Nasopharyngeal Microbiota in SARS-CoV-2 Positive and Negative Patients

Phillip A. Engen
Rush University Medical Center
Ankur Naqib
Rush University Medical Center
Cheryl Jennings
Rush University Medical Center
Stefan J. Green
Rush University Medical Center
Alan Landay
Rush University Medical Center
Ali Keshavarzian
Rush University Medical Center
Robin M. Voigt (robin_voigt@rush.edu)
Rush University Medical Center

Short Report

Keywords: SARS-CoV-2, COVID-19, nasopharyngeal, microbiota, dysbiosis, viral transport media

DOI: https://doi.org/10.21203/rs.3.rs-334241/v1

Abstract

We investigated nasopharyngeal microbial community structure in COVID-19-positive and -negative patients. High-throughput 16S ribosomal RNA gene amplicon sequencing revealed microbial community structure differences between COVID-19-positive and -negative patients. This proof-of-concept study demonstrates that: (1) nasopharyngeal microbiome communities can be assessed using collection procedures for SARS-CoV-2 testing and (2) SARS-CoV-2 infection is associated with altered dysbiotic microbial profiles which could be a biomarker for disease progression and prognosis in SARS-CoV-2.

Main

Since the appearance of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in 2019, cases of coronavirus disease 2019 (COVID-19) have rapidly spread around the world infecting over 113 million people and claimed over 2.5 million lives, as of March 1, 2021 (https://covid19.who.int/). COVID-19 is transmitted through the respiratory tract via aerosolized droplets containing viral particles[1]. Severity of COVID-19 is dictated by immune-inflammatory response of the host to the virus (i.e., cytokine storm). Host-associated microorganisms can influence viral infectivity[2] and is the major player in mediating immune-inflammation[3]. Conversely, studies have shown that viruses can modulate microbiota community in the oropharyngeal and respiratory tract[4]. However, few studies have examined the nasopharyngeal microbiota in COVID-19 patients. One study found no significant microbial diversity outcomes[5], while another study found a significant decrease in microbial diversity, differences between microbial communities, and a higher abundance of species *Propionibacteriaceae* and a reduction in *Corynebacterium accolens* in COVID-19-positive relative to -negative patients[6]. This proof-of-concept study aimed to identify nasopharyngeal microbiota alterations in COVID-19-positive and –negative patients to determine if: (1) nasopharyngeal swab samples collected for PCR testing of SARS-CoV-2 and the associated viral transport media (VTM) are suitable for microbiome analysis and (2) COVID-19-positive and -negative patients have dissimilar nasopharyngeal dysbiotic microbial communities.

We examined COVID-19-positive (n = 9) and -negative (n = 10) subjects recruited from Rush University Medical Center (RUMC) located in Chicago, IL in the early stage of the pandemic (April 2020). The nine COVID-19-positive subjects had mild COVID-19 with no hospitalizations or deaths reported (**Table S1**). The study was approved by RUMC Institutional Review Board [COVID-19 Biorepository (ORA #20032309)], to use random de-identified remnant specimens, from standard care testing of COVID-19, with no a priori exclusion criteria.

Analysis of communities using 16S ribosomal RNA gene amplicon sequencing revealed that microbial alpha-diversity indices were not significantly different between patient groups, but taxonomic feature richness was lower in COVID-19-positive compared to -negative patients (**Table S2**). Significant differences (PERMANOVA: q = 0.016) in nasopharyngeal microbial community structure were observed between positive and negative patients in beta diversity analyses conducted on microbial features (i.e., amplicon sequence variant, ASV) (**Figure 1A**; **Table S3**). Microbial communities in both groups were

dominated by the phyla Proteobacteria, Actinobacteria and Firmicutes, and the genera *Corynebacterium*, *Morganella*, *Moraxella*, *Escherichia-Shigella*, *Proteus*, and *Staphylococcus* (50% of all sequences; **Figure 1B-C**; **Table S4**). The microbial community in COVID-19-positive patients can be characterized as pro-inflammatory as exemplified by significantly higher (Mann-Whitney test: p = 0.0002) Proteobacteria-to-Actinobacteria ratio relative to COVID-19-negative patients (**Figure 1D**). Between-group differences in taxon abundances were identified using ANCOM (**Table 1**) and DeSeq2 (**Tables S5, S6**). ANCOM identified the genus *Streptococcus* as significantly less abundant in COVID-19-positive relative to -negative patients, with nine genera trending towards significance (**Table 1; Figure 1E**). DESeq2 analysis identified 25 significantly differentially abundant genera, including lower abundances of *Rothia* and *Prevotella* in COVID-19-positive patients (**Table S6**). A machine learning approach (Boruta) was employed for feature selection to identify taxa driving differences between COVID-19-positive and -negative patients. This analysis identified nine genera, many of which were also identified in DESeq2 and ANCOM analyses (**Figure 2**).

This study provides evidence that current biorepositories with nasopharyngeal VTM can be utilized to assess microbiome communities. Importantly, this proof-of-concept study is the first to show that COVID-19-positive patients have a dysbiotic “pro-inflammatory” nasopharyngeal microbial community characterized by loss of putative nasal commensal bacteria and an increase in putative pro-inflammatory bacteria. Specifically, this study revealed: (1) a characteristic nasal microbiome in COVID-19-positive individuals with reduced microbial diversity, (2) an increased relative abundance of multiple putative pathogenic Proteobacteria taxa, and (3) decreased relative abundance of nasal commensal organisms such as *Corynebacterium* (Actinobacteria)[7] and *Streptococcus* (Firmicutes), previously shown to be affected by influenza virus infection[8]. These differences need to be further evaluated in future studies, but could reflect differences prior to infection leading to susceptibility to SARS-CoV-2 infection or contributing to COVID-19 severity or may be a result of post-infection microbiome alteration.

In conclusion, this proof-of-concept study provides preliminary data suggesting the presence of a dysbiotic and pro-inflammatory nasopharyngeal microbiota in COVID-19-positive patients. The relatively small sample size may have limited our ability to identify additional significant differences between groups. Nonetheless, our study provides a strong scientific rationale for future studies to investigate the relationship between nasal microbiome and SARS-CoV-2 infection and COVID-19 severity, and also the relationship to the long-lasting effects of COVID-19.

Methods

Sample Collection, Extraction and Sequencing

The feasibility of using nasopharyngeal swabs (NPS) and viral transport media (VTM) collection for SARS-CoV-2 detection and microbiome analysis has been previously reported[9]. The NPS were collected according to CDC guidelines. Sterile synthetic-head, plastic-shaft swabs were used to collect specimens for diagnostic testing and were placed into properly labeled collection tubes containing 3mL of VTM

[REMEL Micro Test™ M4RT®]. The NPS were sent to the RUMC clinical microbiology laboratory and heat inactivated [65˚C for (30) minutes], prior to RT-PCR testing on the Abbott *m*2000 device [Abbott Laboratories][10]. Due to the surge in testing and specimen processing backlog, in the early stage of the pandemic, the NPS tubes were stored at 4˚C for up to nine days prior to processing for biorepository storage. The NPS+VTM tubes were vortexed briefly for (3) seconds and the VTM was aliquoted into multiple 250ul aliquots, and stored at -80˚C.

Isolation of viral nucleic acids and bacterial DNA from nasal swab VTM (200ul) samples were performed using the NucleoMag Pathogen manufacturer's protocol (Macherey-Nagel, Duran, Germany). Microbiome characterization was performed using a PCR-next-generation sequencing (NGS) approach with a two-stage PCR protocol, as described previously[11]. The V4 variable region of microbial 16S rRNA genes was amplified with the 515F/806R primer set (515F:GTGYCAGCMGCCGCGGTAA; 806R:GGACTACNVGGGTWTCTAAT)[12] and using Fluidigm Access Array primers for Illumina sequencers[12]. Negative controls (*i.e.*, PCR reagent blanks; n = 5) were amplified and sequenced with samples. Sequencing was performed using an Illumina MiniSeq with a mid-output kit and paired-end 153 base reads.

16S rRNA V4 Sequencing Analysis

Raw sequences obtained from the sequencer were merged using the PEAR (Paired-End read merger) algorithm (v0.9.11)[13]. Merged sequences were then quality filtered and denoised using the DADA2 algorithm within the QIIME2 (v 2020.8.0) workflow[14,15]. Amplicon sequence variants (ASVs) were generated and utilized for all downstream analyses. Taxonomy was assigned to ASVs by using the naïve Bayes taxonomy classifier trained with the SILVA 138 99% OTU database[16,17]. A total of 527,628 sequencing clusters were generated, with an average depth of 27,770 sequences per sample (median = 15,318; min = 5,164; max = 145,270). Four reagent contaminant ASVs were identified and removed using decontam package based on the prevalence of the ASVs in the reagent negative blank controls, using default parameters[18]. Unassigned ASVs and chloroplast and mitochondrial ASVs were removed from statistical analyses[19].

Statistical Analysis

Analyses of alpha- and beta-diversity were used to compare nasopharyngeal microbial communities in COVID-19-positive and -negative patients; all analyses were performed on feature (ASV) counts. Alpha-diversity metrics were calculated on rarefied data (5,000 sequences/sample). Differences in alpha-diversity indices and bacterial ratios were assessed for significance using unpaired t-test or Mann-Whitney test, based on the outcome of Shapiro-Wilks normality test. Significance levels were set at $p < 0.05$. Permutation Multivariate Analysis of Variance (PERMANOVA) was used to assess global differences in microbial communities between groups[20]. Significance of PERMANOVA values were determined using 9,999 permutations and adjustment for multiple testing was conducted using the Benjamini-Hochberg false-discovery rate correction. Visualization of data was performed using principal

coordinates analysis (PCoA) based on a Bray-Curtis dissimilarity distance matrix within the software package QIIME2.

Random forest models (number of runs = 1,023) were used to predict featured taxa of importance using the R implantation of the Boruta algorithm[21]. Analysis of composition of microbiomes (ANCOM) was performed on nasopharyngeal microbial communities to identify differentially abundant taxa between groups[22]. Additionally, differential abundances of individual taxa between groups were determined using differential abundance analysis (DESeq2) generating a q-value[23], as DESeq2 has increased sensitivity on smaller datasets (< 20).

Declarations

Data Availability

Raw sequence data (FASTQ files) were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA), under the BioProject identifier PRJNA704967.

Acknowledgements: We would like to thank the RUMC COVID-19 Biorepository investigators Drs. Mary Hayden and Nagarjuna Tippireddy for providing clinical data.

Author Contributions: A.K. and R.M.V. conceived and designed the experiments and contributed to writing the manuscript. P.A.E. wrote the manuscript, performed the experiments, and analyzed the data. A.N., S.J.G., and C.J. analyzed the data and contributed to writing the manuscript. A.L. designed the RUMC COVID-19 repository. A.L and C.J. provided the biological samples for experimentation. All authors reviewed the manuscript.

Competing Interests: The authors declare no competing interests.

References

1. Chan, J.F*., et al.* A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **395**, 514-523 (2020).
2. Li, N., Ma, W.T., Pang, M., Fan, Q.L. & Hua, J.L. The Commensal Microbiota and Viral Infection: A Comprehensive Review. *Front Immunol* **10**, 1551 (2019).
3. Belkaid, Y. & Hand, T.W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121-141 (2014).
4. Man, W.H., de Steenhuijsen Piters, W.A. & Bogaert, D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol* **15**, 259-270 (2017).
5. De Maio, F*., et al.* Nasopharyngeal Microbiota Profiling of SARS-CoV-2 Infected Patients. *Biological procedures online* **22**, 18 (2020).

6. Mostafa, H.H.*, et al.* Metagenomic Next-Generation Sequencing of Nasopharyngeal Specimens Collected from Confirmed and Suspect COVID-19 Patients. *mBio* **11**(2020).
7. Hang, J.*, et al.* Composition and variation of respiratory microbiota in healthy military personnel. *PLoS One* **12**, e0188461 (2017).
8. Ding, T.*, et al.* Microbial Composition of the Human Nasopharynx Varies According to Influenza Virus Type and Vaccination Status. *mBio* **10**(2019).
9. Minich, J.*, et al.* Feasibility of using alternative swabs and storage solutions for paired SARS-CoV-2 detection and microbiome analysis in the hospital environment. *Research square* (2020).
10. Moore, N.M., Li, H., Schejbal, D., Lindsley, J. & Hayden, M.K. Comparison of Two Commercial Molecular Tests and a Laboratory-Developed Modification of the CDC 2019-nCoV Reverse Transcriptase PCR Assay for the Detection of SARS-CoV-2. *J Clin Microbiol* **58**(2020).
11. Naqib, A.*, et al.* Making and Sequencing Heavily Multiplexed, High-Throughput 16S Ribosomal RNA Gene Amplicon Libraries Using a Flexible, Two-Stage PCR Protocol. *Methods Mol Biol* **1783**, 149-169 (2018).
12. Caporaso, J.G.*, et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**, 1621-1624 (2012).
13. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620 (2014).
14. Callahan, B.J.*, et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583 (2016).
15. Estaki, M.*, et al.* QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. *Curr Protoc Bioinformatics* **70**, e100 (2020).
16. Bokulich, N.A.*, et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).
17. Quast, C.*, et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590-596 (2013).
18. Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A. & Callahan, B.J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).
19. Hanshew, A.S., Mason, C.J., Raffa, K.F. & Currie, C.R. Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. *J Microbiol Methods* **95**, 149-155 (2013).
20. Kelly, B.J.*, et al.* Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics* **31**, 2461-2468 (2015).
21. Rudnicki, M.K.a.W. Feature Selection with the Boruta Package. *Journal of Statistical Software* **36**, 1-13 (2010).

22. Mandal, S., *et al.* Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial ecology in health and disease* **26**, 27663 (2015).
23. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

Table

Table 1. Results of analysis of compositions (ANCOM) between COVID-19-positive and -negative nasopharyngeal samples.

ANCOM			
Genus (Phylum)	**Median Percentile Abundance: COVID-19-Negative**	**Median Percentile Abundance: COVID-19-Positive**	**W Score**
Streptococcus (Firmicutes)	213.92	1.0	**19**
Genera trending towards significance			
Enterococcus (Firmicutes)	37.25	292	8
Burkholderia-Caballeronia-Paraburkholderia (Proteobacteria)	1.0	91.33	8
Gulbenkiania (Proteobacteria)	1.0	33.67	6
Finegoldia (Firmicutes)	121.83	1.0	5
Anaerococcus (Firmicutes)	173.42	1.0	4
Neisseriaceae Uncultured (Proteobacteria)	371.42	1.0	4
Bacillus (Firmicutes)	47.58	624.67	4
Prevotella (Bacteroidetes)	70.58	1.0	3
Corynebacterium (Actinobacteria)	3,434.08	236.67	3

ANCOM: samples with fewer than 5,000 sequences were removed; features that were not present in at least 4 samples were removed; features with an overall count of less than 500 [10% of feature biom – due to small sample size ≤ 10 per group] were also removed. All listed significant features [genus] rejected the null hypothesis. Median percentile abundance (average of the 25^{th}, 50^{th}, 75^{th} percentiles) and analysis W scores determined using ANCOM. ANCOM values were corrected for multiple testing using the Benjamini-Hochberg method (q-value < 0.05: bold/grey); (p-value < 0.05: bold). Taxa identified by both DESeq2 and ANCOM are bolded above.