

RESEARCH

Open Access



Characterizing the gut virome in ulcerative colitis and crohn's disease: signatures of disease severity

Nasser Ebrahimi Daryani¹, Seyed Mohammad Jazayeri², Nasim Izadi³, Hadi Ahmadi³, Hossein Bannazadeh Baghi³, Masoud Shirmohammadi⁴, Mohammad Sabbaghian³, Ali Akbar Shekarchi⁵, Shima Sadeghipour Marvi², Arezou Azadi³ and Vahdat Poortahmasebi^{3*}

Abstract

Background Inflammatory bowel disease (IBD), including ulcerative colitis (UC) and Crohn's disease (CD), is a chronic disorder marked by intestinal inflammation and immune dysregulation. While bacterial dysbiosis has been widely investigated, the gut virome remains less explored. Altered viral communities, particularly bacteriophages, may destabilize microbial balance and amplify host inflammation.

Methods To characterize virome alterations, we conducted a cross-sectional observational study in Tabriz, Iran, involving fifty participants divided into five groups: mild UC, severe UC, mild CD, severe CD, and healthy controls. Stool samples were processed for viral nucleic acid extraction and analyzed using metagenomic next-generation sequencing. Bioinformatics pipelines included diversity assessment, taxonomic profiling, functional annotation, and discriminant analysis (LEfSe). Predictive modeling was performed with random forest classifiers.

Results Virome richness and diversity were reduced in severe UC and CD compared with controls, whereas mild cases showed values closer to healthy individuals. Taxonomic profiling revealed depletion of crAss-like phages and microviridae in IBD, along with enrichment of Caudovirales families such as siphoviridae and myoviridae. Among eukaryotic viruses, anelloviridae were prominent in severe IBD, and herpesviridae were enriched specifically in severe UC. Functional annotation highlighted enrichment of structural and lytic phage proteins in severe groups, whereas lysogeny-associated domains were more abundant in healthy controls. Random forest models based on viral features achieved appropriate accuracy, with an AUC of 0.89 for distinguishing IBD from controls and 0.83 for classifying mild versus severe disease.

Conclusion Thus, IBD is associated with reduced virome diversity, loss of core protective phages, and selective enrichment of bacteriophages and eukaryotic viruses. These findings suggest that virome features may have potential as biomarkers for non-invasive diagnosis and severity stratification in IBD, requiring validation in larger and longitudinal cohorts.

Keywords Gut virome, Inflammatory bowel disease, Ulcerative colitis, Crohn's disease, Metagenomics

*Correspondence:
Vahdat Poortahmasebi
vahdat2011@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Inflammatory bowel disease (IBD), which includes ulcerative colitis (UC) and Crohn's disease (CD), is a long-term disorder of the gastrointestinal tract that presents with recurrent episodes of inflammation and immune imbalance [1, 2]. Its prevalence has been rising worldwide, particularly in newly industrialized countries, suggesting that environmental and lifestyle changes play a major role alongside genetic predisposition [3]. Despite intensive research, the etiology of IBD remains incompletely understood. Increasing attention has been directed toward the intestinal microbiome, which is now recognized as a central element in disease onset and progression [4]. While bacteria have been the primary focus of most investigations, the gut virome, mainly composed of bacteriophages and a smaller fraction of eukaryotic viruses, has emerged as a potentially important but underexplored component of intestinal ecology [5, 6].

Bacteriophages, through mechanisms such as predation, lysogeny, and horizontal gene transfer, influence bacterial communities and affect microbial structure and function [7]. Perturbations of viral communities may destabilize bacterial populations and indirectly modulate host immune responses [8]. Several studies have shown that the gut virome in IBD patients differs significantly from that of healthy individuals. Specifically, caudovirales families are expanded, while crAss-like phages and microviridae are reduced, indicating disrupted viral stability and loss of protective taxa [8, 9]. These alterations may favor dysbiosis and promote inflammatory conditions by facilitating the growth of pathogenic bacterial groups [10].

The whole gut ecosystem is impacted during disease activity, as evidenced by integrated multi-omics approaches that have demonstrated that viral alterations frequently coincide with bacterial and fungal disruptions [4, 11]. Moreover, large-scale association analyses have identified reproducible viral signatures across populations, suggesting their possible relevance for disease characterization [12]. In addition to bacteriophages, enrichment of eukaryotic viruses such as *anelloviridae* and *herpesviridae* has been reported in IBD patients, particularly in severe forms of disease [5]. These results imply that mucosal damage may be caused by viral reactivation or secondary colonization in inflammatory environments. However, a significant percentage of viral sequences are still unclassified; this is known as “viral dark matter,” and it reflects the limited understanding of the gut virome that is currently available [11].

The present study was conducted in Tabriz, Northwest of Iran, with the aim of investigating gut virome alterations in patients with UC and CD at different disease severities in comparison with healthy controls. This study aimed to investigate the composition and diversity

of viral communities in relation to disease state using metagenomic sequencing coupled with bioinformatics analyses. The results are intended to provide baseline data from this population, enhance the limited literature on IBD-associated virome alterations, and inform future research on microbial contributions to intestinal inflammation. This study characterizes virome alterations across IBD severity levels, highlighting viral signatures as potential non-invasive markers for disease stratification.

Materials and methods

Study design and participants

This cross-sectional observational study was conducted at the Inflammatory Bowel Disease (IBD) Center of Imam Reza Hospital, Tabriz, Iran, between August 2023 and December 2025. Fifty individuals were recruited and divided into five groups ($n=10$ each) including patients with mild ulcerative colitis, severe ulcerative colitis, mild Crohn's disease, severe Crohn's disease, and healthy controls matched for age and sex. The diagnosis of UC and CD was established according to clinical, endoscopic, histological, and radiological criteria. Disease activity in UC patients was assessed using the Mayo score, with mild disease defined as a Mayo score ≤ 2 and severe disease as a Mayo score ≥ 11 . For CD, disease activity was determined by the Crohn's Disease Activity Index (CDAI), where mild disease corresponded to CDAI 150–219 and severe disease to CDAI > 450 .

Eligible participants were adult patients (≥ 18 years) with a confirmed diagnosis of UC or CD who had not received corticosteroids, immunomodulators, or biologic agents during the three months preceding enrollment. Healthy controls had no history of gastrointestinal or systemic inflammatory disorders and no recent use of antibiotics, probiotics, or immunosuppressive drugs. Exclusion criteria included infectious colitis, malignancy, pregnancy, and or refusal to provide consent.

Sample collection

At the time of recruitment, all participants provided fresh stool (~ 10 g) and peripheral blood (~ 5 mL) samples. Stool samples were collected in sterile containers and immediately placed on ice. Samples were transported to the laboratory within two hours of collection and stored at -80°C until nucleic acid extraction. Peripheral blood was collected under fasting conditions in EDTA tubes for laboratory analysis. Serum was separated by centrifugation at 3,000 rpm for 10 min and stored at -80°C for subsequent analyses.

Viral nucleic acid extraction and metagenomic next-generation sequencing

Centrifugation and filtration through $0.22\ \mu\text{m}$ filters were performed to remove host and bacterial cells from fecal

suspensions and to concentrate virus-like particles. Filtrates were treated with Turbo DNase (Thermo Fisher Scientific, USA) to degrade unprotected nucleic acids, ensuring preservation of capsid-protected viral genomes. Viral DNA was then isolated using the QIAamp MinElute Virus Spin Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA yield and quality were assessed with a Qubit 4 fluorometer (Thermo Fisher Scientific), and extracts were stored at -80°C until sequencing. Extracted viral DNA was submitted to Novogene Co., Ltd. (Beijing, China) for metagenomic library preparation and sequencing. Libraries were sequenced on an Illumina NovaSeq platform with paired-end 150 bp reads (2×150 bp). After quality control, an average of 18.3 million clean reads per sample was obtained.

Bioinformatics analysis

Raw reads were checked for quality using FastQC v0.11.9 and trimmed with Trimmomatic v0.39 to remove adapters and low-quality bases. Host sequences were removed by mapping to the human genome (GRCh38) with Bowtie2 v2.4.5. High-quality reads were assembled into contigs using MEGAHIT v1.2.9, and viral contigs were identified by alignment to the NCBI RefSeq viral database with BLASTx v2.13.0 and confirmed with VirSorter2 v2.2.3. Taxonomic classification was assigned using Kraken2 v2.1.2 and VirFinder. Differential abundance of viral taxa was assessed using DESeq2 v1.38.3, applying variance-stabilizing transformation and Benjamini–Hochberg correction ($\text{FDR} < 0.05$). The Pfam database was used for functional annotation of viral proteins. Domain counts were normalized as Reads Per Kilobase (RPK) to account for sequence length and subsequently $\log_2$ -transformed for comparative analyses. Random forest models using discriminant viral taxa were applied to classify controls vs. IBD and mild vs. severe disease, with performance evaluated by 5-fold cross-validated ROC curves (AUC, 95% CI). Bacterial host prediction for viral contigs was performed using CRISPR–spacer detection (MinCED) and BLASTn-based matching (bit-score ≥ 45), complemented by nucleotide-identity alignment against Unified Human Gastrointestinal Genome (UHGG) genomes. Approximately 24% of non-redundant viral operational taxonomic units (vOTUs) obtained host assignments using these criteria [12].

Statistical analysis

All analyses were performed using R software (version 4.2.2). Continuous variables were tested for normality (Shapiro–Wilk) and expressed as mean $\pm$ SD. Group comparisons were made with ANOVA and Tukey's post-hoc test or Kruskal–Wallis with Dunn's correction, while categorical variables were compared using the Chi-square test. For the virome data, alpha diversity

measured by Shannon, Simpson, and Chao1 richness and beta diversity calculated using Bray–Curtis dissimilarity were evaluated in QIIME2 v2023.2. Group differences in beta diversity were assessed with PERMANOVA, and community structures were illustrated using PCoA. To evaluate the assumption of homogeneous dispersion, we performed a PERMDISP test. To ensure fair comparisons, raw reads were normalized to the minimum sequencing depth observed across all samples prior to diversity analysis. Differentially abundant viral taxa were identified with DESeq2 v1.38.3 and LEfSe. A $p < 0.05$ was considered statistically significant. To ensure appropriate control of Type I error across analyses involving multiple group comparisons, explicit correction procedures were applied. For parametric post-hoc tests, Tukey's HSD was used, and for non-parametric post-hoc tests, Dunn's test with Benjamini–Hochberg false discovery rate (FDR) adjustment was applied. In differential-abundance testing with DESeq2, Benjamini–Hochberg FDR correction was applied across all viral taxa, and statistical significance was defined based on adjusted p-values ($\text{FDR} < 0.05$). LEfSe's two-stage testing framework and LDA threshold ($\text{LDA} > 2$) further minimized false positives. PERMANOVA was performed with 999 permutations, and homogeneity of dispersion was confirmed with PERMDISP.

Results

Demographic and clinical findings

Fifty participants were included, divided into five groups (mild UC, severe UC, mild CD, severe CD, and healthy controls; $n = 10$ each). The mean age across all groups was above 36 years with no significant difference, and sex distribution was comparable ($p > 0.05$). In terms of inflammatory markers, ESR and CRP were markedly elevated in severe UC and CD compared with mild disease and controls ($p < 0.001$). Patients with mild UC and CD showed intermediate increases relative to controls. Fecal calprotectin levels were significantly higher in all IBD patients compared with healthy individuals, with the greatest elevations in severe CD and UC ($p < 0.001$). Regarding biochemical parameters, serum albumin concentrations were significantly reduced in patients with severe disease compared with mild subgroups and controls ($p < 0.001$). Creatinine levels remained stable across all groups and showed no significant difference ($p = 0.71$). For hematological indices, NLR, PLR, and MLR were significantly increased in IBD patients compared with controls, with the highest values observed in severe UC and CD (all $p < 0.001$). MPV demonstrated the opposite pattern, being significantly lower in severe subgroups compared with controls ($p = 0.01$) (Table 1).

Table 1 Baseline clinical and laboratory characteristics of patients and controls

Variables	Controls (n = 10)	Mild UC (n = 10)	Severe UC (n = 10)	Mild CD (n = 10)	Severe CD (n = 10)	p-value
Sex, n (%)						
Male	5 (50%)	5 (50%)	5 (50%)	5 (50%)	5 (50%)	1.00
Female	5 (50%)	5 (50%)	5 (50%)	5 (50%)	5 (50%)	
Age (Mean $\pm$ SD)	37.41 $\pm$ 9.53	39.62 $\pm$ 8.73	41.35 $\pm$ 10.29	38.14 $\pm$ 9.87	40.28 $\pm$ 9.66	0.587
ESR (mm/h)	9.29 $\pm$ 3.09	21.74 $\pm$ 7.41	48.64 $\pm$ 12.90	19.86 $\pm$ 6.49	52.43 $\pm$ 13.69	<0.001
CRP (mg/L)	1.85 $\pm$ 0.92	6.29 $\pm$ 2.76	26.95 $\pm$ 9.39	5.77 $\pm$ 2.41	29.44 $\pm$ 10.65	<0.001
Fecal Calprotectin (μ g/g)	24.66 $\pm$ 12.47	176.33 $\pm$ 83.09	628.83 $\pm$ 221.54	192.47 $\pm$ 94.18	711.49 $\pm$ 259.81	<0.001
Albumin (g/dL)	4.44 $\pm$ 0.31	4.06 $\pm$ 0.39	3.18 $\pm$ 0.47	4.09 $\pm$ 0.38	3.12 $\pm$ 0.51	<0.001
Creatinine (mg/L)	0.84 $\pm$ 0.13	0.97 $\pm$ 0.18	0.96 $\pm$ 0.16	0.92 $\pm$ 0.08	0.90 $\pm$ 0.19	0.39
NLR	1.71 $\pm$ 0.63	2.82 $\pm$ 0.90	5.65 $\pm$ 1.43	2.59 $\pm$ 0.82	5.87 $\pm$ 1.57	<0.001
PLR	112.36 $\pm$ 28.93	168.74 $\pm$ 41.50	264.96 $\pm$ 58.24	158.33 $\pm$ 39.38	279.11 $\pm$ 62.66	<0.001
MLR	0.24 $\pm$ 0.08	0.33 $\pm$ 0.11	0.52 $\pm$ 0.14	0.31 $\pm$ 0.09	0.55 $\pm$ 0.15	<0.001
MPV (fL)	10.22 $\pm$ 0.81	9.63 $\pm$ 0.72	9.02 $\pm$ 0.61	9.68 $\pm$ 0.69	8.88 $\pm$ 0.59	0.004

UC, Ulcerative colitis; CD, Crohn's disease; ESR, Erythrocyte sedimentation rate; CRP, C-reactive protein; NLR, Neutrophil-to-lymphocyte ratio; PLR, Platelet-to-lymphocyte ratio; MLR, Monocyte-to-lymphocyte ratio; MPV, Mean platelet volume; FC, Fecal calprotectin; SD, standard deviation

Gut microbiome analysis

With an average of 18.3 million reads per sample, shotgun metagenomic sequencing generated a total of 0.92 billion paired-end reads (150 bp) from 50 fecal samples. After removing host-derived sequences and applying quality filtering, approximately 75.7% of reads were retained. Over 92% of bases achieved a Q30 score, indicating high-quality data suitable for downstream virome analysis, and 7.5% of the reads were assigned to viral contigs.

Alpha diversity indicators showed that virome richness and diversity were lower in severe UC and CD cases compared with healthy controls, while mild cases retained values similar to controls (Fig. 1). Severe UC and CD patients had lower observed richness compared with controls, whereas mild cases were similar to healthy individuals (Fig. 1A). Estimates of total richness based on the Chao1 index were also reduced in severe groups, while mild groups overlapped with controls (Fig. 1B). Shannon diversity, which incorporates both richness and evenness, was lower in severe cases but remained comparable between mild groups and controls (Fig. 1C). Simpson diversity showed the same pattern, with lower values in severe UC and CD groups and little difference between mild cases and controls (Fig. 1D). Overall, these results indicate that reductions in richness and diversity were observed primarily in severe UC and CD, while mild disease showed values that were largely comparable to controls.

Complementary beta diversity analysis based on Bray–Curtis dissimilarity confirmed global differences in community composition (PERMANOVA, $R^2 = 0.021$, $F = 3.24$, $p < 0.001$) (Fig. 2). We also conducted a PERMDISP test, which revealed no significant differences in group dispersion ($p > 0.05$), thereby supporting the robustness of the PERMANOVA results. Severe UC and CD patients formed distinct clusters away from healthy controls, while mild cases overlapped largely with controls. No

separation was detected between UC and CD within the same severity group, although severe groups consistently diverged from mild groups. These results suggest that, in this cohort, both within-sample diversity and between-sample structure of the virome were more related to disease severity than to disease type.

Viral population composition

Taxonomic profiling of the gut DNA virome across all cohorts revealed that the community was strongly dominated by bacteriophages, particularly members of the order Caudovirales (including *myoviridae*, *siphoviridae*, and *podoviridae*) and *microviridae* (Fig. 3A). Together, these accounted for >90% of the total virome, whereas eukaryotic viruses such as *Anelloviridae*, *circoviridae*, *herpesviridae*, and *adenoviridae* were detected at lower relative abundances. Analysis of family-level viral composition using stacked bar plots (Fig. 3A and B) revealed distinct patterns across study groups. Healthy controls carried a higher relative abundance of crAss-like phages and *microviridae*, both considered part of the core human virome and associated with microbial stability. In contrast, patients with severe IBD (particularly severe UC and CD) showed a relative increase in Caudovirales families, especially *siphoviridae* and *myoviridae*, reflecting shifts in bacteriophage composition under inflammatory conditions. Among eukaryotic viruses, *anelloviridae* appeared more prominent in severe UC and CD, while *herpesviridae* were more abundant in severe UC. Other families, such as *podoviridae*, *circoviridae*, *adenoviridae*, and *inoviridae*, exhibited only minor fluctuations without consistent group-specific patterns. Overall, these compositional profiles point to a depletion of core phages in severe disease and a relative prominence of specific eukaryotic viral taxa, particularly in UC, highlighting distinctive changes of the gut virome in IBD.

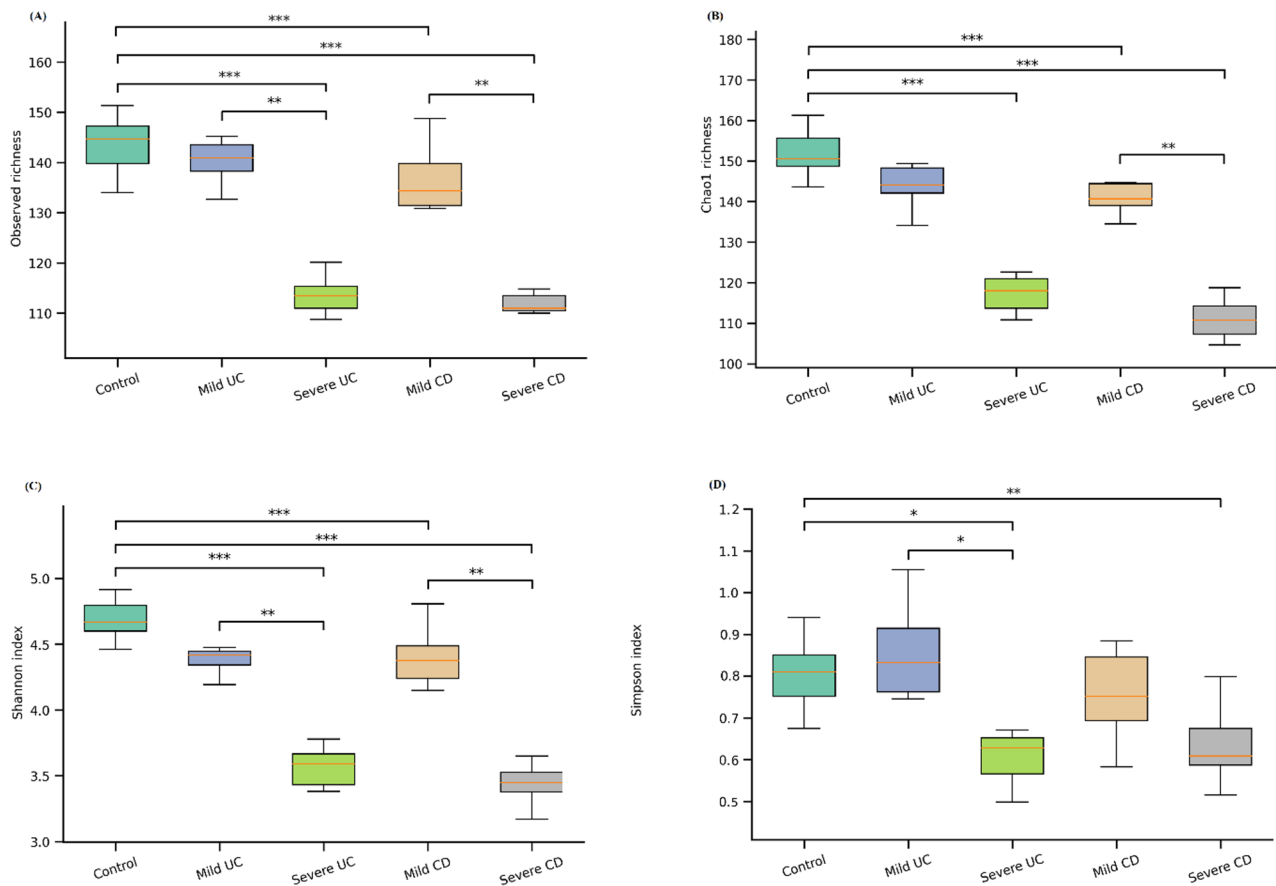


Fig. 1 Alpha diversity of the gut virome across study groups. Boxplots show observed richness (A), Chao1 richness (B), Shannon index (C), and Simpson index (D). Severe UC and CD cases exhibited significantly lower richness and diversity compared with healthy controls, whereas mild UC/CD retained values comparable to controls. Statistical significance was determined using pairwise comparisons (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

We used LEfSe analysis (LDA > 2, Kruskal–Wallis $p < 0.05$; Fig. 3B) to further distinguish viral taxa enriched in health from disease. In the analysis with five groups, few taxa distinguished UC from CD within the same severity level, consistent with β -diversity findings. In the severity-focused analysis (controls vs. IBD-mild vs. IBD-severe), LEfSe identified taxa enriched in IBD patients, with enrichment being most notable in the severe subgroups. Consistent with the stacked bar plots, crAss-like phages and *microviridae* were significantly enriched in healthy controls, supporting their role as core components of a stable gut virome. In contrast, *siphoviridae* and *myoviridae* were enriched across IBD patients, mainly in both severe UC and CD. Among eukaryotic viruses, *anelloviridae* were preferentially enriched in severe IBD, whereas *Herpesviridae* showed a Disease-specific signal, being significantly increased only in severe UC but not in CD. Other viral families, including *podoviridae*, *circoviridae*, *adenoviridae*, and *inoviridae*, showed trends of higher abundance in IBD but did not reach statistical significance. These results indicate that IBD is characterized by depletion of core commensal phages and selective

enrichment of specific bacteriophages and eukaryotic viruses, with *herpesviridae* emerging as a marker of severe UC.

Functional annotations

Functional annotation of viral proteins against the Pfam database revealed top dominant protein domains within the gut virome (Fig. 4). Structural and lytic-associated domains such as phage tail proteins, portal proteins, terminase large subunit, peptidase, glycosyl hydrolases, holin, and endolysin were more abundant in IBD patients, with the highest levels observed in severe IBD. In contrast, lysogenic domains (integrase, recombinase, RT/Retron) were more abundant in healthy controls and Mild IBD. This pattern is consistent with the persistence of temperate phages, such as crAss-like phages, in these groups. Patients with Mild IBD generally exhibited modest increases in structural Pfam domains compared with controls, but consistently lower levels than Severe IBD. Since no significant differences were observed between UC and CD within the same severity class, the data were

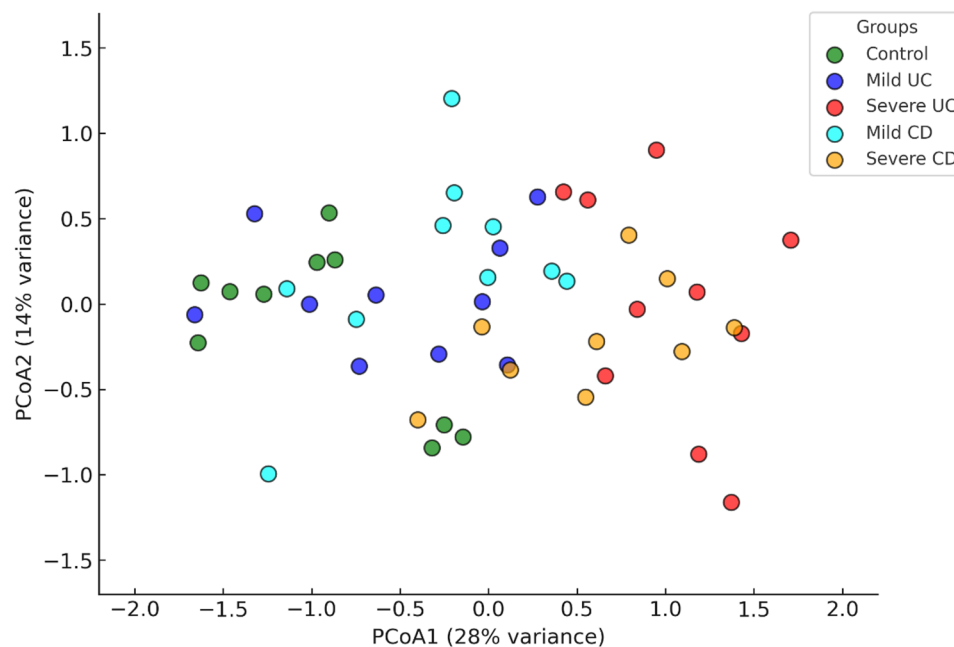


Fig. 2 PCoA (Bray–Curtis) of virome β -diversity across five groups. Ellipses denote 95% confidence intervals around group centroids ($n=10/\text{group}$). Global differences were significant by PERMANOVA ($R^2 = 0.021$, $F=3.24$, $p<0.001$); PERMDISP indicated no significant differences in group dispersion ($p>0.05$). Severe UC/CD samples cluster away from controls, whereas mild cases largely overlap with controls. No clear separation was observed between UC and CD within the same severity class

merged into three categories (Control, Mild IBD, Severe IBD) for clarity of presentation.

Virome-based prediction of disease severity

To assess the diagnostic utility of the gut virome, a random forest classifier was trained using the top discriminant viral taxa identified by LEfSe (LDA >2). Models were built to distinguish (a) controls vs. IBD, and (b) mild vs. severe IBD. Receiver operating characteristic (ROC) curves with area under the curve (AUC) and 95% confidence intervals were generated. Random forest models based on viral relative abundances achieved high discriminatory power. For distinguishing IBD patients from healthy controls, the model yielded an AUC of 0.89 (95% CI 0.81–0.95), while classification of mild vs. severe IBD reached an AUC of 0.83 (95% CI 0.75–0.91). ROC curves demonstrated the potential utility of virome-based biomarkers in non-invasive stratification of IBD. Variable importance analysis further indicated that *siphoviridae*, *anelloviridae*, and crAss-like phages were the strongest predictors (Fig. 5).

Phage–host prediction and ecological assignment

To contextualize the viral signatures observed in this study, genus-level host prediction was performed for all viral operational taxonomic units (vOTUs) using CRISPR spacer matching and nucleotide-identity criteria against UHGG genomes. Approximately 23.7% of vOTUs met confidence thresholds for host assignment, which is

consistent with expectations for metagenomic virome datasets based on short to medium-length contigs.

Host-based comparisons were therefore focused on four phage groups—crAss-like phages, Microviridae, Siphoviridae, and Myoviridae—that showed consistent differential enrichment between IBD patients and healthy controls, whereas other detected phage families were not further examined due to the absence of clear group-specific patterns. CrAss-like phages showed the most consistent and specific host profile, with the majority of annotated crAss-like vOTUs aligning with *Bacteroides* spp., reflecting the well-recognized specialization of this family toward Bacteroidetes lineages in the human gut (Table 2). Microviridae exhibited a similar pattern, with predicted associations primarily involving *Bacteroides* (Table 2). In contrast, Caudovirales families displayed more heterogeneous host associations. Among vOTUs that were differentially enriched between groups and could be confidently linked to a bacterial host, Siphoviridae enriched in IBD samples were predominantly associated with *Escherichia* and *Enterococcus*, whereas IBD-enriched Myoviridae were mainly linked to *Klebsiella* and *Ruminococcus*. Consistent with phage enrichment patterns, control-enriched vOTUs were largely represented by crAss-like phages and Microviridae associated with *Bacteroides*, whereas IBD-enriched vOTUs were mainly affiliated with Caudovirales linked to facultative anaerobic genera. Host-based analyses were restricted to vOTUs with confident genus-level host

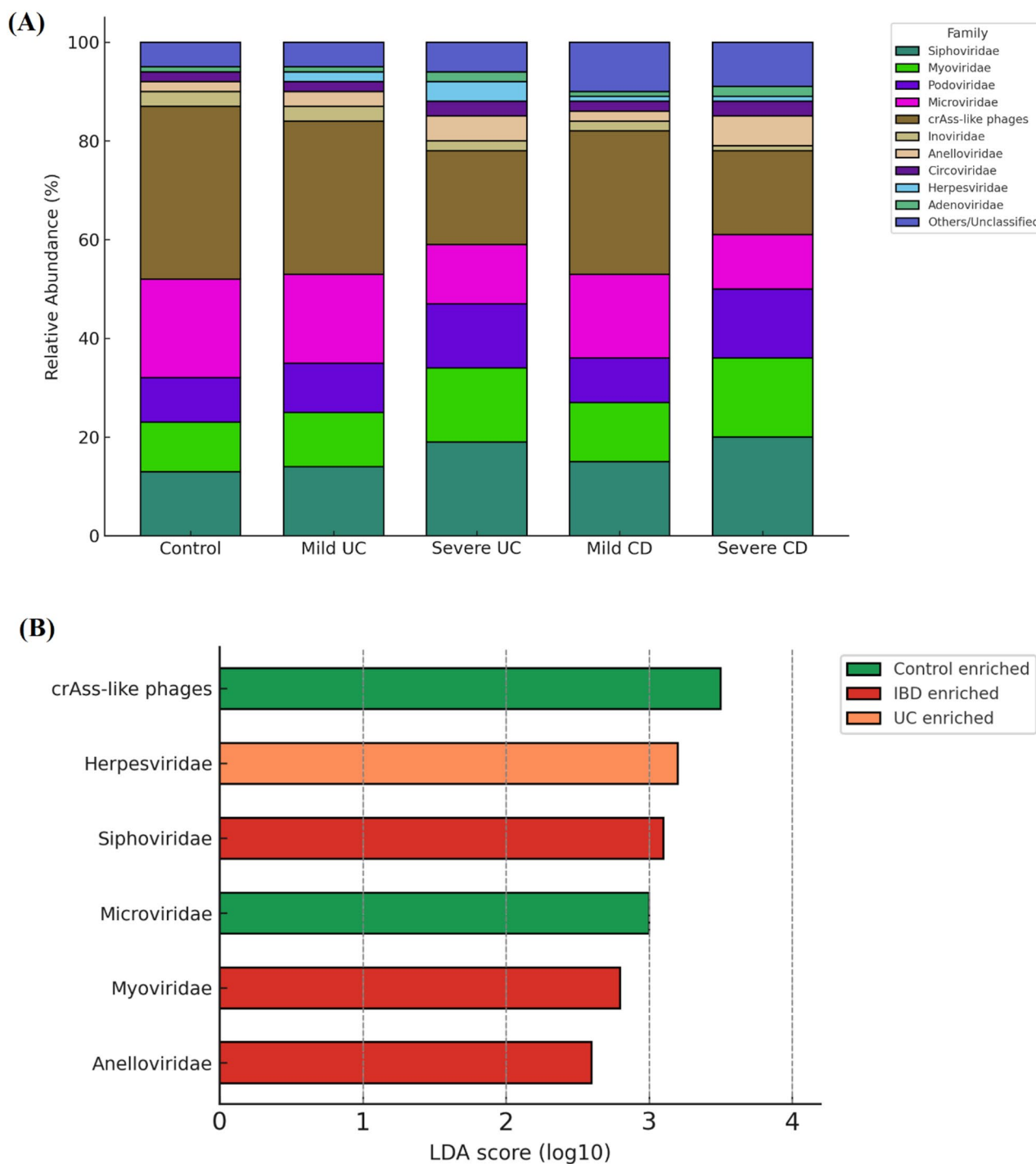


Fig. 3 Viral family profiles and discriminant analysis in IBD. **(A)** Relative abundance of viral families across groups. Controls exhibited higher levels of crAss-like phages and *microviridae*, while IBD patients showed increased Caudovirales and eukaryotic families such as *herpesviridae* and *anelloviridae*. **(B)** Linear discriminant analysis (LEfSe, LDA > 2, $p < 0.05$) of viral families. crAss-like phages and *microviridae* were enriched in controls, whereas *siphoviridae* and *myoviridae* were enriched in IBD patients (notably in severe UC and CD). *Anelloviridae* were predominantly enriched in severe cases, and *herpesviridae* were specifically associated with severe UC

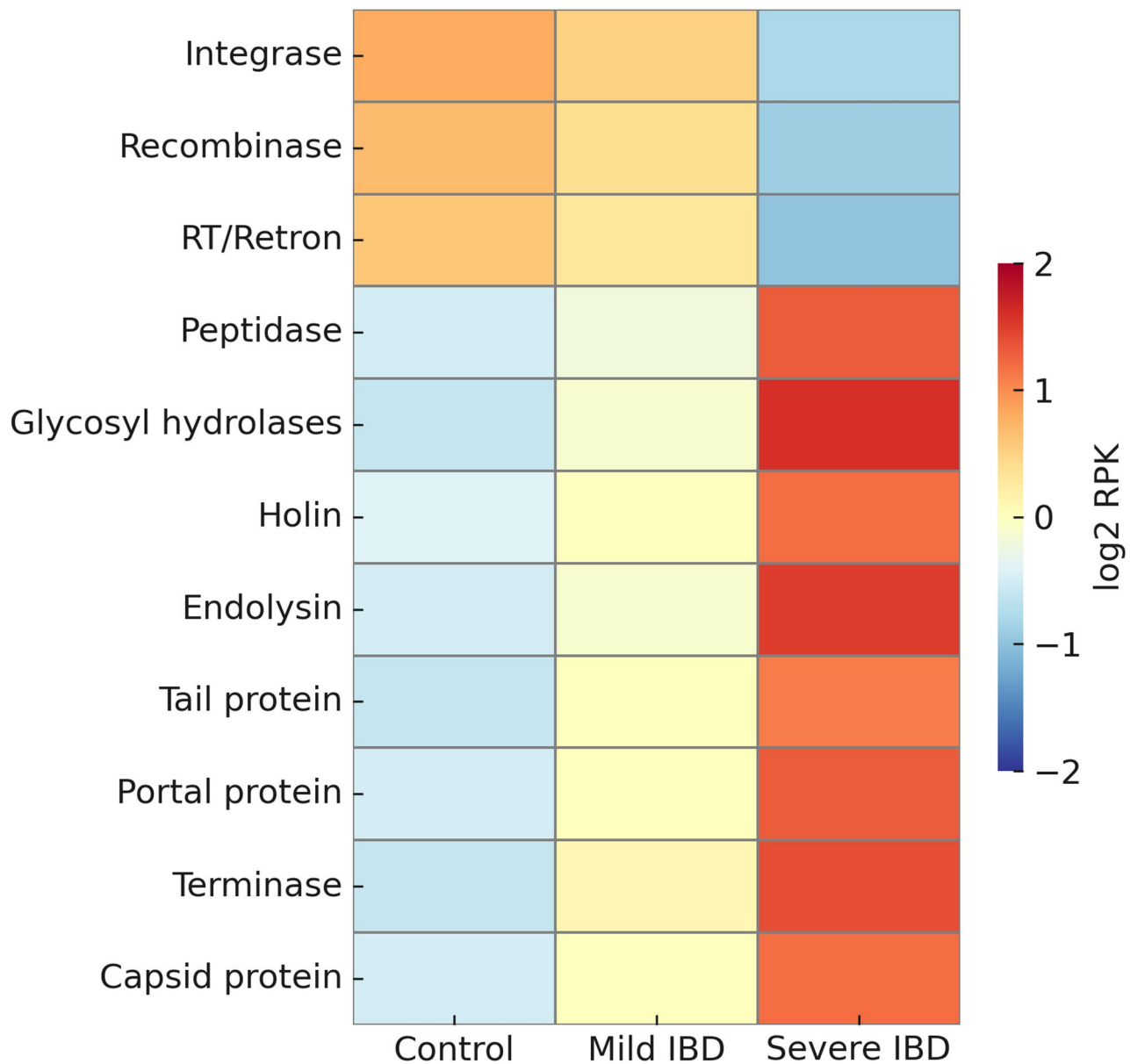


Fig. 4 Functional annotation of viral protein domains in virome samples. Heatmap showing the relative enrichment (log2 RPK) of functional domains across three groups: healthy controls, mild IBD, and severe IBD. Heatmap showing relative enrichment (log2 RPK) of functional domains. Subgroups of UC and CD were merged, as no significant differences were observed between them

assignments, and vOTUs with unclassified or multiple host predictions were excluded. These results provide ecological context for the observed virome alterations by linking phage-level shifts to differences in the distribution of phages infecting distinct bacterial hosts between IBD patients and healthy controls.

Discussion

The findings of this study highlight consistent changes in the gut virome of individuals with inflammatory bowel disease. To our knowledge, this is the first virome-based study in an Iranian IBD cohort that systematically

integrates taxonomic, functional, and predictive modeling analyses. By distinguishing mild from severe disease states through viral signatures, our work extends previous observations of virome alterations and highlights the potential of viral markers for non-invasive disease stratification. One of the strongest signals was the reduction in viral diversity, particularly in patients with severe forms of ulcerative colitis and Crohn's disease. Diversity loss has been described before, and our results align with those reports showing that IBD is consistently marked by a contracted virome [13, 14]. Pediatric and very early onset cases sometimes maintain overall richness, although

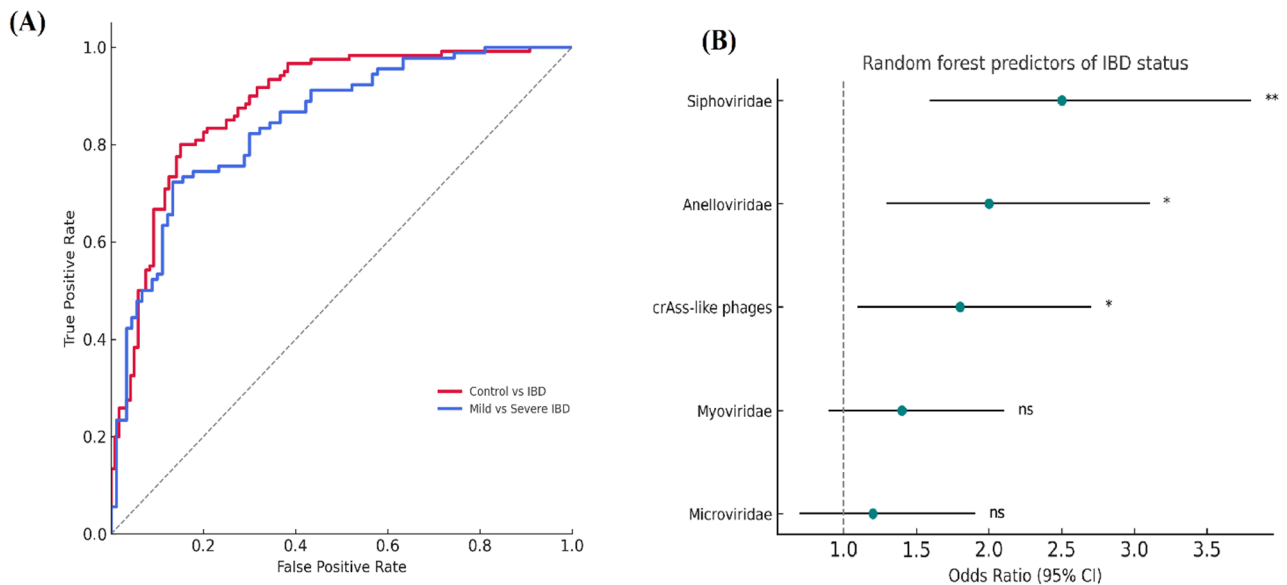


Fig. 5 Predictive modeling of IBD status based on gut virome features. **(A)** Receiver operating characteristic (ROC) curves from random forest models distinguishing **(A)** IBD vs. controls and **(B)** mild vs. severe IBD. AUC values are shown; corresponding 95% confidence intervals are reported in the Results section. **(B)** Variable importance plot indicating *siphoviridae*, *anelloviridae*, and crAss-like phages as the strongest predictors of disease status. Significance levels are denoted as: * $p < 0.05$, ** $p < 0.01$, ns = not significant

Table 2 Host-associated distribution of differentially enriched phage groups in IBD and healthy controls

Phage taxonomic group	Enrichment group	Predicted host category	Number of vOTUs *
crAss-like phages	Control-enriched	Bacteroides	37
Microviridae	Control-enriched	Bacteroides	20
Siphoviridae	IBD-enriched	Escherichia/ Enterococcus	26
Myoviridae	IBD-enriched	Klebsiella/ Ruminococcus	16

Counts represent vOTUs with confident genus-level host assignments; vOTUs with unclassified or multiple host predictions were excluded

their community balance shifts in different ways [15–17]. Taken together, these results indicate that disease severity and host age may affect the degree of reduction in viral diversity.

Beyond overall richness, we observed striking compositional changes. Members of the Caudovirales order, particularly siphoviridae and myoviridae, expanded in IBD patients, whereas crAss-like phages and microviridae were consistently reduced. These patterns have also been reported in other cohorts, which suggests a degree of consistency, although variations between studies exist [13, 14]. The reduction of crAss-like phages is particularly significant, as these phages are abundant and stable components of healthy gut communities. This decline may reflect the loss of ecological stabilizers that regulate bacterial fluctuations and support community resilience [18]. Experimental work has suggested that excessive phage induction can activate innate immune receptors

such as toll like receptors, leading to inflammatory cascades [19]. This biological link may provide a plausible mechanism that connects the observed phage imbalance with clinical inflammation. Not all reports are completely consistent—crAss-like phage levels can vary depending on sampling site or cohort—but this is likely explained by technical and demographic factors rather than a true absence of effect [9].

The virome shifts were not limited to bacteriophages. We also observed increased abundance of certain eukaryotic viruses. Anelloviridae were enriched in severe cases, consistent with their proposed role as markers of immune activation that fluctuate with host immune pressure [5, 20, 21]. Herpesviridae showed a distinct rise in patients with severe ulcerative colitis. These viruses are known for their ability to reactivate during stress and inflammation, raising the possibility that they may exacerbate mucosal injury [22]. The herpesviridae family, including cytomegalovirus and Epstein–Barr virus, has attracted growing interest in IBD research, as several studies have reported their presence in the intestinal lumen of patients more frequently than in healthy controls [23, 24]. Although some datasets have reported higher anelloviridae in healthy individuals [25], the enrichment of herpesviridae specifically in severe UC in our study remains noteworthy and may warrant further mechanistic investigation.

Functional annotation of viral genes provided additional evidence supporting these observations. Genes encoding lytic functions, including structural proteins, peptidase, glycosyl hydrolases, holin, and endolysin, were significantly enriched in IBD samples, particularly

in severe disease. Among these lytic modules, terminase, endolysin, and glycosyl hydrolases showed the most pronounced enrichment in severe IBD. This likely reflects their central roles in phage replication and host cell lysis. By contrast, lysogeny-associated domains, such as integrase, recombinase, and RT/Retron-like reverse transcriptase, were more abundant in controls and Mild IBD. This dual pattern suggests that temperate phages are maintained in the healthy gut, while inflammatory conditions promote their induction, leading to the expression of lytic modules. These findings are consistent with reports highlighting prophage activation during inflammation and illustrate a potential interplay between lysogeny and prophage induction in IBD pathogenesis [25, 26]. A plausible model suggests that inflammation triggers prophage activation, resulting in bacterial lysis and the release of microbial components that further fuel intestinal inflammation, creating a self-reinforcing cycle [11, 14]. This process coincides with the reduced bacterial diversity and counts and the expansion of caudovirales frequently observed in IBD [11, 27]. Prophage induction accelerates bacterial lysis, releasing debris that activates innate immunity and mucosal lymphoid tissues. Consistent with this view, integrated prophages are known to exploit host physiology and population density to modulate the switch from lysogeny to lytic replication [11, 28, 29]. Taken together, our results indicate that lytic functional domains are enriched in IBD, particularly in severe disease, while lysogeny-associated functions are more abundant in controls and mild cases. This pattern aligns with the concept of prophage induction under inflammatory conditions and supports a model in which the virome actively contributes to disease progression, rather than merely reflecting bacterial dysbiosis. Nevertheless, these functional insights should be interpreted with caution. The annotation is based on predicted domain content derived from metagenomic sequences rather than direct evidence of functional activity. As such, the enrichment of structural or lysogeny-associated domains in IBD may primarily reflect shifts in viral community composition. Confirmation of their biological activity would require metatranscriptomic or proteomic data to verify whether these functions are actively expressed in vivo.

Host-prediction analysis provided ecological context for the viral patterns observed in this study. CrAss-like phages and Microviridae—well-recognized core members of the human gut virome—were primarily associated to *Bacteroides* and other *Bacteroidetes* or *Firmicute* genera, consistent with their established host specialization in previous metagenomic and culture-based research [6, 30]. In line with previous work, *Caudovirales* families, particularly *Siphoviridae* and *Myoviridae*, are known to infect a broad range of gut *Proteobacteria* and *Firmicutes*, including clinically relevant genera such as

Escherichia, *Klebsiella* and *Enterococcus* [12, 31]. Reports of temperate phages infecting mucin-degrading *Ruminococcus gnavus* further support the ecological relevance of these host–phage associations in the human gut [32]. Associations involving other phage groups, such as *Podoviridae* and *Inoviridae*, have also been described in prior studies [33], although these families were not further examined in the present host-based analysis. Although 23.7% of vOTUs could be confidently assigned, the overall host–phage distribution observed here aligns with known biological features of major gut phage families and supports the ecological interpretation of the virome signatures identified in this cohort.

One of the most important clinical aspects of this study is predictive modeling. Using the Random Forest algorithm, we were able to reliably distinguish IBD patients from healthy controls (AUC = 0.89) as well as differentiate between mild and severe disease states (AUC = 0.83). In the variable importance ranking, *siphoviridae*, *anelloviridae*, and crAss-like phages emerged as the important predictors, consistent with the results of our taxonomic analysis. These findings align with recent virome-wide association studies, such as the work of Tian et al., which also reported that phages from *siphoviridae* and crAss-like groups are important contributors to IBD classification models with comparable performance metrics [12]. Previous research has demonstrated that crAss-like viruses possess comparatively large genomes and harbor a considerable repertoire of auxiliary functional genes, particularly carbohydrate-active enzymes critical for polysaccharide metabolism. Consequently, a reduction in the abundance of crAss-like viruses may suggest the loss of essential viral functions (such as polysaccharide metabolism) within the gut virome of patients with IBD [18, 34–36]. Consistent with this, Liang et al., reported that elevated *anelloviridae* DNA levels in stool may serve as a potential biomarker of IBD in very early-onset patients and correlate with immunosuppressive treatment [15]. Our findings suggest that viral signatures may help to distinguish IBD patients from healthy controls and could reflect disease activity, although further validation is required. The accuracy of our classification models was similar to, or in some cases better than, what has been reported in previous virome-based studies. When we considered disease severity and included functional features, viral signatures alone showed strong predictive ability. Although combining these data with bacterial information may further improve accuracy, the fact that viruses on their own reached this level of performance highlights their clinical potential [25]. Nevertheless, these predictive models should be interpreted with caution. The sample size was limited, and the models were trained and tested within the same cohort, raising the possibility of overfitting. Independent validation in larger, external

datasets will be essential to confirm the robustness and generalizability.

Previous studies exploring the relationship between virome dynamics and treatment response in IBD are still limited, and only a few longitudinal cohorts have examined pre- and post-therapy viral shifts. Existing evidence suggests that patients who respond to anti-TNF or fecal microbiota transplantation (FMT) therapy tend to exhibit a reduction in Caudovirales and partial restoration of crAss-like phages, whereas non-responders often retain an inflammatory virome signature [15, 37–39]. However, these datasets remain sparse, and virome-based predictors of therapeutic response are not yet well established. Because all individuals in our study were strictly treatment-naïve and the design was cross-sectional, assessing treatment response was not feasible. Nevertheless, the present findings provide a valuable untreated virome baseline that may serve as a reference for future longitudinal investigations aimed at linking viral community shifts with clinical treatment outcomes.

This study has several limitations. The cross-sectional design restricts causal inference, meaning we cannot establish whether virome alterations precede or follow mucosal inflammation. The sample size was 10 participants per group, which was sufficient to detect major differences but limited the statistical power to identify rare taxa or subtle changes. In addition, collecting samples from a single center restricts the generalizability of the findings, as environmental conditions and demographic characteristics may vary across populations. In addition, stool samples were used as the only specimen type, although prior studies indicate that mucosa-associated virome can differ substantially from luminal communities. Technical factors, including sequencing depth and DNA extraction methods, may also introduce detection bias, despite efforts to standardize procedures. Additionally, functional predictions were based on bioinformatic annotations and still require direct experimental validation. Despite these limitations, the study has several strengths. Using balanced groups of 10 participants helped minimize sampling bias, and explicitly stratifying by disease severity allowed for a clearer interpretation of virome changes. In addition, combining taxonomic, functional, and predictive analyses provided complementary insights and strengthened the overall robustness of the findings.

In conclusion, this study showed that in patients with IBD, the gut virome is characterized by the expansion of caudovirales, depletion of crAss-like phages, reduced diversity, and enrichment of certain eukaryotic families. Predictive modeling demonstrated the potential of virome-derived features to distinguish between patients and controls, as well as between mild and severe disease, while functional signatures indicated increased prophage

induction. When considered together, these results are consistent with the possibility that the virome may contribute to the pathophysiology of IBD, although further studies are needed to clarify its role. The current findings provide preliminary evidence that may inform future studies exploring therapeutic approaches targeting the viral component of the gut microbiome and for taking into account virome-based biomarkers in the clinical management of IBD, even though other validation in larger and longer-term cohorts is necessary.

Acknowledgements

The authors gratefully acknowledge all respondents who voluntarily participated in this study.

Author contributions

NED and VP were responsible for the conception and design of the study. NED and MS collected and processed the clinical samples. SMJ, HBB, SSM and AA contributed to VLP preparation and laboratory analyses. HA and AA supervised the statistical analyses. AAS and VP performed data analysis and interpretation. HA, MS and VP drafted the initial manuscript. All authors critically revised the manuscript, approved the final version, and agree to be accountable for the integrity and accuracy of the work.

Funding

This work was supported by the National Institute for Medical Research Development (NIMAD) with grant numbers [979157].

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to ethical and privacy restrictions, but are available from the corresponding author on reasonable request and with permission of the institutional ethics committee.

Declarations

Ethics approval and consent to participate

All participants provided written informed consent prior to enrollment. The study was approved by the Local Ethics Committee of the National Institute for Medical Research Development [IR.NIMAD.REC.1398.236], Iran in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Gastroenterology and Hepatology, School of Medicine, Imam Khomeini Hospital Complex, Tehran University of Medical Sciences, Tehran, Iran

²Research Center for Clinical Virology, Tehran University of Medical Sciences, Tehran, Iran

³Department of Bacteriology and Virology, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran

⁴Liver and Gastrointestinal Diseases Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

⁵Department of Pathology, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran

Received: 30 October 2025 / Accepted: 19 January 2026

Published online: 14 February 2026

References

- Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. *Lancet*. 2017;389(10080):1756–70.
- Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. *Lancet*. 2017;389(10080):1741–55.
- Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet*. 2017;390(10114):2769–78.
- Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*. 2019;569(7758):655–62.
- Wu Y, Cheng R, Lin H, Li L, Jia Y, Philips A, et al. Gut virome and its implications in the pathogenesis and therapeutics of inflammatory bowel disease. *BMC Med*. 2025;23(1):183.
- Shkoporov AN, Hill C. Bacteriophages of the human gut: the known unknown of the microbiome. *Cell Host Microbe*. 2019;25(2):195–209.
- Carding SR, Davis N, Hoyle L. Review article: the human intestinal virome in health and disease. *Aliment Pharmacol Ther*. 2017;46(9):800–15.
- Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell*. 2015;160(3):447–60.
- Yan A, Butcher J, Mack DR, Stintzi A. The colonic mucosal virome in inflammatory bowel disease reveals Crassvirales depletion and disease-specific virome features. *Gut Microbes*. 2025;17(1):2539450.
- Gogokhia L, Buhke K, Bell R, Hoffman B, Brown DG, Hanke-Gogokhia C, et al. Expansion of bacteriophages is linked to aggravated intestinal inflammation and colitis. *Cell Host Microbe*. 2019;25(2):285–99. e8.
- Clooney AG, Sutton TDS, Shkoporov AN, Holohan RK, Daly KM, O'Regan O, et al. Whole-Virome analysis sheds light on viral dark matter in inflammatory bowel disease. *Cell Host Microbe*. 2019;26(6):764–78. e5.
- Tian X, Li S, Wang C, Zhang Y, Feng X, Yan Q, et al. Gut virome-wide association analysis identifies cross-population viral signatures for inflammatory bowel disease. *Microbiome*. 2024;12(1):130.
- Ungaro F, Massimino L, D'Alessio S, Danese S. The gut virome in inflammatory bowel disease pathogenesis: from metagenomics to novel therapeutic approaches. *United Eur Gastroenterol J*. 2019;7(8):999–1007.
- Tun HM, Peng Y, Massimino L, Sin ZY, Parigi TL, Facoetti A, et al. Gut Virome in inflammatory bowel disease and beyond. *Gut*. 2024;73(2):350–60.
- Liang G, Conrad MA, Kelsen JR, Kessler LR, Breton J, Albenberg LG, et al. Dynamics of the stool Virome in very Early-Onset inflammatory bowel disease. *J Crohns Colitis*. 2020;14(11):1600–10.
- Liang G, Gao H, Bushman FD. The pediatric virome in health and disease. *Cell Host Microbe*. 2022;30(5):639–49.
- Tokarz R, Hyams JS, Mack DR, Boyle B, Griffiths AM, LeLeiko NS, et al. Characterization of stool virome in children newly diagnosed with moderate to severe ulcerative colitis. *Inflamm Bowel Dis*. 2019;25(10):1656–62.
- Shkoporov AN, Khokhlova EV, Stephens N, Hueston C, Seymour S, Hryckowian AJ, et al. Long-term persistence of crAss-like phage crAss001 is associated with phase variation in bacteroides intestinalis. *BMC Biol*. 2021;19(1):163.
- Popescu M, Van Belleghem JD, Khosravi A, Bollyky PL. Bacteriophages and the immune system. *Annu Rev Virol*. 2021;8(1):415–35.
- Sabbaghian M, Gheithi A, Shekarchi AA, Tavakoli A, Poortahmasebi V. The mysterious anelloviruses: investigating its role in human diseases. *BMC Microbiol*. 2024;24(1):40.
- Adiliaghdam F, Amatullah H, Digumarthi S, Saunders TL, Rahman RU, Wong LP, et al. Human enteric viruses autonomously shape inflammatory bowel disease phenotype through divergent innate Immunomodulation. *Sci Immunol*. 2022;7(70):eabn6660.
- Hosomi S, Nishida Y, Fujiwara Y. The impact of human herpesviruses in clinical practice of inflammatory bowel disease in the era of COVID-19. *Microorganisms*. 2021. <https://doi.org/10.3390/microorganisms9091870>.
- Ungaro F, Massimino L, Furfaro F, Rimoldi V, Peyrin-Biroulet L, D'Alessio S, et al. Metagenomic analysis of intestinal mucosa revealed a specific eukaryotic gut Virome signature in early-diagnosed inflammatory bowel disease. *Gut Microbes*. 2019;10(2):149–58.
- Sokol H, Seksik P, Rigottier-Gois L, Lay C, Lepage P, Podglajen I, et al. Specificities of the fecal microbiota in inflammatory bowel disease. *Inflamm Bowel Dis*. 2006;12(2):106–11.
- Zuo T, Lu XJ, Zhang Y, Cheung CP, Lam S, Zhang F, et al. Gut mucosal virome alterations in ulcerative colitis. *Gut*. 2019;68(7):1169–79.
- Cornuault JK, Moncaut E, Loux V, Mathieu A, Sokol H, Petit MA, et al. <article-title update="added"> The enemy from within: a prophage of *Roseburia intestinalis* systematically turns lytic in the mouse gut, driving bacterial adaptation by CRISPR spacer acquisition. *ISME J*. 2020;14(3):771–87.
- Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature*. 2017;551(7681):507–11.
- Casjens SR, Hendrix RW. Bacteriophage lambda: early pioneer and still relevant. *Virology*. 2015;479–80:310–30.
- Silpe JE, Bassler BL. A host-produced quorum-sensing autoinducer controls a phage lysis-lysogeny decision. *Cell*. 2019;176(1–2):268–80 e13.
- Guerin E, Shkoporov A, Stockdale SR, Clooney AG, Ryan FJ, Sutton TDS, et al. Biology and taxonomy of crAss-like Bacteriophages, the most abundant virus in the human gut. *Cell Host Microbe*. 2018;24(5):653–64. e6.
- Sutton TDS, Hill C. Gut bacteriophage: current Understanding and challenges. *Front Endocrinol (Lausanne)*. 2019;10:784.
- Buttner C, Khokhlova EV, Stein L, Hueston CM, Govi B, Draper LA, et al. Temperate bacteriophages infecting the mucin-degrading bacterium *Ruminococcus gnavus* from the human gut. *Gut Microbes*. 2023;15(1):2194794.
- Krupovic M, Forterre P. Microviridae goes temperate: microvirus-related proviruses reside in the genomes of bacteroidetes. *PLoS One*. 2011;6(5):e19893.
- Yutin N, Benler S, Shmakov SA, Wolf YI, Tolstoy I, Rayko M, et al. Analysis of metagenome-assembled viral genomes from the human gut reveals diverse putative crAss-like phages with unique genomic features. *Nat Commun*. 2021;12(1):1044.
- Yutin N, Makarova KS, Gussow AB, Krupovic M, Segall A, Edwards RA, et al. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat Microbiol*. 2018;3(1):38–46.
- Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, et al. PhiCrAss001 represents the most abundant bacteriophage family in the human gut and infects bacteroides intestinalis. *Nat Commun*. 2018;9(1):4781.
- Zuo T, Wong SH, Lam K, Lui R, Cheung K, Tang W, et al. Bacteriophage transfer during faecal microbiota transplantation in clostridium difficile infection is associated with treatment outcome. *Gut*. 2018;67(4):634–43.
- Jansen D, Falony G, Vieira-Silva S, Simsek C, Marcelis T, Caenepeel C, et al. Community types of the human gut virome are associated with endoscopic outcome in ulcerative colitis. *J Crohns Colitis*. 2023;17(9):1504–13.
- Jansen D, Deleu S, Caenepeel C, Marcelis T, Simsek C, Falony G, et al. Virome drift in ulcerative colitis patients: faecal microbiota transplantation results in minimal phage engraftment dominated by microviruses. *Gut Microbes*. 2025;17(1):2499575.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.