Vietnam Journal of Biotechnology **22**(1): 169-186, 2024. DOI: 10.15625/vjbt-19787

DIFFERENTIAL ABUNDANCE IN UPPER RESPIRATORY MICROBIOME BETWEEN SMOKING AND NON-SMOKING COVID-19 PATIENTS

Thuy Truc Anh Le[1], Bao Ngoc Mach[2], Hoang Dang Khoa Do[2], Thi Thu Hoai Nguyen[1,3,✉], Minh Thiet Vu[2,✉]

[1]*School of Biotechnology, International University, Vietnam National University Ho Chi Minh City, Quarter 6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam*

[2]*NTT Hi-Tech Institute, Nguyen Tat Thanh University, 300A Nguyen Tat Thanh Street, Ward 13, District 4, Ho Chi Minh City, Vietnam*

[3]*Research Center for Infectious Diseases, International University, Vietnam National University Ho Chi Minh City, Quarter 6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam*

✉To whom correspondence should be addressed. E-mails: ntthoai@hcmiu.edu.vn, and vmthiet@ntt.edu.vn

Received: 04.12.2023
Accepted: 24.03.2024

ABSTRACT

The COVID-19 pandemic caused by SARS-CoV-2 has resulted in significant respiratory morbidity and mortality. Smoking is a well-established risk factor for adverse respiratory outcomes, and it may exacerbate COVID-19 outcomes by altering the upper respiratory tract (URT) microbiome. However, the association between smoking, the URT microbiome, and COVID-19 severity remains controversial. In this study, we investigated the microbial community in the UTR of smoking and non-smoking COVID-19 patients using 16S ribosomal RNA metagenomic datasets. By characterizing the microbial profiles of the URT of 88 COVID-19 patients with and without a smoking history, we examined the differences in the URT microbiome between smokers and non-smokers and identified bacteria that could serve as smoking signatures. Our study found that *Prevotella* was the most abundant genus in the URT microbiome of both smoking and non-smoking COVID-19 patients, followed by five other dominant genera. Notably, *Fusobacterium* was significantly higher in smokers than non-smokers, suggesting that smoking may affect the URT microbial composition. However, the alpha and beta diversity indices showed no significant differences within and across communities, regardless of smoking status or levels of severity. Differential abundance analysis suggested that the impact of smoking on the URT microbiome may lead to an increase in the abundance of *Streptobacillus* in smoking patients. Overall, this study highlights the potential impact of smoking on the URT microbiome and its relevance to COVID-19 outcomes. Thus, it provides insights into the underlying mechanisms that may lead to worsened COVID-19 outcomes due to smoking-induced alterations in the URT microbiome.

Keywords: 16S metagenomics, COVID-19, smoking, upper respiratory tract.

INTRODUCTION

The COVID-19 pandemic has killed millions of people worldwide. SARS-CoV-2, the virus that causes COVID-19, primarily targets the respiratory system and lungs, posing a significant challenge to individuals with pre-existing lung conditions such as asthma and chronic obstructive pulmonary disease (COPD). COVID-19 severity was also correlated with the extent of lung damage (Shi *et al.*, 2020), which can provoke pneumonia, acute respiratory distress syndrome (ARDS), or long-term lung function issues. To assess the severity of COVID-19, the extent of lung damage must be understood.

Furthermore, given that the upper respiratory tract (URT) is the primary entry point for SARS-CoV-2, the interaction between the virus and the URT microbiome is significant. The term "microbiome" is defined as a collection of microorganisms inhabiting a defined environment, including bacteria, fungi, protozoa, viruses, and other micro-eukaryotic colonizers (Berg *et al.*, 2020). Without any infection or chronic disease, a typical URT microbiome primarily consists of Actinobacteria (68%) and Firmicutes (27%), with *Propionibacterium*, *Corynebacterium,* and *Staphylococcus* as representative genera. The remaining bacterial phyla, such as Proteobacteria (4.0%), Bacteroidetes (1.4%), Fusobacteria (0.21%), Cyanobacteria (0.08%), and others, account for a much smaller proportion (Zhu *et al.*, 2020).

However, COVID-19 can alter URT microbial composition and homeostasis, causing diversity shifts and pathological consequences (Bello *et al.*, 2018). SARS-CoV-2 infection disrupts the URT microbiome (Dickson & Huffnagle, 2015), increasing opportunistic pathogens like *Streptococcus* and *Rothia* and decreasing levels of commensal bacteria, such as *Prevotella* and *Veillonella* (Han *et al.*, 2020). Likewise, COVID-19 severity was linked with an alteration in the URT microbiome, which reduced the abundance of *Corynebacterium* relatives (Rosas-Salazar *et al.*, 2021). Thus, understanding the disturbances in the URT microbiome during COVID-19 infection offers a promising way to investigate the repercussions of the disease.

Smoking can harm lung function, making it a risk factor for respiratory infections like COVID-19. Smokers had nearly two times the risk of COVID-19 severity compared to non-smokers (Zhao *et al.*, 2020). Reddy *et al.* (2021) further demonstrated that smokers were more likely to develop severe or critical COVID-19 conditions that required intensive care unit (ICU) admission and mechanical ventilation. Smoking can disrupt the balance of the URT microbial composition and increase the risk of respiratory infections and other respiratory conditions. A study compared the nasal cavity microbiome of smokers and non-smokers and found a significant decrease in bacterial diversity and an increase in pathogenic bacteria (Al Kawas *et al.*, 2021). Several recent studies have focused on the exclusive association between smoking and COVID-19 outcomes, or the URT microbiome, but how smoking-induced microbiome changes affect the severity of COVID-19 has not been explored.

This study examined the microbial profiles of the URT in COVID-19 patients to understand the difference between smoking and non-smoking groups. Our study sought to answer two main questions about the URT microbiome and COVID-19 severity in

smokers and non-smokers: (1) Do smokers and non-smokers have different URT microbial diversity? (2) Does the URT contain smoking-signature microorganisms? To address these questions, we analyzed the 16S metagenomics to look at COVID-19 patients' URT bacterial profiles and examined any microbial shifts associated with smoking status and disease severity.

MATERIALS AND METHODS

Study population

From the BioProject PRJNA751478, we selected a cohort of 88 patients with a confirmed COVID-19 diagnosis out of the original dataset of 335 patients (Galeeva *et al.*, 2022). Participants who lacked information on smoking and the percentage of lung damage were excluded. The selected cohort comprised 78 non-smokers (88.6%) and ten smokers (11.4%). The mean age of this study population was 61.1 (±15.6) years. There was a fairly even gender distribution of females (n = 48) and males (n = 40). The percentages of lung damage were then categorized into three levels based on quartile statistics: mild (below 15% of lung damage, inclusively, n = 29), moderate (between above 15% and under 40%, n = 32), and severe (above 40%, inclusively, n = 27).

Raw data description and processing

Raw 16S rRNA sequencing data generated from the nasopharyngeal samples of these COVID-19 patients was retrieved from the open-source repository NCBI/Sequence Read Archive under the mentioned BioProject (Galeeva *et al.*, 2022). These sequences were generated by the Illumina MiSeq platform (Illumina, San Diego, CA, USA) to amplify the V3-V4 region of the 16S rRNA gene amplicon sequences. The data possesses an average of 4,635 reads per sample. The metadata, providing the personal and clinical information of the patients, such as age, sex, smoking status, the percentage of lung damage (classified into the three levels mentioned), and patient status (hospitalized or on ambulatory treatment), were also included. We then investigated the URT microbiome in the selected cohort using Quantitative Insights Into Microbial Ecology 2 (QIIME2) software version (Bolyen *et al.*, 2019), following the below 16S metagenomics pipeline (Figure 1).

We checked the quality of the paired-end, demultiplexed data of 16S rRNA gene sequences in the form of FASTQ files before being imported into QIIME2. The imported data underwent sequence filtering through the Divisive Amplicon Denoising Algorithm 2 (DADA2) package in QIIME2 (Callahan *et al.*, 2016)**,** which enables high-resolution sequence selection and denoises low-quality sequencing reads. The reads were then merged with VSEARCH (Rognes *et al.*, 2016) and denoised via the UCHIME *de novo* approach to detect and remove chimeric sequences (Edgar *et al.*, 2011). The resulting sequences were clustered into operational taxonomic units (OTUs) with an assumed 97% similarity.

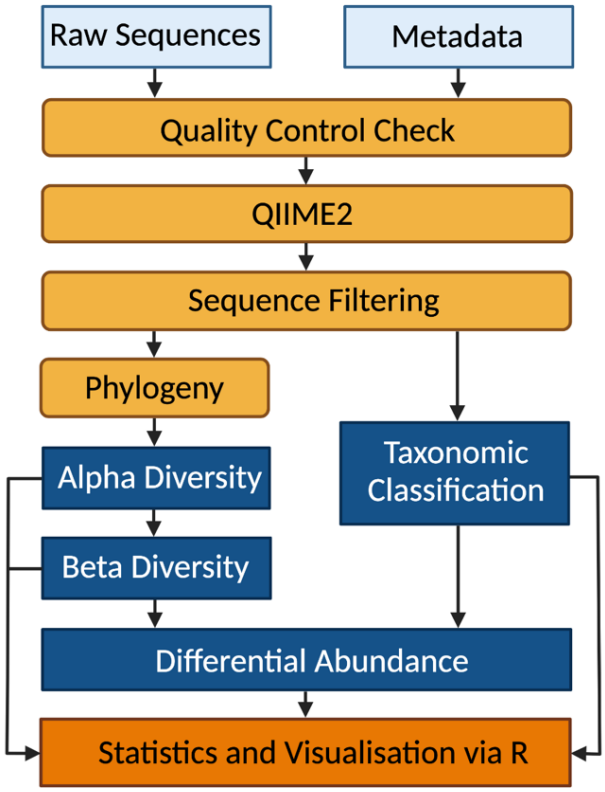

Figure 1. Overview of the metagenomics pipeline used in this study on QIIME2 for 16S rRNA gene sequencing analysis.

Microbiome analysis

Alpha diversity

Alpha diversity evaluates species diversity within a single community (intra-community diversity) through measures of richness and evenness (Thukral, 2017). We measured richness via the Shannon index and observed features, and evenness via the Pielou index, and Faith's phylogenetic diversity (PD). Alpha rarefaction curves constructed via observed features were obtained for each sample to estimate the OTUs detected as a function of sequencing depth. Alpha diversity indices were calculated after rarefying the abundance table to 900 reads per sample.

Beta diversity

Beta diversity (inter-community diversity) measures the variation in species diversity between communities (Anderson *et al.*, 2006). We estimated dissimilarity between groups by incorporating phylogenetic distances between ASVs based on Jaccard, Bray-Curtis, Unweighted UniFrac (qualitative measure), and Weighted UniFrac (quantitative measure) distances. Bray-Curtis distance measures the dissimilarity based on the abundance of shared microbial taxa. Jaccard distance measures dissimilarity based on the presence or absence of microbial taxa. Weighted UniFrac considers both the abundance and phylogenetic relatedness of microbial taxa,

whereas unweighted UniFrac only considers the presence or absence of microbial taxa. We computed the four distance indices, in which the distances between the OTUs were handled by reducing the original distances to principal coordinate analysis (PCoA).

Taxonomic classification

Taxonomy was assigned to the ASVs using a pre-trained classifier of the SILVA database version 138.1 (Quast *et al.*, 2013; Yilmaz *et al.*, 2014). The resulting feature table allowed the quantification and comparison of the microbial community composition. Representative sequences were obtained for each OTU, enabling the construction of a phylogenetic tree to determine the evolutionary relationships between different microbial taxa and identify specific sequences defining each OTU. The relative taxonomic abundance of the groups is presented as a mean percent value, calculated by dividing the read counts of a taxon by the sample size.

Differential abundance

Differential abundance (DA) analysis using sophisticated statistics for multivariate analysis can identify taxa that differ significantly between conditions (Lin & Peddada, 2020). In this study, DA analysis was conducted using the Analysis of Composition of Microbiomes (ANCOM) method (Mandal *et al.*, 2015), which takes into account the compositional nature of microbiome data through Centered Log Ratio (CLR) transformation. The CLR transformation involves taking the logarithm of the ratio of each taxon's abundance to the geometric mean of taxon abundances in the sample. This study used ANCOM with CLR transformation to identify significantly distinct microbial taxa. The W statistic, an output of ANCOM, indicated the differential abundance of microbial taxa between groups, with a higher value indicating a higher likelihood of differential abundance.

Statistical analysis and visualization

The statistical analysis and visualization were conducted using R version 4.2.3 software (*R: The R Project for Statistical Computing*, n.d.), mainly with the QIIME2R (Bisanz, 2023), vegan (Oksanen *et al.*, 2022), tidyverse (Wickham *et al.*, 2019), and ggplot2 (*Ggplot2: Elegant Graphics for Data Analysis | SpringerLink*, n.d.) packages. In addition, the QIIME2 view function was used to visualize alpha-rarefaction and beta-diversity results. Multivariate statistical analysis was conducted by incorporating the OTU table, phylogenetic tree, taxonomic information, and metadata. To compare the relative abundance of the URT microbiome between smokers and non-smokers, an independent t-test was performed after confirming the normality assumptions. The statistical significance of alpha diversity between smoking and non-smoking groups was determined by the Wilcoxon signed rank test. Beta diversity was assessed by permutational multivariate analysis of variance (PERMANOVA) of 999 permutations, which fits linear models to distance matrices and uses a permutation test with pseudo-F ratios. ANCOM, also one of the statistical tools for multivariate analysis, was used to find out the microbial signature of smoking. The W statistic was calculated to measure the magnitude of differences in the relative abundance of a feature between groups while controlling the false discovery rate. A p-value less than 0.05 was considered statistically significant.

RESULTS

Alpha diversity

To account for heterogeneity in sequencing power, all samples were analyzed by rarefaction and diversity measured at a common sampling depth of 1000 reads per sample (Figure 2). As sequencing depth increases, the rarefaction curve reaches a plateau. This indicates that the sequencing data obtained from the samples was reliable for further analysis, given a certain depth and degree of representativeness. Based on the rarefaction curve analysis, we selected a sequencing depth cut-off of 900 reads for all samples for diversity analysis.

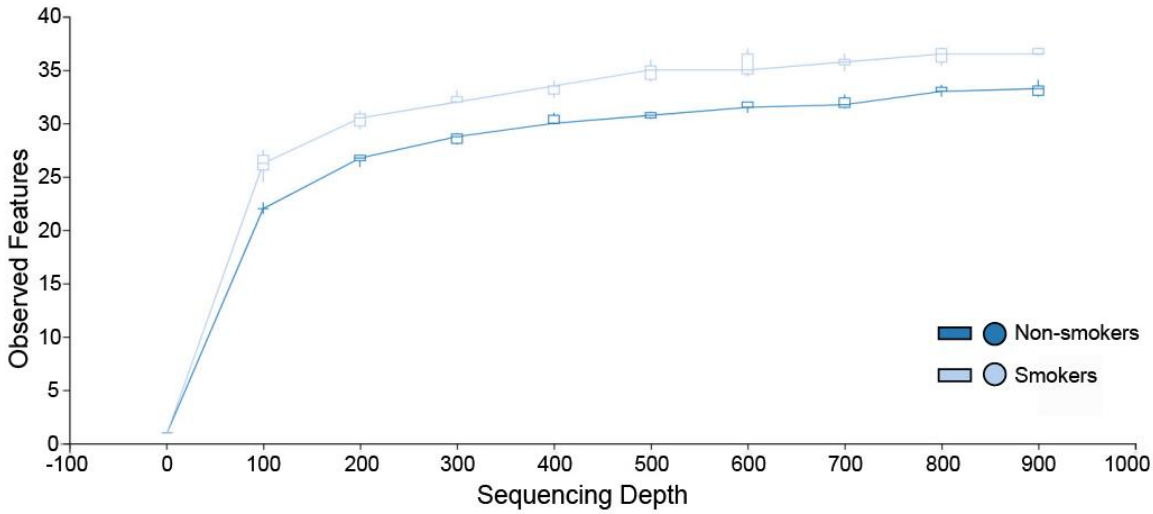

Figure 2. Alpha rarefaction curves constructed based on observed features comparing smoking and non-smoking groups. Alpha rarefaction curves were visualized via QIIME2 view. The x-axis represents the number of sequencing depths, with the cutoff at 900 reads.

Analysis of four alpha diversity indices revealed no significant differences between the smoking and non-smoking groups in terms of richness and evenness. Figure 3 indicates that regardless of smoking status, both groups have a similar level of microbial diversity. In more detail, the richness index revealed that both groups have a comparable number of distinct species, as indicated by Shannon's index (*p-value* = 0.7) and the Observed features index (*p-value* = 0.99) (Figures 3a and 3b). The evenness index indicated that they exhibit a similar complexity, as measured by Pielou's evenness (*p-value* = 0.26) (Figure 3c). Additionally, Faith's PD index, which assesses the complexity of the phylogenetic relationships of species in the samples, showed no significant differences in microbial diversity between the smoking and non-smoking groups (*p-value* = 0.47) (Figure 3d).

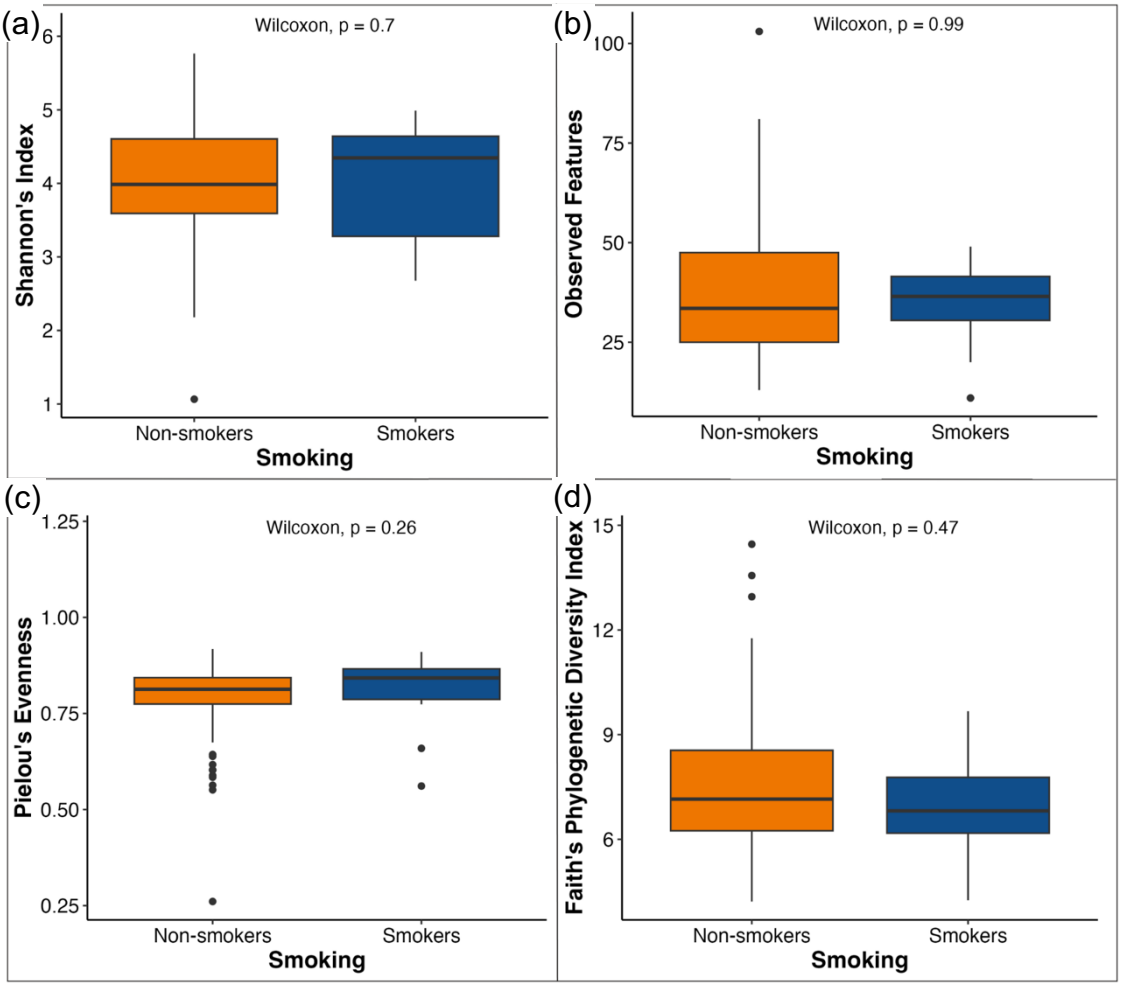

Figure 3. Comparison of alpha diversity indices between smoking and non-smoking groups. Box plots for alpha diversity indices: (a) Shannon's index, (b) Observed Features, (c) Pielou's evenness, and (d) Faith's PD index. P-values were calculated with the Wilcoxon signed rank test between the two groups and displayed on each bar plot. A p-value less than 0.05 is considered statistically significant.

Beta diversity

The analysis of beta diversity showed no significant differences between smoking and nonsmoking groups (*p-value* > 0.05) (Figure 4), nor between patients with mild, moderate, and severe lung damage (*p-value* > 0.05) (Figure 5). Overall, the four beta diversity indices suggested that the microbial composition and abundance patterns of the URT microbiome are relatively comparable across all the groups studied, regardless of smoking status or different levels of lung damage.

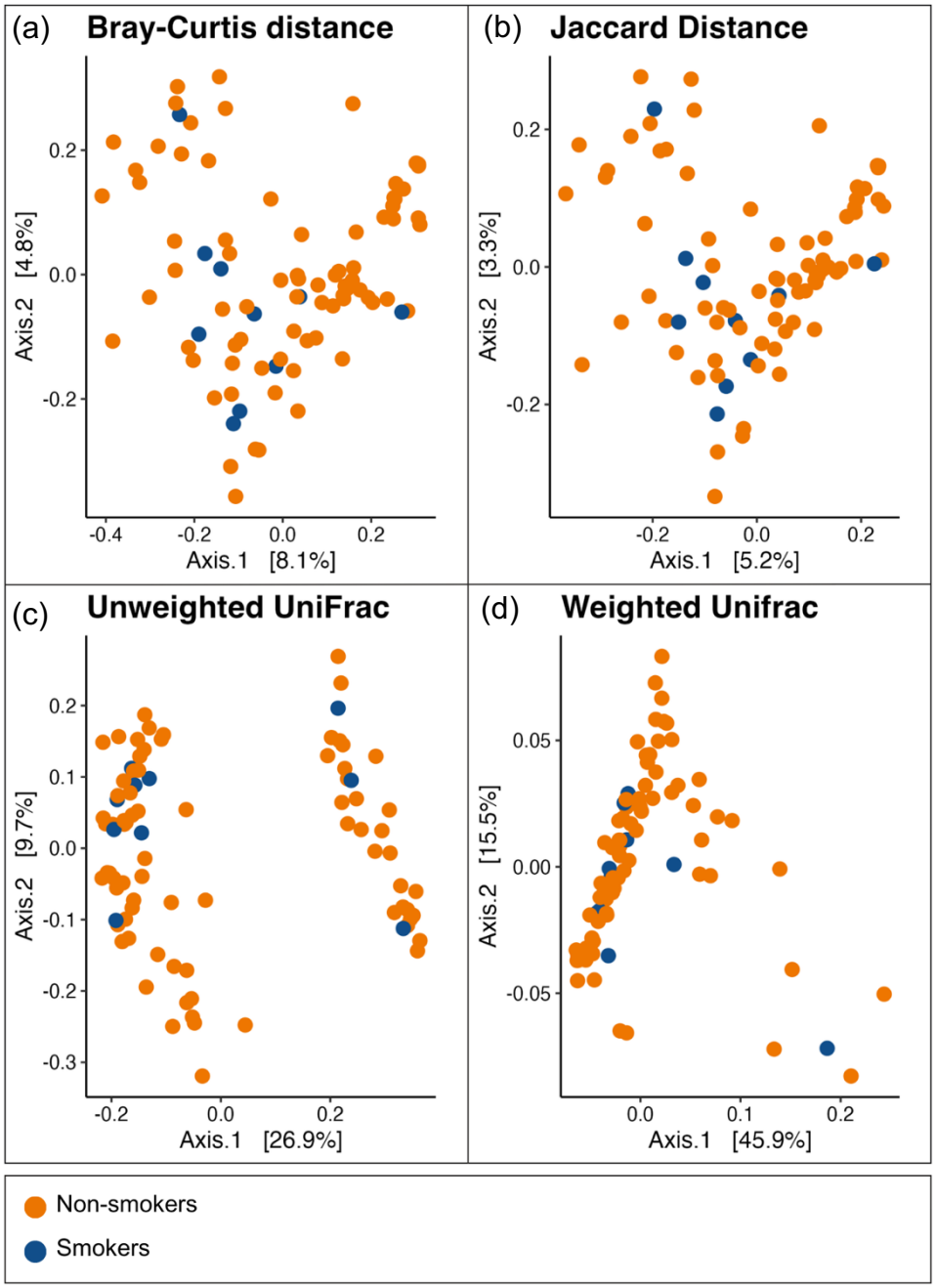

Figure 4. Comparison of beta diversity indices between smoking and non-smoking groups. PcoA plots for beta diversity indices: (a) Bray-Curtis distance, (b) Jaccard distance, (c) weighted UniFrac distance, and (d) unweighted UniFrac distance are displayed. P-values were calculated with a pairwise PERMANOVA between smoking and non-smoking, with 999 permutations. A p-value less than 0.05 is considered statistically significant.

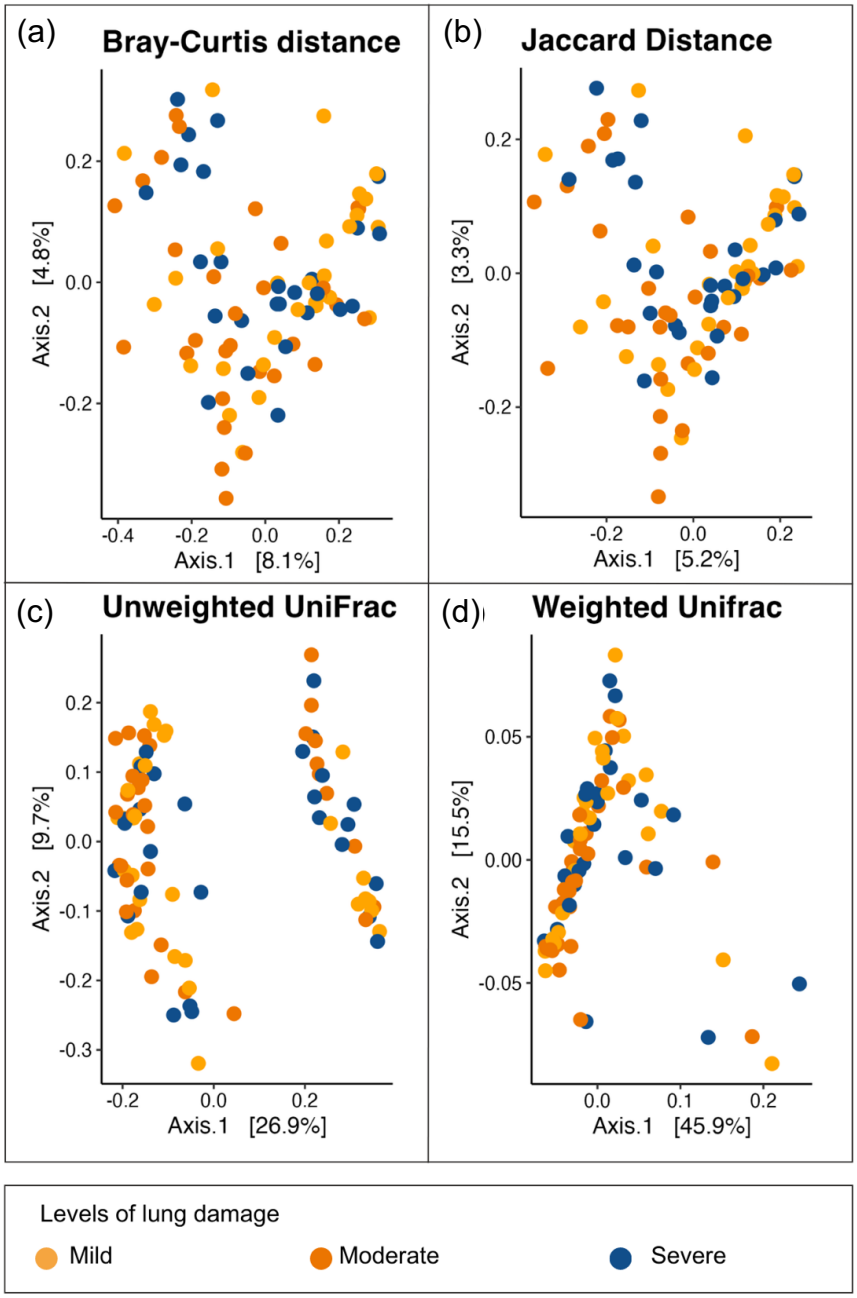

Figure 5. Comparison of beta diversity indices among COVID-19 patients with three levels of lung damage. PcoA plots for beta diversity indices: (a) Bray-Curtis distance, (b) Jaccard distance, (c) weighted UniFrac distance, and (d) unweighted UniFrac distance are displayed. P-values were calculated with a pairwise PERMANOVA between patients with mild, moderate, and severe levels of lung damage, with 999 permutations. A p-value less than 0.05 is considered statistically significant.

However, when visualized in 3-dimension, we observed a cluster of OTU specific for non-smokers with moderate and severe levels of lung damage via the unweighted UniFrac distance analysis, as illustrated in Figure 6. Our findings indicates that non-

smoking COVID-19 patients may possess a distinct microbial community in the URT, which may potentially contribute to the development and progression of lung damage in this subgroup.

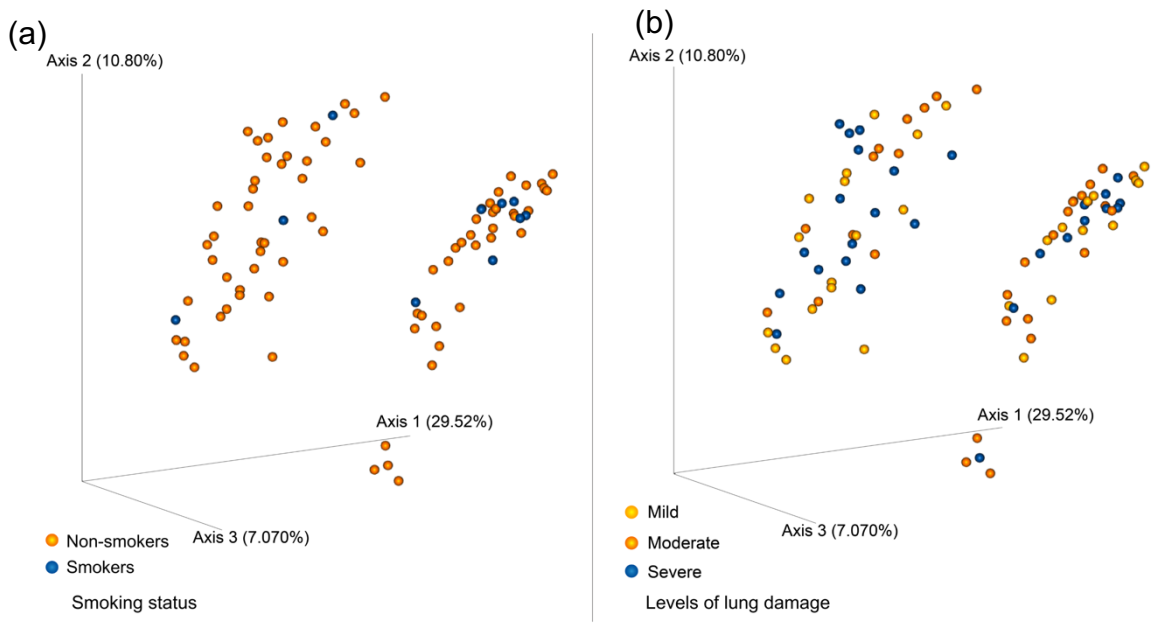

Figure 6. A cluster of OTU specific for non-smokers with moderate and severe lung damage. A 3-dimensional PCoA plot for Unweighted UniFrac distance was visualized via QIIME2 view: (a) for smoking status, and (b) for levels of lung damage.

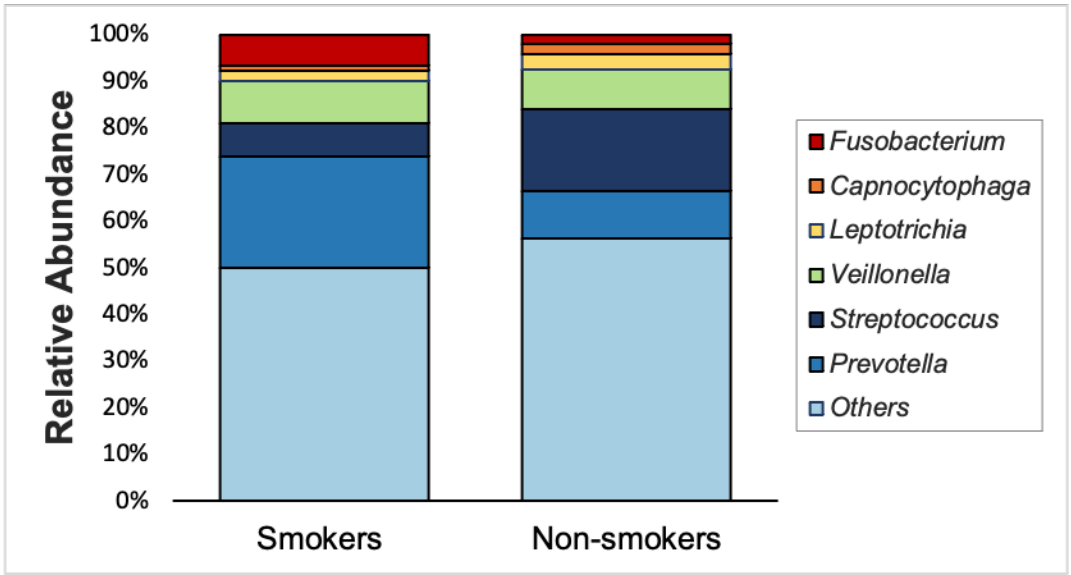

Figure 7. Taxonomic analysis of the URT microbiome at the genus level between smoking and non-smoking groups. Taxonomy- stacked bar plot of the mean relative abundance of the top six most abundant genera for smokers (n = 10) and non-smokers (n = 78). Genera with a mean relative apart from the top six are collapsed into the category “Others”.

Taxonomic classification

The taxonomy-stacked bar plots showed the main taxa in the URT microbiome at the genus level between smoking and non-smoking COVID-19 patients (Figure 7). Overall, the URT microbiomes in both smoking and non-smoking groups were dominated by six bacterial genera. In particular, *Prevotella* was the most abundant genus in both smoking and non-smoking groups, with a mean relative abundance of 23.92% (± 1.81) and 17.54% (± 0.217), respectively. The other five genera included *Streptococcus* (smokers: 7.28% ± 0.674 versus non-smokers: 10.12% ± 0.143), *Veillonella* (8.84% ± 0.578 versus 8.60% ± 0.110), *Leptotrichia* (2.43% ± 0.372 versus 3.07% ± 0.0740), *Capnocytophaga* (0.85% ± 0.106 versus 2.40% ± 0.0650), and *Fusobacterium* (6.73% ± 0.830 versus 1.87% ± 0.0512). Among the six genera, the relative abundance of *Fuscobacterium* in smokers was significantly higher than that in non-smokers (*p*-value = 0.0039), as shown in Figure 8.

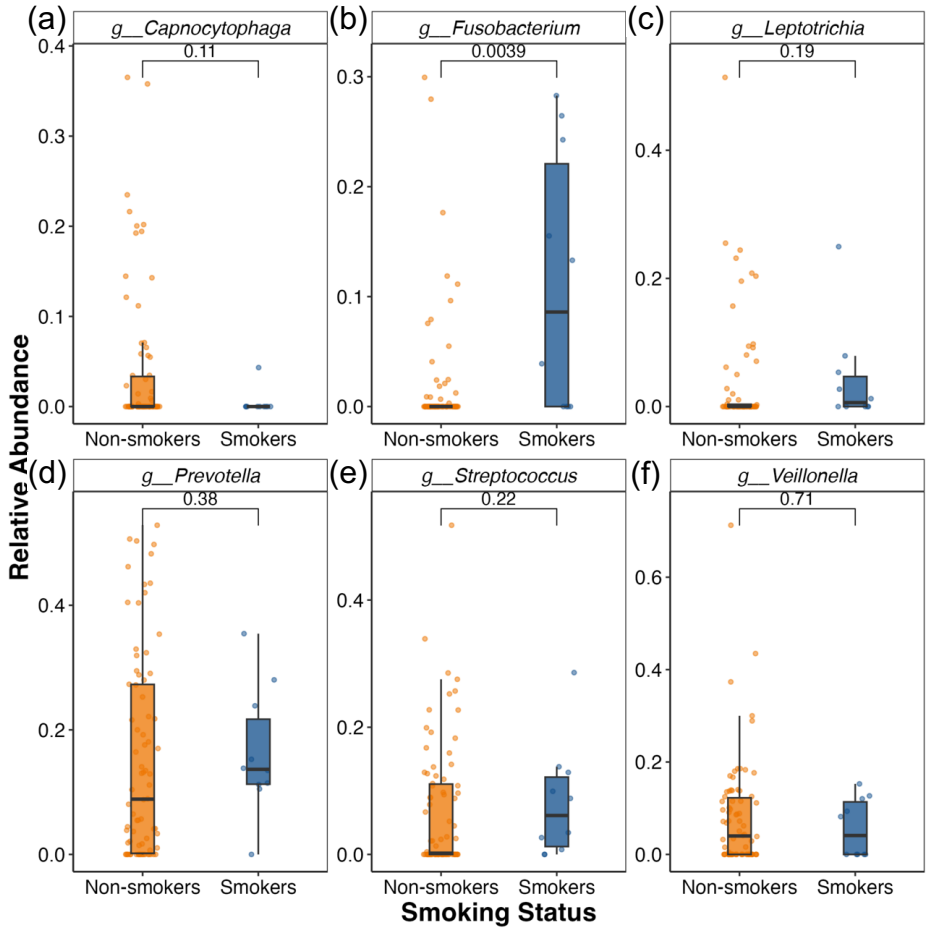

Figure 8. Taxonomic differences among the top six abundant genera in the URT microbiome between smoking and non-smoking groups. Box plots for the mean relative abundance of the main bacterial genera compared between the two groups: (a) Capnocytophaga, (b) Fusobacterium, (c) Leptotrichia, (d) Prevotella, (e) Streptococcus, and (f) Veillonella. A p-value less than 0.05 is considered statistically significant.

Differential abundance

Streptobacillus was identified as the most significantly differential OTU in the URT microbiome between smokers and non-smokers (Figure 9). The ANCOM result showed that *Streptobacillus* was highly represented in the URT microbiome of the smokers (W = 214), though with a low relative frequency (20% of smokers). In addition to this, our study was the first to identify the presence of *Streptobacillus* in COVID-19 patients.

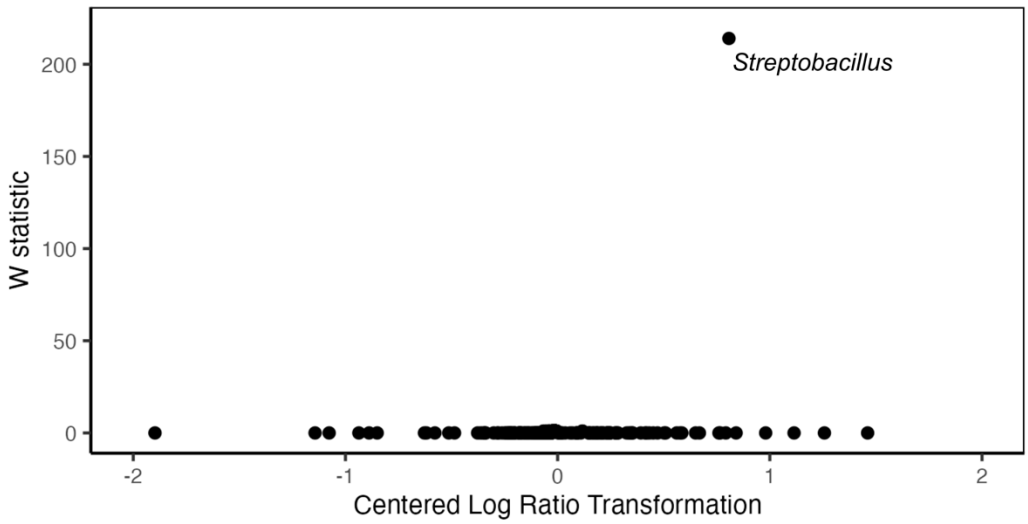

Figure 9. Differentially abundant bacterial taxa were identified by ANCOM. A volcano plot displays differential abundance at the genus level between smoking and non-smoking groups. The x-axis shows the CLR transformation, while the y-axis shows the W statistic. Positive x-axis values indicate the higher abundance in the smoking group, while negative values indicate the higher abundance in the non-smoking group. The higher the W statistic values, the more likely a feature differs statistically. A statistically significant feature, as identified by ANCOM, has been labelled in the plot.

DISCUSSION

The URT microbiome plays a crucial role in respiratory disease by shaping immune responses and potentially acting as the frontline of defense against viral attacks. To investigate the impact of smoking and the URT microbiome on the severity of COVID-19, we analyzed the microbiome of 88 patients using 16S metagenomics. Our findings shed light on the potential link between smoking-related URT microbiome changes and COVID-19 severity.

To examine the impact of smoking on microbial diversity in the URT, we evaluated alpha diversity indices. Prior studies have yielded contradictory results regarding alpha diversity in smokers and non-smokers (Yu *et al.*, 2017; Bach *et al.*, 2023). Our study found no significant differences in microbial diversity between the two groups (Figure 3), similar to a study by Morris *et al.* in 2013. Due to small sample sizes and unequal proportions of smokers and nonsmokers, it is possible that our and other studies' contradictory findings do not reflect the larger population. Smoking-related

perturbations in the URT microbiome may take longer to manifest, which may be missed by alpha diversity indices measured during active COVID-19 infection (Bach *et al.*, 2023). In addition, host genetics and other environmental exposures may influence the URT microbiome, potentially masking the effects of smoking (Hauptmann & Schaible, 2016). These findings imply that the relationship between smoking and the URT microbiome in COVID-19 patients is more complicated than previously anticipated and that a larger sample size of both smokers and non-smokers is necessary to accurately represent the population.

In our study, the severity of COVID-19 was determined using quartile statistics to categorize lung damage levels. Although other clinical factors are typically considered when assessing the severity of COVID-19 (Besutti *et al.*, 2022; M & K, 2020), our classification method, which relies on statistics and COVID-19 severity classification (Cascella *et al.*, 2024), ensures the reliability of our results. The beta diversity analysis did not reveal any significant differences in the URT microbiome between smokers and non-smokers with varying degrees of lung damage (Figures 4 and 5). Nonetheless, a cluster of OTUs specific to non-smokers with moderate to severe lung damage was observed (Figure 6). A recent study discovered that the respiratory microbiome may vary among patients with different levels of COVID-19 severity, with a severe cohort having less bacterial diversity in the URT (Chen *et al.*, 2022). Critical COVID-19 patients also have a URT microbiome profile that is distinct from that of healthy individuals (Bai *et al.*, 2022). This suggests that alterations in the URT microbiome may influence disease susceptibility and severity. The observed clustering, however, could be a coincidental finding or be confounded by other unmeasured factors, emphasizing the need for larger-scale studies with more comprehensive data.

Our results revealed that *Prevotella* was the most abundant genus in the URT microbiome of both smoking and non-smoking COVID-19 patients (Figure 7), consistent with previous research that identified *Prevotella* to be a prominent genus inhabiting the URT (Bassis *et al.*, 2015). While *Prevotella* is typically commensal organism, it can convert into a pathogenic colonizer when exposed to cigarette smoke (Brook & Gober, 2005) or when confronted with respiratory ailments (Dickson & Huffnagle, 2015; Teo *et al.*, 2015). Additionally, the distribution of genera, such as *Prevotella*, *Veillonella*, and *Streptococcus*, has been recognized as dominant yet opportunistic pathogens in the URT microbiome of COVID-19 patients (Lynch, 2016). These genera can modulate the URT immune response and contribute to the progression of respiratory diseases, including COVID-19. Elevated levels of *Prevotella* have previously been observed in COVID-19 patients (Ventero *et al.*, 2021), possibly due to its ability to facilitate mucus accumulation, which is diminished during COVID-19 (Robinot *et al.*, 2021). This enables *Prevotella* to persist in the URT and metabolize mucus, thereby creating conditions that favor the growth of other pathogenic bacteria that are incapable of efficiently utilizing mucus.

Intriguingly, we found that smokers have a higher percentage of *Fusobacterium* present in their URT than non-smokers (Figure 8). *Fusobacterium* is a gram-negative anaerobic bacterium that is commonly found in the URT (Brook, 2002), and its presence has

been linked to smoking (Gopinath *et al.*, 2022) and respiratory diseases (Chakraborty, 2020). Smoking can alter the pH and oxygen levels in the URT, which can lead to increased inflammation and reduced immune surveillance (Stämpfli & Anderson, 2009). Consequently, smoking may contribute to the prevalence of respiratory disease by promoting the proliferation of *Fusobacterium*.

In terms of differential abundance, our study found that *Streptobacillus* was highly abundant in COVID-19 smokers (Figure 9), suggesting that smoking-induced changes in the URT microbiome may facilitate its growth. Note that *Streptobacillus* is naturally exists in the oral cavity (Kimura *et al.*, 2008) and URT (Paegle *et al.*, 1976) of rats and has long been recognized as the causative agent of strepto-bacillary rat-bite fever. In 2016, Lau *et al.* demonstrated that the human URT can serve as a reservoir for a species of *Streptobacillus*, *S. hongkongensis*. Interestingly, differential abundance testing recently showed a higher level of *S. hongkongensis* in the oral microbiome of smokers (Paegle *et al.*, 1976). Nonetheless, because the relative frequency of this genus in our study was generally low among smokers, more research is necessary to establish *Streptobacillus* as a smoking signature.

The study also had some limitations. First, the study was restricted by a lack of personal and clinical information from the open-access repository, including comorbidities, medication use, duration of treatment, and recovery rate of COVID-19 patients. Second, the small sample size (n = 8) and unequal number of COVID-19 patients who smoke (n = 10) versus those who do not (n = 78) may reduce statistical power and hinder the possibility of establishing a correlation between smoking, URT microbiome, and COVID-19 severity in a larger population. Third, the 16S metagenomics used in our study has inherent limitations, such as low accuracy of classification below the genus level, and limited taxonomic resolution among closely related taxa (Poretsky *et al.*, 2014). In addition, the 16S metagenomics approach was unable to provide a functional analysis of the patient's URT microbiome (Greenblum *et al.*, 2012; Franzosa *et al.*, 2014), which could have provided more insights into its role in COVID-19 severity.

CONCLUSION

In conclusion, the URT microbiomes of both smoking and non-smoking COVID-19 patients exhibited similar dominant genera, with *Prevotella* being the most abundant in both groups. However, the higher mean relative abundance of *Fusobacterium* in smokers suggests that smoking may alter the microbial composition of the URT. The levels of microbial diversity did not differ significantly between the two groups based on smoking status. Furthermore, the overall microbial composition and abundance patterns of the URT microbiomes were comparable across the subgroups in terms of smoking status and different levels of lung damage. However, there was a distinct cluster of OTUs specific to non-smokers with moderate and severe lung damage, which may contribute to the progression of lung damage in the COVID-19 non-smokers. More importantly, our observation indicates that *Streptobacillus* exhibited a higher abundance in smokers. The data suggest that it could potentially serve as a marker for distinguishing between smokers and non-smokers in the URT. Nevertheless, further validation is required to ascertain its role as a definitive smoking signature.

ACKNOWLEDGEMENT

We would like to acknowledge the institutional support for this research. We are grateful for the availability of data from BioProject PRJNA751478 (Galeeva et al., 2022).

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Al Kawas S, Al-Marzooq F, Rahman B, Shearston JA, Saad H, Benzina D, et al. (2021) The impact of smoking different tobacco types on the subgingival microbiome and periodontal health: a pilot study. *Sci Rep* 11(1): 1113. https://doi.org/10.1038/s41598-020-80937-3.

Anderson M, Ellingsen K, McArdle B. Anderson MJ, Ellingsen KE, McArdle BH (2006) Multivariate dispersion as a measure of beta diversity. *Ecol Lett* 9: 683–693. https://doi.org/10.1111/j.1461-0248.2006.00926.x.

Bach L, Ram A, Ijaz UZ, Evans TJ, Haydon DT, Lindström J (2023) The Effects of Smoking on Human Pharynx Microbiota Composition and Stability. *Microbiol Spectr* 11(2): e0216621. https://doi.org/10.1128/spectrum.02166-21.

Bai X, Narayanan A, Skagerberg M, Ceña-Diez R, Giske CG, Strålin K, et al. (2022) Characterization of the Upper Respiratory Bacterial Microbiome in Critically Ill COVID-19 Patients. *Biomedicines* 10(5): 982. https://doi.org/10.3390/biomedicines10050982.

Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, et al. (2015) Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *mBio* 6(2): e00037. https://doi.org/10.1128/mBio.00037-15.

Bello MGD, Knight R, Gilbert JA, Blaser MJ (2018) Preserving microbial diversity. *Science* 362(6410): 33–4. https://doi.org/10.1126/science.aau8816.

Berg G, Rybakova D, Fischer D, Cernava T, Vergès MCC, Charles T, et al. (2020) Microbiome definition re-visited: old concepts and new challenges. *Microbiome* 8(1): 103. https://doi.org/10.1186/s40168-020-00875-0.

Besutti G, Djuric O, Ottone M, Monelli F, Lazzari P, Ascari F, et al. (2022) Imaging-based indices combining disease severity and time from disease onset to predict COVID-19 mortality: A cohort study. *PLoS One* 17(6): e0270111. https://doi.org/10.1371/journal.pone.0270111.

Bisanz J (2023) Tutorial: Integrating QIIME2 and R for data visualization and analysis using qiime2R (v0.99.6). Available from: https://github.com/jbisanz/qiime2R.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37(8): 852–7. https://doi.org/10.1038/s41587-019-0209-9.

Brook I (2002) Anaerobic Bacteria in Upper Respiratory Tract and other Head and Neck Infections. *Ann Otol Rhinol Laryngol* 111(5): 430–40. https://doi.org/10.1177/000348940211100508.

Brook I, Gober AE (2005) Recovery of potential pathogens and interfering bacteria in the nasopharynx of smokers and nonsmokers. *Chest* 127(6): 2072–5. https://doi.org/10.1378/chest.127.6.2072.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13(7): 581–3. https://doi.org/10.1038/nmeth.3869.

Cascella M, Rajnik M, Aleem A, Dulebohn SC, Di Napoli R (2023) Features, Evaluation, and Treatment of Coronavirus (COVID-19). *Treasure Island (FL): StatPearls Publishing*

Available from: http://www.ncbi.nlm.nih.gov/books/NBK554776/.

Chakraborty S (2020) Metagenome of SARS-Cov2 patients in Shenzhen with travel to Wuhan shows a wide range of species - Lautropia, Cutibacterium, Haemophilus being most abundant - and Campylobacter explaining diarrhea. *OSF Preprints* Available from: https://osf.io/jegwq/.

Chen J, Liu X, Liu W, Yang C, Jia R, Ke Y, et al. (2022) Comparison of the respiratory tract microbiome in hospitalized COVID-19 patients with different disease severity. *J Med Virol* 94(11): 5284–93. https://doi.org/10.1002/jmv.28002.

Dickson RP, Huffnagle GB (2015) The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease. *PLoS Pathog* 11(7): e1004923. https://doi.org/10.1371/journal.ppat.1004923.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16):2194–200. https://doi.org/10.1093/bioinformatics/btr381.

Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, et al. (2014) Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* 111(22): E2329-2338. https://doi.org/10.1073/pnas.1319284111.

Galeeva J, Babenko V, Bakhtyev R, Baklaushev V, Balykova L, Bashkirov P, et al. (2022) 16S rRNA gene sequencing data of the upper respiratory tract microbiome in the SARS-CoV-2 infected patients. *Data Brief* 40: 107770. https://doi.org/10.1016/j.dib.2021.107770.

ggplot2: Elegant Graphics for Data Analysis. *SpringerLink* Available from: https://link.springer.com/book/10.1007/978-3-319-24277-4.

Gopinath D, Wie CC, Banerjee M, Thangavelu L, Kumar R P, Nallaswamy D, et al. (2022) Compositional profile of mucosal bacteriome of smokers and smokeless tobacco users. *Clin Oral Investig* 26(2): 1647–56. https://doi.org/10.1007/s00784-021-04137-7.

Greenblum S, Turnbaugh PJ, Borenstein E (2012) Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc Natl Acad Sci U S A* 109(2): 594–9. https://doi.org/10.1073/pnas.1116053109.

Han H, Luo Q, Mo F, Long L, Zheng W (2020) SARS-CoV-2 RNA more readily detected in induced sputum than in throat swabs of convalescent COVID-19 patients. *Lancet Infect Dis* 20(6): 655–6. https://doi.org/10.1016/S1473-3099(20)30174-2.

Hauptmann M, Schaible UE (2016) Linking microbiota and respiratory disease. *FEBS Lett* 590(21): 3721–38. https://doi.org/10.1002/1873-3468.12421.

Hernández-Terán A, Mejía-Nepomuceno F, Herrera MT, Barreto O, García E, Castillejos M, et al. (2021) Dysbiosis and structural disruption of the respiratory microbiota in COVID-19 patients with severe and fatal outcomes. *Sci Rep* 11(1): 21297. https://doi.org/10.1038/s41598-021-00851-0.

Kimura M, Tanikawa T, Suzuki M, Koizumi N, Kamiyama T, Imaoka K, et al. (2008) Detection of Streptobacillus spp. in feral rats by specific polymerase chain reaction. *Microbiol Immunol* 52(1): 9–15. https://doi.org/10.1111/j.1348-0421.2008.00005.x.

Koutsakos M, Kedzierska K (2020) A race to determine what drives COVID-19 severity. *Nature* 583(7816): 366–8. https://doi.org/10.1038/d41586-020-01915-3.

Lau SKP, Chan JFW, Tsang CC, Chan SM, Ho ML, Que TL, et al. (2016) Human oropharynx as natural reservoir of Streptobacillus hongkongensis. *Sci Rep* 6: 24419. https://doi.org/10.1038/srep24419.

Lin H, Peddada SD (2020) Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ Biofilms*

Microbiomes 6(1): 60. https://doi.org/10.1038/s41522-020-00160-w.

Lloréns-Rico V, Gregory AC, Van Weyenbergh J, Jansen S, Van Buyten T, Qian J, et al. (2021) Clinical practices underlie COVID-19 patient respiratory microbiome composition and its interactions with the host. *Nat Commun* 12(1): 6243. https://doi.org/10.1038/s41467-021-26500-8.

Lynch S (2016) The Lung Microbiome and Airway Disease. *Ann Am Thorac Soc* 13: S462–5. https://doi.org/10.1513/AnnalsATS.201605-356AW.

Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD (2015) Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis* 26: 27663. https://doi.org/10.3402/mehd.v26.27663.

Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. (2013) Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med* 187(10): 1067–75. https://doi.org/10.1164/rccm.201210-1913OC.

Paegle RD, Tewari RP, Bernhard WN, Peters E (1976) Microbial flora of the larynx, trachea, and large intestine of the rat after long-term inhalation of 100 per cent oxygen. *Anesthesiology* 44(4): 287–90. https://doi.org/10.1097/00000542-197604000-00002.

Oksanen J, Simpson G, Blanchet FG, Kindt R, Legendre P, Minchin P, et al. (2022) vegan community ecology package version 2.6-2 April 2022.

Paropkari AD, Leblebicioglu B, Christian LM, Kumar PS (2016) Smoking, pregnancy and the subgingival microbiome. *Sci Rep* 6(1): 30388. https://doi.org/10.1038/srep30388.

Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT (2014) Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One* 9(4): e93827. https://doi.org/10.1371/journal.pone.0093827.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41: D590-596. https://doi.org/10.1093/nar/gks1219.

R: The R Project for Statistical Computing Available from: https://www.r-project.org/.

Reddy RK, Charles WN, Sklavounos A, Dutt A, Seed PT, Khajuria A (2021) The effect of smoking on COVID-19 severity: A systematic review and meta-analysis. *J Med Virol* 93(2): 1045–56. https://doi.org/10.1002/jmv.26389.

Robinot R, Hubert M, de Melo GD, Lazarini F, Bruel T, Smith N, et al. (2021) SARS-CoV-2 infection induces the dedifferentiation of multiciliated cells and impairs mucociliary clearance. *Nat Commun* 12(1): 4354. https://doi.org/10.1038/s41467-021-24521-x.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4: e2584. https://doi.org/10.7717/peerj.2584.

Rosas-Salazar C, Kimura KS, Shilts MH, Strickland BA, Freeman MH, Wessinger BC, et al. (2021) SARS-CoV-2 infection and viral load are associated with the upper respiratory tract microbiome. *J Allergy Clin Immunol* 147(4): 1226-1233.e2. https://doi.org/10.1016/j.jaci.2021.02.001.

Shi Y, Wang G, Cai XP, Deng JW, Zheng L, Zhu HH, et al. (2020) An overview of COVID-19. *J Zhejiang Univ Sci B* 21(5): 343–60. https://doi.org/10.1631/jzus.B2000083.

Stämpfli MR, Anderson GP (2009) How cigarette smoke skews immune responses to promote infection, lung disease and cancer. *Nat Rev Immunol* 9(5): 377–84. https://doi.org/10.1038/nri2530.

Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, et al. (2015) The infant nasopharyngeal

microbiome impacts severity of lower respiratory infection and risk of asthma development. *Cell Host Microbe* 17(5): 704–15. https://doi.org/10.1016/j.chom.2015.03.008.

Thukral A (2017) A review on measurement of Alpha diversity in biology. *Agric Res* 54:1. https://doi.org/10.5958/2395-146X.2017.00001.1.

vegan: an R package for community ecologists. *Vegandevs* Available from: https://github.com/vegandevs/vegan.

Ventero MP, Cuadrat RRC, Vidal I, Andrade BGN, Molina-Pardines C, Haro-Moreno JM, et al. (2021) Nasopharyngeal Microbial Communities of Patients Infected With SARS-CoV-2 That Developed COVID-19. *Front Microbiol* 12: 637430. https://doi.org/10.3389/fmicb.2021.637430.

Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. (2019) Welcome to the Tidyverse. *J Open Source Softw* 4: 1686. https://doi.org/10.21105/joss.01686.

Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* 42: D643-648. https://doi.org/10.1093/nar/gkt1209.

Yu G, Phillips S, Gail MH, Goedert JJ, Humphrys MS, Ravel J, et al. (2017) The effect of cigarette smoking on the oral and nasal microbiota. *Microbiome* 5(1): 3. https://doi.org/10.1186/s40168-016-0226-6.

Zhao Q, Meng M, Kumar R, Wu Y, Huang J, Lian N, et al. (2020) The impact of COPD and smoking history on the severity of COVID-19: A systemic review and meta-analysis. *J Med Virol* 92(10): 1915–21. https://doi.org/10.1002/jmv.25889.

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 382(8): 727–33. https://doi.org/10.1056/NEJMoa2001017.