Composition of nasopharyngeal microbiota in individuals with SARS-CoV-2 infection across three COVID-19 waves in India

Tungadri Bose*[1], Wasimuddin*[2], Varnali Acharya*[2], Nishal Kumar Pinna*[1], Harrisham Kaur[1], Manish Ranjan[2], Jandhyala SaiKrishna[2], Tulasi Nagabandi[2], Binuja Varma[3], Karthik Bharadwaj Tallapaka[2], Divya Tej Sowpati[2], Mohammed Monzoorul Haque[1], Anirban Dutta[#1], Archana Bharadwaj Siva[#2], Sharmila S. Mande[#1]

* Equal contribution

Corresponding authors

Author Affiliations:

1. TCS Research, Tata Consultancy Services Limited, 54B Hadapsar Industrial Estate, Pune – 411013, Maharashtra, India
2. Centre for Cellular and Molecular Biology (CSIR-CCMB), Uppal Road, Hyderabad-500007, Telangana, India
3. TCS Genomics Lab, Tata Consultancy Services Limited, Embassy Galaxy Business Park, A-44-45, Block A, Industrial Area, Sector 62, Noida - 201309, Uttar Pradesh, India

Email Correspondence:

Sharmila S. Mande (SSM): sharmila.mande@tcs.com

Archana Bharadwaj Siva (ABS): abs@ccmb.res.in

Anirban Dutta (AD): anirban.dutt@tcs.com

Tungadri Bose (TB): tungadri.bose@tcs.com

Wasimuddin (WU): wasim.bt@gmail.com

Varnali Acharya (VA): acharya.varnali@gmail.com

Nishal Kumar Pinna (NKP): nishal.p@tcs.com

Harrisham Kaur (HK): harrisham.kaur@gmail.com

Manish Ranjan (MR): manishranjan@ccmb.res.in

Jandhyala SaiKrishna (JS): krishna@ccmb.res.in

Tulasi Nagabandi (TN): nagabanditulasi@gmail.com

Binuja Varma (BV): binuja.varma@tcs.com

Karthik Bharadwaj Tallapaka (KBT): karthikt@ccmb.res.in

Divya Tej Sowpati (DTS): tej@ccmb.res.in

Mohammed Monzoorul Haque (MMH): mm.haque@tcs.com

Abstract

Background

SARS-CoV-2 is an enveloped RNA virus that is responsible for the global pandemic COVID-19. Multiple variants of this virus have been plaguing the world through successive waves of infection over the past three years. Factors like age, preexisting comorbidities as well as variants of the virus have been attributed to influence disease severity and symptoms. Possible role/ association of the human microbiota in COVID-19 disease has also been investigated in some earlier studies. The compositional changes in microbial diversity during COVID-19 has been reported by independent authors in limited-sized cohorts from different geographical regions, some linking these observations with fungal and bacterial co-infections as well. Here, we study for the first time, the nasopharyngeal microbiota in COVID-19 patients, wherein the samples have been collected across the three COVID-19 waves witnessed in India, which were driven by different variants of concern.

Methods

We profiled the nasopharyngeal microbiota of 589 Indian subjects, across the three waves (First; n=181, Second; n=217, Third; n=191), which were categorized as COVID-19 positives (CP) and COVID-19 negatives (CN). These respective groups were further divided into subgroups based on the symptoms; viz., asymptomatic and symptomatic. The nasopharyngeal swabs were collected from subjects providing samples for diagnostics purposes at Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, India. Using high throughput 16S rRNA gene amplicon-based sequencing, we sequenced and profiled the nasopharyngeal microbiome and subsequently subjected them to diversity, composition, function and network analyses.

Results

Preliminary taxonomic analysis indicated that overall, Firmicutes, Proteobacteria and Actinobacteriota were amongst the dominating phyla in the Indian nasopharyngeal microbiota, while Staphylococcaceae and Corynebacteriaceae were amongst the most abundant families. Patients infected with SARS-CoV-2 showed a reduced microbial alpha diversity compared to the COVID-19 negatives, in a wave-dependent manner. Similarly, compositional change in the microbial community between CP and CN samples were also observed to be wave-dependent. Rickettsiaceae was the only family that was consistently

depleted in CP samples across the waves. The microbiota signatures of the first and third waves were more similar to each other compared to the second wave. On investigating how the compositional changes were associated with the presentation of symptoms, the genera *Staphylococcus*, *Anhydrobacter*, *Thermus*, and *Aerococcus* were observed to be highly abundant in the symptomatic CP samples. In contrast, *Elizabethkingia*, *Leptotrichia* and *Veillonella* were found to be relatively more abundant in asymptomatic CP samples.

Conclusion

Our study examined the nasopharyngeal microbiota of COVID-19 positive as well as COVID-19 negative Indians and reported varying patterns in the microbiota across the three COVID-19 waves. In addition, variations between the microbiota samples collected from symptomatic and asymptomatic patients have also been reported. Any causal association of these microbiota changes with respect to COVID-19 disease onset and/or progression would be interesting to explore in the future.

Keywords

Nasopharyngeal microbiome, COVID-19 disease, SARS-CoV-2, Variants of Concern (Voc), Indian cohort, 16S rRNA gene amplicon - based sequencing

BACKGROUND

SARS-CoV-2 is the causative agent for the pandemic Coronavirus Disease 2019 where there have been 660,378,145 confirmed cases of COVID-19, including 6,691,495 deaths as of 11th January 2023 (WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data). Since its onset, multiple variants of the virus have been reported, some of which have been categorized as variants of concern (VoC) viz. Alpha, Beta, Gamma, Delta, and Omicron [1]. While certain factors such as comorbidities, gender and age can influence the onset of the disease as well as influence the severity of symptoms [2], the effect of the SARS-CoV-2 variants on severity of COVID-19 disease has also been speculated. Differences in transmissibility and entry pathways of the variants [3], [4] could have potentially given rise to the spectrum of symptoms that have been observed in the affected patients across waves [5], [6], [4]. Most patients showed signs of acute respiratory distress syndrome and typical symptoms such as fever, dry cough and tiredness (Coronavirus disease (COVID-19) [who.int] while other groups suffered from pain, anosmia, nasal congestion, sore throat and diarrhea [7]. Moreover, a majority of the population was asymptomatic of the viral infection thus, acting as a hidden carrier [8].

Human microbiota plays a significant role in modulating the host health by forging the immune system and it is well known that the dysbiosis of the same has implications for diseases [9], [10], [11]. Therefore, association of the resident microbiota (or a dysbiosis thereof) with the symptoms and severity of the COVID-19 disease becomes worth exploring in context of different variants of concern. While some prominent studies in this direction have focused on gut microbiome dysbiosis in the COVID-19 infected patients [12], [13], [14]; recent reports also discuss the dynamics and the alterations of the nasopharyngeal microbiome [15], [16], [17]. However, these studies have mostly been restricted to the early waves of the pandemic, and do not include microbiome profiling during subsequent waves, which were caused by distinct virus variants (Fig. 1) and wherein the disease presentations were also different (first wave fueled by variant A2, second wave by Delta variant and third wave by the Omicron variant). Considering the strong association between viral and bacterial co-infections and respiratory disease severity, understanding the nasopharyngeal niche microbiome composition in context of the virulence/infectivity of different SARS-CoV-2 variants becomes imperative.

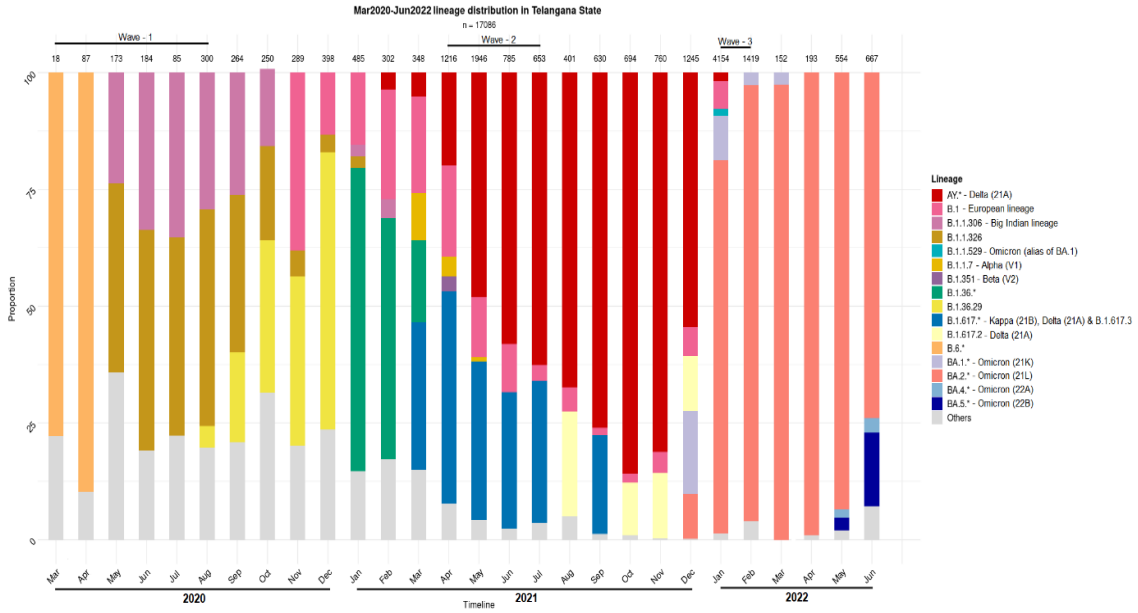

Figure 1: Distribution of the SARS-CoV-2 lineage among the samples collected from COVID-19 positive individuals. The nasopharyngeal samples were collected in the state of Telangana (India) between March 2020 and June 2022.

SARS-CoV-2 primarily enters the human body through ACE-2 and TMPRSS2 receptors present on the alveolar epithelial cells, and nasal epithelial cells of the nasopharyngeal tract and gradually moves towards the lungs [18], [19]). The delta and omicron variants have been observed to utilize different entry pathways for infection [3], and it is likely that during this migration, the resident microbiota can get altered and any inclusion of the pathobionts may aggravate the disease outcome. Studies by independent researchers have reported that nasopharyngeal microbiota during COVID-19 is characterized by a general decrease in the abundance of commensal organisms along with an increase in the abundance of opportunistic pathogens [20], [21], [22]. However, the individual organisms, both commensal as well as opportunistic, reported in each study seemed to vary. While this difference could be attributed to the choice of experimental protocols; geographical/ethnic differences between the subjects involved in each of the studies as well as the relatively lower number of samples analyzed in these studies, are also expected to play a major role in the study outcome. For instance, the opportunistic pathogens identified to be associated with the COVID-19 microbiota in an Indian study [23] differed considerably from those reported in studies conducted in other geographies [17], [21], [22]. Consequently, it is crucial to understand the nasopharyngeal microbiota signature of COVID-19 patients in India in the context of variants as well as waves through study of a larger cohort.

With the above objectives in mind, human nasopharyngeal microbiota was profiled from Indian subjects. The samples were provided for diagnostics and sequencing of SARS-CoV-2 at Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, India. The samples collected from the subjects/ patients were grouped into four categories based on COVID-19 infection status (positive or negative) as well as symptom presentation (symptomatic or asymptomatic), and further processed for high throughput 16S rRNA gene amplicon-based sequencing for microbiome profiling. In addition to uncovering the variations between microbial signatures of COVID-19 positive patients (CP) versus COVID-19 negative individuals (CN), the study also reports how the nasopharyngeal microbiota changed during the course of the pandemic over the three waves in India.

METHODS

Sample collection

This study was conducted for all three waves of COVID-19 in India, from March 2020 to February 2022 in accordance with the guidelines of the Indian Council of Medical Research (ICMR), Government of India and approved by the Institutional Ethics Committee of Centre for Cellular and Molecular Biology (CSIR-CCMB) [IEC-83/2020]. A total of 646 nasopharyngeal swab samples were collected from individual subjects, from different districts of Telangana state in India (data from 589 samples was used for further analysis after discarding 57 samples for various reasons including poor sequencing quality, depth, etc). The samples corresponded to the three COVID-19 waves in India and the time periods for swab collection for each wave were as follows: March 2020 to August 2020 for the first wave, April 2021 to July 2021 for the second, and January 2022 to February 2022 for the third. The swab samples were stored at -80°C until nucleic acid extraction, to ensure DNA quality. Subjects from each wave were divided into COVID-19 positive (CP) and COVID-19 negative (CN) groups. Among the CP group, subjects reporting severe symptoms and/or requiring hospitalization were further categorized as symptomatic COVID-19 positive (sCP). CP with no symptoms were denoted as asymptomatic COVID-19 positive (aCP). CN were further divided into asymptomatic (aCN) and symptomatic (sCN), i.e., COVID-19 negative subjects with presumably respiratory/other infections.

The samples used in this study were collected for SARS-CoV-2 diagnostics and genome sequencing at CSIR-CCMB, also an ICMR-approved COVID-19 testing center. The

diagnosis of COVID-19 patients involved combining the results from the real-time reverse transcription-polymerase chain reaction (qRT-PCR) assay performed on nasopharyngeal swabs, at the BSL-2 facility at CSIR-CCMB.

DNA isolation

DNA was extracted from the viral transport media (VTM) containing the nasopharyngeal swabs using the QIAamp® DNA Microbiome Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. All the extraction procedures were performed in the pre-PCR designated room in the BSL-2 facility. The yield of the extracted DNA was validated by Nanodrop Spectrophotometric (Thermofisher Scientific, Milan, Italy) reading at 260/280 as well as 260/230 and the quality was assessed by running an aliquot of the DNA in a 1% agarose gel.

16S rRNA gene amplification and sequencing

The first round of PCR was performed in a 20μL reaction mixture, with 50 ng of bacterial genomic DNA as the template. The primer pair targeting the V4 hypervariable region (515-F and 806-R) [24] of the 16S rRNA gene was modified by adding the Illumina overhang forward and reverse adaptor sequences.

V4 515-F

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGCCAGCMGCCGCGGTAA

V4 806-R

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACHVGGGTWTCTAAT

The component concentration of each reaction mix included 10μL of 2X EmeraldAmp® GT PCR master mix (Takara Bio, Japan), 1μL (0.5μM) of each primer, and 7μL of Nuclease free water. Thermocycling was performed on a Bio-Rad T100 Thermal cycler and included an initial denaturation at 95°C for 3 min, followed by 35 cycles of 95°C for 30s, 55°C for 30s, and 72°C for 30s, followed by a final extension of 72°C for 5 min. Each PCR reaction mixture was loaded into a 2% agarose gel, stained with ethidium bromide to observe the amplification.

The second round of PCR involved attaching dual indices and Illumina sequencing barcodes to the purified 16S rRNA gene amplicons, which was performed using the Nextera XT Index kit as mentioned in the manufacturer's protocol. The prepared metagenomic libraries were then quantified and normalized using Qubit dsDNA BR Assay before getting pooled in equimolar concentrations. Our constructed libraries were then sequenced (seven sequencing runs, Supplementary Table 1) in a paired-end mode (2 × 300) on the Illumina MiSeq sequencing platform using v3 600 cycles reagent.

Bioinformatic analysis

Data QC and ASV generation: A series of bioinformatic steps were followed to ensure that only high-quality reads generated from the sequencing experiments were used for generating the amplicon sequence variants (ASVs, i.e., sequence differing by as little as one nucleotide). The raw reads were processed using cutadapt (version 3.4) prior to their analysis through the DADA2 pipeline (version 1.20) for ASV generation and subsequent taxonomic assignment. Based on the quality of reads generated across the seven sequencing runs, the forward and reverse reads sequences were trimmed to 170 and 140 bases, respectively in the filterAndTrim step of DADA2. Furthermore, all read pairs with non-standard nucleotide bases, more than two 'expected errors' (maxEE), and lengths lower than 100 bases were discarded. Reads (and their pairs) which encountered atleast one base called with a quality score ≤ 8 (truncQ = 8) were also discarded. To account for the run-based biases, the ASVs were generated separately for each of the runs. The error models for each of the runs in the learnErrors step of DADA2 was performed using the parameters randomize=T and nbase=$5e^{+8}$. To aid in faster computation, the denoising step (wherein the error model was employed for denoising) was preceded by a dereplication step (derepFastq). The initial sequences table generated for each of the seven sequencing runs (through the makeSequenceTable step of DADA2) were finally merged using mergeSequenceTables command of DADA2. Next, the removeBimeraDenovo step (method = "consensus") was employed to remove ASVs originating from chimeric sequences. The retained ASVs were annotated using the assignTaxonomy function in DADA2 following the naïve Bayesian classifier method with the silva_nr_v132 database. Further, the species for these ASVs were assigned using addSpecies function with the silva_species_assignment_v132 database. As a final quality control step, ASVs with extremely low counts (n < 10 across all the sequenced samples) and irrelevant annotations (viz, Order classified as 'chloroplast' or 'na'; Family classified as 'mitochondria' or 'na'; Phylum classified as 'uncharacterized' or 'na') were also

removed. The final microbiota data consisted of 8645 ASV belonging to 589 samples and were used for subsequent analysis.

Alpha diversity: To determine intra individual diversity, we calculated two alpha diversity measures: observed number of ASVs (observed ASVs) and Shannon diversity using the Phyloseq package (*phyloseq*, 1.26.1). Using generalized linear mixed models (GLMMs), we modeled alpha diversity according to COVID-19 status (CN; n=285, CP; n=304), wave (First; n=181, Second; n=217, Third; n=191), interaction between them (COVID-19*Wave), symptoms (Asymptomatic [aCP + aCN]; n=361, Symptomatic [sCP + sCN]; n=228), and sequencing depth to account for differential sequencing effort between samples, sequencing run and to control the random effect of batch processing of samples (*lme4*). In addition, we included the variables age group (25-29y; n=149, 30-34y; n=141, 35-39y; n=122, 40-44y; n=82, 45-50y; n=95), gender (female; n=276, male=313) and Ct-value as these factors could influence the microbiome. To facilitate model convergence, sequencing depth was scaled and to control the run effect we included run as a random factor variable. We used a log distribution when modeling the count data (observed ASVs) and a normal distribution when modeling the continuous data (Shannon diversity). Model selection was based on the information-theoretic (IT) approach using a second order Akaike's information criterion corrected for small sample sizes (AIC_C) as an information criterion and Akaike weights (ω) to determine model support. For all GLMMs, we report both conditional and marginal coefficients of determination of each model ($R^2_{GLMM(c)}$, which explains the variance of both the fixed and random factors, and $R^2_{GLMM(m)}$, which explains the variance of the fixed factors only), which we calculated as the variance explained by the best model and the ΔAIC_C. Finally, we performed Tukey's HSD test to detect differences between each individual category (i.e., COVID-19*Wave) on the above performed GLMMs' outcome.

Beta diversity: To assess the nasopharyngeal bacterial community composition between individuals, we calculated the Jaccard and Bray–Curtis distance matrices using the Phyloseq package (*phyloseq*, 1.26.1). Jaccard accounts for presence-absence of the taxa whereas Bray-Curtis in addition gives weight to taxa abundance. We tested for differences in microbial beta diversity for COVID-19 in different waves using the permutational multivariate analysis of variance (PERMANOVA) test with 999 permutations implemented in the *adonis* function of the vegan package (Vegan, 2.6.2). The fixed variables in our full model were the following: symptoms, age group, gender, sequencing run, Ct-value, and sequencing depth. We retained

the sequencing run variable in our full model to statistically account for its model support. To understand whether COVID-19 reflects true shift in microbial community composition or differential spread (dispersion) of data points from their group centroid, we investigated the homogeneity of the variances of COVID-19 positive and negative category using the PERMDISP test implemented in the *betadisper* function of the vegan package. To visualize patterns of separation between different sample categories, Principal Coordinates Analysis (PCoA) plots were prepared based on the Bray–Curtis dissimilarity coefficient.

Discriminant taxonomy analysis: To identify the differential abundant discriminating taxa between different the nasopharyngeal samples of subjects with varying COVID-19 status, symptoms and across different waves; negative binomial Wald tests were performed using the DESeq2 package (v 1.34.0). For this analysis, only 2120 (out of 8645) ASVs which were present in at least two samples were considered. In order to obtain differentially abundant taxa between Covid-19 positive (CP) and Covid-19 negative (CN) after correcting for the wave effect (i.e., overall) Design-1 (design = ~ Wave + Covid) was adopted for DESeq2 analysis. Further, Design-2 (design = ~ Wave + Covid + Wave:Covid) was implemented to understand the individual and combined effects (interaction) of wave (Wave) and COVID-19 status (Covid) on the microbiota composition. In addition, to ascertain the effect of symptoms (sym_asym) in conjugation with COVID-19 status (Covid), Design-3 (design = ~ sym_asym + Covid + sym_asym:Covid) was used. For each of the designs, DESeq2 was performed with default parameters except for the size factor estimation. The size factor estimation was done using “poscounts”. The discriminating taxonomies were also analyzed for higher taxonomic levels. For this purpose, the ASVs were also aggregated at higher taxonomic levels to generate abundance tables corresponding to Phyla, Family and Genus and a protocol similar to the ASV level analysis was followed. To obtain the significant taxonomic groups across the comparisons log2FoldChange and lfcSE (log2FoldChange Standard Error) were plotted using the ggplot (v 3.3.5) package in R environment (R version 4.1.1, R Development Core Team, 2011) software.

Network construction and identification of driver taxa: Correlational networks were constructed from the microbial abundance data to understand associations among the interacting taxa. All analyses were performed at the Genus level to have a dataset with a reasonable number of nodes for a good display and meaningful interpretation thereof. For generating the networks, a Pearson correlation coefficient cut-off of ±0.5 was used. The

networks were generated using Cytoscope (v 3.9.0) and analyzed using its 'analyze network' module. Further, the taxa driving the major shift in case of COVID-19 infections were identified using NetShift [25]. The tool is hosted on https://web.rniapps.net/netshift/ and uses 'Neighbor Shift (NESH) Score' to identify key taxonomic groups which are more likely to drive changes from a control/ healthy, to a case/disease state.

Discriminant functional analysis: The functional potential of the microbiota samples was predicted using PICRUSt2 [26]. Next discriminant analysis on the inferred functions (enzymes and metabolic pathways) were performed using the DESeq2 package (v 1.34.0). A similar design as elucidated for the discriminant taxonomy analysis was adopted in this case.

Results

Data overview

Nasopharyngeal swabs samples analyzed from 589 individuals between March 2020 – February 2022 (i.e., spanning three 'waves of COVID-19 infection' in India) were subjected to amplicon sequencing of the bacterial V4 hypervariable region of 16S rRNA gene on the Illumina MiSeq platform (details in Methods section). Category and wave-wise distribution of the 589 samples viz., symptomatic COVID-19 positives (sCP), asymptomatic COVID-19 positives (aCP), apparently healthy controls who were COVID-19 negatives as well as asymptomatic (aCN) and COVID-19 negatives exhibiting symptoms (sCN) (i.e., due to other respiratory/other infections) has been provided in Table 1. A total of 8645 amplicon sequence variants (ASVs) were identified in the microbiome sequence data (Supplementary Fig. 1) using DADA2 package (detailed protocol described in Methods section). Preliminary taxonomic analysis indicated the dominance of the phylum Firmicutes, Proteobacteria, and Actinobacteriota across all samples (cumulative abundance over 75%) (Supplementary Fig. 2a & 2b). At the family level, Staphylococcaceae and Corynebacteriaceae were found to be the most abundant taxa (Supplementary Fig. 2c & 2d).

Category	Covid Positives (CP)		Covid Negatives (CN)	
	with symptoms (sCP)	without symptoms (aCP)	with symptoms (sCN)	asymptomatic apparently healthy controls (aCN)
Total	147	157	81	204
Wave1	19	70	12	80
Wave2	81	39	60	37
Wave3	47	48	9	87

Table 1: Statistics on the number of samples analyzed in the study. Categorization of 589 analyzed samples into sub-groups and their distribution across the three COVID-19 waves have been provided.

COVID-19 nasopharyngeal microbiome showed reduced microbial alpha diversity

To test if COVID-19 status has any impact on the nasopharyngeal microbiome, we calculated the microbial alpha diversity using two indices: the observed number of amplicon sequence variants (ASVs) and Shannon diversity. We used generalized linear mixed models, which allowed us to account for differential sequencing effort between samples and sequencing runs. Using model selection based on the information-theoretic (IT) approach, we found strong support for an effect of COVID-19 status, wave and combined effect of COVID-19 status & wave (COVID-19*Wave) on observed ASVs (ΔAIC_C = 38.47, $R^2_{GLMM(m)}$ = 0.312, $R^2_{GLMM(c)}$ = 0.320, Fig. 2a & 2b). Similarly, Shannon diversity, in which microbial alpha diversity is weighed for abundance, was observed to be influenced by COVID-19 status and Wave, individually, but not by COVID-19 status & wave together (ΔAIC_C = 22.85, $R^2_{GLMM(m)}$ = 0.277, $R^2_{GLMM(c)}$ = 0.312, Fig. 2c & 2d). For both alpha diversity indices, in general, lower diversity was observed in the COVID-19 positive (CP) samples as compared to the COVID-19 negative (CN) category (Fig. 2a-d). Tukey's HSD result suggested that for observed ASVs, only third wave samples showed significant ($p<0.05$) lower diversity in the CP category as compared to CN. For Shannon diversity too, we observed a similar trend of significantly lower diversity in the CP samples in the third wave alone, with no significant differences seen in the other waves (all $p>0.05$). Overall, the results show a strong effect of COVID-19 infection on microbial alpha diversity indices, with

the observed ASVs also showing an effect in a wave-dependent manner. Furthermore, the results suggest that this effect may be driven more by rare taxa than abundant ones, since the effect was less pronounced when accounting for abundance (i.e., Shannon diversity). The effects of other variables included in the models such as age group, symptoms, gender, Ct-value were poorly supported by AIC_C model comparison for both alpha diversity indices (data not shown).

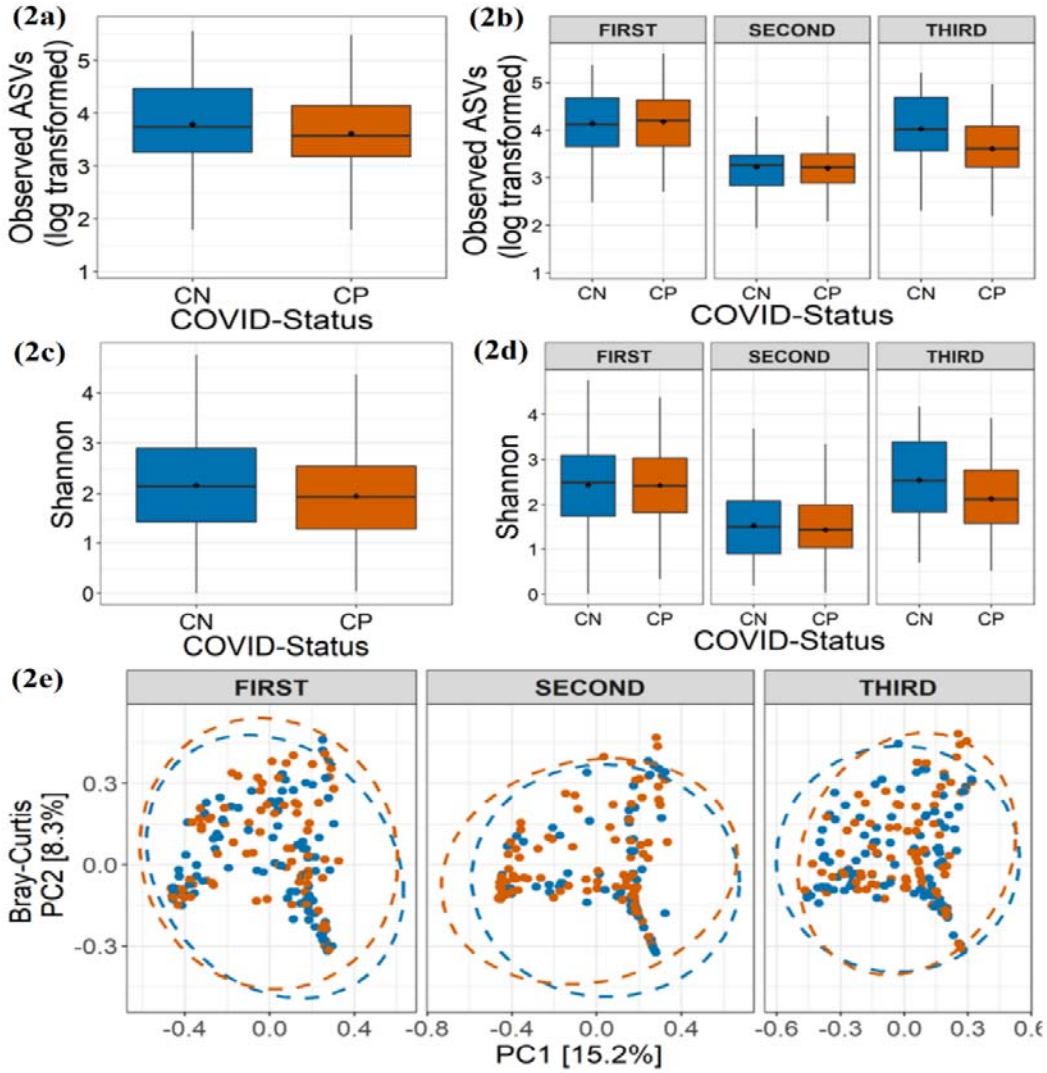

Figure 2: **Alpha and beta diversity of the nasopharyngeal microbiota in COVID-19 positive (CP) and COVID-19 negative (CN) samples.** Number of observed amplicon sequence variants (ASVs) (a) *in all the analyzed samples (*overall), and (b) in the three COVID-19 waves are shown. Shannon diversity in term of ASVs (c) *in all the analyzed samples*, and (d) in the three COVID-19 waves are depicted. (e) Beta diversity of microbiota assessed using principal component analysis (Bray-Curtis distance) represented along first two principal components for the three COVID-19 waves are presented.

Community composition perturbed in COVID-19 nasopharyngeal microbiome

To determine whether COVID-19 status influences the nasopharyngeal microbial community composition, we calculated two beta diversity metrics (Jaccard, Bray-Curtis distance). PERMANOVA models indicated a strong support for, the influence of COVID-19 status (Jaccard: R^2=0.003, p=0.006; Bray-Curtis: R^2=0.003, p=0.007), wave (Jaccard: R^2=0.026, p=0.001; Bray-Curtis: R^2=0.037, p=0.001) and the interaction between them (COVID-19*Wave) (Jaccard: R^2=0.005, p=0.001; Bray-Curtis: R^2=0.006, p=0.002) on the nasopharyngeal microbial beta diversity estimates (Fig. 2e). Additionally, not only Ct-value (Jaccard: R^2=0.003, p=0.005; Bray-Curtis: R^2=0.003, p=0.004) and symptoms (Jaccard: R^2=0.002, p=0.034; Bray-Curtis: R^2=0.002, p=0.017) but sequencing run (Jaccard: R^2=0.010, p=0.002; Bray-Curtis: R^2=0.011, p=0.004) and sequencing depth (Jaccard: R^2=0.011, p=0.001; Bray-Curtis: R^2=0.016, p=0.001) were also shown to influence the community composition. However, no effect of age group or gender (all $p>0.05$) on the nasopharyngeal microbial community composition was observed. PERMDISP tests showed true shift in microbial community composition and no dispersion effect (all $p>0.05$) [Fig. 2e].

Microbes associated with disease status across different COVID-19 waves

Given the observed variation of microbial taxonomic diversity across three COVID-19 waves, negative binomial Wald tests were performed using the DESeq2 package with the intent to identify taxa that are differentially abundant between CP and CN groups corresponding to different waves. Except for a significantly enriched proportion of Proteobacteria in CP compared to CN from third wave (BH corrected $p<0.05$), no significant differences were observed at a phylum level between CP and CN groups in the first and third waves. In contrast, analysis of second wave samples revealed that CP microbiota is characterized by significantly higher abundances of Cyanobacteria and Firmicutes, and a significantly lower abundance of Planctomycetota, Deinococcota, and Bdellovibrionota (Supplementary Fig. 3).

Taxonomic groups discriminating between CP and CN in each of the three waves at the family level are shown in Fig. 3. Similar to the observations made at the phylum level, at the family level only a small number of taxa showed significant differences between CP and CN microbiota from the first wave. This includes an enriched abundance of Hymenobacteraceae

and Spirosomaceae (both belonging to phylum Bacteroidota), and a decrease in the abundance of Thermaceae and Rickettsiaceae in CP compared to CN. Making the comparison at family level for the samples from the second wave CP and CN groups revealed a total of 16 families each to be enriched in the CP and CN groups, respectively. Amongst these, Rhizobiales.Incertae.Sedis, Hydrogenophillaceae, Alteromonadaceae (all three belonging to phylum Proteobacteria), Hymenobacteraceae and Brevibacteriacceae, were among the bacterial families which were found most enriched in CP (w.r.t. CN). Devosiaceae and Thermaceae were the two most enriched families in CN (w.r.t. CP). In case of the third wave, three families (viz. Carnobacteriaceae, Burkholderiaceae, and Corynebacteriaceae) had a higher abundance in CP group whereas seven families (viz. Aeromonadaceae, Rickettsiaceae, Alteromonadaceae, Pasteurellaceae, Spirosomaceae, Aerococcaceae, Fusobacteriaceae) exhibited a lower abundance in CP group with respect to the CN group. The most significant fold changes were observed for Aeromonadaceae, Alteromonadaceae, Rickettsiaceae (all three along with Pasteurellaceae belonging to phylum Proteobacteria) and Spirosomaceae, all of which were significantly enriched in the CN samples when compared to CP samples. The list of the significant changes in abundance of microbes between CP and CN, at genera and ASV levels, across the three waves have been provided in (Supplementary Table 2). The genera *Aliterella* was noted to demonstrate a significant decrease in abundance in the CP group across all the three infection waves. Similarly, the CP samples were also characterized by a significantly reduced abundance of ASVs ASV36 (family Rickettsiaceae), ASV161 (*Gemella*), ASV195 (*Deinococcus*), and ASV431 (*Elizabethkingia*).

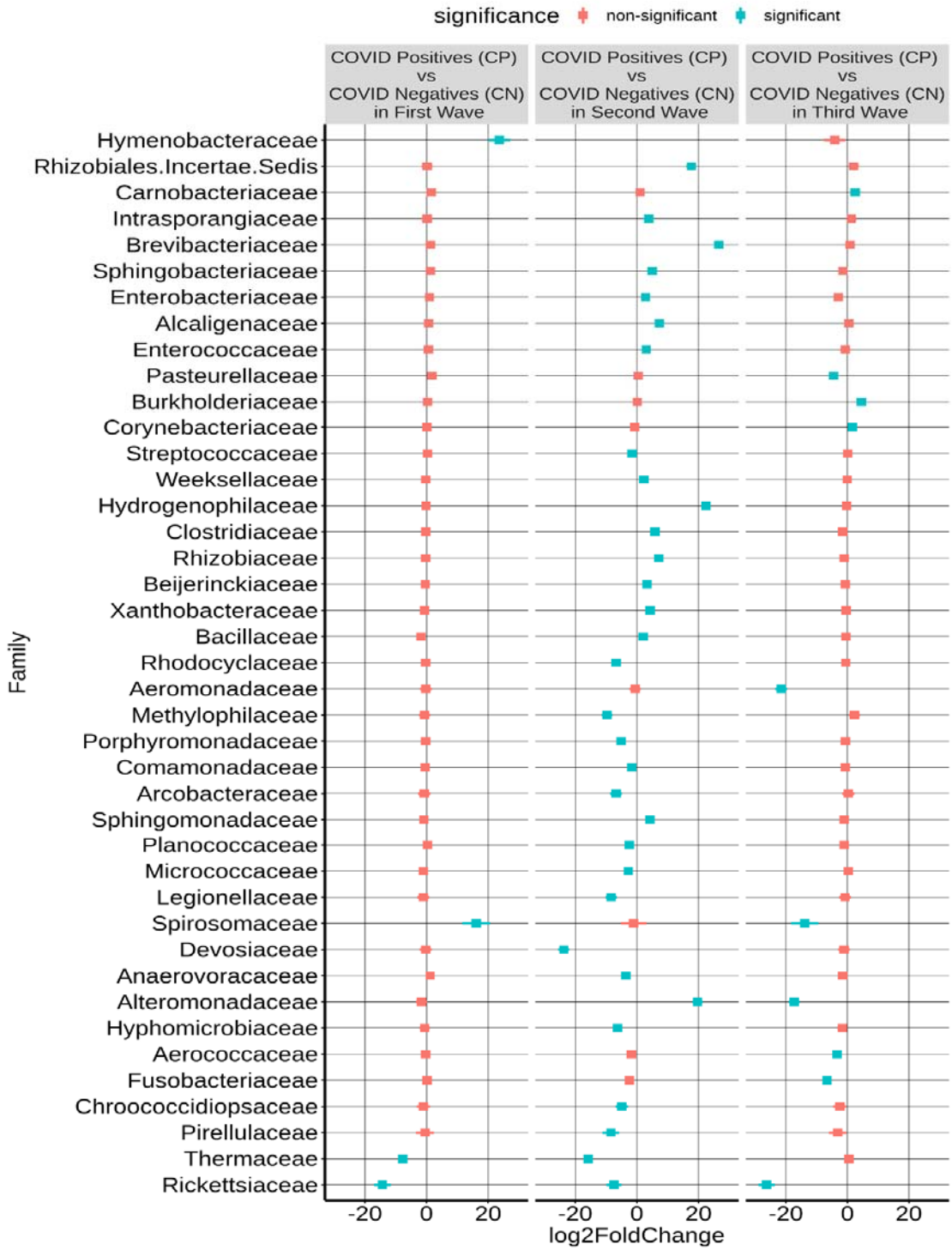

Figure 3: Differential abundance of bacterial families between the COVID-19 positive (CP) and COVID-19 negative (CN) samples. The log2fold change in the mean abundance (along with whiskers representing standard errors) of a bacterial family in CP with respect to CN is depicted in (a) all the analyzed samples (overall) as well as in (b-d) each of the three COVID-19 waves. Significantly different abundance (q-value < 0.05) is indicated with blue colour.

Microbes associated with inter-wave variations in CP and CN groups

(Supplementary Fig. 4) shows the effect of different waves on the bacterial taxonomic proportions observed in CN and CP groups of samples in terms of fold changes at the phylum level. As expected, based on the observations from the alpha and beta diversity tests, the microbiome of the second wave samples was distinct from the other two. When compared to the first wave, the second wave samples were found to be significantly enriched with the phyla Proteobacteria, Campylobacterota and Firmicutes, in both CP and CN groups. In contrast, Cyanobacteria, Planctomycetota, Bdellovibrionota and Fusobacteriota was depleted in both the CP and CN samples from the second wave when compared to those from the first wave. When the third wave samples were compared with those from the second wave, the phyla Cyanobacteria, Plantomycetota and Bdellovibrionota were significantly enriched in the third wave in both CP and CN samples. On the other hand, only Firmicutes were observed to be significantly depleted in the CP and CN samples of the third wave when compared to the second wave. The first wave and the third wave samples were observed to be more similar to each other. The only significant differences pertained to enrichment of Deinococcota and depletion of Bacteriodota in the CP and CN samples of third wave compared to the first wave samples. It was intriguing to note that the phylum Deinococcota showed a gradual (and in most cases significant) increasing trend over the three waves in both CP and CN groups.

Additional results depicting the wave effect on microbiota at genera and ASV levels are provided in Supplementary Table 3. Among the CP group, the genera *Abiotrophia*, *Streptococcus*, *Rheinheimera* along with *X.Eubacterium*..brachy.group and *Escherichia.Shigella* were found to follow a constant decrease in abundance (BH corrected $p<0.05$) between the corresponding waves (i.e., these genera were most abundant in the first wave CP samples and their abundances were least in the third wave CP samples). In the CN group, *Schlegelella* had a significantly increasing trend, whereas *Acinetobacter* and *Megasphaera* showed a significant decreasing pattern. Table 2 shows a list of the ASVs and genera which followed a consistent pattern of significant increase or decrease in abundance in the CP and CN groups.

	COVID-19 Positive grou						
	ASV/ Genus	base Mean	Wave 3 vs Wave 2		Wave 2 vs Wave 1		Phylum
			log2 Fold	padj	log2 Fold	padj	
	ASV22	2229.73	36.39	0.000	4.96	0.004	Proteobacteria
	ASV54	218.86	-23.34	0.000	-4.34	0.014	Firmicutes
	ASV90	94.61	-31.89	0.000	-6.10	0.008	Actinobacteriota
	ASV79	63.90	-23.75	0.000	-7.47	0.001	Firmicutes
	ASV97	61.34	-22.32	0.000	-8.11	0.000	Firmicutes

Table 2: List of the amplicon sequence variants (ASVs) and genera demonstrating a consistent increase or decrease in abundance across the three COVID-19 waves. Significant changes having an adjusted p-value < 0.05 have been reported.

Microbial taxa associated with symptomatic and asymptomatic COVID-19 subjects

Supplementary Fig. 5 shows the overall bacterial taxonomic proportions observed in CN and CP groups of samples categorized based on symptom in terms of fold changes at the phylum levels (abundance plots are seen in Supplementary Fig. 2b & d). Amongst the COVID-19 positive (CP) subgroups, sCP had a significant overabundance of the phyla Camphylobacterota, Patescibacteria and Firmicutes compared to aCP. On the other hand, the phyla Fusobacteria and Proteobacteria were enriched in the asymptomatic COVID-19 positives (aCP). When comparing between the COVID-19 negative (CN) subgroups, Camphylobacterota and Proteobacteria were enriched in the symptomatic group (sCN). In contrast, Deinococcota, Actinobacterota and Bacteroidota were enriched in the apparently healthy subgroup (aCN). When comparing between the symptomatic groups (i.e. sCP vs. sCN), Deinococcota was identified to be enriched in sCP whereas Proteobacteria was significantly depleted. Notably, with respect to aCN, the abundance of Proteobacteria was found to be significantly higher in abundance in aCP, sCN and sCP.

Figure 4 enlists the bacterial families which were found to be significantly different in abundance between each of these subgroups. Maximum number of significantly different families (22 families) were observed between aCN (apparently healthy) and sCN (symptomatic COVID-19 negative) subgroups. In comparison to aCN, the most prominent fold change was noted for Rickettsiaceae, which was enriched in sCN. On the other hand, families with most depleted abundance in sCN included Thermaceae, Dysgonomonadaceae,

Dermacoccaceae, and Alteromonadaceae. Within the CP subgroups, sCP was observed to have a significantly higher abundance of Rickettsiaceae, Aerococcaceae and Thermaceae among others as well as a significantly lower abundance of Leptotrichiaceae, Weeksellaceae, Deinococcaceae, Sphingomonadaceae and Xanthobacteraceae (the latter two belonging to phylum Proteobacteria) when compared to aCP. With respect to the microbiota samples from aCN subgroup, the signature of the aCP microbiota was characterized by a reduced abundance of Rickettsiaceae, Enterobacteriaceae (both belonging to phylum Proteobacteria), Moraxellaceae, Pseudomonadaceae (both belonging to phyla Pseudomonadota), Aerococcaceae, Thermaceae, and Micrococcaceae. At the same time, an enriched abundance of Burkholderiaceae, Xanthobacteraceae (both Proteobacteria) and Carnobacteriaceae were observed in aCP. Further, within the symptomatic subgroups, Thermaceae, Dysgonomonadaceae, Dermacoccaceae, and Alteromonadaceae were amongst the bacterial families with significant enrichments in sCP (w.r.t. sCN). In contrast, w.r.t. sCN, Rickettsiaceae was observed to be the most depleted family in sCP. Notably, while the abundance of Rickettsiaceae decreased during COVID-19 infections, it has been reported to increase in other respiratory infections. Thus, it appears to be a distinguishing feature between COVID-19 and other respiratory infections. The severity of COVID-19 infection (i.e., sCP vs aCP) appeared to be linked to a higher abundance of the family Saccharimonadaceae, and Chitinophagaceae, and a lower abundance of the family Sphingomonadaceae, and Weeksellaceae. Similarly, enrichment of Thermaceae and Enterobacteriaceae, as well as a depleted abundance of Xanthobacteraceae appeared to be markers of a healthy nasopharyngeal microbiota since this trend was observed when aCN group was compared to both the sCN and aCP groups. *Kocuria* and *Liquorilactobacillus* were also linked to a healthy nasopharyngeal microbiota (Supplementary Table 4). Further, the opportunistic pathogenic genera *Stenotrophomonas* was noted to be depleted in a healthy respiratory tract microbiota. The list of the significant changes in abundance of microbes at a genus level between sCP and aCP have been provided in Table 3. *Thermus, Aerococcus, Enhydrobacter, and Staphylococcus* were found to be the most significantly enriched genera in sCP when compared to aCP. ASV108 (*Thermus amyloliquefaciens*) was found to be enriched in sCP in comparison to both aCP as well as sCN (Supplementary Table 4) indicating that its levels in the nasopharyngeal microbiota might act as an early indicator of the severity of the disease outcome. In contrast, *Elizabethkingia, Leptotrichia,* and *Veillonella* were more abundant in aCP (w.r.t. sCP) (Table 3).

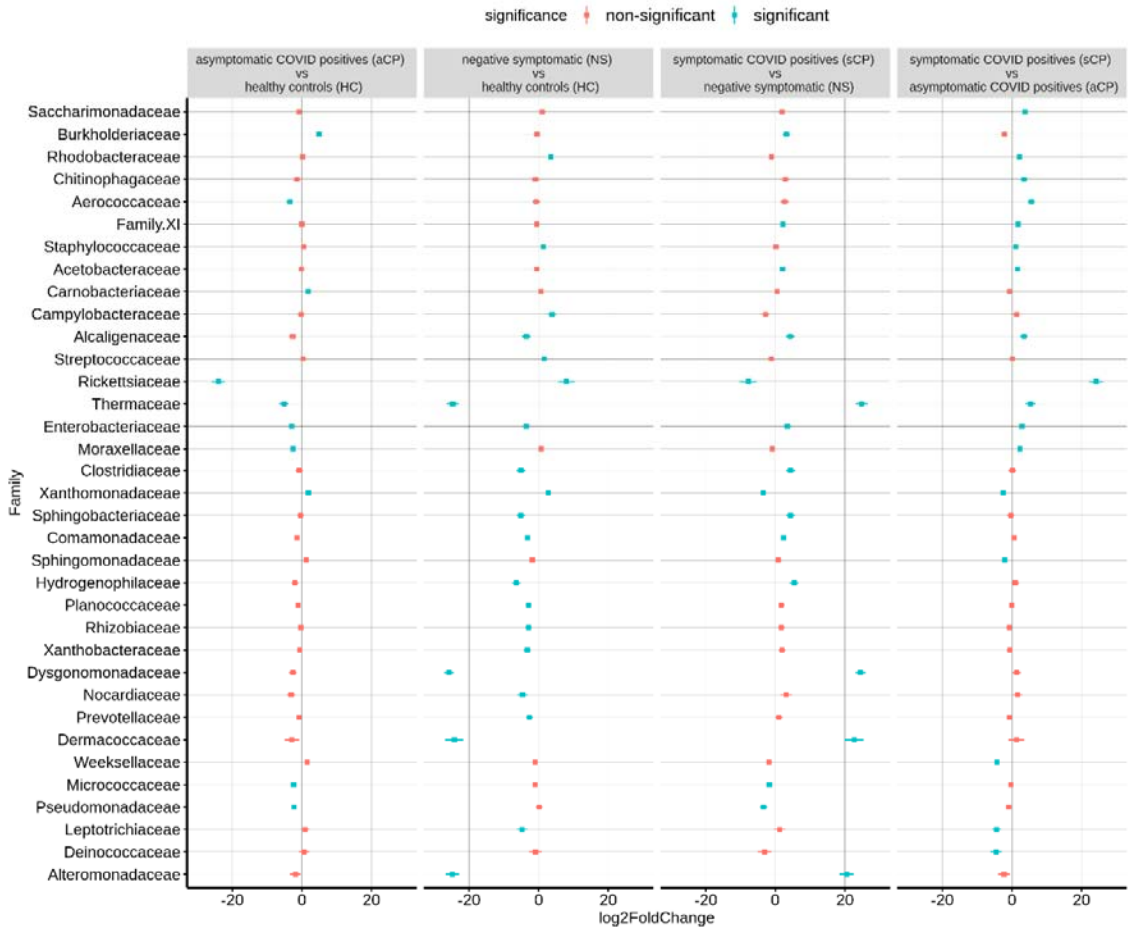

Figure 4: Differential abundance of bacterial families between the four sub-group of samples - asymptomatic COVID-19 positive (aCP), symptomatic COVID-19 positive (sCP), apparently healthy controls who are COVID-19 negatives as well as asymptomatic (aCN), and COVID-19 negatives exhibiting symptoms (sCN). The log2fold change in the mean abundance (along with whiskers representing standard errors) of a bacterial family in (a) aCP with respect to aCN, (b) sCN with respect to aCN, (c) sCP with respect to sCN, and (d) sCP with respect to aCP is depicted. Significantly different abundance (q-value < 0.05) is indicated with blue colour.

Genus	base Mean	log2 Fold Change	padj	Phylum	Family
Elizabethkingia	89.341	-4.8943	0.0019	Bacteroidota	Weeksellaceae
Leptotrichia	49.941	-4.8144	0.0085	Fusobacteriota	Leptotrichiaceae
Veillonella	245.31	-2.6977	0.0145	Firmicutes	Veillonellaceae
Staphylococcus	54139	1.2499	0.003	Firmicutes	Staphylococcaceae
Enhydrobacter	161.48	2.8906	0.0411	Proteobacteria	Moraxellaceae
Thermus	26.63	4.7846	0.0145	Deinococcota	Thermaceae
Aerococcus	713.23	9.1933	3E-11	Firmicutes	Aerococcaceae

Tables 3: List of the bacterial genera which showed significant changes in abundance between symptomatic COVID-19 positive (sCP) and asymptomatic COVID-19 positive (aCP) sub-groups. Significant changes having an adjusted p-value < 0.05 have been reported.

Positive log2foldchange values indicate a higher abundance in sCP while negative log2foldchange values represented genera with higher abundance in aCP.

Network analysis

The microbial association network (Pearson correlation) corresponding to each of the three COVID-19 waves, CP, CN as well as those for symptomatic and asymptomatic conditions were constructed considering the taxonomic abundance data at the genera level. Network properties of each of these networks are provided in Table 4. In line with observations pertaining to diversity indices, minimal number of nodes (genera) and edges (interactions between nodes) were observed in the network corresponding to the second wave. Also, the number of nodes and edges in the third wave network was observed to be considerably higher

than that in the first two waves. The distribution of betweenness centralities of the nodes in the network seemed to indicate a considerable shift in the network architecture between the subsequent waves (Supplementary Fig. 6). Given these preliminary observations, 'NetShift' [25] analysis was used to identify taxonomic groups which might be responsible for driving this wave-to-wave shift. As mentioned earlier, with respect to the first wave, we noted a loss of many nodes in the network corresponding to the second wave, including the genus *Actinomyces* which appeared to be a high betweenness node (degree = 14; betweenness =0.618; stress = 74) in the microbial network corresponding to the first wave (Supplementary Fig. 7). The change in network architecture from the first to the second wave was driven by the genera *Cnuella*, *Marmoricola*, *Paenibacillus*, *Peptostreptococcus*, and *Solobacterium* as indicated by their NESH scores (Supplementary Table 5). Although the abundance of most of these genera (except *Cnuella*) in the second wave was lower in both CP and CN samples w.r.t. First wave (Supplementary Table 3), the driver taxa were found to disrupt the sub-network of the first wave, thereby leading to an altered association among the genera in the network representing the second wave. *Cnuella* was higher in abundance only in the CP group, while its abundance decreased in the CN group. Notably, some organisms belonging to the genera *Paenibacillus*, *Peptostreptococcus*, and *Solobacterium* are known to be opportunistic pathogens [27], [28], [29]. Further, some of these driver taxonomic groups, viz., *Cnuella*, *Marmoricola*, and *Paenibacillus* lost their importance in the microbiome association network corresponding to the third wave (Supplementary Fig. 8). *Leptotrichia* and *Actinomyces* were noted to be taxonomic groups driving the changes during the third wave when compared to the second wave. While the abundance of *Actinomyces* decreased, the abundance of *Leptotrichia* was higher in the third wave w.r.t. the second wave (Supplementary Table 3). Further, *Leptotrichia* appeared as a node with high stress centrality (degree = 18; betweenness =0.317; stress = 236) in the third wave network.

The microbiome association network corresponding to the symptomatic condition (considering data from all three waves as well as both CP and CN groups combined) was found to be represented by a relatively lower number of microbial groups which also interacted less coherently amongst themselves (average number of neighbors), compared to the network corresponding to the asymptomatic condition (Table 4). From the NetShift constructed network (Supplementary Fig. 9), it was intriguing to note that the sub-network composition between the asymptomatic and symptomatic networks largely remained the same, with nodes belonging to the major four clusters showing no change in their cluster

memberships. However, the betweenness centralities of some of these nodes changed between the symptomatic and the asymptomatic condition, indicating a change in the overall network architecture (Supplementary Fig. 9). Such changes were most evident in the clusters 2 and 3. Amongst other noticeable changes, the genera *Methyloversatilis* seemed to gain prominence (both an increased NESH score as well as a higher abundance) in the network representing the symptomatic condition (Supplementary Table 5). In contrast *Porphyromonas*, *Campylobacter*, and *Skermanella* were found to lose their importance in the third wave network due to a decrease in their abundance or that of their interacting partners (Supplementary Tables 5 and Supplementary Fig. 9).

Network	No of Nodes	No of Edges	Avg No of Neighbours	Characteristic Path Length	Clustering Coefficient
Wave 1	85	278	5.5	1.808	0.774
Wave 2	56	92	3.33	1.4	0.667
Wave 3	92	486	4.7	2.284	0.586
Symptomatic	49	128	5.75	1.179	0.94
Asymptomatic	63	214	7.833	1.303	0.854
Covid Positive	69	214	2	3.091	0.091
Covid Negative	75	210	4.727	1.673	0.596

Table 4: Properties of the microbial association networks analyzed in this study.

In case of CP and CN networks, the CP-network was characterized by longer path lengths (3.091) compared to the CN-network (1.673). The network density values indicate a denser CN-network compared to the CP-network. The number of nodes with high betweenness values were also higher in the CN network compared to the CP network. The interacting partners for the microbes were seen to considerably change from the CN to CP. This resulted into the generation of alternate microbial subnetworks between the two conditions which might have an implication of the behavior of the individual microbes (Fig. 5 a & b). Major changes were observed in clusters 2, 3 and 4 of the CN network (control). In particular, several members of cluster 2, such as *[Eubacterium] nodatum* group, *Atopobium*, and *Mogibacterium* were characterized by an increased NESH score in the CP network. Notably, the genera *Schlegelella* which was earlier found to consistently increase in abundance across the three waves of the COVID-19 pandemic in the CN samples (Table 2), witnessed a drop in its betweenness centrality measure in the CP network, thereby indicating at its strong association with the CN nasopharyngeal microbiota (Fig. 5).

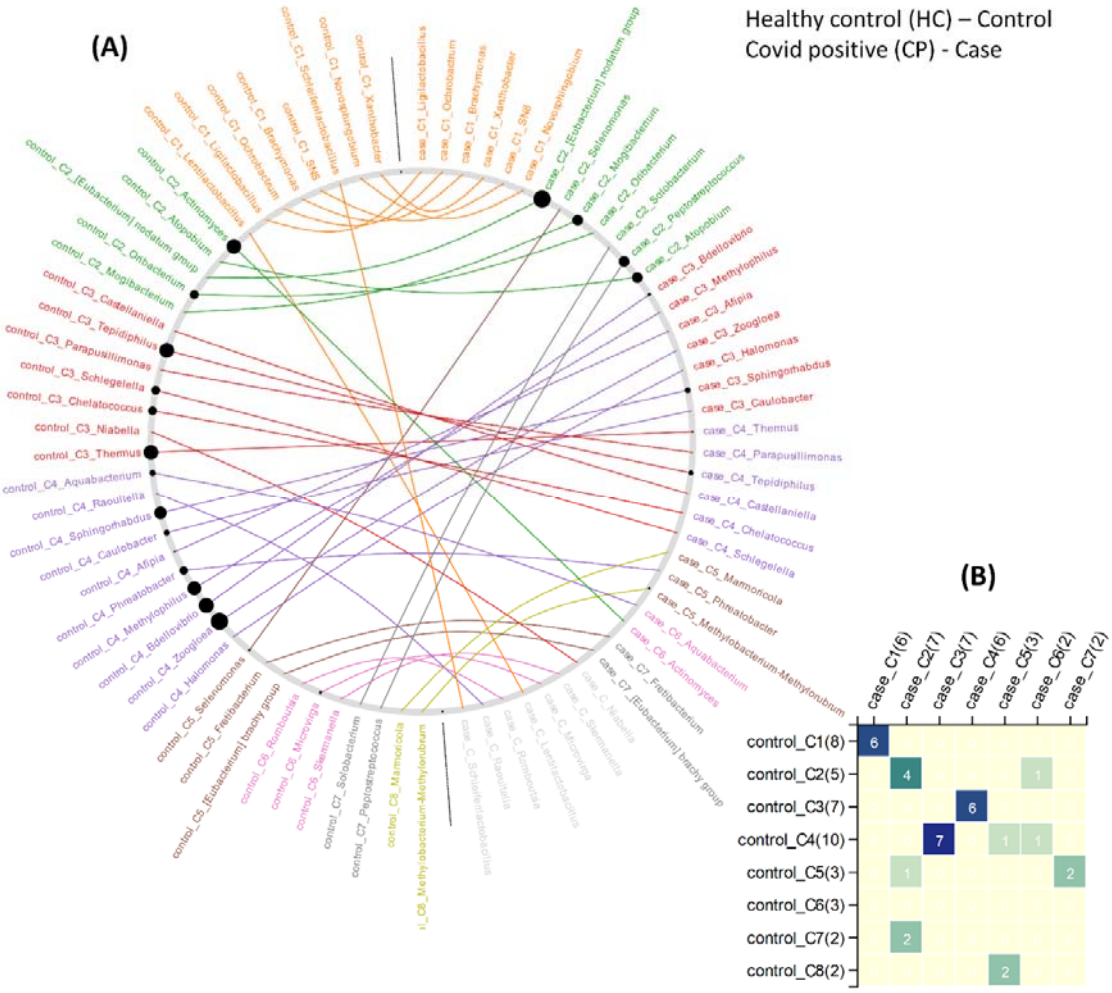

Figure 5: The changes in community structure (community shuffling) between the microbial association networks corresponding to the COVID-19 negative (CN) - 'control' and COVID-19 positive (CP) - 'case' networks. (a) Nodes belonging to the CN - 'control' and CP - 'case' networks are plotted along the left half and right half of the circular frame. Same node (microbe) in the two network is connected by an edge for easy viewing of the community shuffling. Node labels are coloured (at random) based on sub-network/ community affiliations. Grayed out node labels indicate that the node does not interact directly with the common sub-network. The node sizes are proportional to the betweenness centrality measure of the node in the corresponding network. (b) Heatmap representing the communities in the CN - 'control' network (along the vertical axis) and the CP - 'case' network (along the horizontal axis) with counts of the nodes (microbes) in each community in brackets. Numbers (and colour gradient) in the heatmap indicates how the nodes constituting any sub-network/ community of the CN - 'control' network are shared with the sub-networks/ communities of the CP - 'case' network, and vice versa.

Association of inferred functions with microbiota samples during COVID-19

Across the three COVID-19 waves, the metabolic pathway for methanogenesis from acetate (METH-ACETATE-PWY) was found to be significantly enriched in the microbiota corresponding to the CP samples Supplementary Table 6. In contrast, the pathways for anaerobic gondoate biosynthesis (PWY-7663), aerobic respiration I using cytochrome c (PWY-3781), cis-vaccenate biosynthesis (PWY-5973), Kdo transfer to lipid IVA III (PWY-6467), and chlorophyllide a biosynthesis I (CHLOROPHYLL-SYN) appeared to be depleted functions in the CP group as inferred using PICRUSt2. While IgA-specific metalloendopeptidase (EC:3.4.24.13) appeared to be the most enriched enzyme in the CP group, sedolisin (EC:3.4.21.100) was most enriched in the CN group (Supplementary Table 7).

In terms of the functional differences between the CP and CN groups in each of the three COVID-19 waves, while the minimum number of differentially abundant pathways were seen among the first wave samples, the quantum of change among the discriminating pathways was relatively smaller (log2foldchange < ±2) in the third wave samples (Supplementary Table 6). Further, there was no overlap between the metabolic pathways differentially expressed between CP and CN samples in the first two COVID-19 waves with those in the third COVID-19 wave. In the first two COVID-19 waves, spirilloxanthin and 2,2'-diketo-spirilloxanthin biosynthesis pathway (PWY-6581) and coumarins biosynthesis pathway (PWY-7398) were seen to be enriched in the CP group. In contrast, the depleted pathways in the CP group (w.r.t. CN group) included Kdo transfer to lipid IVA III pathway (PWY-6467), L-methionine biosynthesis III pathway (HSERMETANA-PWY), TCA cycle V and VI pathways (PWY-6969 and PWY-5913), aerobic respiration I pathway (PWY-3781), coenzyme B biosynthesis pathway (P241-PWY), and NAD biosynthesis I (from aspartate) pathway (PYRIDNUCSYN-PWY). Notably, of the pathways which were found to be depleted in CP samples from the first two COVID-19 waves, L-methionine biosynthesis III pathway (HSERMETANA-PWY) and TCA cycle V and VI pathways (PWY-6969 and PWY-5913) were found to be marginally enriched in the CP samples from the third COVID-19 wave.

Supplementary Tables 8 & 9 provide the list of functional pathways and enzymes which were found to be differentially enriched among the four disease categories. Irrespective of COVID-19 disease status, the nasopharyngeal microbiota of the symptomatic patients, i.e., sCP and

sCN were seen to be enriched for the following functional pathways w.r.t. the asymptomatic patients – pyrimidine deoxyribonucleotide phosphorylation (PWY-7197), (aerobic) heme biosynthesis I (HEME-BIOSYNTHESIS-II), de novo biosynthesis II of adenosine (PWY-7220) and guanosine (PWY-7222) deoxyribonucleotides, and pyrimidine nucleobases salvage (PWY-7208). The sucrose degradation IV pathway (PWY-5384) was, however, noted to be enriched only in sCN (when compared to aCN) but depleted in sCP (when compared to aCP). 4-hydroxyphenylacetate degradation (3-HYDROXYPHENYLACETATE-DEGRADATION-PWY), and 3-phenylpropanoate degradation (P281-PWY) appeared to be enriched in the microbiota of the CP samples when compared to CN irrespective of symptoms. In contrast, the superpathway of mycolyl-arabinogalactan-peptidoglycan complex biosynthesis (PWY-6404) was found to be enriched in CP (when compared to CN) only in the symptomatic sub-group. Conversely, the pathways for sucrose degradation IV (PWY-5384), O-antigen building blocks biosynthesis (OANTIGEN-PWY), and mixed acid fermentation (FERMENTATION-PWY) were noted to be enriched in the aCP sub-group while compared to aCN.

Discussion

The present study aimed at profiling the nasopharyngeal microbiota of Indian subjects, across the three COVID-19 waves, to identify signatures, if any, which may be distinct between the microbiota sample collected from COVID-19 positive patients (CP) and those tested negative for COVID-19 infection (CN).

Irrespective of the disease status, the nasopharyngeal microbiota in Indian samples was found to be dominated by the phylum Firmicutes, Proteobacteria, and Actinobacteriota and the family Staphylococcaceae and Corynebacteriaceae. This is in line with the reports presented in earlier studies inspecting nasopharyngeal microbiota in other geographies [30], [31], [32]. However, we noted the proportion of Firmicutes to be two to three times higher than Proteobacteria in the nasopharyngeal microbiota in Indian samples, especially in the CN group. Although this concurred with the observations reported in earlier studies from the US and Spain [31] [32], it is in contrast to those observed in Italians wherein Proteobacteria was found to be the most dominant organism [30].

Most earlier studies had shown that the diversity of the microbes constituting the nasopharyngeal microbiota decreased in patients with confirmed COVID-19 infections [17], [30], [23] barring a few exceptions [32]. Our findings also indicate a lower diversity associated with COVID-19, concurring with these findings. An earlier study on a relatively smaller cohort in India had further identified host type (age and gender) and patient condition (symptomatic vs. asymptomatic) as potential factors affecting the enrichment of specific bacterial communities in upper respiratory tract [23]. In the current study, while the condition of the patients (i.e., symptomatic vs. asymptomatic) was found to be associated with bacterial diversity; age and gender did not seem to significantly influence the same (Supplementary Fig. 5). The current study on the other hand found that the diversity of the nasopharyngeal microbiota varied across the three waves (Fig. 2 and Supplementary Fig. 2). The microbes which were found to be distinguishing between CP and CN were also found to vary across the waves in most cases. For instance, while the abundances of Cyanobacteria, Firmicutes, Planctomycetota, Deinococcota and Bdellovibrionota were significantly different between the CP and CN groups in the second wave, Proteobacteria was significantly differing between the third wave CP and CN samples (Supplementary Fig. 5). It may be noted that the depletion of Rickettsiaceae in CP samples was the only consistent pattern that was observed across the three COVID-19 waves at a family level. At an ASV level, this observation was supported by a significant drop in the abundance of ASV36 in CP w.r.t. CN (Supplementary Table 2). To our knowledge, this is the first report on the (negative) association of the abundance of Rickettsiaceae with CP. Further studies into this aspect will be required to see if the presence of Rickettsiaceae in the nasopharyngeal microbiota could provide any advantage in combating the onset/progression of a COVID-19 infection.

Other consistent patterns across the three COVID-19 waves included depletion of the genera *Aliterella* as well as certain ASVs from the genera *Deinococcus* (ASV195) and *Elizabethkingia* (ASV431) in CP. This is consistent with a previous study involving SARS-CoV-2 infected patients which have also shown the absence of microbes belonging to the phylum Deinococcus-Thermus in patients admitted to ICU [33]. It was intriguing to note that despite multiple reports on the risk and incidences of septicemia caused by *Elizabethkingia* in COVID-19 patients [34] [35] the abundance of *Elizabethkingia* was found to be lower in CP compared to CN. While the causation of septicemia in most case has been attributed to *E. meningoseptica* [34] [35] the ASV corresponding to *Elizabethkingia* (ASV431) identified in the current study matched closer to *E. anopheles* (with 100% query coverage and 99.6%

sequence identity) when compared to *E. meningoseptica* (having 100% query coverage and 98.42% sequence identity). Given an earlier report on the contamination of throat swab collection kits with *E. anopheles* we presume that the observed enrichment of *Elizabethkingia* in CN could also be an artifact.

In contrast, *Gemella massiliensis* (ASV161) was enriched in CP across all the three waves. Interestingly there have been contrasting reports regarding the association of *Gemella* with COVID-19 infection. While studies conducted in China had reported a depletion of *Gemella* species like *G. morbillorum* and *G. haemolysans* in swab samples collected from the pharynx of COVID-19 patients [22] a study involving subjects in India had shown an increment in the abundance of *Gemella* to be associated with CP [23]. It is therefore likely that the association of *Gemella* with COVID-19 infection might vary across geographies. The observation pertaining to a lower abundance of genera *Aliterella* in CP across all the three COVID-19 waves was also noteworthy. While none of the previous literature on association of microbiota with COVID-19 reported a similar observation, a few works exploring the antiviral properties of Cyanobacteria drew our attention [36], [37]. Cyanobacteria such as *Aliterella* are a rich source of bioactive compounds and are likely to possess antiviral properties which supports their higher abundance in the CN group.

From the perspective of the potential role of microbiota in the manifestation of COVID-19 infection, our study reports the enrichment of IgA-specific metallo-endopeptidase enzyme (EC:3.4.24.13) in CP samples (Supplementary Table 7 - overall). This metal-dependent enzyme cleaves the Pro-Thr bond in the hinge region of the heavy chain in immunoglobulin A (IgA) and is known to be encoded by pathogenic bacterial groups like *Streptococcus* [38], [39]. Given the role of IgA in mucosal immunity [40] and the reports on the association of IgA with the criticality of COVID-19 disease [41], it appears that CP patients are indeed vulnerable to secondary bacterial infections.

Given the observed variations in the nasopharyngeal microbiota across the COVID-19 waves (Fig. 2, Fig. 3 and Supplementary Fig. 1- 3), we were interested to find which of the changes in the microbiota followed similar trends in both the CP and CN groups. The microbiota abundance signature (at a phylum level) of the first and the third wave seemed to be similar with each other as compared to that of the second wave (Supplementary Fig. 2 & 3). Across both the CP and CN groups, the abundance of the phylum Deinococcota was found to

gradually increase during the course of the pandemic (Supplementary Fig. 3). Although this increase in abundance of Deinococcota was not always statistically significant, the trend was very clear. A number of ASVs (and genera) were also observed to follow a consistent up/ down trend in either CP or CN group across the three waves (Table 2). Most notably, in the CP group, majority of the ASVs which demonstrated a downtrend during the course of the pandemic represented opportunistic pathogens like *Veillonella parvula*, *Rothia mucilaginosa*, *Streptococcus anginosus*, *Acinetobacter baumannii*, etc. In light of the fact that previous studies have reported an increase in the abundance of opportunistic pathogens in the microbiota of COVID-19 patients compared to healthy individuals [22], [16]; this observation might seem counter intuitive. It may be noted that most studies [22], [16], [23] reporting a higher abundance of opportunistic pathogens in microbiota samples from COVID-19 patients were based on data collected during the early part of the pandemic (samples collected in 2020). It is likely that during early days of the pandemic, the human immune system fighting the SARS-CoV-2 virus resulted in an immunocompromised state that could be exploited by the opportunistic pathogens to form secondary infections. The findings by Devi *et al.* [42] regarding the higher transcript abundance of *Achromobacter xylosoxidans* and *Bacillus cereus* in cases of COVID-19 associated mortality, and *Leptotrichia buccalis* in the severe COVID-19 cases highlights the role of co-infecting microbes in the severity and outcome of COVID-19. Thus, despite the observations made in our study that implicated the perturbation of nasopharyngeal microbiota as a result of the viral infection, it cannot be completely ruled out that these changes could be a potential factor aggravating the infection and disease outcome. However, during the later waves, once vaccines against the SARS-CoV-2 virus were introduced, as well as there were many more cases with prior exposure to the pathogen; it is likely that the immune system has become better equipped to deal with the challenges posed by the virus as well as to ward off secondary infections caused by opportunistic pathogens. The lower abundance of bacteria with pathogenic potential like *Streptococcus matis*, *Acinetobacter baumannii*, *Lachnoanaerobaculum orale*, and *Megasphaera micronuciformis* even in the CN group also supports the above notion (Table 2). Indeed, among the genera which were identified to drive the change in microbial network architecture from the first to the second wave (Supplementary Fig. 6), several were noted to be opportunistic pathogens (such as, *Paenibacillus*, *Peptostreptococcus*, and *Solobacterium*) whose abundance decreased in both CP and CN samples from the second wave when compared to the first wave. In addition to vaccination, lower exposure to pollutants due to lockdown and usage of masks as well as practicing home remedies for improving respiratory

health might have also led to the overall decrease in the proportion of opportunistic pathogens in the nasopharyngeal tract of Indians over the course of the pandemic. An exception was noted in the case of the genera *Leptotrichia* which appeared to drive the change in the microbial network architecture (and increased in abundance) from the second to the third wave (Supplementary Fig. 7). Though part of commensal respiratory microbiota, species of the genera *Leptotrichia* genera (eg. *L. buccalis*) possess pathogenic properties and have been shown to have a higher abundance of transcripts in severe COVID-19 patients [42]. In this context, it is also important to note that we did not find any bacterial function which was consistently enriched or depleted in the CP group (w.r.t. CN) across the three COVID-19 waves. As stated above, the changing nature of the underlying microbiota structure through the course of the pandemic due to vaccination, lockdown and/or other interventions might be a reason for this observation. However, we noted an enrichment of the pathway for chitin derivatives degradation (PWY-6906) in sCP w.r.t. sCN. Given the recent interest in the role of chitin derived as anti-viral agents [43], [44] this observation paves way for a deeper introspection into the microbiota's potential in combating viral infection through the production of antiviral metabolites. However, similar observations were not made in the asymptomatic group (aCP w.r.t. aCN).

The wide variation in the symptoms in COVID-19 positive patients led us to investigate whether there were any microbial association with the severity of the infection; i.e., between the symptomatic and the asymptomatic COVID-19 positive sub-groups (sCP; n = 147 vs aCP; n = 157). Our findings (see Table 3) differed from a recently published study by Gupta *et al.* which also inspected the same phenomenon using a smaller cohort of 11 symptomatic and nine asymptomatic COVID-19 positive patients [23]. While this could be due to variations in data analyses protocols, the differences in the number of samples in the two studies might have also influenced the outcome. In this context, we found a set of five ASVs (Table 5) whose abundance seemed to be linked to symptoms in response to infections in the respiratory tract. A thorough investigation into their role in disease manifestation might reveal whether and how these organisms interact with the virus and the host immune system in driving the disease outcome. Details of all the ASVs identified in this study have been provided in Supplementary Table 5.

ASV	baseMean	sCN vs aCN		sCP vs aCP		Phylum	Family
		log2 Fold Change	padj	log2 Fold Change	padj		
ASV244	6.386275	-25.7467	3.08E-10	-34.6709	2.43E-22	Actinobact	Corynebact
ASV107	15.55177	-23.3261	3.14E-06	-35.3004	1.25E-16	Actinobact	Corynebact
ASV4	579.6335	-7.21141	3.71E-05	-4.9978	0.009012	Actinobact	Corynebact
ASV10	148.3991	-3.87589	0.029283	-5.78461	0.000129	Bacteroido	Weeksellac
ASV36	206.5742	10.64582	0.000916	24.7539	1.37E-21	Proteobact	Rickettsiac

Table 5: List of amplicon sequence variants (ASVs) whose abundances were associated with symptoms in response to infections in the respiratory tract. Significant changes having an adjusted p-value < 0.05 have been reported. Positive log2foldchange values indicate a higher abundance of the ASV in symptomatic samples, and vice versa.

In addition, we were also intrigued to find the enrichment of superpathway of mycolyl-arabinogalactan-peptidoglycan (mAGP) complex biosynthesis (PWY-6404) in the sCP when compared to sCN sub-group. Since mAGP complex is best known to be associated with the viability of *Mycobacterium tuberculosis*, we speculate that symptomatic COVID-19 positive (sCP) patients have a likelihood to activate latent tuberculosis infection. Some of the recent reports on also support the notion [45], [46].With the rapid enhancement in microbiome research, it has become evident that the manifestation of a viral infection in the human body is an outcome of a complex interplay between the host, the virus and the resident microbiota [47], [9].

Given the expected variations in the host microbiota across different regions in the world, the present study aimed at presenting the changes in the nasopharyngeal microbiota in COVID-19 positive individuals in an Indian context. This study design also focused on identifying the variations in the microbiota signatures between the symptomatic and asymptomatic individuals and present a set of microbes which might play a role in minimizing the symptoms of a respiratory infection. In addition, possibly for the first time, the study presents the change in the human nasopharyngeal microbiota over the course of the pandemic, which had varied presentation in all the three waves in India, largely caused by different SARS-CoV-2 variants (spanning the first two years). It revealed a perturbation in the microbial diversity and composition with a gradual drop in the abundance of opportunistic pathogens in the human respiratory tract across both the CP and CN groups. Overall, the findings are

expected to enhance the general understanding of the human nasopharyngeal microbiota and aid in devising strategies to ward off other respiratory pathogens in the future.

Conclusions

Most studies on the effect of COVID-19 infection on the human nasopharyngeal microbiota were conducted during the early period of the pandemic. Reports capturing the effect of the changing nature of the virus and/or interventions like vaccination, isolation (due to lockdown) on the host microbiota are largely missing. The current study examined the nasopharyngeal microbiota of COVID-19 positive as well as COVID-19 negative Indians and report the consistent and the varying patterns in the microbiota across the three COVID-19 waves, in terms of microbial diversity, taxonomy and inferred functions. In addition, variations between the microbiota samples collected from symptomatic and asymptomatic patients have also been reported. Whether these changes assist in COVID-19 disease onset and/or progression, would be interesting to explore in future.

Abbreviations

SARS-CoV-2 – Severe Acute Respiratory Syndrome Coronavirus 2

COVID-19 – Coronavirus disease 2019

ACE-2 – Angiotensin Converting Enzyme-2

TMPRSS2 – Transmembrane Serine Protease 2

CP - COVID-19 positive

CN - COVID-19 negative

sCP - symptomatic COVID-19 positive

aCP - asymptomatic COVID-19 positive

aCN - asymptomatic COVID-19 negative

sCN – symptomatic COVID-19 negative

qRT-PCR - Real-Time Quantitative Reverse Transcription PCR

ASV – Amplicon Sequence Variants

Declarations

Ethics approval and consent to participate

This study was conducted for all three waves of COVID-19 in India, from March 2020 to February 2022 in accordance with the guidelines of the Indian Council of Medical Research (ICMR), Government of India and approved by the Institutional Ethics Committee of Centre for Cellular and Molecular Biology (CSIR-CCMB) [IEC-83/2020].

Consent for Publication

All authors have read and approved the paper for submission.

Availability of data and materials

All data generated and analyzed in this study are available within the article and as its Supplementary information files. The datasets supporting the conclusions of this article are available in the National Centre for Biotechnology Information-Sequence Read Archive (SRA) with the BioProject ID PRJNA902495.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors thank TCS CoIN (co-innovation network) programme (CL03) and SBI foundation's CSR Grant (GAP0569) to CSIR-CCMB for supporting the microbiome sequencing experiments performed at CSIR-CCMB, Hyderabad.

Author Contributions

MMH and SSM conceived the idea. ABS, AD, BV, DTS, KBT, MMH, SSM and TB created the study design. ABS, BV, KBT and DTS designed the sequencing protocol. VA and TN performed the wet lab, including sequencing experiments. WU, HK, MR, NKP, TB and JSK performed the bioinformatic analysis, ABS, AD, NKP, TB, VA and WU analyzed the results and wrote the manuscript. ABS and MMH managed the project and oversaw the overall progress. All authors read and approved the final version of the manuscript.

Acknowledgements

Authors also acknowledge the guidance received from Dr. Rakesh Kumar Mishra during the course of the study. We are also grateful to Dr. Shreekant Verma and Dr. Dhiviya Vedagiri for their valuable suggestions during the study and to Venkatesh Madihalli for his cooperation and continued assistance.

References

1. Aleem A, Akbar Samad A B,Slenker A K: Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19). *StatPearls*. Treasure Island (FL); 2022.
2. Abu-Hammad O, Alnazzawi A, Borzangy S S, Abu-Hammad A, Fayad M, Saadaledin S, et al. Factors Influencing Global Variations in COVID-19 Cases and Fatalities; A Review. *Healthcare (Basel).* 2020; 8(3). doi:10.3390/healthcare8030216.
3. Zhao H, Lu L, Peng Z, Chen L L, Meng X, Zhang C, et al. SARS-CoV-2 Omicron variant shows less efficient replication and fusion activity when compared with Delta variant in TMPRSS2-expressed cells. *Emerg Microbes Infect.* 2022; 11(1): 277-283. doi:10.1080/22221751.2021.2023329.
4. Menni C, Valdes A M, Polidori L, Antonelli M, Penamakuri S, Nogal A, et al. Symptom prevalence, duration, and risk of hospital admission in individuals infected with SARS-CoV-2 during periods of omicron and delta variant dominance: a prospective observational study from the ZOE COVID Study. *Lancet.* 2022; 399(10335): 1618-1624. doi:10.1016/S0140-6736(22)00327-0.
5. Ludvigsson J F. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. *Acta Paediatr.* 2020; 109(6): 1088-1095. doi:10.1111/apa.15270.
6. Giacomelli A, Pezzati L, Conti F, Bernacchia D, Siano M, Oreni L, et al. Self-reported Olfactory and Taste Disorders in Patients With Severe Acute Respiratory Coronavirus 2 Infection: A Cross-sectional Study. *Clin Infect Dis.* 2020; 71(15): 889-890. doi:10.1093/cid/ciaa330.
7. Khatiwada S,Subedi A. Lung microbiome and coronavirus disease 2019 (COVID-19): Possible link and implications. *Hum Microb J.* 2020; 17(100073. doi:10.1016/j.humic.2020.100073.
8. Day M. Covid-19: four fifths of cases are asymptomatic, China figures indicate. *BMJ.* 2020; 369(m1375. doi:10.1136/bmj.m1375.
9. Yeoh Y K, Zuo T, Lui G C, Zhang F, Liu Q, Li A Y, et al. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut.* 2021; 70(4): 698-706. doi:10.1136/gutjnl-2020-323020.
10. Zhang H, Ai J W, Yang W, Zhou X, He F, Xie S, et al. Metatranscriptomic Characterization of Coronavirus Disease 2019 Identified a Host Transcriptional Classifier Associated With Immune Signaling. *Clin Infect Dis.* 2021; 73(3): 376-385. doi:10.1093/cid/ciaa663.
11. Wang B, Zhang L, Wang Y, Dai T, Qin Z, Zhou F, and Zhang L. Alterations in microbiota of patients with COVID-19: potential mechanisms and therapeutic interventions. *Signal Transduct Target Ther.* 2022; 7(1): 143. doi:10.1038/s41392-022-00986-0.
12. Zuo T, Zhang F, Lui G C Y, Yeoh Y K, Li A Y L, Zhan H, et al. Alterations in Gut Microbiota of Patients With COVID-19 During Time of Hospitalization. *Gastroenterology.* 2020; 159(3): 944-955 e8. doi:10.1053/j.gastro.2020.05.048.
13. Kazmierczak-Siedlecka K, Vitale E,Makarewicz W. COVID-19 - gastrointestinal and gut microbiota-related aspects. *Eur Rev Med Pharmacol Sci.* 2020; 24(20): 10853-10859. doi:10.26355/eurrev_202010_23448.
14. Zuo T, Liu Q, Zhang F, Lui G C, Tso E Y, Yeoh Y K, et al. Depicting SARS-CoV-2 faecal viral activity in association with gut microbiota composition in patients with COVID-19. *Gut.* 2021; 70(2): 276-284. doi:10.1136/gutjnl-2020-322294.
15. Kumar D, Pandit R, Sharma S, Raval J, Patel Z, Joshi M, and Joshi C G. Nasopharyngeal microbiome of COVID-19 patients revealed a distinct bacterial profile in deceased and recovered individuals. *Microb Pathog.* 2022; 173(Pt A): 105829. doi:10.1016/j.micpath.2022.105829.

16. Hoque M N, Sarkar M M H, Rahman M S, Akter S, Banu T A, Goswami B, et al. SARS-CoV-2 infection reduces human nasopharyngeal commensal microbiome with inclusion of pathobionts. *Sci Rep.* 2021; 11(1): 24042. doi:10.1038/s41598-021-03245-4.
17. Mostafa H H, Fissel J A, 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). doi:10.1128/mBio.01969-20.
18. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell.* 2020; 181(2): 271-280 e8. doi:10.1016/j.cell.2020.02.052.
19. Ou T, Mou H, Zhang L, Ojha A, Choe H,Farzan M. Hydroxychloroquine-mediated inhibition of SARS-CoV-2 entry is attenuated by TMPRSS2. *PLoS Pathog.* 2021; 17(1): e1009212. doi:10.1371/journal.ppat.1009212.
20. Nardelli C, Gentile I, Setaro M, Di Domenico C, Pinchera B, Buonomo A R, 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. doi:10.3389/fcimb.2021.625581.
21. Liu J, Liu S, Zhang Z, Lee X, Wu W, Huang Z, et al. Association between the nasopharyngeal microbiome and metabolome in patients with COVID-19. *Synth Syst Biotechnol.* 2021; 6(3): 135-143. doi:10.1016/j.synbio.2021.06.002.
22. Shi Y L, He M Z, Han M Z, Gui H Y, Wang P, Yu J L, et al. Characterization of Altered Oropharyngeal Microbiota in Hospitalized Patients With Mild SARS-CoV-2 Infection. *Front Cell Infect Microbiol.* 2022; 12(824578. doi:10.3389/fcimb.2022.824578.
23. Gupta A, Karyakarte R, Joshi S, Das R, Jani K, Shouche Y, and Sharma A. Nasopharyngeal microbiome reveals the prevalence of opportunistic pathogens in SARS-CoV-2 infected individuals and their association with host types. *Microbes Infect.* 2022; 24(1): 104880. doi:10.1016/j.micinf.2021.104880.
24. Caporaso J G, Lauber C L, Walters W A, Berg-Lyons D, Lozupone C A, Turnbaugh P J, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A.* 2011; 108 Suppl 1(Suppl 1): 4516-22. doi:10.1073/pnas.1000080107.
25. Kuntal B K, Chandrakar P, Sadhu S,Mande S S. 'NetShift': a methodology for understanding 'driver microbes' from healthy and disease microbiome datasets. *ISME J.* 2019; 13(2): 442-454. doi:10.1038/s41396-018-0291-x.
26. Douglas G M, Maffei V J, Zaneveld J R, Yurgel S N, Brown J R, Taylor C M, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol.* 2020; 38(6): 685-688. doi:10.1038/s41587-020-0548-6.
27. Stein G E,Goldstein E J. Fluoroquinolones and anaerobes. *Clin Infect Dis.* 2006; 42(11): 1598-607. doi:10.1086/503907.
28. Alauzet C, Aujoulat F, Lozniewski A, Ben Brahim S, Domenjod C, Enault C, et al. A New Look at the Genus Solobacterium: A Retrospective Analysis of Twenty-Seven Cases of Infection Involving S. moorei and a Review of Sequence Databases and the Literature. *Microorganisms.* 2021; 9(6). doi:10.3390/microorganisms9061229.
29. Grady E N, MacDonald J, Liu L, Richman A,Yuan Z C. Current knowledge and perspectives of Paenibacillus: a review. *Microb Cell Fact.* 2016; 15(1): 203. doi:10.1186/s12934-016-0603-7.
30. Giugliano R, Sellitto A, Ferravante C, Rocco T, D'Agostino Y, Alexandrova E, et al. NGS analysis of nasopharyngeal microbiota in SARS-CoV-2 positive patients during the first year of the pandemic in the Campania Region of Italy. *Microb Pathog.* 2022; 165(105506. doi:10.1016/j.micpath.2022.105506.
31. Allen E K, Koeppel A F, Hendley J O, Turner S D, Winther B,Sale M M. Characterization of the nasopharyngeal microbiota in health and during rhinovirus challenge. *Microbiome.* 2014; 2(22. doi:10.1186/2049-2618-2-22.

32. Crovetto F, Selma-Royo M, Crispi F, Carbonetto B, Pascal R, Larroya M, et al. Nasopharyngeal microbiota profiling of pregnant women with SARS-CoV-2 infection. *Sci Rep.* 2022; 12(1): 13404. doi:10.1038/s41598-022-17542-z.
33. Rueca M, Fontana A, Bartolini B, Piselli P, Mazzarelli A, Copetti M, et al. Investigation of Nasal/Oropharyngeal Microbial Community of COVID-19 Patients by 16S rDNA Sequencing. *Int J Environ Res Public Health.* 2021; 18(4). doi:10.3390/ijerph18042174.
34. Ong J Y, Wang C H, Tsai Y S, Chen F L, Lee C H,Ou T Y. Nosocomial septicemia in COVID-19 nosocomial K. pneumoniae, A. baumannii, and Elizabethkingia meningoseptica septicemia in a patient of COVID-19. *J Infect.* 2022; 85(1): 90-122. doi:10.1016/j.jinf.2022.04.004.
35. Das A, Kabi S, Kar D,Sahu K K. Prevalence of Elizabethkingia meningoseptica Infections and their Resistant Pattern in Tertiary Care Hospital. *Journal of Pure and Applied Microbiology.* 2022; 16(2): 967-973. doi:10.22207/jpam.16.2.19.
36. Barre A, Van Damme E J M, Simplicien M, Le Poder S, Klonjkowski B, Benoist H, et al. Man-Specific Lectins from Plants, Fungi, Algae and Cyanobacteria, as Potential Blockers for SARS-CoV, MERS-CoV and SARS-CoV-2 (COVID-19) Coronaviruses: Biomedical Perspectives. *Cells.* 2021; 10(7). doi:10.3390/cells10071619.
37. Sami N, Ahmad R,Fatma T. Exploring algae and cyanobacteria as a promising natural source of antiviral drug against SARS-CoV-2. *Biomed J.* 2021; 44(1): 54-62. doi:10.1016/j.bj.2020.11.014.
38. Kornfeld S J,Plaut A G. Secretory immunity and the bacterial IgA proteases. *Rev Infect Dis.* 1981; 3(3): 521-34. doi:10.1093/clinids/3.3.521.
39. Gilbert J V, Plaut A G,Wright A. Analysis of the immunoglobulin A protease gene of Streptococcus sanguis. *Infect Immun.* 1991; 59(1): 7-17. doi:10.1128/iai.59.1.7-17.1991.
40. Chao Y X, Rotzschke O,Tan E K. The role of IgA in COVID-19. *Brain Behav Immun.* 2020; 87(182-183. doi:10.1016/j.bbi.2020.05.057.
41. Zervou F N, Louie P, Stachel A, Zacharioudakis I M, Ortiz-Mendez Y, Thomas K, and Aguero-Rosenfeld M E. SARS-CoV-2 antibodies: IgA correlates with severity of disease in early COVID-19 infection. *J Med Virol.* 2021; 93(9): 5409-5415. doi:10.1002/jmv.27058.
42. Devi P, Maurya R, Mehta P, Shamim U, Yadav A, Chattopadhyay P, et al. Increased Abundance of Achromobacter xylosoxidans and Bacillus cereus in Upper Airway Transcriptionally Active Microbiome of COVID-19 Mortality Patients Indicates Role of Co-Infections in Disease Severity and Outcome. *Microbiol Spectr.* 2022; 10(3): e0231121. doi:10.1128/spectrum.02311-21.
43. Safarzadeh M, Sadeghi S, Azizi M, Rastegari-Pouyani M, Pouriran R,Haji Molla Hoseini M. Chitin and chitosan as tools to combat COVID-19: A triple approach. *Int J Biol Macromol.* 2021; 183(235-244. doi:10.1016/j.ijbiomac.2021.04.157.
44. Ishihara C, Yoshimatsu K, Tsuji M, Arikawa J, Saiki I, Tokura S, and Azuma I. Anti-viral activity of sulfated chitin derivatives against Friend murine leukaemia and herpes simplex type-1 viruses. *Vaccine.* 1993; 11(6): 670-4. doi:10.1016/0264-410x(93)90315-o.
45. Sereda Y, Korotych O, Klimuk D, Zhurkin D, Solodovnikova V, Grzemska M, et al. Tuberculosis Co-Infection Is Common in Patients Requiring Hospitalization for COVID-19 in Belarus: Mixed-Methods Study. *Int J Environ Res Public Health.* 2022; 19(7). doi:10.3390/ijerph19074370.
46. Wang Q, Guo S, Wei X, Dong Q, Xu N, Li H, et al. Global prevalence, treatment and outcome of tuberculosis and COVID-19 coinfection: a systematic review and meta-analysis (from November 2019 to March 2021). *BMJ Open.* 2022; 12(6): e059396. doi:10.1136/bmjopen-2021-059396.
47. Yi H, Yong D, Lee K, Cho Y J,Chun J. Profiling bacterial community in upper respiratory tracts. *BMC Infect Dis.* 2014; 14(583. doi:10.1186/s12879-014-0583-3.

Supplementary Information:

List of Supplementary Figures

Supplementary Figure 1: Venn diagram representing the distribution of 8645 amplicon sequence variants (ASVs) across (a) four disease categories, and (b) three COVID-19 waves.

Supplementary Figure 2: Stacked bar-plot representing the distribution of the bacterial taxonomies across samples. Distribution of bacteria at (a) phylum and (b) family level in COVID-19 positive (CP) and COVID-19 negative (CN) samples across the three COVID-19 waves. Distribution of bacteria at (c) phylum and (d) family level between the four sample sub-groups: asymptomatic COVID-19 positive (aCP), symptomatic COVID-19 positive (sCP), apparently healthy controls who are COVID-19 negatives as well as asymptomatic (aCN), and COVID-19 negatives exhibiting symptoms (sCN).

Supplementary Figure 3: Differential abundance of bacterial phylum between the COVID-19 positive (CP) and COVID-19 negative (CN) samples. The log2fold change in the mean abundance (along with whiskers representing standard errors) of a bacterial phylum in CP with respect to CN is depicted in (a) all the analyzed samples (overall) as well as in (b-d) each of the three COVID-19 waves. Significantly different abundance (q-value < 0.05) is indicated with blue colour.

Supplementary Figure 4: Differential abundance of bacterial phylum between the COVID-19 positive (CP) and COVID-19 negative (CN) samples across the three COVID-19 waves. The log2fold change in the mean abundance (along with whiskers representing standard errors) of a bacterial phylum in the second wave with respect to the first wave in (a) CN samples, (b) CP samples; third wave with respect to the first wave in (c) CN samples, (d) CP samples; third wave with respect to the second wave in (e) CN samples, (f) CP samples. Significantly different abundance (q-value < 0.05) is indicated with blue colour.

Supplementary Figure 5: Differential abundance of bacterial phylum between the four sub-group of samples - asymptomatic COVID-19 positive (aCP), symptomatic COVID-19 positive (sCP), apparently healthy controls who are COVID-19 negatives as well as asymptomatic (aCN), and COVID-19 negatives exhibiting symptoms (sCN). The log2fold change in the mean abundance (along with whiskers representing standard errors) of a bacterial phylum in (a) aCP with respect to aCN, (b) sCN with respect to aCN, (c) sCP with respect to sCN, and (d) sCP with respect to aCP is depicted. Significantly different abundance (q-value < 0.05) is indicated with blue colour.

Supplementary Figure 6: Distribution of betweenness centralities of the nodes (microbes) in each of the analyzed microbial association network.

Supplementary Figure 7: The changes in community structure (community shuffling) between the microbial association networks corresponding to the first COVID-19 wave - ‘control’ and the second COVID-19 wave - ‘case’ networks. Nodes belonging to the ‘control’ and ‘case’ networks are plotted along the left half and right half of the circular frame. Same node (microbe) in the two network is connected by an edge for easy viewing of the community shuffling. Node labels are coloured (at random) based on sub-network/ community affiliations. Grayed out node labels indicate that the node does not interact directly with the common sub-network. The node sizes are proportional to the betweenness centrality measure of the node in the corresponding network.

Supplementary Figure 8: The changes in community structure (community shuffling) between the microbial association networks corresponding to the second COVID-19 wave - ‘control’ and the third COVID-19 wave - ‘case’ networks. Nodes belonging to the ‘control’ and ‘case’ networks are plotted along the left half and right half of the circular frame. Same node (microbe) in the two network is connected by an edge for easy viewing of the community shuffling. Node labels are coloured (at random) based on sub-network/ community affiliations. Grayed out node labels indicate that the node does not interact directly with the common sub-network. The node sizes are proportional to the betweenness centrality measure of the node in the corresponding network.

Supplementary Figure 9: The changes in community structure (community shuffling) between the microbial association networks corresponding to the samples from asymptomatic individuals - ‘control’ and the samples from symptomatic individuals - ‘case’ networks. Nodes belonging to the ‘control’ and ‘case’ networks are plotted along the left half and right half of the circular frame. Same node (microbe) in the two network is connected by an edge for easy viewing of the community shuffling. Node labels are coloured (at random) based on sub-network/ community affiliations. Grayed out node labels indicate that the node does not interact directly with the common sub-network. The node sizes are proportional to the betweenness centrality measure of the node in the corresponding network.

List of Supplementary Tables

Supplementary Table 1: Statistics on the amplicon sequencing runs performed in this study.

Supplementary Tables 2: Discriminatory taxonomic groups identified at a ASV and genera level between COVID-19 positive (CP) and COVID-19 negative (CN) samples across the three COVID-19 waves.

Supplementary Tables 3: Discriminatory taxonomic groups identified at a ASV and genera level between the three COVID-19 waves among samples from the COVID-19 positive (CP) and COVID-19 negative (CN) groups.

Supplementary Tables 4: Discriminatory taxonomic groups identified at a ASV and genera level between the four sample sub-groups based on symptoms and disease status - asymptomatic COVID-19 positive (aCP), symptomatic COVID-19 positive (sCP), apparently healthy controls who are COVID-19 negatives as well as asymptomatic (aCN), and COVID-19 negatives exhibiting symptoms (sCN).

Supplementary Tables 5: Table of 'Neighbor Shift (NESH) Score' corresponding to Figure 5 and Supplementary Figures 7-9.

Supplementary Tables 6: Discriminatory functional pathways identified between COVID-19 positive (CP) and COVID-19 negative (CN) samples across the three COVID-19 waves using inferred functional profiles obtained through PICRUSt2 analysis.

Supplementary Tables 7: Discriminatory enzyme functional potentials identified between COVID-19 positive (CP) and COVID-19 negative (CN) samples across the three COVID-19 waves using inferred functional profiles obtained through PICRUSt2 analysis.

Supplementary Tables 8: Discriminatory functional pathways identified between the four sample sub-groups based on symptoms and disease status - asymptomatic COVID-19 positive (aCP), symptomatic COVID-19 positive (sCP), apparently healthy controls who are COVID-19 negatives as well as asymptomatic (aCN), and COVID-19 negatives exhibiting symptoms (sCN), using inferred functional profiles obtained through PICRUSt2 analysis.

Supplementary Tables 9: Discriminatory enzyme functional potentials identified between the four sample sub-groups based on symptoms and disease status - asymptomatic COVID-19 positive (aCP), symptomatic COVID-19 positive (sCP), apparently healthy controls who are COVID-19 negatives as well as asymptomatic (aCN), and COVID-19 negatives exhibiting symptoms (sCN), using inferred functional profiles obtained through PICRUSt2 analysis.

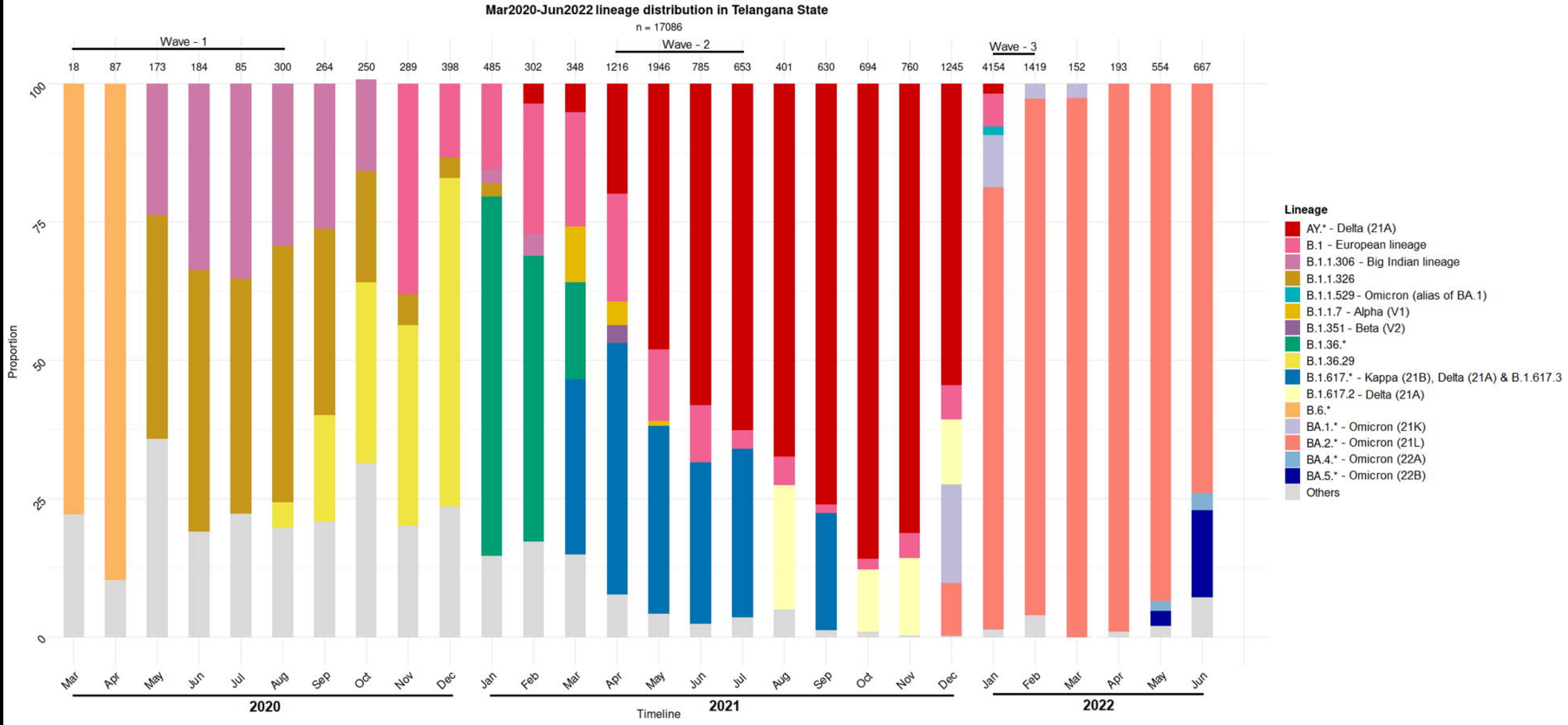

Figure 1: Distribution of the SARS-CoV-2 lineage among the samples collected from COVID-19-positive individuals. The nasopharyngeal samples were collected in the state of Telangana (India) between March 2020 and June 2022.

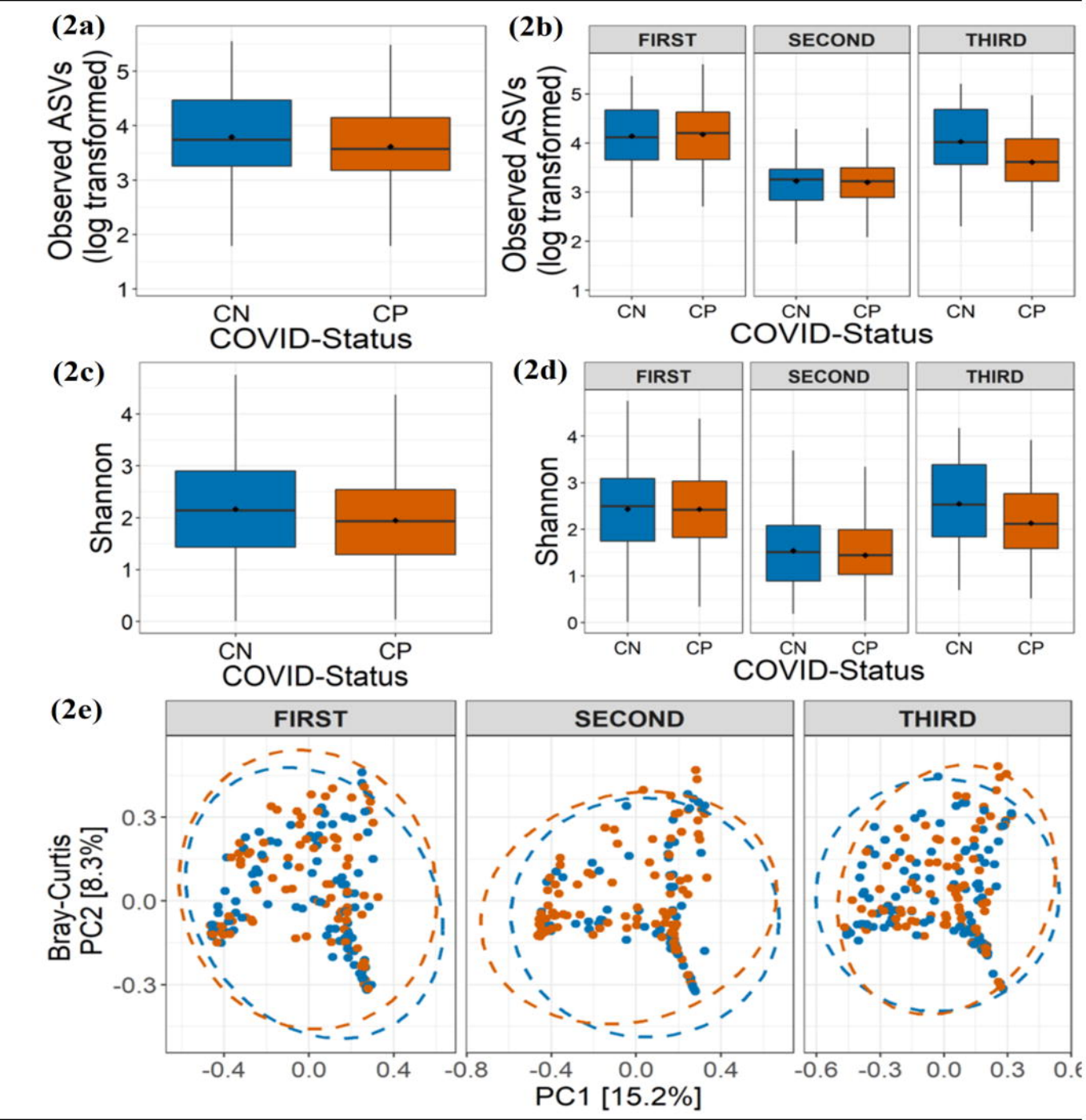

Figure 2: Alpha and Beta diversity of the nasopharyngeal microbiota in COVID-19 positive (CP) and COVID-19 negative (CN) samples. Number of the observed amplicon sequence variants (ASVs) (a) *in all the analyzed samples* (overall) and (b) in the three COVID-19 waves are shown. Shannon Diversity in terms of ASVs (c) *in all the analyzed samples,* and (d) in the three COVID-19 waves are depicted. (e) Beta diversity of microbiota assessed using Principal component analysis (Bray-Curtis distance) represented along the first two principal components for the three COVID-19 waves are presented.

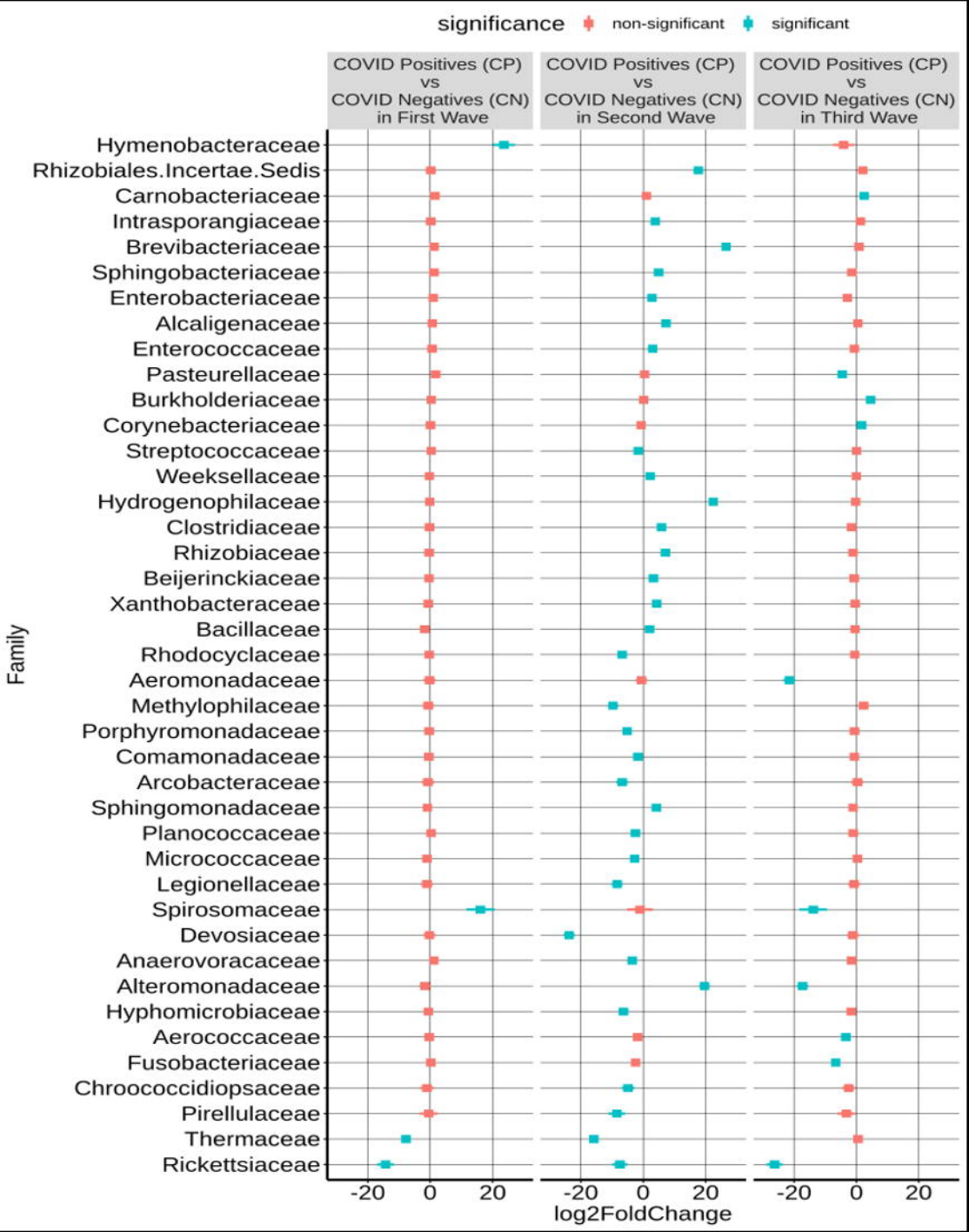

Figure 3: Differential abundance of bacterial families between the COVID-19 positive (CP) and COVID-19 negative (CN) samples. The log2fold change in the mean abundance (along with whiskers representing standard errors) of a bacterial family in CP with respect to CN is depicted in (a) all the analyzed samples (overall) as well as in (b-d) each of the three COVID-19 waves. Significantly different abundance (q-value < 0.05) is indicated with blue colour.

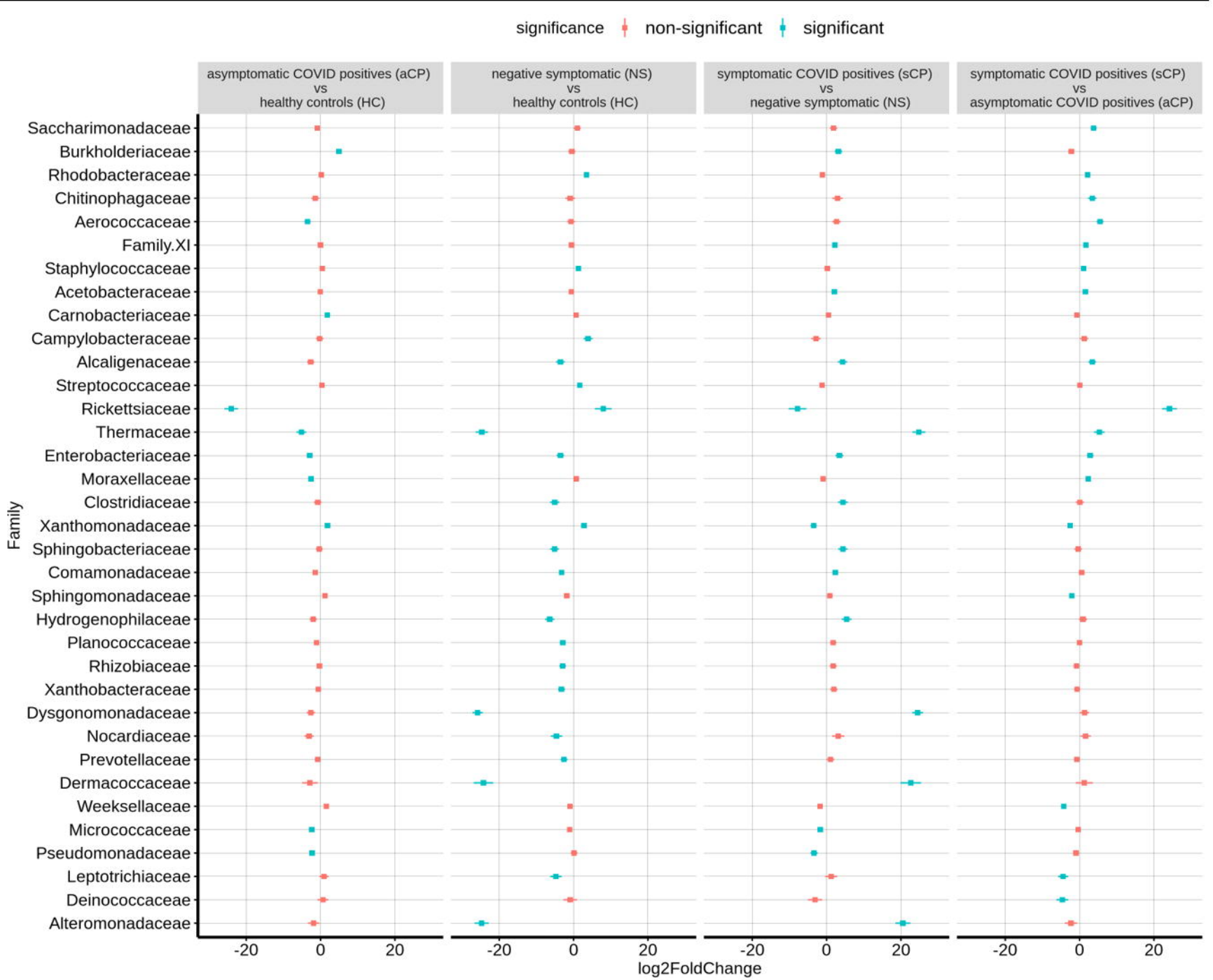

Figure 4: Differential abundance of bacterial families between the four sub-group of samples - asymptomatic COVID-19 positive (aCP), symptomatic COVID-19 positive (sCP), apparently healthy controls who are COVID-19 negatives as well as asymptomatic (aCN), and COVID-19 negatives exhibiting symptoms (sCN). The log2fold change in the mean abundance (along with whiskers representing standard errors) of a bacterial family in (a) aCP with respect to aCN, (b) sCN with respect to aCN, (c) sCP with respect to sCN, and (d) sCP with respect to aCP is depicted. Significantly different abundance (q-value < 0.05) is indicated with blue colour.

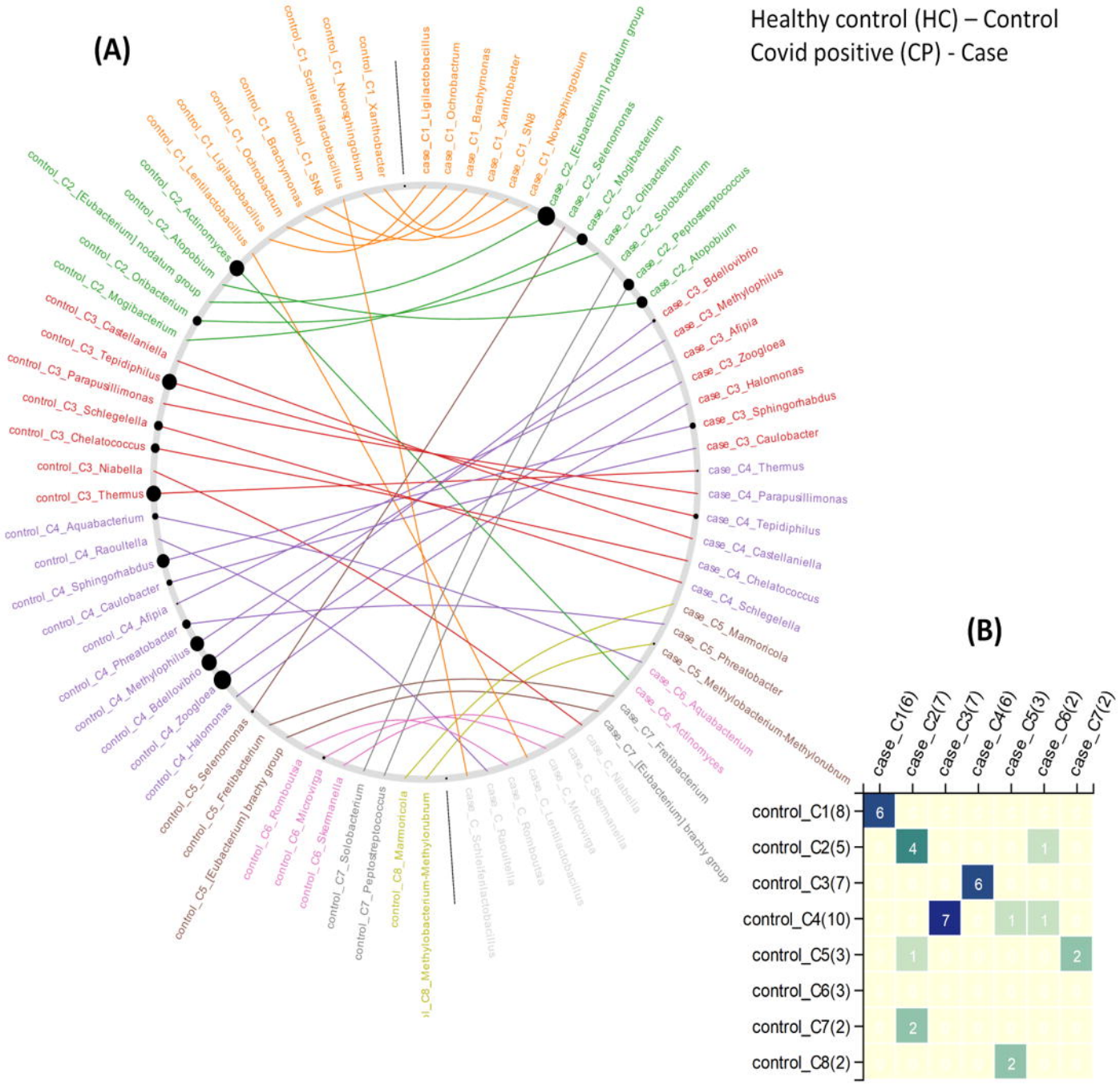

Figure 5: The changes in community structure (community shuffling) between the microbial association networks corresponding to the COVID-19 negative (CN) - 'control' and COVID-19 positive (CP) - 'case' networks. (a) Nodes belonging to the CN - 'control' and CP - 'case' networks are plotted along the left half and right half of the circular frame. Same node (microbe) in the two network is connected by an edge for easy viewing of the community shuffling. Node labels are coloured (at random) based on sub-network/ community affiliations. Grayed out node labels indicate that the node does not interact directly with the common sub-network. The node sizes are proportional to the betweenness centrality measure of the node in the corresponding network. (b) Heatmap representing the communities in the CN - 'control' network (along the vertical axis) and the CP - 'case' network (along the horizontal axis) with counts of the nodes (microbes) in each community in brackets. Numbers (and colour gradient) in the heatmap indicate how the nodes constituting any sub-network/ community of the CN - 'control' network are shared with the sub-networks/ communities of the CP - 'case' network, and vice versa.

	Covid Positives (CP)		Covid Negatives (CN)	
Category	with symptoms (sCP)	without symptoms (aCP)	with symptoms (sCN)	apparently healthy controls (aCN)
Total	147	157	81	204
Wave1	19	70	12	80
Wave2	81	39	60	37
Wave3	47	48	9	87

Table 1: Statistics on the number of samples analyzed in the study. Categorization of 589 analyzed samples into sub-groups and their distribution across the three COVID-19 waves have been provided.

COVID-19 Positive group

ASV/ Genus	base Mean	Wave 3 vs Wave 2		Wave 2 vs Wave 1		Phylum	Family	Genus	Species
		log2 Fold Change	padj	log2 Fold Change	padj				
ASV22	2229.73	36.39	0	4.96	0.004	Proteobacteria	Burkholderiaceae	*Burkholderia-Caballeronia-Paraburk*	#N/A
ASV54	218.86	-23.34	0	-4.34	0.014	Firmicutes	Veillonellaceae	*Veillonella*	*parvula*
ASV90	94.61	-31.89	0	-6.1	0.008	Actinobacteriota	Micrococcaceae	*Rothia*	*mucilaginosa*
ASV79	63.9	-23.75	0	-7.47	0.001	Firmicutes	Streptococcaceae	*Streptococcus*	*anginosus*
ASV97	61.34	-22.32	0	-8.11	0	Firmicutes	Streptococcaceae	*Streptococcus*	*anginosus*
ASV80	51.09	-20.14	0	-8.6	0	Bacteroidota	Porphyromonadaceae	*Porphyromonas*	*pasteri*
ASV53	45.31	-29.74	0	-6.78	0	Proteobacteria	Moraxellaceae	*Acinetobacter*	*baumannii*
ASV135	38.35	-24	0	-5.91	0.005	Firmicutes	Anaerovoracaceae	*[Eubacterium] brachy group*	#N/A
ASV78	21.02	-15.59	0	-10.35	0	Proteobacteria	Rhizobiales Incertae Sedis	*Phreatobacter*	#N/A
ASV206	19.01	-22.96	0	-5.67	0.006	Firmicutes	Streptococcaceae	*Streptococcus*	#N/A
ASV255	17.98	-27.66	0	-8.86	0.001	Firmicutes	Lachnospiraceae	#N/A	#N/A
ASV330	17.36	-21.87	0	-6.18	0.019	Firmicutes	Lachnospiraceae	*Oribacterium*	*parvum*
ASV186	14.78	-34.09	0	-6.58	0.002	Proteobacteria	Xanthobacteraceae	*Afipia*	#N/A
ASV345	13.48	-21.39	0	-6.74	0.001	Firmicutes	Anaerovoracaceae	*[Eubacterium] nodatum group*	#N/A
ASV420	11.7	-29.13	0	-5.59	0.042	Firmicutes	Veillonellaceae	*Veillonella*	#N/A
ASV143	8.3	-21.98	0	-5.72	0.007	Proteobacteria	Pasteurellaceae	*Haemophilus*	#N/A
ASV37	7.7	-17.19	0	-8.1	0.003	Firmicutes	Streptococcaceae	*Streptococcus*	anginosus
Streptococcus	3262.93	-1.94	0.015	-1.57	0.042	Firmicutes	Streptococcaceae	*Streptococcus*	#N/A
X.Eubacterium.. brachy.group	16.22	-19.19	0	-8.88	0	#N/A	#N/A	#N/A	#N/A
Abiotrophia	14.95	-20.77	0	-6.44	0.001	Firmicutes	Aerococcaceae	*Abiotrophia*	#N/A
Rheinheimera	8.46	-34.41	0	-5.01	0.011	Proteobacteria	Alteromonadaceae	*Rheinheimera*	#N/A
Escherichia.Shigella	5.5	-37.8	0	-5.31	0.005	#N/A	#N/A	#N/A	#N/A

COVID-19 Negative group

ASV/ Genus	base Mean	Wave 3 vs Wave 2		Wave 2 vs Wave 1		Phylum	Family	Genus	Species
		log2 Fold Change	padj	log2 Fold Change	padj				
ASV24	96.08	-16.68	0	-11.19	0	Firmicutes	Streptococcaceae	*Streptococcus*	*mitis*
ASV207	47.7	4.79	0	3.77	0.012	Proteobacteria	Comamonadaceae	*Schlegelella*	*aquatica*
ASV53	45.31	-23.72	0	-4.61	0.001	Proteobacteria	Moraxellaceae	*Acinetobacter*	*baumannii*
ASV134	11.26	-20.36	0	-6.61	0.037	Firmicutes	Lachnospiraceae	*Lachnoanaerobaculum*	*orale*
ASV404	6.98	-17.56	0	-7.66	0.01	Firmicutes	Veillonellaceae	*Megasphaera*	*micronuciformis*
Acinetobacter	146.43	-2.6	0.002	-2.28	0.011	Proteobacteria	Moraxellaceae	*Acinetobacter*	#N/A
Schlegelella	31.81	4.3	0.001	3.33	0.015	Proteobacteria	Comamonadaceae	*Schlegelella*	#N/A
Megasphaera	7.33	-17.02	0	-8.89	0	Firmicutes	Veillonellaceae	*Megasphaera*	#N/A

Table 2: List of the amplicon sequence variants (ASVs) and genera demonstrating a consistent increase or decrease in abundance across the three COVID-19 waves. Significant changes having an adjusted p-value < 0.05 have been reported.

Genus	base Mean	log2 Fold Change	padj	Phylum	Family	Genus
Schlegelella	22.58	26.57	0	Proteobacteria	Comamonadaceae	Schlegelella
Proteiniphilum	10.27	24.66	0	Bacteroidota	Dysgonomonadaceae	Proteiniphilum
Ochrobactrum	68.6	26.75	0	Proteobacteria	Rhizobiaceae	Ochrobactrum
Brachymonas	15.31	24.83	0	Proteobacteria	Comamonadaceae	Brachymonas
Parapusillimonas	7.64	25.07	0	Proteobacteria	Alcaligenaceae	Parapusillimonas
Thermus	26.63	24.5	0	Deinococcota	Thermaceae	Thermus
Lactiplantibacillus	4.19	23.81	0	Firmicutes	Lactobacillaceae	Lactiplantibacillus
Castellaniella	5.29	24.83	0	Proteobacteria	Alcaligenaceae	Castellaniella
Aquabacterium	4.03	23.55	0	Proteobacteria	Comamonadaceae	Aquabacterium
Anoxybacillus	3.65	-26.26	0	Firmicutes	Bacillaceae	Anoxybacillus
Cnuella	3.14	25.2	0	Bacteroidota	Chitinophagaceae	Cnuella
Anaerococcus	1064.32	3.67	0	Firmicutes	Family XI	Anaerococcus
Finegoldia	632.01	3.74	0	Firmicutes	Family XI	Finegoldia
Schleiferilactobacillus	84.1	5.71	0	Firmicutes	Lactobacillaceae	Schleiferilactobacillus
Enhydrobacter	161.48	-4.91	0	Proteobacteria	Moraxellaceae	Enhydrobacter
Peptoniphilus	820.56	3.01	0	Firmicutes	Family XI	Peptoniphilus
SN8	164.12	5.91	0.001	Proteobacteria	Xanthomonadaceae	SN8
Lentilactobacillus	119.54	4.19	0.002	Firmicutes	Lactobacillaceae	Lentilactobacillus
Burkholderia.Caballeronia. Paraburkholderia	6130.46	6.3	0.002	#N/A	#N/A	#N/A
Brevundimonas	38.97	4.26	0.002	Proteobacteria	Caulobacteraceae	Brevundimonas
Lactococcus	12.44	5.06	0.002	Firmicutes	Streptococcaceae	Lactococcus
Tepidiphilus	53.54	4.8	0.003	Proteobacteria	Hydrogenophilaceae	Tepidiphilus
Stenotrophomonas	382.15	-4.83	0.006	Proteobacteria	Xanthomonadaceae	Stenotrophomonas
Acetobacter	1443.3	1.97	0.007	Proteobacteria	Acetobacteraceae	Acetobacter
Novosphingobium	66.23	4.12	0.009	Proteobacteria	Sphingomonadaceae	Novosphingobium
Clostridium.sensu.stricto.1	14.74	4.78	0.009	#N/A	#N/A	#N/A
Prevotella	15.18	4.44	0.01	Bacteroidota	Prevotellaceae	Prevotella
Ligilactobacillus	12.61	4.5	0.011	Firmicutes	Lactobacillaceae	Ligilactobacillus
Pseudomonas	104.23	-3.07	0.011	Proteobacteria	Pseudomonadaceae	Pseudomonas
Actinomyces	240.22	-2.61	0.028	Actinobacteriota	Actinomycetaceae	Actinomyces
Brucella	32.64	4.28	0.035	Proteobacteria	Rhizobiaceae	Brucella
Exiguobacterium	36.82	-5.75	0.038	Firmicutes	Exiguobacteraceae	Exiguobacterium
Campylobacter	221.31	-3.43	0.043	Campylobacterota	Campylobacteraceae	Campylobacter

Tables 3: List of the bacterial genera which showed significant changes in abundance between symptomatic COVID-19 positive (sCP) and asymptomatic COVID-19 positive (aCP) sub-groups.

Significant changes having an adjusted p-value < 0.05 have been reported. Positive log2foldchange values indicate a higher abundance in sCP while negative log2foldchange values represented genera with higher abundance in aCP.

Network	No of Nodes	No of Edges	Avg No of Neighbours	Characteristic Path Length	Clustering Coefficient	Network Density
Wave 1	85	278	5.5	1.808	0.774	0.367
Wave 2	56	92	3.33	1.4	0.667	0.667
Wave 3	92	486	4.7	2.284	0.586	0.247
Symptomatic	49	128	5.75	1.179	0.94	0.821
Asymptomatic	63	214	7.833	1.303	0.854	0.712
Covid Positive	69	214	2	3.091	0.091	0.2
Covid Negative	75	210	4.727	1.673	0.596	0.473

Table 4: Properties of the microbial association networks analyzed in this study.

		sCN vs aCN		sCP vs aCP		sCP vs sCN					
ASV	**baseMean**	**log2 Fold Change**	**padj**	**log2 Fold Change**	**padj**	**log2 Fold Change**	**padj**	**Phylum**	**Family**	**Genus**	**Species**
ASV55	121.265	-30.2895	3.64E-09	-21.8624	2.05E-06	-30.2895	3.64E-09	Proteobacteria	Xanthomonadaceae	*Stenotrophomonas*	*maltophilia*
ASV234	5.138502	-23.0848	6.74E-08	-26.832	4.44E-14	-23.0848	6.74E-08	Firmicutes	Bacillaceae	*Bacillus*	#N/A
ASV244	6.386275	-12.4518	0.011673	-34.6709	2.43E-22	-12.4518	0.011673	Actinobacteriota	Corynebacteriaceae	*Corynebacterium*	#N/A
ASV355	11.7301	13.13031	1.48E-07	14.30809	6.64E-12	13.13031	1.48E-07	Proteobacteria	Comamonadaceae	#N/A	#N/A
ASV108	34.01354	24.63862	9.03E-35	5.152616	0.036921	24.63862	9.03E-35	Deinococcota	Thermaceae	*Thermus*	*amyloliquefaciens*

Table 5: List of amplicon sequence variants (ASVs) whose abundances were associated with symptoms in response to infections in the respiratory tract.

Significant changes having an adjusted p-value < 0.05 have been reported. Positive log2foldchange values indicate a higher abundance of the ASV in symptomatic samples, and vice versa.