

Microbiome Characteristics of Induced Sputum compared to Bronchial Fluid and Upper Airway Samples

| | |
|-------------------------------|--|
| Journal: | <i>Pediatric Pulmonology</i> |
| Manuscript ID | PPUL-17-0455.R2 |
| Wiley - Manuscript type: | Original Article: Respiratory Infections |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | An, Shi-Qi; Queen's University Belfast Faculty of Medicine Health and Life Sciences, Wellcome Wolfson Institute for Experimental Medicine Warris, Adilia; University of Aberdeen, Institute of Medical Sciences Turner, Steve; University of Aberdeen, Department of Child Health |
| Keywords: | Infections: Pneumonia, TB, Viral |
| Other Keywords: | Microbiota |
| Abstract: | <p>Objective. The study of the community of microorganisms (the microbiota) in the lower airways in children is restricted to opportunistic sampling in children undergoing elective general anaesthetic. Here we tested the hypothesis that induced sputum is a valid alternative to directly sampling the lower airways to study lower airway microbiota.</p> <p>Methods. Children scheduled for elective operations were recruited. Pre-operatively a sample of induced sputum was obtained. After anaesthesia was induced, a bronchial brushing and swabs of the upper respiratory tract were obtained. Bacterial community analysis was performed by amplification of the V3-V4 16S rRNA gene region.</p> <p>Results. Twenty children were recruited, mean age 10.7 years. Induced sputum samples were obtained from 12 children, bronchial brushing from 14 and nasal, mouth and throat samples in 15, 16 and 17 children. The profile of bacterial communities was similar in the mouth, throat and sputum samples with the nose and bronchial samples being different. Actinobacteria species dominated the nose and mouth, Fusobacteria were the dominant species in the throat and sputum whilst Proteobacteria species dominated in bronchial samples. Forty-one percent of detected bacteria in bronchial samples were unclassified. Bacterial communities from the mouth, throat and induced sputum were tightly clustered and were distinct from nose and those found in bronchial communities.</p> <p>Conclusions. Induced sputum may not be a valid surrogate for microbiome assessment of the lower airways in all individuals. Many bacteria in bronchial samples were not recognised by standard testing, suggesting that our understanding of the lower airway microbiota in children remains rudimentary.</p> |

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SCHOLARONE™
Manuscripts

For Peer Review

1
2
3 Microbiome Characteristics of Induced Sputum compared to Bronchial Fluid and Upper
4
5 Airway Samples
6
7

8 Shi-qi An^{1,2}, Adilia Warris^{3,4}, Steve Turner³
9

10
11
12 ¹Division of Molecular Microbiology, University of Dundee
13

14
15 ² Wellcome Wolfson Institute for Experimental Medicine, Queen's University Belfast
16
17

18
19 ³ Department of Child Health, University of Aberdeen
20

21
22 ⁴ MRC Centre for Medical Mycology, University of Aberdeen
23
24
25
26
27

28
29 Contact details: Dr Steve Turner, Child Health, Royal Aberdeen Children's Hospital,
30
31 Aberdeen, UK, AB25 2ZG. Tel +44 1224 438470. s.w.turner@abdn.ac.uk
32
33

34 Running head: Microbiota of induced sputum
35
36

37
38 Key words: Bronchial fluid, Child, Microbiota, Sputum
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Objective. The study of the community of microorganisms (the microbiota) in the lower airways in children is restricted to opportunistic sampling in children undergoing elective general anaesthetic. Here we tested the hypothesis that induced sputum is a valid alternative to directly sampling the lower airways to study lower airway microbiota.

Methods. Children scheduled for elective operations were recruited. Pre-operatively a sample of induced sputum was obtained. After anaesthesia was induced, a bronchial brushing and swabs of the upper respiratory tract were obtained. Bacterial community analysis was performed by amplification of the V3-V4 16S rRNA gene region.

Results. Twenty children were recruited, mean age 10.7 years. Induced sputum samples were obtained from 12 children, bronchial brushing from 14 and nasal, mouth and throat samples in 15, 16 and 17 children. The profile of bacterial communities was similar in the mouth, throat and sputum samples with the nose and bronchial samples being different. Actinobacteria species dominated the nose and mouth, Fusobacteria were the dominant species in the throat and sputum whilst Proteobacteria species dominated in bronchial samples. Forty-one percent of detected bacteria in bronchial samples were unclassified. Bacterial communities from the mouth, throat and induced sputum were tightly clustered and were distinct from nose and those found in bronchial communities.

Conclusions. Induced sputum may not be a valid surrogate for microbiome assessment of the lower airways in all individuals. Many bacteria in bronchial samples were not recognised by standard testing, suggesting that our understanding of the lower airway microbiota in children remains rudimentary.

INTRODUCTION

The lower airways were traditionally considered sterile in healthy individuals, and this paradigm was based on observations that standard microbiology culture from samples collected from healthy individuals yielded no pathogens¹. The advent of polymerase chain reaction (PCR) testing for bacterial DNA has now identified communities of bacteria in lower airway secretions from healthy individuals^{2,3}. The relevance of bacteria in airway microbiota to respiratory symptoms is unclear, but there is evidence suggesting that the microbiota may be relevant to influence respiratory exacerbations^{2,4} or even to the aetiology of chronic respiratory conditions such as asthma⁵. Exacerbations of chronic conditions such as cystic fibrosis (CF) and chronic obstructive pulmonary disease are linked to a change² (termed “airway dysbiosis”) or reduction⁴ in the dominant bacterial species in the microbiota. In young infants, those carrying *Haemophilus influenzae* or *Moraxella catarrhalis* in the hypopharynx were more likely to have asthma symptoms at three years of age⁵.

A major challenge to studying the lower airway microbiota is obtaining samples without contamination from the upper airways³, and it is known that the microbiota in the naso- and oropharynx differ from that in the bronchus in adults and children⁶⁻⁸. Three papers have used bronchoscopic sampling to compare the upper and lower airway microbiome in children and whilst there are similarities in the dominant phyla identified in some studies, there are differences in genera identified and also in the clustering of bacterial communities⁶⁻⁸. The characteristics of the lower airway microbiome may also vary between individuals depending on their asthma status⁶ and using a nasal or oral approach to the lower airways⁸. . Two studies in young people with CF have compared the microbiota in

1
2
3 induced sputum and oropharyngeal swabs^{9,10}, and both concluded that a throat swab
4
5 sample may be a valid surrogate for sputum for microbiota analysis. Induced sputum is a
6
7 non-invasive alternative to direct bronchial sampling, and is used in the clinical management
8
9 of tuberculosis and CF in children and also in research. The study of the lower airway
10
11 microbiome would be considerably easier if induced sputum was a valid surrogate of
12
13 bronchial fluid. The aim of the current study was to collect induced sputum and samples
14
15 from the bronchus, nose, throat and mouth in children and undertake a comprehensive
16
17 assessment of upper and lower airway microbiota in children. Our hypothesis was that
18
19 induced sputum is a valid alternative to directly sampling the lower airways to study lower
20
21 airway microbiota.
22
23
24
25
26
27

28 **METHODS**

29 **Study design**

30
31
32
33 Children aged 5-16 years and scheduled for elective ear nose and throat operations at Royal
34
35 Aberdeen Children's Hospital between January and June 2015 were invited to take part. All
36
37 children were scheduled for tonsillectomy and some also had grommets inserted. These
38
39 operations require endotracheal intubation, which allows sampling of the lower airways
40
41 without contamination by the upper airways. Children younger than 5 years were excluded
42
43 since they were unlikely to provide a sample of induced sputum. Other exclusion criteria
44
45 included receipt of antibiotics within the last month and CF. In the hours before the
46
47 operation (when the child was fasted), children took part in an assessment which included
48
49 completion of a questionnaire, height and weight measurement, spirometry (in accordance
50
51 with international guidelines¹¹) and induction of sputum using 4% and if required 5% saline
52
53
54
55
56
57
58
59
60

1
2
3 as previously described¹². An adequate quality sputum sample was defined as the presence
4
5 of a white sputum plug on a visual inspection; samples were not sufficiently large enough to
6
7 allow for cell count analysis. Immediately after anaesthesia was induced and the
8
9 endotracheal tube *in situ*, a sterile 2.7 mm interdental brush (Dento Care Professional,
10
11 London UK) was used to collect a sample of nasal secretions and a separate brush
12
13 interdental brush was gently brushed on the buccal surface of the mouth to collect oral
14
15 secretions. A standard bacterial swab (Transwab, Medical Wire Equipment, Corsham, UK)
16
17 was used to obtain a sample of pharyngeal secretions under direct inspection. Finally, a
18
19 bronchoscopy cytology brush (10-mm disposable cytology brush, BC 202D-2010, Olympus,
20
21 Southend-on-Sea, Essex, UK) was passed down the endotracheal tube until resistance was
22
23 met and then rotated and withdrawn to obtain “blind” bronchial samples. Blood was taken
24
25 for serum total IgE assay. The study was approved by the North of Scotland Research Ethics
26
27 Committee (13/NS/0144), written parental consent was obtained and verbal assent was
28
29 obtained from all participants.
30
31
32
33
34
35
36
37
38
39

40 **DNA extraction and sequencing**

41 The bacterial community analysis was carried out as described previously¹³ with some
42
43 modification as detailed below. Briefly, samples taken from the nose, mouth, throat,
44
45 sputum and bronchial brushing were frozen at -80°C prior to analysis. Samples were then
46
47 transported from Aberdeen to Dundee for processing. Upon arrival the sample had thawed
48
49 and were suspended in phosphate buffer saline where genomic DNA was extracted and
50
51 purified using the DNA/RNA All Prep kit¹⁴ (Qiagen) and stored according to manufacturer’s
52
53 instructions. Standard protocol, 16S Metagenomic Sequencing Library Preparation Guide
54
55
56
57
58
59
60

1
2
3 (Illumina), was followed to prepare sequencing libraries targeting the variable V3 and V4
4
5 regions of the 16S rRNA gene and paired-end sequencing was performed on the MiSeq
6
7 System (Illumina). We followed sampling and controls procedures described by the Earlham
8
9 Institute
10
11 [12 \(http://www.earlham.ac.uk/sites/default/files/docs/Genomic%20Services/Sample%20Guide](http://www.earlham.ac.uk/sites/default/files/docs/Genomic%20Services/Sample%20Guide)
13
14 [15 lines%20Aug17.pdf\)](http://www.earlham.ac.uk/sites/default/files/docs/Genomic%20Services/Sample%20Guide) and Illumina
16
17 [18 \(https://support.illumina.com/documents/documentation/chemistry_documentation/16s/1](https://support.illumina.com/documents/documentation/chemistry_documentation/16s/1)
19 [20 6s-metagenomic-library-prep-guide-15044223-b.pdf](https://support.illumina.com/documents/documentation/chemistry_documentation/16s/1)
21
22 which are designed to minimize the risk for contamination. Quality controls and sequencing
23
24 were performed at Earlham Institute (Norwich, UK).
25

26 27 **Data and Statistical analysis**

28
29
30 Bioinformatics analyses on raw data was performed using the Biomedical Genomics
31
32 Workbench version 4.0 (Qiagen) equipped with the Microbial Genomics Module version 2.0
33
34 (Qiagen) plugin. Sequences were imported and processed for optional merge paired reads,
35
36 adapter trimming, fixed length trimming and then the sequences were filtered based on the
37
38 number of reads to obtain sequences that are comparable in length and coverage for
39
40 clustering. Quality and chimera filtering were performed using the recommended
41
42 programme parameters [for complete details see www.qiagenbioinformatics.com]. Samples
43
44 with low coverage were removed from further analysis. Operational Taxonomic Units (OTU)
45
46 clustering and taxonomic assignment were done using Greengenes v13_5 (97%) as
47
48 reference. New OTUs were indicated when similarity percentage was lower than 80% with
49
50 minimum occurrence of 5 reads. Low abundant OTUs were discarded from further analyses
51
52 (minimum combined abundance was set at 10). Summary of processed sequence data is
53
54
55
56
57
58
59
60

1
2
3 described in Table S1. MUSCLE was used for OTUs alignment in order to reconstruct a
4
5 maximum likelihood phylogeny with neighbor joining as construction method and Jukes
6
7 Cantor as nucleotide substitution model. We compared community structures and diversity
8
9 across patient cohorts to determine if inter-cohort differences in structure were seen. To
10
11 achieve this a rarefaction sampling analysis was carried out using a standard methodology¹⁵.
12
13 In order to minimise the risk for contamination, we sought to identify a high (i.e.
14
15 conservative) number of sequences required to characterise bacteria from the samples. In
16
17 the absence of a standard international stringency cut off the default settings in the
18
19 workbench software (Qiagen) were applied; the default settings filters out many reads and
20
21 generates fewer unassigned reads and is therefore highly stringent. These adjustments were
22
23 assessed by analyzing a previously characterized mock community to confirm that the data
24
25 handling was consistent with other studies (sequencing files of the project:
26
27 ERP021973,<https://www.ebi.ac.uk/metagenomics/projects/ERP021973/samples/ERS15889>
28
29 [32](#)). Data generated in the study was assessed using the Qiagen software and also applied to
30
31 EBI Metagenomics analysis pipeline V3.0 to verify results (acknowledging that fewer reads
32
33 would remain in the analysis with Qiagen). The findings in the controls were taken into
34
35 account during the analysis. Alpha diversity was calculated using number of OTUs. Beta
36
37 diversity was obtained using D_{0.5} UniFrac and represented as Principal Coordinate Analysis
38
39 (PCoA). Robustness analysis was performed using PERMANOVA with UniFrac distances. All
40
41 sequencing data have been deposited in NCBI Sequence Read Archive (SRA) database under
42
43 the follow BioProject ID: PRJNA388557
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

RESULTS

Study Subjects

Twenty patients recruited of which 8 were boys, 7 with diagnosed asthma and the mean (standard deviation) age was 10.7 (2.8) years. Descriptives of the children recruited are given in Table 1. Induced sputum was obtained from 12 children and samples of fluid from bronchus, throat mouth and nose were obtained in all children. Three control samples were collected.

Sample quality control

A total of 11,750,879 PCR reads were analysed, and 606,227 high quality reads were obtained. The resulting OTU table contained 1053 OTUs, where 532 OTUs were assigned based on the Greengenes v13_5 (clustered at 97% similarity), and 521 OTUs were novel (see Table S1). These processing criteria allowed data from 74 samples from patients for further analyses. Of these 15 were from the nose, 16 from mouth, 17 from throat, 12 from sputum and 14 from bronchial samples (Table S2). Samples from all five sites were available in six children. The median read counts for the five sites were as follows: bronchus 4438; mouth 4400; throat 4868; nose 8927; and sputum 11250. The results from the mock community analysis using Qiagen and EBI Metagenomics analysis pipelines were highly consistent. However Qiagen software assigned phyla for all samples whereas the EBI Metagenomics did not assign phyla in 15% of samples (see Table S3).

Bacterial abundance

1
2
3 The relative bacterial abundance in each sample cohort is presented in Figure 1. The overall
4
5 profile of bacterial communities was largely similar in the mouth, throat and sputum
6
7 samples with the nose and bronchial being different. All sample cohorts contained five
8
9 major bacterial phyla: Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria and
10
11 Proteobacteria. The three most abundant phyla of bacteria observed by analysis in nose,
12
13 mouth, throat, bronchial and sputum were Actinobacteria, Fusobacteria and Bacteroidetes
14
15 (seen in Figure 1). Actinobacteria dominated in the nose (73% of all total species) and
16
17 mouth (37%) samples, Fusobacteria dominated in throat (31%) and sputum (38%) samples
18
19 and Proteobacteria were dominant in bronchial samples (34%). Both nasal and bronchial
20
21 samples had more unclassified OTUs (12% and 41% respectively) than other samples. The
22
23 relative abundance of each individual sample is shown in Figure 2. Figure S1 presents the
24
25 phylogenetic tree of the microbial community in all analysed samples. In two of the control
26
27 samples there was no identifiable bacteria DNA signal however some DNA signal was
28
29 detected in the one of the three control and this was taken into account during analysis (See
30
31 Figure 2).
32
33
34
35
36
37

38 **Comparison between complete community structures within samples from different sites**

39
40
41 The rarefaction analysis suggested that 2000 sequences would be sufficient for
42
43 characterizing the majority of bacteria present in these communities. The numbers of reads
44
45 were lowest in bacterial communities from mouth and bronchus (Figure 3; Table S2). A
46
47 comparison of individual bacterial communities, using PERMANOVA principal coordinate
48
49 analysis (PCoA), showed a distinct clustering by sample cohorts based roughly on where the
50
51 sample was taken from. Bacterial communities in samples from the mouth, throat and
52
53 induced sputum were tightly clustered together ($p_{adj}=1$) and were distinct from those of
54
55
56
57
58
59
60

1
2
3 the nose and bronchus ($p_{adj} < 0.001$), which independently clustered (Figure 4; Table 2);
4
5 these clustering results were seen when children with asthma and without asthma were
6
7 considered separately (Table S4), Table S5 demonstrates that clustering of bacterial
8
9 communities was not evident between sites among subgroups stratified by asthma/not
10
11 asthma or atopy/not atopy.
12
13

14
15 The bacterial communities in the five sites sampled did not differ between the children with
16
17 and without asthma; there was evidence of some differences in bacterial abundance and
18
19 diversity between groups (see supplement Figure S2). A comparison of the abundant phyla
20
21 of bacteria identified in bronchial samples taken from asthma and non-asthma patients
22
23 showed the profiles were similar, but the abundance of a signal for phyla was different (but
24
25 did not reach significance). In bronchial samples from asthma patients, the abundant phyla
26
27 identified was Proteobacteria while those samples taken from non-asthma patients showed
28
29 a greater abundance for Fusobacteria.
30
31
32
33
34
35
36
37

38 **DISCUSSION**

39
40
41 This study was designed to determine whether induced sputum is a valid alternative to
42
43 directly sampling of the lower airways to study airway microbiota in children. The main
44
45 finding was that the characteristics of the microbiota in induced sputum and in bronchial
46
47 samples were different. A second finding was that a minority of bacterial DNA in bronchial
48
49 samples was unclassified. In two individuals (P12 and P16, see Figure 2) the proportion of
50
51 phyla identified from the bronchial sample was similar to the sputum sample but for the
52
53 remainder of participants there were clear differences in the proportions of different phyla
54
55
56
57
58
59
60

1
2
3 from bronchial and sputum samples. Together these findings suggest that induced sputum is
4
5 not a valid surrogate for direct bronchial sampling to study the lower airway microbiota in
6
7 all children, but that there is still a pressing need to better understand the lower airway
8
9 microbiota in children. Opportunistic bronchial sampling under clinically-indicated general
10
11 anaesthetic should remain the gold standard for studying lower airway microbiome in
12
13 children.
14
15

16
17 To our knowledge there are only three studies which have compared the microbiota of
18
19 upper and lower airway secretions in healthy children or children with asthma⁶⁻⁸. A study
20
21 where the majority of recruits were adults showed that the bacterial community in the
22
23 oropharynx and bronchus were similar in healthy individuals⁶. One solely paediatric study
24
25 found very little overlap between the microbiota of the oropharynx and lower airways⁷
26
27 whilst a second paediatric study found similar predominant phyla in upper and lower airway
28
29 samples but different genera⁸. Both paediatric studies found significant differences in the
30
31 clustering of bacterial communities in the upper and lower airways^{7,8}. The dominant phyla in
32
33 bronchial samples is not consistent between studies but there is consistency for upper
34
35 airway samples. For example, the microbiota of the upper airways is predominated by
36
37 bacteria from the Firmicutes (e.g. Staphylococcus and Streptococcus) and the Actinobacteria
38
39 phyla (e.g. Corynebacteria) in our study and three others⁶⁻⁸. In contrast and in bronchial
40
41 samples, bacterial species within the Proteobacteria phylum (including Moraxella and
42
43 Haemophilus) predominated in our study and one other⁷ but Firmicutes dominated in a
44
45 third study⁸. Differences in lower airway microbiome between studies may be due to
46
47 different sampling methods⁸ and treatment with current or recent antibiotics^{7,8}. Our results,
48
49 where we find no differences between the microbiome of induced sputum and throat swab,
50
51
52
53
54
55
56
57
58
59
60

1
2
3 are also consistent with previous studies which suggested that a throat swab may be a
4
5 useful surrogate of induced sputum in young children with CF^{9,10}. As our study did not
6
7 include children with CF we cannot comment on the validity of induced sputum as an index
8
9 of lower airway infection in this specific clinical setting. The consistency of our results with
10
11 previous studies assures us that our findings are valid, despite the relatively small sample
12
13 size.
14
15

16
17
18 Our study findings were based on samples taken from a group of children who were well.
19
20 Recent work has given insight into the relationship between airway microbiota in the upper
21
22 airways of children with acute infection, with results which differ from the present study.
23
24 For example there is evidence that in the context of acute upper and lower respiratory tract
25
26 infection, different bacterial species predominate in the nasopharynx (e.g. Proteobacteria
27
28 and Firmicutes which contrast with Actinobacteria and Fusobacteria in our study)¹⁵. The
29
30 characteristics of the nasopharyngeal microbiota also change during the course of acute
31
32 otitis media infection¹⁶.
33
34
35

36
37 Our study design minimised the risk of false positive results arising from contamination and
38
39 from samples containing low DNA yields. There is no consensus on the number of control
40
41 samples which should be collected for microbiome studies, and two of the three controls we
42
43 collected contained no bacterial DNA and a third contained only bacteria with a very
44
45 different profile of phyla compared to the samples from study subjects. We set out to have
46
47 a high threshold for reads (>2000) and this filtered out samples with low DNA yields.
48
49 Another source of potential contamination comes from reagents used to analyse the
50
51 samples and this risk was minimised by taking standard precautions were taken and by
52
53 including control samples in the analysis. Whilst we cannot exclude the possibility that
54
55
56
57
58
59
60

1
2
3 some contamination may have occurred, our robust methodology and the consistency of
4
5 our findings with previous studies assure us that any contamination has not substantially
6
7 affected our findings.
8
9

10
11 Finding a small proportion of unassigned bacterial DNA in the nose and bronchial samples
12
13 was unexpected and worthy of further investigation. The presence of “unclassified
14
15 bacteria” has been described in nasopharyngeal samples from children⁷, but not in bronchial
16
17 samples. We have carefully reviewed the unassigned sequences and removed human DNA
18
19 sequences, and we therefore believe that there are a number of bacteria in the respiratory
20
21 tract which are not identified by standard microbiota methodologies currently used. We
22
23 analysed data from a mock community (a “positive control”) and this demonstrated that the
24
25 Qiagen software used for analysis of our samples was able to detect bacteria in similar
26
27 proportions to another software “pipeline”. The Qiagen software identified all phyla
28
29 whereas the alternative software could not assign 15% of bacterial DNA and this may be
30
31 explained by the higher stringency set for the Qiagen software. There is not standard for
32
33 stringency but these results suggest that although the higher stringency reduces the number
34
35 of reads included in the analysis, the results are highly comparable to other methods and
36
37 possibly less affected by potential contaminants. Our data are publicly available and
38
39 colleagues are welcome to apply different stringencies to our data.
40
41
42
43
44
45

46
47 There are a number of settings where upper airway samples have been shown to be valid
48
49 surrogates of lower airway samples for the clinical diagnosis of infective and non-infective
50
51 conditions. Viral aetiology in bronchiolitis (a lower respiratory tract infection) can be
52
53 established by nasopharyngeal aspirate¹⁶ and the nasal mucosa is used for diagnosis of
54
55 primary ciliary dyskinesia¹⁷ and (in some situations) CF¹⁸. In the research setting, sampling
56
57
58
59
60

1
2
3 from the upper airways has been validated as a surrogate for lower airway sampling^{19,20}.
4
5 However, we demonstrated that the microbiota of the nose was distinct to mouth and
6
7 bronchus in most individuals and our results indicating that the microbiota of the upper
8
9 airway is probably not a valid surrogate of lower airway microbiota in children. Previous
10
11 papers have suggested that with regards to lower airway samples, upper airway samples are
12
13 “imperfect but reliable”⁷ and “both similar and different”⁸ and overall, our work finds more
14
15 evidence for the upper airways samples being “imperfect and different” to lower airway
16
17 samples and not “reliable and similar” . .
18
19
20
21

22 The novelty of our study is that we demonstrate how the microbiota of induced sputum
23
24 differs from bronchial samples and very closely matches the microbiota of mouth and
25
26 throat, most likely due to contamination of lower airway secretions as they pass through the
27
28 oropharynx, including sputum which is commonly mixed with expectorant. Our findings are
29
30 consistent with a study of 78 children (mean age 2 years, almost half having received recent
31
32 antibiotic treatment) which also describes differences in the microbiota between nose and
33
34 mouth and between mouth and bronchial fluid⁷. Consistent with our findings, a study of
35
36 adults and children (mean age 11 years) reports Actinobacteria being the predominant
37
38 phyla in the nose but being almost absent from the oropharynx and bronchus. A different
39
40 pattern is shown for Proteobacteria which are increasing from nose through to oropharynx
41
42 and bronchus⁶. The apparent presence of different dominant bacterial communities in the
43
44 upper and lower airways raises the question “where do bacteria in the lower airways
45
46 originate?”
47
48
49
50
51
52

53 In children the lower airways are in direct communication with and adjacent to the
54
55 oropharynx but there are several mechanisms that may lead to the establishment of
56
57
58
59
60

1
2
3 different bacterial communities above and below the vocal cords. Perhaps most obviously,
4
5 a child's mouth is regularly exposed to food and inedible items introduced to the oral cavity
6
7 (including cutlery and fingers) whereas the lower airways are exposed only to inhaled
8
9 exposures. Micro aspiration due to laryngopharyngeal reflux is thought to occur on a
10
11 regular basis²¹ and exposure to acid and other gastric contents (including bacteria) may
12
13 affect the microbiota of the lower airways but not the oropharynx. There are differences in
14
15 the innate defences of the upper and lower airways which may also explain differences in
16
17 their respective microbiota, for example lactoferrin concentrations are twice as high in the
18
19 lower airways relative to upper airways²² whereas nitric oxide concentrations (known to
20
21 have antimicrobial properties²³) are typically one hundred times higher in the upper
22
23 airways.
24
25
26
27
28

29 Our study was not designed to relate microbiota to clinical phenotypic data collected, for
30
31 example asthma or atopy. However, in light of recent studies examining these parameters in
32
33 children with asthma^{6,7} we made the best use of the data available, even with the limited
34
35 number we had at our disposal. When examined the possible relationship between the
36
37 bacteria present in children with and without asthma or atopy, we found no statistically
38
39 differences between the bacteria when the samples were examined as a whole group or as
40
41 specific sample region cohorts (nose, mouth, throat, bronchial and sputum samples), Figure
42
43 S2 and Table S3. A descriptive analysis of the bronchial samples collected found
44
45 Protoeobacteria to be the dominant species and this is consistent with Hilty *et al*⁶.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 There are some limitations to our study. First, the number of participants was relatively
4
5 small. Second, induced sputum was not obtained in all participants and not all samples met
6
7 our quality control criteria for bacterial DNA analysis and this resulted in missing data for
8
9 some individuals. Third, as has been reported in other sputum microbiome studies^{6,7}, we
10
11 were not able to perform cell counts on the sputum pellet to determine the presence of
12
13 neutrophils and absence of squamous epithelial cells. Finally, we did not prospectively
14
15 calculate the DNA yield from samples although the consistency of our results with the small
16
17 number of published studies^{6,7}, the consistency of results using both the Qiagen and EBI
18
19 Metagenomics pipelines and the comparable results from mock communities assure is that
20
21 low DNA yield has not altered the results.
22
23
24

25
26
27 In summary, we report that neither induced sputum nor swabs from nose or mouth give an
28
29 accurate indication of the bronchial microbiota in all children. For very obvious practical
30
31 and ethical reasons, upper airway sampling will be necessary to study airway microbiome in
32
33 many research setting and our findings adds to the evidence describing the limitations of
34
35 this pragmatic approach. Our study was not designed to compare differences in the lower
36
37 airway microbiota of children with and without asthma, and future research is required to
38
39 supplement the little we do know about this potentially important subject.
40
41
42
43
44
45
46
47
48
49

50 **ACKNOWLEDGEMENTS**

51
52
53 The authors would like to thank the study participants and their parents. The study was
54
55 supported by an NHS-Grampian Endowments. AW is supported by the Wellcome Trust
56
57
58
59
60

1
2
3 Strategic Award (grant 097377) and the MRC Centre for Medical Mycology (grant
4
5 MR/N006364/1) at the University of Aberdeen.
6
7
8
9

10
11 **Competing interest statement.** None of the authors has any competing interests.
12
13

14 REFERENCES

- 15
16
17
18 1. Kahn FW, Jones JM. Diagnosing bacterial respiratory infection by bronchoalveolar
19 lavage. *J Infect Dis* 1987; **155**(5): 862-9.
20
21 2. Marsland BJ, Yadava K, Nicod LP. The airway microbiome and disease. *Chest* 2013;
22
23 **144**(2): 632-7.
24
25
26 3. Rogers GB, Shaw D, Marsh RL, Carroll MP, Serisier DJ, Bruce KD. Respiratory microbiota:
27 addressing clinical questions, informing clinical practice. *Thorax* 2015; **70**(1): 74-81.
28
29
30 4. Klepac-Ceraj V, Lemon KP, Martin TR, Allgier M, Kembel SW, Knapp AA, Lory S, Brodie
31 EL, Lynch SV, et al. Relationship between cystic fibrosis respiratory tract bacterial
32 communities and age, genotype, antibiotics and *Pseudomonas aeruginosa*. *Environ*
33
34
35
36
37
38
39
40
41
42
43 5. Bisgaard H, Hermansen MN, Bonnelykke K, Stokholm J, Baty F, Skytt NL, Aniscenko J,
44 Kebabdzic T, Johnston SL. Association of bacteria and viruses with wheezy episodes in young
45 children: prospective birth cohort study. *BMJ* 2010; **341**: 4978.
46
47
48
49
50 6. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L,
51
52
53
54
55
56
57
58
59
60
Pachter L, et al. Disordered microbial communities in asthmatic airways. *PLoS ONE* 2010;
5(1): e8578.

- 1
2
3 7. Marsh RL, Kaestli M, Chang AB, Binks MJ, Pope CE, Hoffman LR, Smith-Vaughan HC. The
4 microbiota in bronchoalveolar lavage from young children with chronic lung disease
5 includes taxa present in both the oropharynx and nasopharynx. *Microbiome* 2016; **4**: 37.
6
7
8
9
10
11
12
13
14 8. Kloepfer KM, Deschamp AR, Ross SE, Peterson-Carmichale SL, Hemmerich CM, Rusch DB,
15 Davies SD. In children, the microbiota of the nasopharynx and bronchoalveolar lavage fluid
16 are both similar and different. *Ped Pulmonol* 2018; DOI: 10.1002/ppul.23953
17
18
19
20
21
22 9. Zemanick ET, Wagner BD, Robertson CE, Stevens MG, Szeffler SJ, Accurso FJ, Sagel SD,
23 Harris JK. Assessment of airway microbiota and inflammation in cystic fibrosis using multiple
24 sampling methods. *Ann Am Thor Soc* 2015; **12**(2): 221-9.
25
26
27
28
29
30 10. Boutin S, Graeber SY, Weitnauer M, Panitz J, Stahl M, Clausznitzer D, Kaderali L,
31 Einarsson G, Tunney MM, Elborn S, et al. Comparison of microbiomes from different niches
32 of upper and lower airways in children and adolescents with cystic fibrosis. *PLoS ONE* 2015;
33 **10**(1): e0116029.
34
35
36
37
38
39
40 11. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P,
41 van der Grinten CP, Gustaffson P, et al. Standardisation of spirometry. *Eur Respir J* 2005;
42 **26**(2): 319-38.
43
44
45
46
47
48 12. Pontin J, Blaylock MG, Walsh GM, Turner SW. Sputum eosinophil apoptotic rate is
49 positively correlated to exhaled nitric oxide in children. *Pediatr Pulmonol* 2008; **43**(11):
50 1130-4.
51
52
53
54
55
56
57
58
59
60

- 1
2
3 13. Twomey KB, Alston M, An S, O'Connell OJ, McCarthy Y, Swarbreck D, Febrer M, Dow
4
5 JW, Plant BJ, Ryan RP. Microbiota and metabolite profiling reveal specific alterations in
6
7 bacterial community structure and environment in the cystic fibrosis airway during
8
9 exacerbation. *PLoS ONE* 2013; **8**(12): e82432.
10
11
12
13 14. Moen AEF, Tannaes TM, Vatn S, Ricanek P, Vatn MH, Jahnsen J, IBD-Character
14
15 Consortium. Simultaneous purification of DNA and RNA from microbiota in a single colonic
16
17 mucosal biopsy. *BMC Research Notes* 2016; **9**: 328.
18
19
20
21 15. Zakharkina T, Martin-Loeches I, Matamoros S, Pova P, Torres A, Kastelijn JB, Hofstra
22
23 JJ, de Wever B, de Jong M, Schultz MJ, et al. The dynamics of the pulmonary microbiome
24
25 during mechanical ventilation in the intensive care unit and the association with occurrence
26
27 of pneumonia. *Thorax* 2017; **72**: 803-10.
28
29
30
31 16. Kelly MS, Surette MG, Smieja M, Pernica JM, Rossi L, Lunistra, Steenhoff AP, Feemster
32
33 KA, Goldfarb DM, Arscott-Mills T, Bioditswe S, Rulaganyang I et al. The Nasopharyngeal
34
35 Microbiota of Children with Respiratory Infections in Botswana. *Pediatr Infect Dis J*
36
37 2017;36:e211-e218
38
39
40
41
42
43
44
45 17. Chonmaitree T, Jennings K, Golovko G, Khanipov K, Pimenova M, Patel JA, McCormick
46
47 DP, Loeffelholz MJ, Fofanov, Y. Nasopharyngeal microbiota in infants and changes during
48
49 viral upper respiratory tract infection and acute otitis media. *PLoS ONE* 2017;12
50
51 (7):e0170630.
52
53
54
55
56
57
58
59
60

- 1
2
3 18. Ralston SL, Lieberthal AS, Meissner HC. Ralston SL, Lieberthal AS, Meissner HC,
4
5 Alverson BK, Baley JE, Gadomski AM, Johnson DW, et al. Clinical Practice Guideline: The
6
7 Diagnosis, Management, and Prevention of Bronchiolitis. *Pediatrics*. 2014;134(5):e1474-
8
9 e1502. *Pediatrics* 2015; **136**(4): 782.
- 10
11
12
13 19. Barbato A, Frischer T, Kuehni CE, Snijders D, Azevedo I, Bakati G, Bartoloni L, Escribano
14
15 A, Haarman E, Messelmar B, et al. Primary ciliary dyskinesia: a consensus statement on
16
17 diagnostic and treatment approaches in children. *Eur Respir J* 2009; **34**(6): 1264-76.
- 18
19
20
21 20. Farrell PM, Rosenstein BJ, White TB, Accurso FJ, Castellani C, Cutting GR, Durie PR,
22
23 Legrys VA, Massie J, Parad RB, et al. Guidelines for diagnosis of cystic fibrosis in newborns
24
25 through older adults: Cystic Fibrosis Foundation consensus report. *J Pediatr* 2008; **153**(2):
26
27 S4-S14.
- 28
29
30
31 21. Pringle EJ, Richardson HB, Miller D, Cornish DS, Devereux GS, Walsh GM, Turner SW.
32
33 Nasal and bronchial airway epithelial cell mediator release in children. *Pediatr Pulmonol*
34
35 2012; **47**(12): 1215-25.
- 36
37
38
39 22. Pitrez PMC, Brennan S, Turner S, Sly PD. Nasal wash as an alternative to
40
41 bronchoalveolar lavage in detecting early pulmonary inflammation in children with cystic
42
43 fibrosis. *Respirology* 2005; **10**(2): 177-82.
- 44
45
46
47 23. Venkatesan NN, Pine HS, Underbrink M. Laryngopharyngeal reflux disease in children.
48
49 *Pediatr Clin North Am* 2013; **60**(4): 865-78.
- 50
51
52
53
54
55
56
57
58
59
60

1
2
3 24. Roca-Ferrer J, Mullol J, Perez M, Xaubet A, Momlins L, de Haro J, Shelhamer J, Picado
4
5 C. Effects of topical glucocorticoids on in vitro lactoferrin glandular secretion: comparison
6
7 between human upper and lower airways. *J Allergy Clin Immunol* 2000; **106**(6): 1053-62.
8
9

10
11 24. Fang FC. Antimicrobial reactive oxygen and nitrogen species: concepts and
12
13 controversies. *Nature Rev Microbiol* 2004; **2**(10): 820-32.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

FIGURE LEGENDS

Figure 1. Relative abundances of bacterial phyla identified as operational taxonomic units (OTUs) from the sequence reads generated from airway samples taken from children. The bar chart illustrates the taxonomic composition of each cohort of samples from a particular site. A detailed summary of the bacteria in each sample is described in Figure 2.

Figure 2. Operational Taxonomic Units based relative sequence abundance of bacterial phyla based on 16S rRNA of all 75 samples. A detailed summary of the bacteria in each sample is described. On the horizontal axis, the number following P is the patient number and NO=nose, MO=mouth, TH=throat, SP=Sputum, BR=bronchial sample. C2=the control sample where bacterial DNA was detected. N/A=not applicable

Figure 3. Alpha diversity rarefaction curves of samples based on total number of observed Operational Taxonomic Units. Colour denotes different sample cohort (“nose” (NO), “mouth” (MO), “throat” (TH), “sputum” (SP) and “bronchial” (BR) samples).

Figure 4. Diversity analysis demonstrating differences in the bacterial phyla community between samples taken from airway samples taken from asthmatic and non-asthmatic children. Principal coordinate (PCo) analysis of all samples based on D_{0.5} UniFrac distance. Colour denotes different sample cohort (“nose” (NO), “mouth” (MO), “throat” (TH), “sputum” (SP) and “bronchial” (BR) samples). Y=sample from child with asthma.

Figure S1. Phylogenetic tree of the microbial community in all analysed samples. All Operational Taxonomic Units are shown in this figure.

Figure S2. Bacterial phyla abundance and diversity between groups. On the horizontal axis, the first letter “Y” indicates that the child had asthma and “N” that they did not have

1
2
3 asthma and NO=nose, MO=mouth, TH=throat, SP=Sputum, BR=bronchial sample. N/A=not
4
5 applicable
6
7

8 **TABLE LEGENDS**

9

10
11 Table 1. Details of the children recruited. SABA=short acting beta agonist, ICS=inhaled
12
13 corticosteroids, LABA=long acting beta agonist
14
15

16
17 Table 2. PERMANOVA analysis results of testing differences in beta-diversity among
18
19 different sample sites.
20

21
22 Table S1. Summary of sequence data used for Operational Taxonomic Units (OTU) clustering
23
24 and analysis. The number of filtered reads and reads in OTUs does not include the reads
25
26 excluded by quality control
27

28
29 Table S2. List of sequence data for individual samples used for Operational Taxonomic Units
30
31 (OTU) clustering and analysis.
32
33

34
35 Table S3. Comparison of results from two pipelines used to analyse data. The Qiagen
36
37 pipeline was used for the present study. The EBA pipeline (full name EBI Metagenomics
38
39 analysis pipeline V3.0) was used as a comparator
40
41

42
43 Table S4. PERMANOVA analysis separately comparing beta-diversity among different sample
44
45 sets for children with asthma and without asthma.
46

47
48 Table S5. PERMANOVA analysis comparing beta-diversity among different sample sets.
49

50
51 Atopy was defined as serum IgE >100 kU/l.
52
53
54
55
56
57
58
59
60

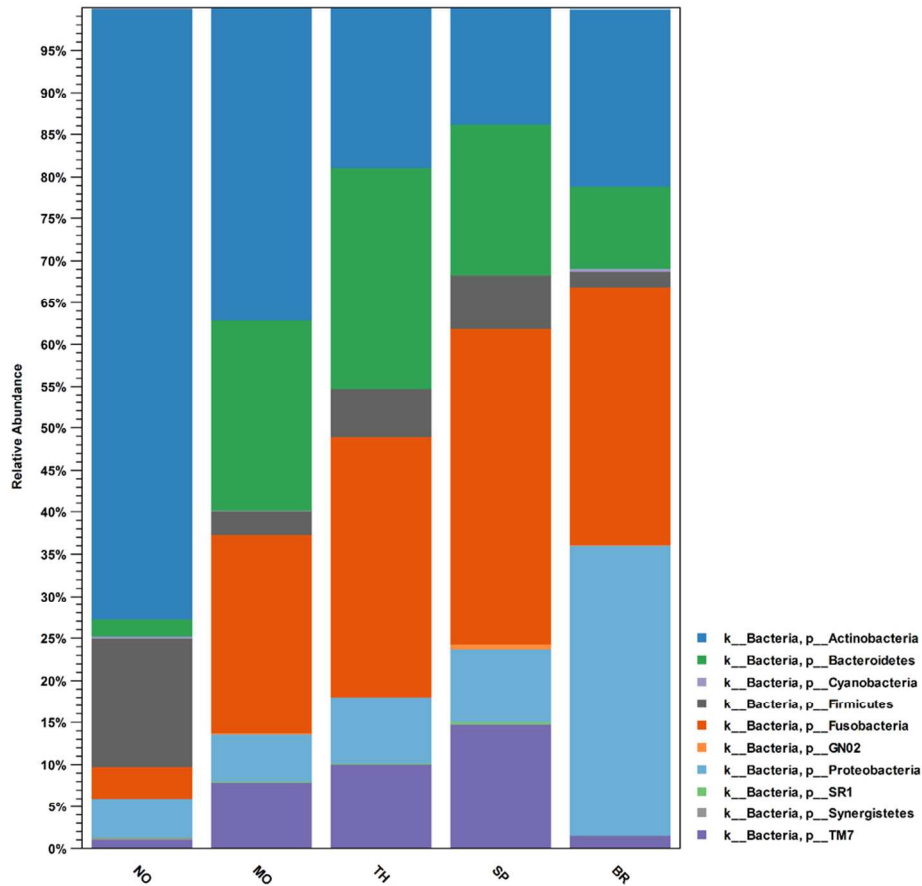


Figure 1. Relative abundances of bacterial phyla identified as operational taxonomic units (OTUs) from the sequence reads generated from airway samples taken from children. The bar chart illustrates the taxonomic composition of each cohort of samples from a particular site. A detailed summary of the bacteria in each sample is described in Figure 2.

193x175mm (150 x 150 DPI)

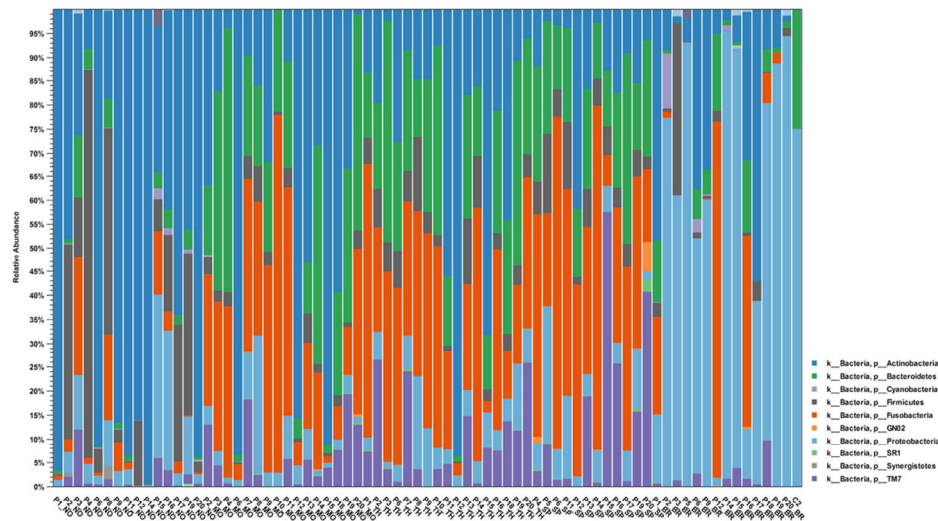


Figure 2. Operational Taxonomic Units based relative sequence abundance of bacterial phyla based on 16S rRNA of all 75 samples. A detailed summary of the bacteria in each sample is described. On the horizontal axis, the number following P is the patient number and NO=nose, MO=mouth, TH=throat, SP=Sputum, BR=bronchial sample. C2=the control sample where bacterial DNA was detected. N/A=not applicable

338x190mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

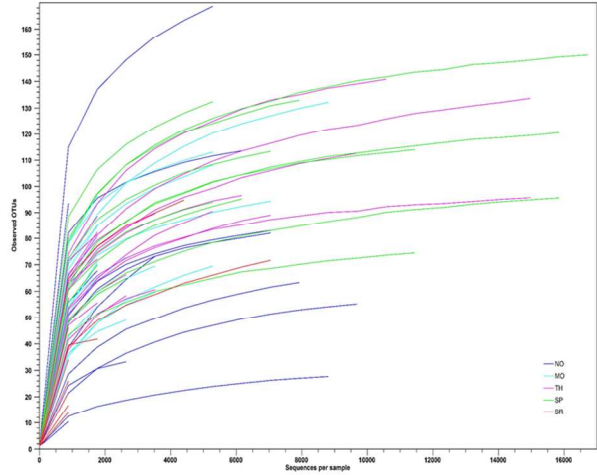


Figure 3. Alpha diversity rarefaction curves of samples based on total number of observed Operational Taxonomic Units. Colour denotes different sample cohort ("nose" (NO), "mouth" (MO), "throat" (TH), "sputum" (SP) and "bronchial" (BR) samples).

338x190mm (96 x 96 DPI)

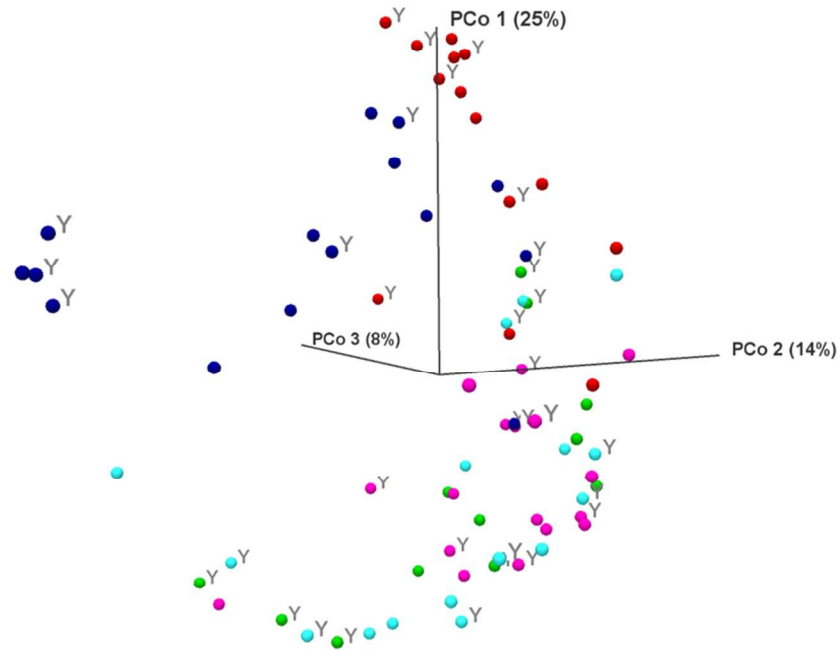


Figure 4. Diversity analysis demonstrating differences in the bacterial phyla community between samples taken from airway samples taken from asthmatic and non-asthmatic children. Principal coordinate (PCo) analysis of all samples based on D_{0.5} UniFrac distance. Colour denotes different sample cohort ("nose" (NO), "mouth" (MO), "throat" (TH), "sputum" (SP) and "bronchial" (BR) samples). Y=sample from child with asthma.

254x190mm (96 x 96 DPI)

Table 1. Details of the children recruited. SABA=short acting beta agonist, ICS=inhaled corticosteroids, LABA=long acting beta agonist

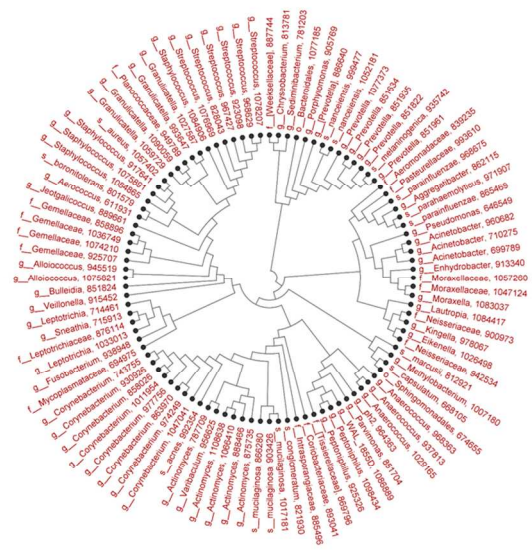
| | | All children (n=20 unless stated) | Children with asthma (n=7 unless stated) | Children without asthma (n=13 unless stated) |
|--|-------------------|--------------------------------------|---|--|
| Mean age (SD), y | | 10.7 (2.8) | 12.1 (2.2) | 9.9 (2.9) |
| Proportion male (number) | | 40% (8) | 43% (3) | 39% (5) |
| Mean % FEV ₁ /FVC ratio (SD) | | 103% (8) n=19 | 101% (9.5) | 104% (8) n=12 |
| Median plasma IgE (SEM), kU/l | | 118 (36) n=19 | 193 (48) n=6 | 82 (49) |
| Mean BMI centile (SD) | | 70% (28) | 64% (28) | 73% (28) |
| Proportion where the sample met the quality control criteria for analysis (n) | Bronchial fluid | 70% (14) | 86% (6) | 62% (8) |
| | Induced sputum | 60% (12) | 86% (6) | 46% (6) |
| | Throat swab | 85% (17) | 100% | 77% (10) |
| | Mouth swab | 80% (16) | 100% | 69% (9) |
| | Nose swab | 75% (15) | 100% (7) | 62% (8) |
| Asthma treatment | | | 2 SABA only 2 ICS plus SABA 3 ICS, LABA plus SABA | |

Table 2. PERMANOVA analysis results of testing differences in beta-diversity among different sample sites. Paired samples were not available in all individuals

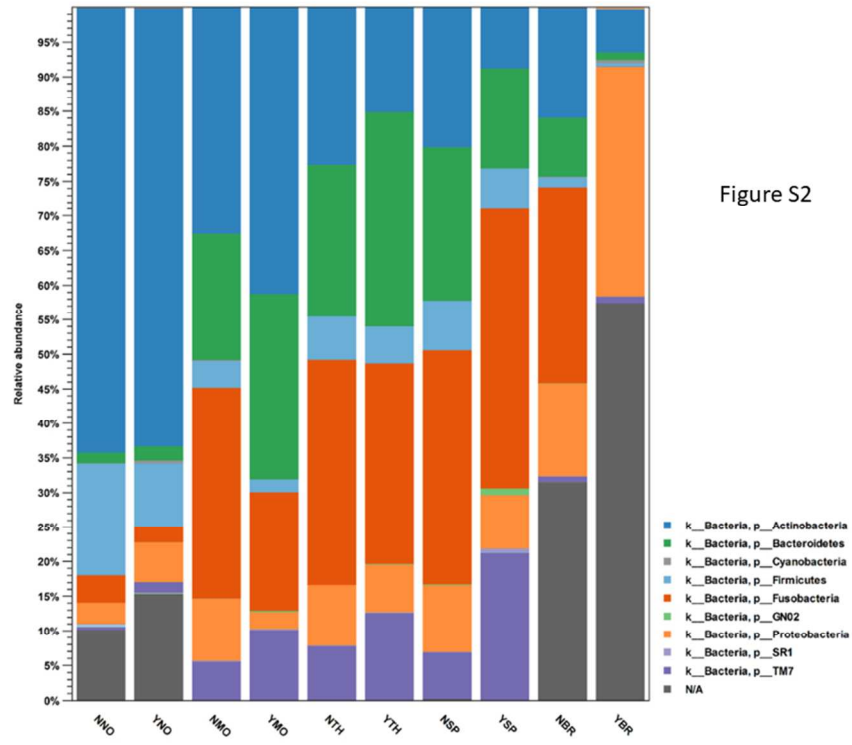
| Sample site 1 | Sample site 2 | Number of individuals with comparable data | Number of samples included in the analysis (site 1&site 2) | p value (unadjusted) | p value (adjusted for multiple testing) |
|---------------|---------------|--|--|----------------------|---|
| Nose | Mouth | 11 | <u>15&16</u> | 0.00001 | 0.00015 |
| Nose | Throat | 13 | <u>15&17</u> | 0.00001 | 0.00015 |
| Mouth | Throat | 14 | <u>16&17</u> | 0.22171 | 1 |
| Nose | Sputum | 10 | <u>15&11</u> | 0.00001 | 0.00015 |
| Mouth | Sputum | 10 | <u>16&11</u> | 0.17806 | 1 |
| Throat | Sputum | 10 | <u>17&11</u> | 0.27851 | 1 |
| Nose | Bronchus | 12 | <u>15&14</u> | 0.00002 | 0.0003 |
| Mouth | Bronchus | 9 | <u>16&14</u> | 0.00001 | 0.00015 |
| Throat | Bronchus | 12 | <u>17&14</u> | 0.00001 | 0.00015 |
| Sputum | Bronchus | 10 | <u>11&14</u> | 0.00003 | 0.00045 |

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure S1



338x190mm (96 x 96 DPI)



254x190mm (96 x 96 DPI)

Whereas some minor issues have been addressed major issues are still not sufficiently replied to. I am still not convinced that one the main messages, namely that bronchial samples differ significantly from induced sputum, is justified by the data and therefore allows to conclude: "Induced sputum may not be a valid surrogate for microbiome assessment of the lower airways."

Before going into a deeper analysis I would like to mention that a considerable part of confusion comes due to erroneous labeling and numbering of the figures and tables. E.g. in my previous review I referred in point 4 to Fig. S2 which according to the legends was the phylogenetic tree (yet the Figure itself had no label) and claimed the absence of Streptococci and Haemophilus. In the reply the authors now refer to bar graphs that of course only show phyla. This is not a reply to my question. Moreover the current manuscript is still superficial in preparation: Fig. S2 as labeled in the legends ("Figure S2. Bacterial phyla abundance and diversity between groups") is shown as Fig. S3 on p.30. In Fig. 1 the legend gives a note "A detailed summary of the bacteria in each sample is described in Figure S1." but this is now Fig. 2. Therefore I now always refer to the page number indicated at the top of the pdf file.

Our response: We apologise for any confusion caused. On the advice of the reviewer we moved one figure from the supplement to the main paper. Therefore the numbering of figures in the supplement changed between the first and the revised manuscript and we failed to capture all of these changes. There are four specific points raised by the reviewer here:

1. Incorrect legend for Figure 1 - We have corrected the error in the legend for Figure 1 (i.e. changed "Figure S1" to "Figure 2").
2. Reply to point 4 in initial review - The reviewer did not refer to Figure S2 in point 4 of their previous review (they did refer to this in point 7). We do not know which question the reviewer invites us to reply to.
3. No label for phylogenetic tree in supplement - We have inserted a label for Figure S1.
4. Figure S2 labelled as Figure S3 - We have corrected this error.

Severe technical shortcomings remain and these are again outlined:

1. *It is not sufficiently clarified how the analysis results in such an enormous loss of reads: Previously of 21.5 million reads only 400,493 were analysed, now it reads that from 11.7 million reads 606,227 could be analysed. Why were now fewer reads available? Why are less than 10% analysed? This is a very unusual yield for 16S sequencing. The authors refer to "high stringency" and p.7 for further explanations in their response but this stringency is nowhere explained. I am afraid that important information is missed.*

Our response: There are actually now more reads and a smaller proportion of "lost" reads included in the analysis reported in the revised manuscript (606,227 compared to 400,493). There is no standard for stringency and, as the reviewer will be aware, the advantage of a higher cut off reduces the potential for "false positive" findings. In this study, we used the default settings in the workbench software (Qiagen). This has a very "high stringency" cut off which means that a lot of reads are removed. Our data are publicly available and colleagues can apply different stringency thresholds to our data. We have expanded the text on page 7 to explain our approach to stringency which explains the threshold used. Finally (as suggested by the reviewer) we have used a mock community to replicate data from our pipeline using second pipeline which applied a different stringency; this last analysis indicates that our stringency has not affected the diversity of microbiome detected but has increased the accuracy of detection (data now included in Table S2).

2. *On "p.29 of 62" the authors again show the phylogenetic tree that should show "all analysed samples". In this tree I do neither see the genus Streptococcus which should be found in all healthy individuals at least in the mouth and throat or Haemophilus which has been reported as a typical proteobacterium in asthma patients. I come to the same conclusion as in (1.) that important*

1
2
3 information might be missing. That is why I requested to analyze a defined mock community but the
4 authors refuse to do so. I do not see why this should not be possible even if the study was closed.
5

6 **Our response:** This phylogenetic tree shows details of phyla and does not provide details of species.
7 Therefore species such as Streptococcus and Haemophilus are not shown. The analysis of mock
8 community indicates that our pipeline more sensitive than a second pipeline and that important
9 information is not missing. We have provided more labels on the phylogenetic tree identifying more
10 phyla.
11
12

13 3. I previously stated that I am afraid that in bronchial samples the bacterial yield might have
14 been low and therefore prone to contamination. Indeed the new Fig. 2 clearly shows that in the one
15 control sample proteobacteria were brought in. I am afraid that in samples with known low yield
16 (bronchial washing from healthy individuals) this might induce an error as previously reported (Salter
17 et al. BMC Biology 2014, 12:87). I therefore requested to do an analysis of bacterial load (e.g.
18 quantitative 16S PCR). The authors now refer me to Table S2 but this is entirely misleading. "Number
19 of reads" has nothing to do with the original bacterial load. During the pipeline of mixing and
20 barcoding for the sequencing process typically equal amounts of PCR products are taken. After the
21 sequencing the number of reads only reflects variation in the sequencing and does not relate to the
22 original bacterial load. Without seeing true data on bacterial load the issue of contamination cannot
23 be discussed in sufficient depth.
24
25

26 **Our response:** We believe that the mock community analysis described previously addresses the
27 issue of low DNA yield in bronchial samples (see Table S2). There was not sufficient sample volume
28 to determine the bacterial yield. The mock community analysis and consistency with published data
29 assure us that low bacterial load in bronchial samples has not substantially affected our results. We
30 have added text acknowledging that we did not calculate DNA yield. We have also added a very
31 recently published paper which also finds different clustering of bacterial communities in the upper
32 and lower airways.
33

34 4. From Table S2 it is visible that for the bronchial samples a number of samples had to be
35 excluded because that did not give enough reads. I am afraid that the majority were from healthy
36 children whereas those from asthmatic were analysed. The latter are known to have a higher
37 Proteobacteria load. Therefore it is mandatory, to show the disease status in Fig. 2. Moreover, I
38 would like to see separate PERMANOVAs for diseased and healthy children with respect to
39 differences between sputa and bronchial washings. This is not covered so far by any of the analysis
40 and from the Fig. on "p. 30 of 62" (mislabelled as S3) it becomes clear that the difference between
41 NBR and NSP is much less as compared to YBR vs. YSP. I do NOT want to make a conclusion on
42 asthma patients but I state that the mixing of healthy and diseased people in on population might be
43 a significant problem. This might be also visible when Fig. 2 would be labeled with the disease status:
44 From a look at Fig. 2 it seems that at least 3-4 samples are much more similar to the throat/sputum.
45 The conclusion of difference between BR and SP is mostly based on the unassigned sequences and for
46 those the data do not sufficiently exclude a contamination issue (see my comment 3).
47
48

49 **Our response:** The reviewer makes a number of points here:

- 50 • Samples need labelling by asthma status in figure 2. We have done this
- 51 • Separate PERMANOVA for children with and without asthma. We have done this (table S4)
- 52 • Difference between BR and SP is mostly based on unassigned sequences. We have now
53 removed unassigned sequences from this figure and the majority of bronchial samples are
54 predominantly the light blue colour corresponding to Proteobacteriae whereas the sputum
55 samples have very few Proteobacteriae seen but are dominated by Fusobacteria and
56
57
58
59
60

1
2
3 Bacteroidites. We do acknowledge that two bronchial samples (P12BR and P16 BR) have
4 few Proteobacteria and have a similar pattern to their corresponding sputum sample.
5 When all patients are considered, there are significant differences between bronchial and
6 sputum samples but we have modified the conclusions to acknowledge that in a minority
7 sputum samples may be a valid index of bronchial samples.
8

9
10 5. *Technically it remains unclear why in Fig. 1 unassigned sequences are shown but in Fig. 2*
11 *excluded. Moreover, on visual inspection there seems to be a discrepancy between proteobacteria*
12 *abundance in Fig. 2 for all BR samples (for many samples >50%) as compared to Fig. 1 (20%). How*
13 *has the data in Fig. 1 been calculated (mean?). All in all the data do not sufficiently justify the main*
14 *conclusion.*

15 **Our response:** We have revised Figure 1 and removed the unassigned phyla. The reviewer is correct
16 and data in Figure 1 are mean values. The proportion of proteobacteria in figure 2 is certainly >50%
17 for most individuals but this figure shows how two individuals (P12BR and p16BR) have very few
18 proteobacteria and results from these two individuals reduce the mean proportion of proteobacteria
19 to the value represented in Figure 1. We have amended the main conclusion (see point 4 above).
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Microbiome Characteristics of Induced Sputum compared to Bronchial Fluid and Upper
4
5 Airway Samples
6
7

8 Shi-qi An^{1,2}, Adilia Warris^{3,4}, Steve Turner³
9

10
11
12 ¹Division of Molecular Microbiology, University of Dundee
13

14
15 ² Wellcome Wolfson Institute for Experimental Medicine, Queen's University Belfast
16
17

18
19 ³ Department of Child Health, University of Aberdeen
20

21
22 ⁴ MRC Centre for Medical Mycology, University of Aberdeen
23
24
25
26
27

28
29 Contact details: Dr Steve Turner, Child Health, Royal Aberdeen Children's Hospital,
30
31 Aberdeen, UK, AB25 2ZG. Tel +44 1224 438470. s.w.turner@abdn.ac.uk
32
33

34 Running head: Microbiota of induced sputum
35
36

37
38 Key words: Bronchial fluid, Child, Microbiota, Sputum
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Objective. The study of the community of microorganisms (the microbiota) in the lower airways in children is restricted to opportunistic sampling in children undergoing elective general anaesthetic. Here we tested the hypothesis that induced sputum is a valid alternative to directly sampling the lower airways to study lower airway microbiota.

Methods. Children scheduled for elective operations were recruited. Pre-operatively a sample of induced sputum was obtained. After anaesthesia was induced, a bronchial brushing and swabs of the upper respiratory tract were obtained. Bacterial community analysis was performed by amplification of the V3-V4 16S rRNA gene region.

Results. Twenty children were recruited, mean age 10.7 years. Induced sputum samples were obtained from ~~induced sputum in~~ 12 children, bronchial brushing ~~from in~~ 14 and , nasal, mouth and throat samples e-in 15, ~~mouth in~~ 16 and ~~throat in~~ 17 children. The profile of bacterial communities was similar in the mouth, throat and sputum samples with the nose and bronchial samples being different. Actinobacteria species dominated the nose and mouth, Fusobacteria were the dominant species in the throat and sputum whilst Proteobacteria species dominated in bronchial samples. Forty-one percent of detected bacteria in bronchial samples were unclassified. Bacterial communities from the mouth, throat and induced sputum were tightly clustered and were distinct from nose and those found in bronchial communities.

Conclusions. Induced sputum may not be a valid surrogate for microbiome assessment of the lower airways in all individuals. ~~A large minority of~~Many bacteria in bronchial samples were not recognised by standard testing, suggesting that our understanding of the lower airway microbiota in children remains rudimentary.

INTRODUCTION

The lower airways were traditionally considered sterile in healthy individuals, and this paradigm was based on observations that standard microbiology culture from samples collected from healthy individuals yielded no pathogens¹. The advent of polymerase chain reaction (PCR) testing for bacterial DNA has now identified communities of bacteria in lower airway secretions from healthy individuals^{2,3}. The relevance of bacteria in airway microbiota to respiratory symptoms is unclear, but there is evidence suggesting that the microbiota may be relevant to influence respiratory exacerbations^{2,4} or even to the aetiology of chronic respiratory conditions such as asthma⁵. Exacerbations of chronic conditions such as cystic fibrosis (CF) and chronic obstructive pulmonary disease are linked to a change² (termed “airway dysbiosis”) or reduction⁴ in the dominant bacterial species in the microbiota. In young infants, those carrying *Haemophilus influenzae* or *Moraxella catarrhalis* in the hypopharynx were more likely to have asthma symptoms at three years of age⁵.

A major challenge to studying the lower airway microbiota is obtaining samples without contamination from the upper airways³, and it is known that the microbiota in the naso- and oropharynx differ from that in the bronchus in adults and children⁶⁻⁸. ~~Three~~ papers have used bronchoscopic sampling to compare describe the upper and lower airway microbiome in children and whilst there are similarities in the dominant phyla identified in some studies, there are differences in genera identified and also in the clustering of bacterial communities⁶⁻⁸. The characteristics of the lower airway microbiome may also vary between individuals depending on their asthma status⁶ and using a nasal or oral approach to the lower airways⁸. ~~reported a greater abundance of Proteobacteria species in those with~~

1
2
3 ~~asthma compared to controls~~^{6,7}. Two studies in young people with CF have compared the
4
5 microbiota in induced sputum and oropharyngeal swabs^{9,10}, and both concluded that a
6
7 throat swab sample may be a valid surrogate for sputum for microbiota analysis. Induced
8
9 sputum is a non-invasive alternative to direct bronchial sampling, and is used in the clinical
10
11 management of tuberculosis and CF in children and also in research. The study of the lower
12
13 airway microbiome would be considerably easier if induced sputum was a valid surrogate of
14
15 bronchial fluid. The aim of the current study was to collect induced sputum and samples
16
17 from the bronchus, nose, throat and mouth in children and undertake a comprehensive
18
19 assessment of upper and lower airway microbiota in children. Our hypothesis was that
20
21 induced sputum is a valid alternative to directly sampling the lower airways to study lower
22
23 airway microbiota.
24
25
26
27
28
29

30 **METHODS**

31 **Study design**

32
33
34
35
36 Children aged 5-16 years and scheduled for elective ear nose and throat operations at Royal
37
38 Aberdeen Children's Hospital between January and June 2015 were invited to take part. All
39
40 children were scheduled for tonsillectomy and some also had grommets inserted. These
41
42 operations require endotracheal intubation, which allows sampling of the lower airways
43
44 without contamination by the upper airways. Children younger than 5 years were excluded
45
46 since they were unlikely to provide a sample of induced sputum. Other exclusion criteria
47
48 included receipt of antibiotics within the last month and CF. In the hours before the
49
50 operation (when the child was fasted), children took part in an assessment which included
51
52 completion of a questionnaire, height and weight measurement, spirometry (in accordance
53
54
55
56
57
58
59
60

1
2
3 with international guidelines¹¹) and induction of sputum using 4% and if required 5% saline
4
5 as previously described¹². An adequate quality sputum sample was defined as the presence
6
7 of a white sputum plug on a visual inspection; samples were not sufficiently large enough to
8
9 allow for cell count analysis. Immediately after anaesthesia was induced and the
10
11 endotracheal tube *in situ*, a sterile 2.7 mm interdental brush (Dento Care Professional,
12
13 London UK) was used to collect a sample of nasal secretions and a separate brush
14
15 interdental brush was gently brushed on the buccal surface of the mouth to collect oral
16
17 secretions. A standard bacterial swab (Transwab, Medical Wire Equipment, Corsham, UK)
18
19 was used to obtain a sample of pharyngeal secretions under direct inspection. Finally, a
20
21 bronchoscopy cytology brush (10-mm disposable cytology brush, BC 202D-2010, Olympus,
22
23 Southend-on-Sea, Essex, UK) was passed down the endotracheal tube until resistance was
24
25 met and then rotated and withdrawn to obtain “blind” bronchial samples. Blood was taken
26
27 for serum total IgE assay. The study was approved by the North of Scotland Research Ethics
28
29 Committee (13/NS/0144), written parental consent was obtained and verbal assent was
30
31 obtained from all participants.
32
33
34
35
36
37
38
39
40

41 **DNA extraction and sequencing**

42
43 The bacterial community analysis was carried out as described previously¹³ with some
44
45 modification as detailed below. Briefly, samples taken from the nose, mouth, throat,
46
47 sputum and bronchial brushing were frozen at -80°C prior to analysis. Samples were then
48
49 transported from Aberdeen to Dundee for processing. Upon arrival the sample had thawed
50
51 and were suspended in phosphate buffer saline where genomic DNA was extracted and
52
53 purified using the DNA/RNA All Prep kit¹⁴ (Qiagen) and stored according to manufacturer’s
54
55
56
57
58
59
60

1
2
3 instructions. Standard protocol, 16S Metagenomic Sequencing Library Preparation Guide
4
5 (Illumina), was followed to prepare sequencing libraries targeting the variable V3 and V4
6
7 regions of the 16S rRNA gene and paired-end sequencing was performed on the MiSeq
8
9 System (Illumina). We followed sampling and controls procedures described by the Earlham
10
11 Institute
12
13
14 (<http://www.earlham.ac.uk/sites/default/files/docs/Genomic%20Services/Sample%20Guide>
15
16 [lines%20Aug17.pdf](http://www.earlham.ac.uk/sites/default/files/docs/Genomic%20Services/Sample%20Guide)) and Illumina
17
18 (https://support.illumina.com/documents/documentation/chemistry_documentation/16s/1
19
20 [6s-metagenomic-library-prep-guide-15044223-b.pdf](https://support.illumina.com/documents/documentation/chemistry_documentation/16s/1)
21
22
23 which are designed to minimize the risk for contamination. Quality controls and sequencing
24
25 were performed at Earlham Institute (Norwich, UK).

26 27 28 29 **Data and Statistical analysis**

30
31
32 Bioinformatics analyses on raw data was performed using the Biomedical Genomics
33
34 Workbench version 4.0 (Qiagen) equipped with the Microbial Genomics Module version 2.0
35
36 (Qiagen) plugin. Sequences were imported and processed for optional merge paired reads,
37
38 adapter trimming, fixed length trimming and then the sequences were filtered based on the
39
40 number of reads to obtain sequences that are comparable in length and coverage for
41
42 clustering. Quality and chimera filtering were performed using the recommended
43
44 programme parameters [for complete details see www.qiagenbioinformatics.com]. Samples
45
46 with low coverage were removed from further analysis. Operational Taxonomic Units (OTU)
47
48 clustering and taxonomic assignment were done using Greengenes v13_5 (97%) as
49
50 reference. New OTUs were indicated when similarity percentage was lower than 80% with
51
52 minimum occurrence of 5 reads. Low abundant OTUs were discarded from further analyses
53
54
55
56
57
58
59
60

1
2
3 (minimum combined abundance was set at 10). Summary of processed sequence data is
4
5 described in Table S1. MUSCLE was used for OTUs alignment in order to reconstruct a
6
7 maximum likelihood phylogeny with neighbor joining as construction method and Jukes
8
9 Cantor as nucleotide substitution model. We compared community structures and diversity
10
11 across patient cohorts to determine if inter-cohort differences in structure were seen. To
12
13 achieve this a rarefaction sampling analysis was carried out using a standard methodology¹⁵.
14
15 In order to minimise the risk for contamination, we sought to identify a high (i.e.
16
17 conservative) number of sequences required to characterise bacteria from the samples ~~was~~
18
19 ~~determined from rarefaction.~~ In the absence of a standard international stringency cut off
20
21 the default settings in the workbench software (Qiagen) were applied; the default settings
22
23 filters out many reads and generates fewer unassigned reads and is therefore highly
24
25 stringent. These adjustments were assessed by analyzing a previously characterized mock
26
27 community to confirm that the data handling was consistent with other studies (sequencing
28
29 files of the project:
30
31 ERPO21973,https://www.ebi.ac.uk/metagenomics/projects/ERPO21973/samples/ERS15889
32
33 32). Data generated in the study was assessed using the Qiagen software and also applied to
34
35 EBI Metagenomics analysis pipeline V3.0 to verify results (acknowledging that fewer reads
36
37 would remain in the analysis with Qiagen). The findings in the controls were taken into
38
39 account during the analysis. Alpha diversity was calculated using number of OTUs. Beta
40
41 diversity was obtained using D_{0.5} UniFrac and represented as Principal Coordinate Analysis
42
43 (PCoA). Robustness analysis was performed using PERMANOVA with UniFrac distances. All
44
45 sequencing data have been deposited in NCBI Sequence Read Archive (SRA) database under
46
47 the follow BioProject ID: PRJNA388557
48
49
50
51
52
53
54
55
56
57
58
59
60

RESULTS

Study Subjects

Twenty patients recruited of which 8 were boys, 7 with diagnosed asthma and the mean (standard deviation) age was 10.7 (2.8) years. Descriptives of the children recruited are given in [Table 1](#). Induced sputum was obtained from 12 children and samples of fluid from bronchus, throat mouth and nose were obtained in all children. Three control samples were collected.

Sample quality control

A total of 11,750,879 PCR reads were analysed, and 606,227 high quality reads were obtained. The resulting OTU table contained 1053 OTUs, where 532 OTUs were assigned based on the Greengenes v13_5 (clustered at 97% similarity), and 521 OTUs were novel (see [Table S1](#)). These processing criteria allowed data from 74 samples from patients for further analyses. Of these 15 were from the nose, 16 from mouth, 17 from throat, 12 from sputum and 14 from bronchial samples ([Table S2](#)). Samples from all five sites were available in six children. The median read counts for the five sites were as follows: bronchus 4438; mouth 4400; throat 4868; nose 8927; and sputum 11250. [The results from the mock community analysis using Qiagen and EBI Metagenomics analysis pipelines were highly consistent. However Qiagen software assigned phyla for all samples whereas the EBI Metagenomics did not assign phyla in 15% of samples \(see \[Table S3\]\(#\)\).](#)

Bacterial abundance

The relative bacterial abundance in each sample cohort is presented in [Figure 1](#). The overall profile of bacterial communities was largely similar in the mouth, throat and sputum samples with the nose and bronchial being different. All sample cohorts contained five major bacterial phyla: Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria and Proteobacteria. The three most abundant phyla of bacteria observed by analysis in nose, mouth, throat, bronchial and sputum were Actinobacteria, Fusobacteria and Bacteroidetes (seen in [Figure 1](#)). Actinobacteria dominated in the nose (73% of all total species) and mouth (37%) samples, Fusobacteria dominated in throat (31%) and sputum (38%) samples and Proteobacteria were dominant in bronchial samples (34%). Both nasal and bronchial samples had more unclassified OTUs (12% and 41% respectively) than other samples. The relative abundance of each individual sample is shown in [Figure 2](#). [Figure S1](#) presents the phylogenetic tree of the microbial community in all analysed samples. In two of the control samples there was no identifiable bacteria DNA signal however some DNA signal was detected in the one of the three control and this was taken into account during analysis (See [Figure 2](#)).

Comparison between complete community structures within samples from different sites

The rarefaction analysis suggested that 2000 sequences would be sufficient for characterizing the majority of bacteria present in these communities. The numbers of reads were lowest in bacterial communities from mouth and bronchus ([Figure 3](#); [Table S2](#)). A comparison of individual bacterial communities, using PERMANOVA principal coordinate

1
2
3 analysis (PCoA), showed a distinct clustering by sample cohorts based roughly on where the
4
5 sample was taken from. Bacterial communities in samples from the mouth, throat and
6
7 induced sputum were tightly clustered together ($p_{adj}=1$) and were distinct from those of
8
9 the nose and bronchus ($p_{adj}<0.001$), which independently clustered (Figure 4; Table 2);
10
11 these clustering results were seen when children with asthma and without asthma were
12
13 considered separately (Table S4). Table S5~~3~~ demonstrates that clustering of bacterial
14
15 communities was not evident between sites among subgroups stratified by asthma/not
16
17 asthma or atopy/not atopy.
18
19

20
21
22 The bacterial communities in the five sites sampled did not differ between the children with
23
24 and without asthma; there was evidence of some differences in bacterial abundance and
25
26 diversity between groups (see supplement Figure S2). A comparison of the abundant phyla
27
28 of bacteria identified in bronchial samples taken from asthma and non-asthma patients
29
30 showed the profiles were similar, but the abundance of a signal for phyla was different (but
31
32 did not reach significance). In bronchial samples from asthma patients, the abundant phyla
33
34 identified was Proteobacteria while those samples taken from non-asthma patients showed
35
36 a greater abundance for Fusobacteria.
37
38
39
40
41
42
43
44

45 DISCUSSION

46
47
48 This study was designed to determine whether induced sputum is a valid alternative to
49
50 directly sampling of the lower airways to study airway microbiota in children. The main
51
52 finding was that the characteristics of the microbiota in induced sputum and in bronchial
53
54 samples were different. A second finding was that a minority of bacterial DNA in bronchial
55
56
57
58
59
60

1
2
3 samples was unclassified. In two individuals (P12 and P16, see Figure 2) the proportion of
4 phyla identified from the bronchial sample was similar to the sputum sample but for the
5 remainder of participants there were clear differences in the proportions of different phyla
6 from bronchial and sputum samples. Together these findings suggest that induced sputum is
7
8 not a valid surrogate for direct bronchial sampling to study the lower airway microbiota in
9 all children, but that there is still a pressing need to better understand the lower airway
10
11 microbiota in children. Opportunistic bronchial sampling under clinically-indicated general
12
13 anaesthetic ~~will be required~~ should remain the gold standard for studying lower airway
14
15 microbiome in children.

16
17
18 To our knowledge there are only ~~three~~ studies which ~~describe~~ have compared the
19
20 microbiota of upper and lower airway secretions in healthy children or children with
21
22 asthma⁶⁻⁸. A study where the majority of recruits were adults showed that the bacterial
23
24 community in the oropharynx and bronchus were similar in healthy individuals⁶ ~~whereas~~
25
26 One ~~a~~ solely paediatric study found very little overlap between the microbiota of the
27
28 oropharynx and lower airways⁷ whilst a second paediatric study found similar predominant
29
30 phyla in upper and lower airway samples but different genera⁸. Both paediatric studies
31
32 found significant differences in the clustering of bacterial communities in the upper and
33
34 lower airways^{7,8}. The ~~difference in dominant~~ phyla in between bronchial ~~and throat~~ samples
35
36 is not consistent between studies but there is consistency for upper airway
37
38 samples previously described^{6,7}. For example, the microbiota of the upper airways is
39
40 predominated by bacteria from the Firmicutes (e.g. Staphylococcus and Streptococcus) and
41
42 the Actinobacteria phyla (e.g. Corynebacteria) in our study and three others⁶⁻⁸. In contrast
43
44 and in bronchial samples, bacterial species within the Proteobacteria phylum (including
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Moraxella and Haemophilus) predominated d-in bronchial samples in our study and one
4 other⁷ but Firmicutes dominated in a third study⁸, and especially in children with asthma⁶,
5 whereas Differences in lower airway microbiome between studies may be due to different
6 sampling methods⁸ and treatment with current or recent antibiotics^{7,8}. in the microbiota of
7 the upper airways is predominated by bacteria from the Firmicutes^{6,7} (e.g. Staphylococcus
8 and Streptococcus) and the Actinobacteria phyla^{6,7} (e.g. Corynebacteria). Our results, where
9 we find no differences between the microbiome of induced sputum and throat swab, -are
10 also consistent with previous studies which suggested that a throat swab may be a useful
11 surrogate of induced sputum in young children with CF^{9,10}. As our study did not include
12 children with CF we cannot comment on the validity of induced sputum as an index of lower
13 airway infection in this specific clinical setting. The consistency of our results with previous
14 studies assures us that our findings are valid, despite the relatively small sample size.

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31 Our study findings were based on samples taken from a group of children who were well.
32 Recent work has given insight into the relationship between airway microbiota in the upper
33 airways of children with acute infection, with results which differ from the present study.
34 For example there is evidence that in the context of acute upper and lower respiratory tract
35 infection, different bacterial species predominate in the nasopharynx (e.g. Proteobacteria
36 and Firmicutes which contrast with Actinobacteria and Fusobacteria in our study)¹⁵. The
37 characteristics of the nasopharyngeal microbiota also change during the course of acute
38 otitis media infection¹⁶.

39
40
41
42
43
44
45
46
47
48
49
50
51 Our study design minimised the risk of false positive results arising from contamination and
52 from samples containing low DNA yields. There is no consensus on the number of control
53 samples which should be collected for microbiome studies, and two of the three controls we
54
55
56
57
58
59
60

1
2
3 collected contained no bacterial DNA and a third contained only bacteria with a very
4
5 different profile of phyla compared to the samples from study subjects. We set out to have
6
7 a high threshold for reads (>2000) and this filtered out samples with low DNA yields.
8
9 Another source of potential contamination comes from reagents used to analyse the
10
11 samples and this risk was minimised by taking standard precautions were taken and by
12
13 including control samples in the analysis. Whilst we cannot exclude the possibility that
14
15 some contamination may have occurred, our robust methodology and the consistency of
16
17 our findings with previous studies assure us that any contamination has not substantially
18
19 affected our findings.
20
21
22
23

24
25 Finding a small proportion of unassigned bacterial DNA in the nose and bronchial samples
26
27 was unexpected and worthy of further investigation. The presence of “unclassified
28
29 bacteria” has been described in nasopharyngeal samples from children⁷, but not in bronchial
30
31 samples. We have carefully reviewed the unassigned sequences and removed human DNA
32
33 sequences, and we therefore believe that there are a number of bacteria in the respiratory
34
35 tract which are not identified by standard microbiota methodologies currently used.
36
37
38

39
40 We analysed data from a mock community (a “positive control”) and this demonstrated that
41
42 the Qiagen software used for analysis of our samples was able to detect bacteria in similar
43
44 proportions to another software “pipeline”. The Qiagen software identified all phyla
45
46 whereas the alternative software could not assign 15% of bacterial DNA and this may be
47
48 explained by the higher stringency set for the Qiagen software. There is not standard for
49
50 stringency but these results suggest that although the higher stringency reduces the number
51
52 of reads included in the analysis, the results are highly comparable to other methods and
53
54
55
56
57
58
59
60

1
2
3 possibly less affected by potential contaminants. Our data are publicly available and
4
5 colleagues are welcome to apply different stringencies to our data.
6
7

8
9 There are a number of settings where upper airway samples have been shown to be valid
10
11 surrogates of lower airway samples for the clinical diagnosis of infective and non-infective
12
13 conditions. Viral aetiology in bronchiolitis (a lower respiratory tract infection) can be
14
15 established by nasopharyngeal aspirate¹⁶ and the nasal mucosa is used for diagnosis of
16
17 primary ciliary dyskinesia¹⁷ and (in some situations) CF¹⁸. In the research setting, sampling
18
19 from the upper airways has been validated as a surrogate for lower airway sampling^{19,20}.
20
21 However, we demonstrated that the microbiota of the nose was distinct to mouth and
22
23 bronchus in most individuals and our results indicating that the microbiota of the upper
24
25 airway is probably not a valid surrogate of lower airway microbiota in children. ~~—and~~
26
27 Previous papers have suggested that with regards to lower airway samples, upper airway
28
29 samples are “imperfect but reliable”⁷ and “both similar and different”⁸ and overall, our work
30
31 finds more evidence for the upper airways samples being “imperfect and different” to lower
32
33 airway samples and not “reliable and similar”. ~~has been shown by other as well.~~
34
35
36
37
38
39

40 The novelty of our study is that we demonstrate how the microbiota of induced sputum
41
42 differs from bronchial samples and very closely matches the microbiota of mouth and
43
44 throat, most likely due to contamination of lower airway secretions as they pass through the
45
46 oropharynx, including sputum which is commonly mixed with expectorant. Our findings are
47
48 consistent with a study of 78 children (mean age 2 years, almost half having received recent
49
50 antibiotic treatment) which also describes differences in the microbiota between nose and
51
52 mouth and between mouth and bronchial fluid⁷. Consistent with our findings, a study of
53
54 adults and children (mean age 11 years) reports Actinobacteria being the predominant
55
56
57
58
59
60

1
2
3 phyla in the nose but being almost absent from the oropharynx and bronchus. A different
4
5 pattern is shown for Proteobacteria which are increasing from nose through to oropharynx
6
7 and bronchus⁶. The apparent presence of different dominant bacterial communities in the
8
9 upper and lower airways raises the question “where do bacteria in the lower airways
10
11 originate?”
12
13

14
15 In children the lower airways are in direct communication with and adjacent to the
16
17 oropharynx but there are several mechanisms that may lead to the establishment of
18
19 different bacterial communities above and below the vocal cords. Perhaps most obviously,
20
21 a child’s mouth is regularly exposed to food and inedible items introduced to the oral cavity
22
23 (including cutlery and fingers) whereas the lower airways are exposed only to inhaled
24
25 exposures. Micro aspiration due to laryngopharyngeal reflux is thought to occur on a
26
27 regular basis²¹ and exposure to acid and other gastric contents (including bacteria) may
28
29 affect the microbiota of the lower airways but not the oropharynx. There are differences in
30
31 the innate defences of the upper and lower airways which may also explain differences in
32
33 their respective microbiota, for example lactoferrin concentrations are twice as high in the
34
35 lower airways relative to upper airways²² whereas nitric oxide concentrations (known to
36
37 have antimicrobial properties²³) are typically one hundred times higher in the upper
38
39 airways.
40
41
42
43
44

45
46 Our study was not designed to relate microbiota to clinical phenotypic data collected, for
47
48 example asthma or atopy. However, in light of recent studies examining these parameters in
49
50 children with asthma^{6,7} we made the best use of the data available, even with the limited
51
52 number we had at our disposal. When examined the possible relationship between the
53
54 bacteria present in children with and without asthma or atopy, we found no statistically
55
56
57

1
2
3 differences between the bacteria when the samples were examined as a whole group or as
4
5 specific sample region cohorts (nose, mouth, throat, bronchial and sputum samples), Figure
6
7 S2 and Table S3. A descriptive analysis of the bronchial samples collected found
8
9 Protoeobacteria to be the dominant species and this is consistent with Hilty *et al*⁶.

10
11
12
13
14
15
16
17
18
19
20 There are some limitations to our study. First, the number of participants was relatively
21
22 small. Second, induced sputum was not obtained in all participants and not all samples met
23
24 our quality control criteria for bacterial DNA analysis and this resulted in missing data for
25
26 some individuals. Third, as has been reported in other sputum microbiome studies^{6,7}, we
27
28 were not able to perform cell counts on the sputum pellet to determine the presence of
29
30 neutrophils and absence of squamous epithelial cells. Finally, we did not prospectively
31
32 calculate the DNA yield from samples although the consistency of our results with the small
33
34 number of published studies^{6,7}, the consistency of results using both the Qiagen and EBI
35
36 Metagenomics pipelines and the comparable results from mock communities assure is that
37
38 low DNA yield has not altered the results. although we included control samples in our
39
40 methodology we did not include samples from commercially available “mock” communities.

41
42
43
44
45
46 In summary, we report that neither induced sputum nor swabs from nose or mouth give an
47
48 accurate indication of the bronchial microbiota in all children. For very obvious practical
49
50 and ethical reasons, upper airway sampling will be necessary to study airway microbiome in
51
52 many research setting and our findings adds to the evidence describing the limitations of
53
54 this pragmatic approach. Our study was not designed to compare differences in the lower
55
56
57
58
59
60

1
2
3 airway microbiota of children with and without asthma, and future research is required to
4
5 supplement the little we do know about this potentially important subject.
6
7
8
9
10
11
12
13

14 **ACKNOWLEDGEMENTS**

15
16
17 The authors would like to thank the study participants and their parents. The study was
18
19 supported by an NHS-Grampian Endowments. AW is supported by the Wellcome Trust
20
21 Strategic Award (grant 097377) and the MRC Centre for Medical Mycology (grant
22
23 MR/N006364/1) at the University of Aberdeen.
24
25
26
27
28
29

30 **Competing interest statement.** None of the authors has any competing interests.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

1. Kahn FW, Jones JM. Diagnosing bacterial respiratory infection by bronchoalveolar lavage. *J Infect Dis* 1987; **155**(5): 862-9.
2. Marsland BJ, Yadava K, Nicod LP. The airway microbiome and disease. *Chest* 2013; **144**(2): 632-7.
3. Rogers GB, Shaw D, Marsh RL, Carroll MP, Serisier DJ, Bruce KD. Respiratory microbiota: addressing clinical questions, informing clinical practice. *Thorax* 2015; **70**(1): 74-81.
4. Klepac-Ceraj V, Lemon KP, Martin TR, Allgier M, Kembel SW, Knapp AA, Lory S, Brodie EL, Lynch SV, et al. Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and *Pseudomonas aeruginosa*. *Environ Microbiol* 2010; **12**(5): 1293-303.
5. Bisgaard H, Hermansen MN, Bonnelykke K, Stokholm J, Baty F, Skytt NL, Aniscenko J, Kebabdzic T, Johnston SL. Association of bacteria and viruses with wheezy episodes in young children: prospective birth cohort study. *BMJ* 2010; **341**: 4978.
6. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, et al. Disordered microbial communities in asthmatic airways. *PLoS ONE* 2010; **5**(1): e8578.
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.

1
2
3 [8. Kloepfer KM, Deschamp AR, Ross SE, Peterson-Carmichale SL, Hemmerich CM, Rusch DB,](#)
4 [Davies SD. In children, the microbiota of the nasopharynx and bronchoalveolar lavage fluid](#)
5 [are both similar and different. *Ped Pulmonol* 2018; DOI: 10.1002/ppul.23953](#)
6
7
8
9

10
11 9. Zemanick ET, Wagner BD, Robertson CE, Stevens MG, Szeffler SJ, Accurso FJ, Sagel SD,
12 Harris JK. Assessment of airway microbiota and inflammation in cystic fibrosis using multiple
13 sampling methods. *Ann Am Thor Soc* 2015; **12**(2): 221-9.
14
15
16

17
18
19 10. Boutin S, Graeber SY, Weitnauer M, Panitz J, Stahl M, Clausznitzer D, Kaderali L,
20 Einarsson G, Tunney MM, Elborn S, et al. Comparison of microbiomes from different niches
21 of upper and lower airways in children and adolescents with cystic fibrosis. *PLoS ONE* 2015;
22 **10**(1): e0116029.
23
24
25
26

27
28
29 11. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P,
30 van der Grinten CP, Gustaffson P, et al. Standardisation of spirometry. *Eur Respir J* 2005;
31 **26**(2): 319-38.
32
33
34
35

36
37 12. Pontin J, Blaylock MG, Walsh GM, Turner SW. Sputum eosinophil apoptotic rate is
38 positively correlated to exhaled nitric oxide in children. *Pediatr Pulmonol* 2008; **43**(11):
39 1130-4.
40
41
42
43

44
45 13. Twomey KB, Alston M, An S, O'Connell OJ, McCarthy Y, Swarbreck D, Febrer M, Dow
46 JW, Plant BJ, Ryan RP. Microbiota and metabolite profiling reveal specific alterations in
47 bacterial community structure and environment in the cystic fibrosis airway during
48 exacerbation. *PLoS ONE* 2013; **8**(12): e82432.
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 14. Moen AEF, Tannaes TM, Vatn S, Ricanek P, Vatn MH, Jahnsen J, IBD-Character
4 Consortium. Simultaneous purification of DNA and RNA from microbiota in a single colonic
5 mucosal biopsy. *BMC Research Notes* 2016; **9**: 328.
6
7
8
9
10
11 15. Zakharkina T, Martin-Loeches I, Matamoros S, Pova P, Torres A, Kastelijn JB, Hofstra
12 JJ, de Wever B, de Jong M, Schultz MJ, et al. The dynamics of the pulmonary microbiome
13 during mechanical ventilation in the intensive care unit and the association with occurrence
14 of pneumonia. *Thorax* 2017; **72**: 803-10.
15
16
17
18
19
20
21 16. Kelly MS, Surette MG, Smieja M, Pernica JM, Rossi L, Lunistra, Steenhoff AP, Feemster
22 KA, Goldfarb DM, Arscott-Mills T, Bioditswe S, Rulaganyang I et al. The Nasopharyngeal
23 Microbiota of Children with Respiratory Infections in Botswana. *Pediatr Infect Dis J*
24 2017;36:e211-e218
25
26
27
28
29
30
31
32
33
34
35 17. Chonmaitree T, Jennings K, Golovko G, Khanipov K, Pimenova M, Patel JA, McCormick
36 DP, Loeffelholz MJ, Fofanov, Y. Nasopharyngeal microbiota in infants and changes during
37 viral upper respiratory tract infection and acute otitis media. *PLoS ONE* 2017;12
38 (7):e0170630.
39
40
41
42
43
44
45 18. Ralston SL, Lieberthal AS, Meissner HC. Ralston SL, Lieberthal AS, Meissner HC,
46 Alverson BK, Baley JE, Gadomski AM, Johnson DW, et al. Clinical Practice Guideline: The
47 Diagnosis, Management, and Prevention of Bronchiolitis. *Pediatrics*. 2014;134(5):e1474-
48 e1502. *Pediatrics* 2015; **136**(4): 782.
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 19. Barbato A, Frischer T, Kuehni CE, Snijders D, Azevedo I, Bakati G, Bartoloni L, Escibano
4
5 A, Haarman E, Messelmar B, et al. Primary ciliary dyskinesia: a consensus statement on
6
7 diagnostic and treatment approaches in children. *Eur Respir J* 2009; **34**(6): 1264-76.
8
9
10
11 20. Farrell PM, Rosenstein BJ, White TB, Accurso FJ, Castellani C, Cutting GR, Durie PR,
12
13 Legrys VA, Massie J, Parad RB, et al. Guidelines for diagnosis of cystic fibrosis in newborns
14
15 through older adults: Cystic Fibrosis Foundation consensus report. *J Pediatr* 2008; **153**(2):
16
17 S4-S14.
18
19
20
21 21. Pringle EJ, Richardson HB, Miller D, Cornish DS, Devereux GS, Walsh GM, Turner SW.
22
23 Nasal and bronchial airway epithelial cell mediator release in children. *Pediatr Pulmonol*
24
25 2012; **47**(12): 1215-25.
26
27
28
29 22. Pitrez PMC, Brennan S, Turner S, Sly PD. Nasal wash as an alternative to
30
31 bronchoalveolar lavage in detecting early pulmonary inflammation in children with cystic
32
33 fibrosis. *Respirology* 2005; **10**(2): 177-82.
34
35
36
37 23. Venkatesan NN, Pine HS, Underbrink M. Laryngopharyngeal reflux disease in children.
38
39 *Pediatr Clin North Am* 2013; **60**(4): 865-78.
40
41
42
43 24. Roca-Ferrer J, Mullol J, Perez M, Xaubet A, Momlins L, de Haro J, Shelhamer J, Picado
44
45 C. Effects of topical glucocorticoids on in vitro lactoferrin glandular secretion: comparison
46
47 between human upper and lower airways. *J Allergy Clin Immunol* 2000; **106**(6): 1053-62.
48
49
50
51 24. Fang FC. Antimicrobial reactive oxygen and nitrogen species: concepts and
52
53 controversies. *Nature Rev Microbiol* 2004; **2**(10): 820-32.
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

FIGURE LEGENDS

Figure 1. Relative abundances of bacterial phyla identified as operational taxonomic units (OTUs) from the sequence reads generated from airway samples taken from children. The bar chart illustrates the taxonomic composition of each cohort of samples from a particular site. A detailed summary of the bacteria in each sample is described in Figure [2S4](#).

Figure 2. Operational Taxonomic Units based relative sequence abundance of bacterial phyla based on 16S rRNA of all 75 samples. A detailed summary of the bacteria in each sample is described. On the horizontal axis, the number following P is the patient number and NO=nose, MO=mouth, TH=throat, SP=Sputum, BR=bronchial sample. C2=the control sample where bacterial DNA was detected. N/A=not applicable

Figure 3. Alpha diversity rarefaction curves of samples based on total number of observed Operational Taxonomic Units. Colour denotes different sample cohort (“nose” (NO), “mouth” (MO), “throat” (TH), “sputum” (SP) and “bronchial” (BR) samples).

Figure 4. Diversity analysis demonstrating differences in the bacterial phyla community between samples taken from airway samples taken from asthmatic and non-asthmatic children. Principal coordinate (PCo) analysis of all samples based on D_{0.5} UniFrac distance. Colour denotes different sample cohort (“nose” (NO), “mouth” (MO), “throat” (TH), “sputum” (SP) and “bronchial” (BR) samples).

Figure S1. Phylogenetic tree of the microbial community in all analysed samples. All Operational Taxonomic Units are shown in this figure.

Figure S2. Bacterial phyla abundance and diversity between groups. On the horizontal axis, the first letter “Y” indicates that the child had asthma and “N” that they did not have

1
2
3 asthma and NO=nose, MO=mouth, TH=throat, SP=Sputum, BR=bronchial sample. N/A=not
4
5 applicable
6
7

8 **TABLE LEGENDS**

9

10
11 Table 1. Details of the children recruited. SABA=short acting beta agonist, ICS=inhaled
12
13 corticosteroids, LABA=long acting beta agonist
14
15

16
17 Table 2. PERMANOVA analysis results of testing differences in beta-diversity among
18
19 different sample sites.
20

21
22 Table S1. Summary of sequence data used for Operational Taxonomic Units (OTU) clustering
23
24 and analysis. The number of filtered reads and reads in OTUs does not include the reads
25
26 excluded by quality control
27

28
29 Table S2. List of sequence data for individual samples used for Operational Taxonomic Units
30
31 (OTU) clustering and analysis.
32
33

34 Table S3. Comparison of results from two pipelines used to analyse data. The Qiagen
35
36 pipeline was used for the present study. The EBA pipeline (full name EBI Metagenomics
37
38 analysis pipeline V3.0) was used as a comparator
39
40

41
42 Table S4. PERMANOVA analysis separately comparing beta-diversity among different sample
43
44 sets for children with asthma and without asthma.
45
46

47 Table S5~~Table S3~~. PERMANOVA analysis comparing beta-diversity among different sample
48
49 sets. Atopy was defined as serum IgE >100 kU/l.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Whereas some minor issues have been addressed major issues are still not sufficiently replied to. I
4 am still not convinced that one the main messages, namely that bronchial samples differ significantly
5 from induced sputum, is justified by the data and therefore allows to conclude: "Induced sputum
6 may not be a valid surrogate for microbiome assessment of the lower airways."

7 Before going into a deeper analysis I would like to mention that a considerable part of confusion
8 comes due to erroneous labeling and numbering of the figures and tables. E.g. in my previous review
9 I referred in point 4 to Fig. S2 which according to the legends was the phylogenetic tree (yet the
10 Figure itself had no label) and claimed the absence of Streptococci and Haemophilus. In the reply the
11 authors now refer to bar graphs that of course only show phyla. This is not a reply to my question.
12 Moreover the current manuscript is still superficial in preparation: Fig. S2 as labeled in the legends
13 ("Figure S2. Bacterial phyla abundance and diversity between groups") is shown as Fig. S3 on p.30. In
14 Fig. 1 the legend gives a note "A detailed summary of the bacteria in each sample is described in
15 Figure S1." but this is now Fig. 2. Therefore I now always refer to the page number indicated at the
16 top of the pdf file.
17

18 **Our response:** We apologise for any confusion caused. On the advice of the reviewer we moved one
19 figure from the supplement to the main paper. Therefore the numbering of figures in the
20 supplement changed between the first and the revised manuscript and we failed to capture all of
21 these changes. There are four specific points raised by the reviewer here:
22

- 23 1. Incorrect legend for Figure 1 - We have corrected the error in the legend for Figure 1 (i.e.
24 changed "Figure S1" to "Figure 2").
- 25 2. Reply to point 4 in initial review - The reviewer did not refer to Figure S2 in point 4 of their
26 previous review (they did refer to this in point 7). We do not know which question the
27 reviewer invites us to reply to.
- 28 3. No label for phylogenetic tree in supplement - We have inserted a label for Figure S1. The
29 Pediatric Pulmonology web site does not allow for a legend to be inserted separately (in
30 contrast with figures for the main paper where a legend is inserted separately).
- 31 4. Figure S2 labelled as Figure S3 - We have corrected this error.
32
33

34 **Severe technical shortcomings remain and these are again outlined:**

35 1. *It is not sufficiently clarified how the analysis results in such an enormous loss of reads:
36 Previously of 21.5 million reads only 400,493 were analysed, now it reads that from 11.7 million reads
37 606,227 could be analysed. Why were now fewer reads available? Why are less than 10% analysed?
38 This is a very unusual yield for 16S sequencing. The authors refer to "high stringency" and p.7 for
39 further explanations in their response but this stringency is nowhere explained. I am afraid that
40 important information is missed.*
41

42 **Our response:** There are actually now more reads and a smaller proportion of "lost" reads included
43 in the analysis reported in the revised manuscript (606,227 compared to 400,493). There is no
44 standard for stringency and, as the reviewer will be aware, the advantage of a higher cut off reduces
45 the potential for "false positive" findings. In this study, we used the default settings in the
46 workbench software (Qiagen). This has a very "high stringency" cut off which means that a lot of
47 reads are removed. Our data are publicly available and colleagues can apply different stringency
48 thresholds to our data. We have expanded the text on page 7 to explain our approach to stringency
49 which explains the threshold used. Finally (as suggested by the reviewer) we have used a mock
50 community to replicate data from our pipeline using second pipeline which applied a different
51 stringency; this last analysis indicates that our stringency has not affected the diversity of
52 microbiome detected but has increased the accuracy of detection (data now included in Table S3).
53
54

55 2. *On "p.29 of 62" the authors again show the phylogenetic tree that should show "all analysed
56 samples". In this tree I do neither see the genus Streptococcus which should be found in all healthy
57
58
59*

1
2
3 *individuals at least in the mouth and throat or Haemophilus which has been reported as a typical*
4 *proteobacterium in asthma patients. I come to the same conclusion as in (1.) that important*
5 *information might be missing. That is why I requested to analyze a defined mock community but the*
6 *authors refuse to do so. I do not see why this should not be possible even if the study was closed.*
7

8 **Our response:** This phylogenetic tree shows details of phyla and does not provide details of species.
9 Therefore species such as Streptococcus and Haemophilus are not shown. The analysis of mock
10 community indicates that our pipeline more sensitive than a second pipeline and that important
11 information is not missing. We have provided more labels on the phylogenetic tree identifying more
12 phyla.
13

14
15 3. *I previously stated that I am afraid that in bronchial samples the bacterial yield might have*
16 *been low and therefore prone to contamination. Indeed the new Fig. 2 clearly shows that in the one*
17 *control sample proteobacteria were brought in. I am afraid that in samples with known low yield*
18 *(bronchial washing from healthy individuals) this might induce an error as previously reported (Salter*
19 *et al. BMC Biology 2014, 12:87). I therefore requested to do an analysis of bacterial load (e.g.*
20 *quantitative 16S PCR). The authors now refer me to Table S2 but this is entirely misleading. “Number*
21 *of reads” has nothing to do with the original bacterial load. During the pipeline of mixing and*
22 *barcoding for the sequencing process typically equal amounts of PCR products are taken. After the*
23 *sequencing the number of reads only reflects variation in the sequencing and does not relate to the*
24 *original bacterial load. Without seeing true data on bacterial load the issue of contamination cannot*
25 *be discussed in sufficient depth.*
26
27

28 **Our response:** We believe that the mock community analysis described previously addresses the
29 issue of low DNA yield in bronchial samples (see Table S3). There was not sufficient sample volume
30 to determine the bacterial yield. The mock community analysis and consistency with published data
31 assure us that low bacterial load in bronchial samples has not substantially affected our results. We
32 have added text acknowledging that we did not calculate DNA yield. We have also added a very
33 recently published paper which also finds different clustering of bacterial communities in the upper
34 and lower airways.
35

36
37 4. *From Table S2 it is visible that for the bronchial samples a number of samples had to be*
38 *excluded because that did not give enough reads. I am afraid that the majority were from healthy*
39 *children whereas those from asthmatic were analysed. The latter are known to have a higher*
40 *Proteobacteria load. Therefore it is mandatory, to show the disease status in Fig. 2. Moreover, I*
41 *would like to see separate PERMANOVAs for diseased and healthy children with respect to*
42 *differences between sputa and bronchial washings. This is not covered so far by any of the analysis*
43 *and from the Fig. on “p. 30 of 62” (mislabeled as S3) it becomes clear that the difference between*
44 *NBR and NSP is much less as compared to YBR vs. YSP. I do NOT want to make a conclusion on*
45 *asthma patients but I state that the mixing of healthy and diseased people in on population might be*
46 *a significant problem. This might be also visible when Fig. 2 would be labeled with the disease status:*
47 *From a look at Fig. 2 it seems that at least 3-4 samples are much more similar to the throat/sputum.*
48 *The conclusion of difference between BR and SP is mostly based on the unassigned sequences and for*
49 *those the data do not sufficiently exclude a contamination issue (see my comment 3).*
50

51 **Our response:** The reviewer makes a number of points here:

- 52 • Samples need labelling by asthma status in figure 2. We have done this
- 53 • Separate PERMANOVA for children with and without asthma. We have done this (table S4)
- 54 • Difference between BR and SP is mostly based on unassigned sequences. We have now
- 55 removed unassigned sequences from this figure and the majority of bronchial samples are
- 56
- 57
- 58
- 59
- 60

1
2
3 predominantly the light blue colour corresponding to Proteobacteriae whereas the sputum
4 samples have very few Proteobacteriae seen but are dominated by Fusobacteria and
5 Bacteroidites. We do acknowledge that two bronchial samples (P12BR and P16 BR) have
6 few Proteobacteriae and have a similar pattern to their corresponding sputum sample.
7 When all patients are considered, there are significant differences between bronchial and
8 sputum samples but we have modified the conclusions to acknowledge that in a minority
9 sputum samples may be a valid index of bronchial samples.
10

11 5. *Technically it remains unclear why in Fig. 1 unassigned sequences are shown but in Fig. 2*
12 *excluded. Moreover, on visual inspection there seems to be a discrepancy between proteobacteria*
13 *abundance in Fig. 2 for all BR samples (for many samples >50%) as compared to Fig. 1 (20%). How*
14 *has the data in Fig. 1 been calculated (mean?). All in all the data do not sufficiently justify the main*
15 *conclusion.*
16

17 **Our response:** We have revised Figure 1 and removed the unassigned phyla. The reviewer is correct
18 and data in Figure 1 are mean values. The proportion of proteobacteria in figure 2 is certainly >50%
19 for most individuals but this figure shows how two individuals (P12BR and p16BR) have very few
20 proteobacteria and results from these two individuals reduce the mean proportion of proteobacteria
21 to the value represented in Figure 1. We have amended the main conclusion (see point 4 above).
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60