Characterising nasal microbiome profiles on the basis of influenza exposure and vaccination among healthy adult volunteers in South India

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Varsha Jaisimha



Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2022

Supervisor: Dr. Gagandeep Kang
© Varsha Jaisimha 2022

All rights reserved

Certificate

This is to certify that this dissertation entitled **"Characterising nasal microbiome profiles on the basis of influenza exposure and vaccination among healthy adult volunteers in South India"** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Varsha Jaisimhaat Indian Institute of Science Education and Research under the supervision of Dr. Gagandeep Kang, Professor of Microbiology. Division of Gastrointestinal Sciences, during the academic year 2021-2022.

Dr. Gagandeep Kang

Committee:

Dr. Gagandeep Kang

Dr. Satyajit Rath

This thesis is dedicated to my family, both genetic and otherwise!

Declaration

I hereby declare that the matter embodied in the report entitled Characterising nasal microbiome profiles on the basis of influenza exposure and vaccination among healthy adult volunteers in South India are the results of the work carried out by me at the Division of Gastrointestinal Sciences, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Gagandeep Kang and the same has not been submitted elsewhere for any other degree.



Varsha Jaisimha

Acknowledgments

This project would not have been possible without the effort and support of many incredible people. First and foremost, I would like to thank the field team and lab team at Wellcome Trust Research Laboratory, CMC Vellore, who have helped make this project possible despite the various hardships and challenges faced.

The field team consisted of Associate Research Officer (ARO), Ms. Ranjitha L; nurses, Ms. Shenbagavalli and Ms. Meghala and field workers, Mr. Karunagaran and Mr. Srinivasan. The team was involved in screening, enrolment, vaccine administration, participant evaluation, sample collection and transportation. The lab team consisting of Associate Research Officers, Ms. Blossom Benny, Ms. Revathi R and Ms. Zayina Zondervenni, was invaluable in processing the samples and familiarising me with the lab environment. I am grateful to Dr. Gayathri, Dr. Ramya and the other co-investigators of the parent study for their support and cooperation.

I would like to thank my mentors Dr Dilip Abraham, Dr Gagandeep Kang and Dr Satyajit Rath, for their constant guidance, encouragement and valuable feedback.

My family has supported me through all of my ups and downs, and without them, I would not have been in a position to produce this thesis. I am extremely grateful for their unwavering love and for putting up with all my tantrums through these years.

My friends have made me cherish every moment of this bittersweet journey and have given me a massive pile of endearing memories to carry on. I think it is about time I said "thank you" to these amazing folks who have had my back and continue to do so.

The support and the resources provided by 'PARAM Brahma Facility' under the National Supercomputing Mission, Government of India at the Indian Institute of Science Education and Research (IISER) Pune are gratefully acknowledged. I would like to extend a special thank you to Mr. Rishabh Umesh Lonjuari for the IT support while using the cluster.

Finally, I would like to sincerely thank every single person who has directly or indirectly contributed to this thesis. No matter how small, your contribution is an important Jenga block on which my work stands.

Abstract

The nasal microbiome and its dynamics appear to play a role in the progression of respiratory illnesses. Understanding these changes would contribute to the mechanistic deconvolution of vaccine-related pathways, open up the possibility of identifying bacterial biomarkers as correlates of protection and may have broader implications on vaccine design. To explore these dynamics further, we characterised changes in the nasal microbiome in response to the trivalent-inactivated influenza vaccine, Cadiflu-S and compared microbial profiles between individuals with high and low immunity to influenza (as inferred from serum antibody titres). Nasal swabs were used to sample the participant's anterior nares on Day 0 before vaccine administration and on Day 8 and Day 30 post-vaccination. A single swab was collected from the unvaccinated participants for cross-sectional analysis of the nasal microbiome in the context of immunity to influenza. Microbiome profiling was done using 16S rRNA gene amplification and next-generation sequencing, and mucosal IgA levels were measured through ELISA.

Significant differences were noted in alpha-diversity (p-value = 0.0025, Shannon) and beta diversity (p-value = 0.004, Weighted UniFrac) between influenza-associated immune and non-immune individuals. Individuals with the highest antibody titres demonstrated a significant decrease in alpha diversity when compared to the non-immune. No changes were observed between samples collected at different time points prior to and post-vaccination. Significant correlation was established between alpha diversity measures and total mucosal IgA concentrations (Correlation coefficient = -0.1835 and p-value = 0.02, Shannon). These results hint that both the local and systemic immune response of an individual may play a major role in shaping the microbiota. However, we are still a long way from establishing the nature of the relationship and the immune-microbiome interplay. While the results look appealing, one must keep in mind that the presence of multiple confounders often leads to misleading results in such clinical studies.

Contents

Acknowledgements **i**

Abstract **vi**

1 Introduction **5**

1.1 Background 5

1.2 Study Design 12

2 Materials and Methods **16**

2.1 Field Work 16

2.2 Sample Processing in lab 20

2.3 Data analysis tools and techniques 25

3 Results **27**

3.1 Performance of taxonomic classifiers 27

3.2 Demographic Distribution 28

3.3 Microbiome comparisons among relevant groups 29

4 Discussion **45**

5 Conclusion **49**

Appendix **57**

List of Figures

1.1 Study Timeline 14

2.1 Sample collection in the field 20

2.2 Sample processing pipeline 20

2.3 Processing nasal samples in the infectious disease room 21

2.4 Data analysis pipeline 25

3.1 Taxa barplot representing taxonomic classification performed by weighted SILVA classification (upper panel) and unweighted SILVA classification (lower panel) at taxonomy level 4 28

3.2 Histogram depicting the age distribution in the participant pool 29

3.3 Boxplot illustrating variation in alpha diversity (Shannon index) among the influenza-associated immune and non-immune groups in baseline samples . . 30

3.4 ANCOM Volcano plot for influenza immunity comparisons of baseline samples. The significant taxon is seen in the upper right corner of the plot. . . . 32

3.5 Model accuracy for Influenza immunity classification 33

3.6 ROC Curves for the influenza immunity classifier 33

3.7 Group significance plots with Bray Curtis dissimilarity matrix and sample collection time-points 34

3.8 Volatility plot of Unweighted UniFrac PCoA metric (Axis 2) with time . . . 35

3.9 Longitudinal trend of Unweighted UniFrac PCoA metric (Axis 2) stratified by influenza immunity status. The orange line represents the Immune group while the blue line represents the Non-Immune group. 35

3.10 Longitudinal trend of Unweighted UniFrac PCoA metric (Axis 2) stratified by HAI status. The blue line in the figure represents HAI titre $\leq 1:10$, the orange line represents a titre = 1:20, the green line represents a titre = 1:40 and the red line indicates a titre $\geq 1:80$. 36

3.11 Comparison of influenza-specific IgA titre among vaccinated samples grouped according to time-point of collection 37

3.12 Participant-wise temporal influenza-IgA trend among the vaccinated group . 37

3.13 Shannon entropy correlation with total IgA concentration in baseline samples 38

3.14 Left Panel illustrates Shannon entropy distribution across genders and the Right Panel illustrates the Weighted UniFrac distance between the males and females microbial communities. 39

3.15 Shannon entropy distribution between the medicated and non-medicated group 40

3.16 Boxplots illustrating alpha diversity variation in light of the influenza immunity trend 41

3.17 Group significance plots with weighted UniFrac distance matrix with the vaccinated and unvaccinated groups 42

3.18 Comparison of total IgA concentration across sampling months 44

1 Shannon entropy distribution among groups differentially Immune to influenza in baseline samples 58

List of Tables

1.1 Grouping of the participants recruited in the study 13

2.1 Distribution of participants across the groups 17

2.2 PCR Conditions for First-Stage PCR 23

3.1 Distribution of participants according to sex and baseline immunity 29

3.2 Results of pairwise KW test done with influenza immunity levels using Shannon index with baseline samples. HAI titre $\leq$ 1: 10 is classified as high-positive, = 1: 20 is equivocal, = 1: 40 is moderate-positive and $\geq$ 1: 80 is high-positive. 31

3.3 Results of pairwise KW test done with Shannon index 34

3.4 Percentile abundances of Sphingomonadales by influenza vaccination and immunity grouping at level 4 in baseline samples 43

1 Comparing the performance of the weighted taxonomic classifier vs the unweighted taxonomic classifier, both of which were trained on the SILVA database. 57

2 Pairwise-comparisons (PERMANOVA) of the different influenza-immunity groups with Unweighted UniFrac distances in baseline samples 58

3 Pairwise comparisons using Wilcoxon rank-sum test of total IgA levels based on our knowledge of recent exposure to the influenza virus 59

4 Pairwise-comparisons (KW test) of the immune and non-immune category stratified based on influenza vaccine status with Shannon's index in baseline samples 59

5 Pairwise-comparisons (PERMANOVA) of the immune and non-immune category stratified based on influenza vaccine status with Unweighted UniFrac distance in baseline samples 60

6 Percentile abundances of Sphingomonadaceae by influenza vaccination and immunity grouping at level 6 in baseline samples 60

7 Percentile abundances of KI89A_clade by influenza vaccination and immunity grouping at level 6 in baseline samples 60

Chapter 1

Introduction

1.1 Background

Our understanding of the biotic world around us has expanded exponentially in the last century. Every day, we learn more and more about the complexity of this web that encompasses the living world. With advancements in microscopy and next-generation sequencing technologies, we have uncovered the "World of the Microbiota"!

1.1.1 The Microbiome

The "microbiota" is a community of microorganisms that inhabit a specific region, such as a body part, an environmental habitat or an ecosystem (Thursby and Juge, 2017). The "microbiome" refers to the collective set of genomes and gene products of all the organisms in a specified microbiotal system (Turnbaugh et al., 2007). However, these two terms are often used interchangeably by the scientific community, and this practice has been followed here as well.

The Human Microbiome Project (HMP) was set up to establish the reference database for the human microbiome. One of the "surprising" discoveries by HMP was that commensal microorganisms contribute more genes vital to human survival than genes derived from the human genome (Methé et al., 2012).

It has been estimated that the human microbiome contributes around 100 times the number of genes than the human genome. However, a single individual may only possess about 15% of the entire collection of microbial genes identified in humans. This illustrates the magnitude of inter-individual variability witnessed (Qin et al., 2010).

Numerous environmental factors have been shown to modulate the composition of the microbiome. While the microbiota appears to be more or less stable for the majority of adult life in a healthy human being, distinct transient changes are seen by increasing temporal resolution. Therefore it is essential to identify and functionally assign the resident microbiota to elucidate key features that are important in health and disease to harness this knowledge to improve human health (Allaband et al., 2019).

1.1.2 Broad implications of microbiome studies

The amount of scientific literature drawing associations between the microbiome and human health has skyrocketed in the past decade. A wide range of diseases have been connected to dysbiosis. Dysbiosis refers to the disruption of the microbiota and its state of homeostasis. Conditions associated with dysbiosis include autoimmune conditions, allergies and asthma (Rivas et al., 2013; Kostic et al., 2015; Zhang et al., 2015), obesity and cardiovascular diseases (Goodrich et al., 2014b), mental illnesses and depression (Luna and Foster, 2015; Dash et al., 2015; Wang et al., 2011). However, integrating these studies becomes challenging given the difference in setup, techniques and analysis. Another challenge is making the leap from associating microbiota and human disease states to establishing a causal relationship. With human studies, it is impossible to distinguish whether the disease drives dysbiosis, the microbiome propels the disease, or if both are affected by a confounding factor such as inflammation (Debelius et al., 2016).

While microbiome studies have immense potential, they do come with numerous limitations. Majority of the techniques used to assign taxonomy and characterize bacteria are not precise enough. Many are able to identify various species, but different strains from the same species may have very different effects on human health. Complete functional annotation is often limited to the set of organisms that have been extensively characterized while the vast majority are yet to be studied. Another challenge is the extent of sampling. Often because of financial restrictions, the number of samples included is very low, which does not allow

for a reasonable approximation of the environment being sampled. Conclusions from studies with a low sample size result in many false negatives and false positives (Liwinski et al., 2021). A fundamental limitation of metagenomic microbiome studies is the inability to link genes of interest to a specific taxa, or phenotype since we are not able to measure actual gene expression. In particular, the big challenge is to associate specific microbial groups, strains or genes to a specific function. Functional readouts as obtained from transcriptomics* or metabolomics** studies have less methodological variation than taxonomic readouts, and are more reproducible and informative (Inkpen et al., 2017). Given that there are so many factors that influence the microbiome, it is especially difficult to find changes in the microbiome composition that actually cause a particular disease. Because of this variability, many studies in the field are inconsistent and yield contradictory results. In most instances, the relationship between the microbiome, health and disease is simply meant as an analysis of associations with disease or functional disturbances. Little is understood about how human-associated microorganisms causally drive host phenotypes or whether the phenotype drives the microbial variations. Longitudinal studies attempt to establish stronger correlations by showing temporal compliance but do not come close to inferring causality given the vast multitude of factors and pathways involved with the microbiome (Neville et al., 2018). Only in a handful of studies have cause-and-effect relations been established (satisfying Koch's postulates) but these are mainly done in animal models (Evans, 1978). Recent research, in some cases guided by hypotheses developed from these correlative studies, has begun to establish strain- and molecular-level connections between the microbiome and disease. A popular approach for providing causal evidence is the use of probiotic interventions. While these are great at assessing the impact of a species on host physiology, they are unable to evaluate interactions on an ecosystem level. Knowledge of causal relationships is critical to understand how the microbiota affects human health and to develop new therapies to treat disease (de Vos and de Vos, 2012; Chaudhari et al., 2021).

*Transcriptomics is the analysis of a complete set of transcripts in a specific type of cell or tissue of interest (Mutz et al., 2013).

** Metabolomics is the identification and quantification of all metabolites in a cellular system of interest (Monteiro et al., 2013).

1.1.3 Techniques used in microbiome studies

Comprehensive examination of enormously complex communities of commensal microbes has been made possible with the development of affordable and accessible next-generation sequencing technologies. This has taken over methods dependent on laboratory culturing, which was limited to the identification of organisms that could be artificially cultured (Suau et al., 1999). Microbiome profiling is done by taxonomic classification performed on sequencing data of the gene encoding the 16S ribosomal ribonucleic acid (rRNA). This gene, found in prokaryotes, codes for the 30S subunit of the ribosomes. It is found to be conserved within archaeal and bacteria species, allowing for selective screening of the microbiome. Apart from this, the gene segment consists of 9 variable regions interspersed between conserved regions, making it vital for determining the microbial community profile. Sequences of the variable regions help identify the phylogenetic classification, such as genus or species in the microbial population of a given sample. In contrast, the conserved regions allow for amplification with common primers (Xu et al., 2012). The choice of which variable region to amplify significantly affects the analysis based on the ability of the variable region to distinguish microbial taxa.

Most microbiome studies look at the differential abundance of microbial taxa, community diversity or functional components such as gene products and pathways between the groups under consideration, such as treatment vs control or diseased vs healthy. Common measures used for microbiome comparisons include alpha diversity, beta diversity and differentially abundant taxa. Alpha diversity measures estimate the diversity within a single sample and are specific to each sample. Different indices reflect different combinations of the richness (number of species) or distribution (evenness of species) in a microbial sample. Commonly used alpha diversity indices are Shannon entropy and Simpson's index, both of which calculate diversity based on richness and evenness. Phylogenetic indices such as Faith's phylogenetic diversity (PD) include phylogenetic information when determining sample diversity. Beta diversity is a measure of the dissimilarity or similarity among two communities. Bray-Curtis dissimilarity is a measure of beta diversity that considers both overall abundance per sample and the abundance of individual taxa of the communities. UniFrac distances estimate beta diversity by taking phylogenetic information into account. Unweighted UniFrac is purely based on calculating the branch length shared between two samples, while weighted UniFrac adds relative abundances as weights to the branch lengths (Galloway-Peña and Hanson, 2020).

The methods used for statistical analysis of the complex microbiome dataset is a continuously evolving field. Common biostatistical methods cannot be employed for microbiome datasets given the high dimensionality, zero inflation, differences in sequencing depth among samples, and the

compositional nature. High dimensionality is because potentially thousands of taxonomic units can be found in the dataset, and zero inflation is due to most taxa being rare. Compositionality refers to the use of relative abundances of taxa since we are not sure of the true bacterial count in the community. Besides the nature of the data, a major limitation is the inability to control intrinsic confounders in clinical cohort studies and the lack of statistical power (Galloway-Peña and Guindani, 2018).

Given the nature of the microbiome data, variation in any step of the protocol of sample processing and data analysis will produce different results and render the two studies incomparable. Different data analysis techniques yield substantial differences in the results obtained even when using the raw sequencing data from the same sequencing instrument (Goodrich et al., 2014a). Given the infancy of metagenomics and microbiome studies, no standard protocols have yet been established.

1.1.4 Nasal Microbiome and Respiratory Tract Infections (RTIs)

Compared to the gut microbiota, which is often the first love of microbiome researchers, much less is known about interactions among the nasal microbiota, respiratory host responses, and viral infection. The predominant phyla observed in the nasal microbiome are Firmicutes and Actinobacteria (Dimitri-Pinheiro et al., 2020).

The nasal microbiome and its dynamics appear to play a role in the progression of respiratory illnesses. A 2016 study in children showed an association between nasal microbiomes rich in Streptococcus spp./Haemophilus influenzae and respiratory syncytial virus (RSV) infection and overexpression of inflammation-associated markers (Steenhuijsen Piters et al., 2016). Nasal commensal bacteria have been found to suppress colonisation of opportunistic pathogens by producing toxic substances that inhibit the pathogen or hinder their growth by competing for limited space and nutrients. A recent molecular-profiling study (Hardy et al., 2019) has shown that species-specific interactions have a significant role in preventing Staphylococcus aureus persistence. It demonstrated that Corynebacterium inoculation in the nasal cavity inhibits colonisation by S. aureus.

Numerous studies done in mice found that probiotic application directly into the nostrils regulates local mucosal immune response against respiratory viral infections (Lehtinen et al., 2018). A randomised controlled trial in human adults (West et al., 2014) discovered that consumption of Bifidobacterium animalis subspp. lactis was associated with a reduced risk of RTIs for a period of 5 months.

1.1.5 Nasal Microbiome and Influenza Infection/Vaccination

The influenza virus infects the human respiratory tract and that of a number of other mammals. It often manifests clinically as respiratory tract inflammation and high fever, among other symptoms. Severe manifestations of the disease include secondary bacteria infection, bronchitis and pneumonia (Moghadami, 2017). Influenza is infamous for seasonal periods of illness and periodic pandemic outbreaks. An estimated 3 to 5 million cases of severe influenza infections leading to hospitalisation, and between 250,000 to 500,000 influenza-associated deaths occur annually, though this is considered an underestimation (Iuliano et al., 2018).

Currently, the most effective strategy to control influenza transmission is vaccination. However, the vaccines available in the market often show suboptimal efficacy due to a mismatch between the mispredicted vaccine strain and the circulating strain (Belongia et al., 2016). The rapid evolution of the influenza virus and immune escape strategies like antigenic variations also renders the vaccines ineffective. Moreover, the current vaccination strategy of a yearly dose is impractical and near impossible to implement in low and middle-income countries.

Our incomplete understanding of the molecular mechanisms of vaccine immunity and a lack of standardised assays to evaluate correlates of protection in clinical studies hinder effective vaccine development [Hirve and WHO (2015)]. This makes it vital to identify host factors and mechanisms that determine susceptibility to influenza. In this matter, the human microbiome shows immense potential. Studies done in both animals and humans suggest that the nasal and oral microbiota may modulate host immunity and vary susceptibility to influenza and other respiratory infections though we are still a long way from establishing solid principles (Tsang et al., 2020). Understanding these changes may have broader implications concerning vaccine design. Apart from contributing to the mechanistic deconvolution of vaccine functioning pathways (a huge feat by itself!), it may open up the possibility of identifying bacterial biomarkers as correlates of protection against influenza upon exposure or vaccination and serve as adjuvants in vaccines.

1.1.6 Relevant background work

Studies have investigated the changes in microbial community composition during infection and its corresponding effect on immune response with regards to influenza infection and vaccination. Often, these studies use live-attenuated influenza vaccines (LAIV) to mimic influenza infection. A study by Salk et al. in 2016 noted changes in the nasal microbiome composition with LAIV. The study found a significant increase in alpha-diversity post-vaccination (on Day 7 and Day 28)

than before vaccination (Day 0). It drew significant associations between the presence of specific microbial species and influenza-specific IgA antibody response. A notable positive correlation was found between influenza-specific IgA titers and two bacterial species - Lactobacillus helveticus and Bacteroides ovatus (Salk et al., 2016).

A paper in 2015 by Tarabichi et al. outlined changes in microbial community structure, diversity and core taxonomic membership and increase in the relative abundances of pathogenic species (members of Staphylococcus/Bacteroides genera) in the nasal mucosa upon LAIV administration. A significant increase in microbial diversity between visit 1 (baseline) and visit 2 (1-2 weeks post-vaccination) and between visit 1 and visit 3 (4-6 weeks post-vaccination) was seen in the vaccinated group but not in the control group. Microbial community structure analysis noted a reduction in variability among the LAIV treated subjects at visits 2 and 3, compared to visit 1, which was not seen in the control group. The study also analysed the nasal epithelium gene expression profiles and associated the activation of Type-I interferon-mediated genes with the emergence of pathogenic genera Staphylococcus and Bacteroides in the nasal community (Tarabichi et al., 2015).

The impact of the nasal microbiome on secondary bacterial infection in the context of influenza infection was discussed in the 2019 paper by Piters et al. Baseline nasal microbiota composition was associated with pneumococcal carriage outcome. The group that showed a low-density prevalence of S. pneumoniae showed the lowest levels of viral replication and microbiota perturbations following LAIV compared to non-carriers and high-density carriers and displayed microbiota profiles dominated by the genera Corynebacterium and Dolosigranulum. This suggests that baseline microbial profiles may serve as predictors of influenza disease progression and co-infection (Steenhuijsen Piters et al., 2019).

Longitudinal studies in humans have only been carried out in the context of live attenuated influenza vaccines (LAIV). Studies characterising the nasal microbiome with respect to the intramuscular inactivated influenza vaccine (the most commonly used type of influenza vaccine!) have been cross-sectional in nature and unable to note any significant trends due to the large inter-individual variability. A retrospective study by Ding et al. noted that among influenza patients, the nasal microbiome in the vaccinated elderly was less diverse than those not vaccinated; however, it pointed out that the retrospective nature of the study resulted in a high variation within each group which served as a limiting factor in establishing significant microbiome changes (Ding et al., 2019). A clinical trial (Rizzardini et al., 2012) demonstrated that pretreatment with probiotics (Bifidobacterium animalis and Lactobacillus paracasei) for six weeks before intramuscular trivalent influenza vaccination (TIV) led to significantly higher vaccine-specific IgG and salivary IgA antibody responses.

1.2 Study Design

This project aims to compare nasal microbiome profiles among individuals with and without immunity to the influenza virus and characterise changes in response to trivalent-inactivated intramuscular influenza vaccine **(Cadiflu-S)** administration.

The project was carried out at **Wellcome Trust Research Laboratory** (WTRL), Division of Gastrointestinal Sciences, Christian Medical College (CMC), Vellore. The work presented in the thesis was done as a sub-study under the study titled **"Study of humoral and cellular immune responses to influenza vaccine in healthy adult volunteers"** at WTRL, CMC Vellore. The project heavily relied upon the parent study for infrastructure, resources and funding.

1.2.1 Correlates of influenza immunity

Immunity to influenza was inferred from the participants' antibody titres determined by Hemagglutinin Inhibition (HAI) Assay. Hemagglutinin Inhibition Assays are used to quantify serum antibody levels against specific influenza viruses. The principle used in the assay is that influenza viruses bind to or hemagglutinate red blood cells (RBCs) thereby forming a "lattice" and preventing the RBCs from clumping (Hobson et al., 1972). Sera from patients with high levels of influenza-specific antibodies prevent hemagglutination of erythrocytes by the influenza virus and hence show more clumping. The HAI titre is the highest dilution of serum that prevents influenza virus-induced hemagglutination of RBCs (Gilbert et al., 2019). While HAI titres do not always accurately depict the complete immunity profile of an individual or determine their susceptibility to influenza infection. However, it is a commonly used correlate of protection (Katz et al., 2011) and will be used as a surrogate for immunity in this study.

For the purpose of this study, HAI titers $\leq$ 1:20 are considered as no or insignificant immunity to influenza (non-immune group), whereas HAI titers $>$ 1:20 are immune and this is considered to be indicative of prior exposure to the influenza virus either through infection or vaccination. This cutoff was chosen because the approved influenza vaccines seek to elicit a HAI titer greater than or equal 1:40 as the primary correlate of protection (Memoli et al., 2016).

1.2.2 Study Objectives

a Note changes in nasal microbiome diversity, structure and core genera in individuals before (baseline) and after vaccination (Day 8 and Day 30) with respect to the administration of the intramuscular trivalent inactivated vaccine (TIV), Cadiflu-S.

b Note differences in nasal microbiome diversity, structure and core genera in individuals with HAI titer $\leq 1:20$ (non-immune group) and $\geq 1:40$ (immune group).

c Draw associations between microbiome composition and local immune response (influenza-associated IgA levels), and identify differentially abundant bacterial genera that may serve as correlates of protection.

1.2.3 Layout of the study

Study population: Healthy adults between 18 to 65 years of age who are residents of Vellore, Tamil Nadu, India and of a low-middle income background.

The population under consideration was restricted to healthy individuals to reduce additional inter-individual variation brought about by dysbiosis, often correlated with the varying levels of inflammation in diseased conditions (Janssens et al., 2018). Healthy adults above the age of 65 years were not recruited since dysbiosis is often observed in ageing individuals and is associated with the onset of geriatric systemic low-grade inflammation and immunosenescence (Bosco and Noti, 2021).

The participants recruited for the study were divided into four different groups, as highlighted in Table 1.1.

Participant Groups	**Non-immune Group**	**Immune Group**
Vaccinated Group	Non-immune and Vaccinated	Immune and Vaccinated
Unvaccinated Groups	Non-immune and Unvaccinated	Immune and Unvaccinated

Table 1.1: Grouping of the participants recruited in the study

Sample collection kits consisting of **nasal swabs** and universal transport media (UTM) were used to collect microbiota samples from the anterior nares of the subjects. The anterior nares show

lesser microbial diversity than the nasopharyngeal region; hence the intragroup diversity is less likely to interfere with the statistical analysis while comparing across groups (Luna et al., 2018).

Three samples were collected from each participant in the **vaccinated group** across different time-points with respect to vaccine administration. The first sample (baseline sample) was collected pre-vaccination on Day 0. The second sample was collected on Day 8 post-vaccination since previous studies have shown that microbial changes appear to be the highest around 1-week post-vaccination (Salk et al., 2016; Tarabichi et al., 2015; Luna et al., 2018) and antibody response at this stage represents the immune response against the vaccine (Kotomina et al., 2020). The human microbiome is known to be robust and resistant to external pressures to an extent. The third sample on Day 30 post-vaccination was collected to allow us to observe the long term changes in the microbiome composition post-vaccination in the nasal environment. Microbiota composition is highly variable between individuals; thus, collecting samples in a longitudinal fashion will allow for comparing changes in the microbiota in the same individual before and after the event rather than comparing the microbiota composition directly between a vaccinated and a control group.

A single sample was collected from each participant in the **unvaccinated group** for cross-sectional analysis of the influence of immunity to influenza on the nasal microbiome.

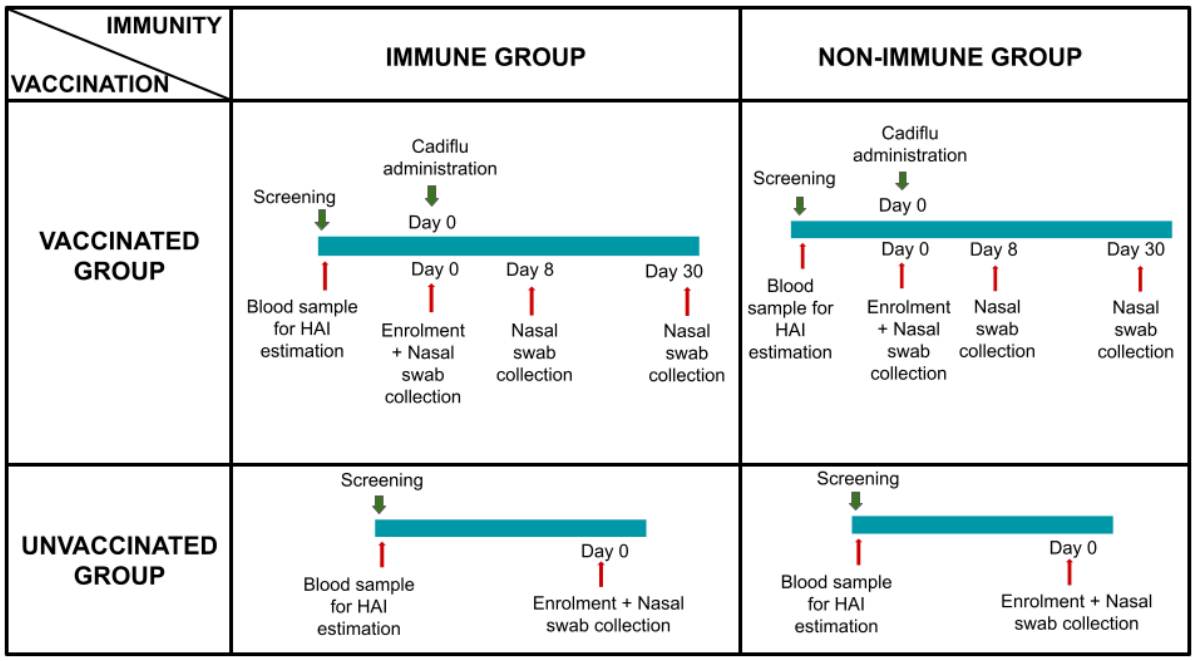


Figure 1.1: Study Timeline

1.2.4 Novelty of the study

To the best of my knowledge, this project is the first to study the nasal microbiome in the context of influenza infection and vaccination in the Indian population and socio-geographical setting. Also, the project explores variations in the nasal microbiota in the context of possible systemic changes brought about by TIV, which has not been well studied. This study looks at TIV-associated microbiome changes in a longitudinal manner, thus reducing the effect of inter-individual variation on the data analysis.

Chapter 2

Materials and Methods

The clinical study was reviewed and approved beforehand by the **Institutional Review Board** (IRB) at CMC Vellore. All revisions to the study design were reviewed and accepted by the IRB.

2.1 Field Work

The field component of the project involved setting up the study, screening, recruitment and consent taking, vaccine administration and sample collection in the clinic and the field. All of this was done with the help of the field team consisting of healthcare workers, nurses, doctors and field coordinators.

2.1.1 Screening

Screening was done to identify healthy adult volunteers through a preliminary health check-up and determine their immunity to influenza. It was carried out by the field team in the vicinity of the Chinnallapuram area in Vellore, Tamil Nadu. Adults between 18 to 65 years of age were screened for their general health status and serum influenza-specific antibody levels. The screening involved the following steps:

1. A questionnaire to obtain contact information, general identifiers and demographic data such as sex, age and socio-economic conditions.

2. Collection of data relevant to the study such as prior medical history, RTI infections in the past month, medication, antibiotic usage and vaccination history.

3. Measuring their blood pressure with a sphygmomanometer, random blood sugar with a glucometer and haemoglobin using a hematocrit. A blood sample was collected from each of them, and influenza-specific serum antibody levels were estimated with the HAI assay. The HAI assay was carried out at the Department of Virology, CMC Vellore.

2.1.2 Enrolment

The study purpose and the study plan were explained to the eligible participants in the regional language, Tamil and their written consent was taken.

Based on the screening results, **94** volunteers were recruited into the **unvaccinated group** to determine microbiota differences among influenza-related immune vs non-immune participants.

30 participants were enrolled into the **vaccination group**. Adult volunteers who are between 18 to 65 years of age, in good health (as evaluated by a physician by clinical history, examination, and investigations prior to enrolment) and those that did not meet any of the exclusion criteria were recruited. A repeat blood sample was collected for HAI estimation among the vaccinated participants prior to vaccine administration if the gap between screening and enrolment was greater than 3 months. This was done to ensure that any recent exposure to the influenza virus would reflect in the HAI titres, and the participants would be correctly classified as immune or non-immune.

Number of participants recruited	Non-immune Group	Immune Group
Vaccinated Group	18	12
Unvaccinated Groups	58	36

Table 2.1: Distribution of participants across the groups

Exclusion criteria

1. Pregnant women (pregnancy was ascertained in women of reproductive age by menstrual history and urine testing if necessary)
2. Volunteers who have been vaccinated against influenza in the past year (as ascertained from volunteer's recall or documentation if available)
3. Antibiotic usage within the last 2 weeks
4. Uncontrolled chronic illness or condition (including diabetes, hypertension)
5. Chronic respiratory conditions including chronic obstructive pulmonary disease (COPD), asthma and interstitial lung disease (ILD)
6. Current regular (daily) smoker
7. A long-term health problem with heart disease, lung disease, kidney disease, neurological disease, liver disease, metabolic disease, anaemia or other blood disorder.
8. Use of drugs for the treatment of rheumatoid arthritis, Crohn's disease, psoriasis, anticancer drugs or radiation treatments.
9. Live with or expect to have close contact with a person whose immune system is severely compromised and in protective isolation.

If the participants had received any vaccination in the last four weeks (usually the COVID-19 vaccine), they were recruited only 1-month post-vaccination. Participants who had received a single dose of the COVISHIELD vaccine were recruited a month after vaccination. This was done since the second dose was scheduled 3 months after the first dose, and hence they received their second dose a month after the final sample collection and two months post influenza vaccination. Participants with ongoing respiratory infections were recruited only after the participant recovered and showed no apparent symptoms (usually one week after). This was done to reduce any transient effect of the other vaccines taken or the RTI on the nasal microbial composition, which may interfere in establishing a baseline profile for the participant.

Patients with chronic illnesses such as diabetes, hypertension and hyperlipidemia were recruited if their condition was considered "in control" (based on clinic laboratory reports) with regular medication. If a participant's haemoglobin (Hb) levels were less than 10 mg/dl, they were given iron supplements and recruited later on when their Hb levels showed improvement.

Since we could not establish or support the above hypothesis, the microbiome composition and changes in medicated patients calculated were separated out and compared with that of other participants to decide if they introduce significant inter-individual variability and show a different immune response.

2.1.3 Vaccine administration

The influenza vaccine was administered intramuscularly to the participants recruited into the vaccine group on the day of enrolment. The vaccine given was the trivalent inactivated influenza vaccine, Cadiflu-S from Cadila Pharmaceuticals, consisting of **hemagglutinin antigen of influenza A/California/04/2009 (H1N1); A/Victoria/361/2011 (H3N2); and B/Massachusetts/02/2012**. Vaccine administration was done at the CMC Chinnallapuram clinic by the nurses.

2.1.4 Sample collection, transport and storage

Before each sample collection, the participants were asked questions regarding recent RTI history, antibiotic usage, presence of nasal piercing and changes in health condition since enrolment. From the participants in the vaccine group, one nasal swab (baseline) was collected just before TIV administration. Two follow-up nasal swabs were collected on Day 8 and Day 30 post-vaccination. A single swab (baseline) was collected from the participants in the unvaccinated group.

Sample collection was done either in the CMC Chinnallapuram clinic or in the field. Sample collection kits consisting of a Flocked Swab Copan® (FLOQSwab) and UTM™ were used to collect microbiota samples from the anterior nares of the subjects. A single swab was used to sample both the nostrils. Once collected, they are immediately placed in an icebox and transported to the lab within 2-3 hours. The samples were stored at 4℃ and transferred to -80℃ within 24 hours.



Figure 2.1: Sample collection in the field

2.2 Sample Processing in lab

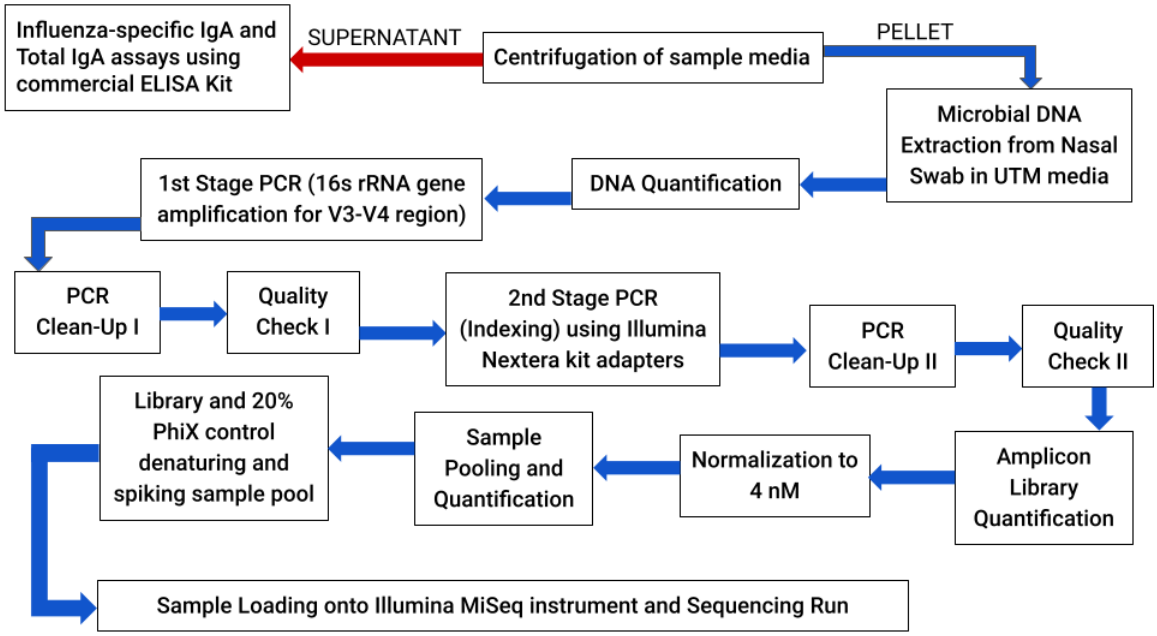


Figure 2.2: Sample processing pipeline

2.2.1 Initial sample processing

The samples were left to thaw at room temperature and vortexed with SPINIX^{TM} Vortex Shaker (TARSONS) for 2-3 minutes. The swabs were discarded and the UTM solution was centrifuged at 14,000 rpm for 20 minutes (Eppendorf Centrifuge 5418R). The supernatant was transferred to another tube for IgA estimation. The pellet was resuspended in 1x PBS for DNA extraction and subsequent 16S rRNA metagenomics.



Figure 2.3: Processing nasal samples in the infectious disease room

2.2.2 IgA ELISA Assays

For mucosal immune response evaluation, nasal IgA antibody levels are recorded from the nasal samples. The measurement of pathogen-specific IgA titres (for the influenza virus) is done using the **Human Influenza A virus IgA antibody (FluA-Ab-IgA) ELISA Kit** from MyBioSource.Inc (Catalogue # MBS2800304). Total IgA titres are measured using **Human Immunoglobulin A, IgA ELISA Kit** from MyBioSource.Inc (Catalogue # MBS704777). The standard kit protocol outlined in the Sample Manual is followed without any modifications.

Briefly, 100 µl of each sample and controls were loaded onto a 96 well plate pre-coated with influenza-A antigen. After a 30 minute incubation at 37°C, the plated was washed with wash buffer three times. HRP-conjugate was added, followed by a second incubation for 30 minutes at 37°C. The plate was washed again 5 times, and the substrate was added, followed by a brief incubation. 50 µl of stop solution was added, and the optical density at 450 nm (with correction at 570 nm) was measured in an Epoch 2 Microplate Spectrophotometer (BioTek Instruments). A similar protocol

was followed for total IgA detection, with a longer incubation time of 2 hours at 37℃ and an ELISA plate pre-coated with IgA-specific antibodies.

2.2.3 Microbiome detection with 16S amplicon sequencing

DNA extraction was done with the resuspended pellet of the processed sample. **QIAamp DNA Mini Kit** from Qiagen (Catalogue # 51306) was used, and the standard kit protocol outlined in the kit handbook was followed without any protocol modifications.

16S rRNA library preparation

Library preparation for sequencing was done using the 16S Metagenomics Sequencing Library preparation protocol outlined by Illumina. The protocol mentioned above (standardised for gut microbiome) was modified to account for low biomass from nasal samples.

The steps involved in library preparation are outlined below.

A **DNA Quantification**

Initial quantification of double-stranded DNA (dsDNA) concentrations in the samples was performed using the Qubit^{TM} dsDNA High Sensitivity Kit and the Qubit 3.0 Fluorometer by Invitrogen (Catalogue # Q33216).

B **First-Stage PCR**

16S amplification PCR was carried out with the primers for the V3-V4 region (341F/805R) of the the16S rRNA gene. We obtained an amplicon of approximately 460 bp in length. The KAPA HiFi HotStart ReadyMix Kit was used. Given the low sample DNA concentration, 10 ul of the sample was added to ensure adequate template DNA. One PCR reaction consisted of 12.5 ul of KAPA HiFi HotStart ReadyMix, 0.5 µl each of 10 µM forward and reverse primers and 1.5 µl of Nuclease Free Water (NFW). The PCR conditions set in the Agilent SureCycler 8800 Thermocycler (Catalogue # G8800) are shown in Table 2.2.

Steps		Temperature	Time
Initial denaturation		95 ºC	3 min
Cycles (35)	Denaturation	95 ºC	30 s
	Annealing	55 ºC	30 s
	Extension	72 ºC	30 s
Final extension		72 ºC	5 min
Hold		4 ºC	∞

Table 2.2: PCR Conditions for First-Stage PCR

C **Post-PCR Cleanup 1 and Quality Check**

Post-PCR cleanup was done with **AMPure XP beads** (Beckman Coulter). 20µl of the beads mix was added to each sample well and incubated at room temperature for 5 minutes and on a magnetic stand for 2 minutes. The clear supernatant was discarded, and two 80% ethanol washes were done with 200µl added to each well and a 30-second incubation in-between. After a 10-minute air dry, 52.5µl of NFW was added. The resuspended samples were allowed to incubate at room temperature for 2 minutes, and 50µl of the sample was eluted. A quality check was done with the **Agilent 4200 TapeStation** System (Catalogue # G2991BA) for verifying the presence of a clear band of length 550 bp.

D **Second-Stage PCR**

To allow for pooling of all samples before sequencing; an indexing PCR is done to add barcode sequences to the samples. The primers used in indexing PCR also had overhang-adapter sequences for compatibility with index and sequencing adapters. A PCR reaction contained 10 µl of NFW, 25 µl of **KAPA HiFi HotStart ReadyMix**, 5 µl each of forward and reverse barcoded primers taken from the **Illumina Nextera XT DNA Library Preparation Kit** and 50 µl of cleaned 16S PCR amplicon. The PCR was set up with the same conditions as above except for 8 cycles instead of 35 to allow for attachment of indexes without much amplification.

E **Post-PCR Cleanup 2 and Quality Check**

A clean-up with AMPure XP beads using the same protocol described in section C. The only protocol modifications were the addition of 56 µl of AMPure beads instead of 20 µl, the addition of 27.5 µl of NFW and the elution of 25 µl of cleaned product. A tape-station run was performed to confirm the presence of a clean band at 630 bp.

F **Library quantification and normalisation**

Library quantification was done with the Qubit Fluorometer. Post-quantification samples were normalised to 4 nM with NFW and were quantified again to ensure the correct concentration.

G **Library pooling, denaturation and PhiX addition**

The normalised library was pooled and was quantified with the Qubit Fluorometer, Agilent Tapestation and real-time PCR (RT-PCR) with the KAPA Library Quantification Kit. The RT-PCR reaction mixture contained 12.4 µl of mastermix, 3.6 µl of NFW and 4 µl of purified indexed amplicon. The PCR conditions set were initial denaturation at 95 °C for 5 minutes and 35 cycles of 30 seconds at 95 °C and 45 seconds at 60 °C. This was carried out in Applied Biosystems **QuantStudio 7 Flex Real-Time PCR** System (Catalogue # 4485688). Denaturation of the pooled library was done with 2N NaOH and 20% denatured PhiX control was spiked in. 600 µl of 8 pM pooled sample was loaded into the flow cell, and a sequencing run on the **Illumina MiSeq™ System** (Catalogue # SY-410-1003) was performed.

2.3 Data analysis tools and techniques

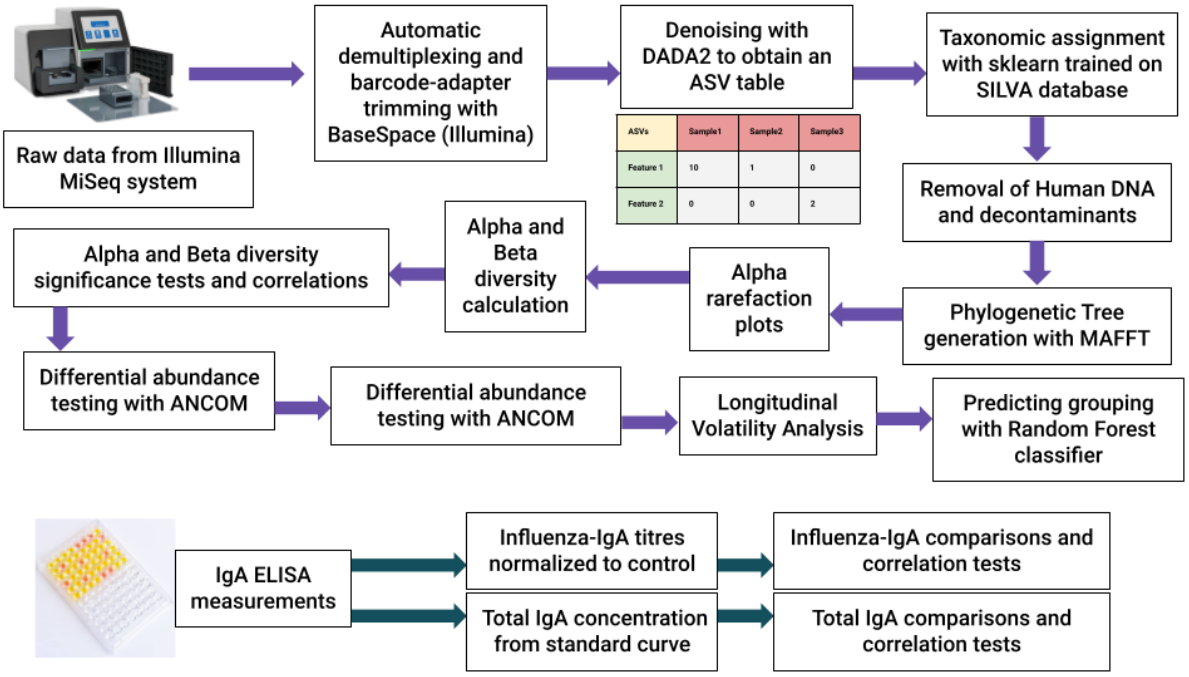


Figure 2.4: Data analysis pipeline

The bioinformatic pipeline, QIIME2 version 2022.2 (Bolyen et al., 2019) on Miniconda 3 version 4.12.0 and Rstudio version = 4.1.3 were used for data analysis. QIIME 2 is an open-source next-generation microbiome bioinformatics platform. The computationally intensive steps were carried out on the 'PARAM Brahma' Super Computer Cluster at the Indian Institute of Science Education and Research (IISER), Pune.

Raw reads obtained from the Illumina MiSeq Sequencer were automatically uploaded onto Illumina BaseSpace Portal. Demultiplexed FastQ files containing the sample IDs were generated with the sequencing adapters and barcodes trimmed out. The FastQ files were processed in QIIME 2 to combine each set of paired-end reads into a single sequence and filter out poor quality sequencing reads. The filtered reads were used to infer **Amplicon Sequence Variants** or ASVs by dividing the amplicon reads into multiple partitions based on the error model considered. This was done using Divisive Amplicon Denoising Algorithm 2 or **DADA2** (Callahan et al., 2016). The ASVs were assigned taxonomy using a naïve Bayes classifier pre-trained on the SILVA 16S database (Pruesse et al., 2007; Glöckner et al., 2017) using the scikit-learn classifier version 0.24.1 (Robeson et al., 2020; Bokulich et al., 2018b).

Removal of human DNA was done in Rstudio with batch BLAST analysis against the Human

RefSeqGene sequences database. The sequences included in the BLAST were taxa with only 1 level of taxonomy assigned (d_Bacteria) and Unassigned sequences. All archaeal samples were removed since primers used for amplicon sequencing had a higher specificity to bacterial species than archaea. Removal of suspected decontaminant sequences was done with the decontam package using the combined method in Rstudio (Davis et al., 2018). The decontam package classifies features as contaminants on two principles. Principle 1: Contaminating sequences are likely to show an inverse correlation of sequence frequencies with the DNA concentration in the sample. Principle 2: Contaminating taxa are likely to be more prevalent in the negative controls than in the true sample. All denoised samples with no sequences were removed, and the feature table was modified to a uniform sequencing depth of 7250 across all samples.

The QIIME 2 pipeline was used to generate alpha and beta diversity measures, perform statistical comparisons and evaluate differentially abundance taxa with ANCOM. Alpha diversity was measured using Shannon, Simpson and Faith PD. To establish significant differences between categorical groups, non-parametric Kruskal Wallis significance test was used (Kruskal and Wallis, 1952). Correlations between microbiome characteristics and numeric variables were calculated by Spearman's correlation (Spearman, 1961). Significance of beta diversity was estimated with PERMANOVA and PERMDISP (Anderson and Walsh, 2013).

Longitudinal analysis for the samples collected from the vaccinated group was done using the q2-longitudinal pipeline(Bokulich et al., 2018a). Predictions to determine sample associated metadata were done with Random Forest algorithms q2-sample-classifier plugin (Pedregosa et al., 2011). The accuracy of the classifier comments on the distinguishing power of microbiome profiles across different groups.

Chapter 3

Results

3.1 Performance of taxonomic classifiers

Taxonomic classifiers compare 16S amplicon sequences against a reference database of annotated sequences to determine the most likely taxonomic origin of each sequence. The unweighted taxonomic classifier (Robeson et al., 2020) considers all the taxa in its reference database to be found with equal probabilities in the sample. However, this is not applicable to real-world samples where the probabilities depend on the ecosystem of origin. The weighted classifier (Kaehler et al., 2019) adds weights to the prevalence of a taxon based on the habitat considered. While the human nasal ecosystem was not one of the 14 environments considered while training the weighted classifier, the authors claim that it would still perform better than the unweighted classifier in an untrained environment. They reason that even inaccurate weights would be more beneficial than using an unweighted equal distribution of taxa prevalence.

We compared the two classifiers and found that the classifier trained on the unweighted SILVA database outperforms the weighted SILVA database. Unweighted SILVA categorises more Human DNA features into the unassigned category, whereas weighted classifies a lot of the Human DNA as Archaea and Bacteria. Overall, unweighted can pick out more Human DNA representative sequences implying that weighted has incorrectly classified atleast 100 sequences with bacterial taxonomy of ≥ 2 levels. The data from the comparisons can be found in the Appendix, Table A. Hence, we decided to proceed with unweighted taxonomic classification using the SILVA database.

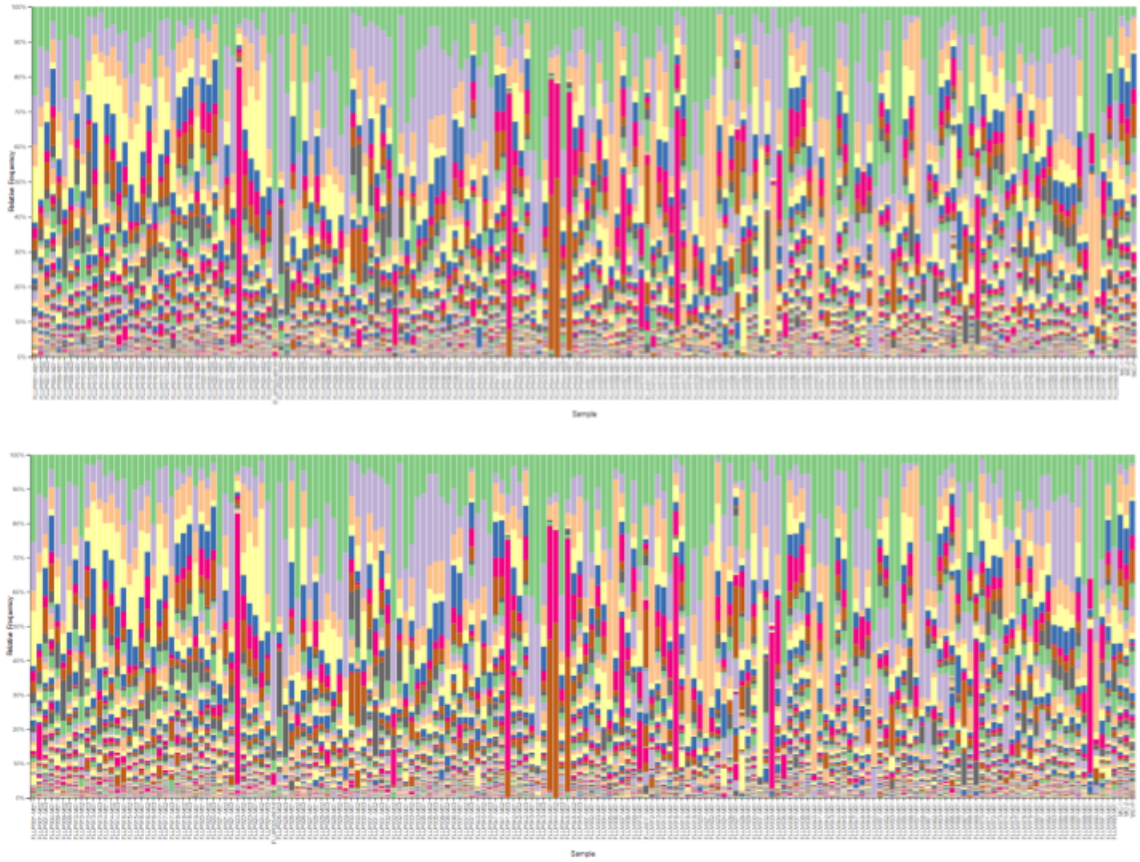


Figure 3.1: Taxa barplot representing taxonomic classification performed by weighted SILVA classification (upper panel) and unweighted SILVA classification (lower panel) at taxonomy level 4

3.2 Demographic Distribution

The demographic distribution of the study participants is highlighted in 3.1 and 3.2.

Number of participants	Non-immune	Immune	Total
Female	40	54	94
Male	14	16	30
Total	54	70	**124**

Table 3.1: Distribution of participants according to sex and baseline immunity

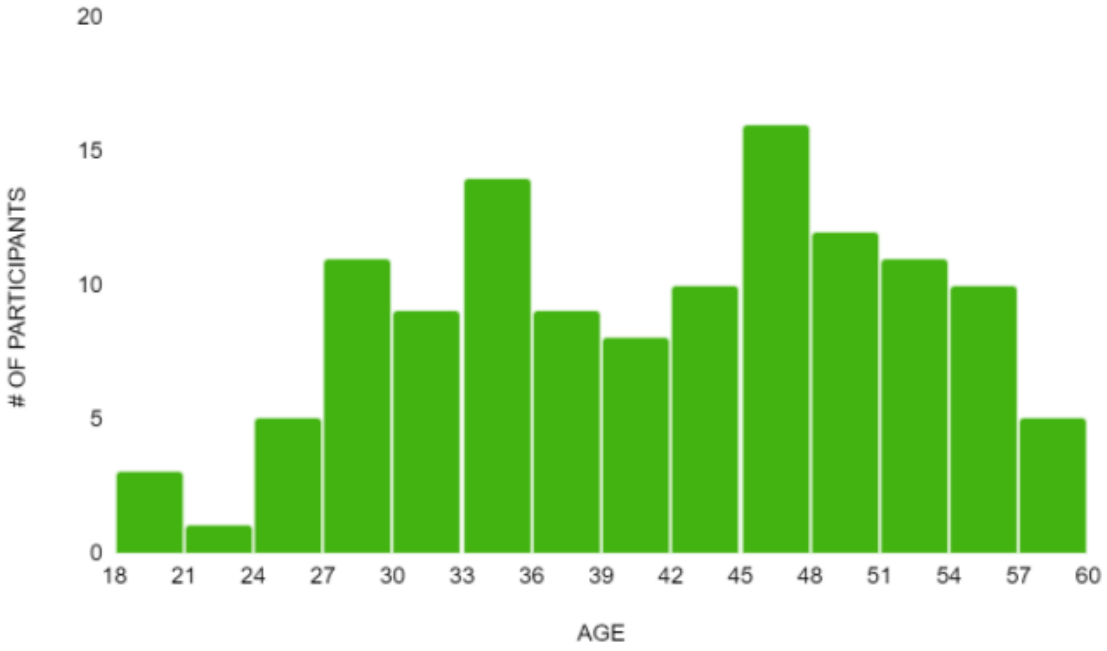


Figure 3.2: Histogram depicting the age distribution in the participant pool

3.3 Microbiome comparisons among relevant groups

All comparisons measure alpha diversity with Shannon entropy, Simpson's index and Faith's PD. Beta diversity distance matrices were Bray-Curtis dissimilarity, Unweighted UniFrac and Weighted UniFrac. Differentially abundant (DAb) taxa were identified through ANCOM. An effect/trend is considered significant if it is found to be significant among two or more diversity metrices for both alpha and beta diversity. If the metrices are not mentioned while presenting a significant result then it implies that all of them have established significant differences/ associations. All differences

pointed out are statistically significant unless mentioned otherwise. P-values mentioned in the text for alpha-diversity comparisons correspond to Shannon entropy while that of beta-diversity corresponds to Weighted UniFrac distance. This is done to maintain uniformity while reporting and due to the fact that the two chosen metrices are generally the conservative estimates among all the metrices considered when calculating significance.

3.3.1 Comparing the "immune" and "non-immune" microbiome

Alpha diversity comparisons of all baseline samples showed consistent significant differences between the influenza-specific immune and non-immune groups. The p-values obtained were 0.0025, 0.001 and 0.011 with Shannon entropy, Simpson's index and Faith's PD respectively. The overall alpha diversity appears to be lowest among the samples with the highest levels of HAI titre ($\geq 1:80$) as seen in Table 5 and Appendix, Figure A. The immune group also appears to show a greater range of intragroup variation as compared to the non-immune group.

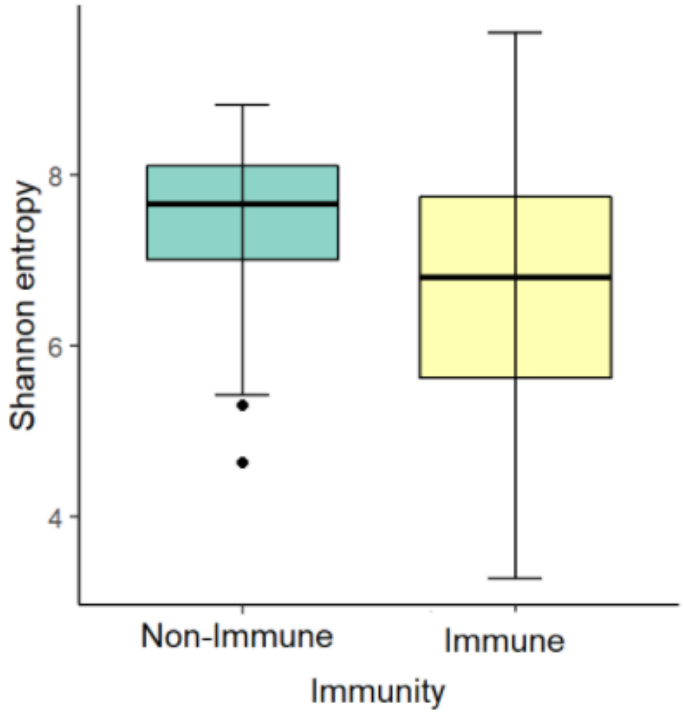


Figure 3.3: Boxplot illustrating variation in alpha diversity (Shannon index) among the influenza-associated immune and non-immune groups in baseline samples

		H	p-value	q-value
Group 1	Group 2			
EQUIVOCAL (n=19)	HIGH-POSITIVE (n=63)	5.200501	0.022580	0.067741
	MODERATE-POSITIVE (n=6)	0.004049	0.949266	0.969785
	NEGATIVE (n=34)	0.099415	0.752533	0.969785
HIGH-POSITIVE (n=63)	MODERATE-POSITIVE (n=6)	2.619501	0.105558	0.211115
	NEGATIVE (n=34)	8.211342	0.004163	0.024978
MODERATE-POSITIVE (n=6)	NEGATIVE (n=34)	0.001435	0.969785	0.969785

Table 3.2: Results of pairwise KW test done with influenza immunity levels using Shannon index with baseline samples. HAI titre $\leq 1:10$ is classified as high-positive, $= 1:20$ is equivocal, $= 1:40$ is moderate-positive and $\geq 1:80$ is high-positive.

Beta diversity comparisons establish significant differences among the immune vs non-immune participants (p-value = 0.002, 0.007 and 0.004; Bray-Curtis, Unweighted and Weighted UniFrac). When comparing across the HAI titres, significant differences seen with all the distance matrices (p-value = 0.002, Weighted UniFrac) were only established between the group with the lowest ($\leq 1:10$) and the highest ($\geq 1:80$) antibody titres. Refer to Table B in Appendix for pairwise comparison among the groups with varying levels of immunity.

ANCOM results point to Sphingomonadaceae (W = 437), a moderately abundant taxon found in higher proportions in non-immune samples. KI89A clade (Gammaproteobacteria) is a DAb rare taxa (W = 180) found in high proportions in samples with HAI titre = 1: 40. Percentage abundance of the features is found in Appendix, Table F and G.

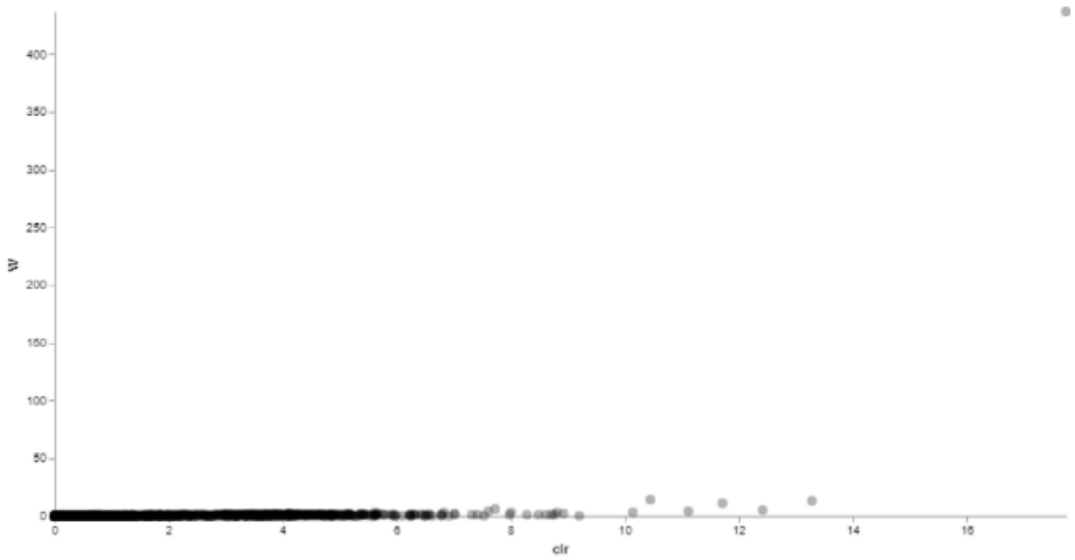
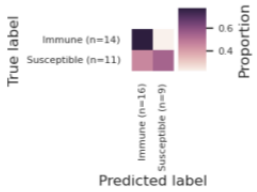


Figure 3.4: ANCOM Volcano plot for influenza immunity comparisons of baseline samples. The significant taxon is seen in the upper right corner of the plot.

To get a sense of the extent of the difference between the two groups, we tried to classify the samples as immune or non-immune based on the microbiome profiles of the baseline samples.

Model Accuracy



Download as PDF

	Immune	Non-Immune	Overall Accuracy
Immune	0.785714	0.214286	
Non-Immune	0.454545	0.545455	
Overall Accuracy			0.68
Baseline Accuracy			0.56
Accuracy Ratio			1.214286

Figure 3.5: Model accuracy for Influenza immunity classification

The model performs above baseline accuracy implying some degree of variation between the groups but the accuracy is not good. This implies that differences in the microbiome between the vaccinated and unvaccinated groups are not very stark and there is a lot of intra-group variation that confuses the classifier.

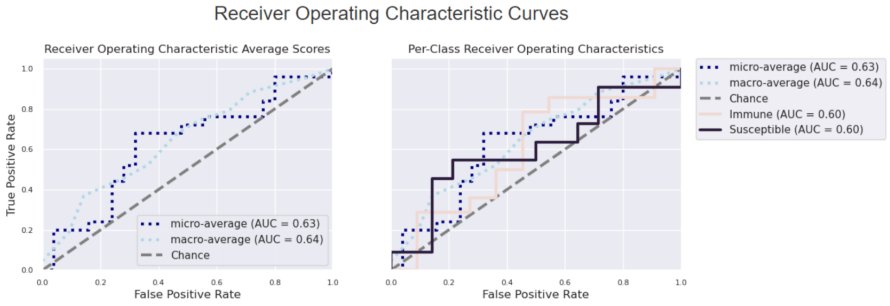


Figure 3.6: ROC Curves for the influenza immunity classifier

3.3.2 Microbiota response to influenza vaccination

The samples collected at different time-points from the vaccinated group were compared. Alpha diversity analysis showed no significant group or pairwise differences between the samples collected at baseline, Day 8 and Day 30 with Shannon, Faith or Simpson indices. Similarly, no difference was found with the beta diversity measures and no differentially abundant taxa were observed.

		H	p-value	q-value
Group 1	**Group 2**			
Baseline (n=33)	**Day 30 (n=30)**	0.318371	0.572588	0.738709
	Day 8 (n=29)	0.906852	0.340951	0.738709
Day 30 (n=30)	**Day 8 (n=29)**	0.111264	0.738709	0.738709

Table 3.3: Results of pairwise KW test done with Shannon index

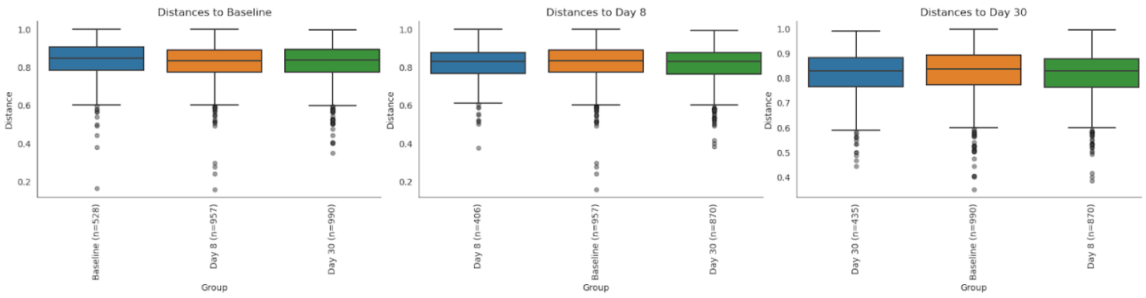


Figure 3.7: Group significance plots with Bray Curtis dissimilarity matrix and sample collection time-points

A better approach to look at the response to vaccination is the longitudinal analysis of the microbiome stability in each participant. This reduces the effect on intragroup variation and allows us to detect transient changes of low magnitude.

PCoA based volatility plots (using Bray-Curtis, Unweighted and Weighted UniFrac metrics) show no apparent changes in the microbiome with time. Statistical analysis found no significant longitudinal variation.

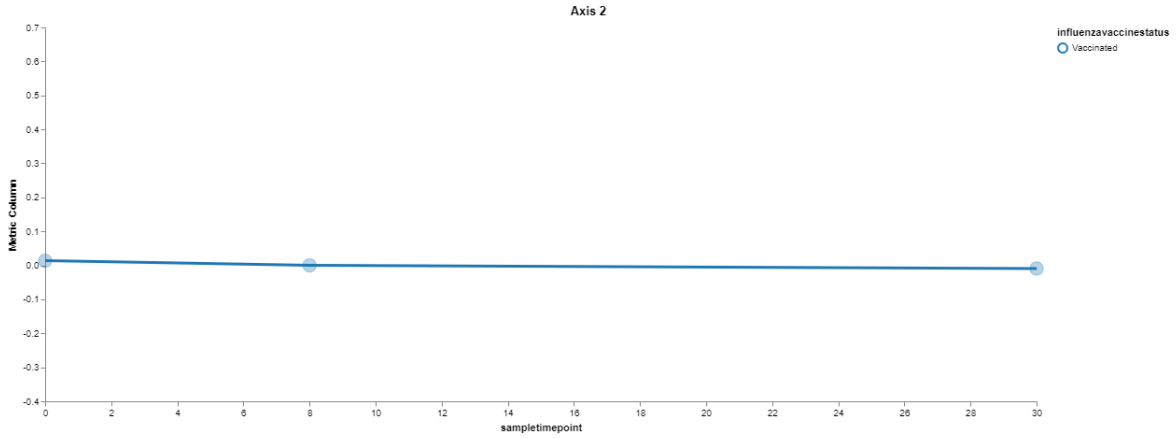


Figure 3.8: Volatility plot of Unweighted UniFrac PCoA metric (Axis 2) with time

Stratifying the data based on HAI titres and influenza immunity does not show significant temporal changes. However, we see in Figure 3.9 that the microbiome of the immune and non-immune groups are distinct and this difference is maintained over time.

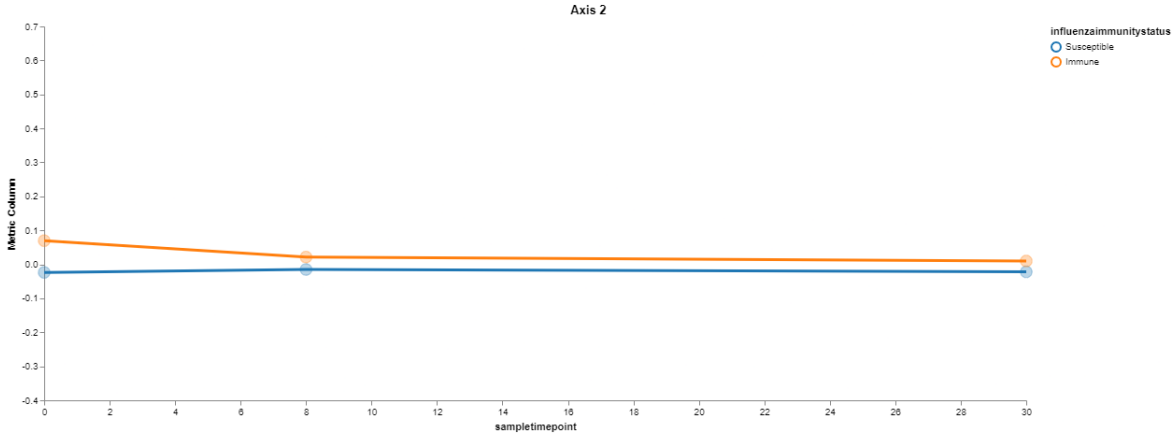


Figure 3.9: Longitudinal trend of Unweighted UniFrac PCoA metric (Axis 2) stratified by influenza immunity status. The orange line represents the Immune group while the blue line represents the Non-Immune group.

In Figure 3.10 we see that a high HAI titre separates from the rest of the group though this difference doesn't seem apparent in the Day 30 samples of the vaccinated group.

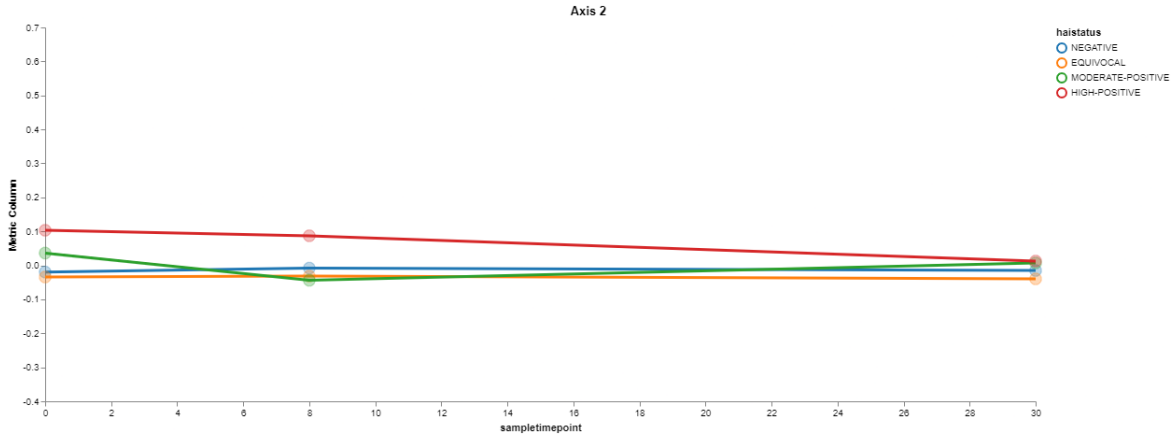


Figure 3.10: Longitudinal trend of Unweighted UniFrac PCoA metric (Axis 2) stratified by HAI status. The blue line in the figure represents HAI titre $\leq 1:10$, the orange line represents a titre = 1:20, the green line represents a titre = 1:40 and the red line indicates a titre $\geq 1:80$.

3.3.3 Variation in IgA antibody levels

The influenza-specific IgA levels were not found to vary significantly across samples collected at different time-points in the vaccinated group (KW test and Pairwise-Wilcox test) and this trend reflects in longitudinal comparisons as well. No significant difference in influenza IgA levels was found between the immune and non-immune groups.

Significant differences were found between moderately positive samples (HAI titre = 1:40) and highly-positive samples (HAI titre $\geq$ 1:80) with a p-value = 0.012 and between moderately positive samples (HAI titre = 1:40) and negative samples (HAI titre $\leq$ 1:20) with a p-value = 0.016. This result is shown in the Appendix, Table C. Total IgA levels were not different across different sample timepoints in the vaccinated group.

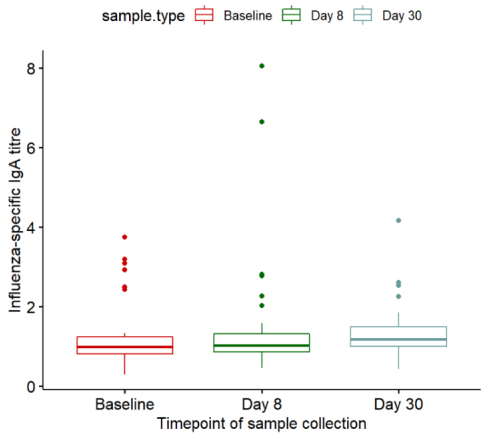


Figure 3.11: Comparison of influenza-specific IgA titre among vaccinated samples grouped according to time-point of collection

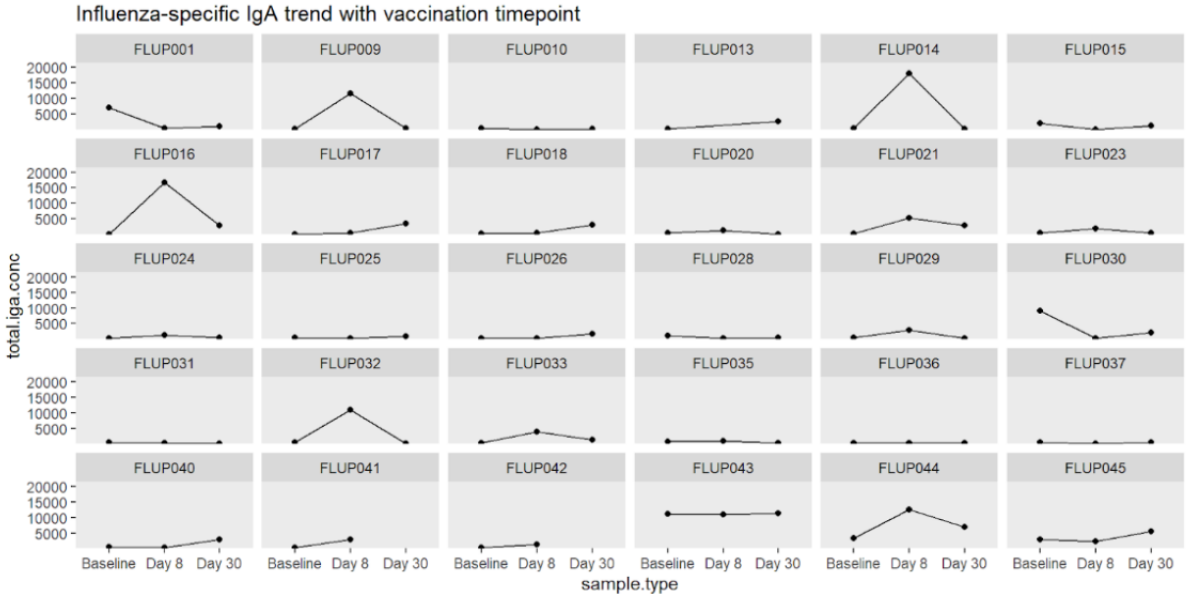


Figure 3.12: Participant-wise temporal influenza-IgA trend among the vaccinated group

3.3.4 Correlation between mucosal antibody levels and the nasal microbiome

No correlation between influenza-specific IgA levels and the nasal microbiome profile was established. A consistent negative correlation between total mucosal IgA levels and alpha diversity was seen, as illustrated in Figure 3.13. (Correlation coefficient = -0.1835 and p-value = 0.02, Shannon). Higher levels of IgA appear to be correlated with lower levels of alpha diversity. No correlation was found between beta-diversity distances and total IgA concentrations.

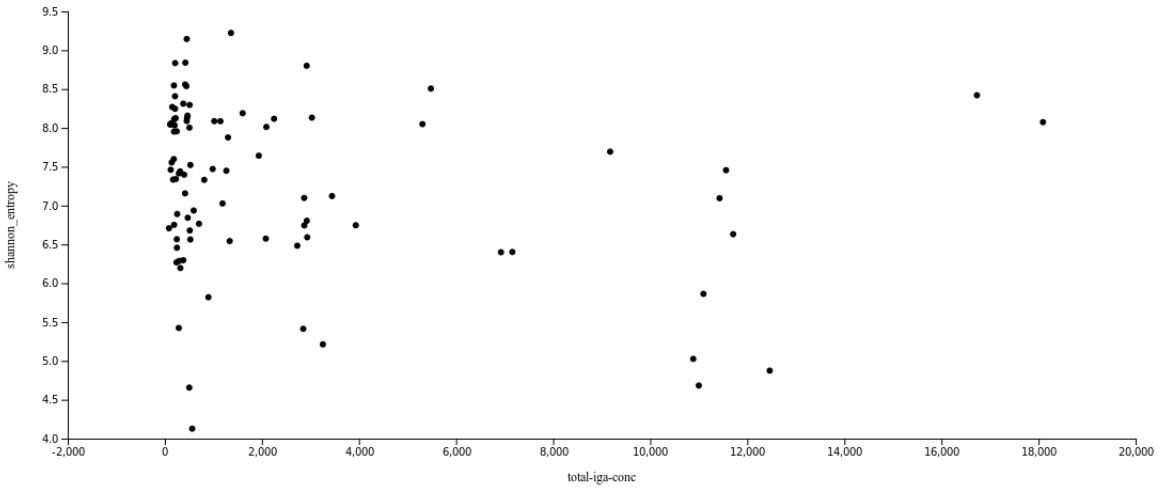


Figure 3.13: Shannon entropy correlation with total IgA concentration in baseline samples

3.3.5 Potential confounders analysis

A myriad of factors affect and bias the data obtained given the nature and scope of clinical studies. This section attempts to identify a few of these factors evident from the data distribution and gain some degree of semblance on how the confounder might interfere with the results of the study.

Gender

From Table 3.1, it is evident that the number of females recruited is thrice that of the males. While efforts were made to recruit more males, it proved to be challenging. The reason for the imbalance

in enrolment was that recruitment and sample collection was done between 9:00 AM to 5:00 PM on weekdays and a majority of the men in the study population were at work during these hours and thus unavailable. Given the time constraint of the project timeline, we ended up recruiting more females than males.

To examine the effect of this imbalance on the analysis, we compared the microbiome between the two groups. No significant differences were found when comparing male and female microbiomes in the context of alpha diversity (Shannon, Simpson and Faith's PD), beta diversity (Bray-Curtis and Unweighted UniFrac) or differential abundance. Weighted UniFrac distances showed significant differences (p-value = 0.017) between the two groups. Since none of the other metrics were able to demonstrate this significance, we did not stratify the data according to gender for analysis.

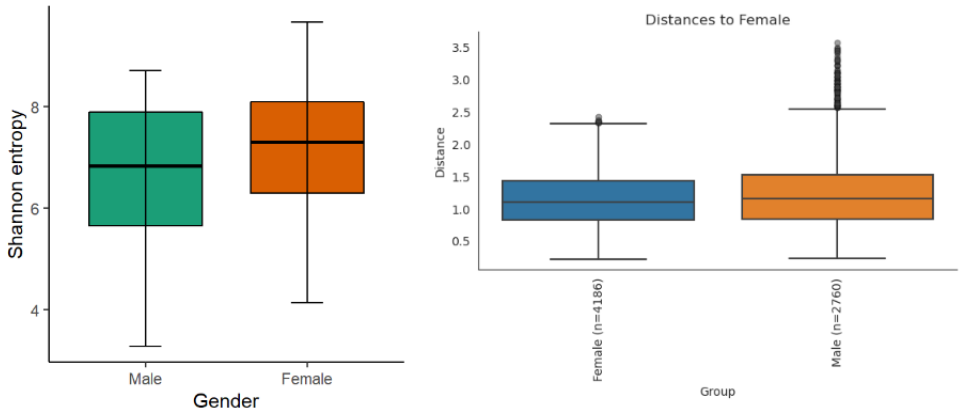


Figure 3.14: Left Panel illustrates Shannon entropy distribution across genders and the Right Panel illustrates the Weighted UniFrac distance between the males and females microbial communities.

Chronic patients on medication

The decision to recruit chronic patients under medication was done for two purposes. The first reason was that hypertension and diabetic patients make up a notable portion of the adult population between 18 to 65 years of age and including such patients would allow us to sample the target population more accurately. The second reason was that we hypothesised that a reduction in inflammation brought about by the medication may partially reduce the dysbiosis associated with these conditions that was aggravated by the prevalent systemic inflammation. Medications for chronic lifestyle diseases are known to have anti-inflammatory effects (Miguel-Carrasco et al., 2010; Cameron et al., 2016). An altered gut microbiome with decreased microbial richness and

diversity is seen in animal studies in both prehypertensive and hypertensive states though this has not been fully investigated in hypertensive patients. We are still in the need of objective evidence and an understanding of mechanisms to support these results (Richards et al., 2017). Consumption of Captopril, an angiotensin-converting enzyme inhibitor commonly used for the treatment of hypertension (HTN), showed a partial reduction in dysbiosis associated with HTN (Al Khodor et al., 2017). Metformin was found to partially revert the differences in the gut microbiome seen in Type-2 diabetic patients in an artificial stimulated gut environment and increase the growth of bacterial species that produce metabolites suspected of glucose-lowering effects. The authors suggest that the metabolites produced might be partly responsible for the better action of metformin when administered orally as compared to intravenous injection (Xourgia et al., 2019). None of the above-mentioned studies have been able to confirm and concretely characterise the effect of medication on disease-associated dysbiosis. Since we are unaware of the impact of including chronic patients in the study and do not have a well-supported hypothesis, we decided to look at the differences between the medicated and non-medicated groups.

Except for Shannon entropy ($p = 0.04$), comparisons with no other metrics yielded significant differences. Since this was not consistently seen with any of the other parameters, the medicated group was not separated out in the analysis.

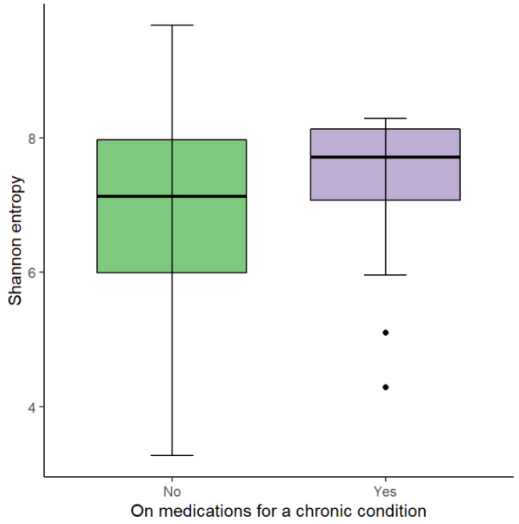


Figure 3.15: Shannon entropy distribution between the medicated and non-medicated group

Recent exposure to influenza virus

In the vaccinated group, a repeat blood sample was taken for HAI estimation prior to vaccination. This provided us with information on recent exposures to influenza within the gap between screening and sample collection. This data split the baseline samples into 4 groups - recently immune (turned immune on rescreening), known non-immune (remained non-immune on rescreening), suspected immune (immune but no rescreening, possibility of reduction in antibody titres) and suspected non-immune (non-immune but no rescreening, possibility of influenza exposure).

Alpha diversity analysis highlights differences between suspected immune and known non-immune (p-value = 0.005, Shannon); and suspected immune and suspected non-immune (p-value = 0.013, Shannon). This is consistent with what is seen in section 3.1 as a majority of the titre levels $\geq$ 1:80 samples are from the suspected immune group. From Figure 3.16, we see that the two immune groups appear to have a lower alpha diversity median than the non-immune groups. Beta diversity comparisons exhibit significant differences among the suspected immune group and known/suspected non-immune (p-value = 0.001/0.03, Weighted UniFrac), and suspected non-immune and known non-immune (p-value = 0.02, Weighted UniFrac). The latter may indicate that some of the participants in the suspected non-immune group have been exposed to influenza in the gap causing a shift in the group profile.

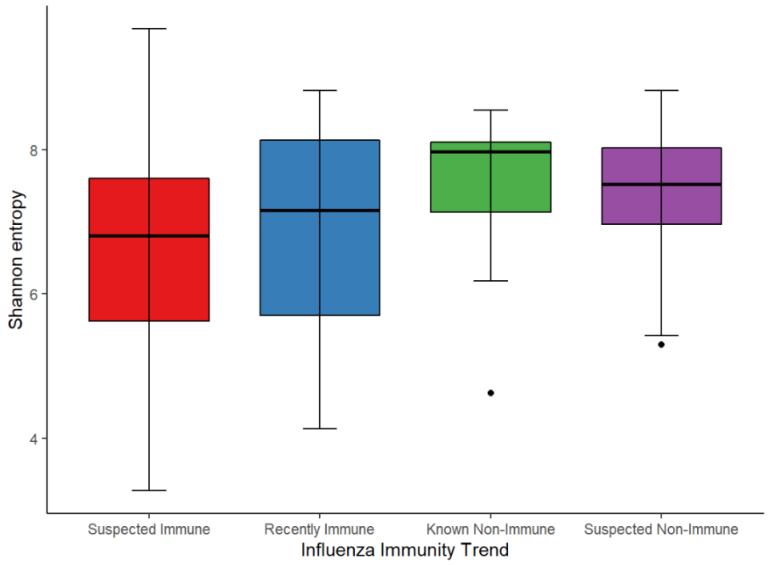


Figure 3.16: Boxplots illustrating alpha diversity variation in light of the influenza immunity trend

The "Vaccinated" and "Unvaccinated" microbiome

The study design underwent major changes due to a multitude of reasons. This introduced differences in the conditions and timing of sample collection between the vaccinated and unvaccinated groups. The vaccinated baseline samples were collected from July to December, while the unvaccinated immune samples were collected in January and the unvaccinated non-immune samples were collected in February. Hence, to estimate the impact of these variations, we compared their microbial profiles.

No difference was observed with alpha diversity measures among the two groups but a consistent significant difference was noted with the beta diversity results (p-value = 0.006, Weighted UniFrac).

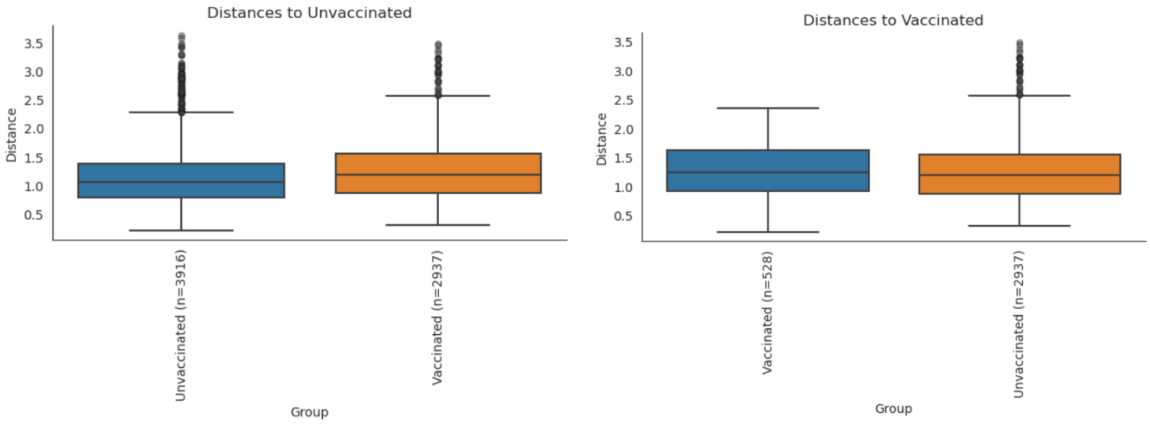


Figure 3.17: Group significance plots with weighted UniFrac distance matrix with the vaccinated and unvaccinated groups

Since the two groups have a different ratio of immune to non-immune participants, this may be a possibility for the difference. Hence, we stratified the data to indicate both immunity and vaccinated status.

Stratified data showed robust alpha diversity differences between the immune unvaccinated and the non-immune unvaccinated as well and between the immune unvaccinated and non-immune vaccinated (p-values reported in table D, Appendix). This trend was also seen with the beta diversity group significance results. An additional significance picked up by the beta diversity tests is the difference between the non-immune vaccinated and the immune vaccinated groups (p-values reported in table E, Appendix). ANCOM revealed a differential abundance of Sphingomonadales (Alphaproteobacteria, W = 81), a taxon of low abundance found in higher amounts in unvaccinated immune samples followed by vaccinated immune samples.

Percentile	0.0	25.0	50.0	75.0	100.0	0.0	25.0	50.0	75.0	100.0
Group	IU	IU	IU	IU	IU	IV	IV	IV	IV	IV
d__Bacteria;p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales	1.0	41.0	89.0	174.0	1823.0	1.0	45.25	87.5	172.5	568.0
	0.0	25.0	50.0	75.0	100.0	0.0	25.0	50.0	75.0	100.0
	SU	SU	SU	SU	SU	SV	SV	SV	SV	SV
	25.0	127.5	228.5	365.0	593.0	76.0	133.0	177.0	281.0	418.0

Table 3.4: Percentile abundances of Sphingomonadales by influenza vaccination and immunity grouping at level 4 in baseline samples

Since the pairwise comparisons reveal that the difference between the group appears to follow the influenza immunity values, we can conclude that there appears to be no difference between the vaccinated and unvaccinated group. The presence of a differentially abundant taxa found in both vaccinated and unvaccinated samples seem to support this argument.

Sample collection month

Microbiome differences based on when the sample was collected has been established with both alpha (p-value = 0.03, Shannon) and beta diversity (p-value = 0.011, Weighted UniFrac). Differences have been noted between multiple combinations of months and are consistent across all metrices. Both influenza-IgA and total IgA were found to significantly vary across the collection month. All of this highlights the strength of the confounder effect of this factor. ANCOM identified Legionellales (Gammaproteobacteria, W = 233), a rare taxon found in high proportions in the August samples and Acetobacteraceae (Alphaproteobacteria, W = 1483), a rare taxon mostly found in the December samples. These bacteria identified are both pathogenic species that were rare, possibly from infections among a few of the participants or sample contamination.

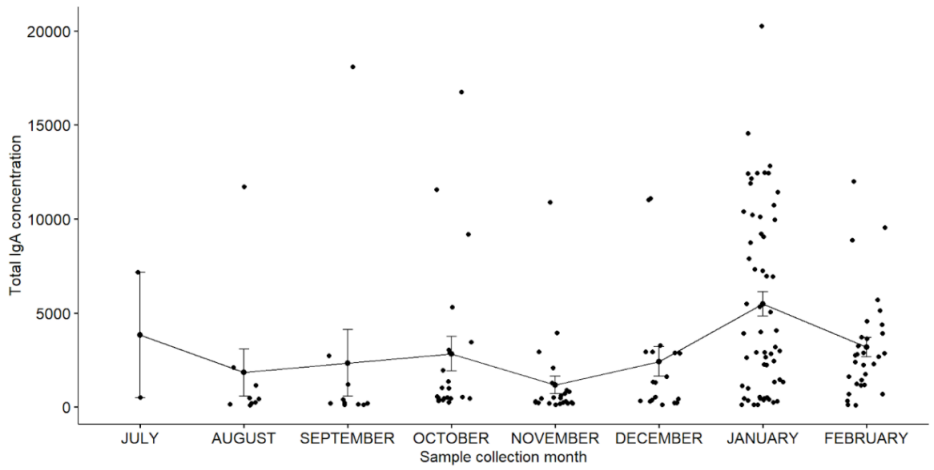


Figure 3.18: Comparison of total IgA concentration across sampling months

COVID vaccination

We hypothesised that an intramuscular vaccine against the SARS-CoV2 virus may follow similar principles of microbiome interactions as an intramuscular vaccine against influenza given that they both target respiratory illnesses and have the same mode of administration. Since there was an overlap in the COVID-19 vaccination campaign and our sample collection timeline, we decided to evaluate the impact of this vaccine on the nasal microbiome and its role as a potential confounder.

No difference was found in alpha diversity with any of the metrices between the COVID-19 vaccinated and unvaccinated groups. This observation was maintained when splitting the data on the basis of the number of doses taken. The same was seen with the beta-diversity measures and differential abundance features. However, ANCOM found Cytophagaceae (Bacteroidota, W = 545) present in samples of unvaccinated COVID-19 participants, though this is very likely present only in one or a few samples and doesn't really contribute to a substantial difference between the two groups.

From this, we concluded that we were unable to spot any variations in the nasal microbiome that may have been due to COVID-19 vaccination. To account for the differences in when the vaccine was taken, a correlation analysis between the time gap since the last COVID-19 vaccine dose in weeks and the nasal microbiome was performed but no significance was established in either alpha, beta diversity or differential abundance analysis. However, it must be noted that the retrospective nature of the investigation may have introduced noise and variations in the data and may have limited our ability to quantify the impact of the COVID-19 vaccine.

Chapter 4

Discussion

In this section, we will attempt to interpret the whirlwind of results mentioned above. All the results will be summarised and interpreted in the context of the null hypotheses that were framed to address the objectives mentioned under the Study Design.

Hypothesis 1: There is no difference between the nasal microbial profiles of individuals immune to influenza and individuals not immune.

Results summary - According to the analysis, alpha diversity (estimated with Shannon, Simpson and Faith PD) shows significant differences among the immune and the non-immune group. By increasing the resolution of data with HAI titres, we observe that samples with ¿1:80 HAI titres are significantly different (of lower diversity) from the rest of the lot. These samples also appears to show a greater range of intragroup variation as compared to the other samples. Differentially abundant taxa, Sphingomonadaceae and KI89A_clade were identified. Sphingomonadaceae was moderately abundant in all sample groups and found in higher proportions in non-immune samples. KI89A clade is a rare taxa found differentially abundant in samples from participants with HAI titre = 1:40. With respect to beta density, influenza immunity groups showed significant differences and again ¿1:80 titres seemed significantly different from the rest of the samples. However, using a random-forest classifier to classify the samples on the basis of immunity was not very accurate implying that the differences between the two groups in more subtle and a high degree of intra-group variation exists which confuses the classifier.

Interpretation - KI89A which belongs to Gammaproteobacteria is usually found in marine environments and the classification was confirmed by BLAST. This is likely contamination as it is

an extremely rare taxa only found slightly abundant in samples that are collected from participants with HAI titre = 1:40 which is the group with the least number of samples.

Several members of the family Sphingomonadaceae, usually isolated from soil and clinical samples, are aerobic organisms. They are known to be generally associated with antibiotic resistance when discovered in clinical samples though only a few cause human infections (Feller et. al, 2021; Vaz-Moreira et al., 2011). This taxa is found to be clearly abundant in non-immune samples and this trend is seen uniformly both in the non-immune participants of the vaccinated and the unvaccinated. Hence it is likely not an artefact of the period of sample collection or contamination, though we cannot comment on the implication of it being differentially abundant in the non-immune samples.

The median alpha diversity appears lowered in immune samples with high HAI titres though this category also appears to have the more intragroup variations than the rest of the samples. Since multiple tests and parameters appear to indicate that non-immune and immune samples differ in nasal microbiome profile, we can reject the null hypothesis. Given the nature of the study, we are unable to explain this result further or point out a specific mechanism of action. **The differences in microbiota may be due to either the immune response determining which bacteria are able to survive in and colonise the upper respiratory tract, or the microbial composition may be driving antigen processing in a manner that leads to a high immune response**. However, it must be kept in mind that this difference may have appeared incorrectly due to a multitude of reasons. A major artefact is likely introduced by the study design that had to undergo multiple changes to which sample collection was not done uniformly. Variation of alpha, beta and differential abundance with sample collection month indicates that this might be a potential confounder affecting the results. Given the waves of COVID, higher incidences of RTIs might be one of the causes of the variation. Both influenza-specific and total IgA levels are also seen to vary across sample collection months and given the total IgA correlation with alpha diversity, this might be driving the difference as well. Alternatively, the fact that we collected samples of different influenza-immunity levels with non-uniform sample sizes in each group across the different months could be the reason for the differences seen across sample collection months. The failure of the classifier may also be due to the low sample size and the fact that a majority of the non-immune samples were not classified correctly due to the lack of rescreening of the unvaccinated participants. This may hint at the non-immune participants not remaining so due to influenza exposure in the gap between screening and sample collection.

Hypothesis 2: Vaccination does not bring about changes in the nasal microbiota, in both the long and short term. Changes brought about by vaccination are not different among folks immune pre-vaccination versus those who are not immune pre-vaccination.

Results summary - Alpha and Beta Diversity analyses and differential abundance reveal no difference among the samples collected across various time-points. To reduce the influence of intra-group diversity, longitudinal analysis was also done, but no significance was established. Stratifying the samples based on the HAI status did not show any distinct trend that might have been lost upon taking the average.

Interpretation - Since we were unable to see any changes, it may be possible that an intramuscular inactivated vaccine does not cause any systemic changes that affects the nasal environment. The small sample size of only 10 samples per time-point may be one of the reasons we saw no effect. Or the timescale of change might be too minute and we might need better temporal resolution to identify these changes if they exist.

Hypothesis 3: Local influenza-specific IgA titre and total IgA concentration do not vary across groups.

Results summary - Local influenza-specific IgA levels vary across influenza immunity status, COVID vaccination status and sample collection month. Total IgA concentrations vary significantly across sample collection months and among vaccinated vs unvaccinated groups.

Interpretation - While total IgA also sees differences between influenza immunity trends with resampling, this is mostly a reflection of the difference between the vaccinated and unvaccinated groups. The vaccinated and unvaccinated sample collections were clearly separated across time and they may be reflecting the difference brought about by sample collection months. Influenza specific IgA also shows variation across groups, however, this may again be due to different sample collection time frames of the immune and susceptible groups. It seems likely that the variation seen in influenza-specific and total IgA levels is an artefact of influenza, COVID-19 and other respiratory infections seasons/waves. Another point to note is that the influenza-specific IgA assays were not very sensitive and only picked up a signal close to background noise in most cases, and thus comparisons may not be accurate.

Hypothesis 4: There is no association between the nasal microbiome composition and influenza-specific IgA titre and total IgA concentration.

Results summary - A correlation between total IgA concentrations and alpha diversity (Shannon, Simpson and Faith) was observed with a correlation coefficient of -0.2274 (Shannon). No

correlation was seen with any of the beta-diversity distance matrices.

Interpretation - The alpha diversity trend is indicative of total IgA levels affecting the nasal microbiome by affecting the local immune environment. The total IgA levels is likely one of the reasons for observing variations among samples collected at different times in the year. However, no strong correlation has been established and there might be another underlying factor (example: the spread of infection) that may be affecting both the microbiome and the total IgA levels.

Chapter 5

Conclusion

A "surprising" result obtained from the study was the association between pathogen-specific serum antibodies and the nasal microbiome! Even if we were to consider the antibody levels as a stand-in for exposure to influenza, the impact of influenza exposure more than 6 months ago still being associated with differences in the nasal microbiome composition is quite startling. One can speculate that the differences in the microbiota seen may arise either due to the variation in antibody levels (and corresponding differences in other arms of the immune system) that determine the microbial landscape, or it may be the microbiota that is shaping the immune response.

Since it is likely that we have a powerful confounder, such as the time frame of sample collection being linked to the circulating RTI seasons and waves, our first step forward should be to reproduce the study in a larger sample size while trying to minimise the potential confounders and incorporate more immune response data. Comparing HAI titre changes pre and post-infection and vaccination in a longitudinal study would be helpful. Use of intranasal live-attenuated influenza vaccines as a substitute for influenza exposure would allow for studying changes to exposure in a controlled environment. While that was the original plan of this study, the study design changed due to the unavailability of the intranasal vaccine amid the pandemic. Moving forward in this direction would help us understand whether the changes were indeed just associated with serum antibody levels (often seen with vaccination) or if the exposure triggers specific changes in the nasal microbiome that appears to persist for extended periods of time post-infection. Only after we have verified the effect in robust clinical studies, will we be able to begin to address the underlying microbiome-immune interplay, and extrapolate the results to clinical applications.

Bibliography

Al Khodor, S., Reichert, B. and Shatat, I. (2017). The microbiome and blood pressure: can microbes regulate our blood pressure? Frontiers in pediatrics *5*, 138.

Allaband, C., McDonald, D., Vázquez-Baeza, Y., Minich, J., Tripathi, A., Brenner, D., Loomba, R., Smarr, L., Sandborn, W., Schnabl, B. and Dorrestein, P. (2019). Microbiome 101: studying, analyzing, and interpreting gut microbiome data for clinicians. Clinical gastroenterology and hepatology *17*, 218–230.

Anderson, M. and Walsh, D. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? Ecological monographs *83*, 557–574.

Belongia, E., Simpson, M., King, J., Sundaram, M., Kelley, N., Osterholm, M. and McLean, H. (2016). Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. The Lancet Infectious Diseases *16*, 942–951.

Bokulich, N., Dillon, M., Zhang, Y., Rideout, J., Bolyen, E., Li, H., Albert, P. and Caporaso, J. (2018a). q2-longitudinal: longitudinal and paired-sample analyses of microbiome data. MSystems *3*, 00219–18.

Bokulich, N., Kaehler, B., Rideout, J., Dillon, M., Bolyen, E., Knight, R., Huttley, G. and Caporaso, J. (2018b). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome *6*, 1–17.

Bolyen, E., Rideout, J., Dillon, M., Bokulich, N., Abnet, C., Al-Ghalith, G., Alexander, H., Alm, E., Arumugam, M., Asnicar, F. and Bai, Y. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology *37*, 852–857.

Bosco, N. and Noti, M. (2021). The aging gut microbiome and its impact on host immunity. Genes & Immunity *22*, 289–303.

Callahan, B., McMurdie, P., Rosen, M., Han, A., Johnson, A. and Holmes, S. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods *13*, 581–583.

Cameron, A., Morrison, V., Levin, D., Mohan, M., Forteath, C., Beall, C., McNeilly, A., Balfour, D., Savinko, T., Wong, A. and Viollet, B. (2016). Anti-inflammatory effects of metformin irrespective of diabetes status. Circulation research *119*, 652–665.

Chaudhari, S. N., McCurry, M. D. and Devlin, A. S. (2021). Chains of evidence from correlations to causal molecules in microbiome-linked diseases. Nature Chemical Biology *17*, 1046–1056.

Dash, S., Clarke, G., Berk, M. and Jacka, F. (2015). The gut microbiome and diet in psychiatry: focus on depression. Current opinion in psychiatry *28*, 1–6.

Davis, N., Proctor, D., Holmes, S., Relman, D. and Callahan, B. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome *6*, 1–14.

de Vos, W. M. and de Vos, E. A. (2012). Role of the intestinal microbiome in health and disease: from correlation to causation. Nutrition reviews *70*, S45–S56.

Debelius, J., Song, S., Vazquez-Baeza, Y., Xu, Z., Gonzalez, A. and Knight, R. (2016). Tiny microbes, enormous impacts: what matters in gut microbiome studies? Genome biology *17*, 1–12.

Dimitri-Pinheiro, S., Soares, R. and Barata, P. (2020). The microbiome of the nose—friend or foe? Allergy & rhinology *11*, 2152656720911605.

Ding, T., Song, T., Zhou, B., Geber, A., Ma, Y., Zhang, L., Volk, M., Kapadia, S., Jenkins, S., Salvatore, M. and Ghedin, E. (2019). Microbial composition of the human nasopharynx varies according to influenza virus type and vaccination status. MBio *10*, 01296–19.

Evans, A. S. (1978). Causation and disease: a chronological journey: the Thomas Parran lecture. American journal of epidemiology *108*, 249–258.

Galloway-Peña, J. and Guindani, M. (2018). Novel approaches in microbiome analyses and data visualisation. Frontiers in Microbiology *2274*.

Galloway-Peña, J. and Hanson, B. (2020). Tools for analysis of the microbiome. Digestive diseases and sciences *65*, 674–685.

Gilbert, P., Fong, Y., Juraska, M., Carpp, L., Monto, A., Martin, E. and Petrie, J. (2019). HAI and NAI titer correlates of inactivated and live attenuated influenza vaccine efficacy. BMC infectious diseases *19*, 1–12.

Glöckner, F., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R. and Ludwig, W. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. Journal of biotechnology *261*, 169–176.

Goodrich, J., Di Rienzi, S., Poole, A., Koren, O., Walters, W., Caporaso, J., Knight, R. and Ley, R. (2014a). Conducting a microbiome study. Cell *158*, 250–262.

Goodrich, J., Waters, J., Poole, A., Sutter, J., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J. and Spector, T. (2014b). Human genetics shape the gut microbiome. Cell *159*, 789–799.

Hardy, B., Dickey, S., Plaut, R., Riggins, D., Stibitz, S., Otto, M. and Merrell, D. (2019). Corynebacterium pseudodiphtheriticum exploits Staphylococcus aureus virulence components in a novel polymicrobial defense strategy. MBio *10*, 02491–18.

Hirve, S. and WHO (2015). Seasonal influenza vaccine use in low and middle income countries in the tropics and subtropics: a systematic review. World Health Organization.

Hobson, D., Curry, R., Beare, A. and Ward-Gardner, A. (1972). The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses. Epidemiology & Infection *70*, 767–777.

Inkpen, S. A., Douglas, G. M., Brunet, T., Leuschen, K., Doolittle, W. F. and Langille, M. G. (2017). The coupling of taxonomy and function in microbiomes. Biology & Philosophy *32*, 1225–1243.

Iuliano, A., Roguski, K., Chang, H., Muscatello, D., Palekar, R., Tempia, S., Cohen, C., Gran, J., Schanzer, D., Cowling, B. and Wu, P. (2018). Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. The Lancet *391*, 1285–1300.

Janssens, Y., Nielandt, J., Bronselaer, A., Debunne, N., Verbeke, F., Wynendaele, E., Van Immerseel, F., Vandewynckel, Y., De Tré, G. and De Spiegeleer, B. (2018). Disbiome database: linking the microbiome to disease. BMC microbiology *18*, 1–6.

Kaehler, B., Bokulich, N., McDonald, D., Knight, R., Caporaso, J. and Huttley, G. (2019). Species abundance information improves sequence taxonomy classification accuracy. Nature communications *10*, 1–10.

Katz, J., Hancock, K. and Xu, X. (2011). Serologic assays for influenza surveillance, diagnosis and vaccine evaluation. Expert review of anti-infective therapy *9*, 669–683.

Kostic, A., Gevers, D.and Siljander, H., Vatanen, T., Hyötyläinen, T., Hämäläinen, A., Peet, A., Tillmann, V.and Pöhö, P., Mattila, I. and Lähdesmäki, H. (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell host & microbe *17*, 260–273.

Kotomina, T., Isakova-Sivak, I., Kim, K., Park, B., Jung, Y., Lee, Y., Mezhenskaya, D., Matyushenko, V., Kang, S. and Rudenko, L. (2020). Generation and characterisation of universal live-attenuated influenza vaccine candiyears containing multiple M2e epitopes. Vaccines *8*, 648.

Kruskal, W. and Wallis, W. (1952). Use of ranks in one-criterion variance analysis. Journal of the American statistical Association *47*, 583–621.

Lehtinen, M., Hibberd, A., Männikkö, S., Yeung, N., Kauko, T., Forssten, S., Lehtoranta, L., Lahtinen, S., Stahl, B., Lyra, A. and Turner, R. (2018). Nasal microbiota clusters associate with inflammatory response, viral load, and symptom severity in experimental rhinovirus challenge. Scientific reports *8*, 1–12.

Liwinski, T., Leshem, A. and Elinav, E. (2021). Breakthroughs and bottlenecks in microbiome research. Trends in molecular medicine *27*, 298–301.

Luna, P., Hasegawa, K., Ajami, N., Espinola, J., Henke, D., Petrosino, J., Piedra, P., Sullivan, A., Camargo, C., Shaw, C. and Mansbach, J. (2018). The association between anterior nares and nasopharyngeal microbiota in infants hospitalised for bronchiolitis. Microbiome *6*, 1–14.

Luna, R. and Foster, J. (2015). Gut brain axis: diet microbiota interactions and implications for modulation of anxiety and depression. Current opinion in biotechnology *32*, 35–41.

Memoli, M. J., Shaw, P. A., Han, A., Czajkowski, L., Reed, S., Athota, R., Bristol, T., Fargis, S., Risos, K., Powers, J. H. et al. (2016). Evaluation of antihemagglutinin and antineuraminidase antibodies as correlates of protection in an influenza A/H1N1 virus healthy human challenge model. MBio *7*, e00417–16.

Methé, B., Nelson, K., Pop, M., Creasy, H., Giglio, M., Huttenhower, C., Gevers, D., Petrosino, J., Abubucker, S., Badger, J. and Chinwalla, A. (2012). A framework for human microbiome research. nature *486*, 215.

Miguel-Carrasco, J., Zambrano, S., Blanca, A., Mate, A. and Vázquez, C. (2010). Captopril reduces cardiac inflammatory markers in spontaneously hypertensive rats by inactivation of NF-kB. Journal of inflammation *7*, 1–9.

Moghadami, M. (2017). A narrative review of influenza: a seasonal and pandemic disease. Iranian journal of medical sciences *42*, 2.

Monteiro, M., Carvalho, M., Bastos, M. and Guedes de Pinho, P. (2013). Metabolomics analysis for biomarker discovery: advances and challenges. Current medicinal chemistry *20*, 257–271.

Mutz, K.-O., Heilkenbrinker, A., Lönne, M., Walter, J.-G. and Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. Current opinion in biotechnology *24*, 22–30.

Neville, B. A., Forster, S. C. and Lawley, T. D. (2018). Commensal Koch's postulates: establishing causation in human microbiota research. Current opinion in microbiology *42*, 47–52.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research *12*, 2825–2830.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B., Ludwig, W., Peplies, J. and Glöckner, F. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic acids research *35*, 7188–7196.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T. and Mende, D. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. nature *464*, 59–65.

Richards, E., Pepine, C., Raizada, M. and Kim, S. (2017). The gut, its microbiome, and hypertension. Current hypertension reports *19*, 1–11.

Rivas, M., Burton, O., Wise, P., Zhang, Y., Hobson, S., Lloret, M., Chehoud, C., Kuczynski, J., DeSantis, T., Warrington, J. and Hyde, E. (2013). A microbiota signature associated with experimental food allergy promotes allergic sensitisation and anaphylaxis. Journal of Allergy and Clinical Immunology *131*, 201–212.

Rizzardini, G., Eskesen, D., Calder, P., Capetti, A., Jespersen, L. and Clerici, M. (2012). Evaluation of the immune benefits of two probiotic strains Bifidobacterium animalis ssp. lactis, BB-12® and Lactobacillus paracasei ssp. paracasei, L. casei 431® in an influenza vaccination model: a randomised, double-blind, placebo-controlled study. British journal of nutrition *107*, 876–884.

Robeson, M. S., O'Rourke, D. R., Kaehler, B. D., Ziemski, M., Dillon, M. R., Foster, J. T. and Bokulich, N. A. (2020). RESCRIPt: Reproducible sequence taxonomy reference database management for the masses. Biorxiv .

Salk, H., Simon, W., Lambert, N., Kennedy, R., Grill, D., Kabat, B. and Poland, G. (2016). Taxa of the nasal microbiome are associated with influenza-specific IgA response to live attenuated influenza vaccine. PLoS One *11*, 0162803.

Spearman, C. (1961). The proof and measurement of association between two things. International journal of epidemiology *39*, 1137–1150.

Steenhuijsen Piters, W., Heinonen, S., Hasrat, R., Bunsow, E., Smith, B., Suarez-Arrabal, M., Chaussabel, D., Cohen, D., Sanders, E., Ramilo, O. and Bogaert, D. (2016). Nasopharyngeal microbiota, host transcriptome, and disease severity in children with respiratory syncytial virus infection. American journal of respiratory and critical care medicine *194*, 1104–1115.

Steenhuijsen Piters, W., Jochems, S., Mitsi, E., Rylance, J., Pojar, S., Nikolaou, E., German, E., Holloway, M., Carniel, B., Chu, M. and Arp, K. (2019). Interaction between the nasal microbiota and S. pneumoniae in the context of live-attenuated influenza vaccine. Nature communications *10*, 1–9.

Suau, A., Bonnet, R., Sutren, M., Godon, J., Gibson, G., Collins, M. and Doré, J. (1999). Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Applied and environmental microbiology *65*, 4799–4807.

Tarabichi, Y., Li, K., Hu, S., Nguyen, C., Wang, X., Elashoff, D., Saira, K., Frank, B., Bihan, M., Ghedin, E. and Methé, B. (2015). The administration of intranasal live attenuated influenza vaccine induces changes in the nasal microbiota and nasal epithelium gene expression profiles. Microbiome *3*, 1–16.

Thursby, E. and Juge, N. (2017). Introduction to the human gut microbiota. Biochemical Journal *474*, 1823–1836.

Tsang, T., Lee, K., Foxman, B., Balmaseda, A., Gresh, L., Sanchez, N., Ojeda, S., Lopez, R., Yang, Y., Kuan, G. and Gordon, A. (2020). Association between the respiratory microbiome and susceptibility to influenza virus infection. Clinical Infectious Diseases *71*, 1195–1203.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R. and Gordon, J. I. (2007). The human microbiome project. Nature *449*, 804–810.

Wang, Z., Klipfell, E., Bennett, B., Koeth, R., Levison, B., DuGar, B., Feldstein, A., Britt, E., Fu, X., Chung, Y. and Wu, Y. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature *472*, 57–63.

West, N., Horn, P., Pyne, D., Gebski, V., Lahtinen, S., Fricker, P. and Cripps, A. (2014). Probiotic supplementation for respiratory and gastrointestinal illness symptoms in healthy physically active individuals. Clinical Nutrition *33*, 581–587.

Xourgia, E., Papazafiropoulou, A., Papanas, N. and Melidonis, A. (2019). Anti-diabetic treatment leads to changes in gut microbiome. Front Biosci (Landmark Ed) *24*, 688–699.

Xu, Z., Zan, H., Pone, E., Mai, T. and Casali, P. (2012). Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. Nature Reviews Immunology *12*, 517–531.

Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y. and Lan, Z. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalised after treatment. Nature medicine *21*, 895–905.

Appendix

Total number of features = **39013**		
Filtering results	**Weighted SILVA**	**Unweighted SILVA**
Number of features classified as d__Bacteria (single-level classification)	773 (**82** Human DNA source matches upon BLAST)	364 (**4** Human DNA source matches upon BLAST)
Number of features classified as Unassigned	158 (**148** Human DNA source matches upon BLAST)	411 (**378** Human DNA source matches upon BLAST)
Number of features classified as d__Archaea (all levels of classification)	167 (**60** Human DNA source matches upon BLAST)	119 (**11** Human DNA source matches upon BLAST)
Total number of features classified as Human DNA	**290 (0.74%)**	**392 (1.00%)**
Number of features classified as Human DNA that wasn't filtered by the other classifier	10	112

Table A: Comparing the performance of the weighted taxonomic classifier vs the unweighted taxonomic classifier, both of which were trained on the SILVA database.

Table 1: Comparing the performance of the weighted taxonomic classifier vs the unweighted taxonomic classifier, both of which were trained on the SILVA database.

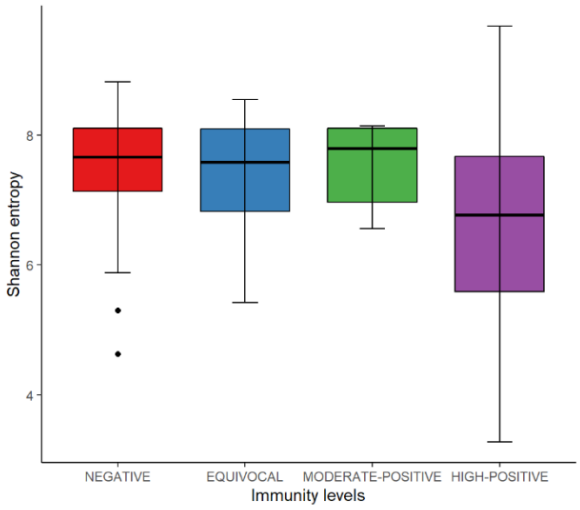


Figure 1: Shannon entropy distribution among groups differentially Immune to influenza in baseline samples

		Sample size	Permutations	pseudo-F	p-value	q-value
Group 1	Group 2					
EQUIVOCAL	HIGH-POSITIVE	82	999	1.821524	0.111	0.333
	MODERATE-POSITIVE	25	999	0.676965	0.571	0.700
	NEGATIVE	53	999	0.557030	0.700	0.700
HIGH-POSITIVE	MODERATE-POSITIVE	69	999	1.039467	0.351	0.700
	NEGATIVE	97	999	4.937408	0.001	0.006
MODERATE-POSITIVE	NEGATIVE	40	999	0.704590	0.595	0.700

Table 2: Pairwise-comparisons (PERMANOVA) of the different influenza-immunity groups with Unweighted UniFrac distances in baseline samples

	Suspected Immune	Recently Immune	Known Non-immune
Recently Immune	0.0112	-	-
Known Non-Immune	0.0022	0.8393	-
Suspected Non-immune	0.2487	0.0184	0.0022

Table 3: Pairwise comparisons using Wilcoxon rank-sum test of total IgA levels based on our knowledge of recent exposure to the influenza virus

		H	p-value	q-value
Group 1	**Group 2**			
IU (n=57)	**IV (n=12)**	0.485213	0.486070	0.493243
	SU (n=32)	6.147661	0.013159	0.039476
	SV (n=21)	7.586085	0.005882	0.035292
IV (n=12)	**SU (n=32)**	0.469444	0.493243	0.493243
	SV (n=21)	0.806723	0.369091	0.493243
SU (n=32)	**SV (n=21)**	0.894180	0.344347	0.493243

Table 4: Pairwise-comparisons (KW test) of the immune and non-immune category stratified based on influenza vaccine status with Shannon's index in baseline samples

		Sample size	Permutations	pseudo-F	p-value	q-value
Group 1	Group 2					
IU	IV	69	999	0.566799	0.697	0.8364
	SU	89	999	2.542953	0.031	0.0620
	SV	78	999	6.667028	0.001	0.0060
IV	SU	44	999	0.384840	0.844	0.8440
	SV	33	999	1.691633	0.169	0.2535
SU	SV	53	999	3.109822	0.020	0.0600

Table 5: Pairwise-comparisons (PERMANOVA) of the immune and non-immune category stratified based on influenza vaccine status with Unweighted UniFrac distance in baseline samples

Percentile	0	25	50	75	100
IMMUNE	1	1	1	4	97
NON-IMMUNE	1	1	1	17	98

Table 6: Percentile abundances of Sphingomonadaceae by influenza vaccination and immunity grouping at level 6 in baseline samples

Percentile	0	25	50	75	100
EQUIVOCAL	1	1	1	1	22
HIGH-POSITIVE	1	1	1	1	32
MODERATE-POSITIVE	1	4.75	18	28.25	45
NEGATIVE	1	1	1	1	32

Table 7: Percentile abundances of KI89A_clade by influenza vaccination and immunity grouping at level 6 in baseline samples