



OPEN ACCESS

EDITED BY

Dong-Qing Ye,
Anhui Medical University, China

REVIEWED BY

Koshy Nithin Thomas,
Sanjay Gandhi Post Graduate Institute
of Medical Sciences (SGPGI), India
Yong-Xin Liu,
Institute of Genetics and
Developmental Biology, Chinese
Academy of Sciences (CAS), China
Wei Liu,
First Teaching Hospital of Tianjin
University of Traditional Chinese
Medicine, China

*CORRESPONDENCE

Wen Sun
✉ sunwen@bucm.edu.cn
Wukai Ma
✉ walker55@163.com

†These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Autoimmune and Autoinflammatory
Disorders: Autoimmune Disorders,
a section of the journal
Frontiers in Immunology

RECEIVED 22 September 2022

ACCEPTED 23 December 2022

PUBLISHED 11 January 2023

CITATION

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 and Ma W (2023)
Alterations of the gut virome in
patients with systemic
lupus erythematosus.
Front. Immunol. 13:1050895.
doi: 10.3389/fimmu.2022.1050895

Alterations of the gut virome in patients with systemic lupus erythematosus

Changming Chen^{1†}, Qiulong Yan^{2†}, Xueming Yao^{1†},
Shenghui Li³, Qingbo Lv^{3,4}, Guangyang Wang², Qin Zhong¹,
Fang Tang¹, Zhengqi Liu¹, Ying Huang¹, Yang An¹, Jing Zhou¹,
Qiongyu Zhang¹, Aiqin Zhang³, Hayan Ullah², Yue Zhang³,
Can Liu¹, Dan Zhu¹, Hufan Li¹, Wen Sun^{5*} and Wukai Ma^{1*}

¹Department of Rheumatology and Immunology, The Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine, Guiyang, China, ²Department of Microbiology, College of Basic Medical Sciences, Dalian Medical University, Dalian, China, ³Puensum Genetech Institute, Wuhan, China, ⁴College of Animal Science and Veterinary Medicine, Heilongjiang Bayi Agricultural University, Daqing, China, ⁵Key Laboratory of Health Cultivation of the Ministry of Education, Beijing University of Chinese Medicine, Beijing, China

Background: Systemic lupus erythematosus (SLE) is a systemic autoimmune disease that has been linked to the dysbiosis of the gut microbiome and virome. However, the potential characterization of the gut virome in SLE patients needs to be explored more extensively.

Methods: Herein, we analyzed the gut viral community of 16 SLE patients and 31 healthy controls using both bulk and virus-like particle (VLP)-based metagenomic sequencing of their fecal samples. A total of 15,999 non-redundant viral operational taxonomic units (vOTUs) were identified from the metagenomic assembled contigs and used for gut virome profiling.

Results: SLE patients exhibited a significant decrease in gut viral diversity in the bulk metagenome dataset, but this change was not significant in the VLP metagenome dataset. Also, considerable alterations of the overall gut virome composition and remarkable changes in the viral family compositions were observed in SLE patients compared with healthy controls, as observed in both two technologies. We identified 408 vOTUs (177 SLE-enriched and 231 control-enriched) with significantly different relative abundances between patients and controls in the bulk virome, and 18 vOTUs (17 SLE-enriched in 1 control-enriched) in the VLP virome. The SLE-enriched vOTUs included numerous *Siphoviridae*, *Microviridae*, and *crAss-like* viruses and were frequently predicted to infect *Bacteroides*, *Parabacteroides*, and *Ruminococcus_E*, while the control-enriched contained numerous members of *Siphoviridae* and *Myoviridae* and were predicted to infect *Prevotella* and *Lachnospirales_CAG-274*. We explored the correlations between gut viruses and bacteria and found that some *Lachnospirales_CAG-274* and *Hungatella_A* phages may play key roles in the virus-bacterium network. Furthermore, we explored the gut viral signatures for disease discrimination and achieved an

area under the receiver operator characteristic curve (AUC) of above 0.95, suggesting the potential of the gut virome in the prediction of SLE.

Conclusion: Our findings demonstrated the alterations in viral diversity and taxonomic composition of the gut virome of SLE patients. Further research into the etiology of SLE and the gut viral community will open up new avenues for treating and preventing SLE and other autoimmune diseases.

KEYWORDS

systemic lupus erythematosus, gut virome, bulk metagenome, virus-like particle-based metagenome, viral diversity, viral dysbiosis

Introduction

Systemic lupus erythematosus (SLE) is a prototypical autoimmune disease. The disease is characterized by chronic inflammation, which leads to tissue and organ damage, and is more common in females (1). The etiology of SLE is unknown, but it is influenced by several confounding factors, including genetic, environmental, and epigenetic factors (2, 3). These factors can cause immune system abnormalities, including abnormal B cells that produce autoantibodies, autoreactive T cells, and pro-inflammatory cytokines (4). A single factor cannot cause SLE, as a combination of microbial communities and environmental stress is also associated with SLE (5, 6). A recent metagenomic study found that imbalances in the microbial community may be associated with the development of SLE (7). Hevia et al. described the imbalance of gut microbiota in SLE patients (8), and Ma et al. proved that the fecal microbiota of SLE mice could stimulate the inflammatory response and change the expression of SLE susceptibility genes through fecal bacteria transplantation experiment (5). These studies suggest that gut microbiota is associated with SLE and may even be one of the inducing factors of SLE. Since the gut microbiota contains a variety of components (e.g., bacteria, viruses, and fungi), an obvious blind spot in the above-mentioned studies is what function each component plays.

Despite the fact that current research focuses on gut bacteria as one of the main components, gut viruses are also noteworthy. The human gut virome is mainly composed of bacteriophages, especially the *Caudovirales* order (9). Bacteriophages infect bacteria, co-evolve with the gut microbiome, and play an important role in shaping the composition of the gut microbiome. They aid nutrient turnover and facilitate horizontal gene transfer in the gut microbiome (10, 11). Virome may have alterations in response to changes in human gut microbiome composition in many diseases closely related to the gut microbiome. Recent studies have highlighted the association of gut virome with many diseases, particularly

diabetes (12), inflammatory bowel disease (IBD) (13), and autoimmune diseases (14). In addition, gut virome can also directly affect the body's health through the immune system (15, 16). Thus, understanding the changes in gut virome diversity and structure during disease is crucial to exploring the potential pathogenesis of the disease.

However, due to metagenomic sequencing technology, some viruses with low abundance or difficult sequencing are easily overlooked. Therefore, gut virome studies will also use virus-like particle (VLP) metagenomic sequencing (17–19). The integration of whole metagenomic and VLP metagenomic sequencing analysis can obtain a more comprehensive map of the gut virus population. To identify changes in the gut virome in SLE, we collected fecal samples and analyzed the viral communities from 16 SLE patients and 31 healthy subjects. Specifically, we compared the viral composition of SLE patients and healthy individuals based on both bulk metagenomic and VLP-based sequencing data, and the relationships between viruses and bacteria were also investigated. A better understanding of the etiology and pathogenesis of SLE based on gut virome exploration will contribute to the development of new prevention and treatment strategies.

Methods

Ethics statement, subject recruitment, and sample collection

The protocol of this study was approved by the Medical Ethics Committees of the Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine. All subjects who took part in this study gave written informed consent. This study enrolled sixteen SLE patients from the Department of Rheumatology and Immunology, the Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine, from August 2020 to August 2021. The

demographic and clinical characteristics of the SLE patients are shown in [Table S1](#). All patients fulfilled the 2019 European League Against Rheumatism/American College of Rheumatology (EULAR/ACR) classification criteria for SLE (20). The exclusion criteria were as follows: (1) SLE patients with neuropsychiatric lupus, renal lupus, vasculitis, pancreatitis, enteritis, myocarditis, or severe major systemic diseases such as malignancy, pyemia, and cardiovascular or metabolic disorders; (2) patients with overlap syndromes; (3) patients with diarrhea; (4) patients who received antibiotics or probiotics treatment in 1 month; (5) patients who had drunk sour milk within 1 week or had smoking or drinking habits. Thirty-one age-matched and gender-matched healthy subjects were recruited based on records available from the Department of Medical Examination Center, Second Affiliated Hospital of Guizhou University of Chinese Medicine. All volunteers' feces were immediately transported to the laboratory on dry ice and preserved in two aliquots at -80°C until analysis.

Virus-like particles enrichment and virome DNA extraction

Virus-like particles were enriched from the fecal samples according to our previous methods (21). Briefly, for each sample, 0.1 g feces were added into 1 ml HBSS buffer (137 mM NaCl, 5.4 mM KCl, 1.3 mM CaCl_2 , 0.3 mM $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, 0.5 mM $\text{MgCl}_2 \cdot 7\text{H}_2\text{O}$, 0.4 mM KH_2PO_4 , 0.6 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 4.2 mM NaHCO_3 , 5.6 mM D-glucose). To remove debris and bacterial cells from fecal suspensions, samples were centrifuged twice at 10,000 g for 2 minutes at 8°C and the supernatants were then passed through 0.45 μm and 0.22 μm filters successively. The sterile filtrate was mixed with an equivalent amount of HBSS buffer and centrifuged for one hour at 750,000 g (Sorvall mTX150, Thermo Scientific) for 1 hour. The remaining, non-encapsulated nucleic acid in centrifugal precipitation was degraded by treating with a mixture of 2.4 μl TURBO DNase (4.8 U, Invitrogen), 8 μl RNase A/T1 Mix (16 μg RNase A, 40 U RNase T1, Thermo Scientific), and 1 μl Benzonase (5 U, EMD Millipore), followed by heat inactivation of nucleases at 65°C for 10 min.

The TIANamp Virus DNA/RNA Kit (TIANGEN) was used to extract viral DNA from the enriched virus-like particles according to the manufacturer's protocols. A mixture with 11.5 μl DEPC H_2O , 1 μl 20 mM random primers D2-8N(5'-AAGCTAAGACGGCGGTTTCGGNNNNNNNN-3'), 1 μl 10xRT mix, 1 μl 10 mM dNTP, and 11.5 μl extracted viral DNA was prepared to synthesize the first strand of viral DNA. Then, we denatured the enzyme of the mixture at 95°C for 5 minutes and added the Klenow fragment solution (0.15 μl 10x Klenow Buffer, 0.5 μl Klenow fragment, 0.85 μl DEPC H_2O) at 37°C . To obtain second-strand viral DNA, the procedure was repeated twice.

Shotgun sequencing of viromes and bulk metagenomes

For bulk metagenomes, total metagenomic DNA was extracted from approximately 170 mg of feces using standard methods (22). For each bulk or virome DNA sample, we prepared a library using the NEB Next[®] Ultra[™] DNA Library Prep Kit for Illumina (NEB, USA). Briefly, the fresh genomic DNA samples were mechanically fragmented by sonication to a size of approximately 350 bp. After end-polished, A-tailed, and ligated with the full-length adapter, the DNA fragments were then amplified using PCR. The AMPure XP system (Beckman Coulter, Beverly, USA) was used to purify the PCR products. After that, the DNA libraries were shotgun sequenced based on the Illumina NovaSeq platform, which generated raw 2×150 bp paired-end reads for further analysis. Initial base calling of the metagenomic dataset was processed using the sequencing platform's system default parameters. The raw paired-end reads for each sample were independently processed for quality control using fastp (23). Fastp processed with the raw reads by trimming the low-quality ($Q < 30$) bases at the end of reads and filtering 'N'-containing (> 3 'N' bases), adapter-contaminated or short length (< 90 bp) reads, to generate the high-quality reads. The human reads were removed from the high-quality reads set based on their alignment to the human reference genome (GRCh38) using Bowtie2 (24).

Identification and clustering of viral sequences

High-quality clean reads for each bulk or virome metagenomic sample were performed for *de novo* assembly via MEGAHIT (25) with a broad range of k-mer sizes ($-k$ -list 21,41,61,81,101,121,141). For bulk metagenome samples, all assembled contigs with a length ≥ 2 kbp were firstly assessed by using CheckV (26), and the non-viral contigs were removed if their viral gene count was less than the number of microbial genes. Then, we identified potential viral sequences from the remaining contigs based on two criteria: 1) contigs with P -value < 0.01 and score > 0.90 in DeepVirFinder (27); and 2) contigs identified as viruses by VIBRANT (28) with default options (-meta mode). Low-quality or "not-determined" viral contigs assessed by CheckV were further removed to avoid contamination. For virome metagenome samples, we identified potential viral sequences from the assembled contigs with length ≥ 2 kbp based on the following criteria: 1) contig whose viral gene was more than the number of microbial genes in CheckV (26); 2) contig with P -value < 0.01 and score > 0.90 in DeepVirFinder (27); and 3) contig identified as a virus by VIBRANT (28) with default options (-virome mode). According to the previous study (29), we searched for bacterial universal single-copy orthologs (BUSCO) (30) within viral sequences using hmmsearch (31)

with the default options and calculated the ratio of the number of BUSCO to the total number of genes in each viral sequence (referred to as the BUSCO ratio). After removing highly contaminated viral sequences with a $\geq 5\%$ BUSCO ratio, the remaining viral sequences were considered the final viral sequences for each sample.

The viral sequences were de-replicated using the following procedures: 1) all viral sequences were aligned pairwise using BLASTN with the options '-evalue 1e-10 -word_size 20 -num_alignments 99999'; 2) viral sequences that shared 95% nucleotide identity across 75% of the sequence were clustered into a viral operational taxonomic unit (vOTU) using in-house scripts.

Taxonomy assignment and host prediction of viruses

Viral protein-coding genes were called from the viral sequences using Prodigal (32). Taxonomic annotation of viral sequences was carried out based on protein sequence alignment to the combined database derived from the Virus-Host DB (downloaded in May 2021) (33), *crAss-like* protein sequences from Guerin's study (34), and viral protein sequences from Benler's study (35). To implement accurate family-level taxonomy, we first aligned proteins of viral sequences from NCBI RefSeq against the combined database using DIAMOND (36) with the parameters '-query-cover 50 -subject-cover 50 -id 30 -min-score 50 -max-target-seqs 10'. A viral sequence was annotated to the viral family level when over a quarter of its proteins were matched to the same family.

The virus-host prediction was performed using two bioinformatic methods that included prophage prediction and clustered regularly interspaced short palindromic repeats (CRISPR)-spacer matches. For prophage prediction, the viral sequence was blasted against the gut prokaryotic genes from the comprehensive unified human gastrointestinal genome (UHGG) database (37), and a host was assigned if the viral sequence matched the host genome at 90% nucleotide identity and 30% viral coverage (29). For CRISPR-spacer matches, we first predicted CRISPR spacer sequences from the UHGG genomes using MinCED (38) with the option '-minNR 2', and then assigned a host to the virus if the host CRISPR spacer sequence was matched to the viral genome (bit-score ≥ 45) using BLASTN with options '-evalue 1e-5 -word_size 8 -num_alignments 99999' (29).

Analyses of the gut bacteriome

The bacteriome composition of each fecal sample was profiled based on the UHGG database (37), which comprised 204,938 nonredundant genomes from 4,644 gut prokaryotes. The bulk metagenomic reads for samples were aligned against

the UHGG genomes to generate the gut bacteriome profiles. Relative abundances of 4,644 prokaryotic species were calculated by normalizing for each sample, and the relative abundances at the phylum and genera levels were obtained by summing the abundances of species from the same taxa.

Statistical analyses

Statistical analyses were implemented on the R platform (<https://www.r-project.org/>). The Shannon diversity index for the gut viral composition was calculated based on the relative abundance profile at the species level using the *vegan* package in the R platform. The *adonis* function of the *vegan* package was used to conduct the permutational multivariate analysis of variance (PERMANOVA), and the *adonis* *P*-value was calculated based on 1,000 permutations. Effect size analysis of the microbiome and virome profiles was performed following the previous study (39). Distance-based redundancy analysis (dbRDA) was performed on normalized taxa abundance matrices with R *vegan* package according to Bray-Curtis distances. Random forest models were trained using the *randomForest* package (1,000 trees) to distinguish between SLE patients and healthy controls. The performance of the predictive model was evaluated using receiver operator characteristic (ROC) analysis, which was implemented on the R platform with the *pROC* package. The area under the receiver operator characteristic curve (AUC) was calculated from the ROC analysis. For multiple testing, *P*-values were adjusted to obtain the false discovery rate (FDR) using the Benjamini-Hochberg procedure.

Correlation analysis of the gut viruses and bacteria was performed based on Spearman's rank correlation coefficient. A correlation coefficient was calculated based on Spearman correlation coefficient $\rho > 0.6$ or < -0.6 and correlation test adjusted $P < 0.05$. The correlation network was visualized using Cytoscape (40).

Results

Overview of the gut virome in SLE patients and healthy subjects

To characterize the gut virome in patients with SLE, we performed bulk metagenomic sequencing and VLP-based viral metagenomic sequencing on fecal samples from 16 SLE patients and 31 healthy controls. Bulk metagenomic sequencing obtained 429.2 Gbp of high-quality data (9.1 ± 2.7 Gbp per sample; Table S2). A total of 780,695 contigs (length ≥ 2 kbp) were generated after *de novo* assembly for each sample, and 2,449 of these contigs were recognized as highly credible viral sequences based on their sequence features and homology to known viral

genomes (see Methods). The average length of the contigs was 36,669 bp, with a minimum and maximum length of 2,126 and 346,124 bp, respectively. 16.4% of viral sequences were assessed as complete viral genomes based on the CheckV algorithm (26), and the remaining 32.6% and 51.0% of sequences were assessed as high-quality (completeness > 90%) and medium-quality (completeness > 50%) viruses, respectively (Figure 1A).

VLP-based viral metagenomic sequencing obtained 219.1 Gbp of high-quality non-human data (4.7 ± 1.8 Gbp per sample; Table S2). The dataset was assembled into 97,797 contigs (length ≥ 2 kbp), and 24,749 of these contigs were recognized as highly credible viral sequences. The average length of the contigs was 6,889 bp (N50 length, 10,404 bp), with a minimum and maximum length of 2,000 and 262,173 bp, respectively. Only 10.1% of viral contigs were assessed as complete (4.8%), high-quality (2.1%), or medium-quality (3.2%) viral genomes based on the CheckV algorithm (26), and the remaining 89.9% of viral sequences were low-quality or undetermined viruses (Figure 1A).

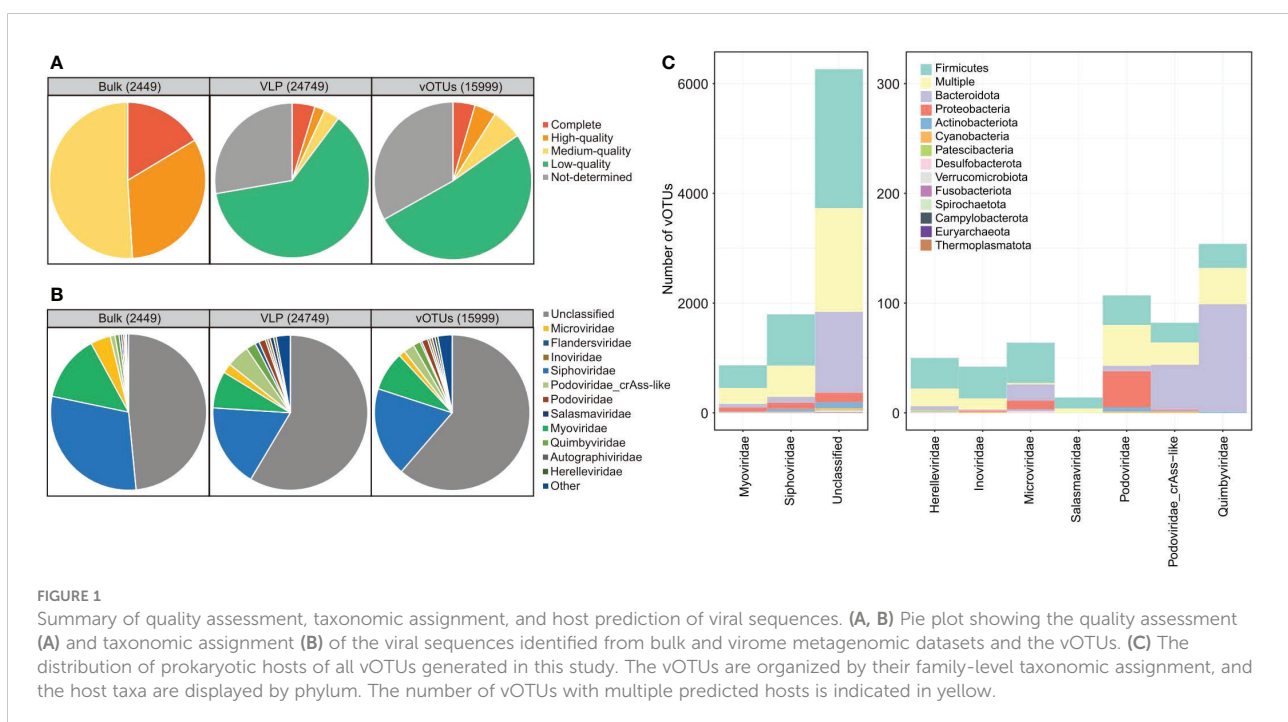
The viral contigs from both bulk and viral metagenomes were then grouped into 15,999 vOTUs by clustering the contigs at 95% nucleotide similarity. These vOTUs included 4.5% complete viral genomes, 4.4% high-quality viral genomes, 6.3% medium-quality viral genomes, and 84.8% low-quality or undetermined viruses (Figure 1A). 38.2% (6,105/15,999) of all vOTUs could be taxonomically assigned to a viral family. *Siphoviridae*, *Myoviridae*, *Microviridae*, *Quimbyviridae*, and *Podoviridae* were the dominant families for these vOTUs (Figure 1B). And the finding was consistent in the viral sequences from both bulk and viral metagenomic datasets. Furthermore, we predicted the microbial hosts for the vOTUs

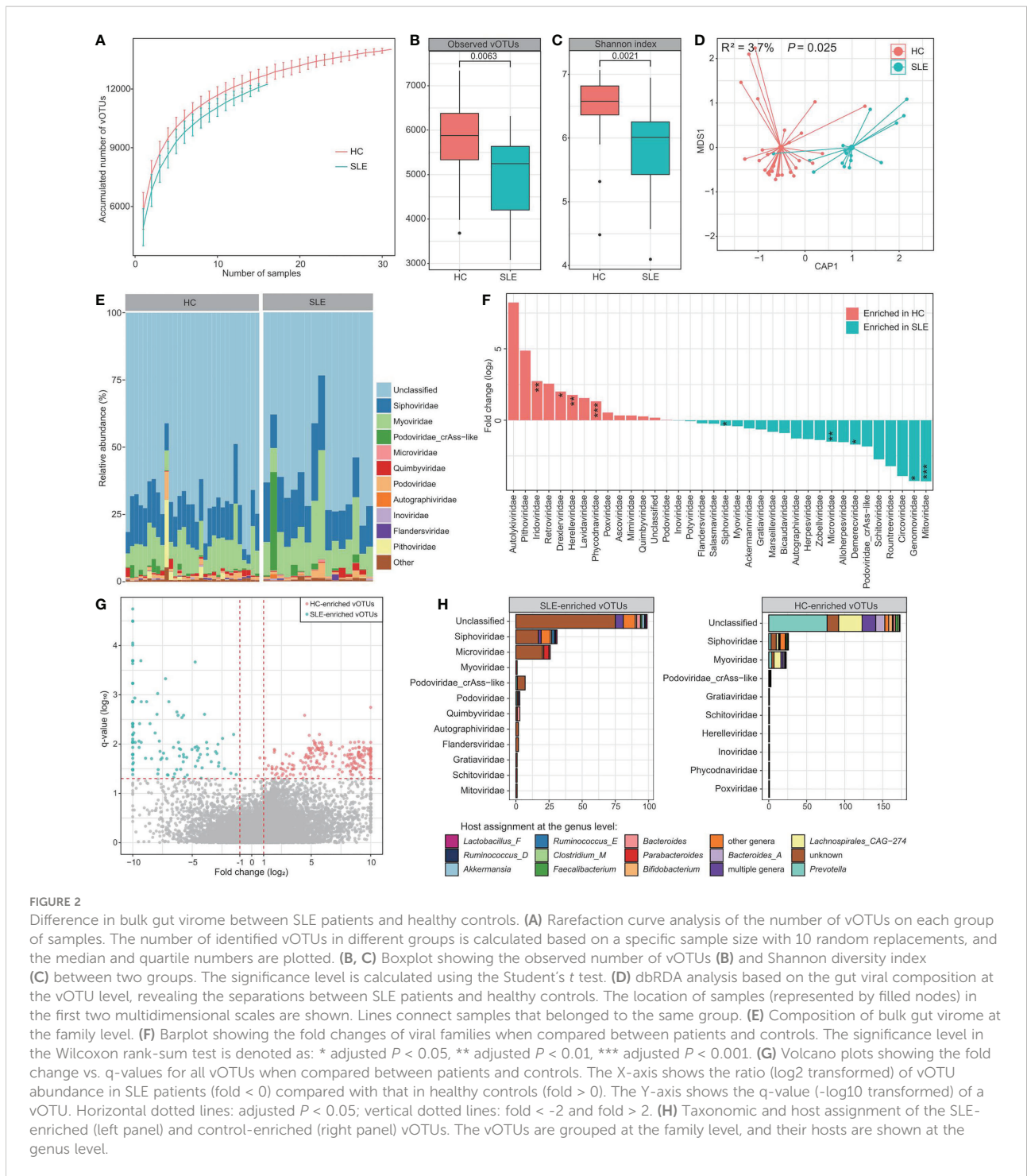
and found that 60.2% (9,636/15,999) of vOTUs could be assigned to one or more host bacteria or archaea. The hosts of these viruses were mostly members of Firmicutes, Bacteroidota, Proteobacteria, and Actinobacteriota (Figure 1C).

Comparison of the viromes between patients and controls in the bulk metagenomic dataset

To explore signatures of the gut virome in SLE, first, we compared the viral profiles between the SLE patients and healthy controls based on the bulk metagenomic dataset. Rarefaction analysis showed that, at the same sample size, the number of vOTUs observed in SLE patients was significantly lower than that in healthy controls (Figure 2A). Consistently, within-sample alpha diversity analysis revealed that both the viral richness (estimated by the observed number of vOTUs) and diversity (estimated by the Shannon index) had significantly decreased in SLE patients compared with controls (Figures 2B, C). Next, dbRDA analysis based on the Bray-Curtis distance at the vOTUs level revealed a clear separation between the two groups (Figure 2D), with the disease state explained an effect size of 3.7% (PERMANOVA $p=0.025$) on the gut virome variations.

Most of the gut virome of both SLE patients and healthy controls could not be assigned to known viral families (average relative abundance, $65.2\% \pm 4.5\%$; Figure 2E). Apart from this, *Siphoviridae* ($11.1\% \pm 1.0\%$), *Myoviridae* ($7.3\% \pm 0.8\%$), *crAss-like* [$5.7\% \pm 1.8\%$; a *Podoviridae* subclade with unique genomic and functional features (41, 42)], and *Microviridae* ($3.6\% \pm$





1.2%) were the most dominant families in the viromes of all investigated samples, followed by a small proportion of *Podoviridae*, *Quimbyviridae*, *Autographiviridae*, and *Inoviridae* (Figure 2E). Using the Wilcoxon rank-sum test, we identified 9 families with significant differences in their relative abundances between the patients and controls (adjusted *P* < 0.05). 4 of these families, including *Iridoviridae*, *Drexlerviridae*, *Herelleviridae*,

and *Phycodnaviridae*, were significantly enriched in the virome of SLE patients, whereas 5 families, including *Siphoviridae*, *Mitoviridae*, *Genomoviridae*, *Demereciviridae*, and *Microviridae*, were enriched in the virome of healthy controls (Figure 2F). At the vOTUs level, 408 vOTUs were identified with significant differences in their relative abundances between patients and controls (Wilcoxon rank-sum test adjusted *P* < 0.05). Among

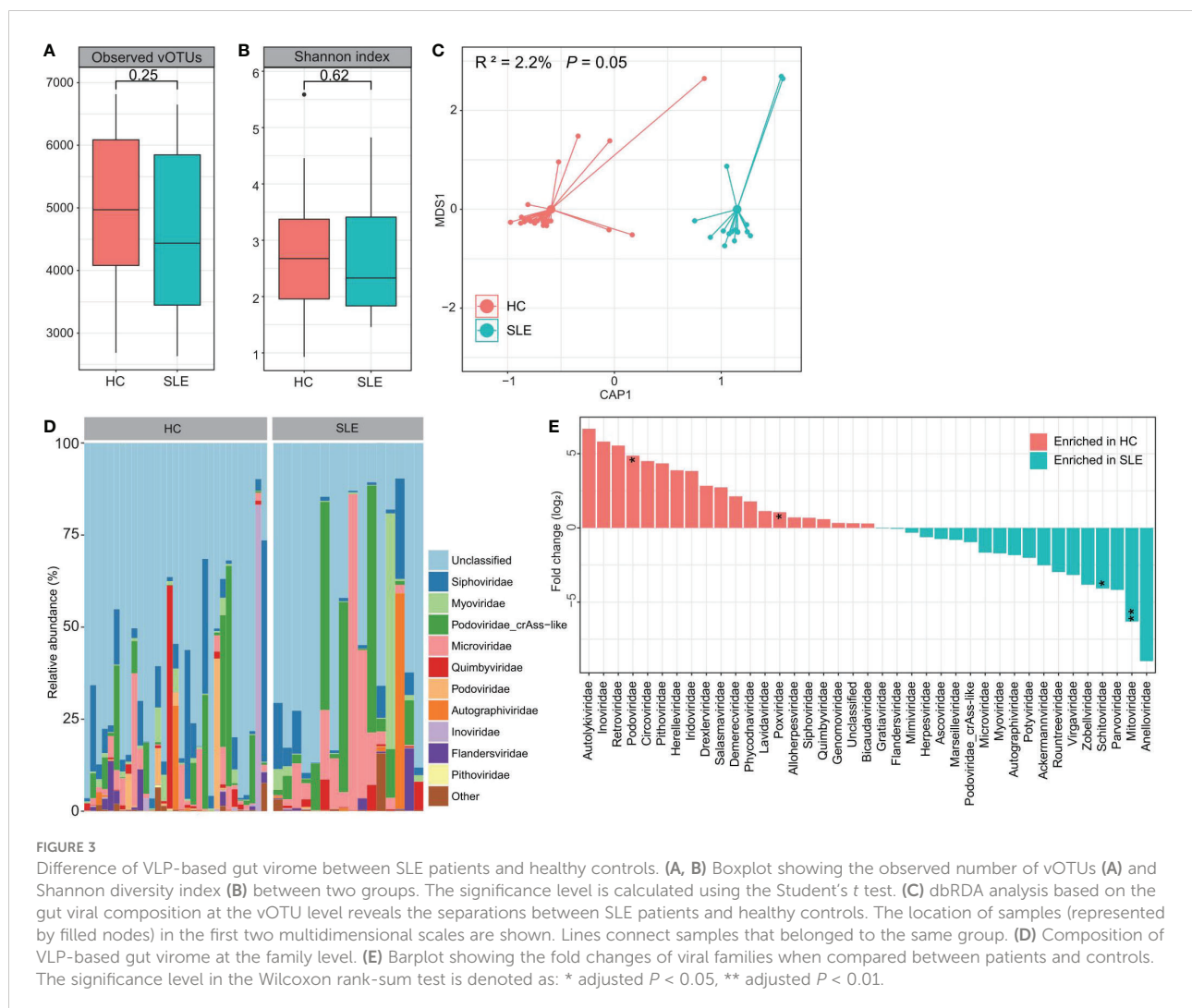
these, 177 vOTUs were significantly enriched in SLE patients, and 231 vOTUs were depleted (Figure 2G; Table S3). The SLE-enriched vOTUs included 31 members of *Siphoviridae*, 26 *Microviridae*, 7 *crAss-like*, 3 *Podoviridae*, 3 *Quimbyviridae*, and 120 unclassified viruses, while the control-enriched vOTUs were composed of 26 *Siphoviridae*, 23 *Myoviridae*, and 173 unclassified viruses (Figure 2H). SLE-enriched vOTUs were frequently assigned to phages of *Bacteroides*, *Bacteroides_A*, *Parabacteroides*, and *Ruminococcus_E*, while the control-enriched vOTUs had remarkably higher proportions of *Prevotella* and *Lachnospirales_CAG-274* phages.

Comparison of the viromes between patients and controls in VLP-based viral metagenomic dataset

Next, we compared the gut viromes between the SLE patients and healthy controls based on the VLP metagenomic

dataset to investigate the signatures of free viral particles. In VLP metagenomic samples, there was no significant difference in the observed number of vOTUs (viral richness) or Shannon diversity index between SLE patients and healthy subjects (Figures 3A, B). However, similar to the bulk metagenomes, dbRDA analysis based on the Bray-Curtis distance of viral metagenomes at the vOTU level also revealed visible separation between the two groups (effect size = 2.2%, PERMANOVA $P = 0.05$; Figure 3C), suggesting a considerable gut viral dysbiosis in these SLE patients.

At the family level, a large proportion of the virome in both groups was assigned to unclassified viruses (average relative abundance = $62.8\% \pm 7.4\%$; Figure 3D). *crAss-like* ($9.8\% \pm 3.0\%$), *Siphoviridae* ($7.6\% \pm 1.4\%$), *Microviridae* ($7.0\% \pm 2.1\%$), *Myoviridae* ($2.6\% \pm 0.9\%$), *Quimbyviridae* ($2.3\% \pm 0.8\%$), and *Autographiviridae* ($2.0\% \pm 0.9\%$) were the most dominant families in the remaining viral communities of all investigated samples. The Wilcoxon rank-sum test showed that *Mitoviridae* and *Schitoviridae* were significantly enriched in the



gut virome of SLE patients compared with that of controls, while *Podoviridae* and *Poxviridae* were enriched in healthy controls (Figure 3E). At the vOTU level, we observed significant differences in the abundances of 18 vOTUs between the SLE patients and healthy controls (Wilcoxon rank-sum test adjusted $P < 0.05$; Table S4). Seventeen of these vOTUs, including 3 members of the *Siphoviridae*, 1 *Autographiviridae*, 1 *Flandersviridae*, 1 *Myoviridae*, 1 *Podoviridae*, and 10 unclassified viruses, were enriched in patients, and only one unclassified virus was enriched in the control group.

Correlation analysis of gut viruses and bacteria

To study the virus-bacterium correlations, first, we performed a PERMANOVA-based analysis to estimate the effect size of variances between the gut virome and bacteriome. The analysis showed that the gut bacteriome accounted for a considerable proportion of the variance of many vOTUs in both the bulk and VLP-based virome datasets. For example, the bacteriome explained over 15% of the variances for 6 vOTUs

in the bulk virome and over 10% of the variances for 3 vOTUs in the VLP-based virome (Figure 4A). On the other hand, several bacterial genera, including *Prevotella*, *CAG-611*, and *Bacillus_A*, had the largest effects on the gut virome (Figure 4B).

Next, we calculated the correlation coefficients between 426 SLE-associated vOTUs (408 identified from bulk metagenomes and 18 from VLP-based viral metagenomes) and 841 bacterial genera. This analysis identified a total of 189 significant correlations between 106 vOTUs and 56 bacterial genera (Figure 4C). A large proportion of vOTUs was positively correlated with *Lachnospirales_CAG-274* and *Hungatella_A*, suggesting their key roles in the virus-bacterium network.

Random forest model predicts SLE state

To assess the discriminating effect of the gut virome on SLE status, we used all vOTUs as predictors of the random forest model. The model obtained an AUC of 0.95 (95% confidence interval [CI], 0.85-1.00; Figure 5A) in distinguishing healthy patients from SLE patients. Retraining the model using 426 SLE-associated vOTUs also showed that the prediction effect is

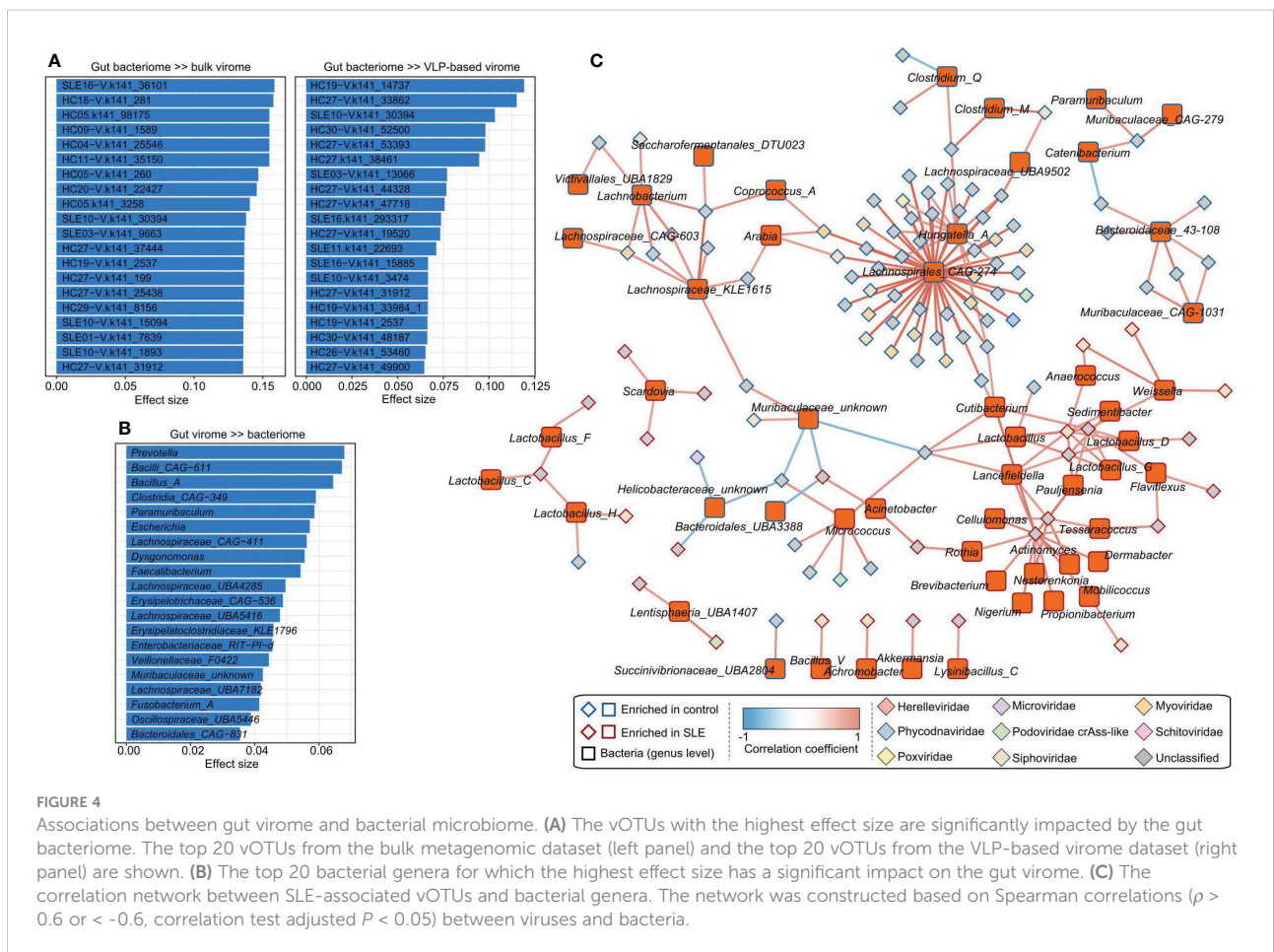


FIGURE 4 Associations between gut virome and bacterial microbiome. (A) The vOTUs with the highest effect size are significantly impacted by the gut bacteriome. The top 20 vOTUs from the bulk metagenomic dataset (left panel) and the top 20 vOTUs from the VLP-based virome dataset (right panel) are shown. (B) The top 20 bacterial genera for which the highest effect size has a significant impact on the gut virome. (C) The correlation network between SLE-associated vOTUs and bacterial genera. The network was constructed based on Spearman correlations ($\rho > 0.6$ or < -0.6 , correlation test adjusted $P < 0.05$) between viruses and bacteria.

slightly improved with an AUC of 0.98 (95% CI, 0.95-1.00). Moreover, to minimize the number of viruses used in the model, we selected a subset of vOTUs to train the random forest model based on the importance of all vOTUs. This analysis showed that the model obtained the highest AUC (0.998) when using a subset of 39 top important vOTUs (Figure 5B). These findings suggest the high diagnostic potential of the gut virome in SLE discrimination.

Discussion

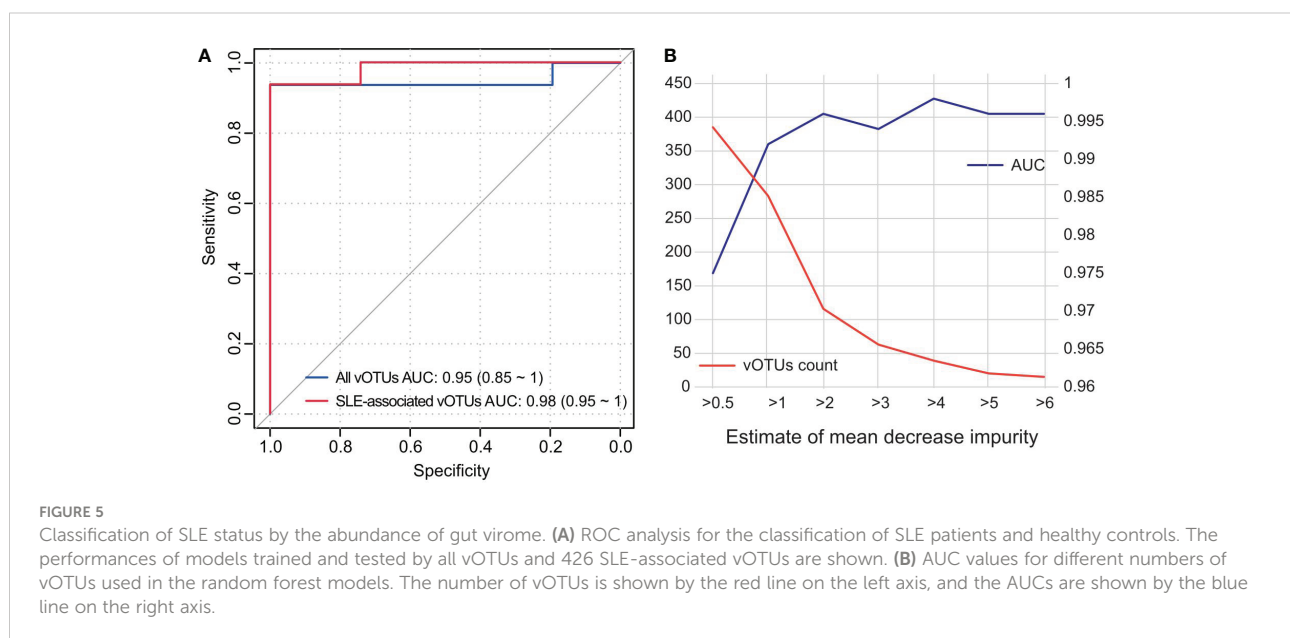
In this study, we revealed the changes in gut viral populations in 16 SLE patients compared with 31 healthy controls. Our study strengthened the previous bulk metagenome-based study on the gut virome of autoimmune diseases (including SLE) (14) by adding more information from both bulk and VLP metagenomic sequencing technologies. VLP-based virome sequencing has improved the diversity of viruses observed, allowing information on previously overlooked viruses to be captured (21).

In the bulk metagenomic dataset, we identified 4 viral families, including *Iridoviridae*, *Drexlerviridae*, *Herelleviridae*, and *Phycodnaviridae*, that were significantly enriched in the SLE virome, whereas 5 families, including *Siphoviridae*, *Mitoviridae*, *Genomoviridae*, *Demereciviridae*, and *Microviridae* were enriched in healthy controls. *Iridoviridae*, a family of large and icosahedral viruses that are known to infect ectothermic vertebrates such as bony fish, amphibians, and reptiles (43), have frequently been found in human fecal or blood samples (44–47) and reported as one of the most abundant viral families in the gut of human immunodeficiency virus (HIV) infected

patients (45). Viral phages belonging to the *Drexlerviridae* family, such as KM18 and IME268, have cleavage activity against *Klebsiella pneumoniae* due to their putative endosialidase (depolymerase) enzyme activity (48, 49), and more relevantly, *Klebsiella* was reported significantly enriched in SLE patients from China (50). *Herelleviridae* is a bacterial virus that infects members of the Firmicutes, especially the *Enterococcus* genus (51). *Phycodnaviridae* is a large icosahedral dsDNA viral family that can infect algae (52), and it has also been found in the oropharyngeal samples obtained from healthy adults (53). The enrichment of these families in SLE patients suggested that they may play potential roles in subjects with an immunocompromised state.

By comparing the viromes between patients and controls in the VLP-based viral metagenomic dataset, we found that *Mitoviridae* and *Schitoviridae* were significantly enriched in the gut virome of SLE patients compared with that of controls, while *Podoviridae* and *Poxviridae* were enriched in healthy controls. *Schitoviridae* was a new family of N4-like phages, and there was a poorly reported association between *Schitoviridae* and human disease (54). Analysis of phages and autoimmune disease (specifically the patients with RA and SLE) in 476 Japanese showed that *Podoviridae* were significantly decreased in the gut of the patients with SLE, and it has a symbiotic relationship with *Faecalibacterium* (14). These results were consistent with our studies.

In the viral host aspect, we found the viruses that are predicted to infect *Bacteroides*, *Bacteroides_A*, *Parabacteroides*, and *Ruminococcus_E* were enriched in the SLE patients. Consistently, Chen et al. found that *Bacteroides fragilis* was enriched in the SLE gut microbiota and reduced after treatment (7). In a cross-sectional cohort study, Azzouz et al.



found that the relative abundance of *Ruminococcus gnavus* in patients with higher SLE disease activity index was overall 5-fold greater than that in healthy subjects (55). These findings suggest that the abundance of bacteria may also affect phages by providing their hosts.

Notably, different research strategies based on bulk and VLP metagenomic datasets have their pros and cons. The enriched VLPs could improve the sequencing depth for viruses, while the bulk metagenome is easier to preprocess and contains bacterial information, which is important for the analysis of the relationship between bacteria and viruses. Until now, many attempts to screen SLE biomarkers have been made. Based on bacterial 16S rRNA gene (56) and various immune cells (57), 0.79 and 0.96 of AUCs were achieved, respectively. We obtained a more accurate and higher AUC of 0.98 (Figure 5) using only 426 gut vOTUs. This analysis showed that the model obtained the highest AUC (0.998) when using a subset of 39 top important vOTUs. These findings are encouraging developments that suggest the high diagnostic potential of the gut virome in SLE discrimination.

A major limitation of this study was the small sample size (16 SLE patients vs. 31 healthy controls). Future studies with a larger cohort of patients and controls will be needed to further uncover the relationship between the gut viral community and SLE. On the other hand, due to the lack of comprehensive reference databases, the majority of viruses in the human gut remain unknown. Many unclassified viruses will be taxonomically classified into known or newly-identified taxa, which will result in more accurate descriptions of the virome characterization of SLE patients.

Conclusion

Overall, based on bulk and VLP-based shotgun metagenomic sequencing datasets, our results systematically characterized the gut virome in SLE patients. Some viral signatures had significantly different abundances between the SLE patients and healthy subjects. Importantly, the excellent predictive model (AUC >0.95) using only a small number of viruses heralded its potential for clinical application. Our research will bring the revelation for future mechanistic and clinical intervention studies.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ebi.ac.uk/ena>, PRJEB55711.

Ethics statement

The studies involving human participants were reviewed and approved by the Medical Ethics Committees of the Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine. The patients/participants provided their written informed consent to participate in this study.

Author contributions

CC, QY, XY, WS and WM contributed to conception and design of the study. CC, QY and SL drafted the manuscript. QZho, FT, ZL, YH, CL, DZ and HL collected the samples and information. SL, QL, GW, AZ, HU, YZ, YA, JZ and QZha performed the data analysis and investigation. All authors revised the manuscript, contributed to the article, and approved the submitted version.

Funding

This work was supported by grants from the Science and Technology Program of Guizhou Province (Guizhou Scientific Foundation-Platform and talent [2020]2202 and [2016]5650) and the National Natural Science Foundation of China (81902037 and 82260894).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.1050895/full#supplementary-material>

References

- Zucchi D, Elefante E, Schilirò D, Signorini V, Trentin F, Bortoluzzi A, et al. One year in review 2022: systemic lupus erythematosus. *Clin Exp Rheumatol* (2022) 40(1):4–14. doi: 10.55563/clinexprheumatol/nolysy
- Picascia A, Grimaldi V, Pignalosa O, De Pascale MR, Schiano C, Napoli C. Epigenetic control of autoimmune diseases: from bench to bedside. *Clin Immunol* (2015) 157(1):1–15. doi: 10.1016/j.clim.2014.12.013
- Tsokos GC. Autoimmunity and organ damage in systemic lupus erythematosus. *Nat Immunol* (2020) 21(6):605–14. doi: 10.1038/s41590-020-0677-6
- Hamilton JA, Hsu H-C, Mountz JD. Autoreactive b cells in SLE, villains or innocent bystanders? *Immunol Rev* (2019) 292(1):120–38. doi: 10.1111/imr.12815
- Ma Y, Xu X, Li M, Cai J, Wei Q, Niu H. Gut microbiota promote the inflammatory response in the pathogenesis of systemic lupus erythematosus. *Mol Med* (2019) 25(1):35. doi: 10.1186/s10020-019-0102-5
- Choi SC, Brown J, Gong M, Ge Y, Zadeh M, Li W, et al. Gut microbiota dysbiosis and altered tryptophan catabolism contribute to autoimmunity in lupus-susceptible mice. *Sci Transl Med* (2020) 12(551):eaax2220. doi: 10.1126/scitranslmed.aax2220
- Chen BD, Jia XM, Xu JY, Zhao LD, Ji JY, Wu BX, et al. An autoimmunogenic and proinflammatory profile defined by the gut microbiota of patients with untreated systemic lupus erythematosus. *Arthritis Rheumatol* (2021) 73(2):232–43. doi: 10.1002/art.41511
- Hevia A, Milani C, Lopez P, Cuervo A, Arboleya S, Duranti S, et al. Intestinal dysbiosis associated with systemic lupus erythematosus. *mBio* (2014) 5(5):e01548–01514. doi: 10.1128/mBio.01548-14
- Santiago-Rodríguez TM, Hollister EB. Human virome and disease: High-throughput sequencing for virus discovery, identification of phage-bacteria dysbiosis and development of therapeutic approaches with emphasis on the human gut. *Viruses* (2019) 11(7):656. doi: 10.3390/v11070656
- De Sordi L, Lourenço M, Debarbieux L. The battle within: Interactions of bacteriophages and bacteria in the gastrointestinal tract. *Cell Host Microbe* (2019) 25(2):210–8. doi: 10.1016/j.chom.2019.01.018
- Shkoporov AN, Hill C. Bacteriophages of the human gut: The "Known unknown" of the microbiome. *Cell Host Microbe* (2019) 25(2):195–209. doi: 10.1016/j.chom.2019.01.017
- Ma Y, You X, Mai G, Tokuyasu T, Liu C. A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* (2018) 6(1):24. doi: 10.1186/s40168-018-0410-y
- 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):764–778 e765. doi: 10.1016/j.chom.2019.10.009
- Tomofuji Y, Kishikawa T, Maeda Y, Ogawa K, Nii T, Okuno T, et al. Whole gut virome analysis of 476 Japanese revealed a link between phage and autoimmune disease. *Ann Rheum Dis* (2022) 81(2):278–88. doi: 10.1136/annrheumdis-2021-221267
- Keen EC, Dantas G. Close encounters of three kinds: Bacteriophages, commensal bacteria, and host immunity. *Trends Microbiol* (2018) 26(11):943–54. doi: 10.1016/j.tim.2018.05.009
- Li Y, Handley SA, Baldrige MT. The dark side of the gut: Virome-host interactions in intestinal homeostasis and disease. *J Exp Med* (2021) 218(5):e20201044. doi: 10.1084/jem.20201044
- Shkoporov AN, Clooney AG, Sutton TDS, Ryan FJ, Daly KM, Nolan JA, et al. The human gut virome is highly diverse, stable, and individual specific. *Cell Host Microbe* (2019) 26(4):527–541 e525. doi: 10.1016/j.chom.2019.09.009
- Coughlan S, Das A, O'Herlihy E, Shanahan F, O'Toole PW, Jeffery IB. The gut virome in irritable bowel syndrome differs from that of controls. *Gut Microbes* (2021) 13(1):1–15. doi: 10.1080/19490976.2021.1887719
- Garmaeva S, Gulyaeva A, Sinha T, Shkoporov AN, Clooney AG, Stockdale SR, et al. Stability of the human gut virome and effect of gluten-free diet. *Cell Rep* (2021) 35(7):109132. doi: 10.1016/j.celrep.2021.109132
- Aringer M, Costenbader K, Daikh D, Brinks R, Mosca M, Ramsey-Goldman R, et al. 2019 European league against Rheumatism/American college of rheumatology classification criteria for systemic lupus erythematosus. *Arthritis Rheumatol* (2019) 71(9):1400–12. doi: 10.1002/art.40930
- Yan Q, Wang Y, Chen X, Jin H, Wang G, Guan K, et al. Characterization of the gut DNA and RNA viromes in a cohort of Chinese residents and visiting pakistanis. *Virus Evol* (2021) 7(1):veab022. doi: 10.1093/ve/veab022
- Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* (2012) 490(7418):55–60. doi: 10.1038/nature11450
- Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* (2018) 34(17):i884–90. doi: 10.1093/bioinformatics/bty560
- Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods* (2012) 9(4):357–9. doi: 10.1038/nmeth.1923
- Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph. *Bioinformatics* (2015) 31(10):1674–6. doi: 10.1093/bioinformatics/btv033
- 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(5):578–85. doi: 10.1038/s41587-020-00774-7
- Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* (2017) 5(1):69. doi: 10.1186/s40168-017-0283-5
- 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(1):90. doi: 10.1186/s40168-020-00867-0
- 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(5):724–740 e728. doi: 10.1016/j.chom.2020.08.003
- Seppy M, Manni M, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol Biol* (2019) 1962:227–45. doi: 10.1007/978-1-4939-9173-0_14
- Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* (2011) 7(10):e1002195. doi: 10.1371/journal.pcbi.1002195
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf* (2010) 11:119. doi: 10.1186/1471-2105-11-119
- Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, et al. Linking virus genomes with host taxonomy. *Viruses* (2016) 8(3):66. doi: 10.3390/v8030066
- 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):653–664 e656. doi: 10.1016/j.chom.2018.10.002
- Benler S, Yutin N, Antipov D, Rayko M, Shmakov S, Gussow AB, et al. Thousands of previously unknown phages discovered in whole-community human gut metagenomes. *Microbiome* (2021) 9(1):78. doi: 10.1186/s40168-021-01017-w
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* (2015) 12(1):59–60. doi: 10.1038/nmeth.3176
- Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* (2021) 39(1):105–114. doi: 10.1038/s41587-020-0603-3
- Skenneron C. *Mined-mining CRISPRs in environmental datasets*. Github (2016). Available at: <https://github.com/ctSkenneron/minced>.
- Wang X, Yang S, Li S, Zhao L, Hao Y, Qin J, et al. Aberrant gut microbiota alters host metabolism and impacts renal failure in humans and rodents. *Gut* (2020) 69(12):2131–42. doi: 10.1136/gutjnl-2019-319766
- Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with cytoscape 3. *Curr Protoc Bioinf* (2014) 47:8 13 11–24. doi: 10.1002/0471250953.bi0813s47
- 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. doi: 10.1038/s41467-021-21350-w
- Gulyaeva A, Garmaeva S, Ruigrok RA, Wang D, Riksen NP, Netea MG, et al. Discovery, diversity, and functional associations of crAss-like phages in human gut metagenomes from four Dutch cohorts. *Cell Rep* (2022) 38(2):110204. doi: 10.1016/j.celrep.2021.110204
- Chinchar VG, Hick P, Ince IA, Jancovich JK, Marschang R, Qin Q, et al. ICTV virus taxonomy profile: Iridoviridae. *J Gen Virol* (2017) 98(5):890–1. doi: 10.1099/jgv.0.000818
- Norman JM, Handley SA, Baldrige 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. doi: 10.1016/j.cell.2015.01.002
- Monaco CL, Gootenberg DB, Zhao G, Handley SA, Ghebremichael MS, Lim ES, et al. Altered virome and bacterial microbiome in human immunodeficiency virus-associated acquired immunodeficiency syndrome. *Cell Host Microbe* (2016) 19(3):311–22. doi: 10.1016/j.chom.2016.02.011
- Ngoi CN, Siqueira J, Li L, Deng X, Mugo P, Graham SM, et al. The plasma virome of febrile adult kenyans shows frequent parvovirus B19 infections and a

novel arbovirus (Kadipiro virus). *J Gen Virol* (2016) 97(12):3359–67. doi: 10.1099/jgv.0.000644

47. Altindis E, Cai W, Sakaguchi M, Zhang F, GuoXiao W, Liu F, et al. Viral insulin-like peptides activate human insulin and IGF-1 receptor signaling: A paradigm shift for host-microbe interactions. *Proc Natl Acad Sci U.S.A.* (2018) 115(10):2461–6. doi: 10.1073/pnas.1721117115

48. Ku H, Kabwe M, Chan HT, Stanton C, Petrovski S, Batinovic S, et al. Novel drexlerviridae bacteriophage KMI8 with specific lytic activity against *klebsiella michiganensis* and its biofilms. *PLoS One* (2021) 16(9):e0257102. doi: 10.1371/journal.pone.0257102

49. Nazir A, Qi C, Shi N, Gao X, Feng Q, Qing H, et al. Characterization and genomic analysis of a novel drexlervirial bacteriophage IME268 with lytic activity against *klebsiella pneumoniae*. *Infect Drug Resist* (2022) 15:1533–46. doi: 10.2147/idr.S347110

50. He Z, Shao T, Li H, Xie Z, Wen C. Alterations of the gut microbiome in Chinese patients with systemic lupus erythematosus. *Gut Pathog* (2016) 8:64. doi: 10.1186/s13099-016-0146-9

51. Barylski J, Kropinski AM, Alikhan NF, Adriaenssens EM, Ictv Report C. ICTV virus taxonomy profile: Herelleviridae. *J Gen Virol* (2020) 101(4):362–3. doi: 10.1099/jgv.0.001392

52. Wilson WH, Van Etten JL, Allen MJ. The phycodnaviridae: the story of how tiny giants rule the world. *Curr Top Microbiol Immunol* (2009) 328:1–42. doi: 10.1007/978-3-540-68618-7_1

53. Yolken RH, Jones-Brando L, Dunigan DD, Kannan G, Dickerson F, Severance E, et al. Chlorovirus ATCV-1 is part of the human oropharyngeal virome and is associated with changes in cognitive functions in humans and mice. *Proc Natl Acad Sci U.S.A.* (2014) 111(45):16106–11. doi: 10.1073/pnas.1418895111

54. Wittmann J, Turner D, Millard AD, Mahadevan P, Kropinski AM, Adriaenssens EM. From orphan phage to a proposed new family—the diversity of N4-like viruses. *Antibiotics (Basel)* (2020) 9(10):663. doi: 10.3390/antibiotics9100663

55. Azzouz D, Omarbekova A, Heguy A, Schwudke D, Gisch N, Rovin BH, et al. Lupus nephritis is linked to disease-activity associated expansions and immunity to a gut commensal. *Ann Rheumatic Dis* (2019) 78(7):947–56. doi: 10.1136/annrheumdis-2018-214856

56. Li Y, Wang HF, Li X, Li HX, Zhang Q, Zhou HW, et al. Disordered intestinal microbes are associated with the activity of systemic lupus erythematosus. *Clin Sci (Lond)* (2019) 133(7):821–38. doi: 10.1042/CS20180841

57. Zhong Y, Zhang W, Hong X, Zeng Z, Chen Y, Liao S, et al. Screening biomarkers for systemic lupus erythematosus based on machine learning and exploring their expression correlations with the ratios of various immune cells. *Front Immunol* (2022) 13:873787. doi: 10.3389/fimmu.2022.873787

COPYRIGHT

© 2023 Chen, Yan, Yao, Li, Lv, Wang, Zhong, Tang, Liu, Huang, An, Zhou, Zhang, Zhang, Ullah, Zhang, Liu, Zhu, Li, Sun and Ma. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.