Title: Application of an ecology-based analytic approach to discriminate signal and noise in low-biomass microbiome studies: whole lung tissue is the preferred sampling method for amplicon-based characterization of murine lung microbiota

Authors: Jennifer M. Baker[1,2], Kevin J. Hinkle[2], Roderick A. McDonald[2], Nicole R. Falkowski[2], Gary B. Huffnagle[1,2,3,4], and Robert P. Dickson[1,2,5*]

Affiliations:

1. Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109, USA
2. Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109, USA
3. Department of Molecular, Cellular, & Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA
4. Mary H. Weiser Food Allergy Center, University of Michigan Medical School, Ann Arbor, MI 48109, USA
5. Michigan Center for Integrative Research in Critical Care; Ann Arbor, Michigan

Corresponding Author Information:

Robert P. Dickson, MD
Pulmonary & Critical Care Medicine
University of Michigan Health System
6220 MSRB III / SPC 5642
1150 W. Medical Center Dr.
Ann Arbor, MI 48109-5642
rodickso@med.umich.edu
(734) 277-1487 (mobile)
(734) 764-2655 (fax)
(734) 936-9368 (lab)
(734) 936-5010 (administrator)

Abstract

Background: Low-biomass microbiome studies (such as those of the lungs, placenta, and skin) are vulnerable to contamination and sequencing stochasticity, which obscure legitimate microbial signal. Since low-biomass microbiome fields have had variable success in establishing the reality and clinical significance of identified microbiota, we sought to develop and apply an analytical approach to discriminate signal from noise in low-biomass microbiome studies. We used this approach to determine the optimal sampling strategy in murine lung microbiome studies, which will be essential for future mechanistic lung microbiome research.

Methods: Using a novel, ecology-based analytic approach, we compared bacterial DNA from the lungs of healthy adult mice collected via two common sampling approaches: homogenized whole lung tissue and bronchoalveolar lavage (BAL) fluid. We quantified bacterial DNA using droplet digital PCR, characterized bacterial communities using 16S rRNA gene sequencing, and systematically assessed the quantity and identity of bacterial DNA in both specimen types. We compared bacteria detected in lung specimens to each other and to potential source communities: negative (background) control specimens and paired oral samples.

Findings: By all measures, whole lung tissue in mice contained greater bacterial signal and less evidence of contamination than did BAL fluid. Relative to BAL fluid, whole lung tissue exhibited a greater quantity of bacterial DNA, distinct community composition, decreased sample-to-sample variation, and greater biological plausibility when compared to potential source communities. In contrast, bacteria detected in BAL fluid were minimally different from those of procedural, reagent, and sequencing controls.

Interpretation: An ecology-based analytical approach discriminates signal from noise in low-biomass microbiome studies and identifies whole lung tissue as the preferred specimen type for

murine lung microbiome studies. Sequencing, analysis, and reporting of potential source communities, including negative control specimens and contiguous biological sites, is crucial for biological interpretation of low-biomass microbiome studies, independent of specimen type.

Funding: National Institutes of Health

Introduction

Though the development of next-generation sequencing has led to heightened interest in the study of microbial communities across biological contexts, the study of low-biomass microbiomes is particularly challenging and requires the development of new methodological approaches. Low-biomass samples - samples with low densities of bacterial cells and therefore low quantities of bacterial DNA - are susceptible to contamination with background-derived signal, which affects the taxonomic composition of low-biomass samples[1,2] and makes it challenging to decipher biological meaning from sequencing data[3]. These methodological challenges exist in all fields that study low-biomass microbial communities across environmental, industrial, and biomedical contexts.

Low-biomass microbiome fields have had variable success in overcoming these methodological challenges. Whereas early findings related to the purported placenta microbiome have subsequently been attributed to contamination[4,5], the lung microbiome field has flourished with robust, validated findings: lung microbiota are detectable in health[6–12], correlated with lung immunity both in health[7,8] and disease[13–15], correlated with disease severity and predictive of response to therapy[16–19], and prognostic of clinical outcome in multiple conditions[20–27]. The lung microbiome field addressed the challenge of low-biomass microbiome sampling by systematically defining methods that collect representative populations of lung microbiota to maximize bacterial DNA content and minimize vulnerability to background contamination[9–12]. As a result, empirically validated sampling approaches such as bronchoalveolar lavage (BAL) fluid, which samples a large surface area and yields high sample volumes, and sputum, which contains concentrated densities of bacterial cells, have been successfully implemented in human lung microbiome studies[28].

Yet despite their routine use in human lung microbiome studies, these sampling methods are not easily adapted for sampling lung microbiota in murine models, which will be critical to understand the mechanisms that govern the relationship between respiratory tract microbiota and pulmonary disease. Anatomic considerations make the application of sequencing-based techniques to murine lung microbiome studies particularly challenging. Collection of BAL fluid is severely limited by the small (~1 mL) volume of the murine lung[29], and sputum collection is not possible in mice. In contrast, analysis of homogenized lung tissue is more feasible in mice than humans, and represents a viable option for maximizing the bacterial DNA content in murine lung samples[30]. The ability to effectively sample low-biomass microbial communities is inherently context-dependent and will require new solutions adapted to the particular context of each study.

We therefore designed an empirical approach to compare microbial signal detected in two distinct sample types collected from the same ecological site (murine lungs) with the following goals: 1) to assess the usefulness of microbial ecology-based analytical techniques repurposed for the discrimination of legitimate microbial signal from background noise and 2) to determine the sampling method that is best suited for the characterization of the murine lung microbiome. To accomplish these goals, we quantified and sequenced the bacterial DNA present in BAL fluid and whole lung tissue from otherwise genetically- and environmentally-identical healthy mice and compared them using a novel analytic approach (Figure 1).

Methods

Ethics approval: The animal studies described in this manuscript were approved by the Institutional Animal Care and Use Committee at the University of Michigan. Laboratory animal care policies at the University of Michigan follow the Public Health Service Policy on Humane Care and Use of Laboratory Animals.

Mice: Eight-week-old female C57BL/6 mice (n = 20) were purchased from Jackson Laboratories and housed under specific pathogen–free conditions. Mice were housed in five-animal cages in a common animal housing room and did not receive independent ventilation. Mice were allowed to acclimate for 1 week before harvest at 9 weeks of age. To avoid batch effect, mice were randomly assigned to specimen type (BAL fluid or whole lung tissue) and evenly sampled across cages. Animal experimentation was performed in compliance with the ARRIVE Guidelines[31,32]. Details regarding tissue collection and processing are reported in the online supplement (Figure 1, step 1).

DNA extraction, quantification, and 16S rRNA gene sequencing: DNA was extracted, amplified, and sequenced according to previously published protocols[33,34] (Figure 1, steps 2,4). Sequencing was performed with the MiSeq platform (Illumina). Bacterial DNA in lung specimens and negative controls was quantified with a QX200 ddPCR system (Bio-Rad, Hercules, CA) according to a previously published protocol[35]. Details are provided in the online supplement.

Data analysis: 16S rRNA gene sequencing data were processed using mothur (v. 1.43.0) according to the Standard Operating Procedure for MiSeq sequence data using a minimum sequence length of 250 base pairs[36,37]. Overall significance was determined as appropriate by the Kruskal-Wallis test and by permutational multivariate ANOVA (PERMANOVA) with 10,000 permutations using Euclidean distances (adonis). Pairwise significance was determined as

103 appropriate by the Wilcoxon test with the Benjamini-Hochberg correction for multiple
104 comparisons, Tukey's HSD test, and two-sample independent Mann-Whitney U test. All
105 statistical tests used p=0·05 as a threshold for significance. Details regarding statistical and
106 ecologic analysis are reported in the online supplement.

Results

Murine whole lung tissue contains more bacterial DNA than BAL fluid and negative controls

Obtaining quality sequencing data depends on the presence of sufficient bacterial DNA in the samples to be analyzed. Therefore, we first compared the quantity of bacterial DNA in whole lung tissue and BAL fluid obtained from healthy C57BL/6 mice (Figure 1, step 3). We hypothesized that whole lung tissue contains more bacterial DNA compared to BAL fluid. To test this hypothesis, we determined the number of 16S rRNA gene copies present in DNA isolated from whole lung tissue, BAL fluid, and negative control specimens using droplet digital PCR (ddPCR). As seen in Figure 2, BAL fluid and whole lung tissue both contained a significantly greater quantity of bacterial DNA than the isolation control (p=0·0084 and 0·0026, respectively). In contrast, BAL fluid did not contain more bacterial DNA than sampling controls or no template controls (p>0·05). Whole lung tissue contained significantly more bacterial DNA than all other groups, including all negative controls (p=0·0001). Whole lung tissue contained 27-fold more 16S rRNA gene copies than BAL fluid (64,110 vs. 2,367 mean copies/mL, respectively; p=0·0002). We thus concluded that murine whole lung tissue contains a greater quantity of bacterial DNA than does BAL fluid.

Having confirmed the presence of detectable bacterial DNA in whole lung tissue and BAL fluid, we proceeded with 16S rRNA gene sequencing according to a standard low-biomass protocol. Along with whole lung tissue and BAL fluid, we sequenced a variety of controls, including cecum as a high-biomass positive control, tongue as a low-biomass positive control and potential source community of the lower respiratory tract, a synthetic mock community as a positive sequencing control, and negative controls for each stage of specimen processing, including sampling, DNA isolation, and sequencing controls. Details regarding adequacy of sequencing are provided in the online supplement.

Murine whole lung tissue has increased alpha diversity and decreased sample-to-sample variation relative to BAL fluid and negative controls

We next determined if the alpha (within-sample) diversity also differed across sampling approaches (Figure 1, step 5). We hypothesized that the increased quantity of bacterial DNA in whole lung tissue would yield greater diversity of bacterial taxa in whole lung tissue compared to BAL fluid. To test this hypothesis, we calculated community richness as measured by the number of unique operational taxonomic units (OTUs) present in each specimen and negative control. As predicted, whole lung tissue had greater community richness than BAL fluid ($p=0{\cdot}001$) and sampling, isolation, and sequencing controls ($p<0{\cdot}001$ for all comparisons) (Figure 3). In contrast, whole lung and BAL specimens did not significantly differ in Shannon diversity index, which reflects both community richness and evenness ($p>0{\cdot}05$; Supplementary Figure 2). We therefore concluded that alpha diversity differs across sampling approaches, with greater alpha diversity in whole lung tissue driven by the detection of greater numbers of unique OTUs relative to BAL fluid.

Since BAL fluid contained low quantities of bacterial DNA and fewer unique OTUs than whole lung tissue, we suspected that incomplete sampling of the respiratory tract via saline lavage may also result in increased sampling and sequencing stochasticity[38], which both lead to decreased specimen-to-specimen reproducibility of cohoused mice (which have similar lung microbiota[6]). We thus hypothesized that whole lung tissue would have decreased sample-to-sample variation relative to BAL fluid, representing greater replicability. To test this hypothesis, we computed the Bray-Curtis dissimilarity index, a beta-diversity metric based on pairwise inter-sample distances between specimens of the same type (i.e. we compared each whole lung tissue specimen to each other whole lung tissue specimen, and likewise for BAL fluid). Whole lung tissue yielded a decrease in average Bray-Curtis dissimilarity index relative to that of BAL fluid and empty well controls ($p<0.0001$) (Figure 4). In contrast, the average Bray-Curtis

dissimilarity index for BAL fluid was not significantly different than the highly dissimilar empty well controls (p=0·27). These results indicate that whole lung tissue displays decreased sample-to-sample variation and samples the lung microbiome of mice more reproducibly than BAL fluid.

The taxonomic composition of murine whole lung tissue is similar to its oral microbiome source community and is distinct from negative controls, whereas that of BAL fluid is not distinct from negative controls

Having identified differences in bacterial quantity and diversity across sampling approaches, we next assessed whether the taxonomic composition of whole lung tissue and BAL fluid differed from each other and from negative controls (Figure 1, step 6). Since whole lung tissue had higher bacterial DNA content and alpha diversity than BAL fluid, we hypothesized that the taxonomic composition of BAL fluid would more closely resemble that of negative control specimens than would whole lung tissue, reflecting background contamination and sequencing noise as predominant sources of taxa in BAL fluid. To test this hypothesis, we used principal component analysis (PCA) to compare the similarity of taxa identified in whole lung tissue, BAL fluid, and negative control specimens. As seen in Figure 5A, the taxonomic composition of whole lung tissue was distinct from that of BAL fluid (p=0·00009) and pooled sampling controls (p=0·0004). In contrast, BAL fluid showed prominent overlap with sampling controls and did not differ in overall community composition (p=0·46). Similar results were obtained when comparing whole lung tissue and BAL fluid with isolation and sequencing controls (Supplementary Figure 3A,B). Overall, these data show that the taxonomic composition of whole lung tissue is distinct from that of BAL fluid and negative controls, whereas BAL fluid is not distinct from most negative controls.

We next assessed the biological plausibility of bacterial taxa by comparing whole lung tissue and BAL fluid communities to their likely source community, the oral microbiome (Figure 1, step

7). We hypothesized that the taxonomic composition of whole lung tissue would more closely resemble that of the oral microbiome source community than does BAL fluid. Principal component analysis confirmed that tongue and whole lung tissue display similar but statistically different (p=0·01) taxonomic compositions, whereas BAL fluid clusters separately both from tongue (p=0·00009) and whole lung tissue (Figure 5B). The clustering of BAL fluid with negative controls and tongue with whole lung tissue is also observed when plotting all lung, tongue, and negative control samples together (Supplementary Figure 3C). We confirmed these results by calculating the Bray-Curtis dissimilarity index for matched (i.e. from the same mouse) tongue and lung samples (Figure 5C). Consistent with the PCA results, whole lung tissue more closely resembled the oral source community than did BAL fluid (p=0·0004). Rank abundance analysis revealed that the prominent taxa in whole lung tissue were also common in tongue specimens, whereas taxa in BAL fluid bore little resemblance to oral taxa and instead resembled taxa in negative controls (Figure 5D). The similarity of taxa in the whole lung and tongue samples and the BAL fluid and negative control samples, respectively, can also be observed when ordering rank abundance plots by the taxa found in the tongue or pooled negative controls (Supplementary Figure 4). Together, these results confirm that the bacterial taxa identified in whole lung tissue are more biologically plausible than those detected in BAL fluid.

Discussion

This study illustrates how an ecology-based analytical approach can determine the reality of bacterial signal in low-biomass microbiome studies. Our approach revealed the superiority of murine whole lung tissue relative to BAL fluid in detecting bacterial signal, and validates the use of whole lung tissue for lung microbiome studies in mice. The bacterial signal in murine whole lung tissue is stronger than that of BAL fluid by all comparisons: increased quantity of bacterial DNA, greater diversity of bacterial taxa, and taxonomic composition that is reproducible across biological replicates, distinct from negative controls, and more similar to the oral microbiome, a biologically plausible source community (Table 1).

This study represents the first systematic comparison of sampling methods appropriate for the study of the murine lung microbiome. The lack of empirically-validated methods for sampling lung microbiota in mice is particularly concerning in light of the current reproducibility crisis[39] and recent controversial low-biomass studies[4,5,40], which highlight the dangers of over-interpreting noisy sequencing data in the absence of rigorous, field-specific standards. A systematic examination of methods for sampling lung microbiota in mice is overdue, especially considering the first report describing the murine lung microbiome was published almost a decade ago[41]. Published murine lung microbiome studies to date have used both whole lung tissue[6,42–47] as well as BAL fluid[48–50], but no study to date has directly compared sample approaches. Based on the findings of the current study, we strongly recommend whole lung tissue as a preferred sampling strategy for subsequent murine lung microbiome studies.

While BAL fluid in mice contains weak bacterial signal relative to lung tissue, in humans the opposite has been observed: human BAL specimens contain consistently stronger bacterial signal than lung tissue acquired via biopsy. This observation is consistent with anatomic and ecologic differences across species. Anatomically, human lungs are much larger than murine

lungs, providing increased surface area for sampling (~75 m^2 vs. 0·008 m^2) and more airspace (6 L vs. 0·001 L) to accommodate the collection of far larger volumes of BAL fluid[29,30,51]. Biopsy specimens of human lungs are typically small in volume and peripheral in anatomic location, meaning they are predominantly composed of interstitium rather than airways and alveolar space (where bacteria are more likely to be found). In contrast, use of whole lung homogenate in mice ensures capture of all bacterial DNA within the entire respiratory tract. Thus, anatomic and ecologic differences between humans and mice necessitates the use of murine-specific sampling approaches, and illustrates why a "one-size-fits-all" approach to low-biomass microbiome sampling is unlikely to work: sampling strategies will need to be tailored to their specific environmental and biologic contexts.

Numerous sources of false signal can confound detection of bacterial communities in low-biomass microbiome studies, including contamination (procedural, reagent, and sequencing) and sequencing stochasticity. Salter and colleagues elegantly demonstrated the susceptibility of low-biomass samples to reagent contamination by sequencing serial dilutions of a pure bacterial culture, where increasingly diluted specimens contained increasing abundances of taxa found in the DNA isolation reagents[1]. Other sources of contamination, such as those introduced during specimen collection (e.g. bronchoscope, surgical instruments, collection tubes) or sequencing (e.g. well-to-well contamination or index switching) may also alter the taxonomic composition of low-biomass samples[52,53]. Additionally, it has recently been demonstrated via the use of sequencing replicates that sequencing stochasticity is itself a major source of variability in microbial signal in low-biomass studies[38]. Given the numerous sources of potential false signal in low-biomass microbiome studies, we do not believe this methodological challenge can be sufficiently addressed with a simple, universal solution (e.g. a single bioinformatic "decontamination" step). Rather, as illustrated in our approach, we believe the reality of

253 microbiologic signal must be assessed within the specific ecologic context from which it is
254 sampled, and anchored in an understanding of microbial ecology.
255
256 Several alternate approaches to false signal in low-biomass microbiome studies have been
257 proposed. Strategies used to detect, interpret, and in some cases, eliminate, contamination
258 have included exclusion of taxa detected in negative controls through statistical packages[54,55] or
259 unbiased subtraction[7], extraction and sequencing technical replicates[38], calculation of
260 abundance ratios[56], correlation analyses[57], hierarchical clustering[58], and building neutral
261 models[59]. In contrast, we propose a simple analytical approach grounded in principles of
262 microbial ecology to discriminate true microbial signal from background-derived signal. While
263 this approach requires the use of several complementary metrics to determine the extent of
264 background-derived signal in each specimen type, it is relatively accessible for those conducting
265 microbiome studies due to its dependence on open-source software, conceptual familiarity to
266 microbial ecologists, and ease of application to other low-biomass sites. Fundamentally, this
267 approach relies on sampling the low-biomass body site of interest and comparing the size,
268 diversity, and taxonomic composition of the microbial community identified at that low-biomass
269 site to all potential source communities, including background signal derived from procedural,
270 reagent, and sequencing contamination and true microbial signal derived from contiguous body
271 sites. This approach can thus be applied to a single specimen type to discern true bacterial
272 signal from background-derived noise, or used to compare multiple specimen types to
273 determine the optimal sampling method in the absence of a gold standard. Our approach does
274 not preclude the use of complementary methods (such as those mentioned above), but rather
275 builds a foundation rooted in thorough experimental design and microbial ecology to support
276 further analyses.
277

There are several limitations to our study. We selected methods of harvesting BAL fluid and whole lung tissue which have been used by our lab and others successfully, and thus cannot directly speak to other approaches (e.g. use of lung portions or pooled BAL specimens from multiple mice). Our study only tested the use of whole lung tissue and BAL fluid for the purposes of amplicon-based sequencing, and may yield different results if other sequencing methods (e.g. metagenomic sequencing) are applied. Whole lung tissue contains much more host DNA than bacterial DNA, which can confound attempts at metagenomic analyses due to the depth of sequencing required to return reliable bacterial data[60]. Given the impossibility of performing both BAL and whole lung homogenization on the same mouse, we could not perform paired analysis on the same mice. We assumed, based on prior results[6], that co-housed mice from the same vendor and shipment should have lung bacterial communities with similar taxonomic composition. Yet it remains possible that mouse-to-mouse variation may have confounded some comparisons. Finally, despite our efforts to thoroughly account for all possible sources of bacterial signal found in both types of lung specimens, it is possible that we have not accounted for all potential source communities, including occult sources of contamination or other body sites in contact with the lungs, such as the nasopharynx and blood.

In conclusion, we here present an ecology-based analytical approach for distinguishing true bacterial signal from background contamination in low biomass microbiome studies and provide evidence supporting the use of whole lung tissue over BAL fluid in murine lung microbiome studies. The use of our ecology-based analytic approach highlights the importance of sequencing, analyzing, and reporting ample negative controls and, to the extent possible, contiguous anatomical sites or other biological source communities to assess the reality of bacterial signal in low-biomass microbiome studies.

Abbreviations

BAL: bronchoalveolar lavage; DNA: deoxyribonucleic acid; OTU: operational taxonomic unit; PCA: principal component analysis; PCR: polymerase chain reaction; rpm: revolutions per minute; PBS: phosphate-buffered saline; ddPCR: droplet digital PCR

Contributors

JMB participated in study design, sample collection, data acquisition, bioinformatic and statistical analysis, data interpretation, and drafted and revised the manuscript. KJH, RAM, and NRF participated in sample collection and processing and data acquisition. GBH participated in study design, data interpretation, and manuscript revision. RPD conceived the study design and participated in data interpretation and manuscript revision. All authors read and approved this version of the manuscript.

Declaration of interests

We declare no competing interests.

Data sharing

The dataset supporting the results of this article has been posted to the NIH Sequence Read Archive (accession number: PRJNA644805). The script used for mothur analysis can be found at https://github.com/piyuranjan/DicksonLabScripts/blob/master/mothurGreatLakes.sh. R code and accompanying files for the microbial community analysis and statistical tests presented in this paper can be found at https://github.com/dicksonlunglab/WholeLungvBALFluid.

Funding

Funding provided by NIH R01 Projects 1R01HL144599 (RPD), 1R01HL121774 (GBH), and 1R01AI138348 (GBH).

329

330 **Acknowledgements**

331 This research was supported in part by the University of Michigan Medical School Host
332 Microbiome Initiative and by work performed by The University of Michigan Microbiome Core.
333 Figures were created in part using BioRender.com. The authors thank Piyush Ranjan,
334 Christopher Brown, Rishi Chanderraj, and John Erb-Downward for assistance with bioinformatic
335 and statistical analyses and William R. Branton for his assistance in generation of pilot data.

336

References

1. Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014; **12(1):** 87. doi: 10.1186/s12915-014-0087-z.

2. Biesbroek G, Sanders EAM, Roeselers G, et al. Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. Gilbert JA, ed. *PLoS ONE* 2012; **7(3):** e32942. doi: 10.1371/journal.pone.0032942.

3. The Lancet Infectious Diseases. Microbiome studies and “blue whales in the Himalayas.” *Lancet Infect Dis* 2018; **18(9):** 925. doi: 10.1016/S1473-3099(18)30503-6.

4. Lauder AP, Roche AM, Sherrill-Mix S, et al. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* 2016; **4(1):** 29. doi: 10.1186/s40168-016-0172-3.

5. Leiby JS, McCormick K, Sherrill-Mix S, et al. Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. *Microbiome* 2018; **6(1):** 196. doi: 10.1186/s40168-018-0575-4.

6. Dickson RP, Erb-Downward JR, Falkowski NR, Hunter EM, Ashley SL, Huffnagle GB. 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* 2018; **198(4):** 497-508. doi: 10.1164/rccm.201711-2180OC.

7. Segal LN, Alekseyenko AV, Clemente JC, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome* 2013; **1(1):** 19. doi: 10.1186/2049-2618-1-19.

8. Segal LN, Clemente JC, Tsay J-CJ, et al. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat Microbiol* 2016; **1(5):** 16031. doi: 10.1038/nmicrobiol.2016.31.

9. Charlson ES, Bittinger K, Haas AR, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 2011; **184(8):** 957-963. doi: 10.1164/rccm.201104-0655OC.

10. Charlson ES, Bittinger K, Chen J, et al. Assessing bacterial populations in the lung by replicate analysis of samples from the upper and lower respiratory tracts. *PLoS ONE* 2012; **7(9):** e42786. doi: 10.1371/journal.pone.0042786.

11. Dickson RP, Erb-Downward JR, Freeman CM, et al. Spatial variation in the healthy human lung microbiome and the adapted island model of lung biogeography. *Ann Am Thorac Soc* 2015; **12(6):** 821-830. doi: 10.1513/AnnalsATS.201501-029OC.

12. Dickson RP, Erb-Downward JR, Freeman CM, et al. Bacterial topography of the healthy human lower respiratory tract. *mBio* 2017; **8(1):** e02287-16. doi: 10.1128/mBio.02287-16.

13. Dickson RP, Erb-Downward JR, Prescott HC, et al. Intraalveolar catecholamines and the human lung microbiome. *Am J Respir Crit Care Med* 2015; **192(2):** 257-259. doi: 10.1164/rccm.201502-0326LE.

14. Wang J, Lesko M, Badri MH, et al. Lung microbiome and host immune tone in subjects with idiopathic pulmonary fibrosis treated with inhaled interferon-γ. *ERJ Open Res* 2017; **3(3):** 00008-02017. doi: 10.1183/23120541.00008-2017.

15. O'Dwyer DN, Ashley SL, Gurczynski SJ, et al. Lung microbiota contribute to pulmonary inflammation and disease progression in pulmonary fibrosis. *Am J Respir Crit Care Med* 2019; **199(9):** 1127-1138. doi: 10.1164/rccm.201809-1650OC.

16. Huang YJ, Nariya S, Harris JM, et al. The airway microbiome in patients with severe asthma: associations with disease features and severity. *J Allergy Clin Immunol* 2015; **136(4):** 874-884. doi: 10.1016/j.jaci.2015.05.044

17. Zhang Q, Cox M, Liang Z, et al. Airway microbiota in severe asthma and relationship to asthma severity and phenotypes. Chalmers JD, ed. *PLoS ONE* 2016; **11(4):** e0152724. doi: 10.1371/journal.pone.0152724.

18. Durack J, Lynch SV, Nariya S, et al. Features of the bronchial bacterial microbiome associated with atopy, asthma, and responsiveness to inhaled corticosteroid treatment. *J Allergy Clin Immunol* 2017; **140(1):** 63-75. doi: 10.1016/j.jaci.2016.08.055.

19. Goleva E, Jackson LP, Harris JK, et al. The effects of airway microbiome on corticosteroid responsiveness in asthma. *Am J Respir Crit Care Med* 2013; **188(10):** 1193-1201. doi: 10.1164/rccm.201304-0775OC.

20. Leitao Filho FS, Alotaibi NM, Ngan D, et al. Sputum microbiome is associated with 1-year mortality after chronic obstructive pulmonary disease hospitalizations. *Am J Respir Crit Care Med* 2019; **199(10):** 1205-1213. doi: 10.1164/rccm.201806-1135OC.

21. Molyneaux PL, Cox MJ, Willis-Owen SAG, et al. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014; **190(8):** 906-913. doi: 10.1164/rccm.201403-0541OC.

22. Han MK, Zhou Y, Murray S, et al. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med* 2014; **2(7):** 548-556. doi: 10.1016/S2213-2600(14)70069-4.

23. Invernizzi R, Barnett J, Rawal B, et al. Bacterial burden in the lower airways predicts disease progression in idiopathic pulmonary fibrosis and is independent of radiological disease extent. *Eur Respir J* 2020; **55(4):** 1901519. doi: 10.1183/13993003.01519-2019.

24. Dickson RP, Schultz MJ, van der Poll T, et al. Lung microbiota predict clinical outcomes in critically ill patients. *Am J Respir Crit Care Med* 2020; **201(5):** 555-563. doi: 10.1164/rccm.201907-1487OC.

25. Panzer AR, Lynch SV, Langelier C, et al. Lung microbiota is related to smoking status and to development of acute respiratory distress syndrome in critically ill trauma patients. *Am J Respir Crit Care Med* 2018; **197(5):** 621-631. doi: 10.1164/rccm.201702-0441OC.

26. Dickson RP, Erb-Downward JR, Freeman CM, et al. Changes in the Lung Microbiome following Lung Transplantation Include the Emergence of Two Distinct Pseudomonas Species with Distinct Clinical Associations. Davis IC, ed. *PLoS ONE* 2014; **9(5):** e97214. doi: 10.1371/journal.pone.0097214.

27. Combs MP, Wheeler DS, Luth J, et al. Increasing relative abundance of pseudomonads predicts chronic rejection after lung transplant. *J Heart Lung Transplant* 2020; **39(4):** S65. doi: 10.1016/j.healun.2020.01.1266.

28. Carney SM, Clemente JC, Cox MJ, et al. Methods in lung microbiome research. *Am J Respir Cell Mol Biol* 2020; **62(3):** 283-299. doi: 10.1165/rcmb.2019-0273TR.

29. Schulte H, Mühlfeld C, Brandenberger C. Age-related structural and functional changes in the mouse lung. *Front Physiol* 2019; **10:** 1466. doi: 10.3389/fphys.2019.01466.

30. Dickson RP, Cox MJ. Sampling the lung microbiome. In: Cox MJ, Ege MJ, von Mutius E, eds. *The Lung Microbiome*. European Respiratory Society; 2019: 1-17. doi: 10.1183/2312508X.10015418.

31. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol* 2010; **8(6):** e1000412. doi: 10.1371/journal.pbio.1000412.

32. Percie du Sert N, Hurst V, Ahluwalia A, et al. Revision of the ARRIVE guidelines: rationale and scope. *BMJ Open Sci* 2018; **2(1):** e000002. doi: 10.1136/bmjos-2018-000002.

33. Mason KL, Erb Downward JR, Mason KD, et al. *Candida albicans* and bacterial microbiota interactions in the cecum during recolonization following broad-spectrum antibiotic therapy. *Infect Immun* 2012; **80(10):** 3371-3380. doi: 10.1128/IAI.00449-12.

34. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013; **79(17):** 5112-5120. doi:10.1128/AEM.01043-13.

35. Sze MA, Abbasi M, Hogg JC, Sin DD. A comparison between droplet digital and quantitative PCR in the analysis of bacterial 16S load in lung tissue samples from control and COPD GOLD 2. *PLoS ONE* 2014; **9(10):** e110351. doi: 10.1371/journal.pone.0110351.

36. Schloss PD. MiSeq SOP - mothur. 2019. https://mothur.org/wiki/miseq_sop/ (accessed on March 26, 2020).

37. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009; **75(23):** 7537-7541. doi: 10.1128/AEM.01541-09.

38. Erb-Downward JR, Falkowski NR, D'Souza JC, et al. Critical relevance of stochastic effects on low-bacterial-biomass 16S rRNA gene analysis. *mBio* 2020; **11(3):** e00258-20. doi: 10.1128/mBio.00258-20.

39. Schloss PD. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. *mBio* 2018; **9(3):** e00525-18. doi: 10.1128/mBio.00525-18.

40. Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. *Sci Transl Med* 2014; **6(237):** 237ra65. doi: 10.1126/scitranslmed.3008599.

41. Barfod K, Roggenbuck M, Hansen L, et al. The murine lung microbiome in relation to the intestinal and vaginal bacterial communities. *BMC Microbiol* 2013; **13(1):** 303. doi: 10.1186/1471-2180-13-303.

42. Singh N, Vats A, Sharma A, Arora A, Kumar A. The development of lower respiratory tract microbiome in mice. *Microbiome* 2017; **5(1):** 61. doi: 10.1186/s40168-017-0277-3.

43. Yun Y, Srinivas G, Kuenzel S, et al. Environmentally determined differences in the murine lung microbiota and their relation to alveolar architecture. *PLoS ONE* 2014; **9(12):** e113466. doi: 10.1371/journal.pone.0113466.

44. Richmond BW, Brucker RM, Han W, et al. Airway bacteria drive a progressive COPD-like phenotype in mice with polymeric immunoglobulin receptor deficiency. *Nat Commun* 2016; **7(1):** 11240. doi: 10.1038/ncomms11240.

45. Gollwitzer ES, Saglani S, Trompette A, et al. Lung microbiota promotes tolerance to allergens in neonates via PD-L1. *Nat Med* 2014; **20(6):** 642-647. doi: 10.1038/nm.3568.

46. Dickson RP, Singer BH, Newstead MW, et al. Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome. *Nat Microbiol* 2016; **1(10):** 16113. doi: 10.1038/nmicrobiol.2016.113.

47. Ashley SL, Sjoding MW, Popova AP, et al. Lung and gut microbiota are altered by hyperoxia and contribute to oxygen-induced lung injury in mice. *Sci Transl Med* 2020; **12(556):** eaau9959. doi: 10.1126/scitranslmed.aau9959.

48. Barfod KK, Vrankx K, Mirsepasi-Lauridsen HC, et al. The murine lung microbiome changes during lung inflammation and intranasal vancomycin treatment. *Open Microbiol J* 2015; **9(1):** 167-179. doi: 10.2174/1874285801509010167.

49. Poroyko V, Meng F, Meliton A, et al. Alterations of lung microbiota in a mouse model of LPS-induced lung injury. *Am J Physiol-Lung Cell Mol Physiol* 2015; **309(1):** L76-L83. doi: 10.1152/ajplung.00061.2014.

50. Yadava K, Pattaroni C, Sichelstiel AK, et al. Microbiota promotes chronic pulmonary inflammation by enhancing IL-17A and autoantibodies. *Am J Respir Crit Care Med* 2016; **193(9):** 975-987. doi: 10.1164/rccm.201504-0779OC.

51. Lutfi MF. The physiological basis and clinical significance of lung volume measurements. *Multidiscip Respir Med* 2017; **12(1):** 3. doi: 10.1186/s40248-017-0084-5.

52. Minich JJ, Sanders JG, Amir A, Humphrey G, Gilbert JA, Knight R. Quantifying and understanding well-to-well contamination in microbiome research. *mSystems* 2019; **4(4):** e00186-19. doi: 10.1128/mSystems.00186-19.

53. Sinha R, Stanley G, Gulati GS, et al. Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. *BioRxiv* 2017; published online April 9. doi: 10.1101/125724.

54. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 2018; **6(1):** 226. doi: 10.1186/s40168-018-0605-2.

55. McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR. microDecon: a highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. *Environ DNA* 2019; **1(1):** 14-25. doi: 10.1002/edn3.11.

56. Robinson KM, Crabtree J, Mattick JSA, Anderson KE, Dunning Hotopp JC. Distinguishing potential bacteria-tumor associations from contamination in a secondary data analysis of public cancer genome sequence data. *Microbiome* 2017; **5(1):** 9. doi: 10.1186/s40168-016-0224-8.

57. Weiss S, Van Treuren W, Lozupone C, et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J* 2016; **10(7):** 1669-1681. doi: 10.1038/ismej.2015.235.

58. Bosch AATM, Levin E, van Houten MA, et al. Development of upper respiratory tract microbiota in infancy is affected by mode of delivery. *EBioMedicine* 2016; **9:** 336-345. doi: 10.1016/j.ebiom.2016.05.031.

59. Venkataraman A, Bassis CM, Beck JM, et al. Application of a neutral community model to assess structuring of the human lung microbiome. *mBio* 2015; **6(1):** e02284-14. doi: 10.1128/mBio.02284-14.

60. Pereira-Marques J, Hout A, Ferreira RM, et al. Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Front Microbiol* 2019; **10:** 1277. doi: 10.3389/fmicb.2019.01277.

Figures and Tables

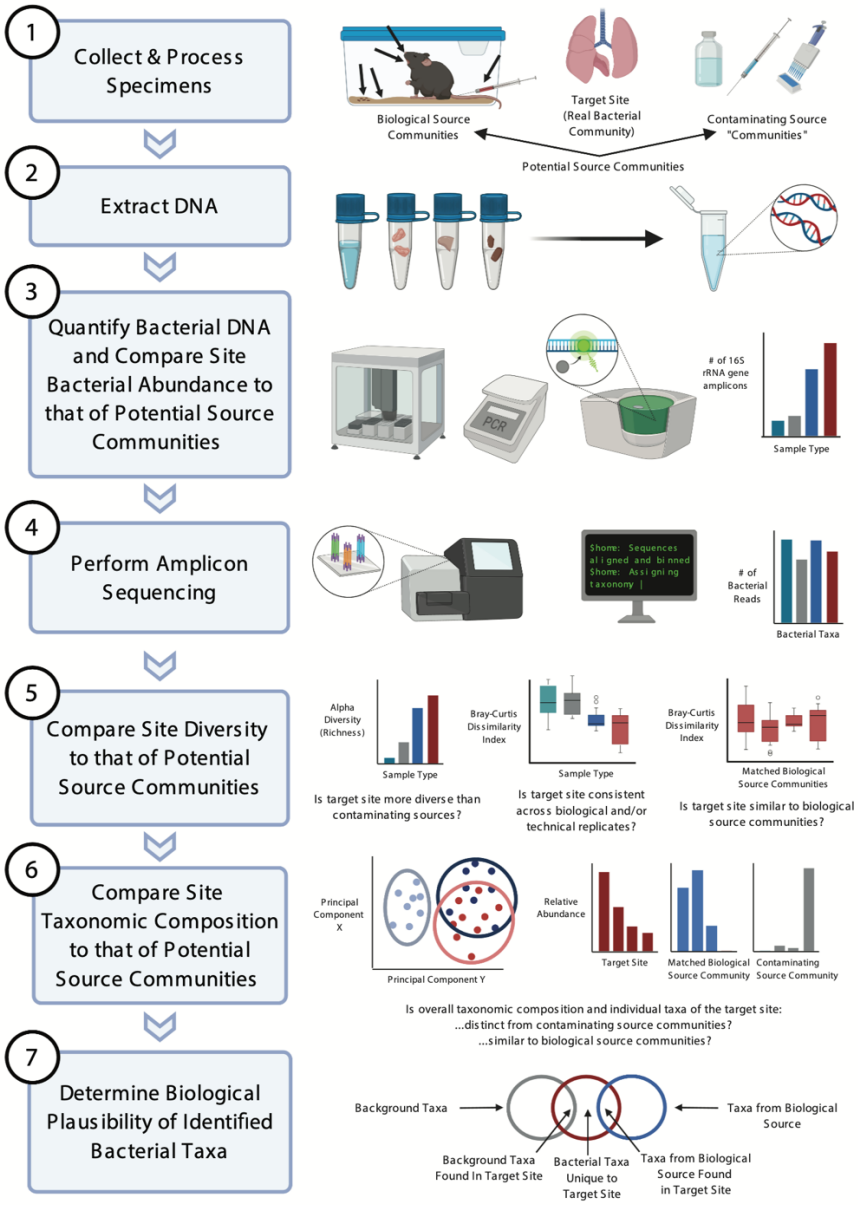

Figure 1: A newly defined experimental and analytic approach can distinguish bacterial signal from noise in low-biomass microbiome studies. Graphical and conceptual outline of an experimental and analytic approach to low-biomass microbiome studies. This approach was applied to murine lung microbiome sampling optimization as a proof-of-concept in this study, and may be useful in other low-biomass microbiome studies across biological contexts.

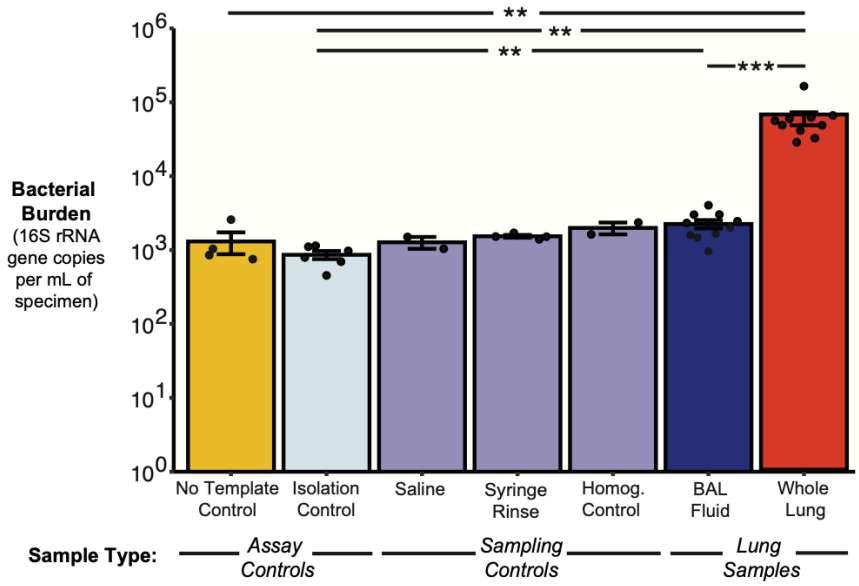

Figure 2: Murine whole lung tissue contains increased bacterial burden relative to BAL fluid and negative controls. Whole lung tissue contains more copies of the bacterial 16S rRNA gene per mL of DNA isolated from lung or control specimens as quantified by ddPCR. Mean ± SEM and individual data points (representing the average of technical duplicates) are shown. Overall significance was determined by the Kruskal-Wallis test ($p = 0.0001$). Pairwise significance was determined by the pairwise Wilcoxon test and corrected for multiple comparisons using the Benjamini-Hochberg method (pairwise comparisons including whole lung or BAL fluid that are not shown were not significant). Significance key: ns $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; ****$p \leq 0.0001$.

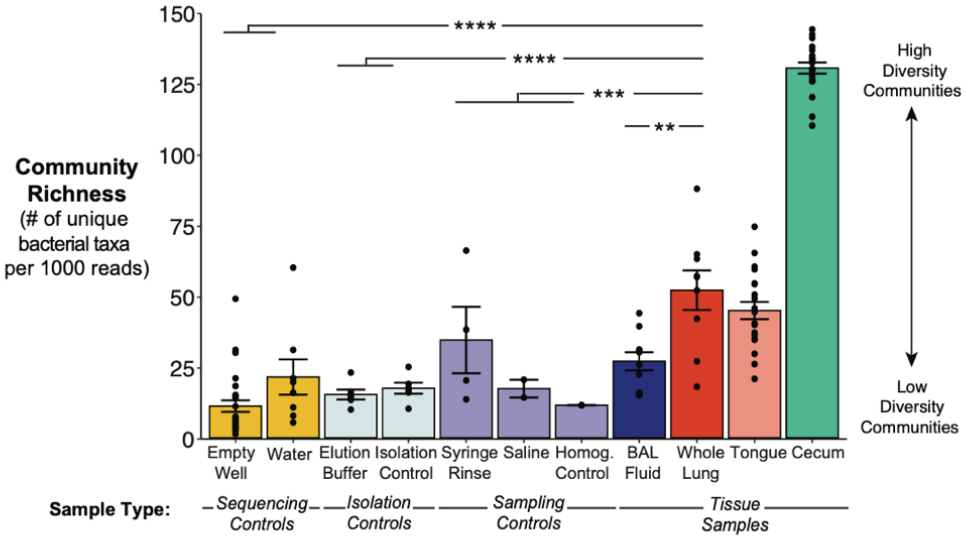

Figure 3: Bacterial communities in murine whole lung tissue have increased alpha diversity relative to BAL fluid and negative controls. A. Whole lung tissue contains a greater number of unique bacterial taxa than BAL fluid and negative controls. Richness of the bacterial community in each tissue or control specimen was determined by clustering reads with species-level similarity (≥ 97% sequence identity) into operational taxonomic units (OTUs) and calculating the number of unique OTUs within each specimen, normalized to 1000 reads per specimen. Mean ± SEM and individual data points are shown. Pairwise significance was determined by comparing whole lung tissue and BAL fluid to pooled sampling, isolation, and sequencing controls (respectively, as shown) using Tukey's HSD test. Significance key: ns $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; ****$p \leq 0.0001$.

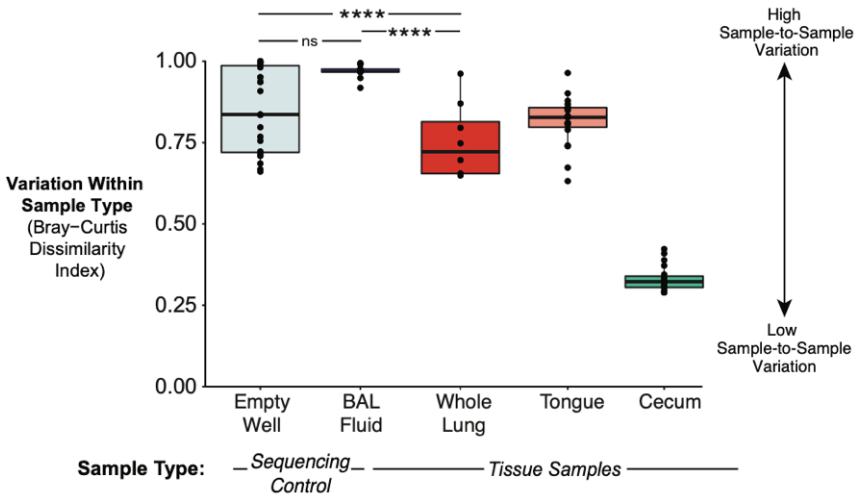

Figure 4: Bacterial communities in murine whole lung tissue show decreased variation among biological replicates compared to those in BAL fluid. Variation among lung bacterial communities of healthy mice from the same shipment was quantified using the Bray-Curtis dissimilarity index. For comparison, Bray-Curtis dissimilarity was also calculated for empty wells as a representative negative control with high variation, cecal communities as a representative body site with low variation, and tongue as a representative seed community for the lower respiratory tract. Median, IQR, and all unique pairwise comparisons (individual data points) are shown. Pairwise significance was determined by pairwise Wilcoxon test and corrected for multiple comparisons using the Benjamini-Hochberg method. Significance key: ns $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; ****$p \leq 0.0001$.

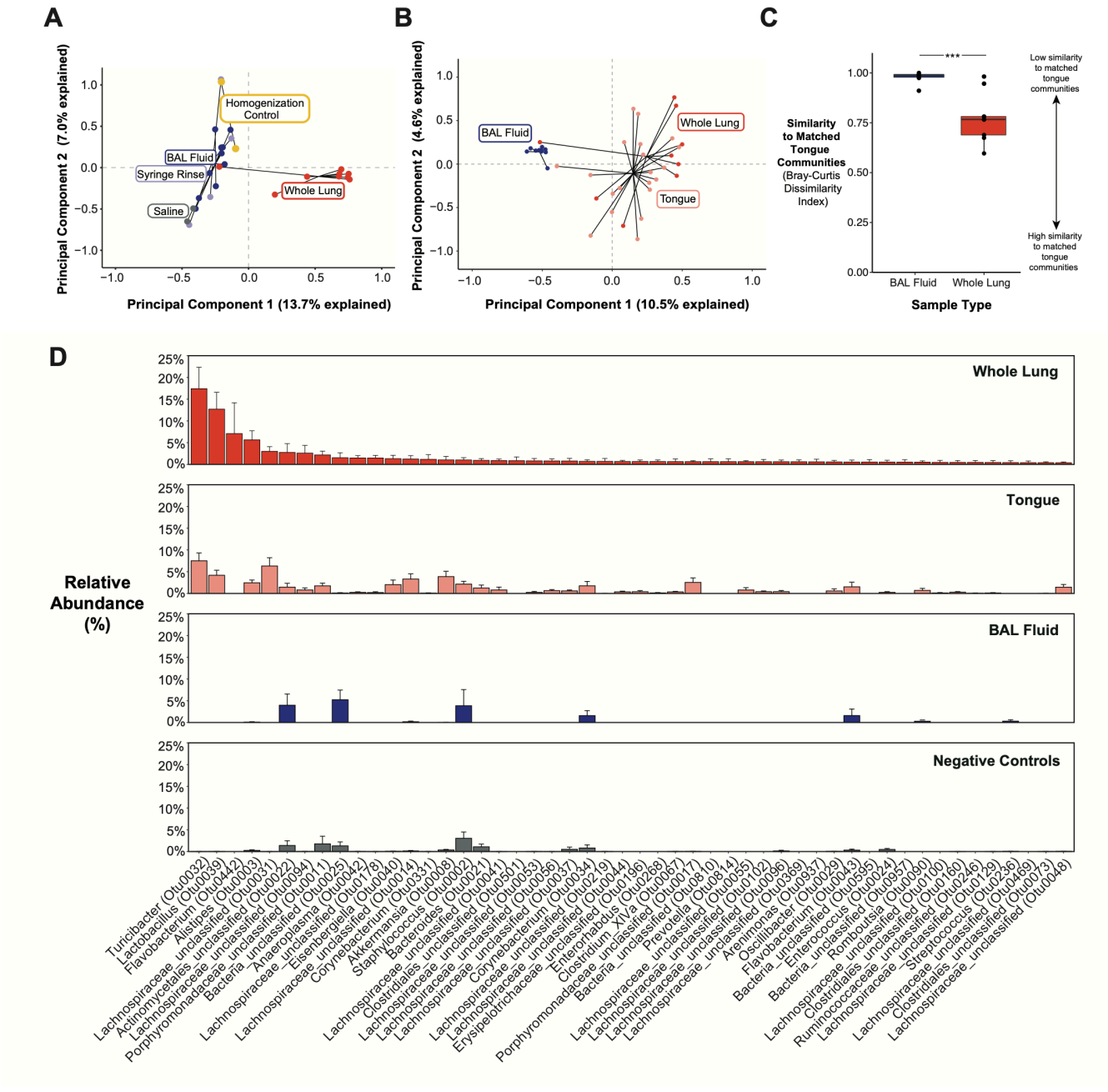

Figure 5: The taxonomic composition of bacterial communities in murine whole lung tissue is distinct from the background-dominant taxonomic composition of BAL fluid and similar to that of the oral microbiome, a biologically plausible source community. A. Whole lung tissue clusters separately from BAL fluid and sampling controls by principal

component analysis of Hellinger-transformed 16S rRNA gene sequencing data. Individual data points represent specimens grouped by sample or control type. B. Whole lung tissue, but not BAL fluid, clusters near tongue samples by principal component analysis of Hellinger-transformed 16S rRNA gene sequencing data. Individual data points represent specimens grouped by sample type. C. Bacterial communities in whole lung tissue are more similar to matched (within-mouse) oral communities than BAL fluid. Similarity of lung bacterial communities, grouped by sampling approach, to matched oral communities was quantified using Bray-Curtis dissimilarity index. Median, IQR, and individual data points representing within-mouse comparisons of oral and lung communities are shown. D. Relative abundance of bacterial taxa in whole lung tissue are similar to that of oral bacterial communities. In contrast, the relative abundance of bacterial taxa in BAL fluid are similar to that of negative controls. Bars are ranked by mean abundance in whole lung tissue and represent mean ± SEM percent relative abundance of the top 50 bacterial taxa (OTUs) in whole lung tissue across sample types. Labels denote genus (or most specific taxonomic level if no genus was assigned) and unique identifier for each OTU. Overall significance was determined by (A, B) permutational multivariate ANOVA ($p = 0.00009$ for both). Pairwise significance was determined by (A, B) two-sample PERMANOVA (A only: pooled sampling controls were compared to each lung sample type), and (C) two-sample unpaired Mann-Whitney U test. Significance key: ns $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; ****$p \leq 0.0001$.

589 **Table 1: Comparison of Sampling Methods for Murine Lung Microbiome Studies.**

	Whole Lung Tissue	**Bronchoalveolar Lavage Fluid**
Sample description	All lung lobes homogenized in sterile water	Dislodged airway and alveolar contents (microbes, leukocytes, epithelial cells) in sterile saline
Biological site sampled	Airway and intra-alveolar space, interstitium, & blood (if not perfused)	Airway and intra-alveolar space only
Bacterial biomass	Low	Low
Host-to-microbe DNA ratio	High	Low
Total DNA content	High	Low
Number of 16S rRNA gene copies	~ 10^4	~ 10^3
Variation among biological replicates	Low	High
Similarity to contaminating source “communities” (negative controls)	Low	High
Similarity to biological source community (oral microbiome)	High	Low

590