The clinical outcome of COVID-19 is strongly associated with microbiome dynamics in the upper respiratory tract

Xie, Linlin; Luo, Gengyan; Yang, Zhongzhou; Wu, Wei-chen; Chen, Jintao; Ren, Yuting; Zeng, Zhikun; Ye, Guangming; Pan, Yunbao; Zhao, Wen-jing; Chen, Yao-qing; Hou, Wei; Sun, Yanni; Guo, Deying; Yang, Zifeng; Li, Jun; Holmes, Edward C.; Li, Yirong; Chen, Liangjun; Shi, Mang

Published in:
Journal of Infection

Published: 01/03/2024

Document Version:
Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

License:
CC BY-NC-ND

Published version (DOI):
10.1016/j.jinf.2024.01.017

Publication details:

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.
The clinical outcome of COVID-19 is strongly associated with microbiome dynamics in the upper respiratory tract

Linlin Xie a,1, Gengyan Luo b,1, Zhongzhou Yang b, Wei-chen Wu b, Jintao Chen c, Yuting Ren c, Zhihun Zeng a, Guangming Ye a, Yunbao Pan a, Wen-jing Zhao b, Yao-qing Chen d, Wei Hou e, Yanni Sun e, Deying Guo b, Zifeng Yang 1, Jun Li f, Edward C. Holmes b, Yirong Li a,*, Liangjun Chen a,*, Mang Shi b,*

a Wuhan Research Center for Infectious Diseases and Tumors of the Chinese Academy of Medical Sciences/Hubei Engineering Center for Infectious Disease Prevention, Control and Treatment/Department of Laboratory Medicine, Zhongnan Hospital of Wuhan University, Wuhan, China
b State key laboratory for biocontrol, Shenzhen Key Laboratory of Systems Medicine for inflammatory diseases, School of Medicine, Shenzhen campus of Sun Yat-sen University, Shenzhen, China
c State Key Laboratory of Virology/Department of Laboratory Medicine/Hubei Province Key Laboratory of Allergy and Immunology, School of Basic Medical Sciences/Zhongnan Hospital, Wuhan University, Wuhan, China
d School of Public Health (Shenzhen), Shenzhen campus of Sun Yat-sen University, Sun Yat-sen University, Shenzhen, China
e Department of Electrical Engineering, City University of Hong Kong, Hong Kong, China
f State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, National Center for Respiratory Medicine, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
g Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, China
h Sydney Institute for Infectious Diseases, School of Medical Sciences, The University of Sydney, Sydney, Australia

Objective: The respiratory tract is the portal of entry for the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Although a variety of respiratory pathogens other than SARS-CoV-2 have been associated with severe cases of COVID-19 disease, the dynamics of the upper respiratory microbiota during disease course, and how they impact disease manifestation, remain uncertain.

Methods: We collected 349 longitudinal upper respiratory samples from a cohort of 65 COVID-19 patients (cohort 1), 28 samples from 28 recovered COVID-19 patients (cohort 2), and 59 samples from 59 healthy controls (cohort 3). All COVID-19 patients originated from the earliest stage of the epidemic in Wuhan. Based on a modified clinical scale, the disease course was divided into five clinical disease phases (pseudotime): “Healthy” (pseudotime 0), “Incremental” (pseudotime 1), “Critical” (pseudotime 2), “Complicated” (pseudotime 3), “Convalescent” (pseudotime 4), and “Long-term follow-up” (pseudotime 5). Using meta-transcriptomics, we investigated the features and dynamics of transcriptionally active microbes in the upper respiratory tract (URT) over the course of COVID-19 disease, as well as its association with disease progression and clinical outcomes.

Results: Our results revealed that the URT microbiome exhibits substantial heterogeneity during disease course. Two clusters of microbial communities characterized by low alpha diversity and enrichment for multiple pathogens or potential pathobionts (including Acinetobacter and Candida) were associated with disease progression and a worse clinical outcome. We also identified a series of microbial indicators that classified disease progression into more severe stages. Longitudinal analysis revealed that although the microbiome exhibited complex and changing patterns during COVID-19, a restoration of URT microbiomes from early dysbiosis toward more diverse status in later disease stages was observed in most patients.

List of abbreviations: SARS-CoV-2, Severe acute respiratory syndrome coronavirus 2; MERS-CoV, Middle East respiratory syndrome-related coronavirus; HSV-1, Human alphaherpesvirus 1; COVID-19, Coronavirus disease 2019; ICU, Intensive care unit; ECMO, Extracorporeal membrane oxygenation; PERMANOVA, Permutational multivariate analysis of variance; RF, Random forest; RPM, Reads per million; rpoB, RNA polymerase beta subunit; gyrB, DNA gyrase subunit B; ERG11, Lanosterol 14-alpha-demethylase; NLR, Neutrophil-to-lymphocyte ratio; NEUT, Neutrophil; LYMHPH, Lymphocytes; WBC, White blood cell/leukocyte; PLT, Platelets; HGB, Hemoglobin; ALT, Alanine aminotransferase; AST, Aspartate aminotransferase; IL-6, Interleukin-6; LDH, Lactate dehydrogenase; CRP, C-reactive protein; DD, D-dimer; PCT, Procalcitonin; SAA, Serum amyloid protein

1 These authors contributed equally to this work.
Introduction

COVID-19 is an acute respiratory disease caused by the betacoronavirus SARS-CoV-2\(^1–3\) that has resulted in large-scale morbidity and mortality with unprecedented economic and social disruption. COVID-19 shows a broad spectrum of clinical manifestations ranging from asymptomatic infection and mild respiratory symptoms to severe pneumonia or death.\(^4\) Many risk factors have been identified in the progression of COVID-19 into a severe and critical stage, including advanced age,\(^5–7\) genetic background,\(^8\) underlying co-morbidities,\(^9\) and elevated levels of inflammatory mediators (i.e., “cytokine storms”).\(^10\)

As the gateway to the respiratory system, the upper airway contains a highly diverse microbiome that could affect health and disease in the lower respiratory system through both direct and indirect mechanisms.\(^11–13\) Previous studies have shown that commensal microbes in the upper airways may protect the host from pandemic respiratory pathogens such as influenza virus.\(^14\) This occurs through mechanisms such as modulating host defenses against pathogen,\(^15\) modulating inflammatory responses,\(^16\) and regulating mucosal homeostasis.\(^17\) In patients infected with influenza viruses, the presence of other pathogens or opportunistic infections can be associated with severe disease and mortality, which is often attributed to dysbiosis and the dysregulation of immune responses.\(^18–20\)

In the case of COVID-19, previous studies have shown that low microbial diversity and elevated levels of opportunistic pathogens in the upper respiratory tract may exacerbate symptoms and contribute to disease severity.\(^21–28\) However, the impact of microbial co-pathogenesis in COVID-19 disease progression remains controversial: some studies have revealed that the dysbiosis and proliferation of pathogens or opportunistic pathogens in the respiratory tract can result in more severe disease manifestations;\(^23–28\) whereas others suggested that secondary infections are not associated with higher mortality.\(^29–32\) While most of these studies investigated the microbiome (mostly bacteriome) based on a cross-sectional design often with single time-point sampling for each patient, there is a general lack of longitudinal characterization of the dynamics of the entire viral, bacterial and eukaryotic microbial population following SARS-CoV-2 infection. Indeed, a highly complex dynamic pattern comprising multiple co-pathogens was observed in four prolonged cases of severe and critical COVID-19.\(^13\) It is therefore unclear how the microbiome as a whole impacts disease progression and clinical outcome.

To reveal microbial dynamics throughout COVID-19 the course of the disease, we conducted meta-transcriptomic sequencing of 349 longitudinal upper respiratory samples collected from 65 COVID-19 patients (cohort 1), as well as 28 samples from 28 recovered COVID-19 patients (cohort 2), and 59 samples from 59 healthy controls (cohort 3). Our analysis revealed that the URT microbiome was associated with COVID-19 disease manifestation, with significant differences between healthy normal flora and the microbial populations present during more severe disease stages. Accordingly, we were able to establish a link between microbial dynamics and disease manifestation and SARS-CoV-2 infection.

Conclusion: This study revealed strong links between URT microbiome dynamics and disease progression and clinical outcomes in COVID-19, implying that the treatment of severe disease should consider the full spectrum of microbial pathogens present.

Materials and methods

Study design and cohort

A total of 152 subjects, including 65 COVID-19 patients (longitudinal cohort 1), 28 healthy individuals recovered from COVID-19 (cohort 2, >3 months since hospital release), and 59 healthy controls with no known history of infection (cohort 3) were enrolled in this study (Fig. 1A and Supplementary Table 1). The sample collections of both COVID-19 patients and healthy controls were carried out at Zhongnan Hospital, Wuhan, between January 1st and September 30th, 2020. All COVID-19 patients enrolled were laboratory confirmed with a positive RT–PCR test. To describe the heterogeneous disease trajectories over time, a modified Clinical Score which considers an 11-point WHO Clinical Progression Scale,\(^33\) lung changes on Chest CT, and several inflammatory markers (serum c-reactive protein [CRP], serum IL-6, DD and SAA) (Supplementary Table 2) was used to classify patients during their disease course. Clinical disease phases were defined as pseudotimes in accordance with Clinical Score to depict the longitudinal course of the disease (Fig. 1B): incremental (pseudotime 1, where clinical symptoms and inflammatory markers were increasing, ICU or non-ICU), critical (pseudotime 2, ICU, mechanically ventilated with signs of ground-glass or consolidations scattered in the lungs), complicated (pseudotime 3, intermittent supplemental oxygen, minor signs of inflammation, gradual resolution of the ground glass and consolidations), convalescent (pseudotime 4, no supplemental oxygen), and long-term follow-up (pseudotime 5, at least three months after hospital discharge).

Sample collection and RNA extraction

Upper respiratory samples (throat swabs) were collected from the COVID-19 patients at multiple time points over the entire course of hospitalization (0–60 days). For each COVID-19 patient, 2–12 (medium 5) samples were collected (Supplementary Table 3), depending on the severity of disease and the duration of hospitalization. For recovered controls, samples were collected at least three months after they were released from hospitals. After sample collection, cotton swabs were immediately immersed in the lysis buffer of the RNA extraction kit (DAANGENE, Guangzhou, China) and stored at −80°C before further processing. Total RNA was subsequently extracted from these samples using the same kit provided by DAANGENE. Negative control groups comprised a cotton swab and collection device directed exposed to air and immersed in lysis buffer, and subsequently subjected to the same RNA extraction procedures.

Meta-transcriptomic library preparation and sequencing

RNA samples were first quantified by a Qubit 2.0 Fluorometer (Invitrogen, USA), dual-indexed and paired-end libraries were then constructed using the Trio RNA-Seq Kit (NuGEN Technologies, USA) designed for low-concentration RNA samples and contained an AnyDeplete probe that removes human ribosomal RNA. Briefly, the
Fig. 1. Overview of the study and clinical disease trajectory analyses of the COVID-19 patients. A, Graphical overview of the study cohorts. B, Definition of disease phase pseudotimes. Clinical disease phases (pseudotimes) were defined based on the WHO Clinical Progression Scale and inflammatory markers. These reflect temporal disease severity and distinguish between incremental and recovering disease stages. See Supplementary Table 2 for a detailed explanation of the scoring and pseudotimes. C, Overview of COVID-19 disease course, intervention, and sample collection of patients in cohort 1. All patients were temporally aligned to the day of COVID-19-related symptom onset. Frames mark days of in-patient care, and the color represents the disease pseudotime. The dates of entering and leaving of the ICU are indicated by arrows, intubations are depicted with an asterisk. Sampling during the hospitalization is marked with triangle symbols, while squares mark sampling after discharge. Upper respiratory specimens (throat swabs) were serially collected for meta-transcriptomic sequencing and RT-qPCR testing for SARS-CoV-2. All COVID-19 patients have at least two consecutive time points, while healthy controls and recovered patients have single time point.
library preparation steps involved removal of DNA, conversion of RNA to cDNA, ligation of adapters with unique barcodes, depletion of ribosomal cDNA, amplification and library clean-up. The quality of the libraries was assessed by Qsep100 (Biop tic, Taiwan, China), and quality-controlled libraries were then sequenced on a NovaSeq platform (Illumina, USA). Four negative controls (NCs) were prepared for library construction and meta-transcriptomic sequencing in parallel with the clinical samples while all the NCs had extremely low biomass.

Identification and quantification of viral, microbial and fungal communities in the respiratory tract

To identify and quantify transcriptionally active microbes from the sequencing results, an in-house pipeline that combines quality trimming, genomic mapping, and similarity searches using the NCBI protein (nr) and nucleotide (nt) databases,\(^1,3,7\) was performed on each of the sequenced samples. Reads containing adaptor sequences and low-complex regions were first removed from the data set using BBmap, version 38.62.\(^4\) The reads were then de-duplicated using CD-HIT, version 4.8.1.\(^5\) The remaining human and microbial rRNA reads were removed by comparisons against a comprehensive rRNA database (28S, 18S, 5.8S; and 23S, 16, and 5S rRNA downloaded from SILVA database\(^6\)), first by using bowtie2 program version 2.3.5.1\(^7\) and then by SortMeRNA version 4.2.0.\(^8\) Human reads were removed by mapping against a human reference genome using the bowtie2 program.

The remaining non-human and non-rRNA reads were compared against the NCBI non-redundant protein and reference virus genome databases for virus identification and quantification. The quantification of existing viral species (including SARS-CoV-2) was also performed by mapping reads against relevant virus genomes.

In the case of bacteria and fungi, sequence reads were first compared against the database implemented in MetaPhlAn2 version 1.0.0\(^9\) to identify eukaryotic and bacterial microbes at the genus level. Reference genome sequences of corresponding species within these genera were subsequently downloaded from GenBank, and rRNA reads in reference sequences were masked by Ns, and then used as a genomic mapping template. Reads from samples were mapped to the reference genome of all species supported by at least 1000 reads and species with a genome coverage of less than 1% were considered false positives. Reads assigned at the genera level were reallocated at the species level according to the relative proportion of each species in that genus. To identify bacterial and eukaryotic microbes to the species level, consensus sequences were derived from reads mapped to specific genes (e.g., rpoB, gyrB and ERG11) for phylogenetic analyses and similarity comparisons. Only those with 98% sequence identity or that clustered within the diversity of an existing species were regarded as the same species.

The relative abundance of transcriptionally active microbes was estimated based on the formula: number of mapped reads/number of total non-rRNA reads *1,000,000 (i.e. reads per million, or RPM). An RPM filter of >1 was used to select taxa for downstream differential analysis.

Statistical analysis

Patient characteristics, including age, gender and health-related characteristics were compared using chi-square tests or independent samples t-tests as appropriate. Microbial diversity was estimated using the phylseq\(^10\) and vegan\(^11\) packages implemented in R. Comparison of the alpha diversity values across different groups was performed using Kruskal-Wallis tests for comparisons across multiple groups. When applicable, pairwise comparisons of alpha diversity were performed using Dunn post hoc tests. Beta diversity between samples was calculated as Bray Curtis dissimilarity, permutational multivariate analysis of variance (PERMANOVA) and W’d test\(^12\) were used to compare differences in distance dissimilarity between groups. Permutational analyses of multivariate dispersions (PERMDISP) was used for assessing multivariate homogeneity of variance (beta dispersion) between groups. Furthermore, a covariate-adjusted principal coordinates analysis (aPCoA)\(^13\) was used to adjust for confounders (age and gender) and for visualization when comparing microbiome differences among all groups (pseudotimes 0–5).

To explore potential associations between the upper respiratory genus-level microbiota composition and the extensive metadata in COVID-19 patients (Cohort 1), distance-based redundancy analyses (dBRDA) were performed using the capsule function from vegan in univariate analyses with 33 metadata variables. Model P-values were corrected using Benjamini-Hochberg’s (BH) multiple-testing correction, to select 31 variables with BH-adjusted p-values < 0.05. These 31 variables were included in a multivariate model, and non-redundant contribution to variation was calculated using forward stepwise variable selection via the ordiR2step function from vegan.

Genus-level microbial community type analysis was performed according to the Dirichlet multinomial mixtures (DMM) using the Dirichlet Multinomial packages.\(^14\) The appropriate microbial community type numbers (DMM clusters) were determined based on the lowest Laplace approximation index. Significant associations between DMM clusters and disease stage were identified using a Fisher exact test.

Pairwise differential abundance analysis using DESeq2 was performed at the genus level to identify the significantly different taxa that contribute to disease progression. Given the possible impact of confounding factors, such as age, gender, and patient ID, the output from DESeq2 was further validated with Multivariate Association with Linear Models (MaAsLin2) by adjusting confounding factors. Therefore, only the taxa revealed by both methods were presented. Random forests analyses were used to identify variables that predict disease aggressiveness for COVID-19 patients using the R package “randomForest.” Random forest models, cross-validation, and Receiver Operating Characteristic (ROC) statistics were performed using stratification.

Since the sampling follows a longitudinal study design, longitudinal generalized additive mixed models (GAMM) were employed to examine changes of microbial diversity and key taxa through COVID-19 disease progression stages (pseudotimes 1–4). A GAMM incorporating age, gender, comorbidities and clinical outcome, a smooth for days post symptom onset or pseudotime, and a random patientID effect, was used to longitudinally model microbial diversity and log-transformed abundances of key taxa. Effect contributions were determined using analysis of variance tests. Models were fit using the gamm4, mgcv and lme4 packages.

Correlations between clinical metadata and the log-transformed abundances of URT microbiome were analyzed using multivariate linear regression test. All statistical analyses were performed using packages under R version 4.1.1.

Results

Subjects and specimens

Patient demographics, health-related characteristics, symptomatology and treatment are described in Table 1. The average age of the COVID-19 patients was 57.2 (range 7–96), comparable to that of controls (average: 50.2, and range 25–83). Most of the patients received antiviral therapy (96.9%) and antibiotic treatment (96.9%). More than half of the patients were given glucocorticoid treatment and oxygen support, with 20 (30.8%) subject to invasive ventilation and 6 (9.2%) being placed on veno-venous extracorporeal membrane oxygenation (ECMO).
To study the dynamics of the upper respiratory tract microbiome during SARS-CoV-2 infection, we collected and sequenced a total of 349 throat swabs from 65 COVID-19 patients collected at multiple time points during the course of their disease progression (Fig. 1C). For comparison, throat swabs were collected from 28 healthy individuals recovered from COVID-19 (>3 months since hospital release) and 59 healthy controls with no known history of infection within the same geographic region (Fig. 1A). Samples were taken at a median of 5 time points (range 2–12) for COVID-19 patients. More than three time points were collected from most patients (96.9%). To describe the heterogenous disease trajectories over time, we used the WHO Clinical Progression Scale, lung changes on Chest CT, and several inflammatory markers (serum c-reactive protein [CRP], interleukin-6 [IL-6], D-dimer [DD] and Serum amyloid protein [SAA]) (Supplementary Table 2) to classify disease phase into six stages denoted: “Healthy” (pseudotime 0), “Incremental” (pseudotime 1), “Critical” (pseudotime 2), “Complicated” (pseudotime 3), “Convalescent” (pseudotime 4), and “Long-term follow-up” (pseudotime 5) (Fig. 1A and C).

Respiratory microbial dysbiosis following SARS-CoV-2 infection

Our meta-transcriptomic characterization of the URT microbiome, which included RNA and DNA viruses, bacteria, and eukaryotic microbes, revealed that alpha diversity differed significantly among disease phases (Kruskal-Wallis test, \( P < 0.05 \)), suggesting that SARS-CoV-2 infection might be accompanied by a decreased microbial diversity in the URT (Fig. 2A and Supplementary Table 4). There were also significant differences in microbial composition among disease phases after adjusting for age and gender (PERMANOVA, \( R^2 = 0.15, P = 0.001 \); \( W*W \) test, \( P = 0.001 \)). These were partially attributed to the different dispersion among groups (PERMDISP, \( P = 0.001 \) (Fig. 2B). Interestingly, the differences were more marked between the Critical and other disease stages (\( R^2 = 0.08–0.30, P < 0.05 \), with PERMANOVA test) than those among the other stages (\( R^2 = 0.01–0.14, P < 0.05 \), with PERMANOVA test) (Fig. 2B, Supplementary Table 4), indicating a more substantial shift in community structure during the more severe stages of COVID-19.

We also explored potential associations between the upper respiratory microbial composition at the genus level and 29 covariates related to patient anthropometrics, clinical variables and medication in cohort 1 (Supplementary Fig. 1). Individually, 27 of 29 covariates showed a significant correlation with microbiota composition in a univariate analysis (dbRDA). Furthermore, 7 of the 27 significant covariates accounted for 53.8% non-redundant variation in this data set based on a multivariate analysis (dbRDA). Notably, patient ID (i.e., patient identity) had the largest impact on the microbiome composition, most likely due to the longitudinal sampling of the same COVID-19 patients, followed by disease stages/severity (Pseudotime and WHO Scale) and anti-fungal treatment before the time of sampling (Supplementary Fig. 1). Intubation also had a strong impact on microbial composition, although this covariate is nested within disease phase or severity-related covariants. Antimicrobial treatment at
the time of or before sampling also showed less but significant impact on the URT microbiome composition. Collectively, our results imply disease progression resulted in profound changes in the composition of URT microbiota.

**Altered microbial communities during disease progression**

Because of the high variability in microbial composition among individuals within disease phases, we also characterized the patterns of microbial community composition using Dirichlet multinomial mixtures (DMMs). Accordingly, we identified six community types (Supplementary Fig. 2A–C) significantly associated with the disease phase (Fig. 3A, \( P = 0.0005 \) with Fisher exact test). The abundance of the 121 detected microbial genera differed significantly between clusters (Kruskal-Wallis test, adjusted \( P < 0.05 \), Supplementary Table 5). DMM1 and DMM2 clusters were mainly associated with samples from Healthy controls and the Long-term follow-up stage, and were dominated by the commensal microbes in the URT described in previous studies\(^2\), such as Streptococcus, Gemella, Prevotella, Veillonella, and Capnocytophaga (Fig. 3B and Supplementary Fig. 2D). Conversely, the DMM3 and DMM4 clusters were mainly associated with samples at Critical stage, and dominated by the respiratory pathogen *Betacoronavirus* (which includes SARS-CoV-2) and potential pathobionts, namely, *Simplexvirus*, *Acinetobacter*, *Enterococcus*, and *Candida*, amongst others (Fig. 3B and Supplementary Fig. 2D).

Our results also revealed a distinction between what we considered “Commensal Type” and “Pathobiont Type” microbes at the genus level during COVID-19 disease. A total of 38 genera of the “Commensal Type” were identified based on commonly accepted classifications (Supplementary Table 5). These represent previously described normal flora of the upper respiratory tract, including *Campylobacter*, *Fusobacterium*, *Prevotella*, *Streptococcus* and *Veillonella*. Notably, the abundance of these microbes was significantly higher in the Healthy and Long-term follow-up stages than in other stages of COVID-19 disease (Fig. 3C, \( P < 0.01 \) with Dunn post hoc tests). Conversely, 30 genera of the “Pathobiont Type” (SARS-CoV-2 excluded), including *Simplexvirus* (HSV-1), *Acinetobacter*, *Klebsiella* and *Candida* were identified at higher relative abundance in the Incremental and Critical stages (Fig. 3D). Strikingly, the total relative abundance of the “Pathobiont Type” microbes was as high as 767,701 RPM (76.77% of total non-ribosomal RNA), indicating their active proliferation within the host respiratory system during critical phase of the disease progression (Supplementary Table 6).

At the species level, using detailed phylogenetic analyses of functionally conserved genes such as rpoB, gyrB and EFG1, we identified 11 “Pathobiont type” microbial species in more than 10% of patients, including *Acinetobacter baumannii*, *Candida albicans*, *Candida glabrata*, *Enterococcus faecalis*, and *Enterococcus faecium* (Supplementary Fig. 3). Several taxa, including *Acinetobacter baumannii*, *Enterococcus faecium*, *Staphylococcus epidermidis*, *Candida albicans*, *Candida glabrata* and Human alphaherpesvirus 1 (HSV-1), exhibited high relative abundance levels in the Incremental-Critical stages but not in the Healthy and Long-term follow-up stages (Supplementary Fig. 4).

**Microbial indicators of disease progression**

We next identified a series of microbial indicators that can be used to classify progression into more severe disease stages (Fig. 4A, absolute Log2FC ≥ 2.0 and \( P < 0.05 \)). Specifically, while *Betacoronavirus* was most enriched during the Incremental stage (Log2FC = 12.9, pseudovalue 1 vs 0), more “Pathobiont Type” microbes were enriched during the Critical stage (pseudovalue 2 vs 0 or 2 vs 1), such as *Simplexvirus* (Log2FC = 11.5, 2 vs 0; Log2FC = 11.5, 2 vs 1), *Acinetobacter* (Log2FC = 12.7, 2 vs 0; Log2FC = 5.9, 2 vs 1), *Candida* (Log2FC = 11.7, 2 vs 0), *Klebsiella* (Log2FC = 8.3, 2 vs 0; Log2FC = 5.2, 2 vs 1), *Citrobacter* (Log2FC = 8.5, 2 vs 0; Log2FC = 5.3, 2 vs 1) and *Enterococcus* (Log2FC = 8.6, 2 vs 0; Log2FC = 4.0, 2 vs 1). Conversely, “Commensal Type” microbes were at reduced abundance during the more severe stages of disease (pseudovalue 2 vs 1 or 2 vs 0), such as *Fusobacterium* (Log2FC = −8.9, 2 vs 0; Log2FC = −9.1, 2 vs 1), *Alloprevotella* (Log2FC = −10, 2 vs 0; Log2FC = −4.1, 2 vs 1) and *Tannerella* (Log2FC = −7.8, 2 vs 0; Log2FC = −4.0, 2 vs 1). Similarly, using MaAsLin2 mixed linear modeling adjusting for confounders (Fig. 4B, \( P < 0.05 \)), we found that *Betacoronavirus* was most positively associated with Incremental stages (coefficient = 5.0, pseudovalue 1 vs 0), while multiple “Pathobiont Type” microbes, including *Acinetobacter* (coefficient = 8.1, 2 vs 0; coefficient = 8.8, 2 vs 1), *Citrobacter* (coefficient = 5.0, 2 vs 0; coefficient = 5.0, 2 vs 1), *Pseudomonas* (coefficient = 5.0, 2 vs 0; coefficient = 4.2, 2 vs 1), *Klebsiella* (coefficient = 4.2, 2 vs 0; coefficient = 4.9, 2 vs 1) and *Simplexvirus* (coefficient = 2.5, 2 vs 0; coefficient = 2.5, 2 vs 1), were positively associated with Critical stages (pseudovalue 2 vs 1 or 2 vs 0).
MaAsLin2 analysis identified *Streptococcus* as the most important microbe negatively associated with the disease progression into critical stages (coefficient = −8.6, pseudotime 2 vs 1).

Based on the microbial changes in COVID-19 patients, we also used a random forest (RF) model to classify critical from incremental stages in COVID-19 patients. We examined differently abundant URT microbiota as potential indicators of disease progression, as well as patient demographics (age, gender and comorbidity) and microbial community types, and conducted random forest analysis (Supplementary Fig. 5). At the level of the individual genus, the URT
pathobiont *Acinetobacter* was most informative (Gini index, accuracy) at classifying critical from incremental stages. The next highest-ranked taxa included *Streptococcus* – the most abundant commensal URT microbiota in our study (Supplementary Fig. 5A). Using the tenfold cross-validation of the RF model, we selected three microbial genera—*Acinetobacter*, *Streptococcus* and *Citrobacter* - for the classification of critical from incremental stages (Supplementary Fig. 5B). In the training set, microbial indicators could differentiate critical from incremental stages with a mean area under the receiver operating characteristic curve (AUC) of 0.878 ([95% CI] = [0.813, 0.942], Supplementary Fig. 5C). For the independent test set, the three indicators clearly stratified critical from incremental stages with an AUC of 0.909 ([95% CI] = [0.806, 1], Supplementary Fig. 5D). This result indicated that URT microbial patterns had the potential to be indicators of disease progression.

**Temporal dynamics of the microbiome in the upper respiratory tract following SARS-CoV-2 infection**

Our study also revealed a complex and highly dynamic microbial composition associated with the more severe phase of disease progression. By monitoring the microbial dynamics (Fig. 5A and Supplementary Fig. 6) and change of microbial community types (DMM clusters) over time (Supplementary Fig. 7) in COVID-19 patients (longitudinal cohort 1), we identified the presence of pathobionts, a constant fluctuation in the presence of prevailing pathobionts, and a suppression of commensals during the severe phase of infection (Fig. 5A, Supplementary Fig. 6). The longitudinal GAMM analysis revealed that the relative abundance of most that pathobionts, including HSV-1, *Acinetobacter baumannii*, *Corynebacterium striatum*, *Klebsiella pneumoniae*, *Candida albicans* and *Candida glabrata*, increased during the critical phase and then decreased during the convalescent phase (Supplementary Fig. 8A, *P* < 0.001), while normal flora showed the reverse trend. Although the abundance of SARS-CoV-2 varied from case-to-case, a general trend was relatively lower or undetectable viral abundance levels towards the later stage of disease progression, even as the disease entered the more critical phase (*P* < 0.001). Although alpha diversity of URT microbiome in COVID-19 patients was stable over the disease progression (Supplementary Fig. 8B, *P* = 0.666), the composition of the URT microbiome changed over the course of the disease, as shown in Bray-Curtis dissimilarity to the healthy controls (*P* < 0.001).
We further stratified our patient cohort 1 on the basis of clinical outcomes; namely, Mild (i.e., patients who only had mild symptoms), Severe (those who become critically ill but recovered) and Deceased (those who died). Subgroup analyses revealed that the magnitude of microbiota change in deceased patients was higher than that of non-deceased patients (Mild and Severe) across the disease course (Fig. 5C, \( P < 0.05 \) for both alpha diversity and beta dissimilarity by longitudinal GAMM model). Specifically, at the early stage (pseudotime 1) DMM3 and DMM4 were enriched in deceased patients, while DMM1 and DMM2 were most enriched in the Mild stage.
Group, less in the Severe Group, and completely depleted in the Deceased Group (Fig. 5C). Indeed, DMM3 and DMM4 were associated with worse outcomes using Cox proportional hazard model analyses, adjusted for age, gender and comorbidities (Fig. 5D). We correlated the URT microbiota with COVID-19 disease mortality using MaAsLin2 analysis adjusting for age, gender, and comorbidities. A total of 38 URT microbiota showed significantly negative correlation with COVID-19 mortality and 35 URT microbiota, which also enriched in DMM3 or DMM4 clusters, and had a significantly positive correlation with COVID-19 mortality (Supplementary Table 7; FDR corrected P < 0.05). Our longitudinal GAMM analysis also revealed that many taxa were associated with clinical outcome over the disease course (Fig. 5E). COVID-19 patients exhibited significant decreases in the prevalence of SARS-CoV-2 over the disease course. Pathobiont species, including Acinetobacter baumannii, Corynebacterium striatum, Klebsiella pneumoniae, Staphylococcus epidermidis, Candida albicans and Candida glabrata, increased in prevalence over the disease course of the Deceased Group. In contrast, commensal taxa, such as Streptococcus and Veillonella, decreased in prevalence over the disease course in the Deceased Group. Overall, these data show that clinical outcome in COVID-19 patients was associated with dynamic changes in the URT microbiome.

**Association between the respiratory tract microbiota and clinical indicators**

We next explored the correlation between URT microbiome and clinical indicators in patients with COVID-19. Notably, the concentration of multiple inflammation indicators, such as neutrophil counts (NEUT), leukocyte counts (WBC), neutrophil to-lymphocyte ratio (NLR), procalcitonin (PCT), D-dimer (DD), interleukin-6 (IL-6) and lactate dehydrogenase (LDH), was positively correlated with mortality at early stage and all stages (Fig. 6A). Revealing exacerbated and prolonged inflammation response in deceased patients. Furthermore, immunological biomarkers, such as CD4 + T, CD8 + T, B cells and NK cells, significantly declined across disease course in deceased patients, suggesting that poor immune response might result in poor outcomes in COVID-19 patients.

We further investigated the associations between individual microbes and clinical features. This revealed a positive correlation between the abundance of pathobiont microbiota and multiple inflammation indicators, such as NLR, IL-6, DD, WBC, NEUT, PCT and LDH. Strikingly, the opposite trend was observed between the normal flora and these indicators. Some of the immunological biomarkers, such as LYMHPH, T cells, CD4+ T cells and CD8+ T cells, showed a strong positive correlation with commensal microbiota, and negative correlations with pathobiont microbiota (Fig. 6B). At the species level in the “Pathobiont Type” group, SARS-CoV-2 only showed positive correlations with NLR, while Candida albicans and Acinetobacter baumannii exhibited the strongest correlations with NLR, WBC and NEUT, and were also negatively associated with the immunological biomarkers (Fig. 6C).

**Discussion**

Determining whether SARS-CoV-2 infection alters URT microbiota to affect COVID-19 disease progression and clinical outcome is a key question in virus pathobiology. Herein, we report a longitudinal investigation of the URT microbiome, with sampling during different times of COVID-19. These data provided three major observations. First, distinct microbiome profiles were seen during the progression of COVID-19 disease compared to healthy controls, with two types of community microbiome associated with lower alpha diversity and enrichment for pathobiont microbes (e.g., Acinetobacter and Candida) during the critical phase of COVID-19 as well as a worse clinical outcome. Second, a series of microbial indicators—including Acinetobacter, Citrobacter, Pseudomonas, Klebsiella, Simplexvirus and Streptococcus—were identified that classified progression into more severe disease stages. Third, the dynamic nature of the URT microbiome was associated with COVID-19 mortality and proinflammatory cytokines.

The respiratory microbiome plays a major role in individual health, and changes in microbial composition have been observed in various chronic and acute diseases. We observed alterations of URT microbiota in COVID-19 throughout disease progression, and found that COVID-19 patients experienced a sustained “microbial dysbiosis” similar to those following the infections with influenza virus, and MERS-CoV, which was consistent with previous reports. More importantly, our data revealed that such dysbiosis is characterized by the active replication of multiple pathobionts at very high abundance, and with constantly shifting combinations throughout the most severe phase of COVID-19, even after the principal pathogen (SARS-CoV-2) became undetectable. This is a clear demonstration of the “Anna Karenina principle”, with dysbiotic microorganisms being more variable and more diverse than those in healthy individuals.

In addition to disease phase or severity, it is important to take into account confounding covariates that include treatment, anthropometrics, and comorbidities that could contribute to the variation of microbial composition. Indeed, a previous study reported that time in the intensive care unit, type of oxygen support, as well as associated treatments with antibiotics explain most of the variation within the upper respiratory tract microbiome in COVID patients. Nevertheless, with longitudinal sampling, our results indicated that disease progression or severity play a major role in the variation of microbial composition, although other covariates such as anti-fungal treatments, intubation, and glucocorticoid treatment are also important contributors. In addition, our results indicate that antibiotic administration is a significant but less important contributor to microbial composition, although it has been shown to be of importance in other studies. However, it is important to note that 96% of the patients in our study received antibiotics, and that because there was a general lack of information on the specific types or dosage of antibiotic usage in many cases, we were unable to perform an in-depth analysis of the impact of antibiotics on the microbiome.

Our study also reveals the community types and a wide spectrum of viral, bacterial, and fungal taxa that are associated with the disease progression of COVID-19. DMMs revealed two clusters in URT communities with enrichment for pathobionts that were more likely to appear in the critical disease phase, itself significantly associated with a worse clinical outcome. Interestingly, while the commensal microbes were similar between different individuals and comparable to those identified from other studies, the pathobionts showed striking differences between individuals and between different studies. For example, Aspergillus fungus was widely recognized as an opportunistic pathogen in COVID-19 cases from India, Europe, and Dongguan, China, but it was not identified here. Furthermore, 11 pathobiont species, including HSV-1, Acinetobacter baumannii and Candida albicans, were identified in more than 10% of patients with high abundance during the critical stage of disease, implying that a broader range of microbes might contribute to disease progression and even death following SARS-CoV-2 infection.

Despite the wide diversity of microbes identified here, our results highlighted the role of herpes virus HSV-1 as a strong and stable indicator, if not contributor, to the progression into critical stage of COVID-19 disease. Several studies have highlighted the occurrence of HSV-1 reactivation in critically ill patients with COVID-19. While it is possible that the proliferation of HSV-1 in severe stage is associated with COVID-19-related psychological and physiological stress, some studies have shown that HSV-1 reactivation is
associated with an increased risk of mortality and pneumonia in critically ill COVID-19 patients. Others have demonstrated that opportunistic HSV-1 infection contributes to systemic diseases such as liver impairment or failure. It is also possible that HSV-1, as a neurotropic and neuroinvasive virus, might be associated with the neurological complications observed in many COVID-19 patients, such as loss of smell or taste, although this clearly needs to be demonstrated with more data.

The microbiome has been suggested to constitute a signaling hub regulating host immunity, mucosal homeostasis and defense against pathogen. It is possible that the replication of SARS-CoV-2 in the respiratory tract may result in a disruption of respiratory physiology and/or an impairment of immunity that facilitates proliferation of pathobiont microbes, although it is unclear whether SARS-CoV-2 infection directly increases receptor availability for other pathogens. Regardless of what causes dysbiosis, the outgrowth of pathobiont microbes in the respiratory tract will likely lead to further tissue damage and immune-mediated pathology, resulting in higher morbidity and mortality long after the acute phase of SARS-CoV-2. Indeed, the microbes identified here showed a strong correlation with multiple inflammation indicators, such as NLR, IL-6, DD, WBC, neutrophils; LYMHP, lymphocytes; NLR, neutrophil/lymphocyte ratio; ALT, alanine aminotransferase; AST, aspartate aminotransferase; LDH, lactate dehydrogenase; CRP, c-reactive protein; SAA, serum amyloid protein; PCT, procalcitonin; IL-6, interleukin-6.

Fig. 6. Correlation between the URT microbiota and clinical indicators. A. Clinical indicators were associated with mortality during hospitalization identified by multivariate linear regression. B. The URT microbiota whose abundances were correlated with clinical indicators are shown in the heatmap. The color in the heat map represents the correlation coefficients. *P < 0.05, **P < 0.01, and ***P < 0.001. WBC, white blood cell; HGB, hemoglobin; PLT, platelets; NEUT, neutrophils; LYMHP, lymphocytes; NLR, neutrophil/lymphocyte ratio; ALT, alanine aminotransferase; AST, aspartate aminotransferase; LDH, lactate dehydrogenase; CRP, c-reactive protein; SAA, serum amyloid protein; PCT, procalcitonin; IL-6, interleukin-6.
microbiota, while the latter could further modulate the local immune response. Hence, the association between microbiota and mortality might be mediated by the host immune response.

Overall, our study revealed that the URT microbiome is highly dynamic and associated with disease progression and hence the clinical outcome of COVID-19. It is therefore important to carefully track microbial composition during severe disease, perhaps adjusting treatment plans accordingly. Additional research is needed to determine the underlying mechanisms of the interactions between SARS-CoV-2 and the URT microbiome, such that an effective strategy can be developed to prevent severe COVID-19 or severe respiratory tract infections generally.

Our study has several limitations. First, there is a lack of consistent sampling time intervals, which makes it difficult to make comparisons across different patients at each stage of the disease. Although pseudotimes are used, they are defined using various scoring systems and clinical data such that it might be subjective. Second, since all subjects in this study were enrolled early in the COVID-19 pandemic when there were insufficient medical resources conduct this study, many of the clinical tests were not performed and much information is either missing or incomplete, which prevents a more detailed investigations of the microbial impact on clinical indicators. Finally, lower respiratory samples were not taken for this study, which would show microbial changes in the lungs.

Conclusions

The data generated here imply that respiratory microbial dysbiosis is a major manifestation during COVID-19. The microbiome dynamics in the upper respiratory tract are strongly associated with progression of COVID-19, clinical indicators and outcome, reflecting the interplay between pathobionts, symbionts, and the host immune status. Combined, our data highlight the significance of the URT microbiome in COVID-19 disease progression and potentially therapeutic.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Zhongnan Hospital, Wuhan University (No. 2021024) and Ethics Committee of Sun Yat-sen University (No. 2020034). Informed consent was obtained from each patient or the immediate family. All sample processing and sequencing procedures were performed in accordance with the relevant guidelines and regulations of the biosafety facility of Zhongnan Hospital, Wuhan University.

Funding

This work was supported by the National Natural Science Foundation of China (82000021), Shenzhen Science and Technology Program (KQT202000820145822203 and RCJ202107060920009004), Shenzhen Key Laboratory of Systems Medicine for inflammatory diseases (ZDSYS202060610083007), Guangdong Province “Pearl River Talent Plan” Innovation and Entrepreneurship Team Project (2019ZT08Y464), the Key Research and Development Program of Hubei Province (2022BCA019), 2021 Hubei Provincial Leading Public Health Talents Project (WSJKRC2022012), the Young Elite Scientist Sponsorship Program By CAST (YESS20200394), Joint Funds of the National Natural Science Foundation of China (U20A20396), Hong Kong Innovation and Technology Fund (ITF) (MRP/071/20X), AIR@InnoHK administered by the Innovation and Technology Commission, Hong Kong (Institute of Data Discovery for Health), the National Health and Medical Research Council (GNT2017197), the Health and Medical Research Fund (COVID190206), Open Project of State Key Laboratory of Respiratory Disease (SKLD-OP-202001) and Guangzhou Institute of Respiratory Health Open Project (2020GIRHHMS01).

Credit authorship contribution statement


Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge all health care workers involved in the diagnosis and treatment of patients in Wuhan. We give special thanks to the Alibaba Cloud Computing Co. Ltd. for providing the computational resources for rapid data processing.

Consent for publication

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

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jinf.2024.01.017.

References

11. Baisis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, et al. Analysis of the upper respiratory tract microbiota as the source of the lung and


