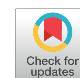
Science of the Total Environment 758 (2021) 143623

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

The environment shapes swine lung bacterial communities



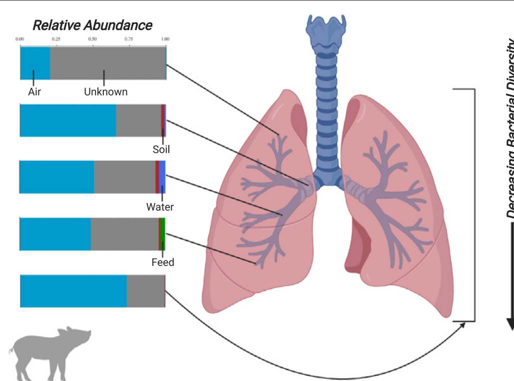
Alexander W. McCumber [a], Yeon Ji Kim [a], Omoanghe Samuel Isikhuemhen [b], Robert M. Tighe [c], Claudia K. Gunsch [a,*]

[a] Civil and Environmental Engineering Department, Duke University, Durham, NC, USA
[b] Department of Natural Resources and Environmental Design, North Carolina Agricultural & Technical State University, Greensboro, NC, USA
[c] Department of Medicine, Duke University, Durham, NC, USA

HIGHLIGHTS

- Most bacteria found in the bronchioles can be sourced from local ambient air.
- There are differences between lung alveolar and bronchiolar microbiomes.
- The environment influences bacteria found at the bronchiolar level but not alveolar.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 6 August 2020
Received in revised form 3 November 2020
Accepted 8 November 2020
Available online 16 November 2020

Editor: Yucheng Feng

Keywords:
Lung
Microbiome
Bacteria
Environment
Pig

ABSTRACT

Previous studies of the lung microbiome have focused on characterizing the community and attempts to understand the role of community membership concerning disease or exposures such as cigarette smoke. However, we still lack an understanding of two critical aspects of the lung microbiome: the origin of the community members and their fate. As we continue to better understand how the lung microbiome influences human health, it is essential to determine how the environment shapes the lung microbiome membership. Using a pig model, we explored the relationship that the surrounding environment has on the resident lung bacteria by collecting environmental samples (soil, air, water, feed) to compare with lung samples (swab, lavage, and tissue). Results suggest that airborne bacteria make up the highest portion of the lung microbiome. Furthermore, bacteria from samples taken from the bronchioles can be correctly identified by which farm they originated, whereas those from alveolar samples are indistinguishable. The findings suggest that while the environment may shape the microbiome of the bronchioles, a distinct community exists within the alveoli. Our findings expand upon the current understanding of the lung microbiome and provide a model of how microbial communities within the lung relate to their surrounding environment.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Lungs were long thought to be sterile, absent of a bacterial community (Dickson and Huffnagle, 2015). This belief persisted contrary to evidence that pathogens such as *Pseudomonas* could not only survive in the lung but could actively replicate and thrive. However, at the time,

Abbreviations: FOF, family owned farm; NCAT, North Carolina Agricultural and Technical State University; COPD, chronic obstructive pulmonary disorder; ASV, amplicon sequence variant; LEfSe, linear discriminant analysis effect size; FPD, Faith's phylogenetic distance; PCoA, principle coordinate analysis; SVM, support vector machine.

* Corresponding author at: 127 Hudson Hall, Durham, NC 27708, USA.
E-mail address: ckgunsch@duke.edu (C.K. Gunsch).

https://doi.org/10.1016/j.scitotenv.2020.143623
0048-9697/© 2020 Elsevier B.V. All rights reserved.

this finding was never extrapolated to include commensal bacteria. While it is believed that the majority of this community of lung bacteria are mostly transient and not actively replicating (Dickson et al., 2015) the existence of a bacterial community residing within the lower respiratory system has been established and the important role these complex bacterial communities in lung health have been the subject of extensive research (Beck et al., 2012; Dickson et al., 2018; Carney et al., 2020).Recently, studies have been able to find specific community members that are associated with increased inflammation or disease progression in chronic obstructive pulmonary disorder (COPD) (Wang et al., 2016; Mayhew et al., 2018), asthma (Huang and Boushey, 2015; Durack et al., 2017) and pulmonary fibrosis (Tong et al., 2019; Han et al., 2014). Still, this information has yet to be translated into clinical practice as there are still many unknowns (Caverly et al., 2019).

The increased understanding of the importance of the lung microbiome community structure and its implications for human health has coincided with humans' transition to a largely urban dwelling, indoor species. Urban indoor spaces have decreased microbial diversity compared to rural spaces (Parajuli et al., 2018), potentially decreasing the diversity of microbes humans interact with. Interactions with higher microbial diversity have been associated with beneficial health outcomes, particularly for respiratory diseases (Ruokolainen et al., 2017; Morgan et al., 2018). Most studies linking respiratory disorders with changes in the microbiome have relied on the inspection of gut or skin bacterial diversity for their analyses. This is likely due to the ease of sampling. However, the respiratory track is an important exposure site. There are approximately 10^5 bacterial like particles per cubic meter of air in indoor spaces (Prussin 2nd et al., 2015) and the average respiration rate for an adult at rest is approximately 12 to 18 breaths per minute (Brrett et al., 2012). Therefore, the lung represents a significant point of contact for human-microbe interactions that has been understudied in comparison to the gut. This topic has also been of recent interest as the National Academies of Science, Engineering, and Medicine brought together a committee dedicated to reviewing and proposing a future research agenda concerning *Microbiomes of the Built Environment* (National Academies of Science, E., and Medicine, 2017). One of these major objectives is to “Characterize interrelationships among microbial communities and built environment systems of air, water, surfaces, and occupants.” Accordingly, this study seeks to deepen the understanding of how the surrounding microbial environment affects and relates to the microbial community in the lung.

Several animal models have been used in the study of the lung microbiome including mice (Dickson et al., 2018), horses (Fillion-Bertrand et al., 2019), and pigs (Huang et al., 2019). While animal models have provided insights into the functions of resident pulmonary microbiota in the lung, the pig lung model provides several advantages in comparison to other animal models. Pig lungs more closely replicate the human lung immune system and physical structure. They also share susceptibility to some of the same pathogens (Meurens et al., 2012; Mair et al., 2014; Rogers et al., 2008). Another advantage is that pigs are commonly reared in a single, contained environment. This facilitates study of lung-environment interactions that may be difficult to capture in humans due to them traveling through multiple distinct environments throughout the day. As a result, the pig lung model may provide insights into how the environment influences the lung bacterial community.

While there have been studies published on the composition of the pig lung bacterial microbiome, these studies have been principally focused on how specific microbiome components alter disease (Huang et al., 2019; Siqueira et al., 2017; Megahed et al., 2019). Fewer studies focus on the impact of housing environment on baseline microbiome community composition which have shown that the type of feed or flooring can play significant roles in impacting community structure (Fillion-Bertrand et al., 2019; Megahed et al., 2019). It has also been shown that mouse lung communities cluster by vendor and cage with the with the similarity increasing as cohousing time increases; suggesting that the local environment has a strong influence on community composition (Dickson et al., 2018). However, to our knowledge, no studies have performed detailed environmental sampling to determine how distinct environmental conditions determine community microbial structure in the lungs. The question of how microbial communities are impacted by distinct environments is critical to define. Given increased urbanization and indoor living, defining these effects has important implications for microbiome-mediated human diseases. In the present study, we performed 16S rRNA gene amplicon sequencing on samples obtained from proximal and distal locations within each pig lung and the surrounding environmental media from two different farms. We then used this data to determine compositional differences in the bacterial community between these distinct locations and the extent to which each location is influenced by the local environment.

2. Methods

2.1. Site selection

Samples were obtained from two distinct farms. The first was a farm owned and operated by North Carolina A&T University (NCAT) located in Greensboro, NC. At NCAT, the swine housing was covered, and the animals were separated by sex. Each pen consisted of a concrete space where feed and water were available along with a larger bedding area with soil and straw. The water was piped in from the local municipal water treatment plant, and the diet was a commercial feed. The second farm in this study was a family-owned farm (FOF) located near Chapel Hill, NC. The swine housing was similar to NCAT, where it consisted of a covered concrete area with straw bedding but unlike NCAT included an uncovered area for access to feed and water. Water was sourced through the local municipal water treatment plant and the diet consisted of commercial feed but also incorporated vegetable scraps and milk from the farm.

2.2. Sample collection

All lung samples were collected at the slaughterhouse during processing, immediately placed on ice, and brought back to the lab for DNA extraction. From each pair of pig lungs ($n = 6$ FOF, $n = 25$ NCAT), three different sample types, each targeting different lung regions: lavage ($n = 6$ FOF, $n = 24$ NCAT), swab ($n = 9$ FOF, $n = 50$ NCAT), and tissue ($n = 12$ FOF, $n = 32$ NCAT) (Fig. 1). Lavage samples were obtained using a vacuum-driven serological pipet. A total of 50 mL of 0.5× normal saline solution was injected through the bronchiole and then withdrawn. In addition, proximal, swab samples of the trachea (proximal A) and bronchioles (proximal B) were collected. Distal tissue samples were excised using a sterile scalpel from two different areas of the lung, the upper (distal A) and lower (distal B) lobes. After the samples were collected, they were frozen at -20 °C until their DNA was extracted.

2.3. Preparation for molecular analysis

The DNA from the lavage samples was extracted using Qiagen PowerWater kit (QIAGEN, Hilden, Germany). The swabs were processed using the Qiagen Bacteremia kit with the protocol alterations performed in a previous study (Volkoff et al., 2020). The distal samples were excised with a sterile scalpel and processed using Qiagen DNeasy Blood and Tissue kit. We selected these extraction kits as they have been designed to be efficient for DNA extraction for the samples in our study.

Environmental samples of the feed ($n = 4$ FOF, $n = 6$ NCAT), water ($n = 2$ FOF, $n = 6$ NCAT), soil ($n = 8$ FOF, $n = 6$ NCAT), and air ($n = 4$ FOF, $n = 4$ NCAT) were collected at each farm on two separate days prior to the pigs being sent for processing. A total volume of 50 mL of the feed and water samples were collected directly from the water and feed troughs. Topsoil samples were collected by removing

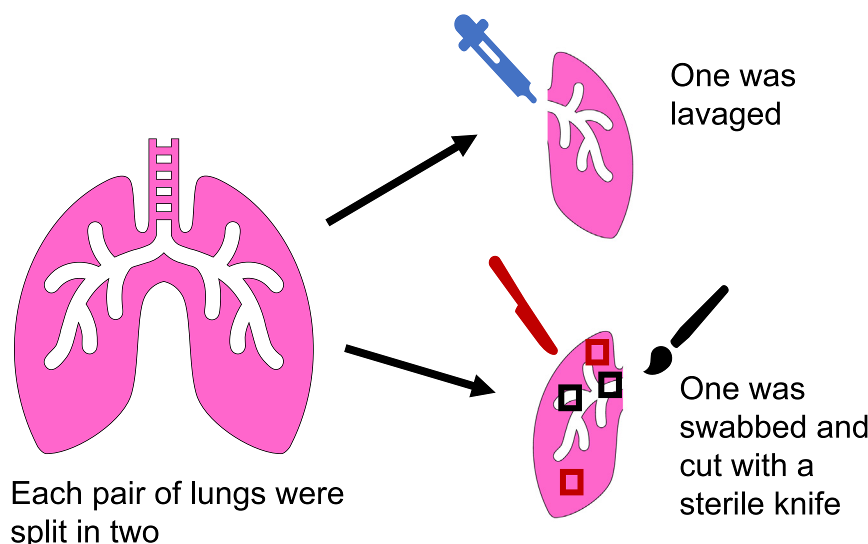


Fig. 1. Outline of sample collection. Boxes indicate approximate location of sample collection.

vegetation from the top of a 2 by 2-inch square and removing the top 1 cm of soil. These soils were later homogenized prior to extraction. Air samples were collected upwind and downwind using a Bobcat ACD-200 (INNOVAPrep, Drexel, Missouri) operated at 100 L/min for 1 h. The air sample was collected on a filter and bacteria were collected by using Bobcat rapid filter elution kit with 1× phosphate buffer saline solution. The eluent was subsequently passed through a 0.2 μm filter prior to beginning the extraction protocol. The feed and soil samples were processed with Qiagen PowerSoil kit. The water and air samples were processed with Qiagen's PowerWater kit. Isolated DNA was stored at −20 °C until further processing.

2.4. High throughput sequencing analysis

The isolated genomic DNA was subsequently used as a template for PCR amplification of the v4 region (515F FWD: GTGYCAGCMGCCGCG GTAA; 806R REV: GGACTACNVGGGTWTCTAAT, Earth Microbiome Project) of the 16S rRNA gene. Samples were prepared using the standard Illumina workflow protocol for 16S amplicon library preparation (Illumina, San Diego, California). Then, 250 paired-end sequencing was performed on an Illumina MiSeq using v3 chemistry at the Duke Sequencing Center (DUGSIM, Duke University, Durham, NC). The raw FASTQ files were demultiplexed and split by sample ID using QIIME. The files were then quality filtered, trimmed, denoised, merged, checked for chimeras, and assigned taxonomy to generate amplicon sequence variants (ASVs) using DADA2 (Callahan et al., 2016). Taxonomic assignments were made using the Silva v132 database (Quast et al., 2013). Removal of contaminating sequences was completed using the R package Decontam (Davis et al., 2018). The sequencing data for this project is available under project number: PRJNA650559. The code used to generate results can be accessed via GitHub through the following url: https://github.com/alexmccumber/PigLungs.

2.5. Data analysis

All data analysis was performed using the R programming suite. Several R packages were used for data analysis including: phyloseq for generating ordination plots and computing alpha and beta diversity, vegan for the generation of rarefaction curves to determine proper sequencing depth was obtained, stats for running statistical analyses, and ggplot2 for the generation of visual plots. linear discriminant analysis (LDA) effect size (LEfSe) (Segata et al., 2011) was used to identify significant differences between the differential abundances of sample groups. The LEfSe algorithm utilizes taxonomic and classification inputs to identify statistical relevance and biological significance between given input conditions. LEfSe was accessed through the Galaxy web interface. The diversity of the microbes within a sample (alpha) and between different samples (beta) were calculated using vegan (Jari Oksanen et al., 2018). SourceTracker was operated with default parameters to assess bacterial environmental origin in the lung samples.

Alpha diversity scores were tested for normality using the Shapiro-Wilkes test prior to statistical testing. For these analyses, normally distributed data significance testing with ANOVA was followed by pairwise testing with Tukey's HSD. If data was not normally distributed, significance testing was conducted with a Kruskal-Wallis test followed by pairwise testing via the Wilcoxon Ranked Sign test with False Discovery Rate (Benjamini-Hochberg) corrected p-values. Beta diversity ordination plots were examined for significance via PERMANOVA followed by pairwise PERMANOVA testing with Bonferroni corrected p-values.

2.6. Classifier development

The support vector machine (SVM) model classifier was built using the R package caret (Kuhn, 2008). To reduce the number of features, ASVs that were not present in at least 20% of the samples with more than 3 counts were removed. The resulting dataset had 291 ASVs. The count table was then transformed using Philr (Silverman et al., 2017). Next the environmental data was split form the pig lung data and used to build the SVM. The best model was selected using 5-fold repeated cross validation as well as tuning the cost and gamma parameters. After training, the model was used to predict the farm origin of the pig lung samples.

3. Results

3.1. The majority of amplicon sequence variants (ASVs) within the lung originate from air sources

From the 180 samples analyzed, after removing contaminating sequences through the R package Decontam, 21,193 ASVs were identified in 12,079,285 reads from the lung and environmental samples. These ASVs represent 1313 different genera. A breakdown of the number of ASVs by farm and sample type can be found in Table S1. The number of ASVs located at NCAT was greater than those at FOF in all samples

except for soil. The three phyla found in highest average relative abundance in the lung samples, listed in decreasing order, were Firmicutes, Proteobacteria and Bacteroidetes. A figure of the relative abundance of each phylum is provided in the supplementary information as Fig. S1. An ASV and taxonomic assignment table are provided as supplementary data.

SourceTracker (Knights et al., 2011) was utilized to identify the environmental origin of the ASVs found in the lung. SourceTracker uses a Bayesian approach to determine the origin of ASVs in "sink" samples from "source" samples. SourceTracker also allows for ASVs to be classified as having an unknown source when the model is not able to confidently classify them. SourceTracker was run on each set of lung samples (lavage, proximal A and B, and distal A and B), which were defined as sinks while the environmental samples were defined as sources. Among these potential sources, air was identified as the potential origin for the greatest proportion of the bacterial community in the lungs (Fig. 2). This was observed consistently at both farms and pairwise Wilcoxon testing showed that the median relative abundance value for air was significantly different ($p < 0.05$) from food, water, and soil. A stacked bar plot of the average source proportion of ASVs for each sample type is shown in Fig. 3. One difference between the two farms was the proportion of feed associated bacteria, which were found to comprise a greater proportion of the lung bacterial community at FOF compared to NCAT. Also, the distal samples at FOF had a higher proportion of feed associated ASVs compared to both the lavage and proximal samples. As for water and soil ASVs, they were not found to be the origin for a large portion of the lung microbiome at either farm. A notable proportion of the bacterial community in all samples from both farms was identified as having an unknown origin.

After finding that air was the origin of most bacteria within the lungs, the alpha diversity, or the total diversity found within each sample, was measured to determine if observed differences in environmental diversity would be reflected within the lung bacterial community (Fig. 4). The two alpha diversity metrics calculated were Faith's PD (FPD), which accounts for phylogeny (Faith, 1992), and the Shannon index, which accounts for the evenness of species across the sample (Abrahamsson et al., 2014). For the lung samples, a Shapiro-Wilk test indicated that the FPD values were not normally distributed ($p < 0.05$) whereas the Shannon values were ($p = 0.24$). Due to the non-normal distribution of the FPD values, a Kruskal-Wallis test with False Discovery Rate (FDR) corrected p-value pairwise testing via the Wilcoxon Ranked Sign test was applied which indicated that the median values of the proximal samples between farms were significantly different ($p < 0.05$). No difference in median alpha diversity was observed between the two farms lavage or distal samples. Contrary to the finding that significant differences existed for the FPD values, testing

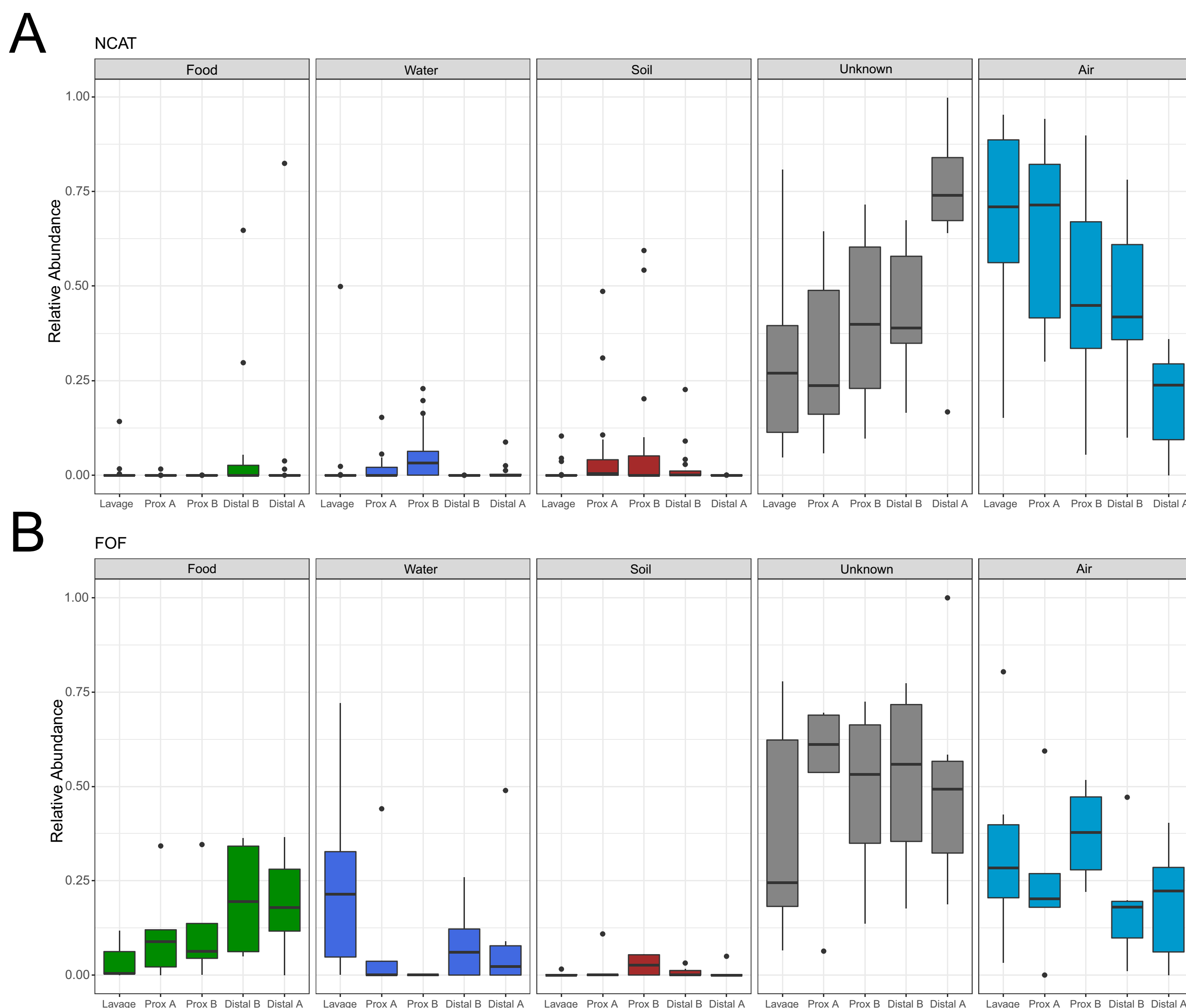


Fig. 2. SourceTracker results showing the proportion of the ASVs coming from each environmental sample by lung sample type for A) NCAT and B) FOF.

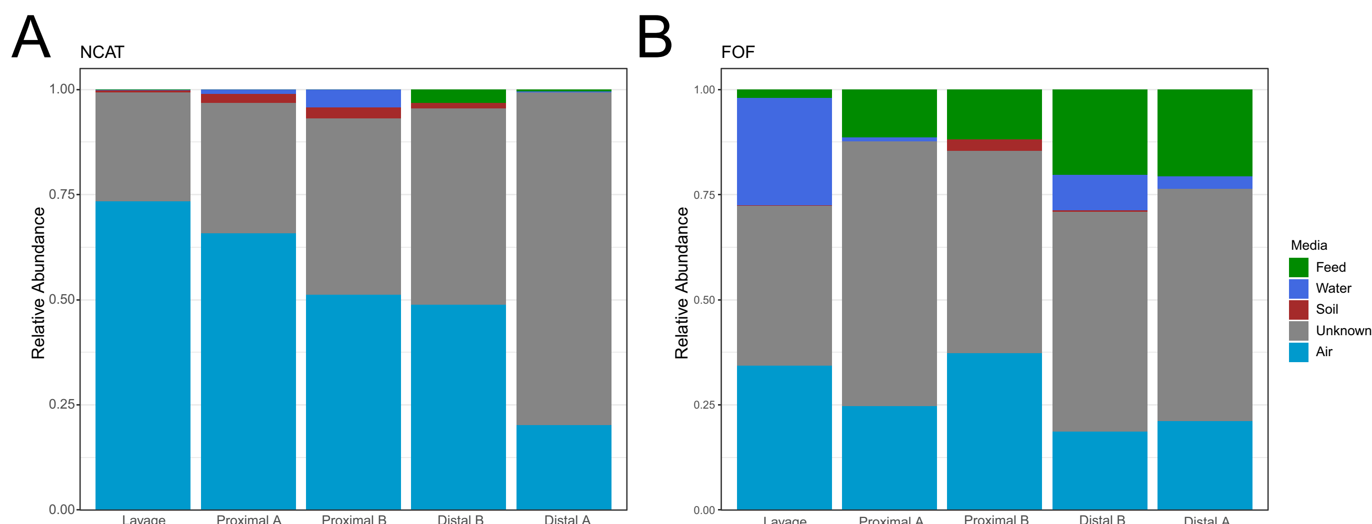


Fig. 3. SourceTracker results for the average proportion of ASVs from each environmental media for each sample type from farm A) NCAT and B) FOF. Outlier values, observed as points in Fig. 2, were removed in the creation of this figure.

with ANOVA showed that there are no significant differences for the Shannon values by farm, sample type, or farm by sample type. This suggests that while the phylogenetic make-up of those bacteria from NCAT may be more diverse, the species distribution is likely similar as measured by Shannon index within the proximal samples. When comparing the sample types between farms, the environmental samples showed no significant difference in either diversity metric. Due to the lack of observable differences in environmental diversity, no association between environmental bacterial alpha diversity and lung bacterial alpha diversity could be drawn.

While overall diversity within the lungs was not reflective of overall environmental diversity, beta diversity measures were calculated to examine overall differences between the farm and sample type's microbial composition. A principal coordinate (PCoA) plot was generated using the UniFrac beta diversity measurement (Fig. 5). The plot shows distinct clustering by sample type (lung vs environment) and by farm (NCAT vs FOF). Statistical testing using PERMANOVA revealed that the clustering due to differences in community compositions of the sample type by farm were significantly different ($p < 0.05$, $R^2 = 0.20$). PERMANOVA was re-run using only the environmental samples, but split by the respective media type (water, soil, air). The resulting pairwise testing showed that community composition did not meet statistical significance ($p = 0.064$). Interestingly, when running this analysis using the lung microbiome samples, significant differences ($p < 0.05$) were observed between the two farms. This result was further backed by analyzing the similarity of the microbial communities using ANOSIM. For the environmental samples, comparisons were found to be significantly different by farm ($p < 0.001$, $R = 0.43$) but not by sample type ($p = 0.066$, $R = 0.08$) whereas lung samples were significantly different by sample type ($p < 0.001$, $R = 0.40$) and farm ($p < 0.001$, $R = 0.52$). Altogether, these findings suggest that the observed environmental differences may lead to distinct community membership within the lungs.

3.2. SVM models can successfully predict lung sample origin using environmental samples

With observed differences in beta diversity established between the farms, a support vector machine (SVM) model was developed using the environmental samples. This model was created to determine whether those taxa driving the environmental differences could be used to successfully classify the lung samples from which farm they originated. After data transformation using Philr and model optimization, the model achieved an overall accuracy of 86% (0.79–0.92, 95% CI) with a sensitivity of 0.99 and specificity of 0.60. For NCAT, 89 of 90 samples were correctly classified. For FOF, 26 of 43 samples were correctly classified. A confusion matrix, or table showing the number of samples by the class that were correctly or incorrectly classified, is shown in Table 1. The high accuracy achieved by this model suggests that the differences in the environmental taxa are being reflected within the lung microbiome. Overall, the model performed better for NCAT than FOF, though sample type was found to influence model accuracy more than sample location. Compared to the lavage and proximal samples (0.89–0.99 Accuracy, 95% CI), the distal samples were more difficult for the model to predict (0.50–0.80 Accuracy, 95% CI) and accounted for 15 of the 17 misclassifications for FOF and the one NCAT misclassification. A summary of the top 20 features and their relative importance are presented in Table S2.

3.3. Expansion of the lung-environment model

The bacterial-lung origin model was further expanded by reprocessing the distal samples through SourceTracker and using the Proximal and Lavage samples as a source. To remove effects occurring unique to each farm, only the samples from NCAT were used. This analysis was completed with two assumptions. First, previous models of particle deposition in the lungs have shown that separation occurs in the lungs based on size. Those particles with a diameter greater than 3 μm are primarily removed in the primary and secondary bronchi, whereas particles with diameters between 0.5 and 3 μm are found in the tertiary bronchi and removed by sedimentation then moved to the alveoli via Brownian diffusion (Klinger-Strobel et al., 2015). Therefore, it is likely that the bacteria are not transiting the airspace; rather they are being transported through settling and/or diffusion. Secondly, while the proximal samples only sampled primary or secondary bronchi, we hypothesize but did not directly test that the lavage samples include mainly those ASVs found in higher order bronchi and not the alveolar space. This assumption is due to the lung being collapsed during sampling and the intrapleural pressure no longer being present. Thus, upon re-running SourceTracker with these assumptions, the known origin of ASVs in the distal sample increased with the primary source being identified as proximal/lavage samples. This simultaneously eliminated nearly all those ASVs previously identified as originating from the air (Fig. 6). Additionally, the unknown origin for many ASVs decreased (from 73% to 60% for Distal A samples and from 41% to 17% for Distal B samples).

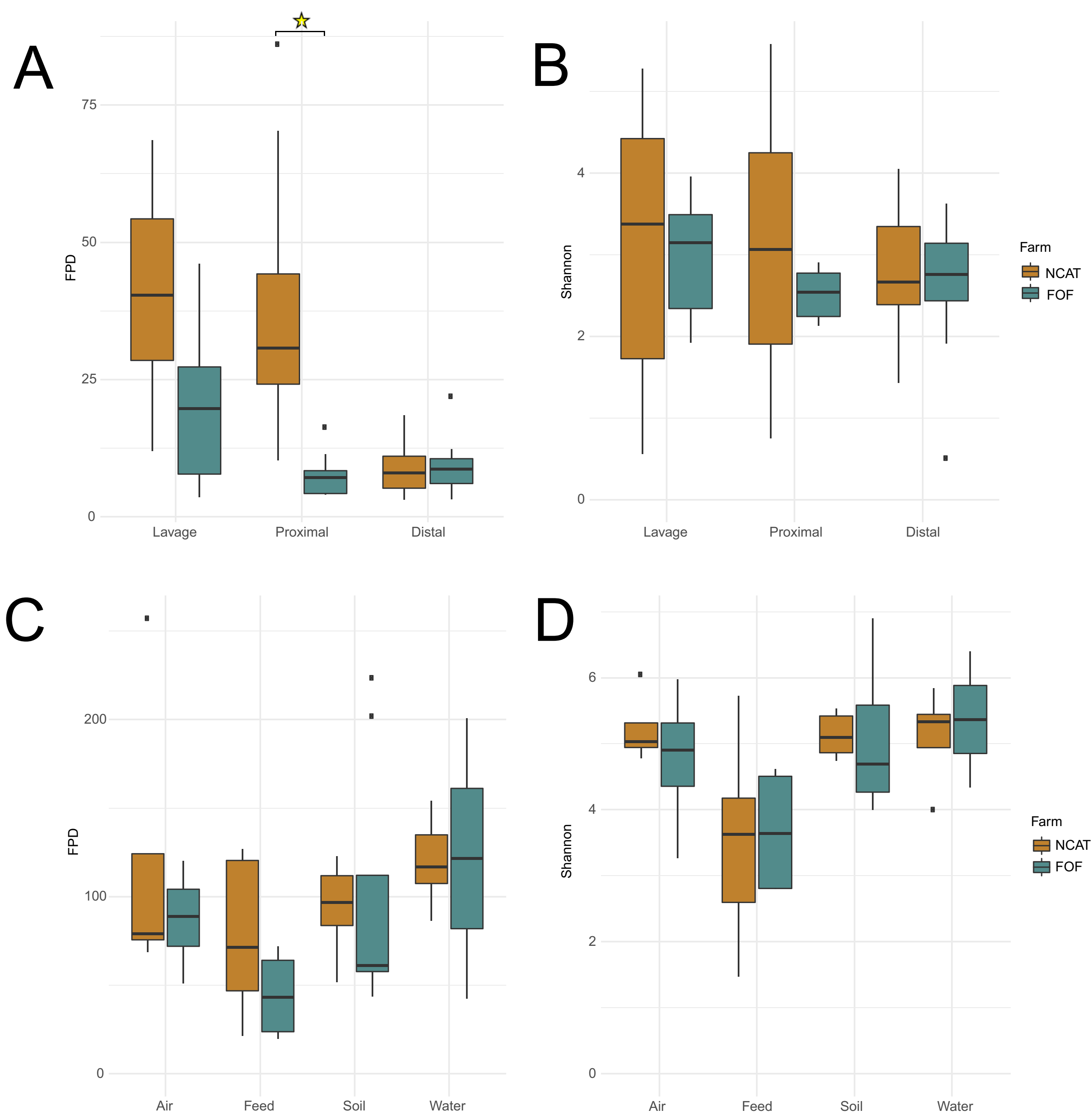


Fig. 4. Box plots of comparing alpha diversity between farms using A) Faith's Phylogenetic Diversity (FDP) by lung sample type and B) Shannon diversity by lung sample type C) FDP by environment sample type and D) Shannon diversity by environment sample type. Star indicates significance ($p < 0.05$).

3.4. Identification of biomarker taxa through LEfSe

Using the results from SourceTracker, which indicated that most of the bacteria in the lungs originated from upstream in the airway, LEfSe was utilized to identify potential biomarker taxa that were different between the proximal and lavage samples from the distal lung samples. LEfSe identified 16 genus level biomarker taxa in the proximal and lavage samples and 7 genus level biomarker taxa in the distal samples. Fig. 7 provides the list of these 23 taxa and their associated LDA scores.

4. Discussion

Overall, this study has shown that differences in lung community membership are significantly influenced by the local environment. This was concluded due to observed differences in beta diversity between the environmental and pig lung samples by farm and the SVM model where the differences in the microbiota in the environmental samples differentiate between the origin of the lung samples. Differences in human lung community membership due to geographical factors has been previously suggested as one study found significant differences between US and UK lung microbiomes (Stressmann et al., 2011). This effect has also been observed to occur between households that do or do not have dogs, as dogs appear to facilitate transport of outdoor microbes into the home (Fujimura et al., 2010). These studies, however, did not include detailed environmental sampling and therefore could not directly identify that the local environment was a factor in the observed differences.

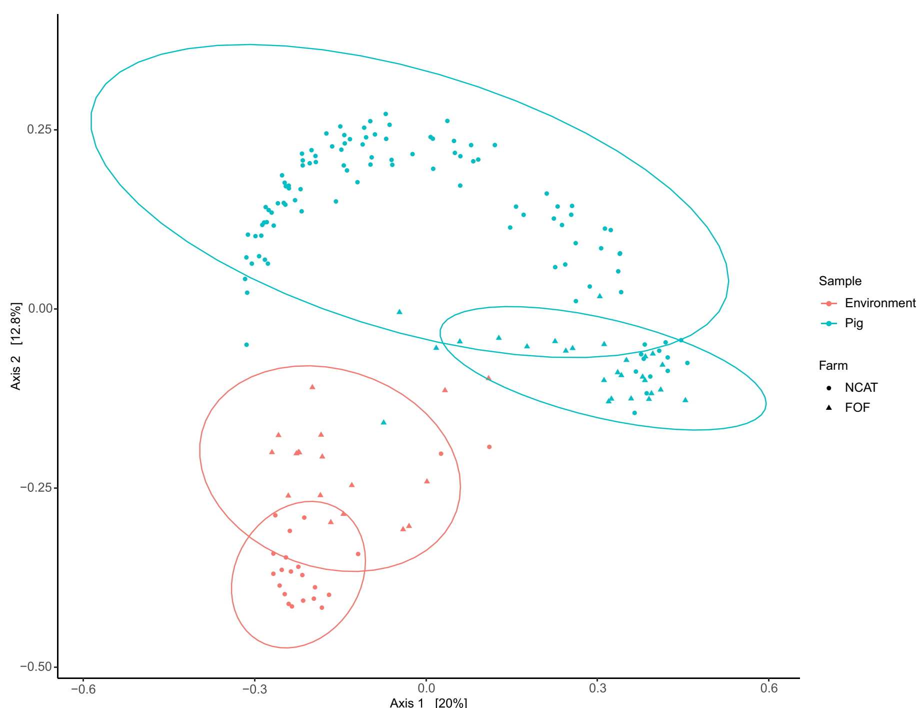


Fig. 5. Principal coordinate analysis of UniFrac distances for all lung and environment samples. The ellipse represents the 95% confidence. interval for the associated farm and sample group. Testing with PERMNOVA showed that each group was significantly different ($p < 0.05$).

Table 1
Confusion matrix for SVM model. Model accuracy was 86% and kappa was 66%.

Reference	Prediction	
	NCAT	FOF
NCAT	89	1
FOF	17	26

The influence of the local environment was confirmed using SourceTracker. SourceTracker analysis indicated that the principal environmental source of most lung associated ASVs in the proximal and lavage samples were air samples. This is an important to consider as we begin to understand how microbial communities in the lungs influence overall health; particularly as the world continues to urbanize (Cohen, 2006) and people begin spending most of the time indoors (Ruiz-Calderon et al., 2016). Studies have already shown that the outdoor and

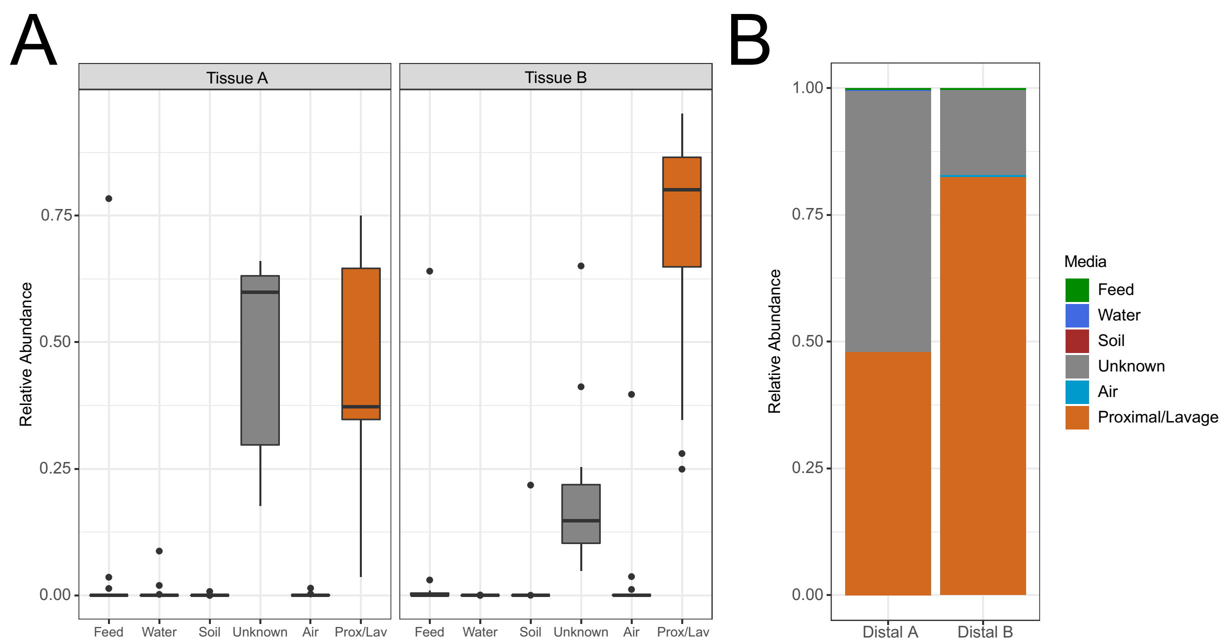


Fig. 6. Results of rerunning SourceTracker with the proximal/lavage samples as a potential source for the distal (tissue) samples as either A) boxplots or B) stacked bar plot of average SourceTracker proportion. Outlier values, identified as points in A) were removed in the creation of B.

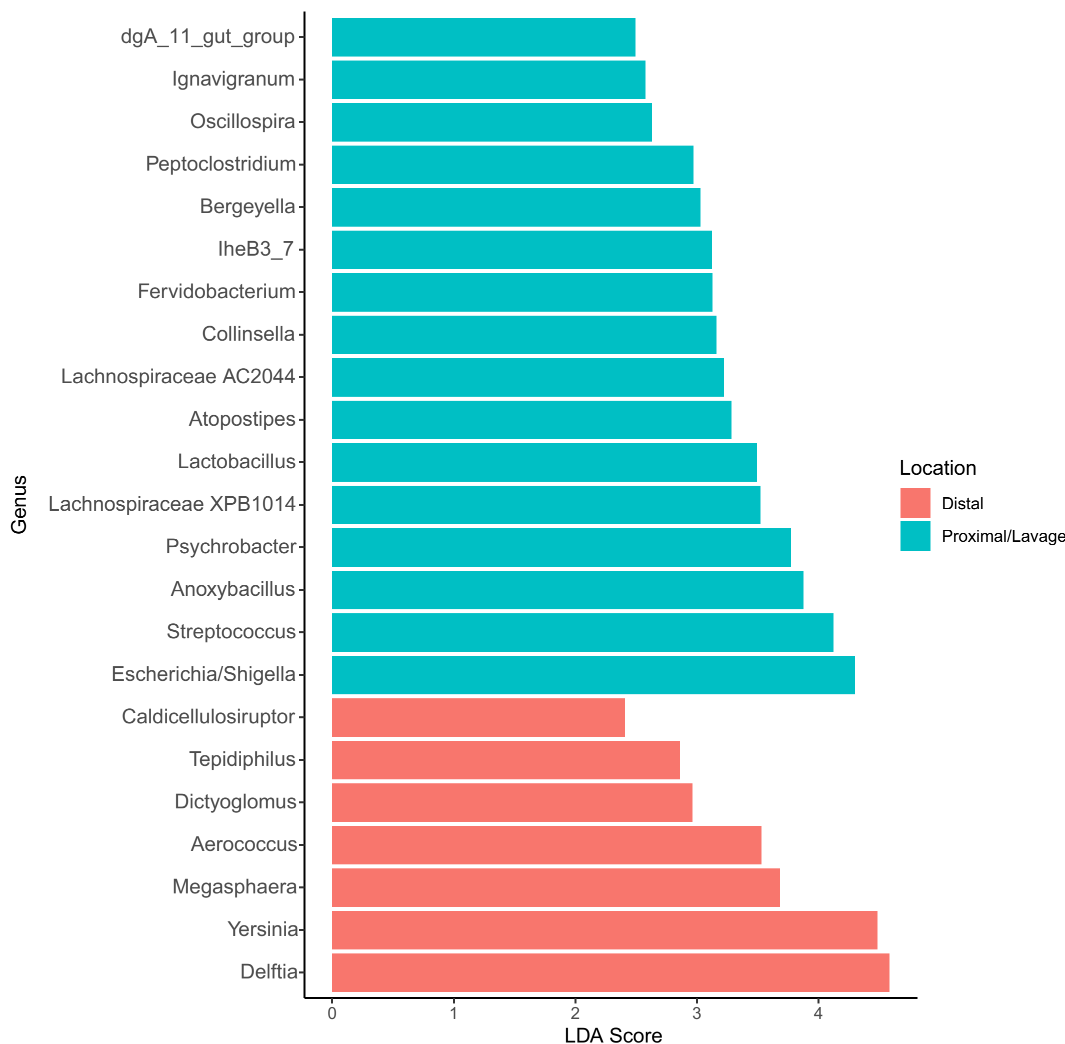


Fig. 7. LDA scores for significant genera identified by LEfSe between the Proximal/Lavage and Distal samples.

indoor airborne microbial communities differ (Miletto and Lindow, 2015; Shin et al., 2015), and the implications of the resulting long-term exposure to indoor microbial communities are not understood, but have been posited as the underlying cause for some diseases (Adams et al., 2016).

This study has also shown that the distal airspace supports a distinct microbial community fed by the proximal airways. Running SourceTracker using the environmental samples as the source revealed a large proportion of ASVs in the distal lung samples as being of unknown origin. This suggests that the distal lung samples are not being directly influences by the external environment. To address potential sources of the distal lung ASVs, we reran SourceTracker using the proximal and lavage samples as the source of the distal lung ASVs. In this setting, the number of unknown ASVs decreased; supporting that the proximal lung ASVs are an important source of the distal ASVs. This agrees with previous models of particle deposition suggesting that bacteria found in the distal samples likely do not originate directly from the air but arrive at the alveolar space through diffusion or sedimentation from the airway (Heyder, 2004; Asgharian et al., 2016). During this transiting process likely a combination of host environmental factors and properties intrinsic to the bacteria allow for their survival at the alveolar space create related, but distinct communities. This is observed through the finding that the total number and phylogenetic diversity of the ASVs found in the distal samples is lower than those found in the proximal samples. LEfSe also found potential biomarkers, or significant taxa for the proximal and distal airways. This was included for comparisons with future and previous studies, as it is beneficial to know which taxa are likely driving the differences between the sample types; however, the answer to the question as why some taxa are found in specific sample types is beyond the scope of this study and is likely better answered through shotgun metagenomic or metatranscriptomic studies, which may explain what is driving the survivability of those groups at the alveolar level.

One of the main limitations with this study is the small sample size, which limited the ability of the study to draw conclusions as to the influence of total environmental diversity being reflected within the lung and distinguish significantly the beta diversity between the farms respective environmental media. Additionally, there are still gaps to fill in this model such as our large unknown ASV proportion. One prevailing theory for this is that most lung bacteria are microaspirated (Dickson et al., 2017) from the mouth into the lung; therefore, in future studies, including more upstream samples such as the mouth as potential sources could further reduce the unknown origin category (Bassis et al., 2015). This could also be accomplished through more intensive environmental sampling.

The novel aspects of this study are that it is the first to our knowledge to attempt to classify lung associated bacterial profiles using environmental samples as inputs and the first to compare distal and proximal samples within the lungs. Overall, this study expands the knowledge of and provides the framework for a model for how the environment influences lung community assembly. Airborne bacteria are inhaled, deposited in the bronchioles, and undergo selection as they

transit toward the alveoli. The predominate phyla found in the lungs in this study, Firmicutes, Proteobacteria and Bacteroidetes, is consistent with what has been observed in previous studies of human (Erb-Downward et al., 2011; O'Dwyer et al., 2016) and pig (Huang et al., 2019; Megahed et al., 2019) lung microbiomes, providing confidence that, while this study was completed in pigs, these results likely provide insight into what is also occurring in human lungs.

5. Conclusion

In summary, this work has shown that the environment influences lung microbiome composition. This is an important consideration as we continue to further our understanding between the interaction of the environment and human health. There are still several questions that exist beyond those already proposed. One assumption in this study is that the lungs are in equilibrium with the environment, so those ASVs that are transient versus those that make up a core community are unable to be distinguished. This work was completed in a pig model continually exposed to the same environment. Therefore, turnover in populations were unable to be detected. Future work should utilize different environments and track temporal changes in the lung bacterial community to identify and to determine the timeline of turnover for those taxa that are transient members.

CRediT authorship contribution statement

Alexander W. McCumber: Conceptualization, Methodology, Investigation, Software, Formal analysis, Writing - original draft, Visualization. **Yeon Ji Kim:** Investigation. **Omoanghe Samuel Isikhuemhen:** Resources, Writing - review & editing. **Robert M. Tighe:** Conceptualization, Writing - review & editing. **Claudia K. Gunsch:** Resources, Conceptualization, Writing - review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to specifically thank Mr. Douglass Jones, Swine Unit Coordinator at NC A&T School farm and Dr. F. N. Anike, Mushroom Biology & Fungal Biotechnology Laboratory, North Carolina A&T State University for their help with experimental pigs and sample collection for this project. We would also like to thank Nick Rogers and Dr. Eric Monson for aiding with the development of the graphical abstract. The graphical abstract was created with BioRender.com.

This work was supported by funding from the National Institutes of Health (grant no. T32GM008555), part of the Center for Biomolecular and Tissue Engineering (CBTE) at Duke University. This work was made possible through the instruction provided by Dr. Joshua Granek, as part of the Integrative Bioinformatics for Investigating and Engineering Microbiomes (IBIEM) with funding provided by the National Science Foundation DGE-1545220.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.143623.

References

Abrahamsson, T.R., et al., 2014. Low gut microbiota diversity in early infancy precedes asthma at school age. Clin. Exp. Allergy 44, 842–850. https://doi.org/10.1111/cea.12253.

Adams, R.I., et al., 2016. Ten questions concerning the microbiomes of buildings. Build. Environ. 109, 224–234. https://doi.org/10.1016/j.buildenv.2016.09.001.

Asgharian, B., et al., 2016. Modeling particle deposition in the pig respiratory tract. J. Aerosol Sci. 99, 107–124. https://doi.org/10.1016/j.jaerosci.2016.01.016.

Bassis, C.M., et al., 2015. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. mBio 6, e00037. https://doi.org/10.1128/mBio.00037-15.

Beck, J.M., Young, V.B., Huffnagle, G.B., 2012. The microbiome of the lung. Transl. Res. 160, 258–266. https://doi.org/10.1016/j.trsl.2012.02.005.

Brett, K., Barman, S., Boitano, S., Brooks, H., 2012. Ganong's Review of Medical Physiology. 26th edn. McGraw-Hill Education.

Callahan, B.J., et al., 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869.

Carney, S.M., et al., 2020. Methods in lung microbiome research. Am. J. Respir. Cell Mol. Biol. 62, 283–299. https://doi.org/10.1165/rcmb.2019-0273TR.

Caverly, L.J., Huang, Y.J., Sze, M.A., 2019. Past, present, and future research on the lung microbiome in inflammatory airway disease. Chest 156, 376–382. https://doi.org/10.1016/j.chest.2019.05.011.

Cohen, B., 2006. Urbanization in developing countries: current trends, future projections, and key challenges for sustainability. Technol. Soc. 28, 63–80. https://doi.org/10.1016/j.techsoc.2005.10.005.

Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., Callahan, B.J., 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 6, 226. https://doi.org/10.1186/s40168-018-0605-2.

Dickson, R.P., Huffnagle, G.B., 2015. The lung microbiome: new principles for respiratory bacteriology in health and disease. PLoS Pathog. 11, e1004923. https://doi.org/10.1371/journal.ppat.1004923.

Dickson, R.P., et al., 2015. Spatial variation in the healthy human lung microbiome and the adapted island model of lung biogeography. Ann. Am. Thorac. Soc. 12, 821–830. https://doi.org/10.1513/AnnalsATS.201501-029OC.

Dickson, R.P., et al., 2017. Bacterial topography of the healthy human lower respiratory tract. mBio, 8 https://doi.org/10.1128/mBio.02287-16.

Dickson, R.P., et al., 2018. The lung microbiota of healthy mice are highly variable, cluster by environment, and reflect variation in baseline lung innate immunity. Am. J. Respir. Crit. Care Med. 198, 497–508. https://doi.org/10.1164/rccm.201711-2180OC.

Durack, J., et al., 2017. Features of the bronchial bacterial microbiome associated with atopy, asthma, and responsiveness to inhaled corticosteroid treatment. J. Allergy Clin. Immunol. 140, 63–75. https://doi.org/10.1016/j.jaci.2016.08.055.

Erb-Downward, J.R., et al., 2011. Analysis of the lung microbiome in the "healthy" smoker and in COPD. PLoS One 6, e16384. https://doi.org/10.1371/journal.pone.0016384.

Faith, D.P., 1992. Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61, 1–10.

Fillion-Bertrand, G., et al., 2019. Lung microbiome is influenced by the environment and asthmatic status in an equine model of asthma. Am. J. Respir. Cell Mol. Biol. 60, 189–197. https://doi.org/10.1165/rcmb.2017-0228OC.

Fujimura, K.E., et al., 2010. Man's best friend? The effect of pet ownership on house dust microbial communities. J. Allergy Clin. Immunol. 126, 410–412 412 e411-413. https://doi.org/10.1016/j.jaci.2010.05.042.

Han, M.K., et al., 2014. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. Lancet Respir. Med. 2, 548–556. https://doi.org/10.1016/s2213-2600(14)70069-4.

Heyder, J., 2004. Deposition of inhaled particles in the human respiratory tract and consequences for regional targeting in respiratory drug delivery. Proc. Am. Thorac. Soc. 1, 315–320. https://doi.org/10.1513/pats.200409-046TA.

Huang, Y.J., Boushey, H.A., 2015. The microbiome in asthma. J. Allergy Clin. Immunol. 135, 25–30. https://doi.org/10.1016/j.jaci.2014.11.011.

Huang, T., et al., 2019. Microbial communities in swine lungs and their association with lung lesions. Microb. Biotechnol. 12, 289–304. https://doi.org/10.1111/1751-7915.13353.

Jari Oksanen, F.G.B., Friendly, Michael, Kindt, Roeland, Legendre, Pierre, McGlinn, Dan, Minchin, Peter R., O'Hara, R.B., Simpson, Gavin L., Solymos, Peter, Stevens, M. Henry H., Szoecs, Eduard, Wagner, Helene, 2018. vegan: Community Ecology Package.

Klinger-Strobel, M., Lautenschlager, C., Fischer, D., Mainz, J., Bruns, T., Tuchscherr, L., Pletz, M., Makarewicz, O., 2015. Aspects of pulmonary drug delivery strategies for infections in cystic fibrosis - where do we stand? Expert Opin. Drug Deliv. 12 (1–24).

Knights, D., et al., 2011. Bayesian community-wide culture-independent microbial source tracking. Nat. Methods 8, 761–763. https://doi.org/10.1038/nmeth.1650.

Kuhn, M., 2008. Caret R package. J. Stat. Softw. 28.

Mair, K.H., et al., 2014. The porcine innate immune system: an update. Dev. Comp. Immunol. 45, 321–343. https://doi.org/10.1016/j.dci.2014.03.022.

Mayhew, D., Devos, N., Lambert, C., et al., 2018. Longitudinal profiling of hte lung micorbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations. Thorax 73, 422–430.

Megahed, A., et al., 2019. Impacts of environmental complexity on respiratory and gut microbiome community structure and diversity in growing pigs. Sci. Rep. 9, 13773. https://doi.org/10.1038/s41598-019-50187-z.

Meurens, F., Summerfield, A., Nauwynck, H., Saif, L., Gerdts, V., 2012. The pig: a model for human infectious diseases. Trends Microbiol. 20, 50–57. https://doi.org/10.1016/j.tim.2011.11.002.

Miletto, M., Lindow, S.E., 2015. Relative and contextual contribution of different sources to the composition and abundance of indoor air bacteria in residences. Microbiome 3, 61. https://doi.org/10.1186/s40168-015-0128-z.

Morgan, B.W., et al., 2018. Asthma and allergic disorders in Uganda: a population-based study across urban and rural settings. J Allergy Clin Immunol Pract 6 (e1582), 1580–1587. https://doi.org/10.1016/j.jaip.2017.11.032.

National Academies of Science, E., and Medicine, 2017. Microbiomes of the Built Environment: A Research Agenda for Indoor Microbiology, Human Health, and Buildings the National Academies Collection: Reports Funded by National Institutes of Health.
O'Dwyer, D.N., Dickson, R.P., Moore, B.B., 2016. The lung microbiome, immunity, and the pathogenesis of chronic lung disease. J. Immunol. 196, 4839–4847. https://doi.org/10.4049/jimmunol.1600279.
Parajuli, A., et al., 2018. Urbanization reduces transfer of diverse environmental microbiota indoors. Front. Microbiol. 9, 84. https://doi.org/10.3389/fmicb.2018.00084.
Prussin 2nd, A.J., Garcia, E.B., Marr, L.C., 2015. Total virus and bacteria concentrations in indoor and outdoor air. Environ. Sci. Technol. Lett. 2, 84–88. https://doi.org/10.1021/acs.estlett.5b00050.
Quast, C., et al., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219.
Rogers, C.S., et al., 2008. The porcine lung as a potential model for cystic fibrosis. Am. J. Physiol. Lung Cell. Mol. Physiol. 295, L240–L263. https://doi.org/10.1152/ajplung.90203.2008.
Ruiz-Calderon, J., Cavallin, H., Song, S., Novoselac, A., Pericchi, L., Hernandez, R., Branch, O., Pereira, H., Paulino, L., Blaser, M., Knight, R., Dominguez-Bello, M., 2016. Walls talk: microbial biogeography of homes spanning urbanization. Sci. Adv. 2 (1–7).
Ruokolainen, L., et al., 2017. Significant disparities in allergy prevalence and microbiota between the young people in Finnish and Russian Karelia. Clin. Exp. Allergy 47, 665–674. https://doi.org/10.1111/cea.12895.
Segata, N., et al., 2011. Metagenomic biomarker discovery and explanation. Genome Biol. 12, R60. https://doi.org/10.1186/gb-2011-12-6-r60.
Shin, S.K., et al., 2015. Metagenomic insights into the bioaerosols in the indoor and outdoor environments of childcare facilities. PLoS One 10, e0126960. https://doi.org/10.1371/journal.pone.0126960.
Silverman, J.D., Washburne, A.D., Mukherjee, S., David, L.A., 2017. A phylogenetic transform enhances analysis of compositional microbiota data. Elife, 6 https://doi.org/10.7554/eLife.21887.
Siqueira, F.M., et al., 2017. Microbiome overview in swine lungs. PLoS One 12, e0181503. https://doi.org/10.1371/journal.pone.0181503.
Stressmann, F.A., et al., 2011. Analysis of the bacterial communities present in lungs of patients with cystic fibrosis from American and British centers. J. Clin. Microbiol. 49, 281–291. https://doi.org/10.1128/JCM.01650-10.
Tong, X., et al., 2019. Alterations to the lung microbiome in idiopathic pulmonary fibrosis patients. Front. Cell. Infect. Microbiol. 9, 149. https://doi.org/10.3389/fcimb.2019.00149.
Volkoff, S.J., McCumber, A.W., Anderson, D.J., Gunsch, CK., 2019. Antibiotic-resistant bacteria on personal devices in hospital intensive care units: Molecular approaches to quantifying and describing changes in the bacterial community of personal mobile devices. Infect Control Hosp Epidemiol. 40 (6), 717–720. https://doi.org/10.1017/ice.2019.56.
Wang, Z., et al., 2016. Lung microbiome dynamics in COPD exacerbations. Eur. Respir. J. 47, 1082–1092. https://doi.org/10.1183/13993003.01406-2015.