Temporal Landscape of Human Gut RNA and DNA Viromes in SARS-CoV-2 Infection and Severity

Tao Zuo
  The Chinese University of Hong Kong

Qin Liu
  The Chinese University of Hong Kong

Fen Zhang
  The Chinese University of Hong Kong

Yun Kit Yeoh
  The Chinese University of Hong Kong

Yating Wan
  The Chinese University of Hong Kong

Hui Zhan
  The Chinese University of Hong Kong

Grace C.Y. Lui
  The Chinese University of Hong Kong

Amy Y.L. Li
  The Chinese University of Hong Kong

Chun Pan Cheung
  The Chinese University of Hong Kong

Nan Chen
  The Chinese University of Hong Kong

Wenqi Lv
  The Chinese University of Hong Kong

Rita W.Y. Ng
  The Chinese University of Hong Kong

Eugene Y.K. Tso
  United Christian Hospital

Kitty S.C. Fung
  United Christian Hospital

Veronica Chan
  United Christian Hospital

Lowell Ling
  The Chinese University of Hong Kong

Gavin Joynt

The Chinese University of Hong Kong

David S.C. Hui
The Chinese University of Hong Kong

Francis K.L. Chan
The Chinese University of Hong Kong

Paul K.S. Chan
The Chinese University of Hong Kong

Siew Ng
Chinese University of Hong Kong

Research

Keywords: RNA and DNA, Coronavirus Disease 2019, infection, chlorotic spot virus

---

Abstract

Background:

Coronavirus Disease 2019 (COVID-19) caused by the enveloped RNA virus SARS-CoV-2 primarily affects the respiratory and gastrointestinal tracts. SARS-CoV-2 was isolated from faecal samples and active viral replication was reported in human intestinal cells. The human gut also harbors an enormous amount of resident viruses (collectively known as virome) that play a role in regulating host immunity and pathophysiology. Understanding gut virome perturbation that underlies SARS-CoV-2 infection and severity is an unmet need.

Methods:

We enrolled 98 COVID-19 patients with varying disease severity (3 asymptomatic, 53 mild, 34 moderate, 5 severe, 3 critical) and 78 non-COVID-19 controls matched for gender and co-morbidities. All study subjects had faecal specimens sampled at inclusion. Blood specimens were sampled for COVID-19 patients at admission to test for inflammatory markers and white cell counts. Among COVID-19 cases, 37 (38%) patients had serially faecal samples collected 2 to 3 times per week from time of hospitalization until after discharge. Using shotgun metagenomics sequencing, we sequenced and profiled the faecal RNA and DNA virome respectively. We investigated alterations and longitudinal dynamics of the gut virome in association with disease severity and blood parameters.

Results:

Patients with COVID-19 showed underrepresentation of Pepper mild mottle virus (RNA virus) and multiple bacteriophage lineages (DNA viruses) and enrichment of environment-derived eukaryotic DNA viruses in faecal samples, compared to non-COVID-19 subjects. Such gut virome dysbiosis persisted up to 30 days after disease resolution. Faecal virome in SARS-CoV-2 infection harboured more stress-, inflammation- and virulence-associated gene encoding capacities including those pertaining to bacteriophage integration, DNA repair, and metabolism and virulence associated with their bacterial host. Human faecal baseline abundance of 10 virus species (1 RNA virus, Pepper chlorotic spot virus, and 9 DNA virus species) inversely correlated with disease severity of COVID-19. These viruses were also inversely associated with blood levels of pro-inflammatory proteins, white cells and neutrophils. Among the 10 COVID-19 severity-associated DNA virus species, 4 showed inverse correlation with age; 5 showed persistent lower abundance both during disease course and after disease resolution relative to non-COVID-19 subjects.

Conclusions:

Both enteric RNA and DNA viromes were perturbed in COVID-19, which prolonged even after disease resolution. Gut virome may calibrate host immunity and regulate severity to SARS-CoV-2 infection. Our observation that gut viruses inversely correlated with both severity of COVID-19 and host age partly

explains that older subjects are prone to severe and unfavorable COVID-19 outcomes. Our data altogether highlight the significance of human gut virome in COVID-19 disease course and potentially therapeutics.

Introduction

A novel RNA virus, SARS-CoV-2 coronavirus, has caused a global pandemic of Coronavirus disease 2019 (COVID-19) since its emergence in December 2019. Although most cases of COVID-19 are mild, disease severity varies substantially across patients, and severe cases can result in respiratory failure or death [1]. A significant amount of patients with COVID-19 suffered gastrointestinal (GI) symptoms such as vomiting and diarrhea [2-5]. In addition, SARS-CoV-2 viral RNA and transcriptional activity were detected in faecal samples [6-8], suggesting that the GI tract is an extra-pulmonary site for virus replication and activity.

The GI tract harbours the most abundant viruses in the human body collectively known as the gut virome [9, 10]. It encompasses both RNA and DNA viruses that chronically infect their eukaryotic (humans, animals, plants) and prokaryotic hosts (bacteria, amoeba) during a steady state[9]. The gut viral communities en masse modulate both innate and adaptive immunity and ecology of the gut microbiota[10]. This "normal" immune state varies from person to person and changes over time. Substantial variation in human immune responses is largely driven by both environmental influences [11] and infection by pathogens [9, 12]. Sequential infection or colonization with acute and chronic viruses can alter host hemostatic immune gene expression and affect response to vaccine[12]. The viral carrier state and its association with host immunophenotype have been under-studied due to a lack of apparopriate tools to detect and quantify extremely diverse members of the virome. With the advancement of high-throughput sequencing technology, the highly individual-specific viromes consisting of members causing acute disease or chronically colonizing in asymptomatic individuals are increasingly being uncovered[13, 14]. We have previously demonstrated that the gut microbiome was significantly perturbed in COVID-19 and associated with disease severity and symptoms [8, 15, 16]. Given the co-residence and co-evolutionary nature of virome and bacterial microbiome in human gut, we herein hypothesize that the gut virome in patients with COVID-19 is altered as well and human baseline virome may be associated with immune defense and severity to SARS-CoV-2 infection. We employed RNA and DNA metagenomics to profile the enormous viral members in faecal samples and correlated with disease immunophenotype to unravel host-viral relationship and potentially mechanisms underlying severity and recovery in COVID-19.

Results

Here, we enrolled 98 hospitalized patients with COVID-19 (mean age 33 years; 53% male) and 78 non-COVID-19 controls matched for gender and co-morbidities (mean age 48 years; 42% male, **Table 1**). All COVID-19 patients and non-COVID-19 controls had stool specimens sampled at inclusion. Blood specimens were additionally sampled for COVID-19 patients at admission to test for pro-inflammatory markers and white cell count (**Supplementary Table 1**). Amongst these COVID-19 patients, 37 (38%) had

serial faecal samples collected from hospitalization until after discharge **(Supplementary Figure 1).** We enriched both faecal RNA and DNA virions from a total of 277 faecal samples and performed non-targeted shotgun metagenomic sequencing on the RNA virome (mostly eukaryotic viruses) and DNA virome (mostly prokaryotic bacteriophages). We reported gut virome profiles in association with SARS-CoV-2 infection, disease severity, and blood parameters.

Table1. Clinical Characteristics of COVID-19 patients and non-COVID-19 controls

Variables	COVID-19 cases	non-COVID-19 controls
Number	99	78
Male	52 (53%)	33 (42%)
Age, years (mean±s.e.)	37 (± 2)	45 (±2)
Co-morbidities	55 (56%)	24 (31%)
Symptoms at admission		
Fever	38 (38%)	
Gastrointestinal symptoms		
Diarrhea	17 (17%)	
Respiratory symptoms		
Cough	40 (40%)	
Sputum	18 (18%)	
Rhinorrhea	19 (19%)	
Shortness of breath	9 (9%)	
Blood result		
Lymphocyte counts (x10^9/L, normal range 1.1-2.9, median (IQR))	1.2 (1.0, 1.7)	
Death	0 (0%)	

Alterations in faecal RNA virome of COVID-19 patients

To understand whether SARS-CoV-2 infection influences the gut RNA virome, we compared faecal RNA virome of COVID-19 patients at baseline (Day 0, the first timepoint of stool collection after hospitalization) with that of non-COVID-19 controls. Among all host factors (SARS-CoV-2 infection, age, gender, medications, co-morbidities), SARS-CoV-2 infection showed the largest effect size on impacting composition of faecal RNA virome (permanova test p<0.01, R^2=0.041, **Figure 1a**) followed by the chronic

hepatitis B (HBV) infection and asthma. At the species level, SARS-CoV-2 was enriched in faecal samples of patients with COVID-19 compared with non-COVID-19 controls (via MaAsLin2 analysis with adjustment for chronic hepatitis B infection and asthma, FDR p<0.05, **Figure 1b**). In contrast, Pepper mild mottle virus (PMMoV), a plant virus known to be prevalent and abundant in human feces [17], was underrepresented in patients with COVID-19 (FDR p<0.05, **Figure 1b&c**). Seven (19%) of the 37 COVID-19 patients with longitudinal follow-up showed prolonged faecal SARS-CoV-2 shedding after nasopharyngeal clearance of the virus (**Supplementary Figure 2a**). PMMoV virus was persistently underrepresented both during hospitalization and after disease resolution in COVID-19 patients (**Figure 1c**, **Supplementary Figure 2b**). Overall, the faecal RNA virome composition of COVID-19 patients markedly differed from that of non-COVID-19 controls during the disease course and after disease resolution (**Figure 1d&e**). Among 16 COVID-19 patients who had nasopharyngeal clearance of SARS-CoV-2 virus (disease resolution as determined by negative PCR result for SARS-CoV-2 on nasopharyngeal swab and/or deep throat saliva), eight (50%) had persistently altered faecal RNA virome after disease resolution and two (13%) lasted up to 30 days (**Figure 1d**).

Alterations in faecal DNA virome of COVID-19 patients

We then investigated the effect of SARS-CoV-2 infection on faecal DNA virome composition at baseline and during disease course. At the community level, viromes of COVID-19 patients at baseline differed significantly from that of non-COVID-19 controls (permanova p<0.01, **Figure 2a**) and were more heterogeneous than that of non-COVID-19 controls (p<0.0001, **Figure 2b**). Among all host factors (SARS-CoV-2 infection, age, gender, medications, co-morbidities), SARS-CoV-2 infection again showed the largest effect size on impacting composition of faecal DNA virome (R^2=0.018, **Figure 2c**) followed by hyperlipidemia and antiviral drug (Lopinavir-ritonavir). A total of 45 DNA virus species were found to be significantly different in faecal DNA virome of COVID-19 patients relative to non-COVID-19 controls (19 virus species enriched in COVID-19 patients versus 26 virus species enriched in non-COVID-19 controls, identified via DESeq while controlling for the factors Hyperlipidemia and Lopinavir-ritonavir, shown in **Figure 2d**). A majority (69%, 18 out of 26 virus species) of the DNA viruses enriched in feces of non-COVID-19 controls were prokaryotic viruses particularly bacteriophages (62%). In contrast, more eukaryotic viruses as opposed to prokaryotic viruses particularly environment-derived eukaryotic viruses with unknown host were enriched in feces of COVID-19 patients. The bacteriophages, Escherichia virus (phage) and Enterobacter phage, were also enriched in COVID-19 patients (**Figure 2d**). Expansion of these phages has been causally implicated in gut inflammation and host interferon response in mice and humans [14, 18]. These data highlight enteric virome alteration is likely a consequence of SARS-CoV-2 infection which may further contribute to the immunological and physiological changes in the host during the disease course.

The differentially enriched gut DNA virus species in COVID-19 patients showed substantial temporal variations during the disease course (**Figure 2e**) suggesting unstable gut DNA virome. Overall, faecal DNA virome composition of COVID-19 patients differed markedly from that of non-COVID-19 controls during the disease course and after clearance of SARS-CoV-2 (**Figure 2f&g**). Among COVID-19 patients who had

follow-up after disease resolution, six (32%) showed marked more dissimilar faecal DNA virome to non-COVID-19 controls at the last follow-up (three patients lasted up to 20-30 days), compared to their dissimilarity to non-COVID-19 controls at baseline (**Figure 2f**).

Alterations in functionality of the enteric virome in COVID-19 patients

We next investigated functionality alterations of the gut virome using HUMAnN2 predication. More gene families were enriched in COVID-19 viromes at baseline than non-COVID-19 controls (28 versus 9 gene families, FDR p<0.05, **Figure 3**). We found significant enhancement in the functional capacity of gene mobilization and phage integration into the host in COVID-19 viromes (**Figure 3**). Features of viral integration (expansion of temperate virions) have been observed in the gut under inflammatory conditions in both humans and mice [18, 19]. In addition, functions involved in host stress/inflammation/virulence response (DNA repair, Arginine repressor, Hemolysin channel protein, DNA polymerase IV), bacterial metabolism and membrane transport were also enriched in the faecal virome of COVID-19 patients (**Figure 3**). These data highlight that a SARS-CoV-2 infection may induce a functionality shift of the gut virome to inflammation- and stress-responsive in association with their hosts (both the commensal bacteria and humans).

Faecal virome alterations correlated with disease severity of COVID-19

Based on COVID-19 disease symptoms and severity classification criteria [20], we stratified our patients into non-severe (N=56; asymptomatic/mild cases) and moderate/severe groups (N=62; moderate/severe/critical cases) (**Figure 4a**). Compared to non-severe cases, moderate/severe cases showed a significantly higher blood levels of LDH, neutrophil count, C-reactive protein (CRP), Alanine aminotransferase (ALT), and lower blood levels of Albumin at admission (all p<0.05, **Figure 4b-f**, **Supplementary Figure 3**). Our data are in line with recent reports highlighting that more severe cases had more pronounced systemic inflammatory responses [2, 21-24]. We then explored association between baseline faecal RNA and DNA virome profiles with COVID-19 severity and blood measurements at hospitalization. Abundance of the plant-derived RNA virus, Pepper chlorotic spot virus (PCSV) was higher in patients with non-severe than those with moderate/severe disease (p=0.013, **Figure 4g**). In addition, a high abundance of PCSV in feces was associated with low blood concentrations of the inflammation markers, LDH and CRP (correlation coefficient Rho=-0.269 and -0.276 respectively, **Figure 4h&i**). Similarly, abundance of 9 DNA virus species (Myxococcus phage, Rheinheimera phage, Microcystis virus, Bacteroides phage, Murmansk poxvirus, Saudi moumouvirus, Sphaerotilus phage, Tomelloso virus, and Ruegeria phage) in feces negatively correlated with COVID-19 severity (all FDR p<0.05, **Figure 4j**). In particular, 8 out of the 9 DNA virus species showed strong negative correlation with blood levels of the inflammation indicators LDH, neutrophil count, white cell count, or CRP (**Figure 4k**). Interestingly, Myxococcus phage, Bacteroides phage, Murmansk poxvirus, and Sphaerotilus phage also inversely correlated with host age (**Figure 4k**), which may partly account for the observation that elderly individuals were at higher risk for unfavorable severe COVID-19 outcomes[2, 25]. These data suggest that RNA and DNA viruses may counteract the effect of SARS-CoV-2 infection predisposing infected subjects to a less

severe COVID-19 course. Five out of the 9 severity-associated DNA virus species showed persistent lower abundance in the feces of COVID-19 patients during disease course and after disease resolution compared to non-COVID-19 controls (all p<0.05, **Figure 5**), indicating an unfavorable effect of SARS-CoV-2 infection on these gut viruses. The cause or consequence of such associations needs to be further explored. In addition, a large amount of DNA virus species in feces (n=132) showed significant correlations with blood parameters in COVID-19 patients, most of which were negative correlations with blood LDH concentrations, neutrophil and white cell counts (**Supplementary Figure 4**). These data underscore the potential significance of gut DNA virome in calibrating host immunity and counteracting infection of SARS-CoV-2.

Discussion

In this study, we provide the first comprehensive characterization of the gut virome landscape for both RNA and DNA virome in patients with COVID-19 and explores the association between host viral features with COVID-19 severity. Dieta-originated plant RNA viruses are known to be prevalent in a healthy human gut[26]. In the present study, two pepper-derived RNA viruses were found to be underrepresented (PMMoV) or correlated negatively with disease severity in COVID-19 patients (PCSV) highlighting that dietary components along with their colonizer viruses might play a role in COVID-19 protection or pathogenesis. In addition, COVID-19 patients had higher abundances of eukaryotic- and environment-derived viruses whereas non-COVID-19 subjects harbored more bacteriophages in feces. Such viral changes might be a consequence of SARS-CoV-2 infection along with its impact on the GI tract and the resident holistic microbiota ecology [8, 15, 16]. Overall, dysbiosis in gut RNA and DNA virome persisted in COVID-19 patients during disease course and after disease resolution indicating a potential long-lasting detrimental effect to the host after SARS-CoV-2 infection.

The immediate response to any infectious-disease outbreak is to approach it from the pathogen perspective because disease severity is assumed to be a direct function of pathogen burden[27]. However, the complexities of SARS-CoV-2 infection serve as an important reminder that this perspective is not sufficient for understanding the survival of infectious diseases[28, 29]. Disease and immune tolerance as opposed to hyperinflammatory (cytokine storm) state is proposed to be critical for limiting damage to the host and for improving patient survival in COVID-19[29-31]. In line with recent studies on blood immune cell and biochemical measurements [21-23], we found elevations of innate immune cells (neutrophils and white cells) and pro-inflammatory indicators (C-reactive protein and LDH) in the blood of more severe cases. We did not find correlations between faecal SARS-CoV-2 levels and COVID-19 severity but a large number of gut colonizing prokaryotic and eukaryotic viruses correlated negatively with blood concentrations of immune cells and pro-inflammatory molecules and COVID-19 severity, suggesting that gut virome may calibrate host physiology and immune response against SARS-CoV-2 infection.

Interestingly, amongst the identified 9 viral species inversely correlated with disease severity of COVID-19 and inflammation markers in blood, 4 (Myxococcus phage, Bacteroides phage, Murmansk poxvirus, and Sphaerotilus phage) also inversely correlated with human age. This finding may at least partly explain

the observation that elderly individuals are prone to unfavorable severe COVID-19 outcomes[2, 25] and shed light on the importance of gut viruses in human pathophysiology.

There are some limitations of this study. First, it is exploratory in nature without clear cause or consequence effect established. Confirmation of gut virome alterations and their impact on disease severity or trajectory will require functional validation in other populations and animal studies. Second, stool collected after hospitalization for virome analysis does not represent the *bona fide* baseline microbiome before or at COVID-19 onset. Further studies should prospectively include healthy subjects, and if infected with SARS-CoV-2, followed up at disease onset, during disease course, and long term after discovery to delineate the role of virome changes in SARS-CoV-2 infection and post-infection recovery.

Conclusion

In summary, both enteric RNA and DNA viromes were perturbed in COVID-19, which prolonged even after disease resolution. Gut virome may calibrate host immunity and regulate severity to SARS-CoV-2 infection. Our observation that gut viruses inversely correlated with both severity of COVID-19 and host age partly explains that older subjects are prone to severe and unfavorable COVID-19 outcomes. Our data altogether highlight the significance of human gut virome in COVID-19 disease course and potentially therapeutics.

Material And Methods

Study subject and design

This prospective study enrolled 98 hospitalized patients with COVID-19 and 78 gender- and co-morbidity-matched non-COVID-19 controls (**Table 1** and **Supplementary Table 1**). All study subjects had a single stool specimen collected at inclusion. All COVID-19 patients had blood samples collected at admission and measured for white cells and biochemical markers in plasma (**Supplementary Table 1**). Among the included COVID-19 patients, 37 COVID-17 patients had longitudinal follow-up from hospitalization till after discharge, and had serial stool specimens collected (2-3 times per week, **Supplementary Figure. 1**). SARS-CoV-2 infection was confirmed by two consecutive RT-PCR test targeting different regions of the RdRp gene performed by the local hospital and Public Health Laboratory Service. All COVID-19 patients were admitted to the Prince of Wales Hospital or the United Christian Hospital, Hong Kong, from February through May, 2020. COVID-19 severity was categorized as (i) asymptomatic, if there was no clinical symptoms; (ii) mild, if there was no radiographic evidence of pneumonia and the clinical symptoms were light; (iii) moderate, if fever, respiratory tract and other symptoms present and imaging suggests pneumonia; (v) severe, if respiratory rate $\geq$30/min, or oxygen saturation $\leq$93% when breathing ambient air; or PaO_2 / FiO_2 $\leq$ 300 mmHg (1mmHg = 0.133 kPa); or (v) critical, if there was respiratory failure requiring mechanical ventilation, shock, or organ failure requiring intensive care[32]. Non-COVID-19 individuals were recruited at the Prince of Wales Hospital before the COVID-19 outbreak, and tested negative for SARS-CoV-2 during the COVID-19 pandemic. The inclusion criteria for non-COVID-19 subjects

are: 1) Aged ≥18 years old; 2) Competent to provide informed consent (no mental illness or dementia); 3) Have no underlying infectious or acute disease; 4) Have lived in the same geographic area for the preceding 6 months. The exclusion criteria for non-COVID-19 subjects are: 1) Use of laxatives or anti-diarrhoeal drugs in the last 3 months; 2) Recent dietary changes (e.g. becoming vegetarian/vegan); 3) Known complex infections or sepsis (excluding uncomplicated infections such as influenza); 4) Known history of severe organ failure (including decompensated cirrhosis, malignant disease, kidney failure, epilepsy, active serious infection, acquired immunodeficiency syndrome); 5) Bowel surgery in the last 6 months (excluding colonoscopy/ procedure related to perianal disease); 6) Presence of an ileostomy/ stoma; 7) Current pregnancy; 8) Colonoscopy in the last month prior to sampling. This study was approved by the Joint Chinese University of Hong Kong–New Territories East Cluster Clinical Research Ethics Committees (Reference number: 2020.076). All subjects provided informed consent to participate in this study and agreed for publication of the research results. Data including demographic, epidemiological, clinical, laboratory results, treatment and medication were extracted from the electronic medical records in Hong Kong Hospital Authority clinical management system. This study was conducted in accordance with the Declaration of Helsinki.

Faecal viral DNA and RNA extraction, isolation and Shotgun metagenomics

The total viral nucleic acid was extracted from faecal sample, using TaKaRa MiniBEST Viral RNA/DNA Extraction Kit (Takara, Japan) following manufacturer's instructions. Then extracted total viral nucleic acid was then purified by DNA and RNA Clean & Concentrator Kits (Zymo Research, CA, USA) to obtain viral DNA AND RNA respectively. After quality control procedure by Qubit 2.0, agarose gel electrophoresis and Agilent 2100, the qualified DNA and RNA was performed library preparation using Nextera DNA Flex Library Preparation kit (Illumina, USA) and KAPA RNA HyperPrep Kit (Roche, Swiss). The qualified libraries were then sequenced (150bp paired end) on Illumina Novaseq platform.

RNA virome classification and quantification

Raw RNA virome metagenomic sequence reads were filtered and quality-trimmed using Trimmomatic v0.36[33] as follows: 1) Trimming low quality base (quality score < 20); 2) Removing reads shorter than 50bp; 3) removing sequencing adapters. Contaminating human reads were filtering using Kneaddata (Reference database: GRCh38 p12) with default parameters. Cleaned RNA viral reads were rarefied to even depth (20 million reads per sample) and profiled via Kraken2 [34] against the NCBI viral Refseq database as of April 20, 2020. Kraken2 has demonstrated appropriate performance in viral metagenomic classification for time-constrained viral diagnostics and profiling purposes, as well as for surveillance and outbreak source tracing[35]. The abundance of constituent RNA virus speices was calculated as count per million sequencing reads (CPM).Differential viral taxa between COVID-19 patients and non-COVID-19 controls were identified using Multivariate Association with Linear Models (MaAsLin2), while adjusting for confounding factors.

DNA virome classification and quantification

Raw sequence reads were filtered utilizing Trimmomatic using the following parameters; SLIDINGWINDOW: 4:20, MINLEN: 60 HEADCROP 15; CROP 225. Contaminating human reads were filtering using Kneaddata (Reference database: GRCh38 p12) with default parameters. Megahit with default parameters, was chosen to assemble the reads into contigs per sample. Assemblies were subsequently pooled and retained if longer than 1kb. Bacterial contamination was removed by using an extensive set of inclusion criteria to select viral sequences only. Briefly, contigs were required to fulfill one of the following criteria; 1) Categories 1-6 from VirSorter when run with default parameters and Refseqdb (−db 1) positive, 2) circular, 3) greater than 3kb with no BLASTn alignments to the NT database (January '19) (e-value threshold: 1e-10), 4) a minimum of 2 pVogs with at least 3 per 1kb , 5) BLASTn alignments to viral RefSeq database (v.89) (e-value threshold: 1e-10), and 6) less than 3 ribosomal proteins as predicted using the COG database . HMMscan was used to search the pVOGs hmm profile database using predicted protein sequences on VLS with and e-value filter of 1e-5, retaining the top hit in each case. Afterwards, a fasta file combining viral contigs was compiled. This viral database includes the viral contigs recovered by the screening criteria from the bulk metagenomic assemblies. Then the paired reads were mapped to the viral contig database with BWA, using default parameters. The viral operational taxonomic unit (OTU) table of viral abundance was pulled from BWA sam output files by customized script, and normalized by the number of metagenomic reads (metagenomic reads were rarefied to even depth, 20 million reads per sample). The contigs that were analyzed according to their open reading frames (ORFs). The ORFs on the contigs were predicted using MetaProdigal (v2.6.3) with the metagenomics procedure (-p meta). To annotate the predicted ORFs, the amino acid sequences of the ORFs were queried by Diamond against the viral RefSeq protein (v84) with an E value $<10^{-5}$ and a bitscore >50. The viral Refseq proteins with the top closest homologies (E value $<10^{-5}$ and bitscore >50) were considered for each ORF. Contigs were taxonomically binned according to the predominant assignment of its constituent ORFs to a taxon. Contigs shared the same taxonomic identity, the contig table was therefore collapsed by taxonomic identity, where the contig abundances were summed as reads per kilobase per million mapped reads (RPKM) if they shared identity. The virome community differences were visualized via PCoA analysis based on the euclidean distance between each individual's virome. Differential DNA viral taxa between COVID-19 patients and non-COVID-19 controls were identified using DESeq, while adjusting for confounding factors.

Virome function analysis

Viral reads were retrieved via BBMap against the collection of assembled viral contigs, followed by functionality enquiry and annotation via *HUMANN2 v0.9.4*. Predicted functions were collapsed by gene family identity, with abundance values expressed in RPK (reads per kilobase). To establish the presence or absence a function within a sample, a stringent RPK threshold value > 10 was used to be defined as present. Differential functions between COVID-19 and non-COVID-19 subjects were identified by DESeq, while adjusting for co-morbidities, antiviral and age.

Correlation between viruses, COVID-19 severity and blood measurements

Spearman correlation analyses were conducted to correlate both RNA and DNA virome abundance with blood parameter measurements. Spearman correlation analyses with multiple comparison adjustment were conducted to correlate the DNA virome abundance profile with COVID-19 severity across COVID-19 patients.

Declarations

Funding

This work was supported by Hui Hoy & Chow Sin Lan Charity Fund Limited, Pine and Crane Company Limited, and Mr. Hui Ming; Center for Gut Microbiota Research (Faculty of Medicine, The Chinese University of Hong Kong); and Health and Medical Research Fund (Hong Kong, China).

Authors' contributions:

TZ, QL, and FZ performed the experiments, data analyses and drafted the manuscript. YKY revised the manuscript and provided critical intellectual contribution. AC, HZ, YW, CPC, and NC assisted in experiments and metagenomics sequencing. AYLL and LL collected the human specimens and data. FKLC, GCYL, GJ, CKLL, ZC, LL and DSCH provided critical comments. GCYL, EYKT, KSCF, VC, and LL recruited study subjects. SN and PKSC designed and supervised the study.

Conflicts of interest: No

Acknowledgement

We would like to thank all healthcare workers working in isolation wards of Prince of Wales Hospital, Hong Kong, China. We thank Apple CM Yeung, Wendy CS Ho, Miu L Chin, Rity Wong, Barry Wong and Vickie Li for their technical contribution in this study.

Availability of data and materials

Virome metagenomic sequencing data has been deposited to the NCBI Sequence Read Archive under BioProject accession number PRJNA657711.

Ethics approval and consent to participate

This study was approved by the Joint Chinese University of Hong Kong–New Territories East Cluster Clinical Research Ethics Committees (Reference number: 2020.076). All subjects provided informed consent to participate in this study and agreed for publication of the research results.

Consent for publication

All subjects provided informed consent to participate in this study and agreed for publication of the research results.

References

1. Onder G, Rezza G, Brusaferro S: **Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy**. *Jama* 2020.

2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X: **Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China**. *The lancet* 2020, **395**(10223):497-506.

3. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y: **Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study**. *The Lancet* 2020, **395**(10223):507-513.

4. Liang W, Feng Z, Rao S, Xiao C, Xue X, Lin Z, Zhang Q, Qi W: **Diarrhoea may be underestimated: a missing link in 2019 novel coronavirus**. *Gut* 2020.

5. Cheung KS, Hung IFN, Chan PPY, Lung KC, Tso E, Liu R, Ng YY, Chu MY, Chung TWH, Tam AR: **Gastrointestinal manifestations of SARS-CoV-2 infection and virus load in fecal samples from the Hong Kong cohort and systematic review and meta-analysis**. *Gastroenterology* 2020.

6. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C *et al*: **Virological assessment of hospitalized patients with COVID-2019**. *Nature* 2020.

7. Xu Y, Li X, Zhu B, Liang H, Fang C, Gong Y, Guo Q, Sun X, Zhao D, Shen J: **Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding**. *Nat Med* 2020:1-4.

8. Zuo T, Liu Q, Zhang F, Lui GC-Y, Tso EYK, Yeoh YK, Chen Z, Boon SS, Chan FKL, Chan PKS: **Depicting SARS-CoV-2 faecal viral activity in association with gut microbiota composition in patients with COVID-19**. *Gut* 2020.

9. Virgin HW: **The virome in mammalian physiology and disease**. *Cell* 2014, **157**(1):142-150.

10. Neil JA, Cadwell K: **The intestinal virome and immunity**. *The Journal of Immunology* 2018, **201**(6):1615-1624.

11. Brodin P, Jojic V, Gao T, Bhattacharya S, Angel CJL, Furman D, Shen-Orr S, Dekker CL, Swan GE, Butte AJ: **Variation in the human immune system is largely driven by non-heritable influences**. *Cell* 2015, **160**(1-2):37-47.

12. Reese TA, Bi K, Kambal A, Filali-Mouhim A, Beura LK, Bürger MC, Pulendran B, Sekaly R-P, Jameson SC, Masopust D: **Sequential infection with common pathogens promotes human-like immune gene expression and altered vaccine response**. *Cell host & microbe* 2016, **19**(5):713-719.

13. Shkoporov AN, Clooney AG, Sutton TDS, Ryan FJ, Daly KM, Nolan JA, McDonnell SA, Khokhlova EV, Draper LA, Forde A: **The human gut virome is highly diverse, stable, and individual specific**. *Cell host & microbe* 2019, **26**(4):527-541. e525.

14. Zuo T, Lu X-J, Zhang Y, Cheung CP, Lam S, Zhang F, Tang W, Ching JYL, Zhao R, Chan PKS: **Gut mucosal virome alterations in ulcerative colitis**. *Gut* 2019, **68**(7):1169-1179.

15. Zuo T, Zhang F, Lui GCY, Yeoh YK, Li AYL, Zhan H, Wan Y, Chung A, Cheung CP, Chen N: **Alterations in Gut Microbiota of patients with COVID-19 during time of hospitalization**. *Gastroenterology* 2020.

16. Zuo T, Zhan H, Zhang F, Liu Q, Tso EYK, Lui GCY, Chen N, Li A, Lu W, Chan FKL *et al*: **Alterations in Fecal Fungal Microbiome of Patients With COVID-19 During Time of Hospitalization until Discharge**. *Gastroenterology* 2020:S0016-5085(0020)34852-34856.

17. Kitajima M, Sassi HP, Torrey JR: **Pepper mild mottle virus as a water quality indicator**. *NPJ Clean Water* 2018, **1**(1):1-9.

18. Gogokhia L, Buhrke K, Bell R, Hoffman B, Brown DG, Hanke-Gogokhia C, Ajami NJ, Wong MC, Ghazaryan A, Valentine JF: **Expansion of bacteriophages is linked to aggravated intestinal inflammation and colitis**. *Cell host & microbe* 2019, **25**(2):285-299. e288.

19. Clooney AG, Sutton TDS, Shkoporov AN, Holohan RK, Daly KM, O'Regan O, Ryan FJ, Draper LA, Plevy SE, Ross RP: **Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease**. *Cell host & microbe* 2019, **26**(6):764-778. e765.

20. Wu J, Liu J, Zhao X, Liu C, Wang W, Wang D, Xu W, Zhang C, Yu J, Jiang B: **Clinical characteristics of imported cases of COVID-19 in Jiangsu province: a multicenter descriptive study**. *Clin Infect Dis* 2020, **10**.

21. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, Akinosoglou K, Antoniadou A, Antonakos N, Damoraki G, Gkavogianni T, Adami M-E, Katsaounou P: **Complex immune dysregulation in COVID-19 patients with severe respiratory failure**. *Cell host & microbe* 2020.

22. Ling W: **C-reactive protein levels in the early stage of COVID-19**. *Med Maladies Infect* 2020.

23. Sun S, Cai X, Wang H, He G, Lin Y, Lu B, Chen C, Pan Y, Hu X: **Abnormalities of peripheral blood system in patients with COVID-19 in Wenzhou, China**. *Clinica Chimica Acta* 2020.

24. Lechien JR, Chiesa-Estomba CM, Place S, Van Laethem Y, Cabaraux P, Mat Q, Huet K, Plzak J, Horoi M, Hans S: **Clinical and epidemiological characteristics of 1,420 European patients with mild-to-moderate coronavirus disease 2019**. *J Intern Med* 2020.

25. Gupta A, Madhavan MV, Sehgal K, Nair N, Mahajan S, Sehrawat TS, Bikdeli B, Ahluwalia N, Ausiello JC, Wan EY: **Extrapulmonary manifestations of COVID-19**. *Nat Med* 2020:1-16.

26. Zhang T, Breitbart M, Lee WH, Run J-Q, Wei CL, Soh SWL, Hibberd ML, Liu ET, Rohwer F, Ruan Y: **RNA viral community in human feces: prevalence of plant pathogenic viruses**. *Plos Biol* 2005, **4**(1):e3.

27. Schneider DS, Ayres JS: **Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases**. *Nat Rev Immunol* 2008, **8**(11):889-895.

28. Ayres JS: **A metabolic handbook for the COVID-19 pandemic**. *Nature Metabolism* 2020:1-14.

29. Ayres JS: **Surviving COVID-19: A disease tolerance perspective**. In.: American Association for the Advancement of Science; 2020.

30. Vabret N, Britton GJ, Gruber C, Hegde S, Kim J, Kuksin M, Levantovsky R, Malle L, Moreira A, Park MD: **Immunology of COVID-19: current state of the science**. *Immunity* 2020.

31. Tufan A, GÜLer AA, Matucci-Cerinic M: **COVID-19, immune system response, hyperinflammation and repurposing antirheumatic drugs**. *Turkish Journal of Medical Sciences* 2020, **50**(SI-1):620-632.

32. Wu J, Liu J, Zhao X, Liu C, Wang W, Wang D, Xu W, Zhang C, Yu J, Jiang B: **Clinical characteristics of imported cases of COVID-19 in Jiangsu province: a multicenter descriptive study**. *Clinical Infectious Diseases* 2020.

33. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.

34. Wood DE, Salzberg SL: **Kraken: ultrafast metagenomic sequence classification using exact alignments**. *Genome biology* 2014, **15**(3):R46.

35. Nooij S, Schmitz D, Vennema H, Kroneman A, Koopmans MPG: **Overview of virus metagenomic classification methods and their biological applications**. *Frontiers in microbiology* 2018, **9**:749.

Figures
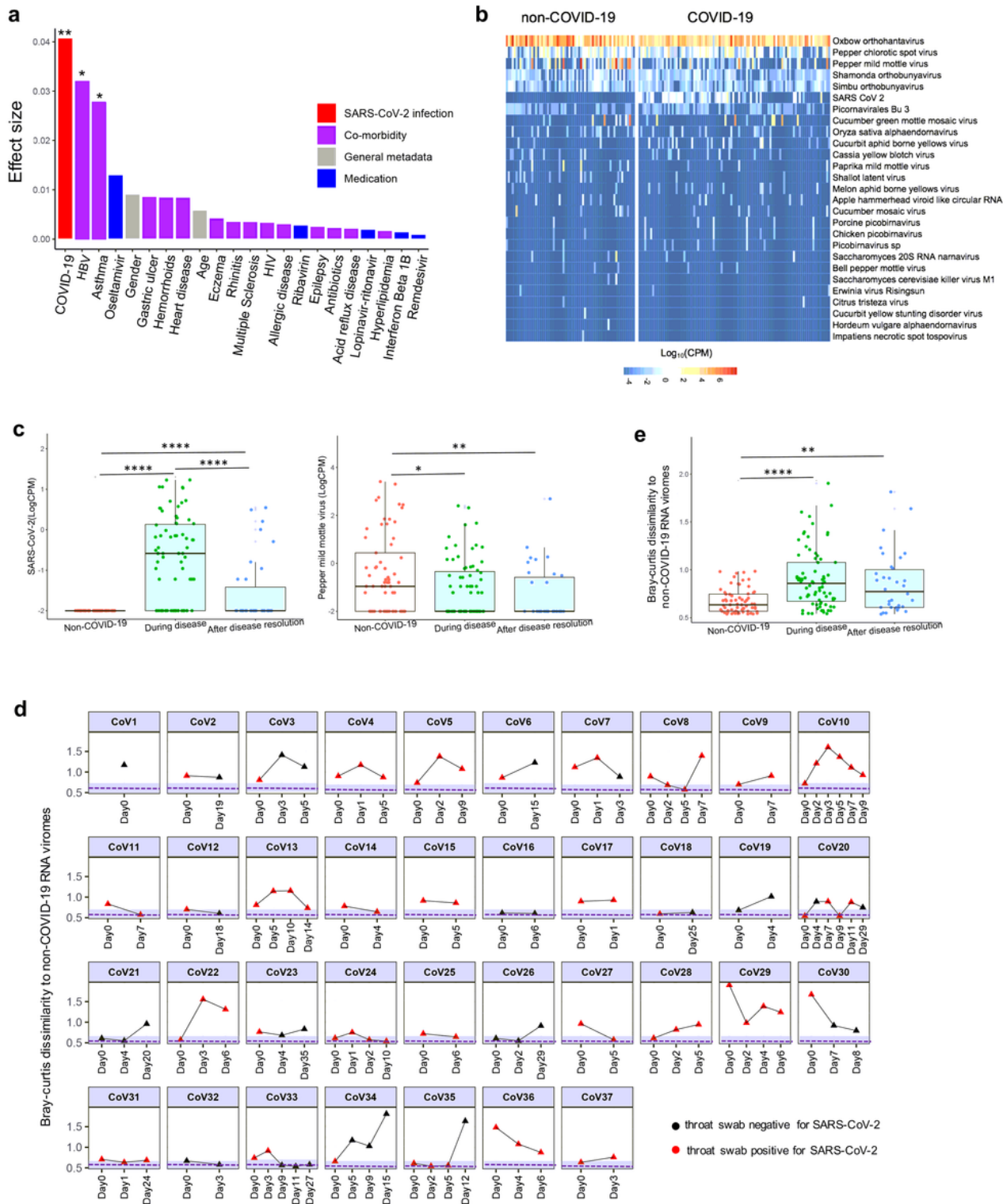
Figure 1

Gut RNA virome in COVID-19 patients and its longitudinal changes during disease course. a, Effect size of SARS-CoV-2 infection (COVID-19) and host factors on faecal RNA virome composition variation. The effect size and statistical significance was determined via PERMANOVA analysis and permutation test (n=999), **p<0.01, *p<0.05. b, Heatmap abundance of faecal RNA virus species in COVID-19 patients and non-COVID-19 controls. CPM, count per million reads. c, The abundance of SARS-CoV-2 and PMMoV in

non-COVID-19 controls and in COVID-19 patients during hospitalization and after disease resolution. Between group comparison was conducted by Mann-Whitney test, ****$p<0.0001$, **$p<0.01$, *$p<0.05$. d, Temporal dissimilarity of patient's faecal RNA virome to non-COVID-19 faecal RNA viromes over the disease course. Virome dissimilarity of the patient to non-COVID-19 subjects was plot as the average bray-curtis dissimilarity between the indicated faecal virome to all non-COVID-19 viromes. The grey area depicts the dissimilarity range (mean±s.e.) between all faecal RNA viromes corresponding to non-COVID-19 controls (the dashed line denotes the mean dissimilarity). "CoV n" denotes COVID-19 patient number. "Day0" denotes baseline date when the first stool was collected after hospitalization; the following time points starting with 'Day' represents days since baseline stool collection. e, The average (median) bray-curtis dissimilarity of patient faecal RNA virome to non-COVID-19 faecal RNA virome during the disease course. For box plots, the boxes extend from the 1st to 3rd quartile (25th to 75th percentile), with the median depicted by a horizontal line. Statistical significance was determined by Mann-Whitney test, ****$p<0.0001$, **$p<0.01$.
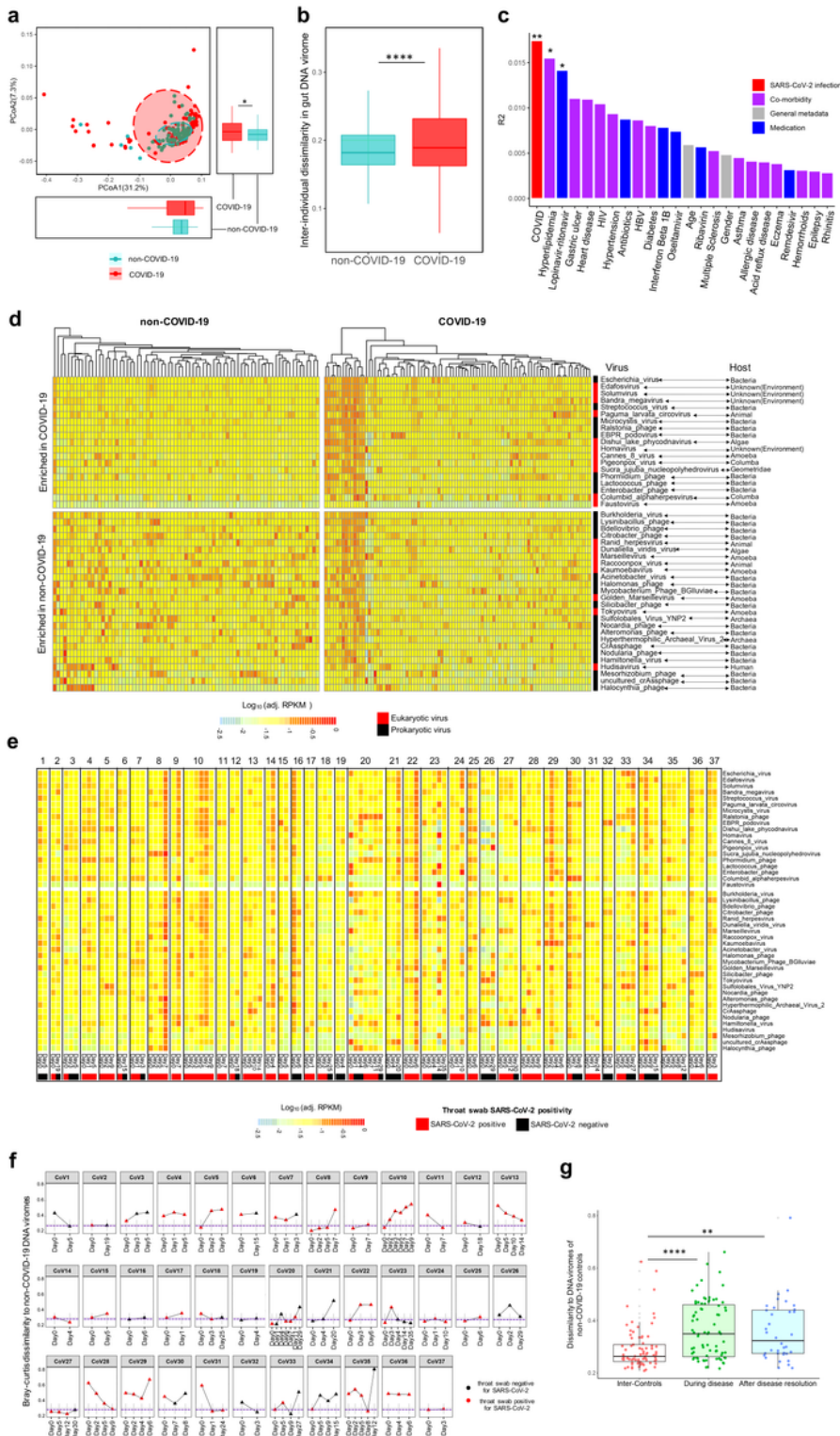
Figure 2

Gut DNA virome alterations in COVID-19 patients and longitudinal changes during disease course. a, PCoA analysis of faecal DNA viromes in COVID-19 patients versus non-COVID-19 controls. Distribution of faecal viromes along each axis between the two groups was statistically determined by Mann-Whitney test, *p<0.05. b, Inter-individual dissimilarity (beta-diversity) of faecal DNA viromes within each group. Between-group comparison was conducted by Mann-Whitney test, ****p<0.0001. c, Effect size of SARS-

CoV-2 infection (COVID-19) and host factors on faecal DNA virome composition variation. The effect size and statistical significance was determined via PERMANOVA analysis and permutation test (n=999), **p<0.01, *p<0.05. d, Differential DNA viruses between COVID-19 patients and non-COVID-19 controls, identified by DESeq, adjusted for antivirals and co-morbidities. Only the significant species (FDR p<0.05) were plotted. e, Temporal changes of the differential speices shown in Figure d in each COVID-19 patient over the disease course. f, Temporal dissimilarity of patient's faecal DNA virome to non-COVID-19 faecal RNA viromes over the disease course. Virome dissimilarity of the patient to non-COVID-19 subjects was plot as the average bray-curtis dissimilarity between the indicated faecal virome to all non-COVID-19 viromes. The grey area depicts the dissimilarity range (mean±s.e.) between all faecal DNA viromes corresponding to non-COVID-19 controls (the dashed line denotes the mean dissimilarity). "CoV n" denotes COVID-19 patient number. "Day0" denotes baseline date when the first stool was collected after hospitalization; the following time points starting with 'Day' represents days since baseline stool collection. g, The average (median) bray-curtis dissimilarity of patient faecal DNA viromes to non-COVID-19 faecal DNA viromes during the disease course. For box plots, the boxes extend from the 1st to 3rd quartile (25th to 75th percentile), with the median depicted by a horizontal line. Statistical significance was determined by Mann-Whitney test, ****p<0.0001, **p<0.01.
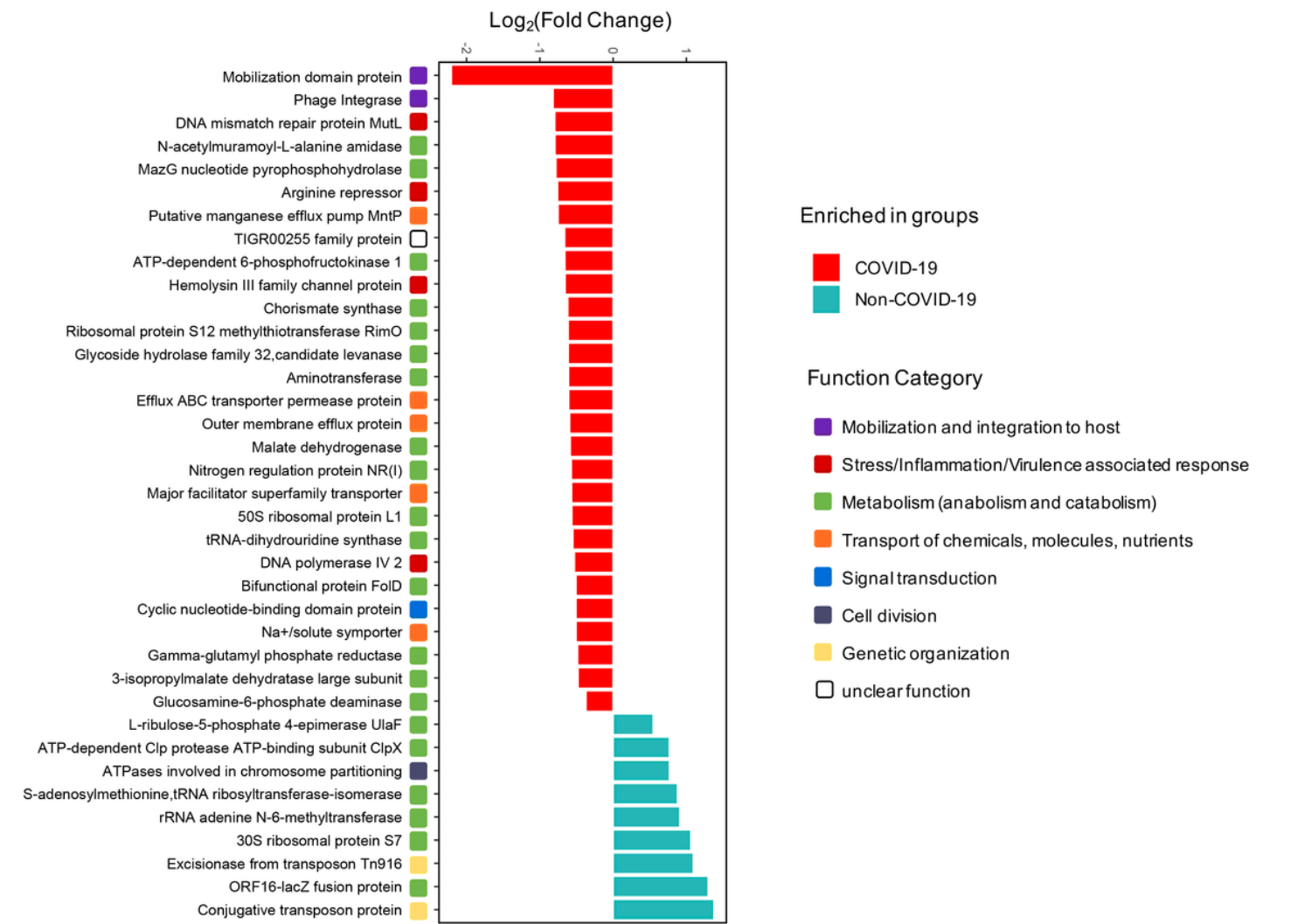
Figure 3

Disparity in the functional capacity of gut virome between COVID-19 patients and Non-COVID-19 controls. The differential functions (gene families) were identified by DESeq. Only the significant function terms with FDR p value <0.05 and abundance (RPK) >10 were shown.
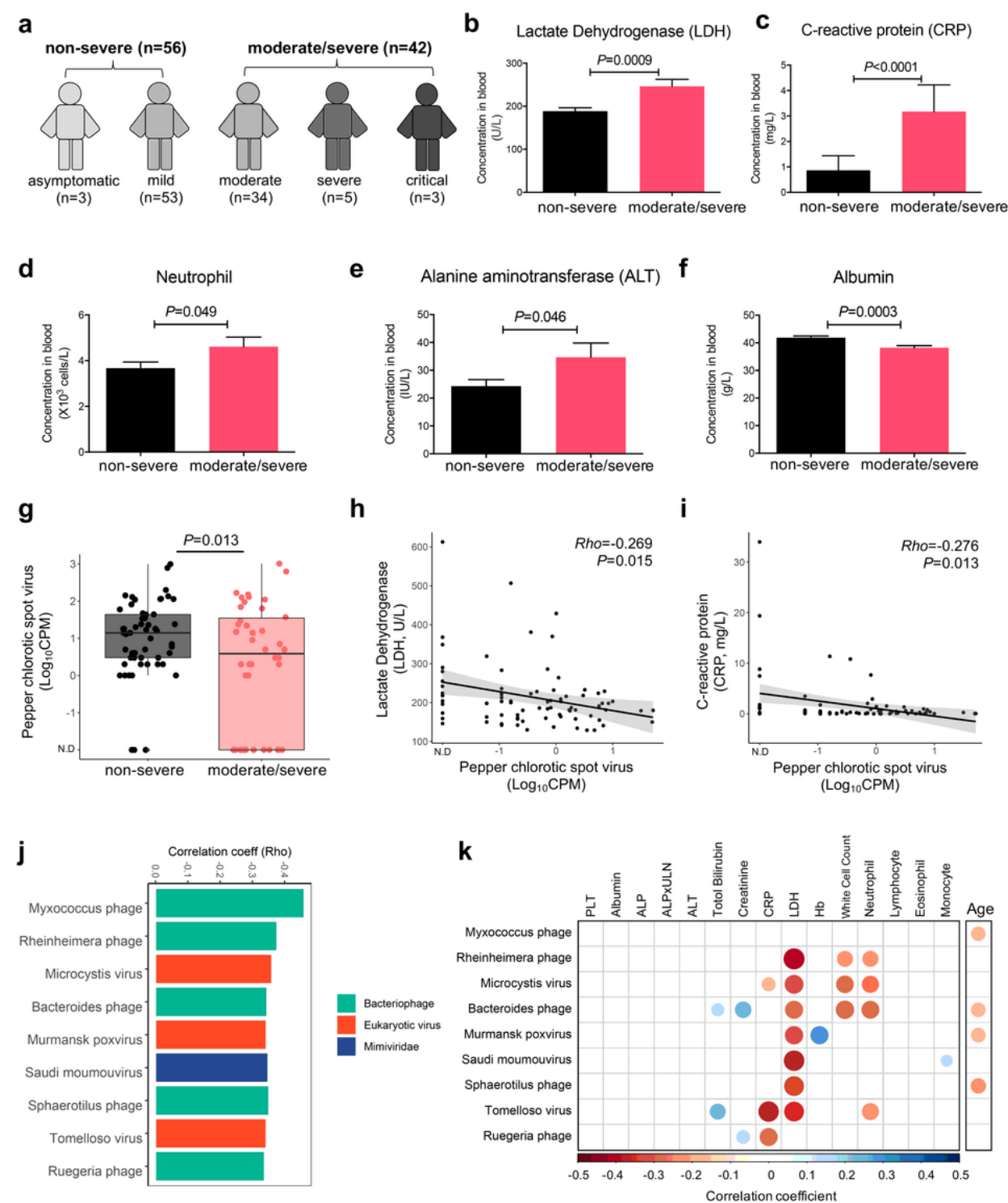
Figure 4

Faecal virus species correlated with COVID-19 severity and blood parameters a. Recruited COVID-19 patients and their symptom severity of COVID-19. Patients were separated in to non-severe (n=56) and moderate/severe (n=42) groups. b-f, Blood measurement results for LDH, c-reactive protein, neutrophil count, ALT and Albumin concentrations. Data are shown in mean±s.e. The comparisons were made between non-severe and moderate/severe groups via Mann-Whitney test. g-i, The abundance of Pepper chlorotic spot virus in faecal RNA virome and its association with disease severity (g), blood concentrations of LDH (h) and C-reactive protein (i). CPM: count per million reads. Between-two group comparison was conducted by Mann-Whitney test. Correlation test was performed by Spearman correlation test. j-k, The 9 DNA virus species that negatively correlated with severity of COVID-19 (j) and their correlations with blood measurements (k). Correlation test was performed by Spearman correlation test. In Figure k, the color and intensity denote the spearman correlation direction and coefficient, where only the significant correlations with blood parameters of |correlation coefficient| > 0.3 were shown.
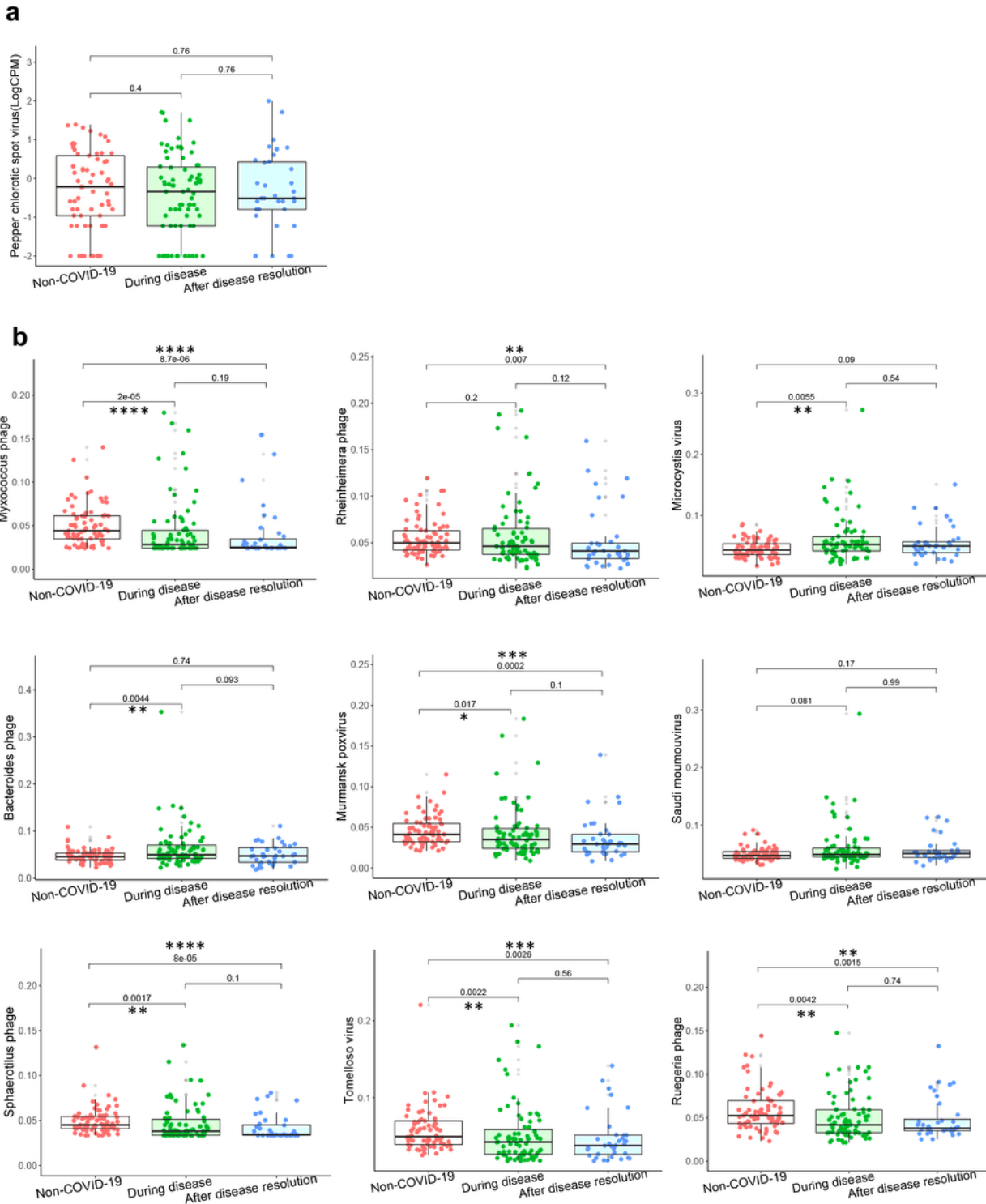
Figure 5

Longitudinal changes in the abundance of RNA (a) and DNA viruses (b) that correlated with COVID-19 severity (shown in Figure 3) in feces of COVID-19 patients during disease course and after disease resolution, as compared to Non-COVID-19 patients. The abundance of RNA virus was expressed in Log10CPM, where CPM denotes count per million reads. The abundance of DNA virus was expressed in normalized abundance (DESeq adjusted RPKM), where RPKM denotes reads per kilobase per million

mapped reads. Statistical significance was calculated by Man-Whitney test with *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.