



OPEN ACCESS

# Respiratory microbiota: addressing clinical questions, informing clinical practice

Geraint B Rogers,<sup>1</sup> Dominick Shaw,<sup>2</sup> Robyn L Marsh,<sup>3</sup> Mary P Carroll,<sup>4</sup>  
David J Serisier,<sup>5,6</sup> Kenneth D Bruce<sup>7</sup>

<sup>1</sup>SAHMRI Infection and Immunity Theme, School of Medicine, Flinders University, Adelaide, Australia

<sup>2</sup>Nottingham Respiratory Research Unit, University of Nottingham, Nottingham City Hospital, Nottingham, Notts, UK

<sup>3</sup>Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia

<sup>4</sup>Cystic Fibrosis Unit, Southampton University Hospitals NHS Trust, Southampton, UK

<sup>5</sup>Immunity, Infection, and Inflammation Program, Mater Research Institute, University of Queensland, and Translational Research Institute, Woolloongabba, Queensland, Australia

<sup>6</sup>Department of Respiratory Medicine, Mater Adult Hospital, South Brisbane, Australia

<sup>7</sup>King's College London, Institute of Pharmaceutical Science, London, UK

## Correspondence to

Dr Dominick Shaw,  
Nottingham Respiratory Research Unit, University of Nottingham, Nottingham City Hospital, Nottingham, Notts, UK;  
Dominic.Shaw@nottingham.ac.uk

Received 3 June 2014  
Accepted 23 June 2014  
Published Online First  
17 July 2014

This is a reprint of a paper that first appeared in *Thorax*, 2015, volume 70, pages 74–81.



Open Access  
Scan to access more  
free content



CrossMark

**To cite:** Rogers GB, Shaw D, Marsh RL, et al. *Thorax* 2015;70:74–81.

## ABSTRACT

Over the last decade, technological advances have revolutionised efforts to understand the role played by microbes in airways disease. With the application of ever more sophisticated techniques, the literature has become increasingly inaccessible to the non-specialist reader, potentially hampering the translation of these gains into improvements in patient care. In this article, we set out the key principles underpinning microbiota research in respiratory contexts and provide practical guidance on how best such studies can be designed, executed and interpreted. We examine how an understanding of the respiratory microbiota both challenges fundamental assumptions and provides novel clinical insights into lung disease, and we set out a number of important targets for ongoing research.

## INTRODUCTION

While efforts to identify microbes associated with airways disease are not new, the sophistication of both our conceptual understanding of these systems and the approaches used to characterise them have increased dramatically over the last decade. A point has now been reached where complex microbial systems in the airways and the manner in which they can influence clinical course increasingly represent a central consideration when trying to understand and treat respiratory disease. However, with the application of ever more sophisticated ecological tools, recent additions to the literature have become less accessible to the non-specialist reader. Furthermore, in order to have the greatest beneficial impact, respiratory microbiota studies must be driven by clinical questions, and the insight gained linked directly with improvements in clinical practice. In this article, we set out the key principles underpinning microbiota research in clinical contexts, as well as providing practical guidance on how best such studies can be designed, executed and interpreted. We examine how an understanding of the respiratory microbiota both challenges fundamental assumptions and provides novel clinical insights into lung disease.

## WHAT IS MEANT BY MICROBIOTA?

In simplest terms, ‘microbiota’ can be defined as the microbes associated with a particular context. For example, the microbes present in the nasopharynx of healthy individuals can be referred to as the healthy (or commensal) nasopharyngeal microbiota. This definition covers all forms of microbe including viruses, bacteria, fungi, archaea and non-fungal microscopic eukaryotes. Most studies of microbiota have focused on their bacterial

component, and it is to this that we refer when using the term within this article. Unlike traditional microbiological approaches that aim to identify individual pathogens, microbiota analysis characterises all of the bacterial species present, both in terms of their identities *and* relative abundance. While used increasingly frequently in the literature, *microbiota* is only one of a number of terms that have been used to describe human-associated microbes (see box 1).

## MICROBIOTA DESCRIBED IN RESPIRATORY CONTEXTS

While the term microbiota has been used for decades, clinical and scientific interest in human-associated microbiota is more recent and has been further stimulated by projects such as the Human Microbiome Project and Metagenomics of the Human Intestinal Tract consortium (METAHit). The number of studies on respiratory microbiota has expanded massively, with PubMed showing more ‘hits’ in papers published from 2012 onwards than the total from all previous years, with microbiota analyses having now being performed in many different respiratory contexts. These studies can be grouped into those focusing on regions of the airways that are colonised by commensal populations under normal circumstances, such as the nasopharynx, and those thought to be free from substantial microbial colonisation in healthy individuals, such as the distal airways. Further, these contexts have been analysed in both healthy airways and those affected by acute and chronic infection. With such a rapidly expanding body of literature, a detailed description of findings is beyond the scope of this article, although a number of comprehensive reviews exist.<sup>1 2</sup>

## WHY ANALYSE THE MICROBIOTA?

From a clinical perspective, the value in microbiota analysis may not be immediately clear. Historically, efforts to characterise the airway microbiota were driven by the need to augment information provided by conventional culture-based microbiology. Based on the rationale that any microbes present in the lower airways represented potential aetiological agents, there was a clear need to detect microbes that might be refractory to culture *in vitro*. Here, there was an expectation that parallels would exist with the culture bias observed in natural environmental microbiology. Indeed, early analyses of samples from patients with cystic fibrosis (CF) sputum revealed many bacterial species that had not been reported through standard diagnostic microbiological culture-based approaches.<sup>3</sup> By

## Box 1 Glossary

- ▶ *Dominance*—the extent to which one or more species is numerically dominant within the microbiota
- ▶ *Dysbiosis*—an imbalance in the microbes present in a particular niche due to a change in conditions
- ▶ *Evenness*—the degree to which the species present are of equal abundance
- ▶ *Metagenome*—the genetic information of the whole microbiota, usually obtained by whole genome sequencing
- ▶ *Microbiome*—the totality of the microbes with their genes that are harboured by the microbiota and the milieu in which they interact
- ▶ *Microbiota*—all the microbes that are found in a particular niche or region
- ▶ *Microflora*—a now defunct term, broadly equivalent to microbiota
- ▶ *Phylogenetic*—relating to the evolution of a species or group of organisms
- ▶ *Resilience*—the rate at which a microbial community returns to its original composition after being disturbed
- ▶ *Resistance*—the degree to which a community withstands change in response to perturbation
- ▶ *Richness*—the number of taxa present in a sample at a particular phylogenetic level
- ▶ *Similarity measures*—a statistical tool to determine the similarity between two profiles based on a given algorithm
- ▶ *Succession*—the gradual and orderly process of change in a microbial community brought about by the progressive replacement of its members
- ▶ *Taxa*—a taxonomic category, such as species or genus

extension, while it is difficult to speculate on the importance of all novel species identified in airways disease, the common detection of strict anaerobes in CF sputum has led to a re-evaluation of their potential to contribute to chronic lung infection.<sup>4 5</sup>

As the complexity of airway microbiota was revealed, it became apparent that detailed microbial data could be informative beyond the simple detection of potentially pathogenic species. One important concept took shape, namely, that the pathogenic potential of a microbiota containing a diverse mix of species can be quite distinct from, and in some cases significantly greater than, that of its individual members. This is due in part to the substantial influence that inter-species interactions can have on the expression of the virulence traits of bacteria when resident within polymicrobial communities.<sup>6</sup> This insight led to an understanding that the pathogenic potential of any single bacterial species requires a consideration of the wider microbial context.<sup>7</sup>

The application of conceptual frameworks from microbial ecology also suggested that the composition of airway microbiota was to a large extent a reflection of the physicochemical characteristics of airway niches.<sup>7</sup> Bacterial species differ in their need for resources and physical conditions in order to grow and different environmental niches therefore select for the growth of certain microbes. It is interesting to speculate that the differential distribution of infections within the respiratory tract, such as the more common finding of upper, not lower, lobe TB infection, may result from such a phenomenon. Assessing airway

microbiota may therefore provide insight into the physical characteristics of the airways, and by extension the degree of disease progression and likelihood of colonisation by a particular pathogen.

### KEY MICROBIOTA DESCRIPTORS

In addition to determining the identities and relative abundance of bacterial species present within a respiratory context, other features of the microbiota can be informative, including *richness*, *evenness* and *dominance* (box 1). In many cases, these measures have been shown to be clinically informative. For example, clinical measures of disease in non-CF bronchiectasis correlate more closely with samples characterised according to dominant species identity than to conventional presence/absence detection of clinically relevant species,<sup>5 8</sup> and represent a predictor of microbial community response to antibiotic therapy.<sup>9</sup> Further, species richness measures have been shown to have a significant inverse correlation with disease severity in a number of conditions, including CF, bronchiectasis and COPD.<sup>10</sup>

### HOW ARE AIRWAYS MICROBIOTA CHARACTERISED?

The DNA sequencing technologies that allow characterisation of complex microbial systems have evolved rapidly, with individual sequencing platforms quickly superseded.<sup>11</sup> However, both the general principles on which DNA sequencing technologies are based and the key considerations when applying them to respiratory samples have remained relatively constant.

When characterising the microbiota, there are three principal considerations: (A) obtaining representative samples, (B) generating accurate microbiota profiles and (C) analysing the resulting data in a manner that is informative and relevant.

#### Obtaining appropriate samples

Respiratory material can be obtained in several ways. It is important to consider the ease, safety and reproducibility of each sampling technique in relation to microbiota profiling, rather than its utility for clinical investigations or other types of analysis. In addition, the issue of contamination requires careful consideration.<sup>12</sup> Obtaining samples from the lower airways involves passage through regions that are typically heavily colonised by microbes such as the nasopharynx and while protected brush specimens obtained at bronchoscopy can be used to limit the introduction of upper respiratory tract microbes, this approach may not always be appropriate.

Regardless of the strategy used to obtain material from an airway niche, the effect of heterogeneity in sample composition, both spatial and temporal, must be considered. For example, two sputa collected consecutively may vary considerably in terms of their composition,<sup>13</sup> a factor particularly important in cross-sectional studies.

#### Generating microbiota profiles

Once sample material has been obtained, there are a number of other stages required before generation of a microbiota profile. The first major step is the extraction of nucleic acids. The ease with which cells of different bacterial species lyse differs greatly; consequently, where stringent cell disruption is not performed, the microbiota data obtained can be distorted with an over-representation of those species most easily lysed (including many Gram negative species) and under-representation of more structurally resilient species (often Gram positive species). A number of studies have examined this topic,<sup>14 15</sup> with the inclusion of enzymatic and physical sample disruption (eg, via bead-beating) commonly considered necessary.

Once nucleic acids have been extracted, a microbiota profile can be generated. Most commonly, approaches involve the PCR amplification of variable regions of the 16S rRNA gene using primers that bind to flanking regions of conserved sequence. Here, the selection of appropriate PCR primers is fundamental to the usefulness of the data obtained.<sup>16</sup> An alternative to analysing amplified 16S rRNA gene analysis is to generate metagenomic data by sequencing all DNA derived from a sample (known as shotgun sequencing) and identifying 16S rRNA gene sequences, or other informative sequences, within the dataset.<sup>17</sup> This technique has been applied successfully in the analysis of CF sputum.<sup>18</sup> While shotgun sequencing has not been applied widely, it has the advantage of providing information on the presence of bacteria, archaea, DNA viruses and eukarya, as well as their potential functionality.

### Data processing

Before microbiota sequence data can be analysed, it must be processed to minimise spurious signal and allow the comparison of profiles with the minimal introduction of bias. Data processing can be divided into a number of stages.

#### Removal of spurious signal

All PCR and sequencing techniques inevitably introduce sources of spurious signal, including amplicon fragments, chimeric sequences (single sequences originating from two organisms) and misreads. It is essential that prior to downstream analysis, data are processed to remove these factors and there are a number of publically available pipelines that can be used to achieve this, including Qiime<sup>19</sup> and mothur;<sup>20</sup> for further detail see.<sup>11</sup>

#### Contamination

Contamination of the analysis pathway can be introduced at any stage. In particular, contamination present in analytical reagents is commonly detectable in 'no template' controls used for PCR amplification prior to sequencing. Here, the contribution of such contamination to the total signal obtained is commonly inversely proportional to the amount of nucleic acids derived from the sample and, by extension, the sample biomass. Sequence data obtained in such controls must be carefully compared with data from clinical samples.

#### Data normalisation

Processing and analysis of clinical samples, even aliquots of the same sample, will give rise to different DNA yields and purities and, in turn, different numbers of sequences will be obtained. When assessing the composition of microbiota, it is important that the number of sequences on which each profile is based (or the 'depth of sequencing') is comparable. It is therefore common practice to normalise the number of sequences per sample to the lowest number obtained for any sample within a set. Consideration must also be given to the depth of sequencing that is most appropriate; sufficient sequences must be obtained to avoid sampling bias, but sequencing to too greater depth will provide little additional information, and only increase costs and processing time. Obtaining pilot data that provide information on the level of diversity present in a particular niche is helpful in determining initial sequencing depths, with rarefaction of the sequence data obtained allowing the proportion of total diversity that is represented within a given microbiota profile to be assessed.

### Taxa identification

Once spurious signal has been removed and profiles normalised, the identities of the microbes present within the samples can be determined. This is achieved by comparing the sequencing obtained to a reference database. These databases can either be public repositories such as the National Institutes of Health sequence database, Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>) or an aligned sequence such as Greengenes (<http://greengenes.lbl.gov/cgi-bin/nph-index.cgi>), SILVA (<http://www.arb-silva.de/>) or RDP-II (<http://rdp.cme.msu.edu/>). Aligned sequence databases are often preferred as they are subject to lower levels of misidentification due to poor quality sequences and inaccurate annotation compared with public repositories.

#### Differentiating resident and transient taxa

In addition to bacteria resident within that niche being sampled, respiratory samples also contain microbes that have entered the lower airway through inhalation or translocation from adjoining regions, or that have been introduced during sample collection. While efforts can be made to limit the contribution of microbes that become associated with the sample as it passes through the upper respiratory tract, determining whether particular species are resident in the lower airways or present only transiently is more challenging. Here, analytical techniques have now been developed that allow differentiation between 'core' species (those that are commonly present in a patient group and in high abundance) and 'satellite' species (those that are present only rarely and at low abundance where detected) based on the distribution of sequences within a particular sample collection.<sup>21</sup>

### DATA ANALYSIS

There are many different approaches that can be taken to microbiota data analysis, with the particular strategy employed depending on the clinical question posed. However, the complexity of microbiota data, even before other factors such as host immunity and treatment are taken into account, means that sophisticated analytical approaches are typically required. Current techniques borrow heavily from environmental microbiology and allow factors such as the relationship between variation in composition and clinical measures to be assessed. Detailed review of microbiota analysis methods is available elsewhere;<sup>22</sup> however, we summarise key concepts below.

In order to reduce the influence of rare or overabundant species, a range of mathematical transformations are available to normalise data.<sup>23</sup> Once sequencing data have been transformed, there are a number of ways in which complex microbiota can be described using relatively accessible metrics. For example, richness is a commonly used measure that refers to the number of taxa, whether defined as species, operational taxonomic units (OTUs) or other phylogenetic classifications, which can be detected in a sample. While simple counts of the number of species present can be influenced by the depth to which sequencing has been performed, measures that take this into account, such as Chao1<sup>24</sup> and ACE<sup>25</sup> can be employed. Importantly, richness measures do not take into consideration the abundance of different taxa. In contrast, diversity measures combine richness metrics with a measure of the evenness of abundance of the different species present. Many different diversity measures exist, each reflecting particular microbiota characteristics. Examples of commonly used diversity measures include the Shannon index,<sup>26</sup> which ranges from 0 for communities with only one taxon to high values for communities with many different taxa of low relative abundance, and the Simpson index<sup>27</sup> that ranges

from 0 (all taxa are equally abundant) to 1 (one taxon dominates the community completely). A detailed description of  $\alpha$ -diversity indices and application in microbiome research can be found in the work by Li *et al.*<sup>28</sup>

In addition to describing the characteristics of the microbiota in an individual sample, it can be useful to compare the characteristics of multiple different samples. Inter-sample measures of similarity or dissimilarity are referred to as  $\beta$ -diversity and, again, can be based on many different facets of microbiota composition.

For example, similarity measures can be based on the presence/absence of particular bacterial species (eg, Sørensen and Jaccard Similarity Indices) or on the abundance of those bacterial species (eg, Bray-Curtis Dissimilarity Index). Further, the inter-sample similarities between multiple samples can be visualised using ordination techniques such as principle coordinate analysis and non-metric multi-dimensional scaling. Hierarchical clustering can also identify groups of samples with similar microbiota. The significance of differences between clusters can be determined using the analysis of similarity or multivariate analysis of variance (eg, NPMANOVA, PERMANOVA). There are numerous methods for examining relationships between observed clusters and clinical metadata, for example, redundancy analysis, correspondence analysis, linear discriminant analysis and regression models. Methods such as similarity percentages analysis can also determine which bacterial species contribute most to observed differences between clusters.

The application of many of the techniques described here to an illustrative set of respiratory samples is shown in figure 1.

Given the complexity of microbiota analyses and the range of analytical tools available, it is recommended that a bioinformatician with specialist expertise in microbial ecology is consulted when preparing an analytic pipeline.

### **PATHOGENICITY, CAUSALITY AND CLINICAL CORRELATES**

Associations between microbiota and airways disease are increasingly being reported. An example is the link between the composition of bronchial microbiota with the degree of bronchial hyper-responsiveness among patients with suboptimally controlled asthma.<sup>29</sup> Here, it has been postulated that asthma and allergy represent interplay among consequences of abnormalities in microbial colonisation, development of immune function and encounter with agents infecting the respiratory tract, especially at a young age.<sup>30</sup> While many such significant associations have been observed, determining causality is challenging. An illustrative observation is the inverse relationship between increasing disease severity and decreasing diversity.<sup>10</sup> Some authors have suggested that the observed link between disease severity and diversity represents a causal relationship (perhaps overgrowth by a pathogen could result both in reduced diversity and increased severity); however, another explanation is that as severity increases so too does antibiotic treatment burden. This antibiotic burden represents a substantial selective pressure acting to exclude all but the species able to grow in the presence of the antibiotics administered.

So how best can causality be assessed? First, an appropriate conceptual framework is needed, and this requires a reconsideration of some of the concepts surrounding infection and disease that predate microbiota analysis. The traditional approaches are still based on Koch's postulates. More recently, these have been modified to make them appropriate for examining the potential role of genes and their products in the pathogenesis.<sup>31</sup> However, factors related specifically to microbiota must also be taken into consideration. Here, rather than the presence of a

particular pathogen, disease may result from the activity of a consortium of species. For example, how do we define an oropharyngeal-associated bacterial species that, while benign when present in isolation, triggers disease when present with cocolonising pathogens such as *Pseudomonas aeruginosa* species?<sup>6</sup> Only when such relationships are understood can the extent to which correlations between microbiota data and clinical outcome are causal be assessed.

In addition to conceptual frameworks, practical systems to assess the nature of associations identified *in vivo* are required. These systems will include *in vitro* models of microbe–host cell interactions, as well as animal models of polymicrobial infections. Using these approaches, the mechanistic basis of observed associations can start to be unravelled.

### **MICROBIOTA DYNAMICS AND CHANGES IN CLINICAL STATUS**

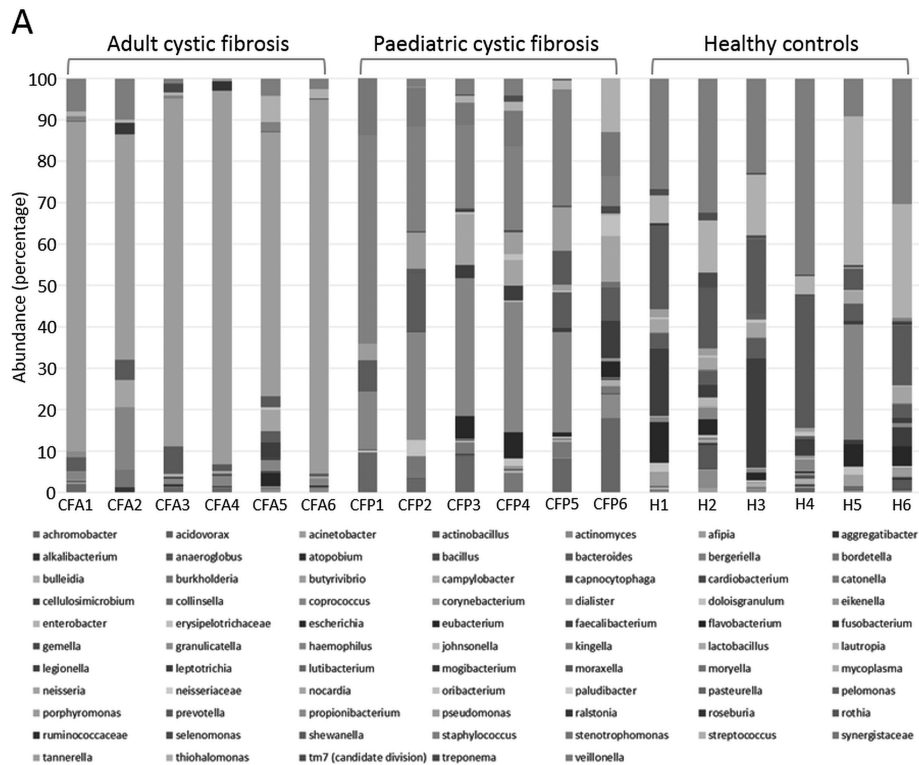
Airway microbiota change in response to a range of factors, most notably host immune response and treatment.<sup>32</sup> The dynamic nature of microbiota mean that cross-sectional studies can be misleading unless analysed with care, with longitudinal studies most appropriate when assessing relationships between microbiota composition and clinical factors.<sup>33</sup> Of particular clinical interest are the changes that take place during the establishment of infective microbiota, the impact of antimicrobial therapy and the changes in microbiota associated with disease progression. In each case, identification of the mechanisms that underpin these processes may offer the opportunity for clinical intervention. Here, there are two important ecological principles that help us understand clinically important processes.

**Succession:** the process of change in microbiota membership over time. Succession can take several forms, and differs depending on whether it takes place within a niche that has not been previously colonised (eg, colonisation of the lower airways in a child with CF) or in a niche where a commensal microbiota is undergoing dysbiosis (eg, the transition that might occur in the oropharyngeal microbiota as a result of H1N1 infection<sup>34</sup>). Importantly, succession is non-random, being influenced by factors such as the nature of the change in the airway environment and interactions between microbes.<sup>35</sup> This ordered basis of succession offers the tantalising prospect of the prediction of clinically important events, such as pathogen acquisition.

**Resistance/resilience:** to understand the effect that perturbations such as antibiotic therapy have on airway microbiota, as well as the stability shown by certain respiratory microbiota over long periods, the principles of resistance (insensitivity to disturbance) and resilience (the rate of recovery after disturbance) are essential. Determining the extent to which a microbiota is resistant and resilient can allow predictions to be made regarding the magnitude of change that an intervention is likely to have, and the period for which microbiota composition will remain disturbed before the community returns to its original composition or a new stable composition is established. Such analyses therefore provide a potential basis for the design of interventions. These concepts are reviewed in detail elsewhere.<sup>36</sup>

### **PHYLOGENETIC STRATA**

In clinical practice, the phylogenetic unit used most commonly is *species*, followed closely by *strain*, classifications that traditionally have been based primarily on phenotypic or morphological characteristics. *Strain* is a unit that reflects subspecies differences, and these can have significant clinical implications on epidemic or antibiotic resistant strains. Finally, there are the



**Figure 1**  $\alpha$ - And  $\beta$ -diversity analysis based on a simulated data set that includes six samples from adult cystic fibrosis (CF) patients, six samples from paediatric CF patients and six healthy controls. Shown are a number of approaches that could be employed to assess the composition of the samples in the different subsets, determine the extent to which differ significantly and identify bacteria that contribute substantially to these differences. (A) Stacked bar plots showing the relative abundance of genera detected in each of the samples. (B)  $\alpha$ -Diversity measures: genus richness, Shannon diversity index and Simpson index calculated for each of the samples. (C) Hierarchical cluster analysis showing the similarity of the microbiota in each sample using the Bray Curtis similarity measure. Red lines indicate no significant dissimilarity, with black lines indicating significant dissimilarity, as calculated using the SIMPROF test. (D) Non-metric multidimensional scaling using Bray Curtis similarity measure. Ellipses indicate thresholds of microbiota similarity. (E) Principal component analysis showing the distribution of the microbiota profiles. Eigen value vectors are included that indicate the contribution of key genera to the observed distribution. (F) Similarity percentages (SIMPER) analysis showing the contribution of different genera to the dissimilarity between first adult CF samples and paediatric CF samples, and then between adult CF samples and healthy controls. Shown are the average dissimilarity between sample groups with the inclusion of each genus, the contribution genera make to dissimilarity and the mean abundance of each genus in the two sample groups carriage return.

The set of six panels provide an overview of approaches to objectively assessing differences in microbiota composition between samples from different patient groups. The stacked bar plots shown in (A) provide an overview of the relative abundance of genera detected in the 18 samples analysed. It can be seen that genus richness in the samples from adult CF patients is lower than in the paediatric CF patients or the healthy controls, with a high relative abundance of *Pseudomonas*. It can also be seen that while *Haemophilus* and *Staphylococcus* are relatively prevalent in the paediatric samples, healthy controls tend to be dominated by *Streptococcus* and *Prevotella* species. The genus richness, Shannon index and Simpson index diversity scores plotted in (B) indicate that samples from healthy controls tend to have more genera detected and that these samples have a more even distribution of abundance compared with paediatric CF samples and, to a greater degree, adult CF samples. The hierarchical cluster analysis shown in (C) indicates that the composition of the three groups of samples is distinct, that is, samples within the three subgroups are more similar to each other than to samples in other subgroups. The black lines indicate clusters of samples that are significantly different, as determined by SIMPROF analysis. The analysis of similarity (ANOSIM) score assesses the significance of the differences between the three sample groups. (D) Shows non-metric multidimensional scaling using the Bray Curtis dissimilarity measure. Here, the closer samples are together, the more similar their composition. It can be seen that the three sample groups cluster together, and that there are further subgroups of samples within the adult CF and paediatric CF sets, with percentage similarity thresholds shown by ellipses. In (E), principle component analysis has been used to project the sample over two axes based on variance in their composition. Again, the samples cluster according to group. Vectors have been overlaid that indicate that *Pseudomonas* contributes most to the separation of adult CF samples from paediatric CF sample or healthy controls along the x axis, whereas *Staphylococcus*, *Haemophilus* and *Moraxella* spp are more prevalent in paediatric samples, and *Prevotella* and *Veillonella* species are more prevalent in healthy controls. The contribution of the detected genera to the differences in microbiota composition between each of the three sample groups was assessed by SIMPER analysis, with the output shown in (F). Genera are listed in order of their contribution to dissimilarity. In this way, the bacteria that drive differences between microbiota in different clinical contexts can be determined objectively.

broad classifications that are used primarily as a guide when selecting antibiotics, for example, Gram positive/negative and 'atypical'. However, when describing a respiratory microbiota, these phylogenetic strata may be inappropriate. In microbiota analysis, differentiation of bacteria is based typically on sequence differences in regions of DNA. In some cases, species will be indistinguishable over the region analysed, while in other

instances, there can be substantial sequence divergence in populations conventionally considered to represent a single strain. A further consideration is the ability of 16S rRNA gene-based techniques to distinguish between species with very different pathogenic potential. For example, in the context of paediatric respiratory infection, *Haemophilus influenzae* is considered an important pathogen, whereas *Haemophilus haemolyticus* is

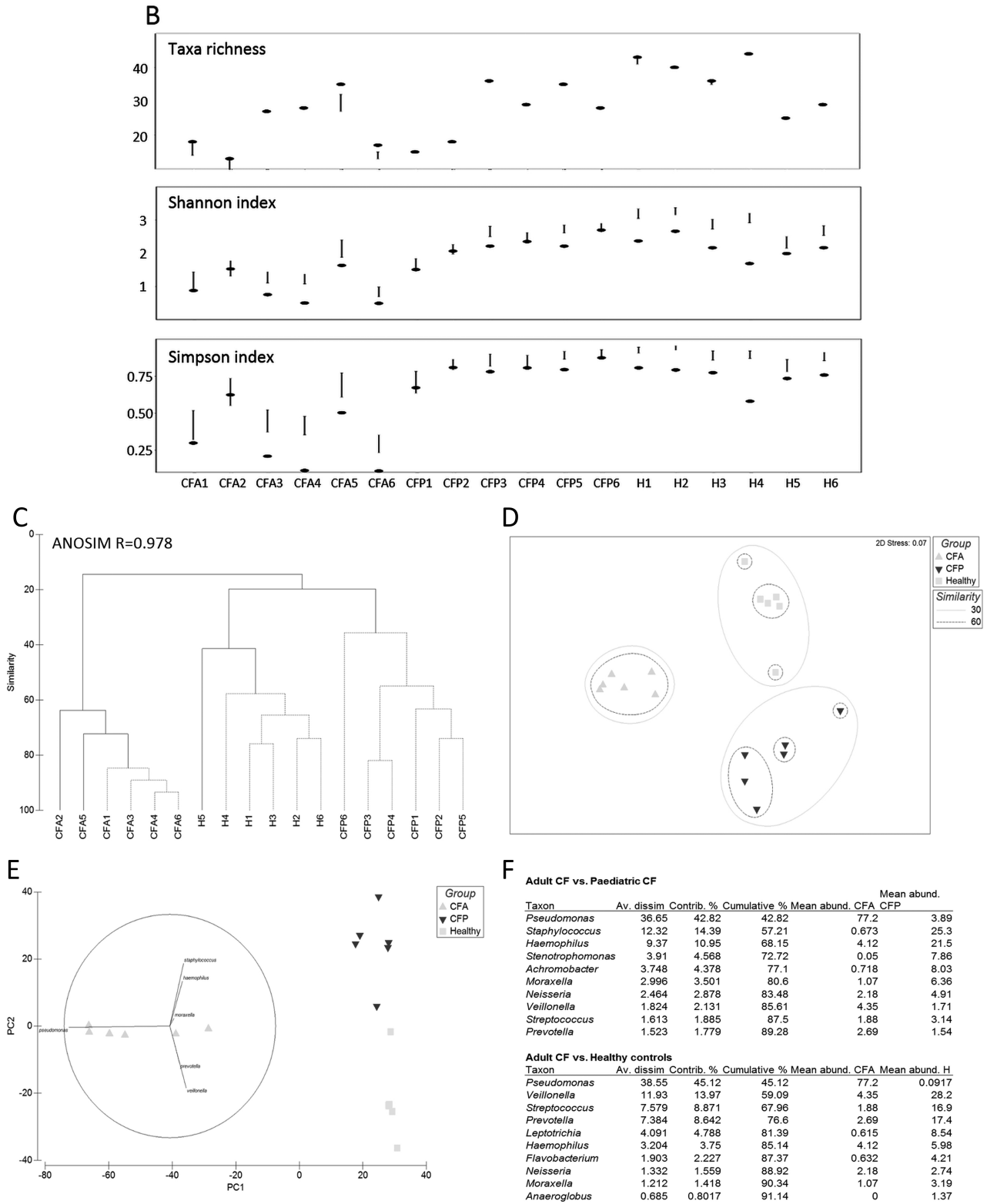


Figure 1 Continued.

considered commensal. However, obtaining satisfactory differentiation of these species based on 16S rRNA gene sequences derived from typical sequencing analysis is not possible.<sup>37</sup>

In order to separate bacterial species in a meaningful way based on sequence data, thresholds of sequence similarity are used. These thresholds are used to define what are referred to as an OTU. For example, although a level of 97% sequences

similarity is commonly applied to 16S rRNA gene sequence data (broadly relating to the species level), traditional strain or species classifications show different degrees of sequence variation. It is therefore important to consider the extent to which the analytical approach used is capable of differentiating particular bacteria of interest, select analytical strategies appropriately and be aware of the limitations of data generated to provide

high-level bacterial identification. Further, it may be most appropriate to consider microbiota data at a higher phylogenetic level, for example, family or phylum, which is not commonly considered in clinical contexts.

### TOWARDS CLINICAL INSIGHT

To date, the analysis of respiratory microbiota has been predominantly in the sphere of academic research, with direct application to clinical practice yet to take place on a significant level. This situation is changing with increasing efforts to link microbiota data with clinically applicable metrics. However, determining the membership of the airway microbiota is only a first step in understanding its interaction with the host and role in both health and disease. Bacteria have many differentially expressed traits relating to virulence or pathogenicity and there is growing interest in determining not just what bacteria are present, but what they are doing. There are a number of different ways in which bacterial behaviour can be assessed, including the analysis of gene transcription (transcriptomics), protein production (proteomics) and metabolic activity (metabolomics). In each case, technological advances are providing a basis for assessing these processes within highly complex microbiota, and their application in respiratory contexts continues to expand.<sup>38–40</sup>

With the rapid expansion of airways microbiota analysis, new applications for the data that they generate are being identified. As above, the composition of the airway microbiota reflects the characteristics of the airway environment, with a continuum between those observed in healthy individuals and in disease. This relationship presents the opportunity to use microbiota data to both track disease progression and assess the efficacy of treatments that aim to retard it, or to reverse underlying defects.

A further important area of application is the assessment of antibiotic impact. The danger of antibiotic therapy promoting resistance in commensal populations has long been recognised; however, such interventions will also promote changes in microbiota composition, for example, by conferring a selective advantage for species that have a natural tolerance for the agent being used. Such species may be pathogenic and, while unable to compete for niche space under normal conditions, are presented with an opportunity to expand during therapy. Characterisation of the effect of therapy on the microbiota, in addition to the intended target, is therefore increasingly considered in trials.

### KEY RESEARCH QUESTIONS

As research into the airway microbiome expands, a number of important questions present themselves. For example, to what extent is the altered airway microbiome present a cause or effect of disease? How do treatments, such as steroids, antibiotics or inhaled medication, affect the airway microbiome and what are the implications of this for disease? Can the airway microbiome be manipulated to change prognosis? Are changes in microbiome composition or behaviour predictive of lower respiratory tract infection? And how best can we standardise airways sampling and analysis to provide comparable datasets and protocols for their interpretation? The challenge now facing researchers is to use recent technological advances to answer these and other pressing clinical questions.

### CONCLUSIONS

The concept of airway microbiota analysis is relatively new and we are only now starting to move beyond an initial phase of simple cataloguing the identities and relative abundances of microbes associated with particular conditions. The number of important clinical correlations that link facets of airways

microbiota with markers of disease progression and clinical outcomes are increasing rapidly, with such investigations starting to be used to provide prognostic insight. We are entering a new era, with researchers trying to determine the mechanisms that underpin these associations with a view to identifying new therapeutic targets. However, with technical advances being achieved at an ever greater rate, it is of fundamental importance that this research continues to be driven by important clinical questions, with the aim of deriving outcomes that can be translated for direct patient benefit.

**Acknowledgements** The authors would like to thank Dr Steve Holden, Nottingham University NHS Hospitals Trust and the Anna Trust.

**Contributors** KDB and DS discussed the original idea for this article and are responsible for the overall content as guarantors. The first draft was written by GBR and KDB. All authors contributed to the work and have revised it critically for important intellectual content following the first draft. All authors have given approval for the final version.

**Competing interests** None.

**Provenance and peer review** Not commissioned; internally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

### REFERENCES

- 1 Segal LN, Rom WN, Weiden MD. Lung microbiome for clinicians. New discoveries about bugs in healthy and diseased lungs. *Ann Am Thorac Soc* 2014;11:108–16.
- 2 Beck JM, Young VB, Huffnagle GB. The microbiome of the lung. *Transl Res* 2012;160:258–66.
- 3 Rogers GB, Carroll MP, Serisier DJ, et al. Characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16S ribosomal DNA terminal restriction fragment length polymorphism profiling. *J Clin Microbiol* 2004;42:5176–83.
- 4 Tunney MM, Field TR, Moriarty TF, et al. Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. *Am J Respir Crit Care Med* 2008;177:995–1001.
- 5 Rogers GB, Zain NMM, Bruce KD, et al. A novel microbiota stratification system predicts future exacerbations in bronchiectasis. *Ann Am Thorac Soc* 2014;11:496–503.
- 6 Sibley CD, Duan K, Fischer C, et al. Discerning the complexity of community interactions using a *Drosophila* model of polymicrobial infections. *PLoS Pathog* 2008;4:e1000184.
- 7 Rogers GB, Hoffman LR, Carroll MP, et al. Interpreting infective microbiota: the importance of an ecological perspective. *Trends Microbiol* 2013;21:271–6.
- 8 Tunney MM, Einarsson GG, Wei L, et al. Lung microbiota and bacterial abundance in patients with bronchiectasis when clinically stable and during exacerbation. *Am J Respir Crit Care Med* 2013;187:1118–26.
- 9 Zhao J, Murray S, Lipuma JJ. Modeling the impact of antibiotic exposure on human microbiota. *Sci Rep* 2014;4:4345.
- 10 Klepac-Ceraj V, Lemon KP, Martin TR, et al. Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and *Pseudomonas aeruginosa*. *Environ Microbiol* 2010;12:1293–303.
- 11 Di Bella JM, Bao Y, Gloor GB, et al. High throughput sequencing methods and analysis for microbiome research. *J Microbiol Methods* 2013;95:401–14.
- 12 Goddard AF, Staudinger BJ, Dowd SE, et al. Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. *Proc Natl Acad Sci USA* 2012;109:13769–74.
- 13 Rogers GB, Skelton S, Serisier DJ, et al. Determining CF lung microbiology: a comparison of spontaneous and serially induced sputum samples using T-RFLP profiling. *J Clin Microbiol* 2010;48:78–86.
- 14 Willner D, Daly J, Whitley D, et al. Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS ONE* 2012;7:e34605.
- 15 Zhao J, Carmody LA, Kalikin LM, et al. Impact of enhanced *Staphylococcus* DNA extraction on microbial community measures in cystic fibrosis sputum. *PLoS ONE* 2012;7:e33127.
- 16 Klindworth A, Pruesse E, Schweer T, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 2013;41:e1.

- 17 Shah N, Tang H, Doak TG, *et al.* Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pac Symp Biocomput* 2011;165–76.
- 18 Lim YW, Evangelista JS III, Schmieder R, *et al.* Clinical insights from metagenomic analysis of sputum samples from patients with cystic fibrosis. *J Clin Microbiol* 2014;52:425–37.
- 19 Caporaso JG, Kuczynski J, Stombaugh J, *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335–6.
- 20 Schloss PD, Westcott SL, Ryabin T, *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75:7537–41.
- 21 van der Gast CJ, Walker AW, Stressmann FA, *et al.* Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J* 2011;5:780–91.
- 22 Ramette A. Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* 2007;62:142–60.
- 23 McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 2013;8:e61217.
- 24 Chao A. Nonparametric estimation of the number of classes in a population. *Scand J Stat* 1984;11:265–70.
- 25 Chao A, Lee S. Estimating the number of classes via sample coverage. *J Am Stat Assoc* 1992;87:210–17.
- 26 Brower JE, Zar JH, von Ende CN. *Field and laboratory methods for general ecology*. Boston: McGraw-Hill, 1998.
- 27 Hutcheson K. A test for comparing diversities based on the Shannon formula. *J Theor Biol* 1970;29:151–4.
- 28 Li K, Bihan M, Yooseph S, *et al.* Analyses of the microbial diversity across the human microbiome. *PLoS ONE* 2012;7:e32118.
- 29 Huang YJ, Nelson CE, Brodie EL, *et al.* Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. *J Allergy Clin Immunol* 2011;127:372–381.e1–3.
- 30 Huang YJ, Boushey HA. The microbiome and asthma. *Ann Am Thorac Soc* 2014;11 (Suppl 1):S48–51.
- 31 Falkow S. Molecular Koch's postulates applied to microbial pathogenicity. *Rev Infect Dis* 1988;10(Suppl 2):S274–6.
- 32 Conrad D, Haynes M, Salamon P, *et al.* Cystic fibrosis therapy: a community ecology perspective. *Am J Respir Cell Mol Biol* 2013;48:150–6.
- 33 Slater M, Rivett DW, Williams L, *et al.* The impact of azithromycin therapy on the airway microbiota in asthma. *Thorax* 2014;69:673–4.
- 34 Leung RK, Zhou JW, Guan W, *et al.* Modulation of potential respiratory pathogens by pH1N1 viral infection. *Clin Microbiol Infect* 2013;19:930–5.
- 35 Fierer N, Nemerugut D, Knight R, *et al.* Changes through time: integrating microorganisms into the study of succession. *Res Microbiol* 2010;161:635–42.
- 36 Shade A, Peter H, Allison SD, *et al.* Fundamentals of microbial community resistance and resilience. *Front Microbiol* 2012;3:417.
- 37 Zhu B, Xiao D, Zhang H, *et al.* MALDI-TOF MS distinctly differentiates nontypable *Haemophilus influenzae* from *Haemophilus haemolyticus*. *PLoS ONE* 2013;8: e56139.
- 38 Cattoir V, Narasimhan G, Skurnik D, *et al.* Transcriptional response of mucoid *Pseudomonas aeruginosa* to human respiratory mucus. *MBio* 2013;3:e00410–12.
- 39 Wiktorowicz JE, Jamaluddin M. Proteomic analysis of the asthmatic airway. *Adv Exp Med Biol* 2014;795:221–32.
- 40 Kozłowska J, Rivett DW, Carroll MP, *et al.* The impact of Pseudomonas growth on cystic fibrosis airway secretion composition in a metabolomic investigation. *Metabolomics* 2013;9:1262–73.