The active lung microbiota landscape of COVID-19 patients

Yang Han[1,2,+], Zhilong Jia[1,2,+,*], Jinlong Shi[1,2], Weidong Wang[1,2,*], Kunlun He[1,2,*]

[1]Key Laboratory of Biomedical Engineering and Translational Medicine, Ministry of Industry and Information

Technology, Medical Innovation Research Division of Chinese PLA General Hospital, Beijing, China

[2]Beijing Key Laboratory for Precision Medicine of Chronic Heart Failure, Medical Innovation Research Division

of Chinese PLA General Hospital, Beijing, China

+ These authors contributed equally

* Correspondence: Dr. Zhilong Jia (jiazhilong@plagh.org), Profs. Weidong Wang (wangwd301@126.com) and

Kunlun He (kunlunhe@plagh.org)

ABSTRACT:

With the outbreak of COVID-19 causing by SARS-CoV-2, the interaction between the host and SARS-CoV-2 was widely studied. However, it is unclear whether and how SARS-CoV-2 infection affects lung microflora, which contributes to COVID-19 complications. Here, we analyzed the metatranscriptomic data of bronchoalveolar lavage fluid (BALF) of 19 COVID-19 patients and 23 healthy controls from 6 independent projects and detailed the active microbiota landscape in both healthy individuals and COVID-19 patients. The infection of SARS-CoV-2 could deeply change the lung microbiota, evidenced by the α-diversity, β-diversity and species composition analysis based on bacterial microbiota and virome. Pathogens (such as *Klebsiella oxytoca* causing pneumonia as well), immunomodulatory probiotics (such as Lactic Acid Bacteria and *Faecalibacterium prausnitzii,* a butyrate producer) and Tobacco mosaic virus (TMV) were enriched in the COVID-19 group, suggesting a severe microbiota dysbiosis. The significant correlation between *Rothia mucilaginosa*, TMV and SARS-CoV-2 revealed drastic inflammatory battles between the host, SARS-CoV-2 and other microbes in the lungs. Notably, TMV only existed in the COVID-19 group, while Human respirovirus 3 only existed in the healthy group. Our study provides insight into the active microbiota in the lungs of COVID-19 patients and will contribute to the understanding of the infection mechanism of SARS-CoV-2 and the treatment of the disease and complications.

Key words: Microbiota; SARS-CoV-2; COVID-19; Lactic Acid Bacteria; *Faecalibacterium prausnitzii*

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Introduction

With the outbreak of severe acute respiratory syndrome coronavirus 2, SARS-CoV-2(1), leading to Coronavirus disease 2019 (COVID-19), many studies on clinical characteristics, three-dimensional structure(2) and interactions between host and virus(3) were reported. The symptoms of COVID-19 patients (COV) usually include fever, vomiting, diarrhea(4). As bacterial or viral infections could also induce several symptoms among them, it is necessary to profile the microbiota of COVID-19 patients to effectively relieve their symptoms(5, 6). The bronchoalveolar lavage fluid (BALF), containing microenvironment information on bronchioles and lung alveoli from the lower respiratory tract, is one of key sample types for characterizing the host inflammatory response and microbiota of COVID-19 patients as lung is one of main organs for the infection of SARS-CoV-2(7, 8).

The human microbiota in BALF, including bacterial microbiota and virome, is a diverse microbial ecosystem associated with beneficial or deleterious physiological functions as well as disease etiologies, including COVID-19(9). It was reported that multiple common respiratory pathogens were co-infected with SARS-CoV-2 in COVID-19 patients(10). Moreover, the role of commensal viruses in human lungs is poorly understood, particularly in the respiratory tract(11). Besides, there are close interactions between microorganisms including symbiosis and competition(12, 13). Therefore, the relationship between SARS-CoV-2 and other microorganisms in the body is of great significance for studying its infection mechanism and developing effective treatments. Shen et al. briefly reported that the bacterial microbiota in COVID-19 patients was dominated by the pathogens and oral and upper respiratory commensal bacteria(14). However, there is no systematic microbiota landscape in BALF samples from COVID-19 patients as well as the healthy individuals.

In this study, we systematically profiled the transcriptionally active microbiota landscape in BALF from COVID-19 patients and healthy individuals, identified microorganism composition in healthy individuals and COVID-19 patients, found disease-specific active microbes in the COVID-19 patient group, revealed the interaction between several bacteria or viruses and SARS-CoV-2. The systematic microbiota landscape provided crucial insight into understanding the mechanism of SARS-CoV-2 infection and the treatment of COVID-19.

Materials and Methods

Data collection and preprocessing

The raw metatranscriptomic fastq data of for 42 BALF samples, consist of 19 COVID-19 patients and 23 healthy controls, and two negative controls (NC) were downloaded from Sequence Read Archive, EMBL-EBI ArrayExpress and National Genomics Data Center, China with project IDs, PRJCA002202, PRJNA601736, PRJCA002326, PRJNA605983, PRJNA615032 and PRJNA434133. The basic information of these samples was described in Supplementary Table 1. The raw fastq data were analyzed using IDSeq webserver(15) (Pipeline v4.9) to obtain the raw read count of each bacterial and viral species. Briefly, the raw reads underwent removal of host reads (Hg38 reference genome) via STAR (Spliced Transcripts Alignment to a Reference), a series of quality control, consist of removal of low-quality, low-complexity and redundant sequences via Paired-Read Iterative Contig Extension

(PRICE), CD-HIT-DUP and Lempel-Ziv-Welch (LZW) compression score, respectively(15). After quality control, the assignment of taxonomic IDs was implemented by aligning to the NCBI nucleotide (nt) and non-redundant protein (nr) databases via GSNAPL and RAPsearch(15), respectively. Then the read counts of nt and reads per million (rPM) of nt at the species level were obtained. To guarantee the quality and credibility of the lung microbiota due to the low biomass, species with rPM value less than 50 in NC were used for downstream analysis. Moreover, bacteria detected in at least two study projects were kept to largely eliminate the reagent and laboratory contamination.

Microbial diversity and differential analysis

Microbial diversity was analyzed by R package vegan (v2.5-6). Shannon diversity index was calculated to evaluate α-diversity. Bacterial or viral loads were computed based on rPM of microorganism. β-diversity was studied by Principal coordinate analysis, PCoA, using R package ape (v5.3) based on Bray-Curtis distance. Differential species were analyzed by DEseq2 (v1.24.0). Differential bacteria with a statistical significance (q-value) < 1e-11 and absolute value of $\log_2$(Fold Change) > 3, virus with a statistical significance (q-value) < 0.01 and absolute value of $\log_2$(Fold Change) > 1 were retained. Student's t test and Mann-Whitney rank test were used in variables with normal and without normal distribution, respectively. The Benjamini-Hochberg correction was used to obtain FDR adjusted p-values (q-values) for multiple hypothesis testing.

Microbiota correlation analysis in BALF

Correlation analysis was implemented by Spearman's rank correlation. Only bacteria and viruses retained in the differential analysis were used for correlation analysis. Bacteria correlations with a coefficient (r) > 0.8 and a statistical significance (q-value) < 1e-13 were considered for further analysis and visualization. Virus correlations with a coefficient (r) > 0.5 and a statistical significance (q-value) < 0.01 were considered for further analysis and visualization. Bacteria and virus correlations with a coefficient (r) > 0.7 and a statistical significance (q-value) < 5e-8 were considered for further analysis and visualization. Correlation network was visualized by Cytoscape 3.7.2(16).

Results

We collected the metatranscriptomic data of 42 BALF samples, consist of 19 COVID-19 patients and 23 healthy controls (HC), from six studies(1, 14, 17-20). The basic information of the samples was shown in Supplementary table 1. 19 COVID-19 patients were from five different studies, representing individual sample source of the samples. In totally, 12413 bacteria and 359 viruses at the species level, which were transcriptionally active, were identified.

Differential microbial diversity in the COVID-19 and healthy groups

The diversity analysis of bacteria and virus, including α- and β-diversity, showed large differences between two groups. The COVID-19 group had higher Shannon's Diversity Index of bacteria than

healthy group (p-value = 0.0268, Student's t test, Figure 1a). Also, Bacterial loads of COVID-19 group were much higher than healthy group (p-value = 9e-4, Mann-Whitney rank test, Figure 1b). For virome, COVID-19 patients had much higher viral loads than healthy controls (p-value < 1e-4, Mann-Whitney rank test) while Shannon's Diversity Index showed no statistical difference (Figure 1cd).

Furthermore, Principal Coordinates Analysis (PCoA) with Bray-Curtis distance was applied to evaluate the β-diversity in BALF microbiota across groups. The PCoA results of bacterial microbiome showed there was a significant difference of β-diversity between two groups (p-value < 1e-3, R = 0.217, ANOSIM Analysis) (Figure 2a). Also, similar results for virome were found (p-value < 1e-3, R = 0.541, ANOSIM Analysis) (Figure 2b), indicating a heterogeneous community diversity in the patient and control groups. The diversity analysis revealed that the infection of SARS-CoV-2 probably caused a different lung microbiota composition in the COVID-19 patient group compared with the healthy group.

Microbial composition of BALF samples

To get an insight of the microbial composition of BALF, we screened out the 10 most abundant bacteria at the species level in healthy controls and COVID-19 patients (Figure 3a), respectively. Concerning the bacteria community, *Porphyromonas gingivalis*, often found in the oral cavity where it is implicated in periodontal disease(21), had a relatively higher abundance in some healthy individuals. The abundance of *Lautropia mirabilis*, which was commonly found in the human oral cavity and the upper respiratory tract(22), and *Enterobacter kobei*, leading to nosocomial infections(23), were increased in a few COVID-19 patients. Notably, *Variovorax* account for a very high proportion in 6 patients. Interestingly, *Staphylococcus xylosus* and *Staphylococcus simulans* had high abundances in three healthy controls (H21, H22 and H23).

The virome composition analysis showed that Human respirovirus 3 and Enterovirus D were higher in a few healthy samples (Figure 3b). Except severe acute respiratory syndrome-related coronavirus, Spodoptera frugiperda rhabdovirus was significantly enriched in five patient samples. (Figure 3b). Severe acute respiratory syndrome-related coronavirus includes Severe acute respiratory syndrome coronaviruses (SARS-CoVs) and SARS-CoV-2. The top abundant species showed that the infection of SARS-CoV-2 changed the bacterial and viral community in the lung of COVID-19 patients.

Microbial differences between the COVID-19 patient and healthy control groups

To identify the species associated with COVID-19, differential abundance analysis was performed between the two groups. 99 bacteria at the species level were significantly different between the two groups (q-value < 1e-11, absolute value of $\log_2$(Fold Change) > 3) (Supplementary Table 2). Meanwhile, 82 species with the maximum relative abundance greater than 0.05% were selected from the 99 bacteria and visualized using heatmap to get a deeper investigation (Figure 4a).

The abundance of Lactic Acid Bacteria (LAB), such as *Lactobacillus fermentum*, *Lactobacillus reuteri*, *Lactobacillus delbrueckii* and *Lactobacillus salivarius*, were higher in the COVID-19 group than in the healthy group (Supplementary Table 2). Of note, *Lactobacillus reuteri* was detected only in

the patient group. Some pathogens were enriched in COVID-19 patients and depleted in healthy controls. For example, *Klebsiella oxytoca*, leading to pneumonia, colitis and sepsis(24), and *Enterobacter cloacae*, positively correlated with COVID-19 severity(25), were increased in COVID-19. The abundance of some nosocomial infection pathogens, such as *Enterobacter kobei*(23), *Enterobacter cloacae*(26) and *Ralstonia pickettii*(27), were higher in patients than healthy controls. Several gut bacteria like *Faecalibacterium prausnitzii*(28), *Enterococcus faecium*(29) and *Citrobacter freundii*(30) and commensal bacteria residing in the mouth and respiratory tracts, such as *Rothia mucilaginosa*, were also enriched in the lung of COVID-19 patients. In addition, a few of pathogens causing severe nausea, vomiting and diarrhea, like *Bacillus cereus*(31) increased in patients significantly.

The abundance of 18 viruses significantly differed between the two groups (q-value < 0.01, absolute value of log_2(Fold Change) > 1) (Figure 4b，Supplementary Table 3). In addition to severe acute respiratory syndrome-related coronavirus, Surprisingly, 13 of 18 differential viruses were only detected in the COVID-19 group, such as Pangolin coronavirus and Tobacco mosaic virus (TMV), which were probably closely related to the infection of SARS-CoV-2, though the function of them is unclear. Notably, Human respirovirus 3 was only detected in healthy controls (12 of 23 individuals).

Microbiota correlation in BALF

To further investigate whether the altered microbiota interacted with each other, we constructed networks of bacteria and viruses based on Spearman rank correlations, respectively. Network within bacteria (Figure 5a and Supplementary Table 4) showed that lots of bacteria were highly correlated. For example, *Enterobacter sp.*, *Pseudomonas sp.* and *Leclercia sp.* were positively correlated to each other. What's more, *Anaerotignum propionicum*, detected only in the patient group, was positively correlated to *Chryseobacterium cucumeris*, *Ruminococcus gnavus* and *Leclercia sp. LSNIH1*.

Network within viruses (Figure 5b and Supplementary Table 5) showed severe acute respiratory syndrome-related coronavirus, TMV and Pangolin coronavirus were positively correlated to each other, while severe acute respiratory syndrome-related coronavirus was negatively correlated to Human respirovirus 3. The Network between bacteria and viruses (Figure 5c and Supplementary Table 6) presented that severe acute respiratory syndrome-related coronavirus was positively correlated to *Rothia mucilaginosa*, *Schaalia meyeri* and *Phyllobacterium myrsinacearum* which was also positively related to TMV.

Taken together, the results of correlation analysis suggest that the altered lung microbiota, particularly species from the *Lactobacillus* and *Rothia*, *Enterobacter* and TMV, actively played a certain role during the infection of SARS-CoV-2. The negative correlation between SARS-CoV-2 and Human respirovirus 3 indicates that the former probably have an inhibitory effect on the latter.

Discussion

The transcriptionally active microbiota landscape of BALF in COVID-19 patients and healthy controls could shed light on the infection mechanism of SARS-COV-2. Our study described the

microbial composition at the species level in BALF of COVID-19 patients and healthy controls. Although *Porphyromonas gingivalis* existed widely in healthy lungs and accounted for a high proportion, lower richness and bacterial loads means it still stay a low level. In addition, though Human respirovirus 3 is one of the most common respiratory viruses, particularly among young children under 5 years of age and immunocompromised patients(32, 33), it was detected in the healthy group with no symptoms.

The infection of SARS-CoV-2 changed the active microbiota in the lungs, characterized by the α-diversity, β-diversity, species composition of bacteria and viruses respectively. The changes should be a result of complex interactions between the host, SARS-CoV-2, bacteria and other viruses. The enrichment of LAB, probiotics usually producing lactic acid, in the COVID-19 group could probably inhibit the inflammatory response caused by SARS-CoV-2 to some extent, as several species of *Lactobacillus* could decrease the inflammatory response, and balance the immune response while limiting the potential of pathogens(34, 35). Besides, *Lactobacillus* could help treat diarrhea. The enrichment of several pathogens, such as bacteria (*Klebsiella oxytoca*, *Enterobacter kobei* and *Bacillus cereus*) and viruses (Pangolin coronavirus and TMV), in the COVID-19 group probably indicated that co-infections or a second infection. Several symptoms, such as gastrointestinal symptoms and extreme fatigue in addition to fever, may be caused by the increase of these pathogens.

Although *Rothia mucilaginosa* was a part of the normal oropharyngeal flora, it was linked to causing disease, including pneumonia, in immunosuppressed humans(36). The enrichment of it in COVID-19 patients may be the result of the impact of SARS-CoV-2 on the patients' immune system. In our study, we also found that some intestinal bacteria (such as *Faecalibacterium prausnitzii* and *Enterococcus faecium*) were enriched in the lung of COVID-19 patients. A study reported that intestinal bacteria could transfer to other organs through the intestinal barrier(37), so we speculated that the infection of SARS-CoV-2 could probably promote the transfer of intestinal bacteria. *Faecalibacterium prausnitzii*, a prototypical IL-10 and butyrate producer in the gut(38), showed a negative correlation with COVID-19 severity(25). Therefore, it has the potential to be a protector of the host for the infection.

Subjects with COVID-19 have unexpected colonization of the lung by TMV and Pangolin coronavirus, suggesting that they may play a role in the infection of SARS-CoV-2. It was reported that TMV could enter and persist in mouse lungs, raising questions about the potential interactions between the virus and human host(39). Accordingly, a further investigation was necessary. Our BALF metatranscriptome results detailed the species in the lungs of COVID-19 patients and shed light on the complex interactions between SARS-CoV-2 and other microbes in the lungs.

Conclusion

The metatranscriptome of human BALF can reveal transcriptionally active microbiota in the lungs. The microbes in the lungs have a great influence on human health and disease. We detailed the active microbiota landscape in both healthy individuals and COVID-19 patients and found the infection of SARS-CoV-2 could deeply change the microbiota in the lungs. This is the first detail description of lung colonization in COVID-19 patients. The enrichment of several microbes, Lactic Acid Bacteria, *Faecalibacterium prausnitzii* and TMV, the significant correlation between *Rothia*, TMV and

SARS-CoV-2 reveals a microbiota disorder and drastic battles between the host, SARS-CoV-2 and other microbes in the lungs. Our study provides insight into the active microbiota in the lungs of COVID-19 patients and will contribute to the understanding of the infection mechanism of SARS-CoV-2 and the treatment of the disease and complications.

Availability of Data

The raw datasets analyzed for this study were downloaded from Sequence Read Archive, EMBL-EBI ArrayExpress and National Genomics Data Center, China with project IDs, PRJCA002202, PRJNA601736, PRJCA002326, PRJNA605983, PRJNA615032 and PRJNA434133.

Funding Statement

Declaration of Competing interest

The authors have declared that no conflicts of interest

Authors' Contributions

Conceptualization: Z. J., K.H. and W.W.; **Data analysis:** Y.H., Z.J., J.S.; **Writing & Editing:** Y.H. and Z.J. All authors read and approved the final manuscript.

References

1. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;**579**(7798):270-3.

2. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020;**581**(7807):215-20.

3. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020;**181**(2):271-80.e8.

4. Li LQ, Huang T, Wang YQ, Wang ZP, Liang Y, Huang TB, et al. COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. J Med Virol. 2020.

5. Zhang H, Ai J-W, Yang W, Zhou X, He F, Xie S, et al. Metatranscriptomic Characterization of COVID-19 Identified A Host Transcriptional Classifier Associated With Immune Signaling. Clinical Infectious Diseases. 2020.

6.   Fan J, Li X, Gao Y, Zhou J, Wang S, Huang B, et al. The lung tissue microbiota features of 20 deceased patients with COVID-19. Journal of Infection. 2020.

7.   Yatera K, Noguchi S, Mukae H. The microbiome in the lower respiratory tract. Respir Investig. 2018;**56**(6):432-9.

8.   Marsh RL, Kaestli M, Chang AB, Binks MJ, Pope CE, Hoffman LR, et al. The microbiota in bronchoalveolar lavage from young children with chronic lung disease includes taxa present in both the oropharynx and nasopharynx. Microbiome. 2016;**4**(1):37.

9.   Matijašić M, Meštrović T, Paljetak H, Perić M, Barešić A, Verbanac D. Gut Microbiota beyond Bacteria-Mycobiome, Virome, Archaeome, and Eukaryotic Parasites in IBD. Int J Mol Sci. 2020;21(8).

10.   Kim D, Quinn J, Pinsky B, Shah NH, Brown I. Rates of Co-infection Between SARS-CoV-2 and Other Respiratory Pathogens. Jama. 2020;**323**(20):2085-6.

11.   Ren L, Zhang R, Rao J, Xiao Y, Zhang Z, Yang B, et al. Transcriptionally Active Lung Microbiome and Its Association with Bacterial Biomass and Host Inflammatory Status. mSystems. 2018;**3**(5).

12.   Pickard JM, Zeng MY, Caruso R, Núñez G. Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. Immunological Reviews. 2017;**279**(1):70-89.

13.   Zuo T, Liu Q, Zhang F, Lui GC-Y, Tso EY, Yeoh YK, et al. Depicting SARS-CoV-2 faecal viral activity in association with gut microbiota composition in patients with COVID-19. Gut. 2020:gutjnl-2020-322294.

14.   Shen Z, Xiao Y, Kang L, Ma W, Shi L, Zhang L, et al. Genomic diversity of SARS-CoV-2 in Coronavirus Disease 2019 patients. Clin Infect Dis. 2020.

15.   Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, et al. IDseq – An Open Source Cloud-based Pipeline and Analysis Service for Metagenomic Pathogen Detection and Monitoring. bioRxiv. 2020:2020.04.07.030551.

16.   Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;**13**(11):2498-504.

17.   Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. Emerg Microbes Infect. 2020;**9**(1):761-70.

18.   Chen L, Liu W, Zhang Q, Xu K, Ye G, Wu W, et al. RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. Emerg Microbes Infect. 2020;**9**(1):313-9.

19.   Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Møller R, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. Cell. 2020;**181**(5):1036-45.e9.

20.   Michalovich D, Rodriguez-Perez N, Smolinska S, Pirozynski M, Mayhew D, Uddin S, et al. Obesity and disease severity magnify disturbed microbiome-immune interactions in asthma patients. Nat Commun. 2019;**10**(1):5711.

21.   Naito M, Hirakawa H, Yamashita A, Ohara N, Shoji M, Yukitake H, et al. Determination of the genome sequence of Porphyromonas gingivalis strain ATCC 33277 and genomic comparison with strain W83 revealed extensive genome rearrangements in P. gingivalis. DNA Res. 2008;**15**(4):215-25.

22.   Gerner-Smidt P, Keiser-Nielsen H, Dorsch M, Stackebrandt E, Ursing J, Blom J, et al. Lautropia mirabilis gen. nov., sp. nov., a gram-negative motile coccus with unusual morphology isolated from the human mouth. Microbiology. 1994;**140** (Pt 7):1787-97.

23.   Hoffmann H, Schmoldt S, Trülzsch K, Stumpf A, Bengsch S, Blankenstein T, et al. Nosocomial urosepsis caused by Enterobacter kobei with aberrant phenotype. Diagn Microbiol Infect Dis. 2005;**53**(2):143-7.

24.   Högenauer C, Langner C, Beubler E, Lippe IT, Schicho R, Gorkiewicz G, et al. Klebsiella oxytoca as a causative organism of antibiotic-associated hemorrhagic colitis. N Engl J Med. 2006;**355**(23):2418-26.

25.   Zuo T, Zhang F, Lui GCY, Yeoh YK, Li AYL, Zhan H, et al. Alterations in Gut Microbiota of Patients With COVID-19 During Time of Hospitalization. Gastroenterology. 2020.

26.   Keller R, Pedroso MZ, Ritchmann R, Silva RM. Occurrence of virulence-associated properties in Enterobacter cloacae. Infect Immun. 1998;**66**(2):645-9.

27.   Ryan MP, Pembroke JT, Adley CC. Ralstonia pickettii: a persistent gram-negative nosocomial infectious organism. J Hosp Infect. 2006;**62**(3):278-84.

28.   Bag S, Ghosh TS, Das B. Complete Genome Sequence of Faecalibacterium prausnitzii Isolated from the Gut of a Healthy Indian Adult. Genome Announc. 2017;**5**(46).

29.   McArthur DB. Emerging Infectious Diseases. Nurs Clin North Am. 2019;**54**(2):297-311.

30.   Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, et al. Genomic variation landscape of the human gut microbiome. Nature. 2013;**493**(7430):45-50.

31.   Kotiranta A, Lounatmaa K, Haapasalo M. Epidemiology and pathogenesis of Bacillus cereus infections. Microbes Infect. 2000;**2**(2):189-98.

32.   Reed G, Jewett PH, Thompson J, Tollefson S, Wright PF. Epidemiology and clinical impact of parainfluenza virus infections in otherwise healthy infants and young children < 5 years old. J Infect Dis. 1997;**175**(4):807-13.

33.   Liu WK, Liu Q, Chen DH, Liang HX, Chen XK, Huang WB, et al. Epidemiology and clinical presentation of the four human parainfluenza virus types. BMC Infect Dis. 2013;13:28.

34.   Llopis M, Antolin M, Carol M, Borruel N, Casellas F, Martinez C, et al. Lactobacillus casei downregulates commensals' inflammatory signals in Crohn's disease mucosa. Inflamm Bowel Dis. 2009;**15**(2):275-83.

35.   Oksaharju A, Kankainen M, Kekkonen RA, Lindstedt KA, Kovanen PT, Korpela R, et al. Probiotic Lactobacillus rhamnosus downregulates FCER1 and HRH4 expression in human mast cells. World J Gastroenterol. 2011;**17**(6):750-9.

36.   Yuan Z, Panchal D, Syed MA, Mehta H, Joo M, Hadid W, et al. Induction of cyclooxygenase-2 signaling by Stomatococcus mucilaginosus highlights the pathogenic potential of an oral commensal. J Immunol. 2013;**191**(7):3810-7.

37.   Manfredo Vieira S, Hiltensperger M, Kumar V, Zegarra-Ruiz D, Dehner C, Khan N, et al. Translocation of a gut pathobiont drives autoimmunity in mice and humans. Science. 2018;**359**(6380):1156-61.

38.   Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux JJ, et al. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A. 2008;**105**(43):16731-6.

39. Balique F, Colson P, Barry AO, Nappez C, Ferretti A, Moussawi KA, et al. Tobacco mosaic virus in the lungs of mice following intra-tracheal inoculation. PLoS One. 2013;**8**(1):e54993.

Figures and legends:

Figure 1. Comparison of microbial α-diversity and loads between COVID-19 patients (COV) and healthy controls (HC). (a) and (c) Shannon index of bacteria and virus respectively. (b) and (d) Bacterial and viral loads respectively. Boxes represent the interquartile ranges (IQRs) between the first and third quartiles, and the line inside the box represents the median; whiskers represent the lowest or highest values within 1.5 times IQR from the first or third quartiles. * p-value < 0.05, ** p-value < 0.01, *** p-value <0.001, **** p-value <0.0001.

Figure 2. β-diversity of bacteria and viruses between COVID-19 patients (COV) and healthy controls (HC). (a) and (b) Principal coordinate analysis plot of bacteria and virus respectively based on Bray-Curtis distance. The pink dot indicates COVID-19 patients and the light green indicates healthy controls. R and p-value are the results of ANOSIM test.

Figure 3. Composition of the top 10 species in COVID-19 patients and healthy controls. (a) and (b) Relative abundance of the top 10 bacteria and viruses respectively in each sample. The size of the point corresponds to the relative abundance value. The color of the sample names represents the sample group, red represents the COVID-19 patient group, and blue represents the healthy control group. The order of the species in the legend from top to bottom represents the relative abundance from high to low. Project bar represents the sample sources (P1, PRJCA002202; P2, PRJNA601736; P3, PRJCA002326; P4, PRJNA605983; P5, PRJNA615032; P6, PRJNA434133).

Figure 4. Heat map of the species with strongly significant differences and important differential pathogens. (a) Heat map of differential bacteria with q-value < 1e-11, absolute value of $\log_2$(Fold Change) >3 and the maximum relative abundance greater than 0.05%. (b) Heat map of virus with q-value < 0.01 and absolute value of $\log_2$(Fold Change) >1. Project bar represents the sample sources (P1, PRJCA002202; P2, PRJNA601736; P3, PRJCA002326; P4, PRJNA605983; P5, PRJNA615032; P6, PRJNA434133). The color of the sample names represents the sample group, red represents the COVID-19 patient group, and blue represents the healthy control group. NC1 and NC2 are negative controls.

Figure 5. Correlation network of differential species between the COVID-19 patient and healthy control groups. (a) The correlation network within bacteria with a coefficient (r) > 0.8 and a statistical significance (q-value) < 1e-13. (b) The correlation network within viruses with a coefficient (r) > 0.5 and a statistical significance (q-value) <0.01. (c) The correlation between bacteria and viruses with a coefficient (r) > 0.7 and a statistical significance (q-value) < 5e-8. Diamonds represent viruses, circles represent bacteria. Blue line represents positive correlations and pink line represents negative correlations. Green represents up-regulation of the abundance of the species and pink represents down-regulation.

Supplementary tables:

Supplementary table 1. The basic information of the samples. The table lists the Project ID, sample ID, sample name, total reads, non-host reads, the percentage of non-host reads, water control and collection location.

Supplementary table 2. Differential bacteria at the species level (q-value < 1e-11, absolute value of $\log_2$(Fold Change) > 3).

Supplementary table 3. Differential viruses at the species level (q-value < 0.01, absolute value of $\log_2$(Fold Change) > 1).

Supplementary table 4. Correlation within differential bacteria (r > 0.8, q-value < 1e-13).

Supplementary table 5. Correlation within differential viruses (r > 0.5, q-value < 0.01).

Supplementary table 6. Correlation of differential bacteria with differential viruses (r > 0.7, q-value < 5e-8).

a

Relative abundance
- 0.0
- 0.2
- 0.4
- 0.6
- 0.8

Top 10 bacteria
- *Lautropia mirabilis*
- *Staphylococcus xylosus*
- *Variovorax paradoxus*
- *Variovorax sp. PMC12*
- *Mesorhizobium loti*
- *Staphylococcus simulans*
- *Enterobacter kobei*
- *Lacrimispora sphenoides*
- *Porphyromonas gingivalis*
- *Prevotella melaninogenica*

b

Project

P1 P2 P3 P4 P5 P6

Project

Relative abundance
- 0.00
- 0.25
- 0.50
- 0.75
- 1.00

Top 10 viruses
- Severe acute respiratory syndrome–related coronavirus
- Human respirovirus 3
- Enterovirus D
- Rhinovirus A
- Bacteriophage sp.
- uncultured human fecal virus
- Betacoronavirus 1
- Propionibacterium virus PHL199M00
- Moraxella phage Mcat14
- Spodoptera frugiperda rhabdovirus

a

	Project
	Actinomyces pacaensis
	Anaerocolumna sp. CBA3638
	Anaerotignum propionicum
	Brevibacterium frigoritolerans
	Campylobacter gracilis
	Campylobacter rectus
	Campylobacter showae
	Chryseobacterium cucumeris
	Citrobacter braakii
	Citrobacter freundii
	Citrobacter freundii complex sp. CFNIH3
	Clostridium beijerinckii
	Comamonas terrigena
	Comamonas testosteroni
	Enterobacter asburiae
	Enterobacter bugandensis
	Enterobacter cancerogenus
	Enterobacter cloacae
	Enterobacter hormaechei
	Enterobacter kobei
	Enterobacter oligotrophica
	Enterobacter roggenkampii
	Enterobacter soli
	Enterobacter sp. ODB01
	Enterobacter sp. T2
	Enterococcus faecium
	Faecalibacterium prausnitzii
	Filimonas lacunae
	Flavonifractor plautii
	Klebsiella oxytoca
	Kocuria sp. KD4
	Lacrimispora saccharolytica
	Lacrimispora sphenoides
	Lactobacillus curvatus
	Lactobacillus delbrueckii
	Lactobacillus fermentum
	Lactobacillus plantarum
	Lactobacillus reuteri
	Lactobacillus salivarius
	Leclercia adecarboxylata
	Leclercia sp. 1106151
	Leclercia sp. 119287
	Leclercia sp. 29361
	Leclercia sp. J807
	Leclercia sp. W17
	Leliottia jeotgali
	Leliottia sp. WB101
	Lysinibacillus sp. B2A1
	Lysinibacillus sphaericus
	Methylobacterium sp. DM1
	Mycoplasma hyorhinis
	Paenibacillus graminis
	Paenibacillus guangzhouensis
	Phyllobacterium myrsinacearum
	Pluralibacter gergoviae
	Pseudomonas alkylphenolica
	Pseudomonas oryzihabitans
	Pseudomonas sp.
	Pseudomonas sp. JY-Q
	Pseudomonas sp. PONIH3
	Rahnella aquatilis
	Ralstonia pickettii
	Rothia aeria
	Rothia mucilaginosa
	Ruminobacter amylophilus
	Schaalia meyeri
	Sphingobacterium sp. B29
	Sphingobacterium sp. G1-14
	Sphingomonas sp. FARSPH
	Stenotrophomonas sp. LM091
	Streptococcus salivarius
	Streptococcus sanguinis
	Streptococcus sp. oral taxon 064
	Succinivibrio dextrinosolvens
	Suiccocus acidiformans
	Treponema putidum
	uncultured Variovorax sp.
	Variovorax boronicumulans
	Variovorax paradoxus
	Variovorax sp. PBL-H6
	Variovorax sp. PMC12
	Vibrio anguillarum

b

	Project
	Autographa californica multiple nucleopolyhedrovirus
	Bamboo mosaic virus
	Bombyx mori nucleopolyhedrovirus
	Brevundimonas phage vB_BsubS-Delta
	Delftia phage RG-2014
	Human respirovirus 3
	Murine leukemia virus
	Nipah henipavirus
	Pangolin coronavirus
	Saccharomyces 20S RNA narnavirus
	Severe acute respiratory syndrome-related coronavirus
	Spodoptera frugiperda rhabdovirus
	Stenotrophomonas virus SMA7
	Sugarcane bacilliform virus
	Tobacco mosaic virus
	uncultured Caudovirales phage
	uncultured human fecal virus
	uncultured phage

P1 P2 P3 P4 P5 P6

Project

Log(rPM)   12 10 8 6 4 2 0

a

Enterobacter kobei *Leclercia sp.*
1106151 *Leclercia sp. J807*
Pseudomonas sp. PONIH3
Lelliottia jeotgali *Leclercia sp. 29361*
Enterobacter sp. ODB01 *Enterobacter oligotrophica*
Pseudomonas sp. XWY-1 *Citrobacter braakii*
Ruminobacter amylophilus *Pantoea sp. SO10*
Leclercia sp. W17 *Chryseobacterium cucumeris*
Succinivibrio dextrinosolvens *Leclercia sp. LSNIH1* *Anaerotignum propionicum*
Paenibacillus guangzhouensis *Enterobacter roggenkampii* *Ruminococcus gnavus*
Enterobacter soli *Pluralibacter gergoviae* *Lachnoclostridium phytofermentans*
Enterobacter cloacae complex sp. *Enterobacter bugandensis* *Lacrimispora sphenoides*
Citrobacter freundii complex sp. CFNIH3 *Lachnoclostridium phytofermentans*
Pseudomonas sp. MRSN12121 *Lactobacillus salivarius*
Enterobacter hormaechei *Lactobacillus reuteri*
Enterobacter cloacae

b

uncultured *Caudovirales* phage
Delftia phage RG-2014 *Brevundimonas* phage vB_BsubS-Delta
Sugarcane bacilliform virus uncultured phage
Bombyx mori nucleopolyhedrovirus Murine leukemia virus
Nipah henipavirus uncultured human fecal virus
Tobacco mosaic virus *Autographa californica* multiple nucleopolyhedrovirus
Spodoptera frugiperda rhabdovirus
Pangolin coronavirus Severe acute respiratory syndrome-related coronavirus *Saccharomyces* 20S RNA narnavirus *Stenotrophomonas* virus SMA7
Human respirovirus 3 Bamboo mosaic virus

c

◇ Virus
○ Bacteria
— Positive correlation
— Negative correlation

Leclercia sp.
Bamboo *Leclercia* Sugarcane *W17* *Delftia* *Clostridium*
mosaic *119287* bacilliform phage *carboxidivorans*
virus virus RG-2014
Enterobacter soli Stenotrophomonas virus SMA7 *Vibrio anguillarum*
Paenibacillus guangzhouensis *Saccharomyces* 20S RNA narnavirus
Pseudomonas sp. XWY-1 *Pseudomonas sp. JY-Q*
Succinivibrio dextrinosolvens Nipah henipavirus
Ruminobacter amylophilus *Lysinibacillus sp. B2A1*
Autographa californica multiple nucleopolyhedrovirus *Bombyx mori* nucleopolyhedrovirus
Leclercia sp. J807 Brevundimonas phage vB_BsubS-Delta
Enterobacter sp. ODB01 *Brevibacterium frigoritolerans*
Schaalia meyeri *Lelliottia jeotgali* *Leclercia sp. 1106151*
Rothia mucilaginosa Severe acute respiratory syndrome-related coronavirus uncultured human fecal virus *Enterobacter sp. E76*
Phyllobacterium myrsinacearum *Paenibacillus graminis* *Enterobacter bugandensis*
Tobacco mosaic virus *Leclercia sp. 29361* *Pseudomonas sp. MRSN12121*
Lactobacillus curvatus *Enterobacter cloacae complex sp.*
uncultured *Variovorax sp.* uncultured *Caudovirales* phage *Leclercia sp.*
Lelliottia sp. WB101 *Spodoptera frugiperda* rhabdovirus *Enterobacter sp. T2* *Leclercia sp. LSNIH1*
Enterobacter sp. 638 *Enterobacter oligotrophica* *Citrobacter freundii complex sp. CFNIH3* *Citrobacter braakii*