy M, Elgaml A, Ahmed TAE, Hassan H, Mottawea W. Metabolic influences of gut microbiota dysbiosis on inflammatory bowel disease. Front Physiol. 2021;12: 715506.

39. Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. Massive expansion of human gut bacteriophage diversity. Cell. 2021;184:1098–109.
40. Nayfach S, Paez-Espino D, Call L, Low SJ, Sberro H, Ivanova NN, Proal AD, Fischbach MA, Bhatt AS, Hugenholtz P, Kyrpides NC. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. Nat Microbiol. 2021;6:960–70.
41. Perez-Brocal V, Garcia-Lopez R, Vazquez-Castellanos JF, Nos P, Beltran B, Latorre A, Moya A. Study of the viral and microbial communities associated with Crohn's disease: a metagenomic approach. Clin Transl Gastroenterol. 2013;4: e36.
42. Monaco CL, Gootenberg DB, Zhao G, Handley SA, Ghebremichael MS, Lim ES, Lankowski A, Baldridge MT, Wilen CB, Flagg M, Norman JM, Keller BC, Luevano JM, Wang D, Boum Y, Martin JN, Hunt PW, Bangsberg DR, Siedner MJ, Kwon DS, Virgin HW. Altered virome and bacterial microbiome in human immunodeficiency virus-associated acquired immunodeficiency syndrome. Cell Host Microbe. 2016;19:311–22.
43. De Vlaminck I, Khush KK, Strehl C, Kohli B, Luikart H, Neff NF, Okamoto J, Snyder TM, Cornfield DN, Nicolls MR, Weill D, Bernstein D, Valantine HA, Quake SR. Temporal response of the human virome to immunosuppression and antiviral therapy. Cell. 2013;155:1178–87.
44. Scher JU, Ubeda C, Artacho A, Attur M, Isaac S, Reddy SM, Marmon S, Neimann A, Brusca S, Patel T, Manasson J, Pamer EG, Littman DR, Abramson SB. Decreased bacterial diversity characterizes the altered gut microbiota in patients with psoriatic arthritis, resembling dysbiosis in inflammatory bowel disease. Arthritis & Rheumatology. 2015;67:128–39.
45. Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, Nalin R, Jarrin C, Chardon P, Marteau P, Roca J, Dore J. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut. 2006;55:205–11.
46. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13:R79.
47. Liang G, Conrad MA, Kelsen JR, Kessler LR, Breton J, Albenberg LG, Marakos S, Galgano A, Devas N, Erlichman J, Zhang H, Mattei L, Bittinger K, Baldassano RN, Bushman FD. Dynamics of the stool virome in very early-onset inflammatory bowel disease. J Crohns & Colitis. 2020;14:1600–10.
48. Fine RL, Vieira SM, Gilmore MS, Kriegel MA. Mechanisms and consequences of gut commensal translocation in chronic diseases. Gut Microbes. 2020;11:217–30.
49. Paczosa MK, Mecsas J. *Klebsiella pneumoniae*: going on the offense with a strong defense. Microbiol Mol Biol Rev. 2016;80:629–61.
50. Jiang W, Wu N, Wang X, Chi Y, Zhang Y, Qiu X, Hu Y, Li J, Liu Y. Dysbiosis gut microbiota associated with inflammation and impaired mucosal immune function in intestine of humans with non-alcoholic fatty liver disease. Sci Rep. 2015;5:8096.
51. Palmela C, Chevarin C, Xu Z, Torres J, Sevrin G, Hirten R, Barnich N, Ng SC, Colombel J-F. Adherent-invasive *Escherichia coli* in inflammatory bowel disease. Gut. 2018;67:574–87.
52. Federici S, Kredo-Russo S, Valdes-Mas R, Kviatcovsky D, Weinstock E, Matiuhin Y, Silberberg Y, Atarashi K, Furuichi M, Oka A, Liu B, Fibelman M, Weiner IN, Khabra E, Cullin N, Ben-Yishai N, Inbar D, Ben-David H, Nicenboim J, Kowalsman N, Lieb W, Kario E, Cohen T, Geffen YF, Zelcbuch L, Cohen A, Rappo U, Gahali-Sass I, Golembo M, Lev V, Dori-Bachash M, Shapiro H, Moresi C, Cuevas-Sierra A, Mohapatra G, Kern L, Zheng D, Nobs SP, Suez J, Stettner N, Harmelin A, Zak N, Puttagunta S, Bassan M, Honda K, Sokol H, Bang C, Franke A, Schramm C, Maharshak N, Sartor RB, Sorek R, Elinav E. Targeted suppression of human IBD-associated gut microbiota commensals by phage consortia for treatment of intestinal inflammation. Cell. 2022;185:2879–98.
53. Zuo T, Lu X-J, Zhang Y, Cheung CP, Lam S, Zhang F, Tang W, Ching JYL, Zhao R, Chan PKS, Sung JJY, Yu J, Chan FKL, Cao Q, Sheng J-Q, Ng SC. Gut mucosal virome alterations in ulcerative colitis. Gut. 2019;68:1169–79.
54. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L, Zhou J, Ni S, Liu L, Pons N, Batto JM, Kennedy SP, Leonard P, Yuan C, Ding W, Chen Y, Hu X, Zheng B, Qian G, Xu W, Ehrlich SD, Zheng S, Li L. Alterations of the human gut microbiome in liver cirrhosis. Nature. 2014;513:59–64.
55. Fujimoto K, Miyaoka D, Uematsu S. Characterization of the human gut virome in metabolic and autoimmune diseases. Inflamm and Regen. 2022;42:32.
56. Chen C, Zhang Y, Yao X, Li S, Wang G, Huang Y, Yang Y, Zhang A, Liu C, Zhu D, Li H, Yan Q, Ma W. Characterizations of the gut bacteriome, mycobiome, and virome in patients with osteoarthritis. Microbiology Spectrum. 2023;11:e01711–22.
57. Lapebie P, Lombard V, Drula E, Terrapon N, Henrissat B. Bacteroidetes use thousands of enzyme combinations to break down glycans. Nat Commun. 2019;10:2043.
58. Fehlner-Peach H, Magnabosco C, Raghavan V, Scher JU, Tett A, Cox LM, Gottsegen C, Watters A, Wiltshire-Gordon JD, Segata N, Bonneau R, Littman DR. Distinct polysaccharide utilization profiles of human intestinal *Prevotella copri* isolates. Cell Host Microbe. 2019;26:680–90.
59. Shkoporov AN, Khokhlova EV, Stephens N, Hueston C, Seymour S, Hryckowian AJ, Scholz D, Ross RP, Hill C. Long-term persistence of crAss-like phage crAss001 is associated with phase variation in *Bacteroides intestinalis*. BMC Biol. 2021;19:163.
60. Yutin N, Benler S, Shmakov SA, Wolf YI, Tolstoy I, Rayko M, Antipov D, Pevzner PA, Koonin EV. Analysis of metagenome-assembled viral genomes from the human gut reveals diverse putative CrAss-like phages with unique genomic features. Nat Commun. 2021;12:1044.
61. Foxman EF, Iwasaki A. Genome-virome interactions: examining the role of common viral infections in complex disease. Nat Rev Microbiol. 2011;9:254–64.
62. Seelbinder B, Chen J, Brunke S, Vazquez-Uribe R, Santhaman R, Meyer A-C, de Oliveira Lino FS, Chan K-F, Loos D, Imamovic L, Tsang C-C, Lam RP-K, Sridhar S, Kang K, Hube B, Woo PC-Y, Sommer MOA, Panagiotou G. Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria. Microbiome. 2020;8:133.
63. Hagan T, Cortese M, Rouphael N, Boudreau C, Linde C, Maddur MS, Das J, Wang H, Guthmiller J, Zheng N-Y, Huang M, Uphadhyay AA, Gardinassi L, Petitdemange C, McCullough MP, Johnson SJ, Gill K, Cervasi B, Zou J, Bretin A, Hahn M, Gewirtz AT, Bosinger SE, Wilson PC, Li S, Alter G, Khurana S, Golding H, Pulendran B. Antibiotics-driven gut microbiome perturbation alters immunity to vaccines in humans. Cell. 2019;178:1313-1328.e1313.
64. Maaser C, Sturm A, Vavricka SR, Kucharzik T, Fiorino G, Annese V, Calabrese E, Baumgart DC, Bettenworth D, Nunes PB, Burisch J, Castiglione F, Eliakim R, Ellul P, Gonzalez-Lama Y, Gordon H, Halligan S, Katsanos K, Kopylov U, Kotze PG, Krustins E, Laghi A, Limdi JK, Rieder F, Rimola J, Taylor SA, Tolan D, van Rheenen P, Verstockt B, Stoker J. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: initial diagnosis, monitoring of known IBD, detection of complications. J Crohns & Colitis. 2019;13:144–64.
65. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:884–90.
66. 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;31:1674–6.
67. Ren J, Song K, Deng C, Ahlgren NA, Fuhrman JA, Li Y, Xie X, Poplin R, Sun F. Identifying viruses from metagenomic data using deep learning. Quantitative biology (Beijing, China). 2020;8:64–77.
68. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome. 2020;8:90.
69. Eddy SR. Accelerated profile HMM searches. PLoS Comp Biol. 2011;7: e1002195.
70. Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, Hingamp P, Goto S, Ogata H. Linking virus genomes with host taxonomy. Viruses-Basel. 2016;8:66.
71. Guerin E, Shkoporov A, Stockdale SR, Clooney AG, Ryan FJ, Sutton TDS, Draper LA, Gonzalez-Tortuero E, Ross RP, Hill C. Biology and taxonomy of crass-like bacteriophages, the most abundant virus in the human gut. Cell Host Microbe. 2018;24:653–64.
72. Ye S, Lu C, Qiu Y, Zheng H, Ge X, Wu A, Xia Z, Jiang T, Zhu H, Peng Y. An atlas of human viruses provides new insights into diversity and tissue tropism of human viruses. Bioinformatics. 2022;38:3087–93.
73. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
74. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60.
75. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457–62.

76. Mistry J, Bateman A, Finn RD. Predicting active site residue annotations in the Pfam database. BMC Bioinformatics. 2007;8:298.
77. Skennerton C. Minced—mining CRISPRs in environmental datasets. ***In: github***. 2016:http://github.com/ctSkennerton/minced.
78. Huo X, Li D, Wu F, Li S, Qiao Y, Wang C, Wang Y, Zhou C, Sun L, Luan Z, Yan Q, Wang J, Zhang Y, Zhao T, An Y, Zhang B, Tian X, Yu Z, Ma X. Cultivated human intestinal fungus Candida metapsilosis M2006B attenuates colitis by secreting acyclic sesquiterpenoids as FXR agonists. Gut. 2022;71:2205–17.

Publisher's Note