Dynamics of the upper respiratory tract microbiota and its association with fatality in COVID-19 patients

Mingkun Li (limk@big.ac.cn)

Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Center for Bioinformation

Lili Ren

NHC Key Laboratory of Systems Biology of Pathogens Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College

Yeming Wang

Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, National Clinical Research Center for Respiratory Diseases, China-Japan Friendship Hospital

Jiaxin Zhong

Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Center for Bioinformation, Beijing, 101300, China;

Dingyu Zhang

Wuhan Jinyintan Hospital

Yan Xiao

Institute of Pathogen Biology, Chinese Academy of Medical Sciences

Jie Li

Tsinghua University

Jing Yang

Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Center for Bioinformation, Beijing, 101300, China;

Guohui Fan

Institute of Respiratory Medicine, Chinese Academy of Medical Science, Beijing

Li Guo

Institute of Pathogen Biology, Chinese Academy of Medical Sciences

Zijie Shen

Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Center for Bioinformation

Wen Liu

Department of Tuberculosis and Respiratory Disease, Jinyintan Hospital, Wuhan

Lu Kang

Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Center for Bioinformation

Leisheng Shi

Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Center for Bioinformation

Xia Li

Jin Yin-tan Hospital

Qiong Li

BaiJiaKangRan

Jizhou Li

School of Life Sciences, Tsinghua-Peking Center for Life Sciences, Tsinghua University

Lin Di

Beijing Advanced Innovation Center for Genomics, Biomedical Pioneering Innovation Center, Peking University

Haibo Li

Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, National Clinical Research Center for Respiratory Diseases, China-Japan Friendship Hospital

Conghui Wang

Institute of Pathogen Biology, Chinese Academy of Medical Sciences

Ying Wang

NHC Key Laboratory of Systems Biology of Pathogens and Christophe Mérieux Laboratory, Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College

Xinming Wang

NHC Key Laboratory of Systems Biology of Pathogens and Christophe Mérieux Laboratory, Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College

Xiaohui Zhou

Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, National Clinical Research Center for Respiratory Diseases, China-Japan Friendship Hospital

Jian Rao

NHC Key Laboratory of Systems Biology of Pathogens and Christophe Mérieux Laboratory, Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College

Li Zhang

Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Center for Bioinformation, Beijing, 101300, China;

Jianbin Wang

Tsinghua University https://orcid.org/0000-0001-6725-7925

Yanyi Huang

Peking University https://orcid.org/0000-0002-7297-1266

Bin Cao

Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, National Clinical Research Center for Respiratory Diseases, China-Japan Friendship Hospital, Beijing
https://orcid.org/0000-0001-6991-0350

Jianwei Wang
Chinese Academy of Medical Sciences and Peking Union Medical College https://orcid.org/0000-0002-1116-4559

Article

Keywords: Human Microbiota Alteration, Metatranscriptome Sequencing, Longitudinal Oropharyngeal Swab Specimens, Streptococcus, Candida, Enterococcus, Prognostic Biomarker

Posted Date: November 30th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-95239/v1

Abstract

The pandemic of Coronavirus disease 2019 (COVID-19) is ongoing globally, which is a big challenge for public health. Alteration of human microbiota had been observed in COVID-19. However, it is unknown how the microbiota is associated with the fatality in COVID-19. We conducted metatranscriptome sequencing on 588 longitudinal oropharyngeal swab specimens collected from 192 COVID-19 patients recruited in the LOTUS clinical trial (Registration number: ChiCTR2000029308) (including 39 deceased patients), and 95 healthy controls from the same geographic area. The upper respiratory tract (URT) microbiota in COVID-19 patients differed from that in healthy controls, while deceased patients possessed a more distinct microbiota. *Streptococcus* was enriched in recovered patients, whereas potential pathogens, including *Candida* and *Enterococcus*, were more abundant in deceased patients. Moreover, the microbiota dominated by *Streptococcus* was more stable than that dominated by other species. In contrast, the URT microbiota in deceased patients showed a more significant alteration and became more deviated from the norm after admission. The abundance of *Streptococcus* on admission, particularly that of *S. parasanguinis*, was identified as a strong predictor of fatality by Cox and L_1 regularized logistic regression analysis, thus could be used as a potential prognostic biomarker of COVID-19. The generalization of the results in other populations and underlying mechanisms needs further investigations.

Introduction

Coronavirus disease 2019 (COVID-19) has infected more than 30 million people worldwide. Although most patients showed mild symptoms or were asymptomatic[1], approximately 14% developed severe diseases, 5% were critically ill[2], and the overall fatality rate is 3.2%. Older people and patients with underlying diseases are at increased risk for severe illness from COVID-19 and have a higher fatality rate. Other risk factors include smoking history, pregnancy, male, and obesity[3]. Genetic variants on toll-like receptor 7, ABO blood group locus, and 3p21.31 gene cluster were also associated with the severe COVID-19[4,5]. Meanwhile, biomarkers to monitor disease progression and predict clinical outcomes have been developed, including antibody concentration[6], serum proteins, metabolites[7], as well as combinations of regular inflammatory and coagulation markers (e.g., procalcitonin, interleukin 6, and D-dimer)[8].

Human microbiota plays a crucial role in individual health; alteration of the human microbiota has been observed in various chronic and acute diseases[9]. In particular, microorganisms residing in the gastrointestinal tract (GIT) and upper respiratory tract (URT) can alter the susceptibility to and outcomes of infectious diseases[10]. The underlying mechanisms include colonization resistance and induction of the immune responses in the host[11]. In recent studies, the diversity and composition of GIT microbiota showed significant differences between COVID-19 patients and healthy controls (HC)[12–15]. Bacterial and fungal opportunistic pathogens were enriched in the feces of COVID-19 patients, and a secondary infection was suspected, suggesting that dysbiosis of the gut microbiota is common in COVID-19 patients[12–16]. Notably, some differential microbes could potentially alter the expression of angiotensin-

converting enzyme 2 (ACE2)[17], which is the receptor used by SARS-CoV-2 to enter the host, thereby influence the susceptibility and the severity of the infection. Besides, some commensal bacteria are able to modify the heparan sulfate on epithelial cells and prevent the virus from binding to the host cell[18]. Respiratory microbiota contributes to the foremost barrier to viral infection[19], how it alters in COVID-19 patients is largely unknown[20–22]. Moreover, there is no study on the association between the microbiota and the fatality risk, which is one of the most crucial questions regarding microbiota's contribution to the health of COVID-19 patients.

The power and validity of the previous studies are limited by the small sample size and overrepresentation of patients with mild symptoms. To obtain a thorough understanding of the association between microbiota and COVID-19, especially those related to the fatality, a large cohort with deceased patients is needed. Lopinavir Trial for Suppression of SARS-CoV-2 in China (LOTUS) was a clinical trial that aims to evaluate the efficacy and safety of lopinavir-ritonavir treatment for SARS-CoV-2 infection[23]. LOTUS recruited 192 severe COVID-19 patients, with an overall mortality rate of 22.1%. Serial oropharyngeal swab samples were collected. By analyzing metatranscriptome data from 588 COVID-19 samples and 95 health controls (HC), we found that the URT microbiota was significantly different between recovered and deceased patients on admission and afterwards. The abundance of *Streptococcus*, particularly *S. parasanguinis*, was strongly associated with the fatality risk. Meanwhile, we found that the restoration of the URT microbiota lags behind the recovery of the disease.

Results

Overview of the samples and sequencing data

Metatranscriptome sequencing was conducted on 588 oropharyngeal swab specimens (OPs) collected from 192 severe COVID-19 patients. Consecutive OPs were obtained on day 1, 5, 10, 14, 21, and 28 after admission when the patient's condition allowed (Fig. 1A). Most patients (147) recovered within 28 days after admission, while 38 patients died within 28 days. The OPs were collected at least twice from 177 patients (Fig. S1A). For comparison, OPs were collected from 95 healthy individuals without any pulmonary diseases in the community from Wuhan in April, 2020. Deionized water was used as the negative control (NC), which was processed following the same protocol as for clinical samples.

The median age of the patients was 58 (IQR 49, 68), and 78 (40.6%) of them were females, while the median age of the HC was 47 (IQR 33, 61) and 57 (60%) of them were females. The median days from symptom onset to the recruitment of the patients were 13 (IQR 11, 17) days. A large fraction of patients had at least one comorbidity or underlying health condition, which included hypertension (64), diabetes(22), heart disease(14), etc. All but four patients took antibiotics. We identified multiple risk factors, including age, severity on admission (severity-A), and corticosteroid use, significantly correlated with the fatality (Fig. S1B).

OPs were sequenced to a median number of reads of 33.0 million (IQR 9.3 million, 63.9 million). After filtering out the low-quality reads, host reads, and ribosome RNA reads, the median number of reads mapped to the microbial genome was 1.0 million (IQR 0.2 million, 2.7 million). We required a minimum of 1,000 microbial reads for the following analysis, which ensured that at least 92% of the microbes with relative abundance (hereinafter referred to as "abundance") greater than 0.1% could be detected (Fig. S1C); 677 samples were qualified.

Microbiota composition in the upper respiratory tract of COVID-19 patients

Bacteria were the most abundant microbes in the OPs of the COVID-19 patients, which accounted for 97.02% of the total microbial reads, followed by fungi (2.76%), viruses (0.22%), and archaea (0.006%). The abundance of fungi in COVID-19 patients was significantly higher than that in HC (median abundance: 2.76% vs. 1.08%, $p<0.001$), which was mainly contributed by the elevated level of *Candida* in COVID-19 patients (1.87% vs. 0.0003%, $p<0.001$). The microbial community was mainly composed of *Streptococcus*, *Veillonella,* and other known commensal microbes in both COVID-19 and HC (Fig. 1B). The overall microbial composition was different between COVID-19 patients and HC ($R^2=0.05$, $p<0.001$ on admission; $R^2=0.04$, $p<0.001$ before discharge/death; PERMANOVA test, Fig. 1C). *Rothia* and *Actinomyces* were significantly enriched in COVID-19 patients, whereas *Streptococcus*, *Capnocytophaga*, and other seven genera were more abundant in HC (Fig. S1D). Notably, the microbiota in deceased patients significantly differed from that in recovered patients, which was more remarkable before discharge/death ($R^2=0.04$, $p<0.001$ on admission; $R^2=0.10$, $p<0.001$ before discharge/death; PERMANOVA test). Additionally, we found no systematic difference between Lopinavir-Ritonavir recipients and standard care controls ($R^2=0.0014$, p=0.56, PERMANOVA test).

The hierarchical clustering algorithm identified 11 clusters (CSs) (Fig. 1D). COVID-19 samples were assigned to all CSs, while HC samples was assigned to CS1, CS3, and CS9. Most samples from patients (403/571) and HC (75/88) were assigned to CS1, which was dominated by *Streptococcus*. CS7 and CS8 were enriched for potential pathogens *Enterococcus* and *Candida*, and all samples in these two CSs were collected from patients, thereby indicating a possible co-infection or secondary infection. CS2, 4-6, 9-10 were dominated by *Prevotella*, *Lautropia*, *Schaalia*, *Rothia*, *Capnocytophaga*, and *Veillonella*, respectively, all of which were known commensal microbes in the oral cavity or URT. In particular, CS11 consisted of outliers that cannot be assigned to any other CSs, which were mostly dominated by opportunistic pathogens (Fig. S1E). All clusters but CS3 showed a great dissimilarity to NCs (Fig. S1F). We found that CS3 was more likely to be observed in samples from the same individual ($p<0.05$, Fig. S1G), reflecting a relatively barren state of microbiota in these samples.

OPs from the same individuals showed a higher similarity than that from different individuals (Fig. 1E), indicating the existence of an individual-specific microbiota. Meanwhile, samples from deceased patients showed larger variance among samples, suggesting that the URT microbiota's alteration was more evident in severe cases. Besides, the similarity between the microbiota of samples from the same

individual was negatively correlated with the sampling interval in recovered patients, which may imply a gradual restoration of the URT microbiota (Fig. 1F).

Association between the upper respiratory tract microbiota and the fatality

To better examine the association between microbiota and clinical features and outcomes, we defined two critical time points, T1 and T2, which referred to the first time point within five days after admission and the last time point within five days before discharge or death, respectively. T1 samples were available for 181 patients (65% samples were collected on the first day of admission, 88% samples were collected within three days after admission), T2 samples were available for 157 patients (80% samples were collected within three days before discharge or death), and 144 patients had both T1 and T2 samples. Among seven demographic and clinical features, the fatality was the only feature significantly associated with the microbiota at both T1 and T2 (Table 1).

All three differential analysis (GAMLSS regression, linear model regression, and LEfSe) identified *Streptococcus* as the most enriched microbe in the recovered group at T1(Fig. 2A, Fig. S2A,C), which was also true for T2 but was ranked behind two low-abundance bacteria in GAMLSS results (*Lautropia* and *Atopobium*, Fig. 2B, Fig. S2B,D). To further identify microbes associated with disease progression, Cox regression was conducted on fatality-associated genera selected by GAMLSS at T1. Finally, seven variables were selected as the best predictors, including age, viral copy number, severity-A, and the abundance of four genera (*Streptococcus*, *Porphyromonas*, *Atopobium*, *Serratia*) (Fig. S2E). Kaplan–Meier survival analysis confirmed that the abundance of these four bacteria was significantly associated with the fatality risk (Fig. 2C).

Moreover, we built a classifier based on L_1 regularized logistic regression model to distinguish the recovered and deceased patients using microbiota data at T1. A good performance was achieved with four variables (Fig. 2D). *Streptococcus* was the only microbe with a non-zero coefficient (regression coefficient =-1.74) besides three host factors, including age, viral copy number, and severity-A. The classifier trained merely on microbiota data also detected *Streptococcus* as the best predictor (regression coefficient =-1.51), further supporting the abundance of *Streptococcus* as a valid predictor for clinical outcomes.

Enterococcus and *Candida* were recognized as the most enriched genera in deceased patients by all three methods at T2, but not at T1 (Fig. 2B, Fig. S2B,D). Analysis at the species level confirmed *Candida albicans* and *Enterococcus faecium* as the two most enriched species in deceased patients besides SARS-CoV-2, suggesting that secondary infections of the two pathogenic microbes could potentially increase the case fatality rate (CFR) of COVID-19 patients.

Association between the abundance of *Streptococcus parasanguinis* and the fatality risk

We then took a closer look at the dynamic change of the *Streptococcus* in COVID-19 patients. The abundance of *Streptococcus* was higher in the recovered group at all time points except day 21 (P<0.01,

Fig. 3A) and significantly increased during the hospitalization (Fig. S3A). In contrast, the abundance of *Streptococcus* was marginally decreased over time in the deceased group (Fig. S3A). Also, CFR declined with the increased abundance of *Streptococcus*, and this trend was more pronounced at T2 (Fig. 3B).

Specifically, four *Streptococcus* species at T1 and seven *Streptococcus* species at T2 were enriched in recovered patients (Fig. 3C, Fig. S3B, Supplementary Text). In samples with the highest abundance of *Streptococcus* (CS1), *S. parasanguinis* was the only *Streptococcus* significantly enriched in the recovered patients at T1 (Fig. S3C). Moreover, Kaplan–Meier survival analysis also identified the abundance of *S. parasanguinis* as the primary factor associated with CFR (Fig. 3D). In summary, our results indicated that a higher level of *S. parasanguinis* at T1 predicted better prognosis in COVID-19 patients.

We then examined whether any confounding factors could explain the association between the fatality and the abundance of *S. parasanguinis* at T1. Neither of the risk factors for mortality observed in our cohort or other studies (age, gender, use of corticosteroid, the severity on admission, and SARS-CoV-2 viral copy number, use of high-grade antibiotics (listed in Table S1), comorbidities) was significantly associated with the abundance of *S. parasanguinis* ($p>0.05$). The fatality was the only factor associated with the abundance of *S. parasanguinis* in multivariable linear regression analysis. Moreover, multivariate logistic regression analysis confirmed that the abundance of *S. parasanguinis* was independently correlated with the fatality in our data (Table 2).

Intriguingly, we noted that the degree of correlation between the abundance of *S. parasanguinis* and the fatality was greater in patients with less severe symptoms on admission (with a severity score of 3 and 4) (Fig. 3E). Specifically, 95.2% (60/63) with a high abundance of *S. parasanguinis* (>10%) and 93% (93/100) patients with an intermediate abundance of *S. parasanguinis* (>1%) on admission in the less severe group got recovered, resulting in a CFR of 7%. On the contrary, patients with negligible *S. parasanguinis* (<0.1%) suffered a higher CFR of 30.8% (8/26) (odds ratio = 6, 95% Confidence Intervals: 2 to 18, p=0.0002). Meanwhile, patients with a severity score of 5 on admission showed the highest CFR (72.7%), which was not correlated with the abundance of *S. parasanguinis* (Fig. 3E), indicating that the predictive power of *S. parasanguinis* was stronger in the less severe patients (AUC=0.703, Fig. S3D).

The authenticity of *S. parasanguinis* was first validated by qPCR using specific primers targeting the groEL gene[24]. A strong positive correlation was found between the abundance inferred from the sequencing data and the copy number estimated from qPCR (Fig. 3F). Moreover, the copy number of *S. parasanguinis* also showed a good performance in predicting the survival of the patients (Fig. S3E). Secondly, reads from ten samples with the highest abundance of *S. parasanguinis* were mapped to two reference genomes of *S. parasanguinis* (NC_017905.1, NC_015678.1). All variants with minor allele frequency greater than 5% were identified, and the variants density was 0.036% and 0.04%, suggesting that the *Streptococcus* in those samples was close relative of known *S. parasanguinis* strains.

Co-detection of other pathogens in the upper respiratory tract of COVID-19 patients

We then screened all samples for 52 common respiratory pathogens (Table S2). Overall, 20.3% of patients had at least one such pathogen, and the proportion of samples with pathogens was higher in deceased patients than that in recovered patients (48.7% vs. 13.1%, $p<0.001$). Also, the number of pathogens identified in each sample was significantly higher in deceased patients than in recovered patients (Fig. 4A). Specifically, *Candida albicans* and *Enterococcus faecium* were more prevalent in deceased patients than in recovered patients (Fig. 4B). More importantly, these two species' abundances on admission were relatively low and similar between recovered and deceased patients, but remarkably increased during hospitalization in deceased patients (Fig. 4C,D), suggesting secondary infections.

E. faecium is a commensal bacteria in human GIT. Translocation of this bacteria had been reported in critically ill patients[25]. We compared the abundance of *E. faecium* between anal swabs, plasma, and OPs in 32 paired samples, including eight samples with the highest abundance of *E. faecium* in OPs, but found no excess *E. faecium* RNA in the plasma of these co-detected patients (p=0.17, Fig. 4E), suggesting that the translocation through the circulatory system was unlikely.

Virus co-infection was rare in this cohort. Only *Rhinovirus A* was detected in one recovered patient with more than 50 reads. Other possible viral co-infections included *Influenza A virus* and *Human mastadenovirus C*, which were revealed by less than 50 reads.

Dynamics of the upper respiratory tract microbiota in COVID-19 patients and its association with the fatality

Although no significant difference in alpha diversity was observed between the recovered and deceased patients (Fig. 5A), an increasing trend of alpha diversity over time was observed in some deceased patients (Fig. S4A). We found this phenomenon was associated with the codominance of both pathogenic microbes and commensal microbes in their microbiotas (Fig. S4B).

The dynamic change of microbiota was then visualized as transitions between different CSs (Fig. 5B). Multivariate logistic regression identified fatality as the only factor associated with CS transition ($p<0.01$). Therefore, the analysis was conducted for recovered and deceased patients, separately (Fig. 5B). Most CSs (73.5%) at T1 transited to CS1 at T2 in recovered patients, while only 33.3% CSs transited to CS1 in deceased patients ($p<0.05$). Notably, CS1 had the lowest CFR among all CSs (12.4% at T1, 4.7% at T2, $p<0.001$) and tended to be more resistant to co-/secondary infections compared to other CSs (6.6% vs. 19.0%, $p<0.001$). Meanwhile, 42.1% CSs transited to the three most fatal CSs (CS7, CS8, and CS11) in deceased patients, whereas only 1.6% CSs transited to the fatal CSs in recovered patients ($p<0.001$, Fig. 5B). Concordantly, expansions of *Streptococcus* (dominant genus for CS1) were more frequently observed in recovered patients than that in deceased patients (47.6% vs.15.0%, $P<0.05$, Fig. 5C). Meanwhile, the expansion of *Enterococcus* and *Candida* (dominant genera for CS7 and CS8) was more frequently observed in deceased patients ($P<0.05$, Fig. 5C). Thus, not only the microbiota compostion, but also the dynamics of the microbiota differed between recovered and deceased patients.

The transition dynamics was further modeled as a Markov chain. The self-transition proportion of CS1 was significantly higher than other CSs (p<0.001, Fig. 5D). Moreover, we found that patients with more samples assigned to CS1 tended to have a low CRF (Fig. S4C), suggesting that CS1 represented a health status. In addition, the transition from deleterious CSs (CS7, CS8, and CS11) to CS1 was less frequent compared to other CSs (Fig. 5D), suggesting severe dysbiotic microbiota cannot recover in a short time. Besides CS1, all other CSs had a low self-transition rate (< 0.29), suggesting that they were unstable and probably represented an intermediate state.

The magnitude of microbiota change in the deceased group was higher than that in the recovered group and tended to be increased over time (r=0.31, p=0.07, Fig. 5E). A similar trend was observed when using the Bray-Curtis distance to HC (Fig. 5F). Strikingly, the distance to HC increased in the deceased patients over time but tended to decline in the recovered group (Fig. 5F). Specifically, *Streptococcus* and *Schaalia* were the two genera whose abundance was significantly approaching the level in HC during hospitalization (Fig. S4D,E). Of note, the distance from the recovered group to HC at T2 was still significantly greater than that within HC, indicating that the URT microbiota was not restored at discharge.

Microbial network in COVID-19 patients

To gain insights into the microbial interactions in COVID-19 patients, we constructed the interaction network based on microbes' co-occurrence. The network analysis revealed that the correlation between genera in deceased patients was sparser and more fragmented than that in recovered patients and HC at both T1 and T2 (Fig. S5), suggesting decreased biotic interactions in deceased patients, which was associated with a more fragile microbiota[26]. The components of a "core" healthy microbiota, including *Prevotella*, *Veillonella*, *Campylobacter*, *Leptotrichia*, *Fusobacterium* and *Selenomonas*[27], were more closely linked in the network of HC and recovered patients than that of deceased patients, suggesting that the aggregation status of the network may reflect the homeostasis and resistance of the microbiota.

Discussion

In this study, we have demonstrated the features and dynamics of the URT microbiota and its association with the fatality in COVID-19 patients using the LOTUS cohort. The metatranscriptomic analysis was conducted, which could provide a real-time snapshot of the microbiota change, as RNA degrades faster than DNA[28]. We noted that the URT microbiota of COVID-19 was more dispersive and partially distinguishable from the HC (Fig. 1C), which is concordant with results in two recent studies[21,22]. However, a significant overlap was observed between the microbiota of two groups in our data, indicating that the influence of COVID-19 on URT microbiota varied among different individuals, which may be associated with the host immune status and the background of the microbiota. Meanwhile, the URT microbiota of deceased patients showed more distinct composition compared to that of the HC and recovered patients, and such tendency is more significant at T2 than that at T1, suggesting that the

extent of the microbiota alteration was correlated with the severity and progression of the disease, while the causality was elusive.

An incomplete restoration of URT microbiota was observed in recovered COVID-19 patients. The microbiota was gradually restoring during hospitalization in recovered patients but was still different from that in HC at discharge (Fig. 5F). Delayed recovery of the GIT and URT microbiota were observed after antibiotic perturbation and virus infection in multiple studies[19,29], suggesting that microbiota recovery lags behind the disease's recovery. A Follow-up study is ongoing to investigate the recovery of URT microbiota and its clinical associations after discharge.

Potential pathogens other than SARS-CoV-2 were co-detected in 20.3% samples in our study, while the previous estimate varied between 6.1% to 94.2%[16,30–32]. *Candida albicans* and *Enterococcus faecium* were the two most common co-detected pathogens, being observed in 11.5% and 6.25% of COVID-19 patients. The increasing abundance of these two pathogens during hospitalization, which may represent nosocomial infections, was significantly correlated with a higher fatality rate (Fig. 4C,D). Meanwhile, several proinflammatory pathways enriched in deceased patients were contributed by these pathogens (Supplementary Text). Co-detection with *C. albicans* has been frequently observed in other studies; thus, special attention should be paid to secondary fungal infection during the management of COVID-19 patients[16,30–32].

Streptococcus in COVID-19 patients was dominated by non-pathogenic species. The most abundant *Streptococcus* in COVID-19 patients were *S. parasanguinis*, *S. sp.oral taxon 431*, *S. pneumoniae_mitis*, and *S. salivarius*, which were also the ones with the highest abundance in HC. The abundance of all these *Streptococcus* species substantially decreased in the COVID-19 patients. Among them, *S. salivarius* showed the most significant decline (Fig. 3C, Fig. S3C), which may relate to its high sensitivity to antibiotics[33]. Notably, *S. salivarius*, specially Stain K12 and M18, are probiotics capable of fostering a balanced healthy oral microbiota[34,35]. Thus, depletion of this species may reduce the resistance and resilience of the microbiota and increase the susceptibility to co-/secondary infections.

For the first time, we found that a higher abundance of *Streptococcus parasanguinis* on admission was strongly associated with better outcomes in COVID-19 patients. We suspected that several mechanisms might be involved. First, *S. parasanguinis* may contribute to a stable niche that resistant to pathogens. The bacteria play a vital role in initiating dental plaque formation, which is critical for maintaining a healthy microbiota in the oral cavity[36]. Moreover, *S. parasanguinis* could directly inhibit the growth of *Pseudomonas aeruginosa* by producing H_2O_2, and induce a varying level of cytokines[37,38], which is essential for maintaining the immune homeostasis. Second, the decreasing of the abundance of *S. parasanguinis* may indicate a disrupted ecosystem in the respiratory tract. As an essential component of a healthy oral/respiratory tract microbiota[39,40], the decreasing or depletion of *S. parasanguinis* may result from the introduction and overgrowth of competing microbes, altered regional conditions, enhanced immune pressure, or use of antibiotics, which have been observed in pneumonia and other pulmonary

diseases[41,42]. We speculate that the disrupted ecosystem, rather than *S. parasanguinis* itself, is associated with a critical condition and poor prognosis. However, it was mysterious why only *S. parasanguinis* but not other commensal microbes were associated with the integrity of the URT microbiota. Third, the abundance of *S. parasanguinis* may be associated with confounding factors that correlate with clinical outcomes. However, we did not find any risk or protective factors associated with the abundance of *S. parasanguinis*. Severely ill patients were typically administered with broad-spectrum or high-grade antibiotics, which may reduce the abundance of *S. parasanguinis*. However, although we found a strong correlation between the administration of the high-grade antibiotics (Table S1) and the fatality (59.0% in deceased patients vs. 6.1% in recovered patients, $p<0.001$), no correlation existed between the administration of high-grade antibiotics and the abundance change of *S. parasanguinis* ($p=0.09$).

There are some limitations in our study. First, the cohort only includes severe COVID-19 patients, while the mild and asymptomatic patients may display distinct features. Nevertheless, as most deaths occurred in severely ill patients, this cohort is appropriate for studying the association between the URT microbiota and fatality. Second, our study is a single-center study. It is unclear whether the findings could be generalized to other populations, particularly considering that URT microbiota could potentially be influenced by geography, ethnicity, host genetics, and medication[40,43]. Thus, more data, especially those from well-controlled clinical trials, is needed to validate findings in this study.

In summary, our study has characterized the features and dynamics of the URT microbiota and its association with clinical features, especially the fatality in COVID-19 patients. The URT microbiota in deceased patients was significantly distinct from that in recovered patients on admission and afterwards. A high abundance of *Streptococcus* was identified as an indicator of a stable and resilient URT microbiota associated with fewer co-infection and a lower fatality rate. In addition, we recommend *S. parasanguinis* as a potential prognostic biomarker related to the fatality in COVID-19. Further studies are needed to unveil the underlying mechanisms responsible for the associations observed in this study. Notably, although several papers have been published to recommend the use of probiotics in COVID-19 patients, which to some extent is supported by our data[44,45], the efficacy of the intervention needs to be validated in rigorous clinical trials.

Methods

Study design and sample collection

Oropharyngeal swab samples were collected on day 1, 5, 10, 14, 21, and 28 after admission from the patient enrolled in Lopinavir Trial for Suppression of SARS-CoV-2 in China (LOTUS) [ChiCTR2000029308]. All patients were positive for SARS-CoV-2 RNA tested by RT-PCR assay (Shanghai ZJ Bio Tec or Sansure Biotech, China) before admission. Demographic and clinical information was shown in the original paper[23]. All patients were assigned a severity score on admission and at discharge or on day 28, ranging from 2 (not hospitalized, not requiring supplemental oxygen) to 7 (death)[23].

The study was approved by the Institutional Review Board of Jin Yin-Tan Hospital (KY2020-02.01). Written informed consent was obtained from all patients or their legal representatives if they were too unwell to provide consent.

Library preparation and sequencing

Total nucleic acids were extracted from each 400 µl oropharyngeal swab sample using the NucliSens easyMAG apparatus (bioMerieux, Marcy l'Etoile, France) following the manufacturer's instructions, and finally eluted in 80 µl elution buffer and stored at -80℃ until use.

The RNA was converted into Illumina-compatible sequencing libraries with MINERVA protocol[46]. Briefly, five µl of nucleic acid was digested with DNase I (New England Biolabs, US), and then the first-strand cDNA was synthesized with Oligo dT primer (5'-T30VN-3'), random decamer primer (5'-N10-3'), and SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA). Direct tagmentation of RNA/DNA hybrids was using Tn5 transposase (Vazyme, China), followed by index PCR to amplify the library. Finally, the product was purified with AMPure XP beads (Beckman Coulter, MA, US). All purified libraries were sequenced on Illumina NovaSeq with 2x150 paired-end mode.

The viral genomes reported in this study have been deposited in the Genome Warehouse in the National Genomics Data Center (under project PRJCA003617, publicly accessible at https://bigd.big.ac.cn/gsa/s/YGLd8t5n, a permanent link will be provided upon acceptance of this manuscript for publication).

Taxonomic profiling and data preprocessing

Reads were first trimmed for quality, adapter, primers, polyX in 3'ends using Fastp (version 0.20.0) (-l 50 -x –detect_adapter_for_pe –overlap_len_require 20 –overlap_diff_limit 5 –overlap_diff_percent_limit 20 –cut_tail –cut_tail_mean_quality 15)[47]. Potiental contaminant reads was discarded by aligning all reads to a customize database which include the human genome (GRCH38) and UNiVec sequences by bmtagger (version 1.0)[48]. The rRNA sequences were removed by SortMeRNA (version 2b)[49]. The clean reads were aligned to NCBI NT database (version 2020.02.25) using Megablast (version 2.9.0)(-evalue 1e-10 -qcov_hsp_perc 60 -perc_identity 60)[50]. Taxonomic assignment was performed by MEGAN (-ms 100 –supp 0 –me 0.01 –mrc 60)(version 6.12.2) [51]. Samples with at least 1000 microbial reads and taxonomies with the relative abundance >0.01 at least in one sample were included in downstream analysis.

An extra step was applied to refine the species results given by MEGAN. Reads from all samples were merged and mapped to the reference genome of all species supported by at least 10,000 reads by Bowtie2 (–sensitive)(version 2.2.3)[52]. Species with a genome coverage of less than 1% were regarded as false positives. Reads assigned at the genera level were reallocated to species level according to the relative proportion of each species in that genus.

Clustering analysis

Principal coordination analysis based on Bray-Curtis distance was applied to the relative abundance matrix at the genus level. The first 56 coordinates, which explained 95% of the total variances, were used as the input for clustering. Then we applied two rounds of hierarchical agglomerative clustering with a complete-linkage strategy. The first clustering removed outliers that might interfere with the estimation of the number of clusters. We determined the number of possible clusters using average silhouette width and performed post examination to remove the outliers within each putative cluster. Specifically, the sample whose average Bray-Curtis distance to other samples in the same cluster greater than 0.7 was reclassified as an outlier. After removing outliers, the optimal number of clusters was re-estimated by 1000 rounds of bootstrap average silhouette width. Since the average silhouette width reached the maximum when k=10 and dropped dramatically at k=11, we determined 10 to be the optimum number of clusters (Fig. S1H).

Association between the microbe and the metadata

Three statistical methods were applied to identify the microbe that was associated with specific metadata, which included Generalized additive models for location, scale, and shape with zero-inflated beta distribution(GAMLSS)[53], Linear model regression implemented in metamicrobiomeR[54], Linear discriminant analysis effect size (LEfSe)[55]. All metadata were input as cofactors in the former two methods.

Network analysis

Network analysis was performed based on reads count data at the genera level using SpeciEasi in deceased, recovered patients at T1, T2, and healthy controls[56]. The density of the networks was calculated by graph.density function in igraph R package[57]. The network was visualized by igraph R package.

Statistical analysis

Alpha-diversity and beta diversity were computed using phyloseq and vegan package[58,59]. Correlation between categorical metadata was calculated by Cramer's V test. Permutational multivariate analysis of variance (PERMANOVA) was used to determine the host factors associated with the microbial community and evaluate the similarity between two microbiotas. For the permutation test, the p-value was calculated based on 2000 permutations. Fisher exact tests and Mann-Whitney U were used to compare categorical variables and continuous variables, respectively. Multiple comparisons were corrected by the FDR algorithm. All statistical analysis was implemented in RStudio[60].

Declarations

Declaration of interests : All authors declare no competing interests.

Acknowledgments: We thank the patients and healthy volunteers who participated in this study. This study was funded in part by the National Major Science & Technology Project for Control and Prevention of Major Infectious Diseases in China (2017ZX10103004, 2018ZX10301401), the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (2019-I2M-2-XX, 2016-I2M-1-014, 2018-I2M-1-003), The Non-profit Central Research Institute Fund of CAMS (2020HY320001, 2019PT310029), Beijing Advanced Innovation Center for Genomics (ICG), and Beijing Advanced Innovation Center for Structural Biology (ICSB).

Role of the funding source

The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Author contributions

The author contributions were as follows. Sample collection and management: Y.M.W., D.Y.Z, G.H.F, L.G., W.L., X.L., H.B.L., and X.H.Z.; Lab work: L.L.R., Y.X., J.L., L.G., Q.L., J.Z.L, L.D., C.H.W., Y.W., L.C., and X.M.W.; Data analysis: Z.J.X., Y.J., S.Z.J., K.L., S.L.S., and L.M.K.; Project administration: L.L.R., W.J.W, C.B., L.M.K., J.B.W., and Y.Y.H.; Writing and editing: L.M.K., L.L.R., W.J.W, C.B, and L.Z.; Conceptualization and supervision: L.M.K., W.J.W, C.B. and L.L.R.
These authors contributed equally: Lili Ren, Jiaxin Zhong, Yeming Wang, Dingyu Zhang, Yan Xiao, Jie Li, Jing Yang, Guohui Fan. These authors jointly supervised this work: Mingkun Li, Jianwei Wang, Bin Cao, Yanyi Huang, and Jianbin Wang.

References

1. Pollán, M. *et al.* Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* (2020) doi:10.1016/S0140-6736(20)31483-5.
2. World Health Organization. Clinical management of severe acute respiratory infection when novel coronavirus (nCoV) infection is suspected. *Who* (2020).
3. People at Increased Risk. https://www.cdc.gov/coronavirus/2019-nCoV/index.html.
4. Ellinghaus, D. *et al.* Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMoa2020283.
5. van der Made, C. I. *et al.* Presence of Genetic Variants Among Young Men With Severe COVID-19. *JAMA* (2020) doi:10.1001/jama.2020.13719.
6. Atyeo, C. *et al.* Distinct early serological signatures track with SARS-CoV-2 survival. *Immunity* (2020) doi:10.1016/j.immuni.2020.07.020.
7. Shen, B. *et al.* Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* (2020) doi:10.1016/j.cell.2020.05.032.

8. Kermali, M., Khalsa, R. K., Pillai, K., Ismail, Z. & Harky, A. The role of biomarkers in diagnosis of COVID-19 – A systematic review. *Life Sciences* (2020) doi:10.1016/j.lfs.2020.117788.
9. Young, V. B. The role of the microbiome in human health and disease: An introduction for clinicians. *BMJ (Online)* (2017) doi:10.1136/bmj.j831.
10. de Steenhuijsen Piters, W. A. A., Binkowska, J. & Bogaert, D. Early Life Microbiota and Respiratory Tract Infections. *Cell Host and Microbe* (2020) doi:10.1016/j.chom.2020.07.004.
11. Libertucci, J. & Young, V. B. The role of the microbiota in infectious diseases. *Nature Microbiology* (2019) doi:10.1038/s41564-018-0278-4.
12. Gu, S. *et al.* Alterations of the Gut Microbiota in Patients With Coronavirus Disease 2019 or H1N1 Influenza. *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciaa709.
13. Zuo, T. *et al.* Alterations in Gut Microbiota of Patients With COVID-19 During Time of Hospitalization. *Gastroenterology* (2020) doi:10.1053/j.gastro.2020.05.048.
14. Zuo, T. *et al.* Alterations in Fecal Fungal Microbiome of Patients With COVID-19 During Time of Hospitalization until Discharge. *Gastroenterology* (2020) doi:10.1053/j.gastro.2020.06.048.
15. Zuo, T. *et al.* Depicting SARS-CoV-2 faecal viral activity in association with gut microbiota composition in patients with COVID-19. *Gut* (2020) doi:10.1136/gutjnl-2020-322294.
16. Zhang, H. *et al.* Metatranscriptomic Characterization of COVID-19 Identified A Host Transcriptional Classifier Associated With Immune Signaling. *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciaa663.
17. Geva-Zatorsky, N. *et al.* Mining the Human Gut Microbiota for Immunomodulatory Organisms. *Cell* (2017) doi:10.1016/j.cell.2017.01.022.
18. Martino, C. *et al.* Bacterial modification of the host glycosaminoglycan heparan sulfate modulates SARS-CoV-2 infectivity. *bioRxiv* (2020).
19. Kaul, D. *et al.* Microbiome disturbance and resilience dynamics of the upper respiratory tract during influenza A virus infection. *Nat. Commun.* (2020) doi:10.1038/s41467-020-16429-9.
20. Shen, Z. *et al.* Genomic diversity of SARS-CoV-2 in Coronavirus Disease 2019 patients. *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciaa203.
21. Xu, R. *et al.* Progressive worsening of the respiratory and gut microbiome in children during the first two months of COVID-19. *medRxiv* (2020) doi:10.1101/2020.07.13.20152181.
22. Xu, R. *et al. Temporal dynamics of human respiratory and gut microbiomes during the course of COVID*. *medRxiv* (2020) doi:10.1101/2020.07.21.20158758.
23. Cao, B. *et al.* A trial of lopinavir-ritonavir in adults hospitalized with severe covid-19. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMoa2001282.
24. Chen, Q. *et al.* Quantification of Human Oral and Fecal Streptococcus parasanguinis by Use of Quantitative Real-Time PCR Targeting the groEL Gene. *Front. Microbiol.* (2019) doi:10.3389/fmicb.2019.02910.
25. Dickson, R. P. *et al.* Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome. *Nat. Microbiol.* (2016) doi:10.1038/nmicrobiol.2016.113.

26. Wagner Mackenzie, B. *et al.* Bacterial community collapse: a meta-analysis of the sinonasal microbiota in chronic rhinosinusitis. *Environ. Microbiol.* (2017) doi:10.1111/1462-2920.13632.
27. Einarsson, G. G. *et al.* Community analysis and co-occurrence patterns in airway microbial communities during health and disease. *ERJ Open Res.* (2019) doi:10.1183/23120541.00128-2017.
28. Ren, L. *et al.* Transcriptionally Active Lung Microbiome and Its Association with Bacterial Biomass and Host Inflammatory Status. *mSystems* (2018) doi:10.1128/msystems.00199-18.
29. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci. U. S. A.* (2011) doi:10.1073/pnas.1000087107.
30. Zhu, X. *et al.* Co-infection with respiratory pathogens among COVID-2019 cases. *Virus Res.* (2020) doi:10.1016/j.virusres.2020.198005.
31. Kim, D., Quinn, J., Pinsky, B., Shah, N. H. & Brown, I. Rates of Co-infection between SARS-CoV-2 and Other Respiratory Pathogens. *JAMA - Journal of the American Medical Association* (2020) doi:10.1001/jama.2020.6266.
32. Hughes, S., Troise, O., Donaldson, H., Mughal, N. & Moore, L. S. P. Bacterial and fungal coinfection among hospitalized patients with COVID-19: a retrospective cohort study in a UK secondary-care setting. *Clin. Microbiol. Infect.* (2020) doi:10.1016/j.cmi.2020.06.025.
33. Tuohy, M. & Washington, J. A. Antimicrobial susceptibility of viridans group streptococci. *Diagn. Microbiol. Infect. Dis.* (1997) doi:10.1016/S0732-8893(97)00140-5.
34. Wescombe, P. A., Hale, J. D., Heng, N. C. & Tagg, J. R. Developing oral probiotics from Streptococcus salivarius. *Future Microbiology* (2012) doi:10.2217/fmb.12.113.
35. Hols, P., Ledesma-García, L., Gabant, P. & Mignolet, J. Mobilization of Microbiota Commensals and Their Bacteriocins for Therapeutics. *Trends in Microbiology* (2019) doi:10.1016/j.tim.2019.03.007.
36. Herrero, E. R. *et al.* Antimicrobial effects of commensal oral species are regulated by environmental factors. *J. Dent.* (2016) doi:10.1016/j.jdent.2016.02.007.
37. Van Den Bogert., B., Meijerink, M., Zoetendal, E. G., Wells, J. M. & Kleerebezem, M. Immunomodulatory properties of streptococcus and veillonella isolates from the human small intestine microbiota. *PLoS One* (2014) doi:10.1371/journal.pone.0114277.
38. Scott, J. E. & O'Toole, G. A. The yin and yang of Streptococcus lung infections in cystic fibrosis: A model for studying polymicrobial interactions. *J. Bacteriol.* (2019) doi:10.1128/JB.00115-19.
39. Liang, X., Chen, Y. Y. M., Ruiz, T. & Wu, H. New cell surface protein involved in biofilm formation by Streptococcus parasanguinis. *Infect. Immun.* (2011) doi:10.1128/IAI.00029-11.
40. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* (2012) doi:10.1038/nature11234.
41. Dickson, R. P. & Huffnagle, G. B. The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease. *PLoS Pathog.* (2015) doi:10.1371/journal.ppat.1004923.

42. Hanada, S., Pirzadeh, M., Carver, K. Y. & Deng, J. C. Respiratory viral infection-induced microbiome alterations and secondary bacterial pneumonia. *Frontiers in Immunology* (2018) doi:10.3389/fimmu.2018.02640.
43. Gupta, V. K., Paul, S. & Dutta, C. Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Frontiers in Microbiology* (2017) doi:10.3389/fmicb.2017.01162.
44. Janda, L., Mihalčin, M. & Šťastná, M. Is a healthy microbiome responsible for lower mortality in COVID-19? *Biologia (Bratisl).* (2020) doi:10.2478/s11756-020-00614-8.
45. Walton, G. E., Gibson, G. R. & Hunter, K. A. Mechanisms linking the human gut microbiome to prophylactic and treatment strategies for COVID-19. *Br. J. Nutr.* (2020) doi:10.1017/S0007114520003980.
46. Chen, C. *et al.* MINERVA: A facile strategy for SARS-CoV-2 whole genome deep sequencing of clinical samples. *bioRxiv* (2020) doi:10.1101/2020.04.25.060947.
47. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. in *Bioinformatics* (2018). doi:10.1093/bioinformatics/bty560.
48. NCBI. BMTagger: Best Match Tagger for Removing Human Reads from Metagenomics Datasets [Online]. Available online at: ftp://ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger/. (2011).
49. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* (2012) doi:10.1093/bioinformatics/bts611.
50. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* (2009) doi:10.1186/1471-2105-10-421.
51. Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* (2016) doi:10.1371/journal.pcbi.1004957.
52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* (2012) doi:10.1038/nmeth.1923.
53. Rigby, R. A., Stasinopoulos, D. M. & Lane, P. W. Generalized additive models for location, scale and shape. *J. R. Stat. Soc. Ser. C Appl. Stat.* (2005) doi:10.1111/j.1467-9876.2005.00510.x.
54. Ho, N. T., Li, F., Wang, S. & Kuhn, L. MetamicrobiomeR: An R package for analysis of microbiome relative abundance data using zero-inflated beta GAMLSS and meta-analysis across studies using random effects models. *BMC Bioinformatics* (2019) doi:10.1186/s12859-019-2744-2.
55. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* (2011) doi:10.1186/gb-2011-12-6-r60.
56. Kurtz, Z. D. *et al.* Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLoS Comput. Biol.* (2015) doi:10.1371/journal.pcbi.1004226.
57. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Syst.* (2006).

58. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* (2013) doi:10.1371/journal.pone.0061217.
59. Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, H. W. vegan: Community Ecology Package. R package version 2.5-6. https://cran.r-project.org/package=vegan (2019).
60. RStudio Team. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. *2020* http://www.rstudio.com/.

Tables

Table 1. Association between the metadata and the upper respiratory tract microbiota composition in COVID-19 patients.

	T1		T2	
	R^2	p value	R^2	p value
Fatality	0.031	<0.001 ***	0.045	<0.001 ***
Viral Copy Number	0.009	0.135	0.007	0.259
Age	0.008	0.142	0.009	0.139
Corticosteroid	0.007	0.243	0.009	0.139
Gender	0.006	0.355	0.006	0.393
Severity-A	0.004	0.555	0.007	0.246
Treatment	0.004	0.650	0.002	0.916

R^2, which represents the proportion of variance explained by the factor, and the p-value were calculated by PERMONAVA analysis.

Table 2. Multivariate logistic regression analysis of factors associated with fatality

	Univariable OR (95% CI)	p.value	Multivariable OR (95% CI)	p.value
S. parasanguinis relative abundance	0.30(0.13-0.66)	0.004**	0.17(0.04-0.6)	0.009**
Gender	1.70(0.79-3.84)	0.185	1.58(0.45-5.97)	0.484
Age	1.08(1.04-1.12)	<0.001***	1.12(1.06-1.2)	<0.001***
Treatment	0.50(0.23-1.05)	0.071	0.7(0.22-2.24)	0.542
Severity_A(4 vs 3)	0.56(0.20-1.87)	0.312	0.71(0.16-3.66)	0.658
Severity_A(5 vs 3)	11.73(3.26-49.99)	<0.001***	7.25(1.17-55.99)	0.042*
Viral copy number	1.45(1.25-1.74)	<0.001***	1.1(0.86-1.41)	0.451
High-grade antibiotics Usage	21.16(8.50-57.07)	<0.001***	9.46(2.41-42.78)	0.002**
Corticosteroid Usage	5.78(2.68-12.92)	<0.001***	3.37(1.04-11.62)	0.045*
Comorbidity	3.09(01.37-7.68)	0.009**	0.85(0.22-3.29)	0.817

For numerical variables, two levels were classified by comparing to the median value.

Figures

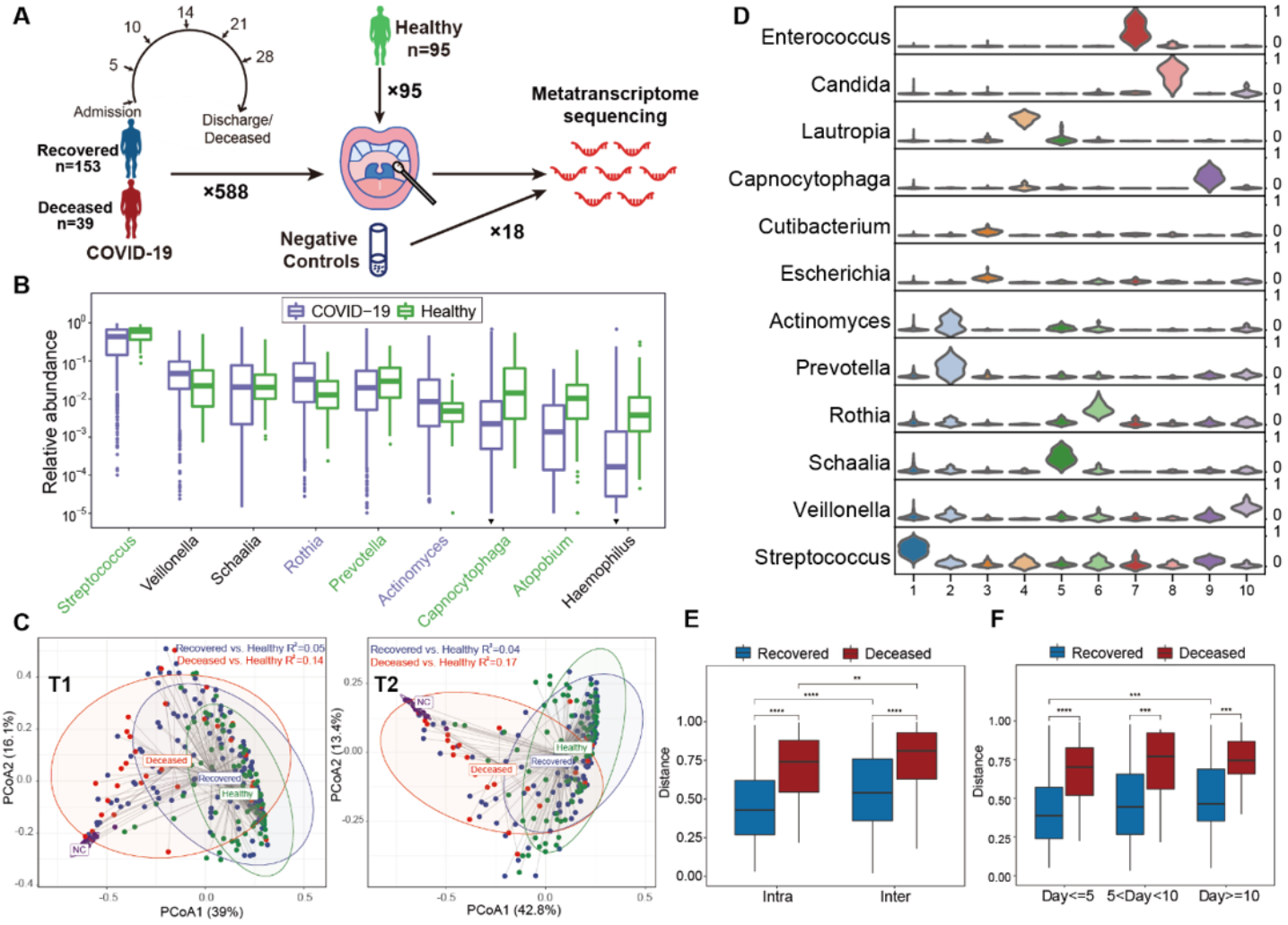

Figure 1

Project design and the overview of the upper respiratory tract microbiota. A. The design of the study. B. Relative abundance of the ten most abundant genera in COVID-19 and HC. The genera whose abundances in both groups were not significantly higher than NC were discarded. The genera were ordered by their abundances in COVID-19. Genera significantly enriched in COVID-19 patients and HC were labeled in purple and green, respectively. The inverted triangle indicates the genus whose abundance was not significantly higher than that in NC. C. Principal coordinate analysis (PCoA) plot of samples at the genus level on admission (left, T1) and before discharge/death (right, T2), R2 calculated by PERMONAVA analysis is shown in the figure and all p-values were less than 0.001. D. The abundances of compositional signatures for 10 clusters. Genera with average relative abundances greater than 10% in at least one cluster are shown. E. Bray-Cutis distances between samples from the same individual and between samples from different individuals. F. Bray-Cutis distances between samples with different sampling intervals from the same individual. The boxes represent 25th-75th percentiles, and the black lines indicate the median. * p < 0.05, ** p<0.01, *** p<0.001, and ****p<=0.0001.

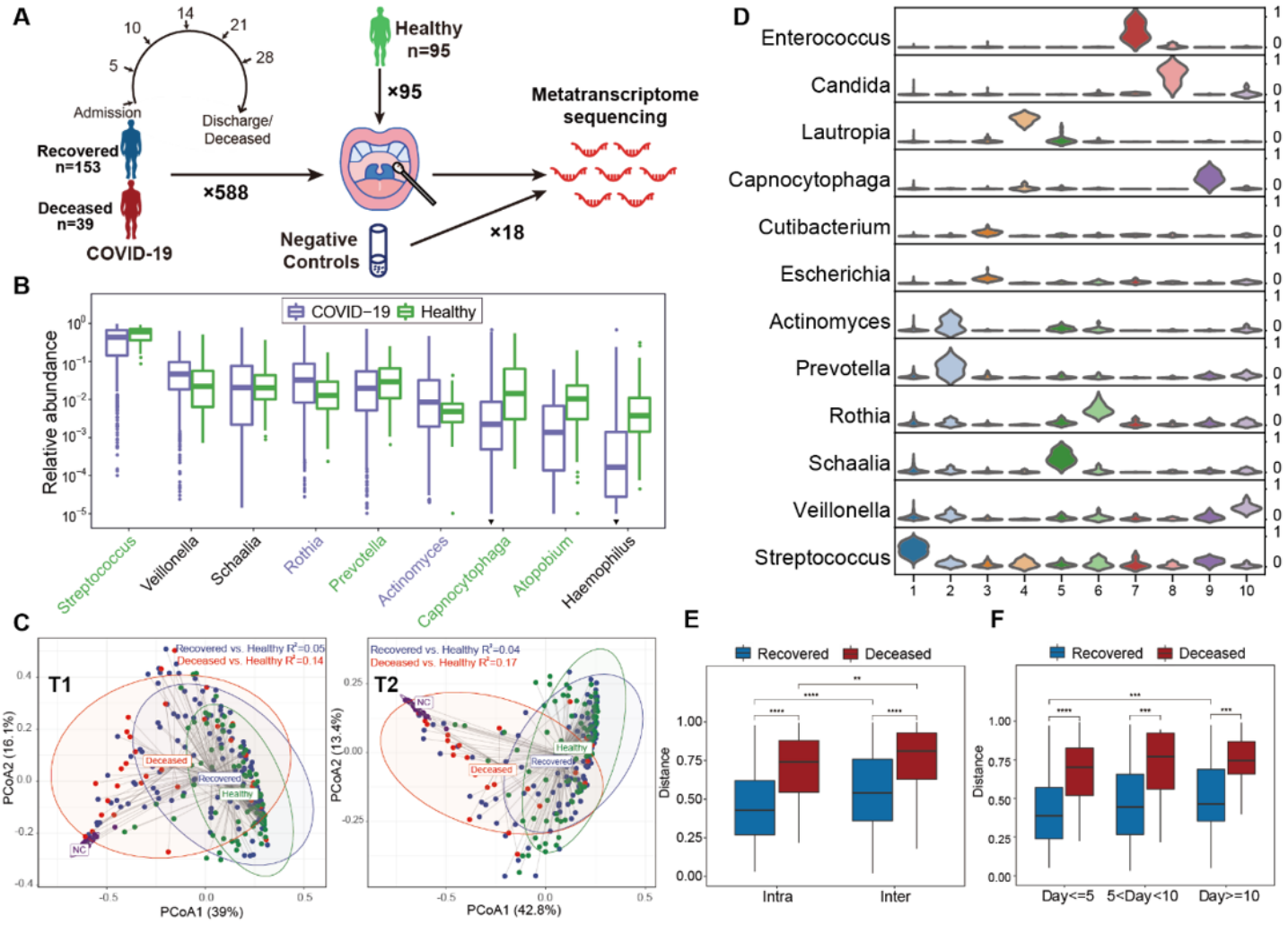

Figure 1

Project design and the overview of the upper respiratory tract microbiota. A. The design of the study. B. Relative abundance of the ten most abundant genera in COVID-19 and HC. The genera whose abundances in both groups were not significantly higher than NC were discarded. The genera were ordered by their abundances in COVID-19. Genera significantly enriched in COVID-19 patients and HC were labeled in purple and green, respectively. The inverted triangle indicates the genus whose abundance was not significantly higher than that in NC. C. Principal coordinate analysis (PCoA) plot of samples at the genus level on admission (left, T1) and before discharge/death (right, T2), R2 calculated by PERMONAVA analysis is shown in the figure and all p-values were less than 0.001. D. The abundances of compositional signatures for 10 clusters. Genera with average relative abundances greater than 10% in at least one cluster are shown. E. Bray-Cutis distances between samples from the same individual and between samples from different individuals. F. Bray-Cutis distances between samples with different sampling intervals from the same individual. The boxes represent 25th-75th percentiles, and the black lines indicate the median. * p < 0.05, ** p<0.01, *** p<0.001, and ****p<=0.0001.

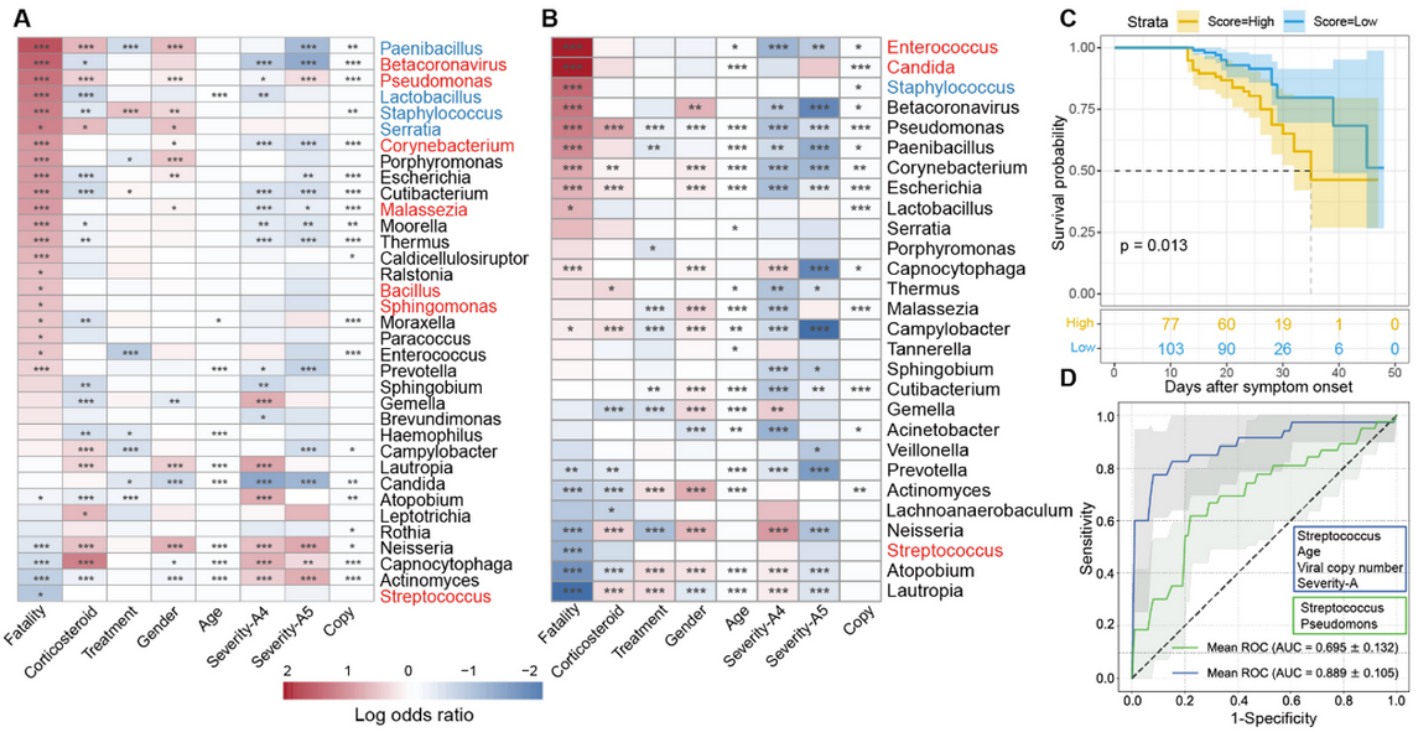

Figure 2

The associations between the upper respiratory tract microbiota and the fatality. Genera that associated with different metadata identified by GAMLSS analysis are shown in A (at T1) and B (at T2) ($p<0.05$). Genera were ordered by the correlation coefficients with the fatality. Fatality-associated genera identified by the other two methods (multivariate linear regression (MLR) and LEfSe) were marked in red, fatality-associated genera identified by GAMLSS and MLR were marked in blue. The color in the heatmap represents regression coefficients (log odds ratio) estimated by GAMLSS. C. Kaplan–Meier survival curves for the two groups classified by four genera selected by Cox regression, which included Streptococcus, Porphyromonas, Atopobium, and Serratia. Log-rank p-value is labeled in the figure. D. ROC (Receiver Operator Characteristic) curves for the fatality classifier based on both host factors and the microbiota composition (blue line), and merely on microbiota composition (green line).

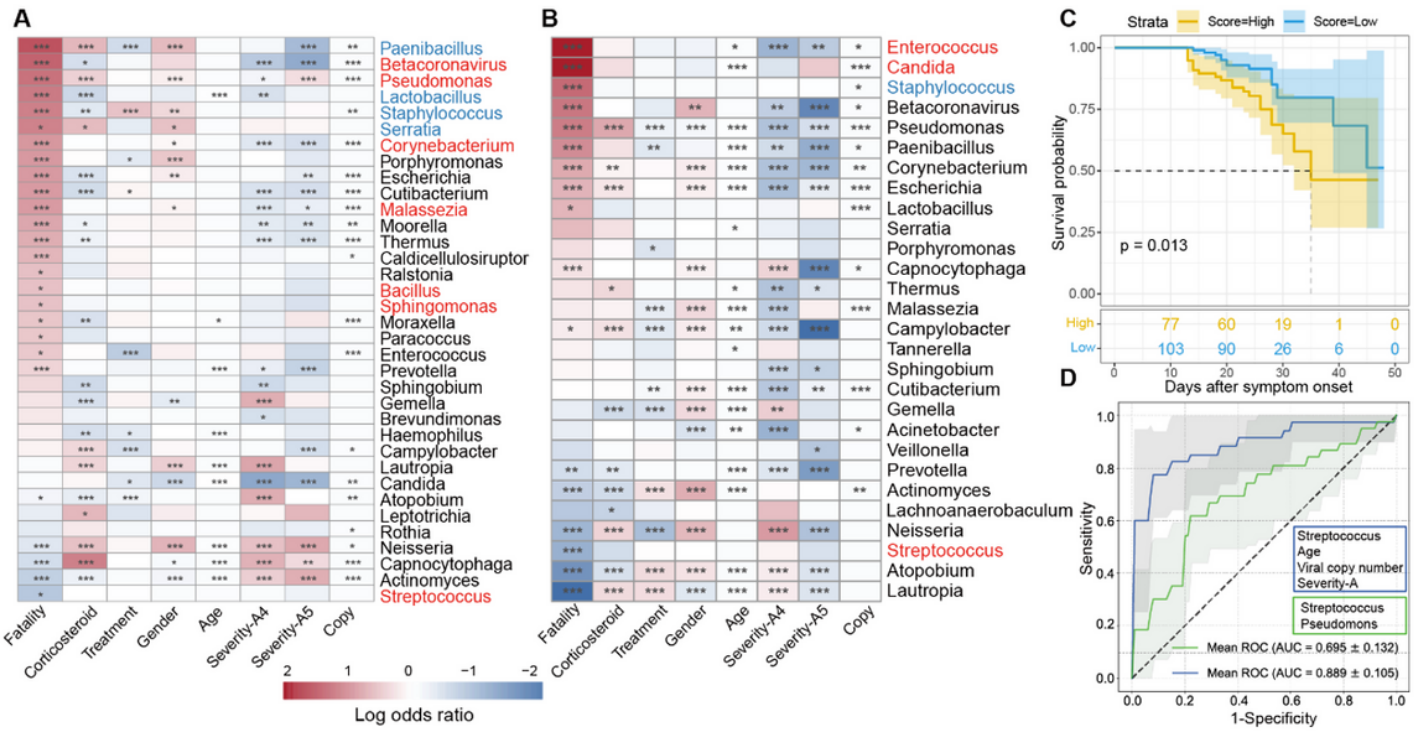

Figure 2

The associations between the upper respiratory tract microbiota and the fatality. Genera that associated with different metadata identified by GAMLSS analysis are shown in A (at T1) and B (at T2) ($p<0.05$). Genera were ordered by the correlation coefficients with the fatality. Fatality-associated genera identified by the other two methods (multivariate linear regression (MLR) and LEfSe) were marked in red, fatality-associated genera identified by GAMLSS and MLR were marked in blue. The color in the heatmap represents regression coefficients (log odds ratio) estimated by GAMLSS. C. Kaplan–Meier survival curves for the two groups classified by four genera selected by Cox regression, which included Streptococcus, Porphyromonas, Atopobium, and Serratia. Log-rank p-value is labeled in the figure. D. ROC (Receiver Operator Characteristic) curves for the fatality classifier based on both host factors and the microbiota composition (blue line), and merely on microbiota composition (green line).

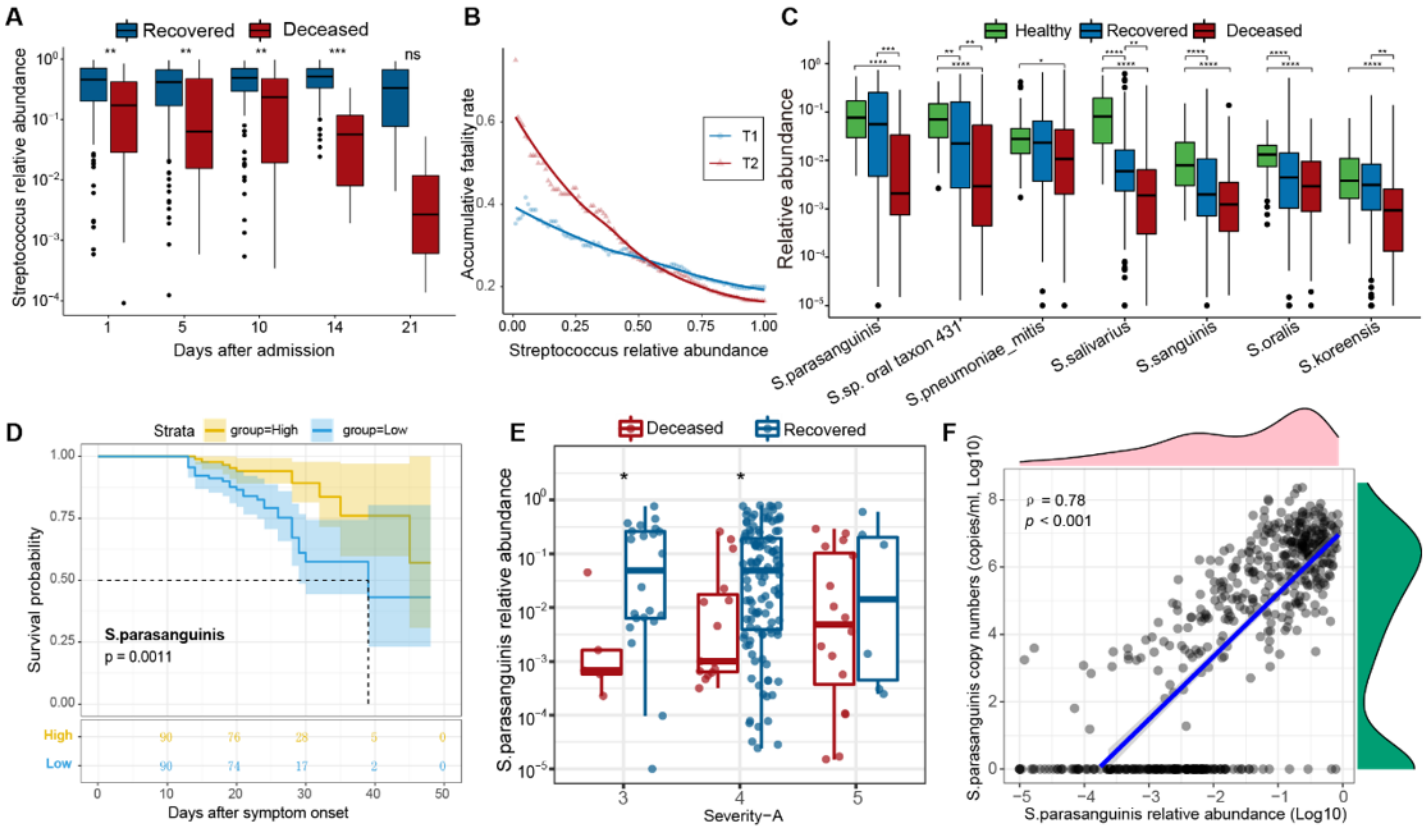

Figure 3

Streptococcus is a potential prognostic marker for COVID-19. A. The abundance of Streptococcus in recovered and deceased patients at different time points. B. The accumulative fatality rate for samples with different Streptococcus abundances. C. The abundances of Streptococcus species with mean abundances greater than 1% at T1. D. The Kaplan-Meier curves for two groups classified by the abundance of S. parasanguinis. The p-value was calculated by log-rank test. E. The abundance of S. parasanguinis in different severity groups. F. Correlation between the abundance inferred from the sequencing data and that estimated from qPCR. The regression line and Spearman's rank correlation coefficient and p-value are shown in the figure.

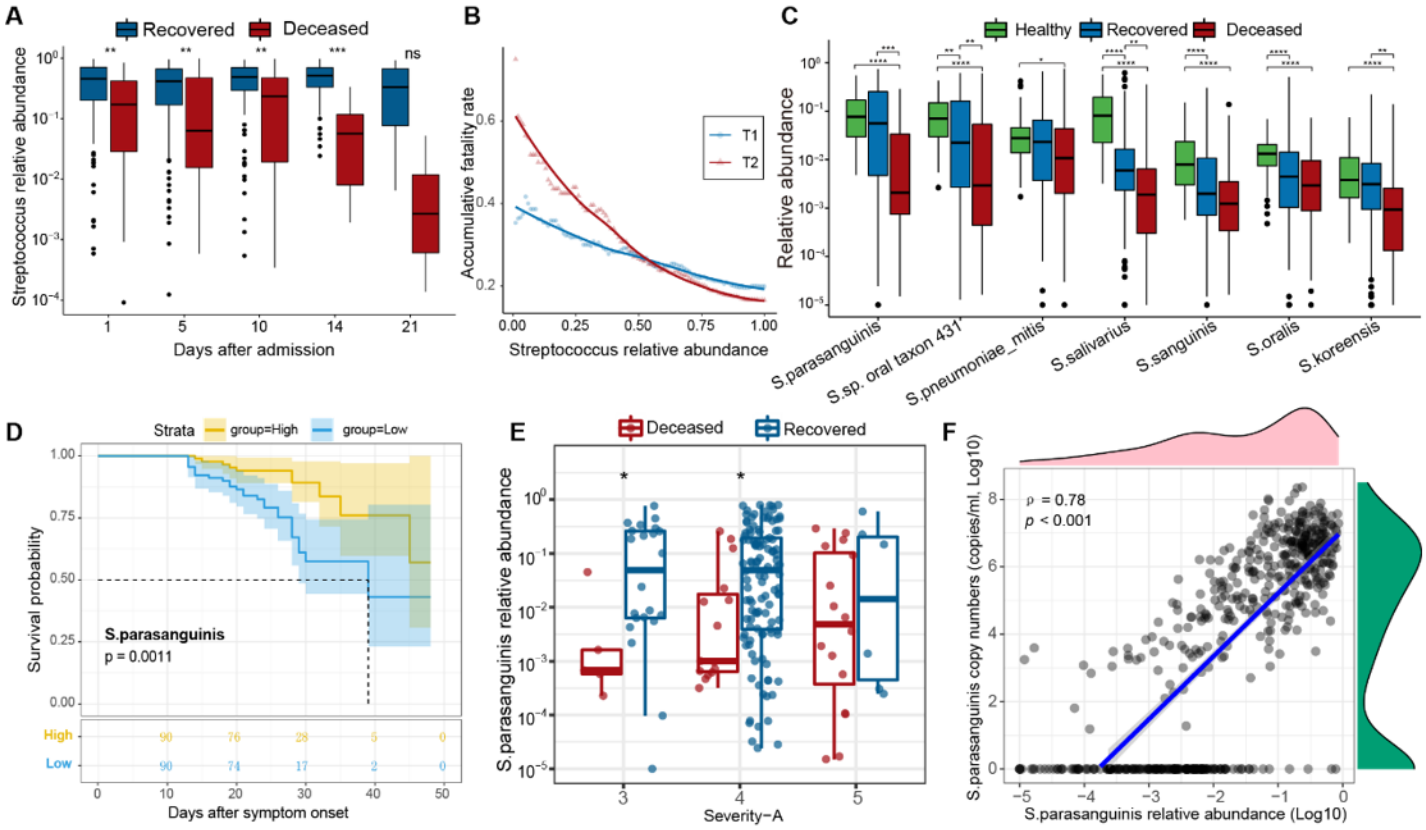

Figure 3

Streptococcus is a potential prognostic marker for COVID-19. A. The abundance of Streptococcus in recovered and deceased patients at different time points. B. The accumulative fatality rate for samples with different Streptococcus abundances. C. The abundances of Streptococcus species with mean abundances greater than 1% at T1. D. The Kaplan-Meier curves for two groups classified by the abundance of S. parasanguinis. The p-value was calculated by log-rank test. E. The abundance of S. parasanguinis in different severity groups. F. Correlation between the abundance inferred from the sequencing data and that estimated from qPCR. The regression line and Spearman's rank correlation coefficient and p-value are shown in the figure.

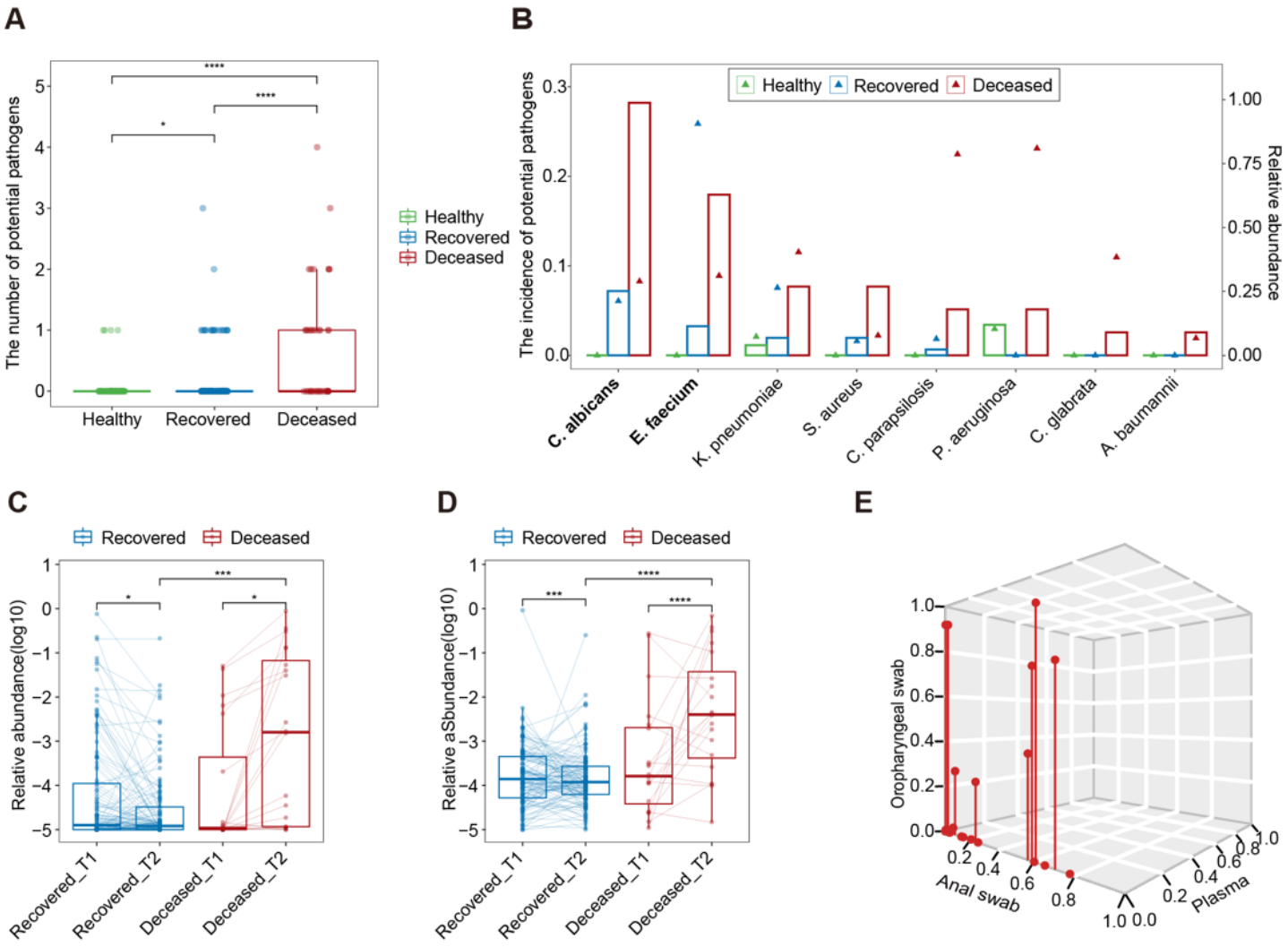

Figure 4

Co-detection of other pathogens in COVID-19 patients. A. The number of potential pathogens detected in each sample. B. The incidence of potential pathogens in different groups. Barplot shows the incidence of each potential pathogen, and the triangle shows the median abundance of the pathogen. The potential pathogens enriched in decreased COVID-19 patients were labeled in bold (Fish Exact test). C-D. The abundance of Candida albicans (C) and Enterococcus faecium (D) at T1 and T2 in paired samples from the same patient. Samples from the same patient were connected by a solid line. E. The abundance of Enterococcus faecium in the oropharyngeal swab, plasma, and anal swab in 32 paired samples.

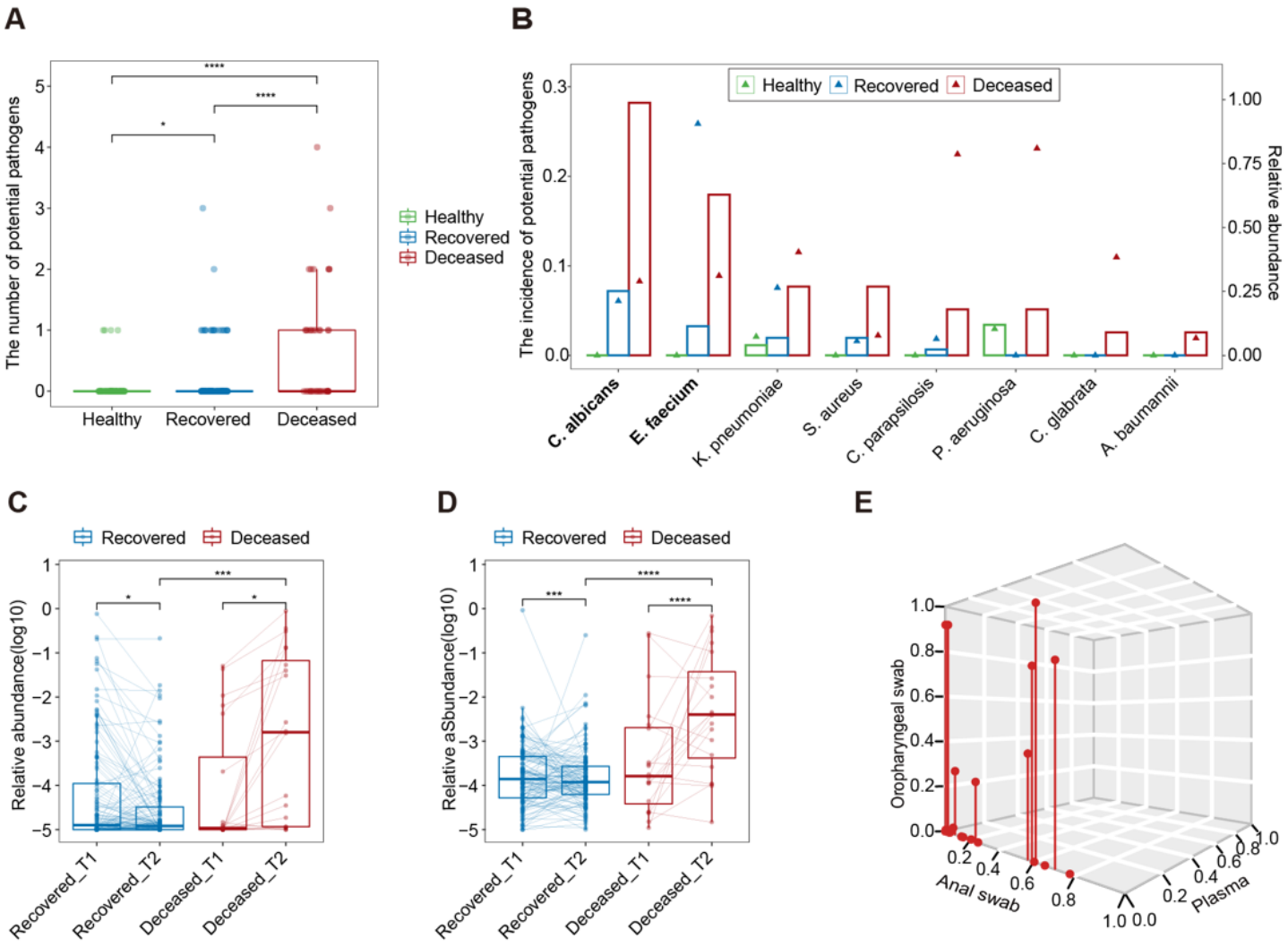

Figure 4

Co-detection of other pathogens in COVID-19 patients. A. The number of potential pathogens detected in each sample. B. The incidence of potential pathogens in different groups. Barplot shows the incidence of each potential pathogen, and the triangle shows the median abundance of the pathogen. The potential pathogens enriched in decreased COVID-19 patients were labeled in bold (Fish Exact test). C-D. The abundance of Candida albicans (C) and Enterococcus faecium (D) at T1 and T2 in paired samples from the same patient. Samples from the same patient were connected by a solid line. E. The abundance of Enterococcus faecium in the oropharyngeal swab, plasma, and anal swab in 32 paired samples.

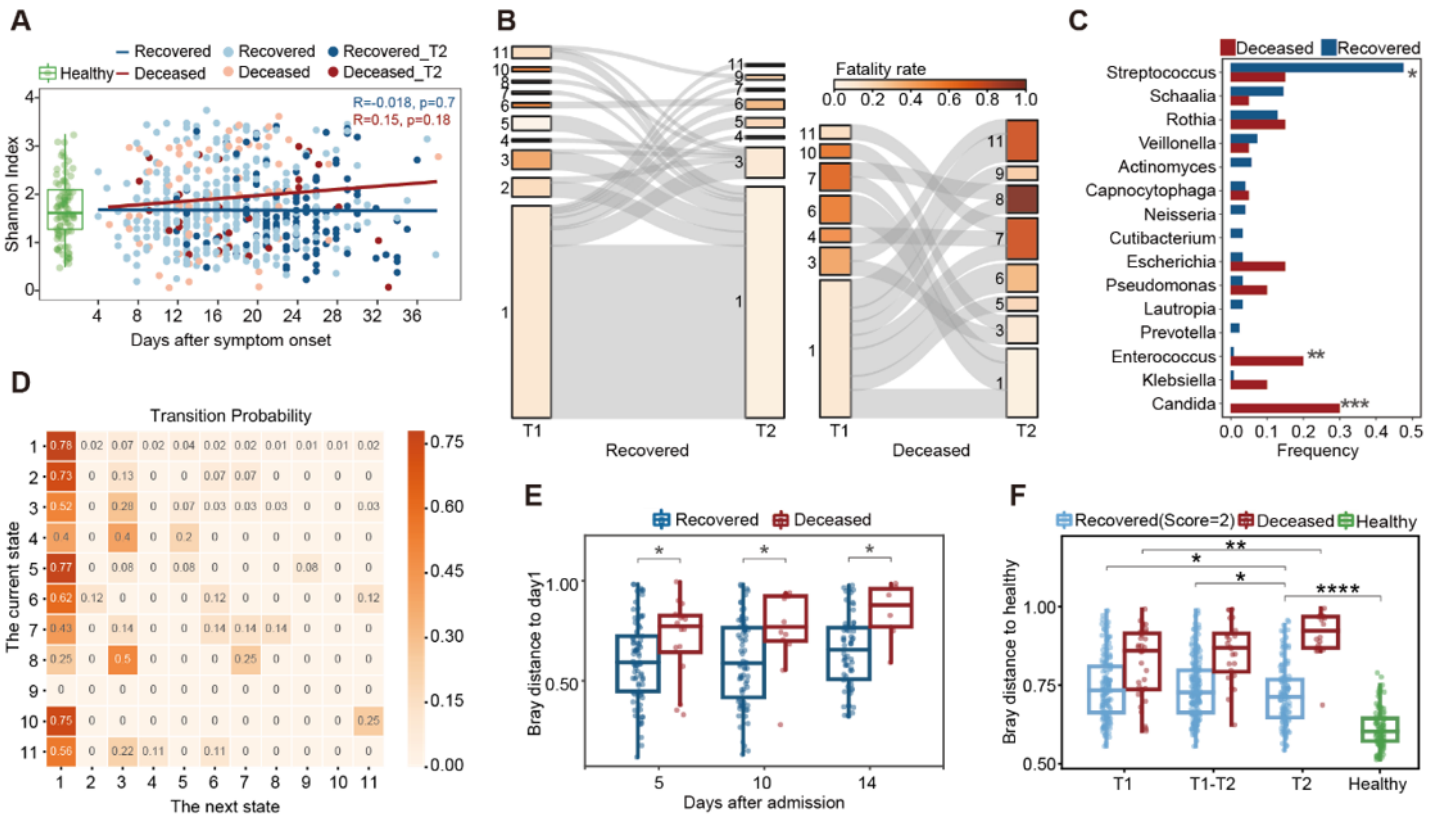

Figure 5

Dynamics of the upper respiratory tract microbiota and its association with the fatality risk. A. Shannon diversity index in COVID-19 patients and healthy controls. Samples collected at T2 from deceased and recovered patients were labeled as Deceased_T2 and Recovered_T2, respectively. Regression lines were drawn for both recovered and decreased patients. B. The transitional Sankey diagram of 11 clusters from T1 to T2. The color of rectangles indicates the fatality rate, and the height is proportional to the number of samples assigned to each cluster. C. The frequency of genera with abundance markedly increased from T1 to T2 (the increase of abundance > 0.1) in all samples. The significance was calculated by the Fisher Exact test. D. Transition rate between each cluster. E. Bray-Curtis distance to the sample of the first day after admission for each patient at different time points. F. Bray-Curtis distance to the HC at different time points. Only patients recovered within 28 days were included. The distances within healthy controls are shown for comparison.

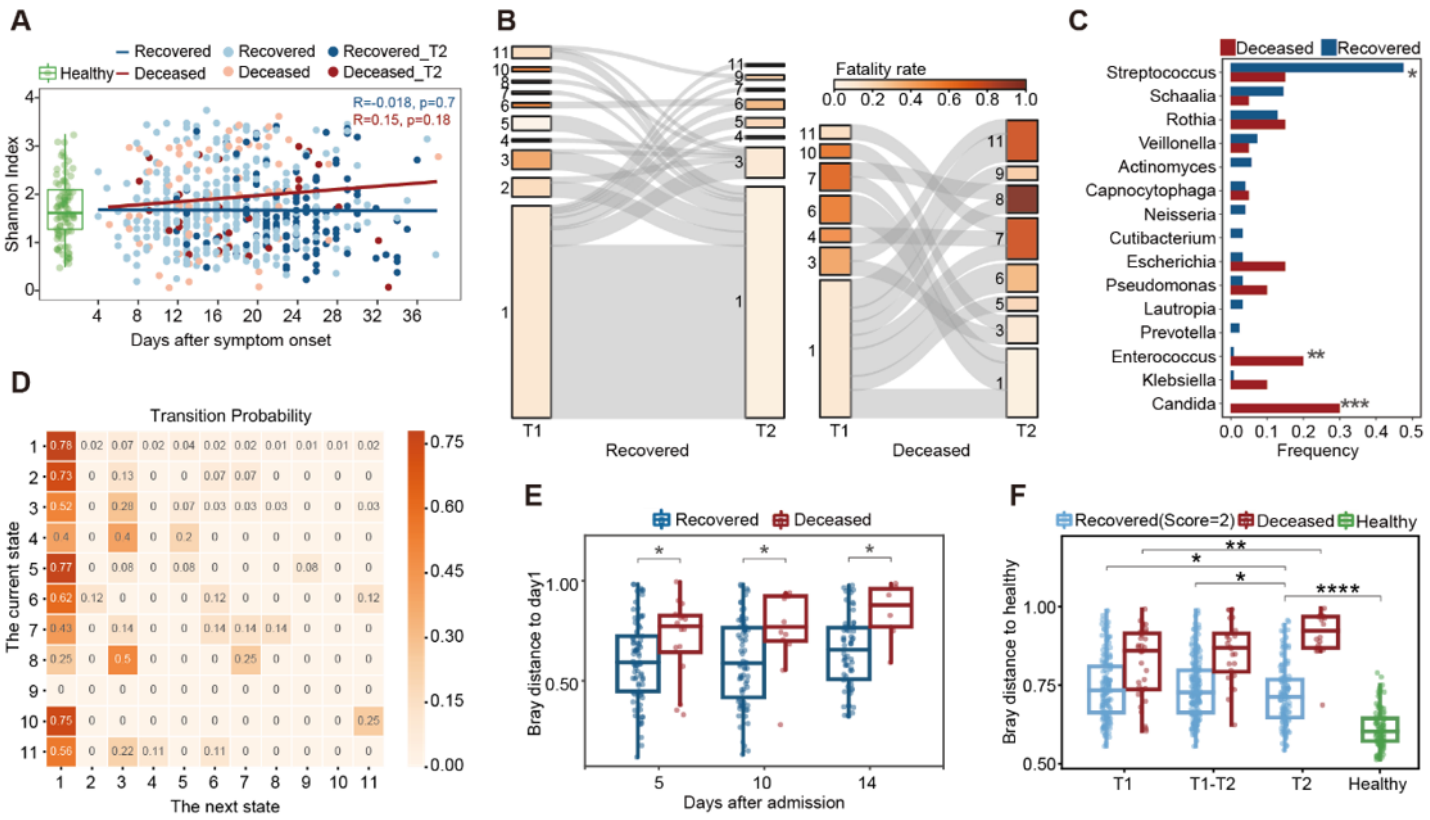

Figure 5

Dynamics of the upper respiratory tract microbiota and its association with the fatality risk. A. Shannon diversity index in COVID-19 patients and healthy controls. Samples collected at T2 from deceased and recovered patients were labeled as Deceased_T2 and Recovered_T2, respectively. Regression lines were drawn for both recovered and decreased patients. B. The transitional Sankey diagram of 11 clusters from T1 to T2. The color of rectangles indicates the fatality rate, and the height is proportional to the number of samples assigned to each cluster. C. The frequency of genera with abundance markedly increased from T1 to T2 (the increase of abundance > 0.1) in all samples. The significance was calculated by the Fisher Exact test. D. Transition rate between each cluster. E. Bray-Curtis distance to the sample of the first day after admission for each patient at different time points. F. Bray-Curtis distance to the HC at different time points. Only patients recovered within 28 days were included. The distances within healthy controls are shown for comparison.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytext.docx
- Supplementarytext.docx
- Supplementaryfigures.docx

- Supplementaryfigures.docx
- Supplementarytables.pdf
- Supplementarytables.pdf