

RESEARCH

Open Access

Whole lung tissue is the preferred sampling method for amplicon-based characterization of murine lung microbiota



Jennifer M. Baker^{1,2}, Kevin J. Hinkle², Roderick A. McDonald², Christopher A. Brown², Nicole R. Falkowski², Gary B. Huffnagle^{1,2,3,4} and Robert P. Dickson^{1,2,5*} 

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. While human lung microbiome studies have rigorously identified sampling strategies that reliably capture microbial signal from these low-biomass microbial communities, the optimal sampling strategy for characterizing murine lung microbiota has not been empirically determined. Performing accurate, reliable characterization of murine lung microbiota and distinguishing true microbial signal from noise in these samples will be critical for further mechanistic microbiome studies in mice.

Results: Using an analytic approach grounded in microbial ecology, 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. 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.

(Continued on next page)

* Correspondence: rodickso@med.umich.edu

¹Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109, USA

²Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan Health System, 6220 MSRB III/SPC 5642, 1150 W. Medical Center Dr, Ann Arbor, MI 48109-5642, USA

Full list of author information is available at the end of the article



© The Author(s). 2021, corrected publication July 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusions: An ecology-based analytical approach discriminates signal from noise in this low-biomass microbiome study 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, are crucial for biological interpretation of low-biomass microbiome studies, independent of specimen type.

Keywords: Lung microbiome, 16S rRNA gene amplicon sequencing, Bronchoalveolar lavage, Whole lung tissue

Background

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]. This challenge of background DNA, contamination, and sequencing stochasticity (here collectively referred to as “noise”) intermingled with legitimate bacterial signal originating from a biological specimen (here referred to as “signal”) 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 [6–13], correlated with lung immunity both in health [7, 8, 14, 15] and disease [16–18], correlated with disease severity and predictive of response to therapy [19–22], and prognostic of clinical outcomes in multiple conditions [23–30]. The lung microbiome field addressed the challenge of low-biomass microbiome sampling in humans 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 (in humans) samples a large surface area and yields high sample volumes, have been successfully implemented in lung microbiome studies [31].

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, as collection of BAL fluid is severely limited by the small (~1 mL) volume of the murine lung and terminal nature of the procedure in mice [32]. 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 [33]. Thus, 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 goal: to determine the sampling method that is best suited for the characterization of the murine lung microbiome. To accomplish this goal, 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 microbial ecology-based analytic approach (Fig. 1).

Methods

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 [34, 35].

Tissue collection and processing

On the day of harvest, mice were randomized to either the whole lung tissue or BAL fluid groups (Fig. 1, step

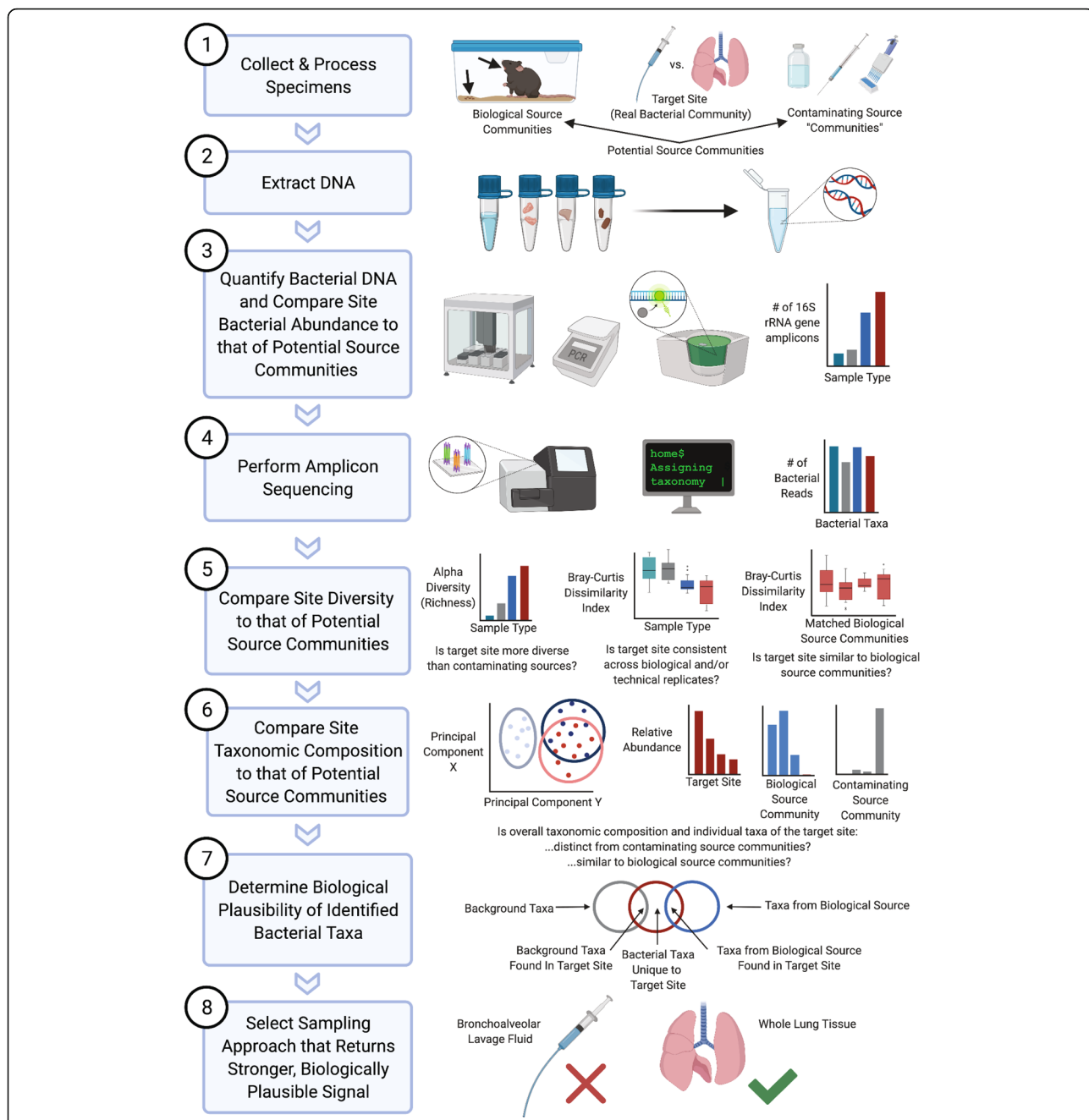


Fig. 1 An ecology-based experimental and analytic approach can distinguish bacterial signal from noise in low-biomass microbiome studies. Graphical and conceptual outline of an experimental and analytic approach to low-biomass microbiome studies. This approach was applied to murine lung microbiome sampling optimization in this study and may be useful in other low-biomass microbiome studies across biological contexts

1). To account for cage effect during downstream analysis, two or three mice per five-mouse cage were randomly assigned to each sampling group, for a total of 10 mice from each of the four cages in the whole lung tissue and BAL fluid groups. Mice were sacrificed via CO₂ asphyxiation, and organs were harvested in the following

order: tongue, whole lung tissue or BAL fluid, and cecum. A summary of all tissue samples and controls used in this study is provided in Supplementary Table 1. All tissue samples were collected using sterile technique, and instruments were rinsed with ethanol and flamed between each organ. Tongue (*n*=20) and cecum (*n*=20)

samples, which were collected to serve as low- and high-biomass positive controls, respectively, were immediately snap-frozen using liquid nitrogen and stored at -80°C until DNA isolation.

Both types of lung samples collected for this study were harvested and processed according to previously published protocols [6, 36], described in brief below. Murine whole lung tissue ($n=10$) was excised, placed in tubes containing 1 mL sterile water, and mechanically homogenized using a Tissue-Tearor (Biospec Products, Bartlesville, OK). The tissue homogenizer was cleaned and rinsed in ethanol and water between each tissue sample. Water control specimens from homogenization ($n=2$) rinsed with clean instruments were included as procedural controls for whole lung tissue. Lung homogenate was centrifuged at 13,000 rpm, supernatant was removed, and the cellular pellet was snap-frozen with liquid nitrogen and stored at -80°C until DNA isolation.

BAL fluid ($n=10$) was collected by (1) sterile dissection to expose and make a small incision in the trachea, (2) insertion of a piece of sterile tubing (BD Intramedic polyethylene tubing, 0.58-mm internal diameter, catalog no. 427410) and connected sterile syringe needle (BD PrecisionGlide™ 23 gauge needle, catalog no. 305145) into the incision, (3) tightening of a piece of sterile surgical thread around the intubated trachea to create an air-tight seal, and (4) two rounds of instillation and retrieval of 1 mL of sterile phosphate-buffered saline (PBS) into the lungs using a sterile syringe (BD 1 mL sterile syringe, catalog no. 309659). Each tubing-needle-syringe setup was rinsed thoroughly with sterile PBS between the collection of each sample. Sterile PBS ($n=2$) used for lavage and PBS rinses ($n=4$) of the syringe, needle, and tubing (pre- and post-lavage) were collected as procedural controls. BAL fluid was prepared by pooling the two serial lavages from each mouse, yielding up to 2 mL total BAL fluid per mouse. Pooled BAL fluid was centrifuged at 13,000 rpm for 30 min, supernatant was removed, and the cellular pellet was snap-frozen with liquid nitrogen and stored at -80°C until DNA isolation.

DNA extraction, quantification, and 16S rRNA gene sequencing

DNA was extracted, amplified, and sequenced according to previously published protocols [37, 38] (Fig. 1, steps 2 and 4). DNA isolation was performed with a single kit according to a modified protocol previously demonstrated to isolate bacterial DNA [37]. Briefly, genomic DNA was extracted from mouse tissue samples using a DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany, catalog no. 69506) and homogenized in PowerBead Tubes (Qiagen, Hilden, Germany, catalog no. 13123-50). To detect contamination introduced by the DNA

isolation kit, elution (AE) buffer ($n=6$) and specimen-free DNA isolations using empty bead tubes ($n=6$) were collected and sequenced as negative controls. Samples were processed in a randomized order to reduce false pattern formation due to reagent contamination [1].

Bacterial DNA quantification

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 [39]. Sampling and DNA isolation controls and sterile PCR-grade water used for sample dilution as a no template control ($n=4$) were run alongside lung specimens. All lung specimens and negative controls were run with two technical replicates. Droplets were generated using an automated droplet generator (Bio-Rad, catalog no. 1864101). PCR amplification was performed with the Bio-Rad C1000 Touch Thermal Cycler (catalog no. 1851197). Primers were 5'-GCAGGCCTAACACA TGCAAGTC-3' (63F) and 5'-CTGCTGCCTC CCGTAGGAGT-3' (355R). The cycling protocol was 1 cycle at 95°C for 5 min, 40 cycles at 95°C for 15 s and 60°C for 1 min, 1 cycle at 4°C for 5 min, and 1 cycle at 90°C for 5 min, with all steps at a ramp rate of $2^{\circ}\text{C}/\text{s}$. Droplets were detected using the automated droplet reader (Bio-Rad, catalog no. 1864003), quantified using Quantasoft™ Analysis Pro (version 1.0.596), and imported to R (version 4.0.2) for visualization and statistical analysis.

16S rRNA gene sequencing

The V4 region of the 16S rRNA gene was amplified using published primers [40] and the dual-indexing sequencing strategy developed by the laboratory of Patrick D. Schloss [38], according to the manufacturer's instructions with modifications found in the Schloss standard operating procedures (SOP) [41] as published previously [42, 43]. For primary PCR, each 20 μL PCR reaction contained the following: 5 μL of a 4 μM equimolar primer set, 2 μL 10X AccuPrime PCR Buffer II (Life Technologies, catalog no. 12346094), 9.85 μL sterile PCR-grade water, 0.15 μL Accuprime High Fidelity Taq Polymerase (Life Technologies catalog no. 12346094), and 3 μL of template DNA. PCR cycling conditions were 95°C for 2 min, followed by 20 cycles of touchdown PCR (95°C 20 s, 60°C 15 s and decreasing 0.3° each cycle, 72°C 5 min), then 20 cycles of standard PCR (95°C for 20 s, 55°C for 15 s, and 72°C for 5 min), and finished with 72°C for 10 min. PCR products were visualized using an E-Gel 96 with SYBR Safe DNA Gel Stain, 2% (Life Technologies catalog no. G7208-02). Samples that did not amplify during the first round of PCR were reamplified during a single round of troubleshooting using the same PCR cycling protocol and reaction

composition as described above, except for the following modifications: increasing template DNA volume from 3 to 5 μL and decreasing the sterile PCR-grade water volume by 2 μL to yield a total reaction volume of 20 μL .

After confirming successful amplification of all samples, libraries were normalized using SequalPrep Normalization Plate Kit (Life Technologies, catalog no. A10510-01) following the manufacturer's protocol for sequential elution. The concentration of the pooled samples was determined using Kapa Biosystems Library Quantification kit for Illumina platforms (Kapa Biosystems, catalog no. KK4824), and amplicon size was determined using the Agilent Bioanalyzer High Sensitivity DNA analysis kit (catalog no. 5067-4626). Libraries were prepared according to Illumina's "Preparing Libraries for Sequencing on the MiSeq" protocol for 2 nM libraries (part no. 15039740 Rev. D). The final library consisted of equimolar amounts from each of the plates normalized to the pooled plate at the lowest concentration. The final load concentration was 5 pM, spiked with PhiX at 15% to add diversity. Sequencing reagents were prepared according to the Schloss SOP, and custom read 1, read 2, and index primers were added to the reagent cartridge. Amplicons were sequenced using the Illumina MiSeq platform (San Diego, CA) using a MiSeq Reagent Kit V2 (Illumina, catalog no. MS102-2003) for 500 cycles. Sterile water ($n=8$) and empty wells ($n=28$) were sequenced as negative controls, and a synthetic community ($n=4$; ZymoBIOMICS Microbial Community DNA Standard, Zymo Research catalog no. D6306) was sequenced as a positive control. FASTQ files were generated with paired end reads and retained for further analysis.

Adequacy of sequencing

The full dataset obtained from the sequencing run included 5,560,120 total reads, with a mean \pm SD of 46,334 \pm 63,233 reads per specimen and a range of 53–287,832 reads per specimen. The dataset post-*mothur* processing included 2,062,759 identified bacterial reads, with a mean \pm SD of 17,190 \pm 23,515 reads per specimen and a range of 12–113,592 reads per specimen. One whole lung tissue sample, one isolation control, one elution buffer control, two water controls, and four empty well controls were identified as having an insufficient number of sequencing reads (< 150 reads) and were excluded from further analysis. All other tissue and control samples met the minimum requirement for number of sequencing reads (≥ 150 reads) and were retained for further analysis (Supplementary Figure 1). No major differences between the conclusions drawn from hypothesis tests conducted with the full sequencing dataset compared to the trimmed, quality-checked sequencing dataset were observed (Supplementary Table 2).

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 [41, 44]. A shared community file and a genus-level phylotyping file were generated using operational taxonomic units (OTUs) binned at 97% identity, using SILVA (v. 132) for sequence alignment (*silva.nr_v132.regionV4.align*). OTU numbers were arbitrarily assigned in the binning process and are referred to throughout the manuscript in association with their most specified level of taxonomy (typically genus or family). OTUs were classified using the *mothur* implementation of the Ribosomal Database Project (RDP) classifier and RDP taxonomy training set 16 (*trainset16_022016.rdp.fasta*, *trainset16_022016.rdp.tax*), available on the *mothur* website [41]. After data processing with *mothur*, shared community (OTU) and taxonomy files were imported to R for trimming and quality checks. OTUs that composed greater than 0.1% of reads in all samples were retained in the trimmed dataset for further analysis. One sample (WVB_Lung_L3) yielded less than 150 reads and was removed from the quality-checked dataset; all other experimental and control samples were retained in the quality-checked dataset, which was used for the main analysis.

Microbial community analysis of the quality-checked dataset was performed in R (version 4.0.2) [45] and relied primarily on the *tidyverse* (v. 1.3.0) [46], *ggplot2* (v. 3.3.0) [47], *vegan* (v. 2.5-6) [48], and *cmbtools* (v. 0.0.09025) [49] packages. For relative abundance, samples were normalized to the percent of total reads, and analysis was restricted to OTUs that were present at greater than 0.1% of the sample population. No OTUs were excluded from the dataset to account for background contamination. Diversity comparisons were performed by calculating community richness rarefied to 100 reads per sample, Shannon diversity index, and the Bray-Curtis dissimilarity index. Ordinations were performed using principal component analysis on Hellinger-transformed OTU count tables generated using Euclidean distances [50].

Overall significance was determined as appropriate by the Kruskal-Wallis test and by permutational multivariate analysis of variance (PERMANOVA) with 10,000 permutations using Hellinger-transformed OTU count tables and Euclidean distances with the *adonis* function in the R package *vegan*. Pairwise significance was determined as appropriate by the Wilcoxon test with the Benjamini-Hochberg correction for multiple comparisons, Tukey's honest significant difference (HSD) test, two-sample independent Mann-Whitney *U* test, and

two-group PERMANOVA as described for overall significance testing. All statistical tests used $p=0.05$ as a threshold for significance.

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 (Fig. 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 Fig. 2, BAL fluid and whole lung tissue both contained a significantly greater quantity of bacterial DNA than the isolation control ($p=0.008$ and 0.003 , 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 ($p=0.00005$). Whole lung tissue contained 27-fold more 16S rRNA gene copies than BAL fluid (42,740 vs. 1578 mean copies/specimen, 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. Despite the increased amount of bacterial DNA in whole lung tissue, BAL fluid returned a greater number of reads than whole lung tissue. These results are consistent with the compositional, rather than quantitative, nature of amplicon-based sequencing, and are likely

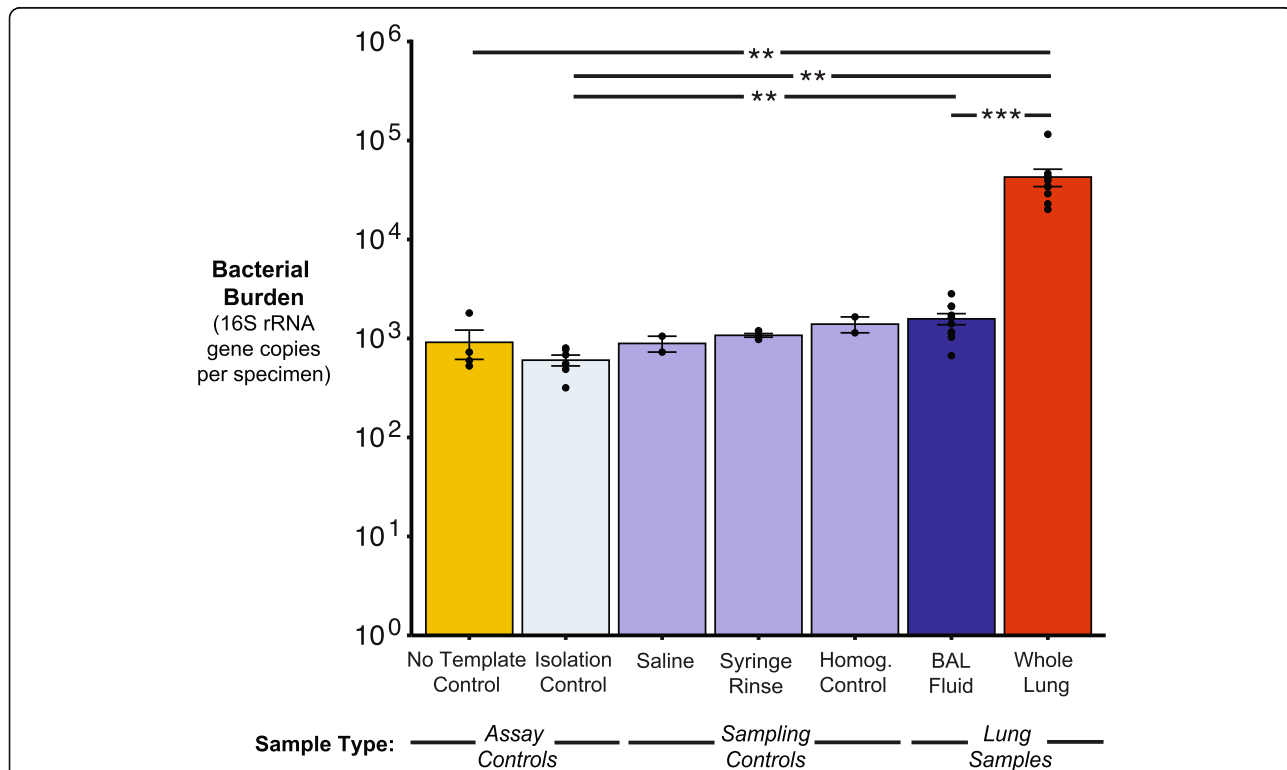


Fig. 2 Murine whole lung tissue contains increased bacterial burden relative to BAL fluid and negative controls. Whole lung tissue contains more copies of the bacterial 16S rRNA gene per mL of DNA isolated from lung or control specimens as quantified by ddPCR. Mean ± SEM and individual data points (representing the average of technical duplicates) are shown. Overall significance was determined by the Kruskal-Wallis test ($p = 0.00005$). 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: * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$

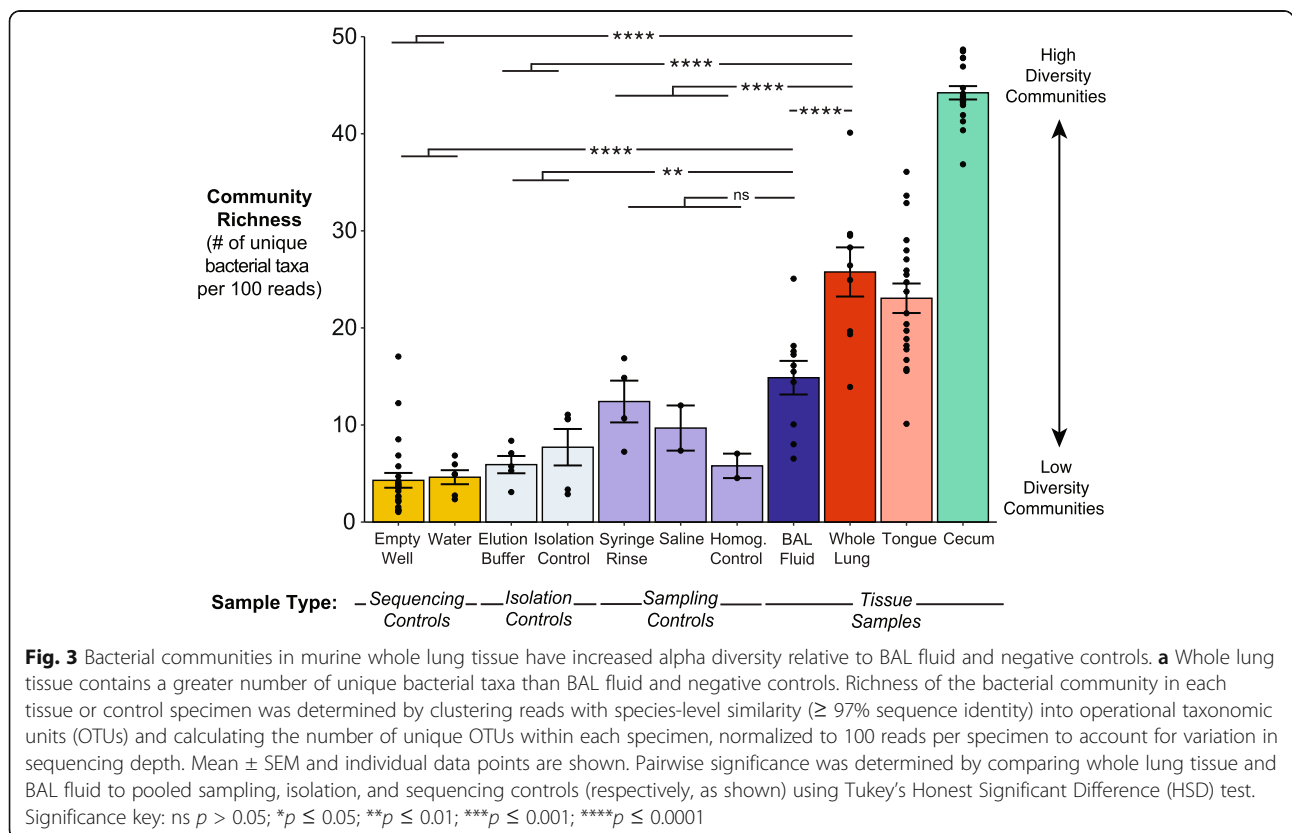
attributable to the high host-to-bacteria ratio in whole lung tissue (Supplementary Figure 1). With the exception of one whole lung tissue sample and five negative controls, all other specimens returned an adequate number of reads for further analysis. More details regarding adequacy of sequencing are provided in the “Methods” section.

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 (Fig. 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) per 100 reads present in each specimen and negative control. As predicted, whole lung tissue had greater community richness than BAL fluid ($p=0.00002$) and sampling, isolation, and sequencing controls ($p<0.0001$ for all comparisons) (Fig. 3). In contrast, whole lung and BAL specimens did not significantly differ in Shannon diversity index, which reflects both community richness and evenness ($p>0.05$; 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/or sequencing stochasticity [51], which both lead to decreased specimen-to-specimen reproducibility of cohoused mice. We have previously shown that mice from the same shipment and vendor have similar lung microbiota [6], and thus made this assumption in the following comparison. We 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 ($p<0.0001$) (Fig. 4). In contrast, the average Bray-Curtis dissimilarity index for BAL fluid was not significantly different than the highly dissimilar empty well controls ($p=0.3$). The high dissimilarity among BAL fluid samples is likely attributable to the low amount of bacterial DNA present in this sample type, as previous



work from our group showed an increase in sequencing stochasticity as bacterial biomass of the sample decreases [51]. Notably, the average Bray-Curtis dissimilarity index for whole lung tissue still showed a large amount of variation (median = 0.72). Since technical replicates were not sequenced in this study, the degree to which sequencing stochasticity contributes to the high amount of variation observed among lung microbiota from either sample type remains unknown. However, this variation observed in whole lung tissue is at least partly attributable to the sampling of mice from four different cages and the innate microbiological variation that occurs when sampling a single timepoint in the lung, which experiences rapid turnover of microbiota due to constant immune surveillance [14]. By comparison, the average Bray-Curtis dissimilarity index for whole lung tissue was lower than that of oral samples, supporting the biological plausibility of the observed variation between whole lung tissue samples. Taken together, these results suggest that whole lung tissue displays decreased sample-to-sample variation and likely 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 (Fig. 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 Fig. 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 3AB). 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 (Fig. 1, step 7). We hypothesized that the taxonomic composition of whole lung tissue would

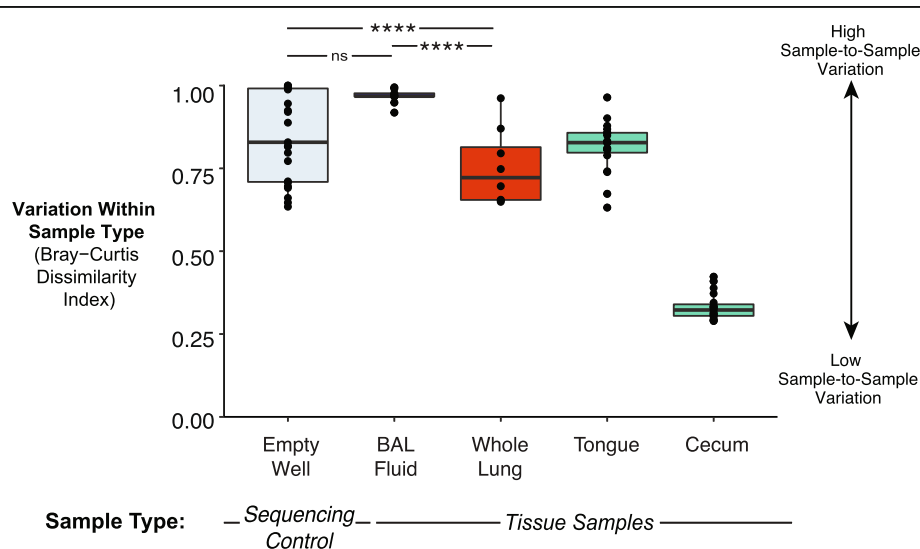


Fig. 4 Bacterial communities in murine whole lung tissue show decreased variation among biological replicates compared to those in BAL fluid. Variation among lung bacterial communities of healthy mice from the same shipment was quantified using the Bray-Curtis dissimilarity index. For comparison, Bray-Curtis 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 $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$

(See figure on previous page.)

Fig. 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 is similar to that of negative controls. Bars are ranked by mean abundance in whole lung tissue and represent mean \pm 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 $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$

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 (Fig. 5b). We confirmed these results by calculating the Bray-Curtis dissimilarity index for matched (i.e., from the same mouse) tongue and lung samples (Fig. 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 (Fig. 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 (Fig. 1, step 8).

Discussion

This study illustrates how an ecology-based analytical approach can interrogate 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 [52] and recent controversial low-biomass studies [4, 5, 53], which highlight the dangers of overinterpreting 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 [54]. Published murine lung microbiome studies to date have used both whole lung tissue [6, 36, 55–59] as well as BAL fluid [60–62], but no study to date has directly compared sampling 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 ($\sim 75 \text{ m}^2$ 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 [32, 33, 63]. 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

Table 1 Comparison of sampling methods for murine lung microbiome studies

		Whole lung tissue	Bronchoalveolar lavage fluid
Sample description	Sample content	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, and blood (if not perfused)	Airway and intra-alveolar space only
	Bacterial biomass	Low	Low
	Host-to-microbe DNA ratio	High	Low
Summary of findings (relative to each other)	Total DNA content	High	Low
	Bacterial DNA content	High	Low
	Variation among biological replicates	Low	High
	Similarity to contaminating source "communities" (negative controls)	Low	High
	Similarity to biological source community (oral microbiome)	High	Low

ecologic differences between humans and mice necessitate the use of murine-specific sampling approaches and illustrate 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 [64, 65]. 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 [51]. 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 microbiologic signal must be assessed within the specific ecologic context from which it is sampled, and anchored in an understanding of microbial ecology.

Several approaches to false signal in low-biomass microbiome studies have been proposed. Strategies used

to detect, interpret, and, in some cases, eliminate contamination have included exclusion of taxa detected in negative controls through statistical packages [66, 67] or unbiased subtraction [7], extraction and sequencing technical replicates [51], calculation of abundance ratios [68], correlation analyses [69], hierarchical clustering [70], and building neutral models [71]. In this study, we implement an experimental and analytical approach grounded in principles of microbial ecology to discriminate true microbial signal from background-derived signal. Fundamentally, this approach relies on sampling the low-biomass body site of interest and comparing the size, diversity, and taxonomic composition of the microbial community identified at that low-biomass site to all potential source communities, including background signal derived from procedural, reagent, and sequencing contamination and true microbial signal derived from contiguous body sites. This approach can thus be applied to a single specimen type to discern true bacterial signal from background-derived noise or to compare multiple specimen types to determine the optimal sampling method in the absence of a gold standard. We advocate for this approach to be applied to all low-biomass studies, as the choice of sampling method and resulting sample type should reflect that which yields the strongest biological signal. Of note, this approach does not preclude the use of complementary methods (such as those intended to handle contamination mentioned above), but rather builds a foundation rooted in thorough experimental design, which can then be subjected to further analysis with other bioinformatic tools.

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). While we implemented a strategy for lung lavage that would maximize volume returned, thereby maximizing bacterial signal from the lungs (and minimizing the likelihood of a false negative conclusion), the dilution of BAL fluid is a recognized source of variability in both humans and mice and may have impacted our results. 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 [72], and may contribute to the lower numbers of reads observed in whole lung tissue in our study. Given the impossibility of performing both BAL and whole lung homogenization on the same mouse, as lavaging the lungs before the collection of whole lung tissue would confound the comparison of the two specimen types, we could not perform paired analysis on the same mice. Based on prior results [6], we assumed that co-housed mice from the same vendor and shipment should have lung bacterial communities with similar taxonomic composition, but 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.

Conclusions

We here present evidence supporting the use of whole lung tissue over BAL fluid in murine lung microbiome studies. The use of an ecology-based experimental and 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; PBS: Phosphate-buffered saline; ddPCR: Droplet digital PCR; SEM: Standard error of the mean; HSD: Honest significant difference

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-021-01055-4>.

Additional file 1. Online Data Supplement

Acknowledgements

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

Authors' contributions

JMB participated in study design, sample collection, data acquisition, bioinformatic and statistical analysis, and data interpretation, and drafted and revised the manuscript. KJH, RAM, and NRF participated in sample collection and processing and data acquisition. CAB participated in bioinformatic and statistical analysis. 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.

Funding

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

Availability of data and materials

The dataset supporting the conclusions of this article is available in the NIH Sequence Read Archive (accession number: PRJNA644805) at <https://www.ncbi.nlm.nih.gov/bioproject/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>.

Declarations

Ethics approval and consent to participate

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. No human samples or subjects were used in this study.

Consent for publication

All authors read and approved this version of the manuscript for publication.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109, USA. ²Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan Health System, 6220 MSRB III/SPC 5642, 1150 W. Medical Center Dr, Ann Arbor, MI 48109-5642, USA. ³Department of Molecular, Cellular, & Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA. ⁴Mary H. Weiser Food Allergy Center, University of Michigan Medical School, Ann Arbor, MI 48109, USA. ⁵Michigan Center for Integrative Research in Critical Care, Ann Arbor, MI, USA.

Received: 17 December 2020 Accepted: 22 March 2021

Published online: 05 May 2021

References

- Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 2014;12(1):87. <https://doi.org/10.1186/s12915-014-0087-z>.
- Biesbroek G, Sanders EAM, Roeselers G, Wang X, Caspers MPM, Trzcinski K, et al. Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. *PLoS ONE.* 2012; 7(3):e32942. <https://doi.org/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. [https://doi.org/10.1016/S1473-099\(18\)30503-6](https://doi.org/10.1016/S1473-099(18)30503-6).
4. Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, et al. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome.* 2016;4(1):29. <https://doi.org/10.1186/s40168-016-0172-3>.
5. Leiby JS, McCormick K, Sherrill-Mix S, Clarke EL, Kessler LR, Taylor LJ, et al. Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. *Microbiome.* 2018;6(1):196. <https://doi.org/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. <https://doi.org/10.1164/rccm.201711-21800C>.
7. Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Chen H, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome.* 2013;1(1):19. <https://doi.org/10.1186/2049-2618-1-19>.
8. Segal LN, Clemente JC, Tsay J-CJ, Korolov SB, Keller BC, Wu BG, 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. <https://doi.org/10.1038/nmicrobiol.2016.31>.
9. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med.* 2011;184(8):957–63. <https://doi.org/10.1164/rccm.201104-0655OC>.
10. Charlson ES, Bittinger K, Chen J, Diamond JM, Li H, Collman RG, 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. <https://doi.org/10.1371/journal.pone.0042786>.
11. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Beck JM, Huffnagle GB, 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):821–30. <https://doi.org/10.1513/AnnalsATS.201501-029OC>.
12. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, et al. Bacterial topography of the healthy human lower respiratory tract. *mBio.* 2017;8(1):e02287–16. <https://doi.org/10.1128/mBio.02287-16>.
13. Sulaiman I, Wu BG, Li Y, Tsay J-C, Sauthoff M, Scott AS, et al. Functional lower airways genomic profiling of the microbiome to capture active microbial metabolism. *Eur Respir J.* 2021;2003434:2003434. <https://doi.org/10.1183/13993003.03434-2020>.
14. Wu BG, Sulaiman I, Tsay J-CJ, Perez L, Franca B, Li Y, et al. Episodic aspiration with oral commensals induces a MyD88-dependent, pulmonary Th17 response that mitigates susceptibility to *Streptococcus pneumoniae*. *Am J Respir Crit Care Med.* 2020;Epub ahead of print. <https://doi.org/10.1164/rccm.202005-1596OC>.
15. Mac Aogáin M, Baker JM, Dickson RP. On bugs and blowholes: why is aspiration the rule, not the exception? *Am J Respir Crit Care Med.* 2021; Epub ahead of print. <https://doi.org/10.1164/rccm.202011-4257ED>.
16. Dickson RP, Erb-Downward JR, Prescott HC, Martinez FJ, Curtis JL, Lama VN, et al. Intraalveolar catecholamines and the human lung microbiome. *Am J Respir Crit Care Med.* 2015;192(2):257–9. <https://doi.org/10.1164/rccm.201502-0326LE>.
17. Wang J, Lesko M, Badri MH, Kapoor BC, Wu BG, Li Y, 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):00008–2017. <https://doi.org/10.1183/23120541.00008-2017>.
18. O'Dwyer DN, Ashley SL, Gurczynski SJ, Xia M, Wilke C, Falkowski NR, et al. Lung microbiota contribute to pulmonary inflammation and disease progression in pulmonary fibrosis. *Am J Respir Crit Care Med.* 2019;199(9):1127–38. <https://doi.org/10.1164/rccm.201809-1650OC>.
19. Huang YJ, Nariya S, Harris JM, Lynch SV, Choy DF, Arron JR, et al. The airway microbiome in patients with severe asthma: associations with disease features and severity. *J Allergy Clin Immunol.* 2015;136(4):874–84. <https://doi.org/10.1016/j.jaci.2015.05.044>.
20. Zhang Q, Cox M, Liang Z, Brinkmann F, Cardenas PA, Duff R, et al. Airway microbiota in severe asthma and relationship to asthma severity and phenotypes. *PLOS ONE.* 2016;11(4):e0152724. <https://doi.org/10.1371/journal.pone.0152724>.
21. Durack J, Lynch SV, Nariya S, Bhakta NR, Beigelman A, Castro M, 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. <https://doi.org/10.1016/j.jaci.2016.08.055>.
22. Goleva E, Jackson LP, Harris JK, Robertson CE, Sutherland ER, Hall CF, et al. The effects of airway microbiome on corticosteroid responsiveness in asthma. *Am J Respir Crit Care Med.* 2013;188(10):1193–201. <https://doi.org/10.1164/rccm.201304-0775OC>.
23. Leitao Filho FS, Alotaibi NM, Ngan D, Tam S, Yang J, Hollander Z, 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):1205–13. <https://doi.org/10.1164/rccm.201806-1135OC>.
24. Molyneux PL, Cox MJ, Willis-Owen SAG, Mallia P, Russell KE, Russell A-M, et al. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med.* 2014;190(8):906–13. <https://doi.org/10.1164/rccm.201403-0541OC>.
25. Han MK, Zhou Y, Murray S, Tayob N, Noth I, Lama VN, et al. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med.* 2014;2(7):548–56. [https://doi.org/10.1016/S2213-2600\(14\)70069-4](https://doi.org/10.1016/S2213-2600(14)70069-4).
26. Invernizzi R, Barnett J, Rawal B, Nair A, Ghai P, Kingston S, 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. <https://doi.org/10.1183/13993003.01519-2019>.
27. Dickson RP, Schultz MJ, van der Poll T, Schouten LR, Falkowski NR, Luth JE, et al. Lung microbiota predict clinical outcomes in critically ill patients. *Am J Respir Crit Care Med.* 2020;201(5):555–63. <https://doi.org/10.1164/rccm.201907-1487OC>.
28. Panzer AR, Lynch SV, Langelier C, Christie JD, McCauley K, Nelson M, 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):621–31. <https://doi.org/10.1164/rccm.201702-0441OC>.
29. Dickson RP, Erb-Downward JR, Freeman CM, Walker N, Scales BS, Beck JM, et al. Changes in the lung microbiome following lung transplantation include the emergence of two distinct *Pseudomonas* species with distinct clinical associations. *PLoS ONE.* 2014;9(5):e97214. <https://doi.org/10.1371/journal.pone.0097214>.
30. Combs MP, Wheeler DS, Luth J, Falkowski NR, Chanderraj R, Walker NM, et al. Increasing relative abundance of pseudomonads predicts chronic rejection after lung transplant. *J Heart Lung Transplant.* 2020;39(4):565. <https://doi.org/10.1016/j.healun.2020.01.1266>.
31. Carney SM, Clemente JC, Cox MJ, Dickson RP, Huang YJ, Kitsios GD, et al. Methods in lung microbiome research. *Am J Respir Cell Mol Biol.* 2020;62(3):283–99. <https://doi.org/10.1165/rmb.2019-0273TR>.
32. Schulte H, Mühlfeld C, Brandenberger C. Age-related structural and functional changes in the mouse lung. *Front Physiol.* 2019;10:1466. <https://doi.org/10.3389/fphys.2019.01466>.
33. Dickson RP, Cox MJ. Sampling the lung microbiome. In: Cox MJ, Ege MJ, von Mutius E, editors. *The lung microbiome*. Sheffield, United Kingdom: European Respiratory Society; 2019. p. 1–17. <https://doi.org/10.1183/2312508X.10015418>.
34. 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. <https://doi.org/10.1371/journal.pbio.1000412>.
35. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Altman DG, Avey MT, et al. Revision of the ARRIVE guidelines: rationale and scope. *BMJ Open Sci.* 2018;2(1):e000002. <https://doi.org/10.1136/bmjos-2018-000002>.
36. Ashley SL, Sjoding MW, Popova AP, Cui TX, Hoostal MJ, Schmidt TM, 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. <https://doi.org/10.1126/scitranslmed.aau9959>.
37. Mason KL, Erb Downward JR, Mason KD, Falkowski NR, Eaton KA, Kao JY, et al. *Candida albicans* and bacterial microbiota interactions in the cecum during recolonization following broad-spectrum antibiotic therapy. *Infect Immun.* 2012;80(10):3371–80. <https://doi.org/10.1128/IAI.00449-12>.
38. 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):5112–20. <https://doi.org/10.1128/AEM.01043-13>.
39. 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. <https://doi.org/10.1371/journal.pone.0110351>.
 40. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci*. 2011;108(Supplement 1):4516–22. <https://doi.org/10.1073/pnas.1000080107>.
 41. Schloss PD. MiSeq SOP - mothur. 2019. Available from: https://mothur.org/wiki/miseq_sop/.
 42. Koenigsnecht MJ, Theriot CM, Bergin IL, Schumacher CA, Schloss PD, Young VB. Dynamics and establishment of *Clostridium difficile* infection in the murine gastrointestinal tract. *Infect Immun*. 2015;83(3):934–41. <https://doi.org/10.1128/IAI.02768-14>.
 43. Seekatz AM, Theriot CM, Molloy CT, Wozniak KL, Bergin IL, Young VB. Fecal microbiota transplantation eliminates *Clostridium difficile* in a murine model of relapsing disease. *Infect Immun*. 2015;83(10):3838–46. <https://doi.org/10.1128/IAI.00459-15>.
 44. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75(23):7537–41. <https://doi.org/10.1128/AEM.01541-09>.
 45. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019. Available from: <https://www.R-project.org/>
 46. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. *J Open Source Softw*. 2019;4(43):1686. <https://doi.org/10.21105/joss.01686>.
 47. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016. Available from: <https://ggplot2.tidyverse.org>. <https://doi.org/10.1007/978-3-319-24277-4>.
 48. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, et al. Vegan: community ecology package (R package version 2.5-6). 2019. Available from: <https://CRAN.R-project.org/package=vegan>.
 49. cbmtools: General purpose package containing useful functions for microbiome analysis pipelines. Available from: <https://github.com/cb-42/cbmtools>.
 50. Legendre P, Gallagher ED. Ecologically meaningful transformations for ordination of species data. *Oecologia*. 2001;129(2):271–80. <https://doi.org/10.1007/s004420100716>.
 51. Erb-Downward JR, Falkowski NR, D'Souza JC, McCloskey LM, McDonald RA, Brown CA, et al. Critical relevance of stochastic effects on low-bacterial-biomass 16S rRNA gene analysis. *mBio*. 2020;11(3):e00258–20. <https://doi.org/10.1128/mBio.00258-20>.
 52. Schloss PD. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. *mBio*. 2018;9(3):e00525–18. <https://doi.org/10.1128/mBio.00525-18>.
 53. 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. <https://doi.org/10.1126/scitranslmed.3008599>.
 54. Barfod K, Roggenbuck M, Hansen L, Schjørring S, Larsen S, Sørensen S, et al. The murine lung microbiome in relation to the intestinal and vaginal bacterial communities. *BMC Microbiol*. 2013;13(1):303. <https://doi.org/10.1186/1471-2180-13-303>.
 55. Singh N, Vats A, Sharma A, Arora A, Kumar A. The development of lower respiratory tract microbiome in mice. *Microbiome*. 2017;5(1):61. <https://doi.org/10.1186/s40168-017-0277-3>.
 56. Yun Y, Srinivas G, Kuenzel S, Linnenbrink M, Alnahas S, Bruce KD, et al. Environmentally determined differences in the murine lung microbiota and their relation to alveolar architecture. *PLoS ONE*. 2014;9(12):e113466. <https://doi.org/10.1371/journal.pone.0113466>.
 57. Richmond BW, Brucker RM, Han W, Du R-H, Zhang Y, Cheng D-S, et al. Airway bacteria drive a progressive COPD-like phenotype in mice with polymeric immunoglobulin receptor deficiency. *Nat Commun*. 2016;7(1):11240. <https://doi.org/10.1038/ncomms11240>.
 58. Gollwitzer ES, Saglani S, Trompette A, Yadava K, Sherburn R, McCoy KD, et al. Lung microbiota promotes tolerance to allergens in neonates via PD-L1. *Nat Med*. 2014;20(6):642–7. <https://doi.org/10.1038/nm.3568>.
 59. Dickson RP, Singer BH, Newstead MW, Falkowski NR, Erb-Downward JR, Standiford TJ, et al. Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome. *Nat Microbiol*. 2016;1(10):16113. <https://doi.org/10.1038/nmicrobiol.2016.113>.
 60. Barfod KK, Vrankx K, Mirsepasi-Lauridsen HC, Hansen JS, Hougaard KS, Larsen ST, et al. The murine lung microbiome changes during lung inflammation and intranasal vancomycin treatment. *Open Microbiol J*. 2015;9(1):167–79. <https://doi.org/10.2174/1874285801509010167>.
 61. Poroyko V, Meng F, Meliton A, Afonyushkin T, Ulanov A, Semenyuk E, 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):L76–83. <https://doi.org/10.1152/ajplung.00061.2014>.
 62. Yadava K, Pattaroni C, Sichelstiel AK, Trompette A, Gollwitzer ES, Salami O, et al. Microbiota promotes chronic pulmonary inflammation by enhancing IL-17A and autoantibodies. *Am J Respir Crit Care Med*. 2016;193(9):975–87. <https://doi.org/10.1164/rccm.201504-0779OC>.
 63. Lutfi MF. The physiological basis and clinical significance of lung volume measurements. *Multidiscip Respir Med*. 2017;12(1):3. <https://doi.org/10.1186/s40248-017-0084-5>.
 64. 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. <https://doi.org/10.1128/mSystems.00186-19>.
 65. Sinha R, Stanley G, Gulati GS, Ezran C, Travaglini KJ, Wei E, et al. Index switching causes “spreading-of-signal” among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. *BioRxiv*. 2017;125724. <https://doi.org/10.1101/125724>.
 66. 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. <https://doi.org/10.1186/s40168-018-0605-2>.
 67. 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. <https://doi.org/10.1002/edn3.11>.
 68. 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. <https://doi.org/10.1186/s40168-016-0224-8>.
 69. Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J*. 2016;10(7):1669–81. <https://doi.org/10.1038/ismej.2015.235>.
 70. Bosch AATM, Levin E, van Houten MA, Hasrat R, Kalkman G, Biesbroek G, et al. Development of upper respiratory tract microbiota in infancy is affected by mode of delivery. *EBioMedicine*. 2016;9:336–45. <https://doi.org/10.1016/j.ebiom.2016.05.031>.
 71. Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, Huffnagle GB, et al. Application of a neutral community model to assess structuring of the human lung microbiome. *mBio*. 2015;6(1):e02284–e02214. doi: <https://doi.org/10.1128/mBio.02284-14>.
 72. Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn L-J, 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. <https://doi.org/10.3389/fmicb.2019.01277>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.