Pneumonia

How low can we go? The implications of low bacterial load in respiratory microbiota studies

Robyn L. Marsh[1*], Maria T. Nelson[2], Chris E. Pope[2], Amanda J. Leach[1], Lucas R. Hoffman[2], Anne B. Chang[1,3] and Heidi C. Smith-Vaughan[1]

Abstract

Background: Culture-independent sequencing methods are increasingly used to investigate the microbiota associated with human mucosal surfaces, including sites that have low bacterial load in healthy individuals (e.g. the lungs). Standard microbiota methods developed for analysis of high bacterial load specimens (e.g. stool) may require modification when bacterial load is low, as background contamination derived from sterile laboratory reagents and kits can dominate sequence data when few bacteria are present.

Main body: Bacterial load in respiratory specimens may vary depending on the specimen type, specimen volume, the anatomic site sampled and clinical parameters. This review discusses methodological issues inherent to analysis of low bacterial load specimens and recommends strategies for successful respiratory microbiota studies. The range of methods currently used to process DNA from low bacterial load specimens, and the strategies used to identify and exclude background contamination are also discussed.

Conclusion: Microbiota studies that include low bacterial load specimens require additional tests to ensure that background contamination does not bias the results or interpretation. Several methods are currently used to analyse the microbiota in low bacterial load respiratory specimens; however, there is scant literature comparing the effectiveness and biases of different methods. Further research is needed to define optimal methods for analysing the microbiota in low bacterial load specimens.
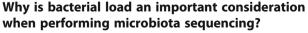
Keywords: Microbiota, 16S rRNA gene sequencing, Bacterial load, Low bacterial load, Background contamination

Background

The emergence of high throughput sequencing technologies over the past 10 years has enabled deep understanding of the microbiota associated with human mucosal surfaces. Much of this work has focused on the gut, as the digestive tract has the highest microbial density in the human body. As a result, current microbiota methods were largely optimized for analysis of stool specimens [1]. Application of these methods to other specimen types is not necessarily straightforward, as the bacterial load in stool is several orders of magnitude higher than that at other mucosal sites, including the respiratory tract [2, 3]. The potential for large variation in microbial density between different specimen types prompts consideration of how bacterial load may affect microbiota studies. This review considers important implications of low bacterial load in respiratory microbiota studies. The review highlights the importance of residual bacterial DNA as a source of background contamination in sterile laboratory reagents and kits, discusses how this contamination can dominate the microbiota profiles of low bacterial load specimens, and reviews current strategies for overcoming this limitation.

Why is bacterial load an important consideration when performing microbiota sequencing?

An inverse power relationship exists between the number of bacterial genomes used to prepare sequencing libraries and

* Correspondence: robyn.marsh@menzies.edu.au
[1]Child Health Division, Menzies School of Health Research, Darwin, Northern Territory, Australia
Full list of author information is available at the end of the article

the proportion of erroneous reads present in microbiota sequence data. This relationship was clearly demonstrated by studies that used microbiota sequencing to analyze pure bacterial cultures. An analysis of serially diluted *Salmonella bongori* culture found that the relative abundance of erroneous reads in the microbiota data progressively increased when $< 10^8$ *S. bongori* genome equivalents were sequenced; when $\leq 10^3$ genome equivalents were analysed, $> 90\%$ of resulting sequences were not *S. bongori* [4]. Similar findings were reported for pure *Staphylococcus aureus* and *Escherichia coli* cultures [5].

In vitro studies have identified residual 16S rDNA in sterile laboratory reagents and DNA extraction kits as the primary source of background contamination in microbiota data [4–6]. Taxa suggestive of environmental contamination have also been detected in respiratory specimens with low bacterial load, but not in the corresponding negative controls [7, 8]. In one study, environmental taxa with relative abundance inversely proportional to total bacterial load (strongly suggestive of background contamination) were detected exclusively in the clinical specimens [7]. It is currently unclear whether detection of such taxa indicates unsampled sources of background contamination or unidentified amplification and sequencing errors [9, 10].

To date, more than 200 genera have been reported as background contamination in microbiota data from negative controls and low bacterial load specimens (Table) [4, 8, 11]. This list includes genera previously reported amongst the microbiota of low bacterial load human mucosal and tissue specimens [12, 13]. The potential for misinterpretation of background contamination in data from low bacterial load specimens was recently highlighted in placental microbiota studies [14]. Early reports suggested the presence of a commensal placental microbiota with potential impacts for the developing infant microbiome [15, 16]; however, these early studies were limited, as tests for background contamination were not performed. This limitation was emphasized by a subsequent study that included robust negative controls and bacterial load measures and found that the placental microbiota of women with uncomplicated pregnancies could not be differentiated from background contamination present in negative controls [17]. This example highlights the importance of bacterial load measures and negative controls when studying the microbiota of environments or specimens with potentially low numbers of bacteria. Other examples of anatomic sites that are expected to have low bacterial load or be sterile in the absence of clinical infection include the middle ear [18], eyes [19], urinary system [13], central nervous system [20] and the lower airways [3].

Background contamination in respiratory microbiota studies

Several studies have reported background contamination in microbiota data generated from respiratory specimens with low bacterial load ($< 10^6$ bacterial cells/mL of specimen) [3, 4, 7, 8, 18, 21]. Factors that determine the bacterial load in respiratory specimens include the specimen type (e.g. sputum, bronchoalveolar lavage [BAL], swabs), the specimen volume, the anatomic site sampled (e.g. upper or lower airways) and clinical parameters (e.g. age, acute infection, exacerbation, antibiotic use).

Many respiratory microbiota studies have analyzed sputum from patients with lower airway disease. Spontaneously expectorated sputum has relatively high bacterial load (typically $> 10^8$ 16S copies per gram or mL sputum [22–24]), and thus, sputum microbiota data are expected to be less affected by background contamination than specimens containing lower bacterial densities (e.g. BAL specimens).

The bacterial load reported for BAL specimens varies depending on the volume tested and the disease state. BAL from healthy adults has low bacterial load ($< 10^6$ 16S copies/mL) [3, 25], and thus, concentration of cellular components from large BAL volumes (> 100 mL) may be needed prior to DNA extraction to avoid background contamination issues [26]. When lower BAL volumes (< 2 mL) [3] or acellular BAL [25] were tested, taxa consistent with background contamination were present in the microbiota data. BAL bacterial load may also vary in patients with lower airways disease. For example, high BAL bacterial load (up to 10^8 16S copies/mL) was reported for patients with cystic fibrosis [27] and interstitial pulmonary fibrosis [28], whereas low bacterial load was reported for BAL from children with protracted bacterial bronchitis or bronchiectasis [7].

Studies of oropharyngeal swabs have reported bacterial load at levels high enough to minimise adverse effects from background contaminants ($> 10^6$ 16S copies/mL) [3, 7, 27]; however, lower bacterial load (and therefore higher risk of bias due to detection of background contamination) was reported following antibiotic treatment [27]. Nasopharyngeal (NP) bacterial load is lower than that in the oropharynx, and varies depending upon the age [29] and disease status [30] of the study population. Several microbiota studies have reported low bacterial load in a proportion of NP specimens [18, 31, 32], with some studies reporting low bacterial load in up to a third of the NP swabs tested [7, 21].

What can be done to reduce background contamination effects in respiratory microbiota studies?

Given the potential for variation in the amount of bacteria present in clinical specimens, this review recommends that quantitative polymerase chain reaction (qPCR)-based measures of bacterial load [33] be included in respiratory microbiota studies: firstly, to aid investigation of background contamination issues [17];

and secondly, to allow critical review of published data. Here, we discuss key stages in the design and implementation of microbiota analyses that can be affected by bacterial load issues.

Study design and specimen collection

Establishing a standardized methodology for low bacterial load specimens requires consideration of differences in patient populations and specimen types. Many microbiota studies are designed to compare the microbiota present at high bacterial load in all study groups, e.g. gut microbiota studies. In respiratory studies, bacterial load can be more variable with low bacterial load specimens more common in some groups than others. For example, lower airway microbiota in healthy controls compared to that of patients with chronic lung disease [34], or bacterial load variations between groups defined by antibiotic exposure [27, 35], or respiratory sites [7, 31]. Bacterial load is also important because specimens containing few bacteria may be excluded from downstream analyses if insufficient amplicons or sequence reads are generated. This exclusion risks loss of statistical power by reducing the sample size or by introducing imbalances to the study design that may affect downstream analyses.

Optimal specimen type and volume should also be considered when planning respiratory microbiota studies. For example, in lower airway studies where sputum collection is not possible, protected bacterial brush specimens may yield higher bacterial load than BAL [3], and thus may help abrogate background contamination effects. In BAL studies, recovery of bacterial DNA can be improved by concentrating large sample volumes [26]; however, this may not be possible where lower saline volumes are instilled (e.g. paediatric studies [36]) or when limited BAL volume is available. Opportunities to increase the volume of upper airway specimens can be limited, especially when swabs are collected. Swab-based respiratory microbiota analyses are typically performed using < 0.5 mL of the preservative media [7, 21, 29, 37]. Increasing this sample volume (e.g. by using the entire swab [31]) will improve recovery of bacterial DNA; however, it is important to note that linear increments in the swab volume will not achieve the $\log_{10}$-scale increases in bacterial load that may be required to minimize the proportion of background contamination in microbiota data.

Quality control

Where low bacterial load specimens are unavoidable, care must be taken to avoid introduction of exogenous DNA during specimen collection and laboratory processing. This includes taking steps to avoid introduction of exogenous DNA originating from the airways of laboratory staff (e.g. processing of specimens in biosafety cabinets). Negative controls should be processed alongside clinical specimens through all analytic stages (including specimen collection and sequencing) to ensure all potential sources of background contamination are sampled [1, 4, 17, 38]. Relevant negative controls include those that sample any equipment or media used during specimen collection. For BAL studies, bronchoscope saline wash controls should be collected prior to insertion into the airways, as these controls can be critical to the appropriate interpretation of microbiota data when bacterial load is low [3, 31] or when it overlaps with that of negative controls [25]. Negative controls should be included through all stages of laboratory processing, including sequencing [4, 17]. Where possible, nucleic acid-free laboratory reagents should be used. It is also recommended that laboratory processing be done in clean environments (e.g. decontaminated and UV-treated hoods) using dedicated reagents and pipettes to minimise contamination [1, 4]. Minimizing the number of liquid handling events during DNA extraction may help to reduce introduction of exogenous DNA and may also improve yield by preventing DNA loss during processing. Given the potential risks for confounding of microbiota data by background contamination, there is growing recognition of the need for negative control results and bacterial load measures to be included when publishing microbiota data from low bacterial load specimens [17, 38].

Optimizing DNA extraction for low bacterial load specimens

A range of DNA extraction methods have been used to prepare respiratory specimens for microbiota analysis; however, few studies have compared the bacterial DNA yield achieved by different methods. Biesbroek et al. [21] reported a 100-fold difference in bacterial DNA yield between extraction methods when applied to both high and low bacterial load upper airway specimens. This study reported a magnetic bead-based DNA purification method outperformed commonly used silica columns, including the MoBio Powersoil DNA Isolation kit (Qiagen Pty Ltd., Dandenong, Australia), which is widely used and has been recommended by large international sequencing projects (e.g. the United States National Institutes of Health Human Microbiome Project [39]).

Optimal DNA extraction pre-treatment conditions are not well defined for low bacterial load specimens. Mechanical disruption by bead-beating is widely accepted to improve bacterial coverage by lysing cells that resist chemical or enzymatic processes [21, 40]; however, some studies have reported reduced DNA yield following mechanical disruption [41, 42], potentially due to DNA shearing. Bead-beating has been combined with chemical and/or enzymatic lysis; however, optimal pre-treatment conditions have not been determined or standardized. Some studies have included bead-beating with phenol-chloroform

extraction methods [21, 43]; however, these methods can be difficult to perform (risking loss of specimen volume), include toxic chemicals and are not easily automated. Other studies have combined bead-beating and enzymatic digestion to improve bacterial coverage [29, 44]; however, there is little standardization in either the types of enzymes used or whether the enzymatic digestion is performed before or after bead-beating. Further research is needed to inform optimal pre-treatment strategies for maximum recovery and coverage of bacterial DNA from low load respiratory specimens.

Variation in background contamination among batches of DNA extraction kits or reagents must also be considered when studying the microbiota in low bacterial load specimens. Salter et al. [4] reported batch-specific variation in background contamination from DNA extraction kits used to longitudinally assess NP microbiota in children. In this study, NP swabs were processed in chronological order, with different batches of the DNA extraction kit used for specimens from different time-points. Initial analyses suggested age-specific changes in the NP microbiota; however, these associations were not detected once the batch-specific background contamination was removed [4]. To prevent this type of confounding, it is recommended that a single batch of DNA extraction reagents is used when analyzing the microbiota of low bacterial load specimens [4]. Where use of multiple batches of extraction reagents cannot be avoided, it is recommended that specimens are processed in a random order with concurrent processing of batch-specific negative controls [4].

Optimizing 16S rDNA amplification from low bacterial load specimens

Even with optimized, high-performing DNA extraction methods, there may be instances where bacterial load is low. For example, diminished loads may result from testing low volume specimens from healthy lower airways. In these instances, optimization of the 16S rDNA PCR may be needed to achieve sufficient amplicons for sequencing.

Many different strategies have been used to increase amplicon yield from low bacterial load respiratory specimens. Some studies have improved the amplicon yield by using > 30 PCR cycles [21]; however, this approach is not recommended as it can increase amplification errors [45, 46] as well as increasing detection of background contamination [4]. Some respiratory studies [31, 47] have used touchdown-PCR in which the annealing temperature is varied to improve amplicon specificity [48]; however, this method may not improve amplicon yield. Several nested PCRs have also been used in studies where a single round of PCR did not generate sufficient amplicons [49–52]. Nested PCR is expected to improve

amplicon yield and also to reduce heteroduplex and chimera formation [9], but requires additional liquid handling steps which risks further introduction of exogenous DNA. Whole specimen amplification techniques (e.g. multiple displacement amplification) may also be used to increase the amount of DNA available for sequencing studies [53, 54]; however, this method may increase the overall sequencing error rate.

As with DNA extraction methods, there is currently scant literature comparing the strengths and weaknesses of different approaches for increasing amplicon yield, including assessment of the impacts of different amplification strategies on overall sequence error rates. It is also important to recognize that while optimized amplification methods may improve amplicon yield, background contamination issues are likely to remain as exogenous DNA present in sampling equipment and laboratory reagents will be co-amplified alongside the target DNA.

Optimizing library preparation

16S sequencing libraries are routinely prepared using a standardized input mass of PCR amplicons. The Earth Microbiome Project (www.earthmicrobiome.org) currently recommends library preparation with 240 ng of amplicons from each sample [55]; however, this amount may be difficult to generate when applying standard protocols to respiratory specimens, especially when the bacterial load is low [4]. To overcome this limitation, some studies have excluded specimens generating too few amplicons when preparing sequencing libraries [56, 57]. Others have successfully used a lower mass of amplicons to prepare the library (25–100 ng) [5, 58]. A third approach has been to pool multiple PCR replicates of low-load specimens to ensure a minimum amplicon mass is achieved. This approach was successfully used by Salter et al. [8] to characterize the microbiota in NP swabs from infants; however, up to 18 replicates were needed for some specimens.

Identifying and excluding background contamination

As discussed above, microbiota studies that include specimens with low bacterial load need to test for and exclude background contamination prior to data analysis. Failure to identify and exclude background contamination risks confounding of the microbiota analysis and can result in incorrect conclusions, as recently demonstrated by studies of the placental [14, 17] and nasopharyngeal [4] microbiota.

The first step in assessing background contamination is to ensure that negative controls are included through all stages of specimen collection and laboratory processing [4, 17, 38]. Although negative controls and low bacterial load specimens may generate low amplicon yields, it is recommended that these specimens not be excluded

from sequencing libraries to ensure that background contamination is fully sampled [17, 38]. It is also recommended that microbiota studies report results of negative controls and bacterial load data, especially when specimens with low bacterial load are tested [17, 38].

The second step is to review sequence data from negative controls and clinical specimens to identify and exclude background contamination. Several studies have removed background contamination by excluding all operational taxonomic units (OTUs) identified in negative controls [47, 52, 59, 60]; however, this approach risks exclusion of biologically significant OTUs that may be present in negative controls as a result of barcoding errors [5, 8, 11, 61, 62]. For example, a *Haemophilus* read from a clinical specimen may have a barcode error that results in it being incorrectly assigned to a negative control; exclusion of all OTUs detected in negative controls would exclude this taxon from downstream analyses. This is an important point, as OTUs consistent with constituents of the human respiratory microbiota have been detected in negative controls (e.g. *Streptococcus, Haemophilus, Pseudomonas* and *Burkholderia*, Table 1). Thus, summary exclusion of all OTUs detected in negative controls may impact the accuracy of microbiota data from the clinical specimens.

Many different strategies have been used to overcome this limitation, each of which has different strengths and weaknesses (Table 2). Salter et al. [8] identified background contaminants by conservatively assessing OTUs detected in negative controls to avoid exclusion of biologically relevant taxa, as well as reviewing data from replicate DNA extractions and different DNA extraction batches. Lazarevic et al. [63] sequenced serial dilutions of pure bacterial cultures and bioinformatically tested for background contamination by determining the ratio of mean OTU abundance in negative controls compared to data from the culture specimens, then excluding OTUs with ratios > 0.001 as probable contaminants. Jervis-Bardy et al. [56] excluded OTUs detected in negative controls or clinical specimens when a strong negative correlation was observed between the relative abundance and the bacterial load. Bosch et al. [32] used complete linkage hierarchical clustering to identify and exclude clinical specimens with microbiota profiles indistinguishable from negative controls. Venkataraman et al. [64] used a neutral model to identify probable background contamination reads in microbiota data from respiratory specimens with low bacterial load; reads that were identified by the neutral model as likely to have originated from background contamination were excluded prior to OTU clustering.

The authors are unaware of any studies to date that have compared the strengths and weaknesses of these different methods. In the absence of such comparisons,

the authors currently recommend using one of the unsupervised methods that exclude contaminant reads or OTUs (Table 2). Cluster analysis may be appropriate; however, this method requires careful consideration of the impact of specimen exclusion on the study design and may be problematic if negative control data do not form a distinct cluster. Inclusion of mock community controls at different bacterial loads to enable assessment of the selected contaminant exclusion strategy is recommended. It is not recommended to exclude all OTUs detected in negative controls, as this risks exclusion of potentially biologically relevant OTUs present in negative control data because of sequencing errors (e.g. barcoding errors).

Current recommendations

Salter et al. [4] and Lauder et al. [17] systematically assessed background contamination in microbiota data from low bacterial load specimens and made a series of recommendations that help define current best practice. In combination, these recommendations are as follows:

1. Negative controls should be processed alongside clinical specimens through all stages of laboratory testing (including specimen collection and sequencing) to ensure that all potential sources of exogenous DNA are sampled.
2. Where high volume specimens are collected (e.g. BAL from adult patients), the specimen should be concentrated prior to DNA extraction to maximise recovery of bacterial DNA.
3. Use of a single batch of DNA extraction reagents. If this is not possible, specimens should be processed in a random order to reduce the risk of false pattern formation, with negative controls processed with each batch.
4. Where possible, include replicate DNA extractions and sequencing.
5. Bacterial load should be determined prior to sequencing to determine whether specimens with low bacterial load are present.
6. Negative controls should be sequenced to maximise sampling of background contamination.
7. Testing to identify and exclude taxa suggestive of background contamination should be done prior to downstream analyses of microbiota from clinical specimens.

This review recommends that these parameters are reported in a standardized manner when publishing microbiota data. It is also recommended that the following parameters be detailed when publishing data from specimens with potentially low bacterial load:

Table 1 Examples of genera detected previously as background contamination in negative controls from respiratory microbiota studies [4, 8, 11, 52]

Abiotrophia[a]	Bifidobacterium	Curvibacter[a]	Halocella	Microlunatus	Pseudoclavibacter	Staphylococcus
Acidocella	Blastococcus	Deinococcus	Herbaspirillum[a]	Moryella	Pseudoflavonifractor	Stenotrophomonas[a]
Acidovorax[a]	Blautia	Delftia	Hoeflea	Neisseria	Pseudomonas[a]	Streptococcus[a]
Acinetobacter[a]	Bosea	Devosia	Hydrotalea[a]	Nevskia	Pseudorhodoferax	Sulfuritalea[a]
Actionmyces	Bradyrhizobium[a]	Dialister	Ilumatobacter	Niastella	Pseudoxanthomonas[a]	Suttonella
Adhaeribacter	Brevibacillus[a]	Dietzia[a]	Intestinibacter	Nocardioides	Psychrobacter	Tericibacter
Aerococcus	Brevibacterium	Dorea	Janibacter	Novosphingobium	Quinella	Thauera
Aeromicrobium[a]	Brevundimonas[a]	Duganella	Janthinobacterium[a]	Ochrobactrum[a]	Ralstonia[a]	tm7
Afipia[a]	Brochothrix	Dyadobacter	Johnsonella	Olivibacter[a]	Rhizobium[a]	Treponema
Agrococcus	Brooklawnia	Dyella	Kingella[a]	Oscillospira	Rhodococcus[a]	Trichococcus
Alcaligenes	Burkholderia[a]	Eggerthella	Klebsiella	Oxalobacter[a]	Roseateles	Tsukamurella
Alcanivorax	Cadidatus pelagibacter	Elusimicrobium	Kocuria	Paenibacillus[a]	Roseburia	Tumebacillus
Alicycliphilus	Cadidatus_alysiospaera	Enhydrobacter[a]	Lachnoanaerobaculum	Papiibacter	Roseomonas[a]	Tyzzerella
Alistipes	Campylobacter	Enterobacter	Lactobacillus	Parabacteroides	Rothia	Unclassified Gp 2 Acidobacteria
Aminobacter	Capnocytophagia	Enterococcus	Lactococcus	Paracoccus[a]	Rubellimicrobium	Undibacterium[a]
Anaerococcus	Cardiobacterium[a]	Erysipelatoclostridium	Lanoclostidium	Parasutterella	Ruegeria	Vampirovibrio
Anaeromyxobacter	Catenibacterium	Escherichia_Shigella[a]	Lanospira	Parvimonas	Ruminococcus	Variovorax[a]
Anaerotrucus	Caulobacter[a]	Eubacterium	Leptothrix	Patulibacter	Runiniclostridium	Veillonella[a]
Anoxybacillus	Cellvibrio	Facklamia	Leptotrichia	Pedobacter[a]	Saccharopolyspora	Wautersiella
Aquabacterium[a]	Christensenella	Faecalibacterium	Limnobacter	Pedomicrobium	Salinicoccus	Xanthomonas
Arthrobacter[a]	Chryseobacterium[a]	Fastidiosipila	Massilia[a]	Pelomonas[a]	Schlegelella	Xlyanibacter
Asticcacaulis	Cloacibacterium	Flavobacterium	Megasphaera[a]	Peptococcus	Seleomonas	
Atopobium	Clostridium	Flavonifractor	Meiothermus	Petoniphilus	Serratia	
Aurantimonas[a]	Collinsella	Fusobacterium	Mesorhizobium[a]	Phyllobacterium[a]	Shewanella	
Azoarcus	Comamonas[a]	Gemella	Methylobacterium[a]	Pilimelia	Slackia	
Azospira	Coprococcus	Geobacillus	Methylophilus	Polaromonas	Solobacterium	
Bacillus[a]	Corynebacterium[a]	Geodermatophilus	Methylopila	Porphyromonas	Sphingobacterium[a]	
Bacteroides	Craurococcus	Gordonia	Methyloversatilis	Prevotella[a]	Sphingobium[a]	
Beijerinckia	Cupriavidus	Granulicatella[a]	Microbacterium	Propionibacterium[a]	Sphingomonas[a]	
Beutenbergia	Curtobacterium	Haemophilus	Micrococcus	Pseudobutyrivibrio	Sphingopyxis	

It is important that negative controls are included in all studies of the microbiota in low bacterial load specimens, as background contamination in sterile reagents and kits is likely to be batch-specific

[a]Indicates genera detected as contaminants in multiple studies

Table 2 Summary of methods used to identify and exclude background contamination in microbiota datasets

Method	How does it work?	Strengths and weaknesses
Exclusion method (e.g. [47, 52])	All OTUs detected in negative controls are excluded as background contaminants	Background contaminants will be excluded; however, the method is not recommended as it risks exclusion of biologically relevant OTUs that can occur in negative controls (e.g. because of barcoding errors) [5, 8, 11, 61, 62].
Replicate method [8]	Contaminant OTUs excluded based on replicate data from different DNA extraction batches and negative controls.	This method is not suitable if replicate extractions are not possible. Requires subjective interpretation of negative control data.
Abundance ratio [5]	The ratio of mean OTU abundance in negative controls and study specimens is calculated. OTUs with a ratio > 0.001 are excluded as probable background contaminants.	Unsupervised method developed using pure bacterial cultures. Optimal threshold for defining contamination in more complex microbiota data may need to be determined.
Correlation analysis [56]	Spearman correlation between OTU relative abundance and bacterial load is determined. OTUs showing a strong and significant negative correlation with bacterial load (Spearman rho < − 0.7) are excluded as probable contaminants.	Unsupervised method. Limited when applied to OTUs present in < 20 specimens. Significance testing requires adjustment for multiple measures.
Cluster analysis [32]	Hierarchical cluster analysis is used to identify and exclude specimens with high similarity to negative controls.	Unsupervised method. Exclusion of specimens (instead of reads or OTUs) may impact study design. Background contamination clusters may be difficult to define where similarity between negative controls and low bacterial load specimens is variable.
Neutral model [64]	Reads that are neutrally distributed and enriched in negative controls are excluded prior to OTU clustering. Additionally, only reads that are unique to clinical specimens are included in downstream analyses.	Unsupervised method. It is unclear whether restriction of downstream analyses to reads unique to clinical specimens may result in exclusion of biologically relevant taxa (e.g. because of barcoding errors).

1. The types of negative controls tested.
2. The methods used to collect and process negative controls.
3. Negative control results and bacterial load data.
4. The methods used to identify and exclude background contamination.
5. A list of taxa and/or OTUs excluded prior to downstream analyses.
6. Results of any downstream testing used to confirm the presence of biologically unexpected taxa (e.g. extended culture, species-specific PCR, fluorescent in situ hybridization [FISH]).

Summary

Modern DNA sequencing technologies provide previously unimagined capacity to interrogate the complex bacteriology present on human mucosal surfaces; however, it is increasingly recognized that standard methods require modification when applied to low bacterial load specimens. As bacterial load in respiratory specimens can be highly variable, it is critical that respiratory microbiota studies include robust negative controls and report bacterial load measures. Where low bacterial load specimens are analyzed, it is essential that background contamination is tested and accounted for and, if possible, excluded to avoid confounding. A range of methods are currently used to process DNA from low bacterial load specimens and to identify and exclude background contamination; however, there is scant literature comparing the effectiveness and biases of different methods. Priority areas for future research include studies to determine optimal methods for analysis of low bacterial load specimens, including specimen collection parameters, DNA extraction and PCR amplification conditions, and assessment of the different bioinformatic methods used to identify and exclude background contamination.

Authors' contributions
The manuscript was drafted by RLM with input from all authors. All authors read and approved the final manuscript.

Author details
[1]Child Health Division, Menzies School of Health Research, Darwin, Northern Territory, Australia. [2]Respiratory Medicine, Seattle Children's Hospital and University of Washington, Seattle, Washington, USA. [3]Department of Respiratory and Sleep Medicine, Children's Health Queensland and Queensland University of Technology, Brisbane, QLD, Australia.

References

1. Weiss S, Amir A, Hyde ER, Metcalf JL, Song SJ, Knight R. Tracking down the sources of experimental contamination in microbiome studies. Genome Biol. 2014;15:564.
2. Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria cells in the body. PLoS Biol. 2016;14:e1002533.
3. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, Bushman FD, Collman RG. Topographical continuity of bacterial populations in the healthy human respiratory tract. Am J Respir Crit Care Med. 2011;184:957–63.
4. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12:87.
5. Lazarevic V, Gaia N, Girard M, Schrenzel J. Decontamination of 16S rRNA gene amplicon sequence datasets based on bacterial load assessment by qPCR. BMC Microbiol. 2016;16:73.
6. Willner D, Daly J, Whiley D, Grimwood K, Wainwright CE, Hugenholtz P. Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. PLoS One. 2012;7:e34605.
7. Marsh RL, Kaestli M, Chang AB, Binks MJ, Pope CE, Hoffman LR, Smith-Vaughan HC. The microbiota in bronchoalveolar lavage from young children with chronic lung disease includes taxa present in both the oropharynx and nasopharynx. Microbiome. 2016;4:37.
8. Salter SJ, Turner C, Watthanaworawit W, de Goffau MC, Wagner J, Parkhill J, Bentley SD, Goldblatt D, Nosten F, Turner P. A longitudinal study of the infant nasopharyngeal microbiota: the effects of age, illness and antibiotic use in a cohort of South East Asian children. PLoS Negl Trop Dis. 2017;11: e0005975.
9. Thompson JR, Marcelino LA, Polz MF. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. Nucleic Acids Res. 2002;30:2083–8.
10. Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. Wrinkles in the rare biosphere: pyrosequencing errors lead to artificial inflation of diversity estimates. Environ Microbiol. 2009:118–23.
11. Glassing A, Dowd SE, Galanduik S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathog. 2016;8:24.
12. Huang Y, Yang B, Li W. Defining the normal core microbiome of conjunctival microbial communities. Clin Microbiol Infect. 2016;22:643.
13. Lewis DA, Brown R, Williams J, White P, Jacobson SK, Marchesi JR, Drake MJ. The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults. Front Cell Infect Microbiol. 2013;3:41.
14. Perez-Munoz ME, Arrieta MC, Ramer-Tait AE, Walter J. A critical assessment of the "sterile womb" and "*in utero* colonization" hypotheses: implications for research on the pioneer infant microbiome. Microbiome. 2017;5:48.
15. Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. Sci Transl Med. 2014;6:237ra65.
16. Collado MC, Rautava S, Aakko J, Isolauri E, Salminen S. Human gut colonisation may be initiated *in utero* by distinct microbial communities in the placenta and amniotic fluid. Sci Rep. 2016;6:23129.
17. Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, Leite R, Elovitz MA, Parry S, Bushman FD. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. Microbiome. 2016;4:29.
18. Jervis-Bardy J, Rogers GB, Morris PS, Smith-Vaughan HC, Nosworthy E, Leong LE, Smith RJ, Weyrich LS, De Haan J, Carney AS, Leach AJ, O'Leary S, Marsh RL. The microbiome of otitis media with effusion in indigenous Australian children. Int J Pediatr Otorhinolaryngol. 2015;79:1548–55.
19. Doan T, Akileswaran L, Andersen D, Johnson B, Ko N, Shrestha A, Shestopalov V, Lee CS, Lee AY, Van Gelder RN. Paucibacterial microbiome and resident DNA virome of the healthy conjunctiva. Invest Ophthalmol Vis Sci. 2016;57:5116–26.
20. Simon TD, Van Yserloo B, Nelson K, Gillespie D, Jensen R, McAllister JP 2nd, Riva-Cambrin J, Stockmann C, Daly JA, Blaschke AJ. Use of quantitative 16S rRNA PCR to determine bacterial load does not augment conventional cerebrospinal fluid (CSF) cultures among children undergoing treatment for CSF shunt infection. Diagn Microbiol Infect Dis. 2014;78:188–95.
21. Biesbroek G, Sanders EAM, Roeselers G, Wang X, Caspers MPM, Trzciński K, Bogaert D, Keijser BJF. Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. PLoS One. 2012;7:e32942.
22. Zemanick ET, Harris JK, Wagner BD, Robertson CE, Sagel SD, Stevens MJ, Accurso FJ, Laguna TA. Inflammation and airway microbiota during cystic fibrosis pulmonary exacerbations. PLoS One. 2013;8:e62917.
23. Cox MJ, Turek EM, Hennessy C, Mirza GK, James PL, Coleman M, Jones A, Wilson R, Bilton D, Cookson WO, Moffatt MF, Loebinger MR. Longitudinal assessment of sputum microbiome by sequencing of the 16S rRNA gene in non-cystic fibrosis bronchiectasis patients. PLoS One. 2017;12:e0170622.
24. Brill SE, Law M, El-Emir E, Allinson JP, James P, Maddox V, Donaldson GC, McHugh TD, Cookson WO, Moffatt MF, Nazareth I, Hurst JR, Calverley PM, Sweeting MJ, Wedzicha JA. Effects of different antibiotic classes on airway bacteria in stable COPD using culture and molecular techniques: a randomised controlled trial. Thorax. 2015;70:930–8.
25. Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Gao Z, Chen H, Berger KI, Goldring RM, Rom WN, Blaser MJ, Weiden MD. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. Microbiome. 2013;1:19.
26. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, Curtis JL. Bacterial topography of the healthy human lower respiratory tract. MBio. 2017;8:e02287–16.
27. Pittman JE, Wylie KM, Akers K, Storch GA, Hatch J, Quante J, Frayman KB, Clarke N, Davis M, Stick SM, Hall GL, Montgomery G, Ranganathan S, Davis SD, Ferkol TW, ARESTfC F. Association of antibiotics, airway microbiome, and inflammation in infants with cystic fibrosis. Ann Am Thorac Soc. 2017;14:1548–55.
28. Molyneaux PL, Cox MJ, Willis-Owen SA, Mallia P, Russell KE, Russell AM, Murphy E, Johnston SL, Schwartz DA, Wells AU, Cookson WO, Maher TM, Moffatt MF. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2014;190:906–13.
29. Stearns JC, Davidson CJ, McKeon S, Whelan FJ, Fontes ME, Schryvers AB, Bowdish DM, Kellner JD, Surette MG. Culture and molecular-based profiles show shifts in bacterial communities of the upper respiratory tract that occur with age. ISME J. 2015;9:1246–59.
30. Smith-Vaughan H, Byun R, Nadkarni M, Jacques NA, Hunter N, Halpin S, Morris PS, Leach AJ. Measuring nasal bacterial load and its association with otitis media. BMC Ear Nose Throat Disord. 2006;10:10.
31. Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, Beck JM, Curtis JL, Huffnagle GB. Analysis of the upper respiratory tract microbiota as the source of the lung and gastric microbiotas in healthy individuals. MBio. 2015;6:e00037.
32. Bosch AA, Levin E, van Houten MA, Hasrat R, Kalkman G, Biesbroek G, de Steenhuijsen Piters WA, de Groot PK, Pernet P, Keijser BJ, Sanders EA, Bogaert

D. Development of upper respiratory tract microbiota in infancy is affected by mode of delivery. EBioMedicine 2016;9:336–345.

33. Nadkarni MA, Martin FE, Jacques NA, Hunter N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. Microbiology. 2002;148:257–66.

34. Huang YJ, Nelson CE, Brodie EL, DeSantis TZ, Baek MS, Liu J, Woyke T, Allgaier M, Bristow J, Wiener-Kronish JP, Sutherland ER, King TS, Icitovic N, Martin RJ, Calhoun WJ, Castro M, Denlinger LC, DiMango E, Kraft M, Peters SP, Wasserman SI, Wechsler ME, Boushey HA, Lynch SV. Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. J Allergy Clin Immun. 2011;127:372–81.

35. Peleg AY, Choo J, Langan K, Edgeworth D, Keating D, Wilson J, Rogers GB, Kotsimbos T. Antibiotic exposure and interpersonal variance mask the effect of ivacaftor on respiratory microbiota composition. J Cyst Fibros. 2018;17:50–6.

36. de Blic J, Midulla F, Barbato A, Clement A, Dab I, Eber E, Green C, Grigg J, Kotecha S, Kurland G, Pohunek P, Ratjen F, Rossi G. Bronchoalveolar lavage in children. ERS Task Force on bronchoalveolar lavage in children European Respiratory Society. Eur Res J. 2000;15:217–31.

37. Bogaert D, Keijser B, Huse S, Rossen J, Veenhoven R, van Gils E, Bruin J, Montijn R, Bonten M, Sanders E. Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. PLoS One. 2011;6:e17035.

38. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, Lauder A, Sherrill-Mix S, Chehoud C, Kelsen J, Conrad M, Collman RG, Baldassano R, Bushman FD, Bittinger K. Optimizing methods and dodging pitfalls in microbiome research. Microbiome. 2017;5:52.

39. McInnes P, Cutting M. Manual of procedures for Human Microbiome Project. Core microbiome sampling protocol A. HMP protocol # 07–001. Version Number: 12.0. 2010.

40. Salonen A, Nikkila J, Jalanka-Tuovinen J, Immonen O, Rajilic-Stojanovic M, Kekkonen RA, Palva A, de Vos WM. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. J Microbiol Methods. 2010;81:127–34.

41. Gill C, van de Wijgert JH, Blow F, Darby AC. Evaluation of lysis methods for the extraction of bacterial DNA for analysis of the vaginal microbiota. PLoS One. 2016;11:e0163148.

42. Abusleme L, Hong BY, Dupuy AK, Strausbaugh LD, Diaz PI. Influence of DNA extraction on oral microbial profiles obtained via 16S rRNA gene sequencing. J Oral Microbiol. 2014;6

43. van der Gast CJ, Cuthbertson L, Rogers GB, Pope C, Marsh RL, Redding GJ, Bruce KD, Chang AB, Hoffman LR. Three clinically distinct chronic pediatric airway infections share a common core microbiota. Ann Am Thorac Soc. 2014;11:1039–48.

44. Zhao J, Carmody LA, Kalikin LM, Li J, Petrosino JF, Schloss PD, Young VB, LiPuma JJ. Impact of enhanced *Staphylococcus* DNA extraction on microbial community measures in cystic fibrosis sputum. PLoS One. 2012;7:e33127.

45. Jensen MA, Straus N. Effect of PCR conditions on the formation of heteroduplex and single-stranded DNA products in the amplification of bacterial ribosomal DNA spacer regions. PCR Methods Appl. 1993;3:186–94.

46. Lazarevic V, Whiteson K, Gaia N, Gizard Y, Hernandez D, Farinelli L, Osteras M, Francois P, Schrenzel J. Analysis of the salivary microbiome using culture-independent techniques. J Clin Bioinformatics. 2012;2:4.

47. Dickson RP, Erb-Downward JR, Freeman CM, Walker N, Scales BS, Beck JM, Martinez FJ, Curtis JL, Lama VN, Huffnagle GB. Changes in the lung microbiome following lung transplantation include the emergence of two distinct *Pseudomonas* species with distinct clinical associations. PLoS One. 2014;9:e97214.

48. Korbie DJ, Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nat Protoc. 2008;3:1452–6.

49. Kim RJ, Biswas K, Hoggard M, Taylor MW, Douglas RG. Paired analysis of the microbiota of surface mucus and whole-tissue specimens in patients with chronic rhinosinusitis. Int Forum Allergy Rhinol. 2015;5:877–83.

50. Liu HX, Tao LL, Zhang J, Zhu YG, Zheng Y, Liu D, Zhou M, Ke H, Shi MM, Qu JM. Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects. Int J Cancer. 2018;142:769–78.

51. Neeff M, Biswas K, Hoggard M, Taylor MW, Douglas R. Molecular microbiological profile of chronic suppurative otitis media. J Clin Microbiol. 2016;54:2538–46.

52. Sillanpaa S, Kramna L, Oikarinen S, Sipila M, Rautiainen M, Aittoniemi J, Laranne J, Hyoty H, Cinek O. Next-generation sequencing combined with specific PCR assays to determine the bacterial 16S rRNA gene profiles of middle ear fluid collected from children with acute otitis media. mSphere. 2017;2

53. Ahsanuddin S, Afshinnekoo E, Gandara J, Hakyemezoglu M, Bezdan D, Minot S, Greenfield N, Mason CE. Assessment of REPLI-g multiple displacement whole genome amplification (WGA) techniques for metagenomic applications. J Biomol Tech. 2017;28:46–55.

54. Huang L, Ma F, Chapman A, Lu S, Xie XS. Single-cell whole-genome amplification and sequencing: methodology and applications. Annu Rev Genomics Hum Genet. 2015;16:79–102.

55. The Earth Microbiome Project. [cited November 2017]; Available from: http://www.earthmicrobiome.org/protocols-and-standards/16s/.

56. Jervis-Bardy J, Leong LE, Marri S, Smith RJ, Choo JM, Smith-Vaughan HC, Nosworthy E, Morris PS, O'Leary S, Rogers GB, Marsh RL. Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data. Microbiome. 2015;3:19.

57. Mika M, Mack I, Korten I, Qi W, Aebi S, Frey U, Latzin P, Hilty M. Dynamics of the nasal microbiota in infancy: a prospective cohort study. J Allergy Clin Immunol. 2015;135:905–12. e11

58. Cope EK, Goldberg AN, Pletcher SD, Lynch SV. Compositionally and functionally distinct sinus microbiota in chronic rhinosinusitis patients have immunological and clinically divergent consequences. Microbiome. 2017;5:53.

59. Einarsson GG, Comer DM, McIlreavey L, Parkhill J, Ennis M, Tunney MM, Elborn JS. Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers. Thorax. 2016;71:795–803.

60. Cuthbertson L, Rogers GB, Walker AW, Oliver A, Green LE, Daniels TW, Carroll MP, Parkhill J, Bruce KD, van der Gast CJ. Respiratory microbiota resistance and resilience to pulmonary exacerbation and subsequent antimicrobial intervention. ISME J. 2016;10:1081–91.

61. Schnell IB, Bohmann K, Gilbert MT. Tag jumps illuminated–reducing sequence-to-sample misidentifications in metabarcoding studies. Mol Ecol Resour. 2015;15:1289–303.

62. Deakin CT, Deakin JJ, Ginn SL, Young P, Humphreys D, Suter CM, Alexander IE, Hallwirth CV. Impact of next-generation sequencing error on analysis of barcoded plasmid libraries of known complexity and sequence. Nucleic Acids Res. 2014;42:e129.

63. Lazarevic V, Gaia N, Emonet S, Girard M, Renzi G, Despres L, Wozniak H, Yugueros Marcos J, Veyrieras JB, Chatellier S, van Belkum A, Pugin J, Schrenzel J. Challenges in the culture-independent analysis of oral and respiratory samples from intubated patients. Front Cell Infect Microbiol. 2014;4:65.

64. Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, Huffnagle GB, Schmidt TM. Application of a neutral community model to assess structuring of the human lung microbiome. MBio. 2015;6:e02284-14.