Exploring the Variability of the Sheep Lung Microbiota

Laura Glendinning, Steven Wright, Jolinda Pollock, Peter Tennant, David Collie, Gerry McLachlan

The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, Midlothian, United Kingdom; Monogastric Science Research Centre, Scotland's Rural College (SRUC), Edinburgh, Midlothian, United Kingdom

#Address correspondence to Laura Glendinning, Laura.Glendinning@roslin.ed.ac.uk

Running Head: The Sheep Lung Microbiota
Abstract

Sequencing technologies have recently facilitated the characterisation of bacterial communities present in lungs during health and disease. However, there is currently a dearth of information concerning the variability of such data in health both between and within subjects. This study seeks to examine such variability using healthy adult sheep as our model system.

Protected specimen brush samples were collected from three spatially disparate segmental bronchi of six adult sheep (age 20 months) on three occasions (day 0, one month and three months). To further explore the spatial variability of the microbiota, more extensive brushings (n=16) and a throat swab were taken from a separate sheep. The V2-V3 hypervariable regions of the bacterial 16S rRNA gene were amplified and sequenced via Illumina MiSeq. DNA sequences were analysed using the MOTHUR software package. Quantitative PCR was performed to quantify total bacterial DNA.

Some sheep lungs contained dramatically different bacterial communities at different sampling sites whereas in others airway microbiota appeared similar across the lung. In our spatial variability study, clustering was observed related to the depth within the lung from which samples were taken. Lung depth refers to increasing distance from the glottis progressing in a caudal direction. We conclude that both host influence and local factors have an impact on the composition of the sheep lung microbiota.

Importance

Until recently, it was assumed that the lungs were a sterile environment which were only colonised by microbes during disease. However, recent studies using sequencing technologies...
have found that there is a small population of bacteria which exists in the lung during health, referred to as the ‘lung microbiota’. In this study we characterise the variability of the lung microbiota of healthy sheep. Sheep are not only economically important animals but are also often used as large animal models of human respiratory disease. We conclude that, whilst host influence does play a role in dictating the types of microbes which colonise the airways, it is clear that local factors also play an important role in this regard. Understanding the nature and influence of these factors will be key to understanding the variability in, and functional relevance of the lung microbiota.

Introduction

Within the past five years, a diverse array of bacteria has been detected in healthy lungs through the use of non-culture based methods (1, 2). These bacterial communities are commonly referred to as the lung microbiota and are thought to originate predominantly from the upper respiratory tract (3, 4). The presence of particular bacterial communities in the lung has been associated with several human diseases including cystic fibrosis (5), chronic obstructive pulmonary disease (6), bronchiectasis (7) and lung transplant rejection (8).

Whilst variation in the microbial communities present in the human lung exists at both a large and small scale, based upon the location of the bacteria within the lungs (9) and the host cell-types present (10), intra-individual variation has been found to be significantly less than inter-individual variation, indicating that each individual may play host to a specific lung microbiota (9).
The lung microbiota of healthy domestic sheep has previously been investigated using culture
based methods (11-14) but these studies have shown conflicting descriptions of the extent of
lung colonisation by bacteria. A study in pneumonic Bighorn sheep lungs found that for most
sheep studied, bacterial 16S rRNA gene amplification and sequencing was able to identify
additional bacterial species which were not found by culturing (15). Previous studies have also
examined the upper respiratory tract of healthy sheep by culture based methods (11, 12, 14, 16).
These studies are highly variable in the types and proportions of microbes identified.

Previously, our group studied the composition of the lung microbiota in sheep pre- and post-
infected with *Pseudomonas aeruginosa* (17). This study included the first description of the
lung microbiota communities of healthy domestic sheep by next generation sequencing. A
diverse community of microbes was identified and variability was seen to be high, both within
and between animals. The variability of the healthy lung microbiota at specific lung sites over
time has not been reported in any animal, although serial sampling of non-diseased human is
planned as part of the Lung HIV Microbiome Project (LHMP) (18).

In the present study, protected specimen brush samples were collected from three spatially
disparate segmental bronchi at three time-points (baseline, one month and three months) to
examine the composition and variability of the lung microbiota in healthy domestic sheep. In
addition, samples were also taken from a separate sheep from a greater number of respiratory
tract locations, to further explore the extent of spatial variability.

Such studies are fundamental to understanding the functional relevance of lung microbiota in
health and disease in ruminants. Indeed bacterial pneumonia is well recognised in cattle and
sheep and is often associated with high morbidity and mortality. Notably, regional predilection is
evident in that infection by *Pasteurella* occurs most frequently in the apical and cardiac lobes in

both sheep (12, 19) and cattle (20, 21). Co-infections with other respiratory pathogens are commonplace, it is already well known that infection by *Bordetella parapertussis* and *Mycoplasma ovipneumoniae* can lead to more severe disease caused by *Mannheimia (Pasteurella) haemolytica* (22-25), and there are well recognised links to stressful events such as housing or transport. As it is conceivable that changes in the lung microbiota may precipitate or associate with such events it is vital to ground future disease-related studies on a firm basis of understanding normal variation in health. Whilst the immediate focus of such studies relates to animal health it is also important to acknowledge that sheep are frequently used as models for human respiratory research (26, 27) and that there is an ongoing need to highlight any comparative contrasts and consistencies as and when they arise.

**Materials and methods**

**Animals and airway sampling**

Six twenty month old Suffolk-cross sheep were used in this study (Table 1) (5 females, 1 castrated male) and were housed indoors in pens for the trial duration. All animals had not undergone bronchoscopic examination during the four months preceding the study. Animal procedures were subject to the Animals (Scientific Procedures) Act 1986 and were approved by the Roslin Institute Animal Welfare and Ethics Committee.

Anaesthesia was performed as described previously (28). Sheep were sampled by protected specimen brushings (ConMed Disposable Microbiology Brush, New York, NY, USA) at 0 days (baseline), one month and three months. Sampling sites are shown in Fig 1. Bronchoscopy was
performed via an endotracheal tube by the same operator for all sheep at all time-points. The sample harvest dates can be found in Table S1. Before sampling of every sheep on any given day, 7.5 ml of phosphate buffered saline (PBS) was passed through the bronchoscope channel to act as an environmental qPCR control. Bronchoscope washings were centrifuged at 13,000 g for 15 minutes and the pellet was resuspended in 500 µl of PBS.

A throat swab and brushing samples (harvested as above) were also taken from a further sheep (female; age 36 months; 60 kg bodyweight) at a single time-point to further explore the spatial variability of the lung microbiota (sampling date 01/05/2015). Brushing sites were dorsal and ventral trachea and paired sites from either side of airway bifurcations progressing along the anterior to caudal lung axis (Fig 2).

DNA extraction, amplification and sequencing

DNA extraction was performed using the MOBIO PowerSoil® DNA Isolation Kit (Carlsbad, CA, USA). Brushes were transferred into PowerSoil Bead Tubes with PowerSoil Solution C1 and PowerSoil Bead Solution. Bead Tubes were heated at 65°C for 10 minutes then placed in a FastPrep FP120 Cell Disrupter (Qbiogene Inc., Cedex, France) for 45 seconds at 5.0 m/sec. From this point onwards the manufacturer’s instructions were followed, except for the final elution step. Purified DNA was eluted into 50 µl of PowerSoil Solution C6 rather than 100 µl to increase the DNA concentration.

All PCR steps used Q5® High-Fidelity 2X Master Mix (New England Biolabs, Beverly, MA, USA). A nested PCR reaction was performed with Illumina adaptor sequences and barcodes (Table S2) included only on the primers for the second round in an attempt to reduce bias.
caused by barcoded primers when amplifying low biomass samples (29). The conditions for the first round of PCR, amplifying the V1-V4 16S hypervariable regions (primers: 28F (5'–GAGTTTGTACNTGGCTCAG–3') and 805R (5'–GACTACCAGGGGTATCTAATC–3')), were: 94°C for 2 minutes followed by 20 cycles of 94°C for 1 minute, 55°C for 45 seconds and 72°C for 1.5 minutes followed by 72°C for 20 minutes. The conditions for the second round of PCR, amplifying the V2-V3 16S hypervariable regions (primers: 104F (5'–GGCGVACGGGTGAGTAA–3') and 519R (5'–GTNTTACNGCGGCKGCTG–3')), were: 98°C for 30 seconds followed by 20 cycles of 98°C for 10 seconds, 67°C for 30 seconds and 72°C for 10 seconds followed by 72°C for 2 minutes. Amplicons from both rounds of PCR were purified using the AMPure XP PCR Purification System (Beckman Coulter, Brea, CA, USA). Amplicons were sequenced using an Illumina MiSeq or HiSeq (Illumina, San Diego, CA) run producing paired end 250-nucleotide reads (30). Those samples sequenced by two MiSeq runs are listed in Dataset S1 and those sequenced by HiSeq are listed in Dataset S2. Where samples from the MiSeq runs were found to have low read numbers, they were sequenced again on a separate MiSeq run (Samples: 2D618 RA 3 months and 2D619 RA 3 months). We previously confirmed cross-run stability by comparing separate runs made on the same samples (Fig S1).

Extraction kit controls were produced by carrying out a reagent-only extraction using the MOBIO PowerSoil® DNA Isolation Kit. PCR reagent controls were constructed by adding 20 µl of nuclease free water to the PCR reaction mixture. The Human Microbiome Project Mock Community HM-782D (100,000 copies per organism per µl, BEI Resources, ATCC, Manassas, VA, USA) , extraction kit controls, PCR reagent-only controls and positive controls (DNA extracted from Pseudomonas aeruginosa Strain PA0579) were amplified and sequenced by the same methods as were used for samples.
A separate mock community sample was sequenced using an Illumina HiSeq. For this sample, the solution produced from the first round of PCR was diluted 1:100 in nuclease free water before being used in the second round of PCR. This was carried out to ascertain the effect of placing different concentrations of DNA into the second PCR round on PCR bias.

The unassembled reads, with primers removed, are publicly available through the NCBI Sequence Read Archive (SRA) under the Bioproject ascension: PRJNA298882.

Bioinformatic and statistical analysis

Primers were removed using Cutadapt (31). Sequences which contained more than one base error per 10 primer bases were removed from further analysis. The following steps were carried out in MOTHUR (32) and were based upon a protocol developed for MiSeq by the MOTHUR creators (30). Forward and reverse reads were aligned to form one continuous DNA sequence; any sequences which failed to align were discarded. Sequences which contained ambiguous bases; were less than 369 base pairs in length or contained homopolymers of greater than 9 base pairs were also discarded. Chimeras were identified and removed using UCHIME (33). Sequences were aligned to the SILVA reference alignment (34) and were classified using MOTHUR’s Bayesian classifier against the Greengenes database (35), which was trimmed to the V2-V3 hypervariable region of the 16S rRNA gene to improve classification depth (36). Sequences identified as not originating from bacteria were removed from further analysis. Operational taxonomic units (OTUs) were clustered into phylotypes using a database-dependent approach then sub-sampled.
Distance matrices were created using Yue and Clayton theta values (37). Analysis of molecular variance (AMOVA) (38) was used to determine significant differences between the bacterial compositions of groups. Principal co-ordinate analysis (PCOA) graphs were constructed to visualise similarities between samples. The Inverse Simpson’s index was used to quantify diversity. Where data was non-parametric the Friedman test was used to identify significant differences in diversity, using Minitab® 16 for Windows (Minitab, Coventry, UK). All other statistical tests were carried out within MOTHUR. Metastats (39) was used to identify OTUs which were different between groups. Good’s coverage (40) was used to estimate sample coverage and the Chao 1 index was used to calculate richness. Indicator OTUs (OTUs which are indicative of a particular group of samples) were identified using the indicator metric within MOTHUR (41). Repeated measures ANOVAs were carried out using the Vegan package in R (42-44).

**qPCR**

qPCR reactions were performed using the LightCycler® 480 SYBR Green I Master Mix (Roche Applied Science, Indianapolis, IN, USA), 1 µl of extracted DNA solution and the 16S rDNA qPCR primers UniF340 (5’–ACTCCTACGGGAGGCAGCAGT–3’) and UniR514 (5’–ATTACCGCGGCTGCTGGC–3’) at a final concentration of 0.4 µM.

The qPCR run consisted of a pre-incubation step of 50°C (ramp rate: 4.80°C/s for 2 minutes) then 95°C (ramp rate: 4.80°C/s for 10 seconds) and an amplification step consisting of 45 cycles of 95°C (ramp rate: 4.80°C/s for 30 seconds) then 63°C (ramp rate: 2.50°C/s for 30 seconds). This was followed by a melting cycle consisting of 95°C (ramp rate: 4.80°C/s for 5 seconds) then
65°C (ramp rate: 4.80°C/s for 1 minute) followed by 97°C (ramp rate: 0.11°C/s, acquisition mode, continuous).

Negative controls consisted of both water and extraction kit reagent controls. For water controls, 1 µl of nuclease free water was added to the qPCR reaction mixture. For extraction kit controls, DNA extractions were carried out using the MOBIO PowerSoil® DNA Isolation Kit (Carlsbad, CA, USA) following the same protocol as was used to extract DNA from samples, except no sample was added meaning that any bacterial DNA in the final elution must have been derived from the extraction kit reagents. 1 µl of this elution was added to the qPCR reaction mixture.

In order for us to compare the quantity of bacterial DNA found in bronchoscope wash and brushing samples it was necessary to use a unit of measurement which could be applied to both sample types. Bacterial DNA concentrations are therefore reported as the 16S copy numbers present per µl of eluent produced from samples by the MOBIO PowerSoil® DNA Isolation Kit.

Statistical analysis was carried out in Minitab® 16 for Windows. Where data was non-parametric the Mann-Whitney U test was used to statistically compare groups.

Results

Quality control and adequacy of sequencing

After constructing DNA sequences from the forward and reverse reads generated by sequencing, various quality control steps were performed to decrease the number of artefacts and poor quality sequences used in subsequent analyses.
For the Miseq runs, these steps resulted in a 15% loss of sequences (sequencing error rate = 0.39%). On average, samples contained 205625 ± 27232 (mean ± standard error of the mean SEM) sequences and a total of 925 bacterial OTUs were identified (Dataset S1). Sequences were assigned to OTUs based on their taxonomic classifications. Each OTU does not necessarily represent an individual bacterial species but instead represents the lowest taxonomic level to which its bacterial sequences could be assigned. For example, 77.4% of reads could be identified to genus while 31.1% could be assigned to species. If two species from the same genus could only be assigned to genus level then they would both be binned into the same OTU.

For the HiSeq run, samples contained on average 233505 ± 69735 (mean ± SEM) and the sequencing error rate was 0.39%. 633 OTUs were identified (Dataset S2) and the total reduction in sequence numbers due to quality control was 5%.

Good’s coverage estimate values exceeded 97% for all samples. This indicates that at least 97% of the bacteria present in our original samples are likely to have been identified, demonstrating that the depth of sequencing was adequate.

Of the twenty bacteria contained in the mock community, all could be taxonomically identified down to genus level except *Bacillus cereus*, *Escherichia coli* and *Listeria monocytogenes* which could only be identified at family level. This indicates that the primers were able to amplify a wide diversity of bacteria. Whilst the proportions of bacterial DNA were different to the proportions anticipated if no PCR bias was present (Table 2) this was less apparent in the sample which had been diluted 1:100 after the first round of PCR. In the undiluted mock the proportions of bacterial orders differed from the expected proportions by an average of 9.48% (SEM: 2.24%, range: 0.99% - 19.48%) whereas the orders in the diluted mock differed on average by 4.33% (SEM: 1.12%, range: 0.29% - 12.71%). This diluted mock community may be more comparable...
to the kind of biases we would find in our samples as the undiluted mock community contained a far higher concentration of template DNA (2000000 16S copies per µl) than our samples did on average (13133 16S copies per µl).

The assumption was made that PCR bias would reasonably be expected to apply equally across all samples and therefore any statistical tests between samples should still be valid. The two bacterial species most overrepresented in the undiluted mock community (*Deinococcus radiodurans* and *Helicobacter pylori*) are not commonly associated with the respiratory tract and bacteria from these genera were very rare within our dataset.

**Longitudinal study in six sheep over three months**

To examine the spatial, longitudinal and inter-individual variation of the sheep lung microbiota, lung brushings were taken from three spatially disparate lung locations (RA, RCD and LCD) in six sheep at three time-points (baseline, one month and three months). Estimates of total bacterial yield from qPCR analysis indicated that sheep lung brushings contained an average of 13133 ± 894 (mean ± SEM) 16S copy numbers/µl (range: 1032-37627 16S copy numbers/µl).

Bronchoscope wash controls contained significantly lower bacterial 16S rDNA concentrations than lung brushings (Mann-Whitney U test: p < 0.0001), containing an average of 1471 ± 279 (mean ± SEM) 16S copy numbers/µl (range: 397-4792 16S copy numbers/µl) ([Fig 3](#)). The qPCR negative water controls were found to contain 190, 479 and 739 16S copy numbers/µl and the extraction kit controls were found to contain 347 and 511 16S copy numbers/µl.

After sequencing and sub-sampling, bacterial communities isolated from extraction kit and 16S PCR negative controls were found to cluster separately from those found in sheep lung brushings.
(AMOVA: p < 0.001). Extraction kit controls were included from two different lots. The most abundant OTUs found in the first extraction kit control were Corynebacterium (36%), Enterobacteriaceae (13%), Mycobacterium llaterense (7%) and Staphylococcus haemolyticus (5%). The most predominant OTUs in the second extraction kit control were Aerococcus (13%), Dermabacteraceae (11%), Micrococcus (10%), Enhydrobacter (9%) and Leuconostoc (7.2%). The predominant bacterial order present in both extraction kit controls was Actinomycetales (50.1% and 40.5% respectively).

The bacteria isolated from lung brushings predominantly belonged to the orders Bacillales (26%), Actinomycetales (21%), Clostridiales (11%) and Lactobacillales (9%) while common genera included Staphylococcus (16%), Corynebacterium (9%), Jeotgalicoccus (5%) and Streptococcus (5%).

The underlying changes in bacterial OTUs between sampling points were examined. The bacterial communities found in lung brushings clustered significantly by time-point (AMOVA: P < 0.001) (Fig 4). The OTUs causing this clustering were identified by applying Metastats (Tables S3 and S4). The largest difference observed between the first and second time-points was an 11% increase in the abundance of an OTU identified as Corynebacterium. This is also the most abundant OTU in one of our extraction kit controls. OTU 12: Mycobacterium llaterense was also significantly more abundant at the one month time-point and was the third most abundant OTU in the same extraction kit control. It is therefore likely that our time-points were affected to different degrees by reagent contamination and therefore the analysis of segments over time is not possible. However, all samples taken in the same sheep at the same time-point were processed using the same extraction kit; therefore, an analysis of spatial variability can be performed.
Visual perceptions of community structure indicated that in some sheep, samples taken from separate lung sites differed appreciably, whereas in other sheep there appeared relative concordance between such samples (example shown in Fig 5). A full visual summary of the results can be found in Fig S2. There were no significant differences between the diversity of communities located at different lung sites (Inverse Simpson’s index: Friedman test: P > 0.5).

Sheep clustered separately by the composition of their lung bacterial communities at the baseline time-point (AMOVA: P = 0.001), and at the three month time-point (AMOVA: P = 0.045), indicating that samples taken from within the same sheep were more similar to one another than to samples taken from other sheep. At the one month time-point, sheep did not cluster in this manner (AMOVA: P = 0.394), though this is likely due to the presence of contamination causing a homogenisation of our one month samples. Pairwise comparisons of samples showed no significant results. The similarity of samples to one another can be visualised using PCOA graphs (Fig 6).

Spatial variability of the lung microbiota in an individual sheep

The observed variability between spatially disparate lung sites in some sheep prompted enquiry as to the consistency of bacterial communities sampled from sites in close spatial apposition.

Further samples were derived by systematically sampling multiple sites of the lungs of an individual animal at one time-point. Whilst the three month experiment did not include a control for every lot of extraction kit used, emerging literature and opinion within the field has since
indicated the value of using the same extraction kit for all samples. This strategy was therefore adopted for these latter samples which were all processed at the same time.

The extraction kit control was mainly composed of one OTU (OTU 18: 79%) which was also present in our brushing samples (mean ± SEM: 51.1 ± 3.3). We felt confident in removing this OTU from all of our samples prior to analysis as it could be identified to species level (*Methylobacterium komagatae*) and was considered highly unlikely to be found within the sheep lung. No further OTUs were removed before analysis.

Lung brushings contained on average 2116 16S copy numbers per ul (SEM = 365) while the throat swab and extraction kit control contained 42480 and 43 16S copy numbers per ul respectively. The richness and diversity of the lung samples (Chao = 103.77 ± 7.32, Inverse Simpson’s index = 14.24 ± 2.14) was found to be far lower than in the throat swab (Chao = 257.038, Inverse Simpson’s index = 9.19). Sample A1, taken from the ventral aspect of the trachea just caudal to the bifurcation with the right apical lobe segmental bronchus, had the second highest richness (Chao = 155.024) and diversity (Inverse Simpson’s index = 8.713).

However, sample A2, which was taken at the same level as sample A1 but from the dorsal aspect of the trachea, had a much lower richness (Chao = 76.038) and diversity (Inverse Simpson’s index = 4.925).

The composition of the communities taken from the respiratory tract showed some variation, even between paired samples located very closely to one another (**Fig 7**). Sub-tracheal samples paired to their most proximate neighbour did not cluster together significantly when OTUs were defined at the lowest taxonomic depth (AMOVA: P = 0.30). However, paired samples did cluster significantly by the bacterial orders which they contained (AMOVA: P = 0.046). Sub-tracheal samples also clustered significantly (by order) based upon the depth in the lung from which
samples were taken (AMOVA: P = 0.033) (Fig 8) (lung depth in this context refers to increasing
distance from the glottis progressing in a caudal direction). An indicator OTU for the group
which included the samples A4, A5, A14 and A15 was found to be OTU 4: Pseudomonadales
(P = 0.042). The most abundant bacterial orders identified from brushings were Clostridiales
(25.8%), Pseudomonadales (18.3%) and Actinomycetales (16.0%) while the throat swab was
dominated by Pasteurellales (36.5%) and Pseudomonadales (15.1%). The extraction kit control
was predominantly composed of Actinomycetales (31.1%) and Pseudomonadales (31.0%).

As the Pasteurellales order contains several species which are known to act as sheep lung
pathogens and which display regional patterns of infection we felt it would be interesting to
investigate where OTUs belonging to this order were found within the respiratory tract (Table
3). By far the largest proportion of these OTUs was found in the throat swab and one of the
tracheal brushings (A1).

Discussion

In order to better understand the variability present in the sheep lung microbiota, we compared
the lung bacterial communities of six sheep at three different lung sites over a duration of three
months. To further explore the extent of spatial variability, we also took 17 samples from the
respiratory tract of one sheep.

Previously, the bacteria in healthy domestic sheep lungs had been investigated by culture-based
methods which seemed to indicate that bacterial colonisation of the sheep lung was rare or did
not occur in all sheep (11, 12, 14). In contrast, using non-culture based methods we have found
that all of the sampled sites in our seven sheep harboured diverse communities of bacteria,
although in far smaller numbers than is generally found in other niches such as the gut or upper
respiratory tract.

Bacteria belonging to genera previously isolated from goat and sheep lungs (11, 12) were found
in our samples. These included *Corynebacterium*, *Bacillus*, *Enterococcus*, *Klebsiella*,
*Mannheimia*, *Micrococcus*, *Moraxella*, *Pasteurella*, *Pseudomonas*, *Staphylococcus* and
*Streptococcus*. Of the most common genera observed within our animals *Staphylococcus*,
*Streptococcus* and *Corynebacterium* are commonly isolated from the upper respiratory tract and
skin of many animals whereas *Jeotgalicoccus* is a less well known genus (45) which has not
been found to make up a substantial part of the lung microbiota communities in any previous
studies. However, it has been isolated from the small intestinal mucosa of calves (46), the canine
oral cavity (47), aerosols in a poultry house (48, 49), cattle teats (50), lamb meat (51), the rumen
of cattle (52) and from aerosol samples near a dairy (53).

The most common bacterial orders found in the sheep lung during the three month study were
Bacillales, Actinomycetales and Clostridiales. This agrees with the findings of a previous study
carried out by our group which examined the sheep lung microbiota before and after infection
with *P. aeruginosa* (17). Pseudomonadales (mainly *Pseudomonas*) was also commonly found in
the lungs during our single sheep study while the throat swab from this study was dominated by
*Pasteurellales* and Pseudomonadales.

Co-infection with *Bordetella parapertussis* or *Mycoplasma ovipneumoniae* has been shown to
lead to more severe disease caused by *Mannheimia (Pasteurella) haemolytica* (22-25).

Mycoplasmas were very rare within our dataset with only one sheep segment containing reads
from this genus at one time-point. We did not identify any OTUs as *Bordetella*; however, we did
find an OTU designated as Alcaligenaceae (the family to which \textit{Bordetella} belongs) though these were uncommon and occurred in low abundance. We identified several OTUs which were classified as members of the Pasteurellaceae family including \textit{Mannheimia} and \textit{Bibersteinia} and less commonly: \textit{Aggregatibacter segnis}, \textit{Haemophilus parainfluenzae}, \textit{Bibersteinia trehalosi} and \textit{Actinobacillus parahaemolyticus}. All of these microbes have previously been isolated from the lungs or upper respiratory tract (54-58). Despite the fact that disease by members of this family is often located in the apical and cardiac lobes (12, 19), we observed members of this family to be present across the lung.

The composition of the lung microbiota found in our sheep shows some differences to that previously identified in humans where Bacteroidales are found in higher numbers and there are generally less members of the Actinomycetales and Clostridiales (2, 9, 59). Segal \textit{et al.} identified various bacterial taxa that were commonly found in high relative abundance in human lungs (1). These included taxa which were found in all of our sheep samples in high relative abundance (\textit{Streptococcus}, \textit{Staphylococcus}, \textit{Corynebacterium}); taxa which were found in the majority of our samples but in lower abundances (\textit{Propionibacterium}, \textit{Pseudomonas}) and taxa which were only found sporadically in our samples and were usually in low abundance (\textit{Stenotrophomonas}, \textit{Prevotella}, \textit{Veillonella}, \textit{Fusobacterium}, \textit{Porphyromonas}).

Such differences may at least in part reflect the different surroundings in which sheep live as well as behavioural or physiological features such as rumination. A study using buccal swabs to identify bacteria originating from the rumen suggested that as the time between regurgitation and sampling increases the orally associated bacterial populations in the buccal cavity will increase and the rumen associated bacteria will decrease, potentially contributing to inter-animal variation.
In future studies it may be useful to take rumen and upper respiratory tract samples alongside lung samples to explore if the variability between these sites and the lung is related.

Regardless of the highlighted differences between sheep and the human lung microbiota, there is a pressing need to understand the mechanisms that underlay the spatial and temporal variability of microbiota in the mammalian lung. These fundamental studies are difficult to facilitate in healthy human subjects as a consequence of the invasive nature of the repeated sampling protocol, as well as controlling for the influence of environmental and/or lifestyle factors. Large animal models can however play an important role in filling this need. Indeed, the physiological and immunological similarities between sheep and human lungs (60, 61) have contributed to the widespread use of sheep as translational models for human lung research (26, 27) including asthma (62-65), the delivery of drugs via the upper respiratory tract (66-68), emphysema (69-71), pulmonary hypertension (72-74), physical lung injury (75-78), lung infection (28, 79-81), respiratory distress syndromes (82-85), asbestosis (86-88) and lung cancer (89, 90).

In our study we examined the variability of the lung microbiota in sheep. Bacterial populations were often different between lung segments and between individuals, which confirms our previous observations (17). There was more similarity between samples from the same sheep at the baseline and three month time-points than between samples taken from different sheep but this was not found to be the case at the one month time-point. Lung sample clustering by individuals has previously been identified in humans (9) and sheep (17).

Clearly, large differences can exist in the microbiota sampled from different lung segments at the same time-point. This spatial variability of lung microbial populations can be observed in *P. aeruginosa* infections in cystic fibrosis patient lungs (91). The mechanisms underlying such observations have yet to be elucidated; however, possible candidate influences may include...
regional variability of physiological parameters such as gas concentrations, osmolality, temperature, pH and blood flow (92-96) which may lead to the creation of ‘microhabitats’ providing a selective advantage to certain bacteria (97). It has previously been demonstrated that differences in pH can lead to changes in the colonic microbiota (98) and that temperature combined with humidity can lead to changes in the composition of the skin microbiota (99).

A longitudinal analysis of the lung microbiota at specific lung sites in healthy individuals has not previously been reported. Our goal was to define the variability of the lung microbiota over time and to detect whether there was a sheep lung microbiota ‘signature’ which remains stable. Unfortunately, at the time of carrying out this study the extent of the variability of bacterial DNA found within different lots of extraction kits was not yet known (100). While we therefore did include some extraction kit controls for our longitudinal study we did not include controls for all lots which were used. Samples from different time-points were also processed at different times. Due to our small sample sizes and the fact that samples clustered significantly by time-point, we do not feel that accurate conclusions can be drawn about the temporal stability of the microbiota from our data. However, all samples taken from the same time-point in the same sheep were processed at the same time. Therefore, we can be confident that the spatial variability we observe within animals is not due to our methodology.

In some individuals, samples taken from different lung segments were found to be highly different from one another whereas in others the lung microbiota appeared to be quite stable across the lung. Another finding was the disappearance of the significantly separate clustering of sheep samples at the one month time-point. This was correlated with an increase in the proportions of several OTUs found in sheep lungs, the most noticeable increase arising from an OTU classified as Corynebacterium which was also the most abundant OTU in one of our
extraction kit controls. It is likely that the disappearance of significant clustering by individual at
the one month time-point is due to the increased presence of contamination in our samples.
OTUs that were identified in both samples and negative PCR and extraction kit controls were not
removed from the analysis for the three month sheep study. The reason for this decision was that
a number of bacteria commonly associated with the upper and lower respiratory tract were
present in these controls, including the genera *Streptococcus* and *Pseudomonas*, and it was
judged that their removal would merely introduce another source of bias.
Equally, any specific *a priori* manipulation based around assumptions gleaned from the human
literature regarding microbiota in the upper and lower respiratory tract, are potentially ill advised.
Indeed, it has been demonstrated that the microbes found in the lungs of animals often match
those found in their bedding and hay (101). It is therefore not possible to dismiss environmental
microorganisms as being only due to the contamination of samples.
In our spatial variation study, one OTU was removed before analysis as we felt confident that its
presence was due to contamination of our extraction kit. Clustering of lung brushings by the lung
depth from which they were taken was observed when OTUs were defined by bacterial order.
Samples paired with their proximate neighbour were also found to cluster significantly separately
from brushings taken elsewhere in the lung but this may just be due to the fact that these samples
were taken from the same lung depth. Certainly further research to explore the relationship
between lung depth and community composition appears warranted.
After sequencing a mock community of bacteria which contained equimolar concentrations of
each bacterial species we did find some bias present, with some bacterial species being over or
underrepresented. These biases, which may be caused by various factors including primer
mismatching, PCR cycle number and the bioinformatic pipeline used, are quite common in 16S sequencing (102-105). We also sequenced a 1:100 dilution of the same mock community and found that the apparent biases were far less. As the concentration of bacterial DNA in our samples was far lower than that of the undiluted mock community, we feel that the 1:100 dilution is likely to better represent the biases which may be present in our samples as it is closer to their bacterial DNA concentrations. We believe that this vindicates our choice of DNA amplification strategy, including the use of nested PCR.

It may not be possible to claim that the bacterial abundances identified via 16S sequencing quantitatively represent the relative abundances of bacteria in the sample. Indeed, this is made even more difficult as different bacterial taxa contain different copy numbers of the 16S gene (106). However, it seems logical to assume that if the same methodology is used for all samples within a study then the biases present will be the same for all samples and therefore comparisons between groups or claims about the types of microbes present in samples would still be valid.

In conclusion, we observed variability in the sheep lung microbiota both between and within individuals. In some animals different lung segments contained highly different bacterial communities whereas other animals showed similar communities at all lung sites. While spatial variation was observed to occur over both large and small distances across the lung, samples taken at the same lung depth clustered together separately from those taken at different lung depths. Further studies are needed to explore the stability of the healthy lung microbiota over time.
Funding Information

Funding for this study was supplied by the BBSRC (Biotechnology and Biological Sciences Research Council) under the EASTBIO doctoral training program (Grant number: BB/J01446X/1). The funders had no input on study design; data collection and interpretation or the decision to submit this study for publication.

Acknowledgments

The authors would like to thank the administrative staff of the BBSRC EASTBIO DTP program for their encouragement and support. We would like to thank the staff at Dryden farm for providing the everyday care of our animals. We would also like to thank Dr. Helen Brown for her advice on statistical methods and sample sizes and Tina Baker for providing the qPCR protocol. The Human Microbiome Project Mock Community HM-782D was obtained through BEI Resources, NIAID, NIH as part of the Human Microbiome Project: Genomic DNA from Microbial Mock Community B (Even, Low Concentration), v5.1L, for 16S rRNA Gene Sequencing, HM-782D.

References


methicillin-resistant *Staphylococcus aureus*-induced pneumonia and sepsis. Shock 29:642-649.


Figure Legends

Fig 1: Diagram of the sheep lung, divided into anatomical segments: Boxes indicate the segments from which lung protected specimen brushings were taken in six sheep at three time-points; these correspond to the RA: right apical, RCD: right caudal diaphragmatic, and LCD: left caudal diaphragmatic segments.

Fig 2: Location of brushings within the sheep lung: Protected specimen brushings were taken from the sections of the lung labelled A1-A9 and A13-A19 in one sheep at one time-point.

Fig 3: qPCR of lung brushings and controls: The bronchoscope channel was flushed through with 7.5 ml PBS and the wash collected (Wash Control n=18) prior to protected specimen lung brushings being taken from sheep (Lung Brushings n=54). DNA was extracted from Wash Controls and Lung Brushings and the quantity of bacterial DNA calculated using 16S rDNA qPCR. Lung Brushings were found to contain significantly higher quantities of bacterial DNA than Wash Controls (Mann-Whitney U test: P < 0.0001). Negative controls consisted of either water (n=3) or extraction kit controls (n=2). Boxes indicate interquartile ranges and outliers are represented as diamonds.

Fig 4: Clustering of time-points by lung microbiota composition: PCOA graph showing the similarities between bacterial communities sampled from three sheep lung segments in six sheep at three time-points. Samples were found to cluster significantly by the time-point at which they were taken (AMOVA: P < 0.001).

Fig 5: The bacterial communities found in three separate lung segments within two sheep: Protected specimen brushings were taken from the lungs of sheep at three different lung segments (RA: right apical, RCD: right caudal diaphragmatic and LCD: left caudal diaphragmatic) at day 0. Sheep A (2S066) had highly different bacterial communities at each
lung segment whereas Sheep B (2D644) had similar bacterial communities at all three lung
sites.

**Fig 6: Clustering of individuals by lung microbiota composition:** PCOA graphs showing
the similarities between the bacterial communities extracted from protected specimen
brushing samples taken from sheep lungs at three time-points (baseline (0 days), one month
and three months). Samples were taken from three separate lung segments (RA: right apical,
RCD: right caudal diaphragmatic and LCD: left caudal diaphragmatic). Samples from within
the same sheep were found to cluster significantly at baseline (AMOVA: P = 0.001) and three
months (AMOVA: P = 0.045) but not at one month. This is likely to be due to the presence of
contaminants originating from the extraction kits in the one month samples.

**Fig 7: Diagram of the bacterial orders found in the sheep lung:** Bacterial orders found in
protected specimen brushings from the lung and trachea (A1-A9 and A13-A19), throat swab
and an extraction kit control taken during a study of one sheep at one time-point.

**Fig 8: Clustering of lung brushings by depth within the lung:** PCOA graph showing the
similarity of samples taken at different lung depths based upon the bacterial orders present.
Lung depths are represented by colour and correspond to different distances from the glottis
progressing in a caudal direction. Adjacent pairs of samples are represented by the same
symbol and colour. For the exact location of each sampling site see **Fig 2.** Sub-tracheal
samples (≥ A3) clustered significantly by lung depth (AMOVA: P= 0.033) as did paired
samples (AMOVA: P = 0.046).
### Table 1: Sheep used in this study

<table>
<thead>
<tr>
<th>Sheep ID</th>
<th>Gender</th>
<th>Mean weight (kg) ± SD</th>
<th>Mean rectal temperature (°C) ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D618</td>
<td>Female</td>
<td>51 ± 3.1</td>
<td>39.0 ± 0.06</td>
</tr>
<tr>
<td>2S066</td>
<td>Male (castrated)</td>
<td>69 ± 2.6</td>
<td>39.6 ± 0.20</td>
</tr>
<tr>
<td>2D619</td>
<td>Female</td>
<td>59 ± 1.7</td>
<td>39.3 ± 0.20</td>
</tr>
<tr>
<td>2D620</td>
<td>Female</td>
<td>64 ± 4.6</td>
<td>39.1 ± 0.21</td>
</tr>
<tr>
<td>2D644</td>
<td>Female</td>
<td>65 ± 1.0</td>
<td>39.3 ± 0.06</td>
</tr>
<tr>
<td>2D645</td>
<td>Female</td>
<td>70 ± 2.0</td>
<td>39.4 ± 0.06</td>
</tr>
</tbody>
</table>

### Table 2

Proportion of DNA sequence reads belonging to bacterial members of a mock community

<table>
<thead>
<tr>
<th>Taxonomy</th>
<th>Expected proportion of reads</th>
<th>Actual proportion (undiluted)</th>
<th>Actual proportion (1:100)</th>
<th>Mock community species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deinococcales</td>
<td>5%</td>
<td>24.48%</td>
<td>7.65%</td>
<td><em>Deinococcus radiodurans</em></td>
</tr>
<tr>
<td>Campylobacterales</td>
<td>5%</td>
<td>22.05%</td>
<td>12.65%</td>
<td><em>Helicobacter pylori</em></td>
</tr>
<tr>
<td>Bacteroidales</td>
<td>5%</td>
<td>19.59%</td>
<td>10.91%</td>
<td><em>Bacteroides vulgatus</em></td>
</tr>
<tr>
<td>Class</td>
<td>Percent</td>
<td>Genus/Species</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>---------</td>
<td>-------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacillales</td>
<td>20%</td>
<td><em>Bacillus cereus, Listeria monocytogenes, Staphylococcus aureus, Staphylococcus epidermidis</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lactobacillales</td>
<td>25%</td>
<td><em>Enterococcus faecalis, Lactobacillus gasseri, Streptococcus agalactiae, Streptococcus mutans, Streptococcus pneumoniae</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridiales</td>
<td>5%</td>
<td><em>Clostridium beijerinckii</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhodobacterales</td>
<td>5%</td>
<td><em>Rhodobacter sphaeroides</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudomonadales</td>
<td>10%</td>
<td><em>Acinetobacter baumannii, Pseudomonas aeruginosa</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterobacteriales</td>
<td>5%</td>
<td><em>Escherichia coli,</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseriales</td>
<td>5%</td>
<td><em>Neisseria meningitidis</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actinomycetales</td>
<td>10%</td>
<td><em>Actinomyces odontolyticus,</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other/unclassified</td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Genus</strong>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Deinococcus</em></td>
<td>5%</td>
<td><em>Deinococcus radiodurans</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Helicobacter</em></td>
<td>5%</td>
<td><em>Helicobacter pylori</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bacteroides</em></td>
<td>5%</td>
<td><em>Bacteroides vulgatus</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Rhodobacter</em></td>
<td>5%</td>
<td><em>Rhodobacter sphaeroides</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Clostridium</em></td>
<td>5%</td>
<td><em>Clostridium beijerinckii</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Staphylococcus</em></td>
<td>10%</td>
<td><em>Staphylococcus aureus, Staphylococcus epidermidis</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 3

Abundance of the OTUs within the Pasteurellaceae family found in different locations of the sheep respiratory tract

<table>
<thead>
<tr>
<th>Group</th>
<th>OTU 5: Mannheimia (%)</th>
<th>OTU 6: Pasteurellaceae (%)</th>
<th>OTU 7: Bibersteinia (%)</th>
<th>OTU 9: Bibersteinia trehalosi (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Throat swab</td>
<td>23.7</td>
<td>10.1</td>
<td>1.8</td>
<td>0.7</td>
</tr>
<tr>
<td>Trachea</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>5.5</td>
<td>4.5</td>
<td>28.4</td>
<td>5.3</td>
</tr>
<tr>
<td>A2</td>
<td>0</td>
<td>0.03</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Left Lung</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>0</td>
<td>3.04</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>A4</td>
<td>1.2</td>
<td>0.006</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A5</td>
<td>0</td>
<td>0.2</td>
<td>0.006</td>
<td>0</td>
</tr>
<tr>
<td>A6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A7</td>
<td>0.7</td>
<td>1.4</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>A8</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>0</td>
</tr>
<tr>
<td>A9</td>
<td>0.006</td>
<td>0.02</td>
<td>0.2</td>
<td>0</td>
</tr>
</tbody>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Right Lung</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A13</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>0.006</td>
</tr>
<tr>
<td>A14</td>
<td>0.006</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A15</td>
<td>0</td>
<td>0.6</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>A16</td>
<td>2.3</td>
<td>0.6</td>
<td>0.006</td>
<td>0</td>
</tr>
<tr>
<td>A17</td>
<td>0.10</td>
<td>0</td>
<td>1.3</td>
<td>0.01</td>
</tr>
<tr>
<td>A18</td>
<td>0</td>
<td>0.02</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>A19</td>
<td>3.2</td>
<td>0</td>
<td>1.3</td>
<td>0</td>
</tr>
</tbody>
</table>
A1: Ventral trachea just caudal to bifurcation with right apical (RA) lobe segmental bronchus
A2: Dorsal trachea just caudal to bifurcation with right apical (RA) lobe segmental bronchus
A3: Medial aspect of left main bronchus immediately caudal to carina
A4: Medial aspect of first left ventral diaphragmatic segmental bronchus (LVD1), just caudal to bifurcation from left main bronchus
A5: Left main bronchus, lateral wall just after bifurcation to LVD1
A6: Medial aspect of second left ventral diaphragmatic segmental bronchus (LVD2), just caudal to bifurcation from left main bronchus
A7: Left main bronchus, lateral wall just after bifurcation to LVD2
A8: Third left ventral diaphragmatic segmental bronchus (LVD3)
A9: Left caudal diaphragmatic (LCD) segmental bronchus
A13: Medial aspect of right main bronchus immediately caudal to carina
A14: Medial aspect of first right ventral diaphragmatic segmental bronchus (RVD1), just caudal to bifurcation from right main bronchus
A15: Right main bronchus, lateral wall just after bifurcation to RVD1
A16: Medial aspect of second right ventral diaphragmatic segmental bronchus (RVD2), just caudal to bifurcation from right main bronchus
A17: Right main bronchus, lateral wall just after bifurcation to RVD2
A18: Third right ventral diaphragmatic segmental bronchus (RVD3)
A19: Right caudal diaphragmatic (RCD) segmental bronchus