Long-read 16S-seq reveals nasopharynx microbial dysbiosis and enrichment of Mycobacterium and Mycoplasma in COVID-19 patients: A source of co-infection

Punit Prasad (punit@ils.res.in)
Institute of Life Sciences https://orcid.org/0000-0002-9132-6078

Soumendu Mahapatra
Institute of Life Sciences

Rasmita Mishra
Institute of Life Sciences

Krushna Chandra Murmu
Institute of Life Sciences

Shifu Aggarwal
Institute of Life Sciences

Manisha Sethi
Institute of Life Sciences

Priyanka Mohapatra
Institute of Life Sciences

Arup Ghosh
Institute of Life Sciences

Rina Yadav
Institute of Life Sciences

Hiren Dodia
Institute of Life Sciences

Shamima Azma Ansari
Institute of Life Sciences

Saikat De
Institute of Life Sciences

Deepak Singh
Institute of Life Sciences

Amol Suryawanshi
Institute of Life Sciences

Rupesh Dash
Institute of Life Sciences

Shantibhushan Senapati
Institute of Life Sciences
Tushar K. Beuria
Institute of Life Sciences
Soma Chattopadhyay
Institute of Life Sciences
Gulam Hussain Syed
Institute of Life Sciences
Rajeeb Swain
Institute of Life Sciences
Sunil K. Raghav
Institute of Life Sciences
Ajay Parida
Institute of Life Sciences

Research Article

Keywords: Nasopharyngeal microbiome, Nanopore, COVID-19, 16S rRNA, Mycobacterium and Mycoplasma

Posted Date: December 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1149233/v1

Long-read 16S-seq reveals nasopharynx microbial dysbiosis and enrichment of *Mycobacterium* and *Mycoplasma* in COVID-19 patients: A source of co-infection

Punit Prasad[1*$], Soumendu Mahapatra[1*], Rasmita Mishra[1*], Krushna Chandra Murmu[1], Shifu Aggarwal[1], Manisha Sethi[1], Priyanka Mohapatra[1], Arup Ghosh[1], Rina Yadav[1], Hiren Dodia[1], Shamima Azma Ansari[1], Saikat De[1], Deepak Singh[1], Amol Suryawanshi[1], Rupesh Dash[1], Shantibhushan Senapati[1], Tushar K. Beuria[1], Soma Chattopadhyay[1], Gulam Hussain Syed[1], Rajeeb Swain[1], Sunil K. Raghav[1], Ajay Parida[1$]

[1]Institute of Life Sciences, Bhubaneswar, Odisha, India.

* These authors contributed equally to this work

[$] Correspondence: **Ajay Parida, Ph.D.**
Institute of Life Sciences,
Nalco Square, Chandrasekharpur,
Bhubaneswar, Odisha – 751023
Phone: +91-674-2304324
Email: ajayparida@ils.res.in drajayparida@gmail.com

Punit Prasad, Ph.D.
Institute of Life Sciences,
Nalco Square, Chandrasekharpur,
Bhubaneswar, Odisha – 751023
Phone: +91-674-2304319
Email: punit@ils.res.in punit.ils@gov.in

Running title: 16S rRNA-sequencing of nasopharyngeal microbiome identifies *Mycobacterium* and *Mycoplasma* in COVID-19 patients

Keywords: Nasopharyngeal microbiome, Nanopore, COVID-19, 16S rRNA, *Mycobacterium* and *Mycoplasma*

Abstract

Background

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) is a major global health concern. This virus infects the upper respiratory tract and causes pneumonia-like symptoms. So far, few studies have shown alterations in nasopharyngeal (NP) microbial diversity, enrichment of opportunistic pathogens and their role in co-infections during respiratory infections. Therefore, we hypothesized that microbial diversity changes, with increase in the population of opportunistic pathogens, during SARS-CoV2 infection in the nasopharynx which may be involved in co-infection in COVID-19 patients.

Methods

The 16S rRNA variable regions, V1–V9, of NP samples of control and COVID-19 (symptomatic and asymptomatic) patients were sequenced using the Oxford Nanopore™ technology. Comprehensive bioinformatics analysis for determining alpha/beta diversities, non-metric multidimensional scaling, correlation studies, canonical correspondence analysis, linear discriminate analysis, and dysbiosis index were used to analyze the control and COVID-19-specific NP microbiomes.

Results

We observed significant dysbiosis in COVID-19 NP microbiome with increase in abundance of opportunistic pathogens at genus and species levels in asymptomatic/symptomatic patients. The significant abundance of *Mycobacteria* spp. and *Mycoplasma* spp. in symptomatic patients suggest their association and role in co-infections in COVID-19 patients.

Furthermore, we found strong correlation of enrichment of *Mycobacteria* and *Mycoplasma* with the occurrences of chest pain and fever in symptomatic COVID-19 patients.

Conclusion

This is the first study from India to show the abundance of *Mycobacteria* and *Mycoplasma* opportunistic pathogens in non-hospitalized COVID-19 patients and their relationship with symptoms, indicating the possibility of co-infections.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic, a global health threat, is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The symptoms range from fever, throat pain, and loss of taste and smell to severe congestion in the chest, drop in oxygen levels, pneumonia, and acute respiratory distress syndrome (1). A significant population worldwide has been found to be asymptomatic and to act as spreaders of the infection (2). The virus enters the host via the upper respiratory tract (URT) where the spike protein binds to the angiotensin I converting enzyme 2 (ACE2) receptor, an essential step in invading host cells to cause progressive disease (3, 4). Random mutations in the SARS-CoV2 spike protein and receptor-binding domain promote efficient invasion and enhance pathogenicity (5).

The nasopharyngeal (NP) tract is inhabited by a large number of microbial communities which maintain normal homeostasis (6). Studies have revealed association between microbial communities that influence viral infections of the lung, such as chronic rhinosinusitis, asthma, pneumonia, and cystic fibrosis in the URT (7, 8). URT infections lead to decrease in microbial diversity, resulting in dysbiosis, which potentiate increase in the opportunistic pathogen population and promote the growth of other pathogens in immunocompromised hosts (9, 10).

Reports have shown that NP swabs in virus transport media can be used to investigate the NP microbial composition in COVID-19 patients (11, 12). Recent studies have revealed overall compositional changes in the NP microbiota and increase in the number of opportunistic pathogens such as *Rothia* and *Veillonella* in COVID-19 patients with shortness of breath (11, 13, 14). Secondary infections in patients with COVID-19 are associated with the abundance of opportunistic pathogens such as *Moraxella*, *Corynebacterium*, *Haemophilus*, *Stenotrophomonas*, *Acinetobacter*, *Fusobacterium periodonticum*, *Mycobacterium* spp., *Mycoplasma pneumonia* and *Pseudomonas aeruginosa* (15-19).

The treatment modalities of SARS-CoV2 infection are not yet established and patients are administered immunosuppressants to regulate cytokine storm, thereby creating an environment conducive for the growth of opportunistic pathogens and co-infections in COVID-19 patients (19, 20). Several case reports have highlighted fungal, viral, and bacterial co-infections in COVID-19 patients with severe symptoms (19). However, information regarding the crosstalk between SARS-CoV-2 infection and NP microbiota is scarce. The absence of an animal model makes it difficult to test and validate the role of NP microbiota in SARS-CoV-2 infection. Various studies so far have shown differences in the abundance of various opportunistic pathogens in the NP microbiota of patients, which is one of the bottlenecks in this area of research. Hence more studies on the NP microbiome are required for understanding its role in COVID-19 patients and its relation with symptom severity, as microbial composition is dynamic and changes with multiple extrinsic factors.

Here, we hypothesized that similar to other respiratory diseases (asthma, pneumonia, and cystic fibrosis) dysbiosis might occur in the COVID-19 NP microbial community (high virus-bacteria interaction zone), resulting in emergence of opportunistic pathogens, which may be

associated with symptoms of COVID-19. In this regard, we also predicted that the alterations in NP microbiota may occur in both symptomatic and asymptomatic COVID-19-patients and that the microbial composition may differ between these two categories of patients, studies on which are limited. In this study, we have investigated the alterations in the NP microbial community of patients with active COVID-19 (n = 46) and compared them with those of the healthy individuals (n = 12). We also analyzed the differences in the microbial community composition between asymptomatic (n = 25) and symptomatic (n = 21) COVID-19 patients. We have used the metataxonomic approach of 16S rRNA long-read sequencing (V1-V9) with the Nanopore sequencing method to elucidate the reduction in microbial diversity in patients with COVID-19. We found enrichment of *Mycobacteria* and *Mycoplasma* in both asymptomatic and symptomatic patients with COVID-19, however, the abundance in symptomatic patients were much higher and it strongly correlated with chest pain and fever.

Materials and Methods

Ethical approval: Ethical permission for nasopharyngeal microbiome study and the biorepository was obtained from the Institutional Ethical Committee (IEC)/Institutional Review Board (IRB) of the Institute of Life Sciences [(102/HEC/2020) and (100/HEC/2020)]. Approval was also obtained from the Institutional Biosafety Committee (IBSC) (V-122-MISC/2007-08/01/2/2.1) for this study and the biorepository (V-122-MISC/2007-08/01) and from the Review Committee on Genetic Manipulations (RCGM) under Department of Biotechnology, Ministry of Science and Technology.

Sample collection and reverse transcription-polymerase chain reaction (RT-PCR): In total, 60 NP samples were collected for 16S rDNA amplicon sequencing from the Institute of Life Science (ILS) COVID-19 sample biorepository unit. The COVID-19-positive samples (n

= 47) were confirmed by amplifying the genes encoding SARS-CoV-2 nucleocapsid, spike, and ORF1ab/RdRP using either TaqPath™ COVID-19 combo kit (Invitrogen, A47814) or Meril COVID-19 one-step RT-PCR kit (Meril Diagnostics, NCVPCR-02). All samples were collected in the hospital setup prior to the medication. These patients were not treated with antibiotics as they were not aware of their COVID-19 test results. The patients were grouped as symptomatic (n = 22) or asymptomatic (n = 25) based on their clinical data. The control samples (n = 13) were negative for SARS-CoV-2 virus RNA, and none of the subjects from whom the samples were obtained had any flu-like symptoms. All samples were collected in viral transport media (VTM) and stored at –80□C until DNA isolation.

DNA extraction and PCR amplification: DNA was isolated using the PureLink™ microbiome DNA purification kit (Invitrogen, A29790) according to the manufacturer's protocol and eluted in 40 µl elution buffer. The quality and quantity of DNA were determined using the Multiskan™GO spectrophotometer (Thermo Scientific).

16S rDNA amplification, library preparation, and sequencing: The V1–V9 variable regions of the 16S rRNA gene were amplified using 130-F (5′-GGCGGATCCAAGGAGGTGTTCCAGCCGC-3′) and 139-R (5′-GGCCTCGAGAGAGTTTGATCCTGGCTCAGG-3′) primers. PCR (50 µl) was set up using total DNA (10 ng) isolated from NP samples, primers (5 nM), and NEB Q5® High-Fidelity 2 X master mix (NEB, M0492L) per the manufacturer's protocol. The amplicons (~1.6 kb) were analyzed on 0.8% agarose gel and cleaned using DNA Clean and Concentrator-25 kit (Zymo Research, D4034). The PCR products were quantified using a Qubit 4 fluorometer (Thermo Scientific) using the Qubit® dsDNA BR assay kit (Thermo Scientific, Q32853). Amplicon libraries were generated following the Oxford Nanopore 1D library preparation protocol using

115 the PCR barcoding (96) genomic DNA kit (Oxford Nanopore™, SQK-LSK109). Equimolar
116 amounts of amplicon libraries were pooled and sequenced using the MinION OXFORD
117 NANOPORE™ device at the ILS DNA sequencing facility.

118 **Microbiome data processing:** RAW fast5 files were generated using the MinKNOW™ tool
119 for individual samples. Base calling was performed using the Guppy base-caller and fastq files
120 were generated. FastQC of each sample was performed using the Babraham fastqc suite
121 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), followed by trimming of low
122 quality reads using nanoflit. Operational taxonomic units (OTUs) were generated using
123 Kraken2 (https://ccb.jhu.edu/software/kraken2/index.shtml) (21) and the unclassified reads
124 were filtered for downstream analysis using the 'phyloseq' 'R" package to generate combined
125 OTUs for all the samples and metadata (Supplemental Table 1). Read counts for mitochondria
126 and chloroplast were discarded. Normalization and differential OTU abundance were
127 determined between control, and symptomatic and asymptomatic subjects using the DESeq2
128 function (cutoff of p-value ≤ 0.05) (22).

129 **In-depth microbiome data analysis:**

130 Diversity analysis: Alpha diversity was assessed using the Shannon and Simpson diversity
131 indices. Statistical significance was estimated using the Wilcoxon rank sum test. The beta
132 diversity significance among groups was examined using PERMANOVA (p-value 0.001).
133 Ordination analysis was performed using principal coordinate analysis (PCoA), non-metric
134 multidimensional scaling (NMDS) and canonical correspondence analysis (CCA). R
135 packages, 'microbiome','Vegan','ade4', and ggpubr, were used for analysis and 'ggplot2' was
136 used for visualization.

Dysbiosis index: Microbiome dysbiosis in each sample was calculated based on Bray-Curtis distances. All samples were subjected to PCoA using Bray-Curtis distances. Next, the centroid (median) of the control subjects was calculated along PCoA axes. The dysbiosis score for each sample was calculated as a Euclidian distance between its position in the PCoA space and control centroid (DI(X, HC) = $\sqrt{(Xi - HCi)^2 + (Xj - HCj)^2}$ | (DI: Dysbiosis Index, X: Samples, HC: Control Centroid). Their significance was assessed using Wilcoxon and Kruskal-Wallis test (23).

Sample correlation: Correlation matrix between samples and OTUs for each taxonomic level (phylum, order, family, and genus) from differential OTUs was obtained using Spearman's correlation method and it was visualized as a heat map. Correlation coefficients for each sample correlation pair and each classification level and density plot were plotted with mean and median. The Kolmogorov-Smirnov test (KS) was used to determine the significance in sample groups (control, asymptomatic, and symptomatic).

Linear discriminant analysis (LDA) effect size (LEfSe) analysis: The LEfSe was calculated using the online Galaxy web application and the Huttenhower lab's tool (https://huttenhower.sph.harvard.edu/galaxy/). LDA effect size was calculated using the Kruskal-Wallis sum rank test (p-value ≤ 0.05) and it detected differential abundant features at genus and species level within three sample groups. Significance at the taxonomic level was then tested using the pairwise Wilcoxon rank-sum tests (p-value ≤ 0.05). Finally, the effect size of each differentially abundant feature was estimated using LDA. One-against-all sample groups were compared and a linear discriminant analysis score greater than 3.6 was set as the threshold; all-against-all sample groups were compared and a linear discriminant analysis

score greater than 2.0 was set as the threshold. Cladogram was used for identification of taxa at different levels of the taxonomic hierarchy between sample groups (LDA score > 2).

Network analysis: Network was constructed using weighted correlation network analysis or weighted gene co-expression network analysis (WGCNA) (24). Briefly, pairwise Spearman correlation between OTUs (which was generated from LefSe analysis) was calculated using the WGCNA function. Network metrics such as betweenness, closeness, Eigen centrality, and PageRank centrality of the resulting network were calculated and visualized using 'Gephi', (https://gephi.org/) (25).

Results

Study design and subject attributes

The role of the microbiome in viral infections is an emerging field and we were interested in understanding NP microbial diversity and changes in its composition in patients during SARS-CoV-2 infection. We collected NP samples from COVID-19 patients between 11th May 2020 and 10th October 2020 to study alterations in the NP microbiome. The schematic representation of the study is shown in Figure 1A. In total, 60 NP samples subjects (infected, n = 47 and control, n = 13 subjects, positive and negative for SARS-CoV2 RT-qPCR test respectively) were obtained from the Institute of Life Sciences biorepository. Out of 47 SARS-CoV-2-positive subjects, 25 were asymptomatic and 22 were symptomatic with mild symptoms (Figure 1B). In total, 179, 59, 691 reads were generated. Two samples with low read counts (1 from control and other from symptomatic category) were excluded and the final study was performed with 58 subjects, including the control (C) [n = 12 (21%)], and asymptomatic [IA, infected asymptomatic; n = 25 (43%)] and symptomatic patients [IS,

infected symptomatic; n = 21 (36%)]. The details of the participants considered for this study are shown in Table 1. Differential OTUs (n = 795, $p \leq 0.05$) were obtained from 3482 OTUs using the deseq2 function by comparing with control NP subjects. For downstream analysis differential, 795 OTUs were considered. We used the t-distributed stochastic neighbor embedding (t-SNE) dimension reduction method to obtain the overall distribution of NP samples with 795 OTUs (Figure 1C). We found that the control and SARS-CoV-2-infected subjects showed distinct segregation of OTUs in the NP microbiome, while asymptomatic and symptomatic subjects showed modest separation. This indicated that the abundance of the 795 differential OTUs potentially determines the compositional distribution patterns.

NP microbiome diversity was significantly altered in SARS-CoV-2-infected patients

Distinct distribution of OTUs from control and infected patients prompted us to compare the alpha diversities (evenness and richness of bacterial community compositions) using Shannon and Simpson indices. The Shannon and Simpson microbial diversity indices differed significantly between control and SARS-CoV-2-infected participants (p-value $\leq$ 0.05) in pairwise Wilcoxon rank test (Shannon p-value $=3.0*10^{-4}$ and Simpson p-value$=3.3*10^{-3}$) (Figure 2A, B). Although the alpha diversity indices for samples from symptomatic and asymptomatic patients compared to control subjects were found to be significantly reduced, no difference was observed between symptomatic and asymptomatic samples (Figure 2C, D). Furthermore, we used a linear regression model to establish the association between total OTU read counts for each sample and Shannon/Simpson alpha diversity indices. We found negative correlation for both Shannon (IA - R = -0.35, $R^2 = 0.44$, p = 0.083; IS - R = -0.54, $R^2 = 0.48$, p = 0.012) and Simpson (IA - R = -0.58, $R^2 = 0.68$, p = 0.0028; IS - R = -0.77, $R^2 = 0.63$, p = $7.7*10^{-5}$) alpha diversity indices with 95% confidence intervals with total OTU counts (Figure

2E, F). To further understand the microbial composition dissimilarity within the samples, we analyzed beta diversity using PCoA and applied both unweighted (microbial richness) and weighted (microbial richness and abundance) unifrac distance methods. The first two components of PCoA showed 60.3% and 80.1% variance for the unweighted and weighted unifrac method. The overall difference in microbial population showed two different clusters of control and SARS-CoV-2-infected patients (IA and IS) in the unifrac weighted method, while the unifrac unweighted method showed more clear segregation between symptomatic and asymptomatic samples (Figure 2G). We assessed the significance of beta diversity to calculate unifrac distance matrix (PERMANOVA test with 999 permutations) for both unweighted and weighted methods and found that the three sample groups (C, IA, and IS) differed significantly ($P = 0.001$) with 18% variance explained ($R^2 = 0.18842$).

NP microbiome dysbiosis in patients with COVID-19

Alterations in the microbial diversity prompted us to determine microbial dysbiosis index (DI) (variations in the microbial community) across the three groups (C, IA, and IS). We performed PCoA using the Bray Curtis distance matrix and found that NP microbiota was significantly altered ($p = 0.001$) with 61% variation in distances explained ($R^2 = 0.6136$), as assessed using the ADONIS test. Next, we calculated the Euclidean distance from the centroid for samples from control (median = 0.3404), asymptomatic (median = 0.1881) and symptomatic (median = 0.1511) individuals and calculated the DI (Supplementary figure 1B). The overall observed DI was significant (Kruskal-Wallis test, p = 1.317E-07) across all the groups. Pairwise comparison showed significant dysbiosis between control vs. symptomatic ($p = 5.6E{-}09$) and control vs. asymptomatic ($p = 1.1E{-}09$) groups; however, dysbiosis between asymptomatic and symptomatic ($p = 0.016$) pair was also significant, although the value was moderately low

compared to those of the other comparisons (Figure 2H). We also observed highly significant dysbiosis (p = 2.2E-12) between the control and infected groups (Supplemental Figure 1A, 1C). This showed that compared to control subjects, the NP microbial community was severely altered in both symptomatic and asymptomatic COVID-19 patients.

Distinct microbial composition and abundance at phylum and family levels in patients with SARS-CoV-2 infection

The alpha and beta diversities, and DI showed that the NP microbiome was significantly altered in COVID-19 patients. Next, we aimed to identify the microbial communities that were altered at different taxonomic levels in three sample groups. We found 795 differential OTUs, out of which 12 phyla, 65 orders, 126 families, and 240 genera were present in all three groups (C, IA, and IS) (Supplemental Table 1). The 12 phyla and their significance is shown in Table 2. The most significant phyla in the SARS-CoV-2-infected groups (symptomatic and asymptomatic) were Actinobacteria (p = 9.96E-07), Firmicutes (p = 4.65E-02), and Proteobacteria (p = 9.61E-07), including 8 other phyla assessed using the Kruskal-Wallis test. In contrast, Bacteroidetes (p = 1.48E-06) and Proteobacteria (p = 6.56E-07) were highly abundant in the control group (non-infected) (Supplemental Figure 2A). Furthermore, we analyzed the relative abundance of the top 30 families and found enrichment of *Mycobacteriaceae*, *Propionibacteriaceae*, and *Streptomycetaceae* (Supplemental Figure 2B). These families contain opportunistic pathogens, which were found in both symptomatic and asymptomatic COVID-19 patients, while they were absent in the healthy controls. The top families and their significance is shown in Table 3.

Taxonomic classifications based on OTU abundance showed sample group segregation at the genus level

To further our understanding regarding the 795 differentially abundant OTUs, we used the NMDS approach at phylum, order, family, and genus levels for C, IA, and IS sample groups using the Bray-Curtis distance matrix. Analysis of statistical significance using analysis of similarities (ANOSIM) for phylum (R = 0.262, p = 1.7E-03), order (R = 0.322, p = 3E-04), family (R = 0.3461, p = 3E-04), and genus (R = 0.3507, p = 3E-04) showed gradual increase in R-value for genus. This indicated that as we go lower in taxonomic classification, the variance in the OTUs provides better sample segregation. The differential OTUs present at the genus level in three sample groups showed significant level of dissimilarity (p = 3E-04) with R = 0.3507 and clear sample segregation (Figure 3A). We used sample correlation (Spearman matrix) to further validate the NMDS findings and identify the NP OTU differences between C, IA, and IS sample groups (Supplemental Figure 3A-D). To further reconcile the distinct sample segregation at higher to lower taxonomic level based on OTU abundances, we plotted density histogram of correlation coefficient values (obtained in sample correlation). The mean and median value of each density plot revealed lack of difference between the C, IA and IS groups at the phylum level and subtle differences were observed at the order and family levels. However, at the genus level, we found comprehensible differences between C (mean = 7.95E-01; median = 6.39E-01), IA (mean = 5.65E-01; median = 8.33E-01), and IS (mean = 6.51E-01; median = 7.01E-01) (Table 4) (Figure 3B). To evaluate the statistical significance of densities based on sample segregation, we calculated cumulative distribution distance (D) and significance between C, IA, and IS groups using the KS test for each taxonomic rank (Table 5). Based on the 'D' value, the samples were well distributed in C vs. IA (D = 5.94E-01; p-value < 2.2E-16), C vs. IS (D = 5.06E-01; p = 3.308E-14), and IA vs. IS (D = 2.28E-01; p = < 2.2E-16) at the genus level. The differences in overall sample distribution were more highly

enriched at the genus level than between these three groups. Although the 'D' value between IA and IS groups was less, their distribution pattern differed significantly. Based on the above observations, we considered the genus level OTUs (n = 240) for downstream analysis.

Cluster-specific OTUs at genus level revealed unique sample-specific OTUs

To gain insights regarding how the bacterial genera were segmented among three groups of samples, we performed genus level OTU correlation (n = 240) and calculated the correlation coefficient (Spearman correlation), followed by unsupervised hierarchical clustering (Figure 3C). We identified five distinct clusters, C1 (n = 23), C2 (n = 109), C3 (n = 59), C4 (n = 33) and C5 (n = 16), with variable number of OTUs (Supplemental Table 2). The heat map corresponding to each cluster is shown in Supplemental Figure 4. The relative abundance density maps of cluster-wise OTUs were constructed, which distinguished OTUs that were enriched in IA/IS (C1, C3, C4, and C5) and in the control samples (C2) (Figure 3D). Some of the enriched cluster-specific OTUs in IA/IS are *Mycobacterium* (1763), *Mycolicibacterium* (1766), *Mycobacteroides* (1774), *Halothiobacillus* (927), *Flavobacterium* (986), *Bifidobacterium* (1695), *Streptomyces* (1884), *Rothia* (2047), and *Mycoplasma* (2100). C2, a control-specific cluster, contained OTUs such as *Thermomicrobium* (500), *Kingella* (502), *Enterobacter* (547), *Bacteroides* (821), and *Prevotella* (840). Thus, this analysis shows the distinction in genus-specific OTUs for both SARS-CoV-2-infected and control subjects. Next, we performed CCA on each of the clusters (C1 to C5) to eliminate sample heterogeneity and enhance the stringency of our analysis pipeline (Figure 3E). We considered the first two components of CCA that explained cumulative variance for the clusters. Cluster C4 explained the highest cumulative variance of 94.8%, while cluster C3 showed the lowest variance of

12.9% (Figure 3F, Supplemental Figure 5A-E). CCA showed efficient sample clustering, which is reminiscent of the density plot (Figure 3D-E).

LefSe analysis identified unique OTUs at genus level in patients with COVID-19

The CCA analysis prompted us to select clusters with maximum variance explained. Therefore, we considered all clusters with ≥ 30% variance, which includes all the clusters, except C3, for LDA. OTUs (n = 181) were extracted from clusters (C1, C2, C4, C5) and plotted in a heat map with their abundance (Figure 4A). Different genera could be clearly distinguished between C, IA, and IS sample groups. Next, we performed LefSe to distinguish the most significant microbiomes from C, IA, and IS groups. In a one-against-all comparison (C with IA and IS), we obtained 40 genera in the control group, 34 genera in the symptomatic group, and 4 genera in the asymptomatic group (LDA score [log10] > 3.6). The genera obtained from one-against-all are highlighted in the heatmap (Figure 4A and Supplemental Figure 6A). Relative abundance of each OTU obtained from C, IA, and IS groups are shown in stack plots, with clear segregation in OTUs for individual samples (Supplemental Figure 6B-D). The DI calculated from these genera showed high dysbiosis between control and SARS-CoV-2-infected patients (Supplemental Figure 6E-F). We further increased the LefSe stringency by using all-against-all (each sample group compared with each other) comparisons and constructed a cladogram and a bar plot (Figure 4B-C). All the genera obtained from LefSe (one-against-all and all-against-all) with their LDA scores and comparison are listed in Supplemental Table 3 and Table 6. We obtained 12 significantly enriched genera of *Gallibacterium, Orientia, Acidocella,* and *Citrobacter* in control samples (LDA score [log10] > 2.0), *Mycoplasma, Streptosporangium, Mycobacterium, Mycolicibacterium, Mycolicibacillus,* and *Mycobacteroides* in symptomatic samples, and *Oerskovia* and

Cellulosimicrobium in asymptomatic samples (Figure 4C). The histogram showing the relative abundance of the 12 genera for C, IA, and IS sample groups clearly distinguished each sample type (Figure 4D). Finally, we used weighted correlation network analysis to construct a network (Spearman correlation) with 12 genera identified using the LDA analysis. The network creates two distinct modules, one for control groups and another for both symptomatic and asymptomatic groups. We obtained strong association within the genera of C, IA, and IS sample groups (Table 7). However, the correlation between C vs. IA was extremely weak and correlation was not obtained for C vs. IS groups. The network analysis suggested that the NP microbiota of the control group was clearly distinct from that of the asymptomatic and symptomatic groups. The DI of the 12 genera showed the highest significance between C vs. IS ($p = 4.7E-05$), while significant dysbiosis was not observed between IA and IS groups (Figure 4F). Overall, our analysis confirmed the significance of the genera identified and their associations with symptomatic and asymptomatic COVID-19 patients.

Enrichment of *Mycobacterium* spp. and *Mycoplasma* spp. in patients with COVID-19

After successful identification of 12 genera based on LDA score from all-against-all comparisons, we investigated different species-specific OTUs associated with each of the 12 genus and identified 54 species represented in the heatmap (Figure 6A). The differential abundance of several known opportunistic pathogens such as *Mycobacterium tuberculosis, Mycobacterium avium Mycoplasma pneumoniae, Mycobacterium leprae,* and *Mycobacterium kansasii* were significantly high in the SARS-CoV2 infected patients (Figure 6A-D). The pairwise comparison and significance of the relative abundance of 12 previously reported pathogens are shown in Figure 6B-D and Table 8. We found that all 12 pathogenic species, including *Mycobacterium tuberculosis, Mycobacterium avium, Mycobacterium intracellulare*

342 *(MAC), Mycoplasma pneumoniae, Mycobacterium leprae, Mycobacterium Kansasii,*
343 *Mycobacterium simiae, Mycobacterium xenopi, Mycobacterium gordonae, Mycobacterium*
344 *asiaticum, Mycobacterium szulgai,* and *Mycobacterium ulcerans* were more abundant in
345 symptomatic patients than in asymptomatic patients (Figure 6D). Although abundance of these
346 pathogens was obtained in both IA and IS subjects compared to that in the controls, the overall
347 enrichment in the IS subjects were higher, indicating their possible role in secondary
348 infections.

349 **Distinct correlation of OTUs with clinical symptoms in patients with COVID-19**

350 To evaluate the accuracy of LDA classification that identified eight bacterial genera in the IA
351 and IS sample group, we tested the ROC (receiver operating characteristics)–AUC (area under
352 the curve) score. We obtained a value of 0.8 with 95% confidence interval for true positive
353 classification, showing 80% sensitivity and specificity of data obtained from LDA analysis
354 (Figure 5A). Next, we used the Spearman correlation matrix to identify the association of
355 symptoms with the genera. Interestingly, chest pain showed high positive correlation with the
356 presence of *Mycoplasma, Mycobacterium, Mycolicibacterium, Mycolicibacillus,* and
357 *Mycobacteroides,* which were related to the IS group, and weak correlation with *Oerskovia*
358 and *Cellulosimicrobium,* which were associated with the IA group. *Mycoplasma,* however,
359 showed strong correlation with both chest pain (0.4446) and fever (0.4214) (Figure 5B). ROC-
360 AUC analysis for chest pain and fever yielded 0.90 and 0.79 scores, respectively, with eight
361 bacterial genera (Figure 5C-D). At the species level, fever correlated strongly with most
362 *Mycobacterium* spp. and *Mycoplasma* spp., which were significantly abundant in symptomatic
363 COVID-19 patients (Figure 6E). In short, we established the association of pathogenic
364 microbes with COVID-19 and showed susceptibility to alterations in the NP microbiome in

case of co-infections in SARS-CoV-2. We also identified the compositional difference in NP microbiota between symptomatic and asymptomatic group.

Discussion

Although evidence regarding changes in NP microbiota in COVID-19 patients is available, data regarding association of opportunistic pathogens with disease symptoms, indicative of co-infections, are limiting; in addition, evidence regarding the presence of asymptomatic- and symptomatic patient-specific bacterial communities is still awaited. Previous studies have revealed the association of disease-specific opportunistic pathogens in respiratory diseases such as chronic rhinosinusitis, asthma, pneumonia, and cystic fibrosis (7-10). Therefore, we aimed to identify the NP microbial alterations, opportunistic pathogens, and their association with COVID-19-positive asymptomatic and symptomatic subjects. In this study, we were able to distinguish the microbial community associated with the control, symptomatic, and asymptomatic COVID-19 patients.

Scientists worldwide are attempting to understand the pathophysiology of SARS-CoV-2 infection and the associated co-infections/secondary infections that may arise due to the alterations in the host microbiome. As SARS-CoV-2 infection initiates in the upper respiratory tract, we investigated the alterations in the NP microbiota of COVID-19 patients. We amplified the 16S rRNA gene of variable regions (V1–V9) and performed long-read sequencing using Oxford Nanopore technology. Subsequently, we have used multiple bioinformatics approaches to cross-validate our data sets at various levels and identify the most significant bacterial population in the NP microbiome of COVID-19 patients. We found significant changes in microbial abundance/diversity and DI in SARS-CoV-2-infected patients compared to those of the control. The IA and IS groups also showed overall significant

388 alterations in microbiota composition. We found abundance of opportunistic pathogens such
389 as *Mycoplasma* and *Mycobacterium* in symptomatic patients, which correlated strongly with
390 patient symptoms such as chest pain and fever. Insights into species level revealed the
391 abundance of 12 different bacterial species belonging to the genus *Mycoplasma* and
392 *Mycobacterium* in the SARS-CoV-2-infected patients. To the best of our knowledge, this is
393 the first comprehensive and systematic cohort-based study to report the abundance of
394 opportunistic pathogens *Mycoplasma* and *Mycobacterium* using long read sequencing of 16S
395 rRNA variable regions in patients with SARS-CoV-2 infection.
396 Respiratory infections alter the NP microbiota, reducing microbial diversity, and promotes the
397 growth of opportunistic pathogens (26). Proteobacteria, Firmicutes, and Actinobacteria are
398 present in all NP samples at the phylum level, however, the abundance of Firmicutes and
399 Actinobacteria was significantly higher in both symptomatic and asymptomatic COVID-19
400 patients in our study. Our results are in partial agreement with those of Ventero et al., who
401 found the abundance of Firmicutes, Bacteroidota, Proteobacteria, and Actinobacteria in the
402 NP samples of COVID-19 patients (13). Maio *et al*. and Braun *et al*. did not find any
403 significant alterations in NP microbial composition (12, 27). Other studies have shown the
404 prevalence of opportunistic pathogens such as *Staphylococcus*, *Anelloviridae*, *Pseudomonas,*
405 *Haemophilus, Stenotrophomonas, Redondoviridae,* and *Pseudomonas aeruginosa* in COVID-
406 19 patients (11, 13, 15-18). Compared to earlier reports, our study also revealed overall
407 changes in the composition of the NP microbial community, reduction in bacterial diversity
408 and the presence of opportunistic pathogens.
409 Most of the NP microbial studies involve amplification of the shorter 16S rRNA variable
410 region using the Illumina platform, which is more accurate, but is limited by taxonomic

411 resolution owing to sequencing of shorter reads and specific variable region (28). The
412 taxonomic resolution can be improved to genus, species, and even at the strain level by
413 sequencing the V1□V9 (~ 1600 bp) variable regions of the 16S rRNA gene (28, 29). In this
414 study, we have successfully used the Oxford Nanopore long read sequencing platform and
415 sequenced V1 to V9 (~1.6 kb) of the 16S rDNA variable regions to obtain taxonomic
416 resolution to genus and species level. This has provided us immense advantage in determining
417 the abundance of opportunistic pathogens in the NP of the COVID-19 patients. Until now,
418 only Mostafa et al. has used metagenomics for COVID-19 NP samples using Oxford Nanopore
419 technology. They have sequenced both RNA and DNA from the NP samples without any PCR
420 amplification. They not only identified the SARS-CoV-2 virus in the samples, but also
421 potential pathogens that may lead to co-infections (18). This is the first comprehensive study
422 on high throughput analysis of the NP microbiome in COVID-19 patients from an Odisha
423 cohort and the second in India after Gupta et al., who used the Illumina platform for 16S
424 amplicon sequencing and found enrichment of several opportunistic pathogens (17).
425 Interestingly, *Mycoplasma, Mycolicocibacterium,* and *Mycobacterium* were not present in
426 their list of opportunistic pathogens. This could be due to the short reads sequenced, and
427 analysis pipeline and/or geographical region-specific differences. Nevertheless, the
428 identification of opportunistic pathogens and increase in their abundance in COVID-19
429 patients is one of the important aspects of this study.
430 SARS-CoV-2 infections elicits strong immune responses in the host, leading to cytokine
431 storm, which also affects normal cells. The administration of immune suppressants
432 compromise patient's immunity and render them prone to co-infections with other
433 opportunistic pathogens, worsening the pathophysiology of the COVID-19 patients (19, 20).

Our comprehensive bioinformatics analysis with sample-sample and OTU-OTU correlation analysis distinguished COVID-19-infected and control specific OTUs at the genus level. Furthermore, LDA and network analysis identified significantly high abundance of *Mycobacterium* and *Mycoplasma* in symptomatic patients, which correlated well with the occurrence of fever and chest pain. Significantly high relative abundance of members of family *Mycobacteriaceae* in the symptomatic COVID-19 group indicates the presence of both pathogenic and non-pathogenic bacteria. Members of the *Mycobacterium* genera are well associated with several pulmonary diseases; for example, *Mycobacterium tuberculosis* is responsible for tuberculosis in humans and is associated with pulmonary infection (30), while *Mycobacterium avium* is highly associated with lung disease (31). Interestingly, several case reports from China, Mexico, India, and Singapore have identified *Mycobacterium tuberculosis* as co-infections in COVID-19 patients and were treated with drugs to combat tuberculosis (32-36). Two meta-analysis studies clearly showed *Mycobacterium tuberculosis* co-infection in patients with COVID-19, which renders the patients vulnerable to severe COVID-related complications and death (37-41). As per WHO report, around 10 million people were affected by tuberculosis in 2020 and eight countries accounted for 2/3rd of patients with tuberculosis infection. Reports show susceptibility for tuberculosis and activation of latent tuberculosis during COVID-19 infections (36-38). Therefore, this study is particularly important for countries such as China, India, Indonesia, South Africa, Bangladesh, Pakistan, Philippines, and Nigeria, where tuberculosis is endemic and possibly aggravates the COVID-19 symptoms and severity. Studies have revealed that patients suffering from tuberculosis are more susceptible for SARS-CoV-2 infection, faster disease progression, and complications associated with COVID-19 disease (36, 38).

Mycoplasma pneumoniae, a member of genus *Mycoplasma,* is responsible for pneumonia and other respiratory infections in humans (42, 43). Co-infections and/or secondary infections by *Mycoplasma pneumoniae* is not uncommon in COVID-19 patients. Several case reports from different parts of the world have shown that *Mycoplasma pneumoniae* co-infections in COVID-19 patients result in disease severity (44-51). Patients with atypical respiratory pathogens such as *Mycoplasma pneumoniae* or *Chlamydia pneumoniae* co-infections have high propensity of developing acute respiratory distress syndrome (ARDS) (45). Both COVID-19 pneumonia and pulmonary pneumonia elicit fever, cough, and shortness of breath, and present similar diagnostic features in chest X-ray and HRCT, because of which they escape logical diagnosis by clinicians. Serological tests can reveal the presence of *Mycoplasma pneumoniae* in COVID-19 patients. The results of our study and those of other reports have proved the association of opportunistic pathogens with alterations in the diversity of the microbial communities in symptomatic and asymptomatic COVID-19 patients. This study establishes a new set of opportunistic pathogens in the context of the NP microbiome in COVID-19-infected patients. Moreover, this study clearly distinguishes between the NP microbial composition of symptomatic and asymptomatic groups using LefSe with AUC-ROC validation. Thus, we believe that SARS-CoV-2 virulence may promote the growth of opportunistic pathogens and may lead to co-infection or secondary infection in COVID-19 patients. Our study also highlights the importance of incorporating routine screening for common opportunistic pathogens in COVID-19 patients in public health policies. Timely identification and medications may regulate the occurrence of co-infections and may reduce disease burden, along with faster recovery from SARS-CoV-2 infections.

Our study has certain limitations. The subject size is limited and a larger cohort would have strengthened our findings. The clinical manifestations are limited, and therefore, we have attempted not to over interpret our data. Future studies should include NP samples from vaccinated, asymptomatic, and hospitalized COVID-19 patients, the detailed associated pathophysiology of whom are known, to understand the role of opportunistic pathogens in COVID-19 disease progression and severity. Furthermore, blood biochemistry and serum metabolite studies will boost conclusions regarding the functional aspects of the NP microbiome.

Conclusions

In this study, we report the abundance of *Mycobacteria* and *Mycoplasma* opportunistic pathogens in the nasopharynx of non-hospitalized COVID-19 patients. Using long-read sequencing of the 16S rRNA variable regions V1-V9, we extended our analysis to species level and found enrichment of several *Mycobacterium* spp. and *Mycoplasma pneumoniae* in both asymptomatic and symptomatic patients with COVID-19. However, their abundance was significantly higher in the symptomatic COVID-19 patients, with strong association with chest pain and fever. Our study adds to the growing list of opportunistic pathogens in patients with COVID-19 and indicates the possibility of co-infections during disease progression. Hence, we believe that it is important to incorporate routine screening for common opportunistic pathogens in SARS-CoV-2-infected patients and modulate treatment modalities to prevent co-infections and secondary infections.

Availability of data and materials

The 16S rRNA gene sequencing data has been deposited in NCBI repository with the Bioproject ID for PRJNA774098 (http://www.ncbi.nlm.nih.gov/bioproject/774098).

References

1. He Y, Wang J, Li F, Shi Y. Main Clinical Features of COVID-19 and Potential Prognostic and Therapeutic Value of the Microbiota in SARS-CoV-2 Infections. Front Microbiol. 2020;11:1302.
2. Khatiwada S, Subedi A. Lung microbiome and coronavirus disease 2019 (COVID-19): Possible link and implications. Hum Microb J. 2020;17:100073.
3. Zou X, Chen K, Zou J, Han P, Hao J, Han Z. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. Front Med. 2020;14(2):185-92.
4. Hou YJ, Okuda K, Edwards CE, Martinez DR, Asakura T, Dinnon KH, 3rd, et al. SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. Cell. 2020;182(2):429-46 e14.
5. Mlcochova P, Kemp SA, Dhar MS, Papa G, Meng B, Ferreira I, et al. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. Nature. 2021.
6. Belkaid Y, Harrison OJ. Homeostatic Immunity and the Microbiota. Immunity. 2017;46(4):562-76.
7. Fazlollahi M, Lee TD, Andrade J, Oguntuyo K, Chun Y, Grishina G, et al. The nasal microbiome in asthma. J Allergy Clin Immunol. 2018;142(3):834-43 e2.

8. de Steenhuijsen Piters WA, Sanders EA, Bogaert D. The role of the local microbial ecosystem in respiratory health and disease. Philos Trans R Soc Lond B Biol Sci. 2015;370(1675).

9. Kumpitsch C, Koskinen K, Schopf V, Moissl-Eichinger C. The microbiome of the upper respiratory tract in health and disease. BMC Biol. 2019;17(1):87.

10. Yildiz S, Mazel-Sanchez B, Kandasamy M, Manicassamy B, Schmolke M. Influenza A virus infection impacts systemic microbiota dynamics and causes quantitative enteric dysbiosis. Microbiome. 2018;6(1):9.

11. Engen PA, Naqib A, Jennings C, Green SJ, Landay A, Keshavarzian A, et al. Nasopharyngeal Microbiota in SARS-CoV-2 Positive and Negative Patients. Biol Proced Online. 2021;23(1):10.

12. De Maio F, Posteraro B, Ponziani FR, Cattani P, Gasbarrini A, Sanguinetti M. Nasopharyngeal Microbiota Profiling of SARS-CoV-2 Infected Patients. Biol Proced Online. 2020;22:18.

13. Ventero MP, Cuadrat RRC, Vidal I, Andrade BGN, Molina-Pardines C, Haro-Moreno JM, et al. Nasopharyngeal Microbial Communities of Patients Infected With SARS-CoV-2 That Developed COVID-19. Front Microbiol. 2021;12:637430.

14. Feehan AK, Rose R, Nolan DJ, Spitz AM, Graubics K, Colwell RR, et al. Nasopharyngeal Microbiome Community Composition and Structure Is Associated with Severity of COVID-19 Disease and Breathing Treatment. Applied Microbiology. 2021;1(2):177-88.

15. Rhoades NS, Pinski AN, Monsibais AN, Jankeel A, Doratt BM, Cinco IR, et al. Acute SARS-CoV-2 infection is associated with an increased abundance of bacterial pathogens, including Pseudomonas aeruginosa in the nose. Cell Rep. 2021;36(9):109637.

16. Nardelli C, Gentile I, Setaro M, Di Domenico C, Pinchera B, Buonomo AR, et al. Nasopharyngeal Microbiome Signature in COVID-19 Positive Patients: Can We Definitively Get a Role to Fusobacterium periodonticum? Front Cell Infect Microbiol. 2021;11:625581.

17. Gupta A, Karyakarte R, Joshi S, Das R, Jani K, Shouche Y, et al. Nasopharyngeal microbiome reveals the prevalence of opportunistic pathogens in SARS-CoV-2 infected individuals and their association with host types. Microbes Infect. 2021:104880.

18. Mostafa HH, Fissel JA, Fanelli B, Bergman Y, Gniazdowski V, Dadlani M, et al. Metagenomic Next-Generation Sequencing of Nasopharyngeal Specimens Collected from Confirmed and Suspect COVID-19 Patients. mBio. 2020;11(6):1-13.

19. Abdoli A, Falahi S, Kenarkoohi A. COVID-19-associated opportunistic infections: a snapshot on the current reports. Clin Exp Med. 2021.

20. Fishman JA. Opportunistic infections--coming to the limits of immunosuppression? Cold Spring Harb Perspect Med. 2013;3(10):a015669.

21. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20(1):257.

22. Varet H, Brillet-Gueguen L, Coppee JY, Dillies MA. SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. PLoS One. 2016;11(6):e0157022.

23. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569(7758):655-62.

568 24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network
569 analysis. BMC Bioinformatics. 2008;9:559.
570 25. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and
571 Manipulating Networks. Proceedings of the International AAAI Conference on Web and
572 Social Media. 2009;3(1):361-2.
573 26. Santacroce L, Charitos IA, Ballini A, Inchingolo F, Luperto P, De Nitto E, et al. The
574 Human Respiratory System and its Microbiome at a Glimpse. Biology (Basel). 2020;9(10).
575 27. Braun T, Halevi S, Hadar R, Efroni G, Glick Saar E, Keller N, et al. SARS-CoV-2 does
576 not have a strong effect on the nasopharyngeal microbial composition. Sci Rep.
577 2021;11(1):8922.
578 28. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, et al.
579 Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis.
580 Nat Commun. 2019;10(1):5029.
581 29. Kaul D, Rathnasinghe R, Ferres M, Tan GS, Barrera A, Pickett BE, et al. Microbiome
582 disturbance and resilience dynamics of the upper respiratory tract during influenza A virus
583 infection. Nat Commun. 2020;11(1):2537.
584 30. Peto HM, Pratt RH, Harrington TA, LoBue PA, Armstrong LR. Epidemiology of
585 extrapulmonary tuberculosis in the United States, 1993-2006. Clin Infect Dis.
586 2009;49(9):1350-7.
587 31. Hwang JA, Kim S, Jo KW, Shim TS. Natural history of Mycobacterium avium complex
588 lung disease in untreated patients with stable course. Eur Respir J. 2017;49(3):1600537.

589 32. Yao Z, Chen J, Wang Q, Liu W, Zhang Q, Nan J, et al. Three Patients with COVID-19 and
590 Pulmonary Tuberculosis, Wuhan, China, January-February 2020. Emerg Infect Dis.
591 2020;26(11):2755-8.
592 33. Sreenath K, Batra P, Vinayaraj EV, Bhatia R, SaiKiran K, Singh V, et al. Coinfections with
593 Other Respiratory Pathogens among Patients with COVID-19. Microbiol Spectr.
594 2021;9(1):e0016321.
595 34. Tham SM, Lim WY, Lee CK, Loh J, Premkumar A, Yan B, et al. Four Patients with
596 COVID-19 and Tuberculosis, Singapore, April-May 2020. Emerg Infect Dis.
597 2020;26(11):2764-6.
598 35. Visca D, Ong CWM, Tiberi S, Centis R, D'Ambrosio L, Chen B, et al. Tuberculosis and
599 COVID-19 interaction: A review of biological, clinical and public health effects.
600 Pulmonology. 2021;27(2):151-65.
601 36. Gerstein S, Khatri A, Roth N, Wallach F. Coronavirus disease 2019 and extra-pulmonary
602 tuberculosis co-infection - A case report and review of literature. J Clin Tuberc Other
603 Mycobact Dis. 2021;22:100213.
604 37. Gao Y, Liu M, Chen Y, Shi S, Geng J, Tian J. Association between tuberculosis and
605 COVID-19 severity and mortality: A rapid systematic review and meta-analysis. J Med
606 Virol. 2021;93(1):194-6.
607 38. Song WM, Zhao JY, Zhang QY, Liu SQ, Zhu XH, An QQ, et al. COVID-19 and
608 Tuberculosis Coinfection: An Overview of Case Reports/Case Series and Meta-Analysis.
609 Front Med (Lausanne). 2021;8:657006.

610 39. Martinez Orozco JA, Sanchez Tinajero A, Becerril Vargas E, Delgado Cueva AI, Resendiz
611 Escobar H, Vazquez Alcocer E, et al. COVID-19 and Tuberculosis Coinfection in a 51-
612 Year-Old Taxi Driver in Mexico City. Am J Case Rep. 2020;21:e927628.
613 40. He G, Wu J, Shi J, Dai J, Gamber M, Jiang X, et al. COVID-19 in tuberculosis patients: A
614 report of three cases. J Med Virol. 2020;92(10):1802-6.
615 41. Tadolini M, Codecasa LR, Garcia-Garcia JM, Blanc FX, Borisov S, Alffenaar JW, et al.
616 Active tuberculosis, sequelae and COVID-19 co-infection: first cohort of 49 cases. Eur
617 Respir J. 2020;56(1).
618 42. Beeton ML, Zhang XS, Uldum SA, Bebear C, Dumke R, Gullsby K, et al. Mycoplasma
619 pneumoniae infections, 11 countries in Europe and Israel, 2011 to 2016. Euro Surveill.
620 2020;25(2).
621 43. Foy HM. Infections caused by Mycoplasma pneumoniae and possible carrier state in
622 different populations of patients. Clin Infect Dis. 1993;17 Suppl 1:S37-46.
623 44. Amin D, McKitish K, Shah PS. Association of mortality and recent Mycoplasma
624 pneumoniae infection in COVID-19 patients. J Med Virol. 2021;93(2):1180-3.
625 45. Chaudhry R, Sreenath K, Batra P, Vinayaraj EV, Rathor N, Saikiran K, et al. Atypical
626 bacterial co-infections among patients with COVID-19: A study from India. J Med Virol.
627 2022;94(1):303-9.
628 46. Chen S, Zhu Q, Xiao Y, Wu C, Jiang Z, Liu L, et al. Clinical and etiological analysis of
629 co-infections and secondary infections in COVID-19 patients: An observational study. Clin
630 Respir J. 2021;15(7):815-25.
631 47. Tang ML, Li YQ, Chen X, Lin H, Jiang ZC, Gu DL, et al. Co-Infection with Common
632 Respiratory Pathogens and SARS-CoV-2 in Patients with COVID-19 Pneumonia and

633 Laboratory Biochemistry Findings: A Retrospective Cross-Sectional Study of 78 Patients
634 from a Single Center in China. Med Sci Monit. 2021;27:e929783.
635 48. Huang AC, Huang CG, Yang CT, Hu HC. Concomitant infection with COVID-19 and
636 Mycoplasma pneumoniae. Biomed J. 2020;43(5):458-61.
637 49. Chaudhry R, Sreenath K, Vinayaraj EV, Sahoo B, Vishnu Narayanan MR, Kiran K, et al.
638 Mycoplasma pneumoniae co-infection with SARS-CoV-2: A case report. Access
639 Microbiol. 2021;3(3):000212.
640 50. Gayam V, Konala VM, Naramala S, Garlapati PR, Merghani MA, Regmi N, et al.
641 Presenting characteristics, comorbidities, and outcomes of patients coinfected with
642 COVID-19 and Mycoplasma pneumoniae in the USA. J Med Virol. 2020;92(10):2181-7.
643 51. Choubey A, Sagar D, Cawley P, Miller K. Retrospective review analysis of COVID-19
644 patients co-infected with Mycoplasma pneumoniae. Lung India.
645 2021;38(Supplement):S22-S6.
646 52. Operario DJ, Pholwat S, Koeppel AF, Prorock A, Bao Y, Sol-Church K, et al.
647 Mycobacterium avium Complex Diversity within Lung Disease, as Revealed by Whole-
648 Genome Sequencing. Am J Respir Crit Care Med. 2019;200(3):393-6.
649 53. Martins AB, Matos ED, Lemos AC. Infection with the Mycobacterium avium complex in
650 patients without predisposing conditions: a case report and literature review. Braz J Infect
651 Dis. 2005;9(2):173-9.
652 54. Shah NM, Davidson JA, Anderson LF, Lalor MK, Kim J, Thomas HL, et al. Pulmonary
653 Mycobacterium avium-intracellulare is the main driver of the rise in non-tuberculous
654 mycobacteria incidence in England, Wales and Northern Ireland, 2007-2012. BMC Infect
655 Dis. 2016;16:195.

55. Daley CL, Winthrop KL. Mycobacterium avium Complex: Addressing Gaps in Diagnosis and Management. J Infect Dis. 2020;222(Suppl 4):S199-S211.

56. Yang JJ, Mohallem DF, Cardoso TA, Lima Junior CL, Tebcherani AJ, Vidigal Mdo R. Case for diagnosis. An Bras Dermatol. 2014;89(5):837-8.

57. Cerqueira S, Deps PD, Cunha DV, Bezerra NVF, Barroso DH, Pinheiro ABS, et al. The influence of leprosy-related clinical and epidemiological variables in the occurrence and severity of COVID-19: A prospective real-world cohort study. PLoS Negl Trop Dis. 2021;15(7):e0009635.

58. Zhang M, Feng M, He JQ. Disseminated Mycobacterium kansasii infection with cutaneous lesions in an immunocompetent patient. Int J Infect Dis. 2017;62:59-63.

59. Johnston JC, Chiang L, Elwood K. Mycobacterium kansasii. Microbiol Spectr. 2017;5(1).

60. van Ingen J, Boeree MJ, Dekhuijzen PN, van Soolingen D. Clinical relevance of Mycobacterium simiae in pulmonary samples. Eur Respir J. 2008;31(1):106-9.

61. Hamieh A, Tayyar R, Tabaja H, S ELZ, Bou Khalil P, Kara N, et al. Emergence of Mycobacterium simiae: A retrospective study from a tertiary care center in Lebanon. PLoS One. 2018;13(4):e0195390.

62. Okano Y, Shinohara T, Imanishi S, Takahashi N, Naito N, Taoka T, et al. Miliary pulmonary nodules due to Mycobacterium xenopi in a steroid-induced immunocompromised patient successfully treated with chemotherapy: a case report. BMC Pulm Med. 2016;16(1):92.

63. Gochi M, Kaneko Y, Seki A, Saitou Z, Samejima T, Seki Y, et al. [Mycobacterium xenopi lung infection in a patient with multiple lung cysts that responded well to chemotherapy: a case report]. Kekkaku. 2012;87(11):733-6.

64. Research Committee of the British Thoracic S. First randomised trial of treatments for pulmonary disease caused by M avium intracellulare, M malmoense, and M xenopi in HIV negative patients: rifampicin, ethambutol and isoniazid versus rifampicin and ethambutol. Thorax. 2001;56(3):167-72.

65. Youssef D, Shams WE, Elshenawy Y, El-Abbassi A, Moorman JP. Pulmonary infection with caseating mediastinal lymphadenitis caused by Mycobacterium gordonae. Int J Mycobacteriol. 2014;3(3):220-3.

66. Mazumder SA, Hicks A, Norwood J. Mycobacterium gordonae pulmonary infection in an immunocompetent adult. N Am J Med Sci. 2010;2(4):205-7.

67. Grech M, Carter R, Thomson R. Clinical significance of Mycobacterium asiaticum isolates in Queensland, Australia. J Clin Microbiol. 2010;48(1):162-7.

68. Arttawejkul P, Kongpolprom N. A case of pulmonary infection caused by Mycobacterium asiaticum: Difficulties on diagnostic and therapeutic approaches. Respir Med Case Rep. 2018;24:150-2.

69. van Ingen J, Boeree M, Janssen M, Ullmann E, de Lange W, de Haas P, et al. Pulmonary Mycobacterium szulgai infection and treatment in a patient receiving anti-tumor necrosis factor therapy. Nat Clin Pract Rheumatol. 2007;3(7):414-9.

70. Maloney JM, Gregg CR, Stephens DS, Manian FA, Rimland D. Infections caused by Mycobacterium szulgai in humans. Rev Infect Dis. 1987;9(6):1120-6.

71. Anagonou EG, Johnson RC, Barogui YT, Sopoh GE, Ayelo GA, Wadagni AC, et al. Decrease in Mycobacterium ulcerans disease (Buruli ulcer) in the Lalo District of Benin (West Africa). BMC Infect Dis. 2019;19(1):247.

72. Sizaire V, Nackers F, Comte E, Portaels F. Mycobacterium ulcerans infection: control, diagnosis, and treatment. Lancet Infect Dis. 2006;6(5):288-96.

Acknowledgments

We acknowledge Biorepository, BSL-3, and BSL-2 laboratories, qPCR, and DNA-sequencing institutional central core facilities. R.M., S.M., and K.C.M received their fellowships from Ramalingaswami, ILS Flagship, and SERB core research grant, respectively. We would like to thank Dr. Moumita Biswas for scientific input and professional scientific editing of this manuscript. We thank all the volunteers who provided samples for research purposes.

Funding

The funding for this research was obtained from institute's core funding from the Department of Biotechnology (DBT), Government of India. This work was also supported by the ILS flagship project (BT/ILS/Flagship/2019) from DBT, Ramalingaswami Re-entry fellowship (BT/RLF/Re-entry/25/2015), and SERB core research grant (CRG/2018/002052).

Author information

Punit Prasad[1*$], Soumendu Mahapatra[1*], and Rasmita Mishra[1*], First co-authors

Ajay Parida[1$] and Punit Prasad[1*$] are co-corresponding authors

Krushna Chandra Murmu[1], Shifu Aggarwal[1], Manisha Sethi[1], Priyanka Mohapatra[1], Arup Ghosh[1], Rina Yadav[1], Hiren Dodia[1], Shamima Azma Ansari[1], Saikat De[1], Deepak Singh[1], Amol Suryawanshi[1], Rupesh Dash[1], Shantibhushan Senapati[1], Tushar K. Beuria[1], Soma Chattopadhyay[1], Gulam Hussain Syed[1], Rajeeb Swain[1], Sunil K. Raghav[1], Ajay Parida[1$]

[1]Institute of Life Sciences, Bhubaneswar, Odisha, India.

Contributions

P.P. and A.P. conceptualized the study and secured funding. P.P and S.M. initiated the work, directed overall workflow, interpreted data, and troubleshooted the experiments. R.M. did most of the bioinformatics analysis and S.M., K.C.M. and A.G. helped in bioinformatics analysis and troubleshooting. S.M., S.A., M.S., P.M., R.Y., H.D., S.A.A., S.D., and D.S., helped with the preprocessing of the samples in Biosafety level 3 (BSL3) facility and nucleic acid extractions. S.M., S.A., M.S., P.M., and R.Y. were involved in amplicon library preparations. R.S. provided samples from the Biorepository. A.S., R.D., S.S., T.K.B., S.C., G.H.S., R.S., S.K.R., P.P., and A.P. coordinated with COVID-19 sampling and testing at BSL3. P.P., S.M., and R.M wrote the manuscript.

Ethics declarations

Ethics approval and consent to participate

Ethical permission for nasopharyngeal microbiome study and the biorepository was obtained from the Institutional Ethical Committee (IEC)/Institutional Review Board (IRB) of the Institute

of Life Sciences [(102/HEC/2020) and (100/HEC/2020)]. Approval was also obtained from the Institutional Biosafety Committee (IBSC) (V-122-MISC/2007-08/01/2/2.1) for this study and the biorepository (V-122-MISC/2007-08/01) and from the Review Committee on Genetic Manipulations (RCGM) under Department of Biotechnology, Ministry of Science and Technology.

Consent for publication

All authors have given consent for publication of this manuscript.

Competing interest

The authors declare no competing commercial or financial interests in relation to this work.

Figure Legends:

Figure 1: Schema of nasopharyngeal sample processing, 16S sequencing, and OTU-based sample distribution. (A) Flow chart showing nasopharyngeal sample processing for DNA extraction, amplicon library preparation, Oxford Nanopore sequencing, and bioinformatics analysis pipeline. (B) Pie chart showing the nasopharyngeal samples (controls, symptomatic, and asymptomatic) used in this study. (C) t-SNE plot showing the OTU-based sample distribution and ordination points for control, symptomatic, and asymptomatic samples.

Figure 2: Alpha/beta diversities and dysbiosis index in COVID-19-positive and negative nasopharyngeal samples. (A-B) Alpha diversity index (Shannon/Simpson) between control and COVID-19-infected samples (pairwise Wilcoxon rank-sum test $p = \leq 0.05$). (C-D) Same analysis

as above where the COVID-19-infected samples are classified as asymptomatic and symptomatic compared to the control group. (E-F) Linear regression model showing the association between total OTU count and Shannon/Simpson diversity index for each sample; the shaded grey region represents 95% confidence intervals of two groups, symptomatic and asymptomatic, with correlation (Spearman) regression line [Shannon: R = -0.35 (asymptomatic), R = -0.54 (symptomatic) and Simpson: R = -0.58 (asymptomatic), R = -0.77 (symptomatic)]. (G) Principal coordinate analysis (PCoA) showing beta diversity in asymptomatic, symptomatic, and control sample groups based on unifrac (weighted/unweighted) distance (p = 0.001, PERMANOVA). (H) Violin plot showing dysbiosis indexes of samples from control, asymptomatic, and symptomatic participants (pairwise Wilcoxon rank-sum test p = ≤0.05).

Figure 3: Taxonomic classification of bacterial communities using NMDS, correlation, and CCA. (A) NMDS ordination of Bray-Curtis distance matrix based on all samples and bacterial communities of each taxonomy level (phylum, order, family, and genus) (ANOSIM p = <0.05). (B) The density plot representing the Spearman correlation coefficient at each taxonomy level (phylum, order, family, and genus); dotted line indicates the mean value of each sample group (Kolmogorov–Smirnov (KS) Test p ≤ 0.05). (C) Heatmap of Spearman correlation for genus level with sample correlation (lower) and OTU correlation (upper). Five clusters (C1, C2, C3, C4, and C5) were generated using unsupervised hierarchical clustering from the OTU correlation plot. (D) Sample-wise OTU density plot for each cluster (C1, C2, C3, C4, and C5) showing relative abundance. (E-F) CCA plot of microbial community composition for each cluster and bar plot representing cumulative variation percentage from two components [C1 (92.4%), C2 (80.5%), C3

784 (12.9%), C4 (94.8%), and C5 (39.4%)]. Dotted line shows 30% variance cut-off for downstream
785 analysis.

786

787 **Figure 4: Linear discriminant analysis effect size (LefSe) analysis revealed distinct genus-**
788 **level OTUs in control, asymptomatic, and symptomatic participants.** (A) Heatmap showing
789 genus level OTU (n = 181) abundance distribution from four clusters (C1, C2, C4, and C5)
790 identified from CCA analysis in control, asymptomatic and symptomatic samples. The OTUs
791 marked on either side of the heatmap were obtained from one-against-all and all-against-all
792 comparison in LDA analysis (B) The cladogram shows the output of the LEfSe (LDA score >2.0),
793 which identified taxonomic differences between sample groups. Each circle represents a bacterial
794 taxon, and each ring of taxonomy level starting with kingdom in the innermost circle is followed
795 by phylum, class, order, family, and genus in the outermost circle. The different color intensities
796 indicate the different taxonomy levels, and the diameter of each circle is proportional to the taxon's
797 abundance and correlates with the LDA score. (C) The histogram of the LDA scores (score >2.0
798 and all-against-all) was computed for differentially abundant taxa between sample groups. The
799 effect size of specific taxa in the particular group at the genus level. (D) Histogram of the all LefSe-
800 specific taxa (*Mycoplasma*, *Streptosporangium*, *Citrobacter*, *Acidocella*, *Mycolicibacterium*,
801 *Mycolicibacillus*, *Mycobacterium*, *Mycobacteroides*, *Orientia*, *Gallibacterium*,
802 *Cellulosimicrobium,* and *Oerskovia*) showing relative abundance across sample groups. Solid and
803 dotted lines show median and mean relative abundance respectively. (E) Weighted correlation
804 network analysis (WGCNA) was used for network construction and plotted using Gephi. Each
805 node of the network represents the individual bacterial genera with their respective abundance size,
806 and the edges represent correlation strength with edge weight by thickness. The pie chart within

each node represents abundance for each genus. The dotted line shows two distinct modules (control and infected) created in the network analysis. (F) Violin plot showing the dysbiosis indices of LefSe sample groups (pairwise Wilcoxon rank-sum test p-value < 0.05).

Figure 5: Area under the curve-receiver operating characteristic (AUC-ROC) validation and correlation of genera with the symptoms of COVID-19 subjects. (A) ROC curve for LDA classified symptomatic and asymptomatic group. AUC of 0.80 with 95% confidence interval (CI). (B) Correlation between bacteria at genus level and clinical symptoms of patients. (C-D) ROC curve for chest pain and fever in the symptomatic and asymptomatic groups. The AUCs were 0.904 (chest pain) and 0.793 (fever) with 95% confidence interval (CI).

Figure 6: Relative abundance and species-specific OTUs reveal *Mycobacterium* spp. and *Mycoplasma* as key opportunistic pathogens. (A) Heatmap showing the taxonomic distribution of 54 species obtained from 12 genus based on LDA scores. (B-D) Boxplots showing relative abundance of opportunistic pathogens (Statistical parameters (Wilcoxon Rank test): ns: p = >0.05, * p< = 0.05, ** p < = 0.01, *** p < = 0.001, **** p < = 0.0001) in C vs. IA, C vs. IS, and IA vs. IS groups. (E) Spearman correlation showing relationship between opportunistic pathogenic bacteria and patient's clinical symptoms.

Table 1: Details of samples included in this study

	Control (n=12)	**Asymptomatic (n=25)**	**Symptomatic (n=21)**
Sex			
Male	5 (41.66%)	18 (72%)	19 (90.47%)
Female	7 (58.33%)	7 (28%)	2 (9.52%)

Age (years)	31 (median)	26 (median)	32 (median)
Symptoms			
Dry Cough	NA	NA	7 (33.3%)
Fever	NA	NA	17 (80.95%)
Tiredness	NA	NA	8 (38.09%)
Sore throat	NA	NA	11(52.38%)
Body pain	NA	NA	13(61.9%)
Chest pain	NA	NA	9(42.85%)
Fever + Body pain	NA	NA	4(19.04%)
Fever + multiple symptoms*	NA	NA	16(76.19%)
Fever + chest pain	NA	NA	8(38.09%)
Loss of smell/taste + multiple symptoms* without fever	NA	NA	2(9.52%)

825 *Multiple symptoms refer to having more than one symptoms from symptoms list.

826 **Table 2: Phylum based on relative abundance and their respective values**

Phylum	**Mean**	**1st Quartile**	**Median**	**3rd Quartile**	**p_value**	**BH_FDR**
Proteobacteria	1.51E-01	3.79E-02	4.53E-02	7.29E-02	6.56E-07	6.56E-07
Fusobacteria	2.25E-03	1.31E-03	1.72E-03	2.71E-03	9.00E-07	9.00E-07
Actinobacteria	6.83E-01	7.09E-01	7.62E-01	7.99E-01	9.96E-07	9.96E-07
Bacteroidetes	8.77E-03	5.71E-03	6.66E-03	1.00E-02	1.48E-06	1.48E-06
Tenericutes	9.37E-04	1.35E-04	5.60E-04	1.21E-03	2.21E-06	2.21E-06
Chloroflexi	2.25E-03	6.47E-04	9.84E-04	1.46E-03	6.54E-06	6.54E-06
Chlamydiae	5.93E-04	2.82E-04	4.98E-04	7.76E-04	7.27E-06	7.27E-06
Fibrobacteres	5.65E-04	2.86E-04	4.05E-04	7.10E-04	4.77E-05	4.77E-05
Thermodesulfobacteria	2.62E-02	2.09E-02	2.84E-02	3.24E-02	8.13E-04	8.13E-04
Aquificae	2.31E-04	1.19E-04	2.30E-04	2.93E-04	1.59E-03	1.59E-03
Firmicutes	1.23E-01	1.19E-01	1.28E-01	1.44E-01	4.65E-02	4.65E-02

Chlorobi	1.02E-03	6.81E-04	9.32E-04	1.27E-03	8.58E-01	8.58E-01

827

828 **Table 3: Top 30 family based on relative abundance and their respective values**

Family	**Mean**	**1st quartile**	**Median**	**3rd quartile**	**p_value**	**BH_FDR**
Acetobacteraceae	5.08E-02	5.00E-02	5.12E-02	5.28E-02	9.14E-07	9.14E-07
Actinomycetaceae	2.23E-02	2.10E-02	2.18E-02	2.32E-02	8.18E-03	8.18E-03
Aeromonadaceae	2.96E-02	2.54E-02	3.07E-02	3.38E-02	8.18E-03	8.18E-03
Alcaligenaceae	8.14E-02	8.50E-02	8.71E-02	9.43E-02	1.41E-06	1.41E-06
Bifidobacteriaceae	7.69E-02	4.59E-02	6.52E-02	1.04E-01	2.99E-03	2.99E-03
Brevibacteriaceae	4.37E-02	2.72E-02	3.82E-02	5.98E-02	6.35E-04	6.35E-04
Cellulomonadaceae	6.75E-02	4.60E-02	6.29E-02	8.52E-02	8.64E-05	8.64E-05
Chromobacteriaceae	2.30E-02	2.27E-02	2.30E-02	2.32E-02	9.37E-07	9.37E-07
Corynebacteriaceae	5.49E-02	4.18E-02	5.05E-02	6.52E-02	2.36E-05	2.36E-05
Enterobacteriaceae	1.26E-01	3.61E-02	1.65E-01	2.17E-01	1.73E-06	1.73E-06
Erwiniaceae	3.19E-02	2.97E-02	3.10E-02	3.46E-02	2.01E-06	2.01E-06
Erysipelotrichaceae	2.26E-02	2.21E-02	2.26E-02	2.30E-02	1.69E-06	1.69E-06
Eubacteriaceae	2.99E-02	2.58E-02	2.84E-02	3.28E-02	1.85E-06	1.85E-06
Lactobacillaceae	2.12E-02	2.05E-02	2.10E-02	2.18E-02	5.39E-03	5.39E-03
Micrococcaceae	3.52E-02	2.65E-02	3.37E-02	4.36E-02	3.49E-03	3.49E-03
Micromonosporaceae	2.41E-02	2.25E-02	2.43E-02	2.53E-02	2.65E-06	2.65E-06
Morganellaceae	5.01E-02	4.51E-02	5.30E-02	5.93E-02	1.90E-05	1.90E-05
Mycobacteriaceae	2.19E-01	1.85E-01	2.42E-01	2.69E-01	7.93E-07	7.93E-07
Neisseriaceae	5.24E-02	5.08E-02	5.23E-02	5.43E-02	1.29E-06	1.29E-06
Nocardiaceae	3.15E-02	2.46E-02	3.41E-02	3.63E-02	8.90E-07	8.90E-07
Pasteurellaceae	4.73E-02	3.34E-02	5.14E-02	5.75E-02	5.89E-05	5.89E-05
Pectobacteriaceae	2.17E-02	2.05E-02	2.15E-02	2.27E-02	2.09E-06	2.09E-06

Peptostreptococcaceae	3.01E-02	2.56E-02	2.82E-02	3.23E-02	1.89E-06	1.89E-06
Propionibacteriaceae	1.17E-01	9.00E-02	1.20E-01	1.33E-01	8.90E-07	8.90E-07
Pseudonocardiaceae	2.94E-02	2.70E-02	2.98E-02	3.27E-02	1.44E-06	1.44E-06

Table 4: Mean and median value of density plots.

Taxonomic rank	Statistic value	Control	Asymptomatic	Symptomatic
Phylum	Mean	8.96E-01	8.68E-01	8.90E-01
	Median	9.30E-01	8.95E-01	9.16E-01
Order	Mean	8.58E-01	7.70E-01	8.36E-01
	Median	8.78E-01	8.21E-01	8.93E-01
Family	Mean	8.22E-01	6.86E-01	7.60E-01
	Median	8.63E-01	7.46E-01	8.06E-01
Genus	Mean	7.95E-01	5.65E-01	6.51E-01
	Median	6.39E-01	8.33E-01	7.01E-01

Table 5: Result of Kolmogorov–Smirnov (KS) test between the densities of each taxonomic rank.

Taxonomic rank	Control vs Asymptomatic	Control vs Symptomatic	Asymptomatic vs Symptomatic
Phylum	D = 1.88e-01	D = 1.07e-01	D = 1.01e-01
	p-value = 3.03e-02	p-value = 0.4834	p-value = 9.78e-04
Order	D = 3.18e-01	D = 9.97e-02	D = 2.30e-01
	p-value = 1.237e-05	p-value = 0.5706	p-value < 2.2e-16
Family	D = 4.14e-01	D = 2.93e-01	D = 2.25e-01
	p-value = 2.706e-09	p-value = 4.75e-05	p-value < 2.2e-16
Genus	D = 5.94e-01	D = 5.06e-01	D = 2.28e-01
	p-value < 2.2e-16	p-value = 3.308e-14	p-value < 2.2e-16

Table 6: Linear discriminate analysis (LDA) score for all-against-all analysis

Genus	Highest mean among all the classes	Samples	LDA score (log 10)	pvalue
Oerskovia	3.83	Asymptomatic	3.44	1.96E-02
Cellulosimicrobium	3.85	Asymptomatic	3.5	1.73E-02
Gallibacterium	3.84	Control	3.36	8.46E-04
Orientia	3.88	Control	3.46	3.07E-04
Acidocella	4.07	Control	3.81	4.68E-07
Citrobacter	4.08	Control	3.89	2.82E-07
Mycobacteroides	3.88	Symptomatic	3.61	3.19E-07
Mycolicibacillus	3.98	Symptomatic	3.67	3.19E-07
Mycolicibacterium	3.93	Symptomatic	3.71	2.55E-07
Mycobacterium	3.9	Symptomatic	3.79	2.55E-07
Streptosporangium	4.16	Symptomatic	3.83	1.65E-04
Mycoplasma	4.26	Symptomatic	3.95	1.64E-06

835 **Table 7: WGCNA network data table**

ID	Genus	OTUs	Module	absolute abundance	Degree	Page ranks	Eigen centrality	Modularity class	Clustering	Triangles
1	*Acidocella*	525	M1	1362	5	0.0673	0.4963	0	1	10
2	*Citrobacter*	546	M1	3463	5	0.0673	0.4963	0	1	10
3	*Gallibacterium*	750	M1	285	5	0.0673	0.4963	0	1	10
4	*Orientia*	784	M1	263	5	0.0673	0.4963	0	1	10
5	*Cellulosimicrobium*	1710	M2	11832	10	0.1202	1	1	0.5555	25
6	*Oerskovia*	1713	M2	13210	10	0.1202	1	1	0.5555	25
7	*Mycobacterium*	1763	M2	139017	7	0.0852	0.8208	1	0.9047	19
8	*Mycolicibacterium*	1766	M2	81430	7	0.0852	0.8208	1	0.9047	19

9	*Mycobacteroides*	1774	M2	7172	7	0.0852	0.8208	1	0.9047	19
10	*Mycolicibacillus*	1798	M2	2749	7	0.0852	0.8208	1	0.9047	19
11	*Streptosporangium*	2002	M2	163	5	0.0642	0.5732	1	1	10
12	*Mycoplasma*	2100	M2	475	7	0.0852	0.8208	1	0.9047	19

836

837 **Table 8: List of opportunistic pathogenic bacteria**

Sl. No.	Species	Clinical manifestation	References
1	*Mycoplasma pneumoniae*	Mycoplasma pneumonia, a form of atypical bacterial pneumonia.	(42, 44, 49)
2	*Mycobacterium tuberculosis*	Tuberculosis (TB)	(35, 38-41)
3	*Mycobacterium avium*	MAC lung disease	(52, 53)
4	*Mycobacterium intracellulare*	MAC lung disease	(54, 55)
5	*Mycobacterium leprae*	Hansen's disease (Leprosy)	(56, 57)
6	*Mycobacterium Kansasii*	Chronic pulmonary infection	(58, 59)
7	*Mycobacterium simiae*	Pulmonary infection	(60, 61)
8	*Mycobacterium xenopi*	NTM lung disease	(62-64)
9	*Mycobacterium gordonae*	Pulmonary infection	(65, 66)
10	*Mycobacterium asiaticum*	Pulmonary and extrapulmonary disease	(67, 68)
11	*Mycobacterium szulgai*	Pulmonary infection	(69, 70)
12	*Mycobacterium ulcerans*	Buruli ulcer	(71, 72)

838

Figures

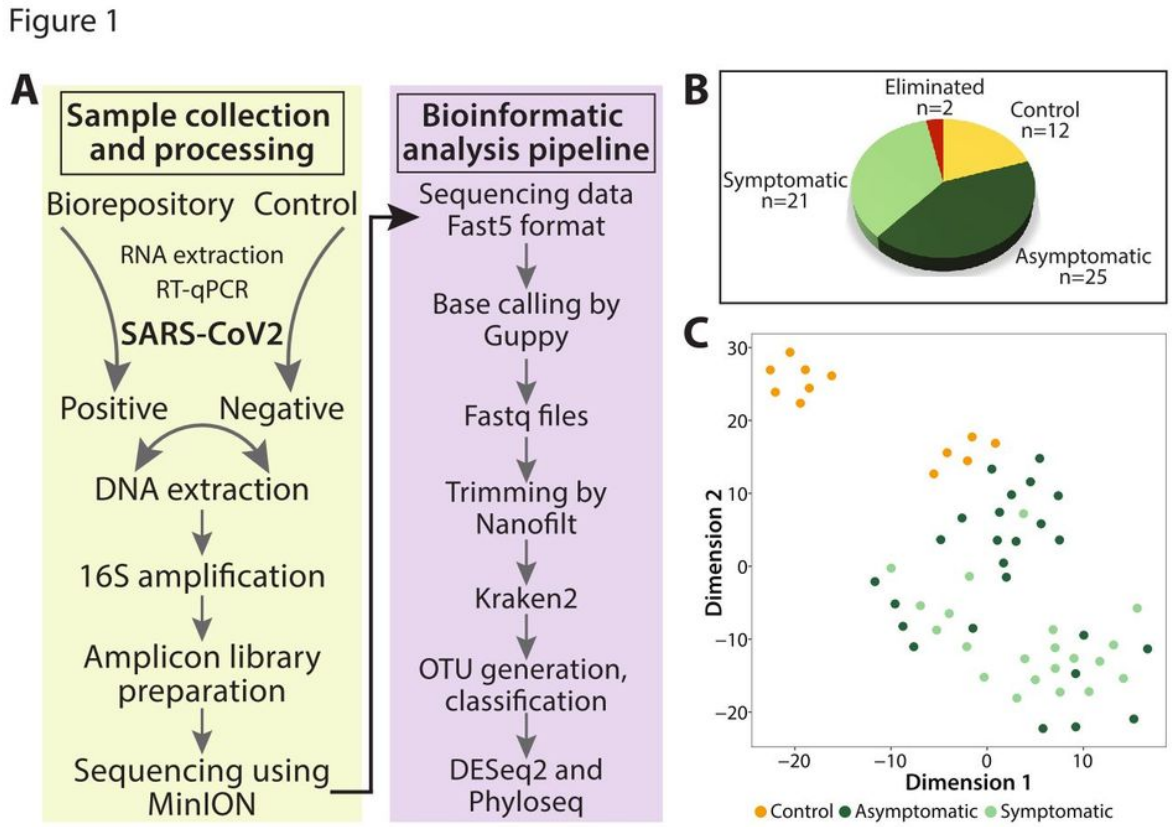

Figure 1

Schema of nasopharyngeal sample processing, 16S sequencing, and OTU-based sample distribution. (A) Flow chart showing nasopharyngeal sample processing for DNA extraction, amplicon library preparation, Oxford Nanopore sequencing, and bioinformatics analysis pipeline. (B) Pie chart showing the nasopharyngeal samples (controls, symptomatic, and asymptomatic) used in this study. (C) t-SNE plot showing the OTU-based sample distribution and ordination points for control, symptomatic, and asymptomatic samples.

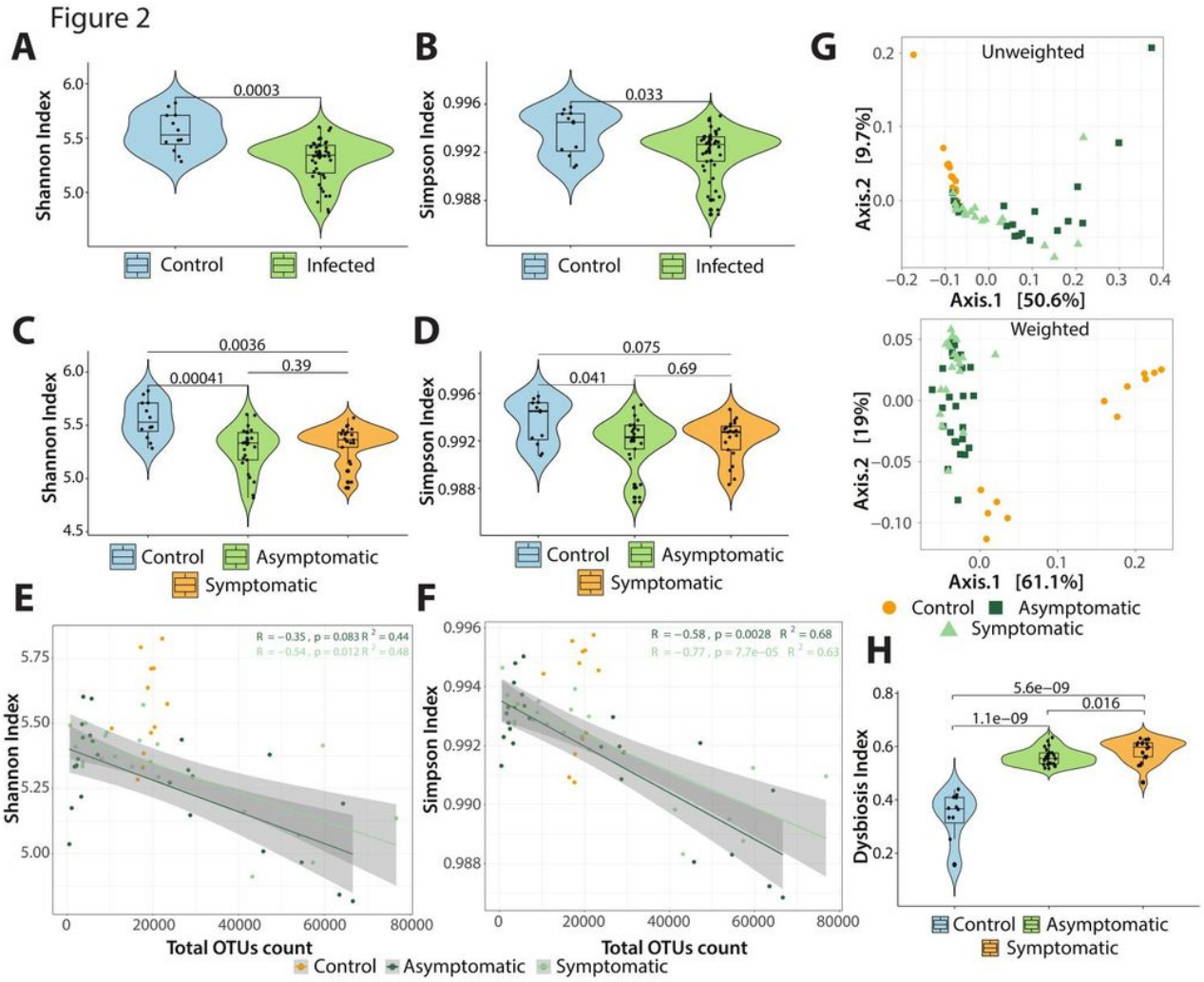

Figure 2

Alpha/beta diversities and dysbiosis index in COVID-19-positive and negative 760 nasopharyngeal samples. (A-B) Alpha diversity index (Shannon/Simpson) between control and COVID-19-infected samples (pairwise Wilcoxon rank-sum test p = ≤0.05). (C-D) Same analysis 35 as above where the COVID-19-infected samples are classified as asymptomatic and symptomatic compared to the control group. (E-F) Linear regression model showing the association between total OTU count and Shannon/Simpson diversity index for each sample; the shaded grey region represents 95% confidence intervals of two groups, symptomatic and asymptomatic, with correlation (Spearman) regression line [Shannon: R = -0.35 (asymptomatic), R = -0.54 767 (symptomatic) and Simpson: R = -0.58 (asymptomatic), R = -0.77 (symptomatic)]. (G) Principal coordinate analysis (PCoA) showing beta diversity in asymptomatic, symptomatic, and control sample groups based on unifrac (weighted/unweighted) distance (p = 0.001, PERMANOVA). (H) Violin plot showing dysbiosis indexes of samples from control, asymptomatic, and symptomatic participants (pairwise Wilcoxon rank-sum test p = ≤0.05).

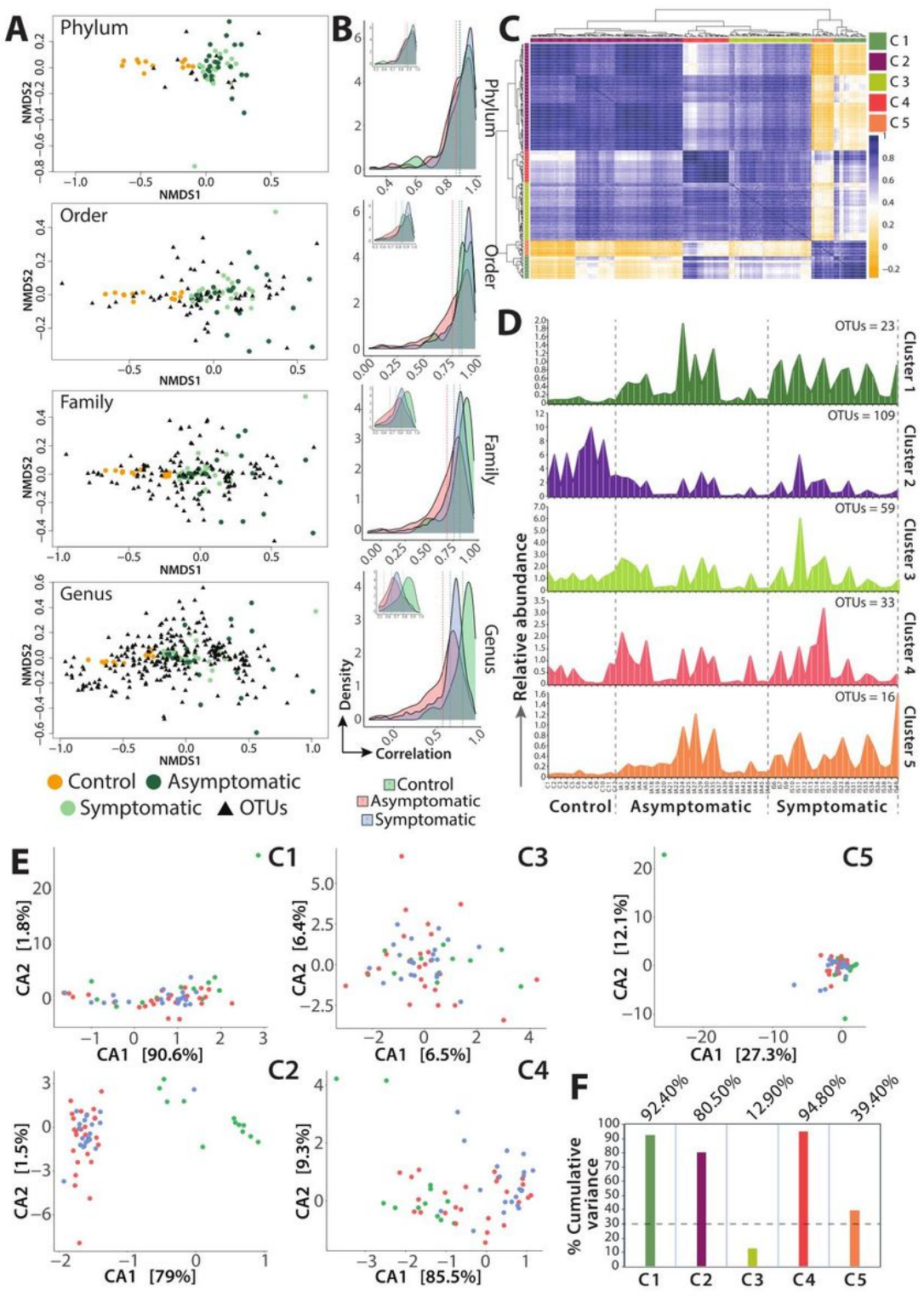

Figure 3

Taxonomic classification of bacterial communities using NMDS, correlation, andCCA. (A) NMDS ordination of Bray-Curtis distance matrix based on all samples and bacterial communities of each taxonomy level (phylum, order, family, and genus) (ANOSIM p = <0.05). (B) The density plot representing the Spearman correlation coefficient at each taxonomy level (phylum, order, family, and genus); dotted line indicates the mean value of each sample group (Kolmogorov–Smirnov (KS) Test $p \leq 0.05$). (C)

Heatmap of Spearman correlation for genus level with sample correlation (lower) and OTU correlation (upper). Five clusters (C1, C2, C3, C4, and C5) were generated using unsupervised hierarchical clustering from the OTU correlation plot. (D) Sample-wise OTU density plot for each cluster (C1, C2, C3, C4, and C5) showing relative abundance. (E-F) CCA plot of microbial community composition for each cluster and bar plot representing cumulative variation percentage from two components [C1 (92.4%), C2 (80.5%), C3 36 784 (12.9%), C4 (94.8%), and C5 (39.4%)]. Dotted line shows 30% variance cut-off for downstream analysis.

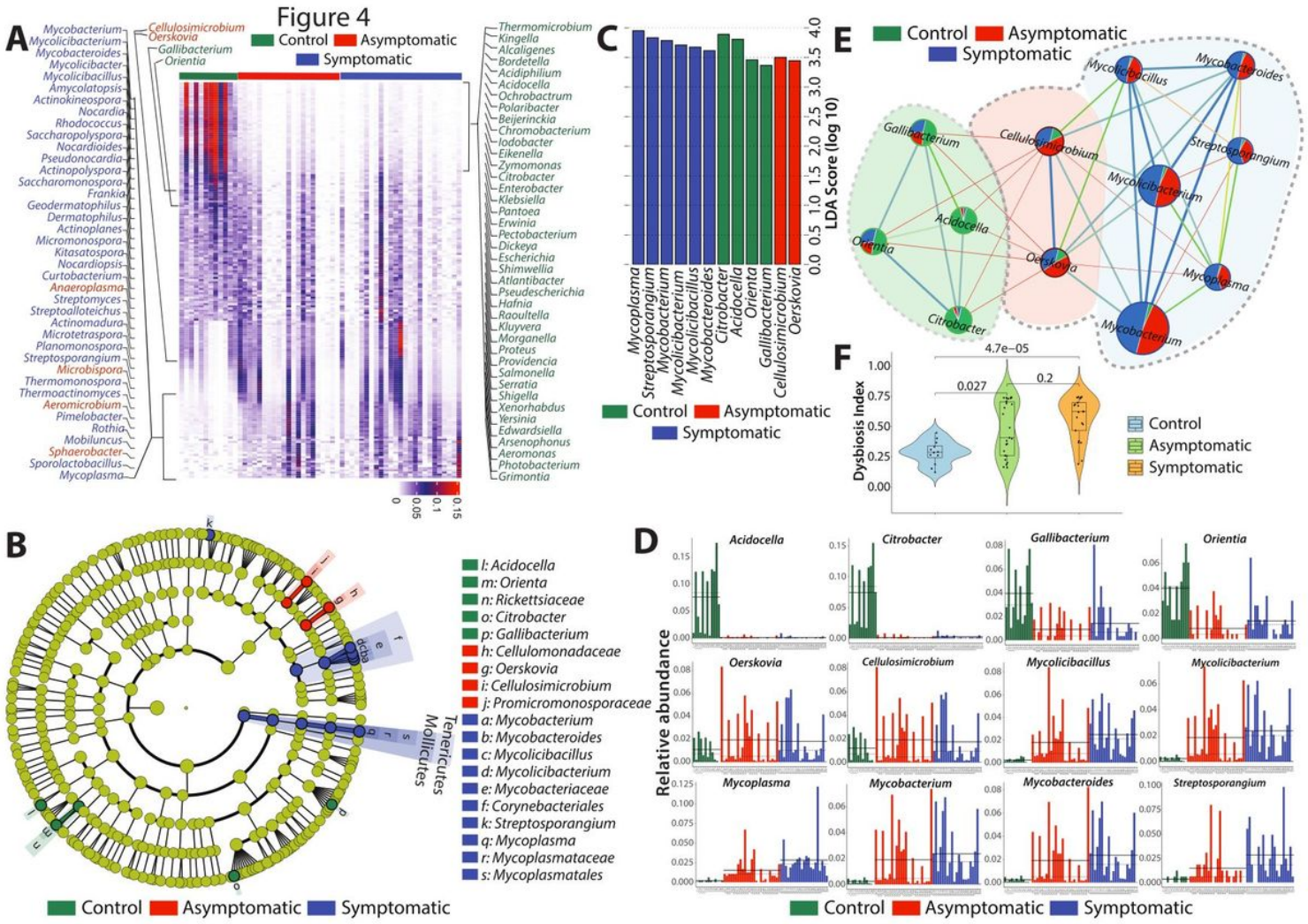

Figure 4

Linear discriminant analysis effect size (LefSe) analysis revealed distinct genus level OTUs in control, asymptomatic, and symptomatic participants. (A) Heatmap showing genus level OTU (n = 181) abundance distribution from four clusters (C1, C2, C4, and C5) identified from CCA analysis in control, asymptomatic and symptomatic samples. The OTUs marked on either side of the heatmap were obtained from one-against-all and all-against-all comparison in LDA analysis (B) The cladogram shows the output of the LEfSe (LDA score >2.0), which identified taxonomic differences between sample groups. Each circle represents a bacterial taxon, and each ring of taxonomy level starting with kingdom in the innermost circle is followed by phylum, class, order, family, and genus in the outermost circle. The

different color intensities indicate the different taxonomy levels, and the diameter of each circle is proportional to the taxon's abundance and correlates with the LDA score. (C) The histogram of the LDA scores (score >2.0 and all-against-all) was computed for differentially abundant taxa between sample groups. The effect size of specific taxa in the particular group at the genus level. (D) Histogram of the all LefSe specific taxa (Mycoplasma, Streptosporangium, Citrobacter, Acidocella, Mycolicibacterium, Mycolicibacillus, Mycobacterium, Mycobacteroides, Orientia, Gallibacterium, Cellulosimicrobium, and Oerskovia) showing relative abundance across sample groups. Solid and dotted lines show median and mean relative abundance respectively. (E) Weighted correlation network analysis (WGCNA) was used for network construction and plotted using Gephi. Each node of the network represents the individual bacterial genera with their respective abundance size, and the edges represent correlation strength with edge weight by thickness. The pie chart within 37 each node represents abundance for each genus. The dotted line shows two distinct modules (control and infected) created in the network analysis. (F) Violin plot showing the dysbiosis indices of LefSe sample groups (pairwise Wilcoxon rank-sum test p-value < 0.05).

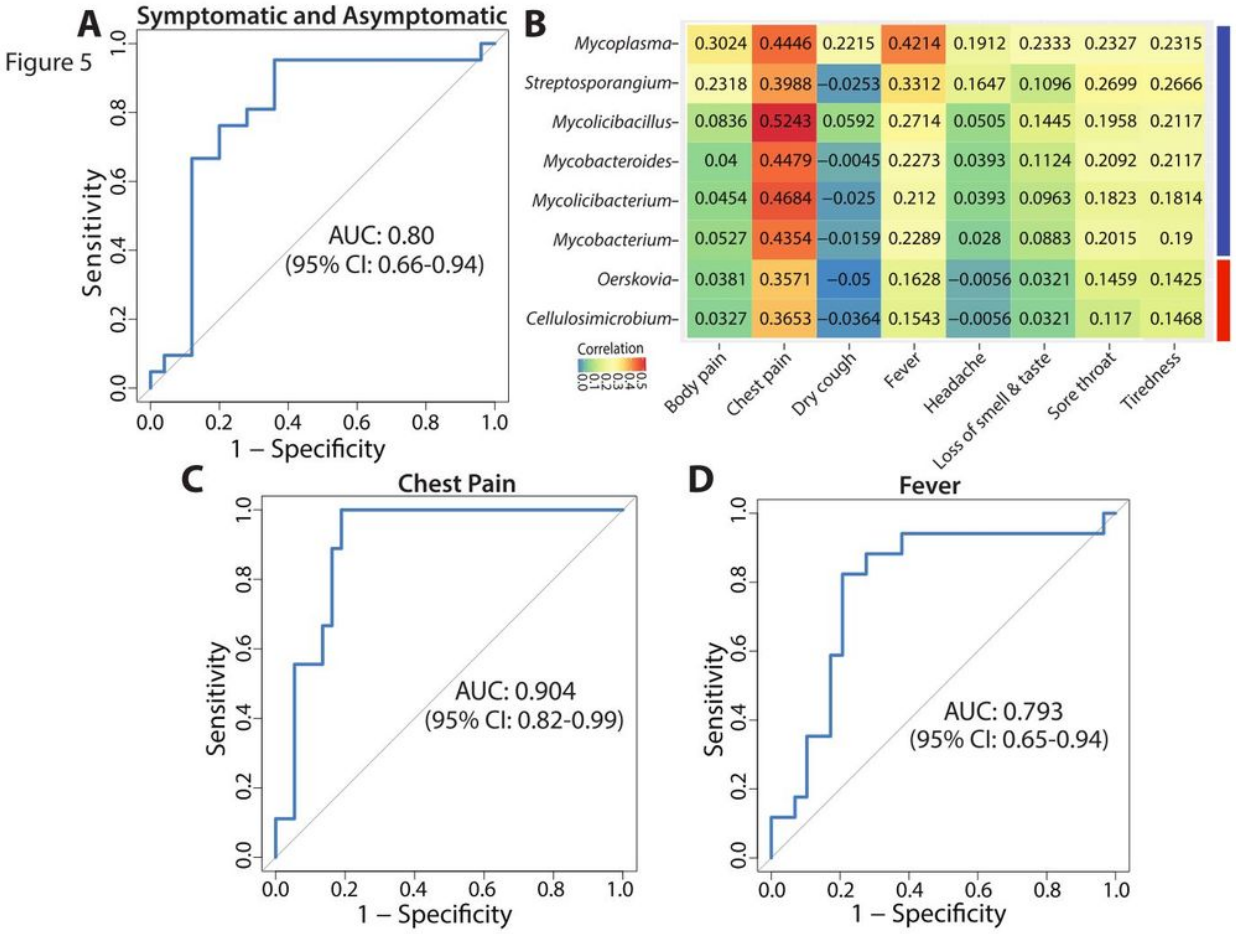

Figure 5

Area under the curve-receiver operating characteristic (AUC-ROC) validation and correlation of genera with the symptoms of COVID-19 subjects. (A) ROC curve for LDA classified symptomatic and asymptomatic group. AUC of 0.80 with 95% confidence interval (CI). (B) Correlation between bacteria at genus level and

clinical symptoms of patients. (C-D) ROC curve for chest pain and fever in the symptomatic and asymptomatic groups. The AUCs were 0.904 (chest pain) and 0.793 (fever) with 95% confidence interval (CI).

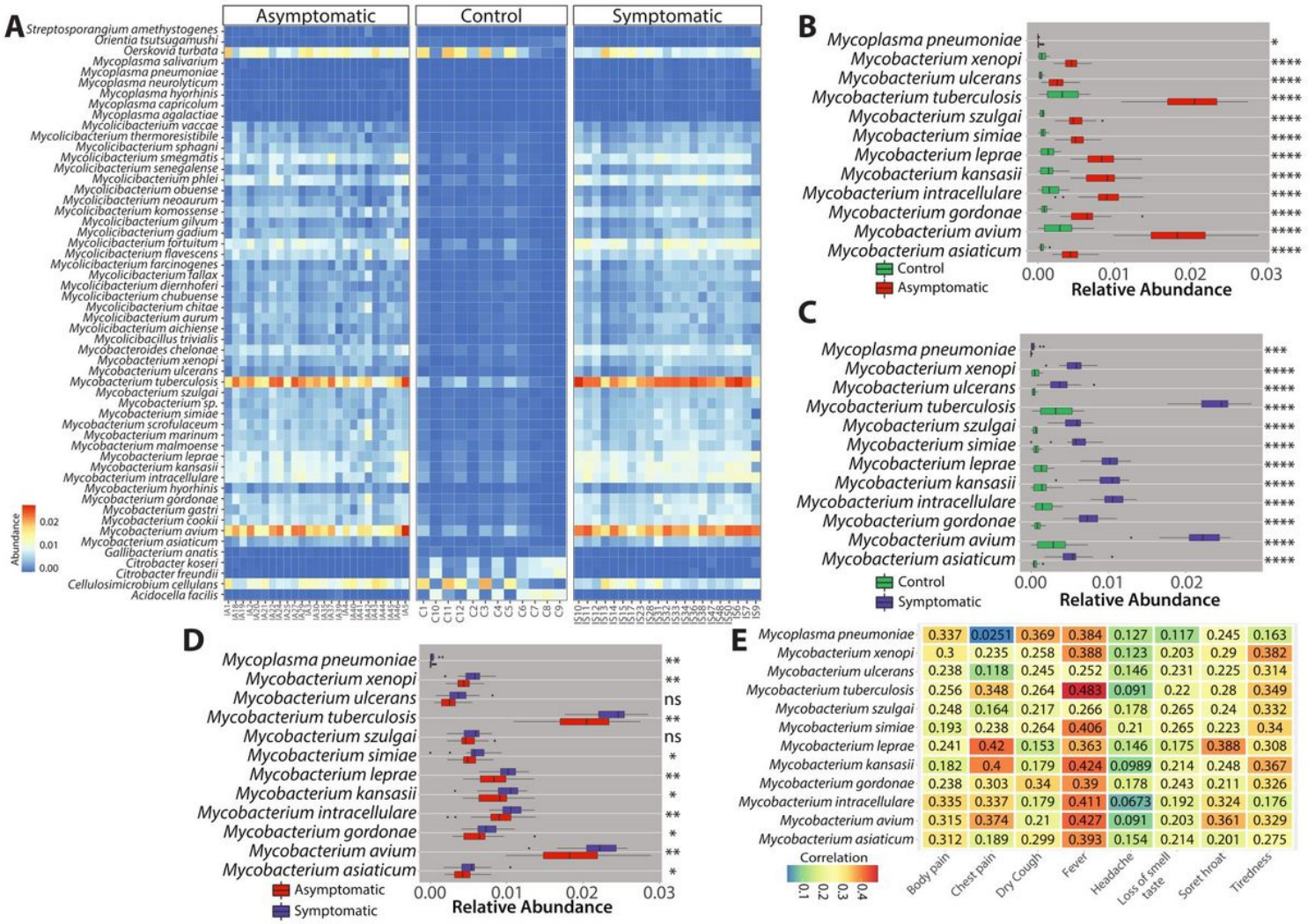

Figure 6

Relative abundance and species-specific OTUs reveal Mycobacterium spp. and Mycoplasma as key opportunistic pathogens. (A) Heatmap showing the taxonomic distribution of 54 species obtained from 12 genus based on LDA scores. (B-D) Boxplots showing relative abundance of opportunistic pathogens (Statistical parameters (Wilcoxon Rank test): ns: p = >0.05, 821 * p< = 0.05, ** p < = 0.01, *** p < = 0.001, **** p < = 0.0001) in C vs. IA, C vs. IS, and IA vs. IS groups. (E) Spearman correlation showing relationship between opportunistic pathogenic bacteria and patient's clinical symptoms.

Supplementary Files

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

- MahapatraMishraPrasadetalSupplementaldataBMC.pdf

- SupplementalTable1.xlsx
- SupplementalTable2.xlsx
- SupplementalTable3.xlsx