

Longitudinal Investigation of the Airway Microbiota in Children with Chronic Suppurative Lung Disease

Thesis submitted for the degree of
Doctor of Philosophy
September 2017

Dr Bushra Ahmed
MBBS BSc (Hons) MRCPCH
National Heart and Lung Institute,
Imperial College London

Abstract

Recurrent airway infection and its complications are the leading cause of death in Cystic Fibrosis (CF) but remain poorly understood. Molecular microbiology demonstrates that the CF airways are more polymicrobial than previously thought. The role of the microbiota in disease prognosis is unclear.

There have been few longitudinal studies of the airway microbiota in children, partly because sampling the lower airways is difficult as many children cannot expectorate. Part one of this thesis determined whether upper airway samples could act as surrogates for the lower airway microbiota in children. Cough swabs [CS], throat swabs [TS] and lower airway samples (bronchoalveolar lavage and bronchial brushings) were collected for 16S rRNA gene sequencing. There was good correlation between TS and the lower airways at a community but not an individual level; TS distinguished disease differences, suggesting that they could be used in longitudinal studies of the airway microbiota in children. CS sequenced poorly, thus precluding their use for molecular microbiology.

The longitudinal study had two arms; infants with CF diagnosed on newborn screening (NBS), in whom molecular microbiology revealed a small increase in diversity of the airway microbiota until 2 years of age; *Streptococcus* spp. and *Haemophilus* spp. were the most common organisms and showed an inverse relationship in their relative abundances over time. The second compared older children with CF with Primary Ciliary Dyskinesia (PCD), which shares a similar

pathology to CF but rarely the same rate of disease progression. This revealed similar trends to those seen in NBS in the relative abundance of *Streptococcus* spp. and *Haemophilus* spp. in PCD but not in CF. This suggests that early changes in CF mirror those seen in later childhood in PCD and a switch occurs in CF during childhood from a milder to a more pathological community composition.

Acknowledgements

Firstly, I would like to express special gratitude to my team of supervisors, Professor Andy Bush, Professor Miriam Moffatt, Professor Jane Davies and my assistant supervisor Dr Mike Cox, each one of whom I personally admire and from whom I have learnt immensely. Not only are they exceptional scientists and clinicians but also incredibly supportive and compassionate people and I feel privileged to have been mentored by them. The example they have set I hope to carry with me in my future career.

I would also like to thank all the staff at Molgen. In particular, Professor Bill Cookson, Dr Leah Cuthbertson and Dr Phill James for their vast patience and kindness, particularly in helping me learn to perform laboratory and analytical techniques. The support from the entire Molgen team was invaluable. I would also like to thank the staff at the RBH BRU for providing me with the opportunity to undertake this PhD and their support throughout. During my Paediatric training, I have seldom worked in a friendlier team than that of the Paediatric Respiratory Department at RBH. I owe thanks to the entire multi-disciplinary team and of course, the amazing patients, whose incredible enthusiasm in providing samples for the “cough swab lady” in exchange for as many stickers as they could get made my job even more enjoyable. A special mention goes to Dr Chris Grime, Dr Beccy Thursfield, Dr Julie Duncan and Dr Rishi Pabary for not only their support with my study but all their friendship. And to my best friends, Ying Ying, Sam and K for always being there for me no matter what.

Finally and most importantly, I would like to thank my family. To my parents and sisters, without whose unwavering belief in me and encouragement, I might never have believed that I could do a PhD. I owe the majority of my achievements in life to them. And to my wonderful husband Andy, who never stops supporting me in my career, never complained when I spent more time with my laptop than with him and always knows how to make me laugh, even in tough times. My greatest inspiration.

Declaration of originality

I declare that all the work within this thesis is my own unless stated otherwise and that all external sources are appropriately acknowledged.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Table of contents

Abstract	2
Acknowledgements	4
Declaration of originality.....	5
List of figures.....	11
List of tables.....	18
List of abbreviations	20
 Chapter 1: Introduction.....	 24
1.1 The human microbiome.....	24
1.1.1 Lower respiratory tract infections	24
1.1.2 Studying bacteria and the 16S rRNA bacterial gene.....	24
1.1.3 The microbiome in health and disease.....	30
1.1.4 The airway microbiota in healthy individuals.....	30
1.1.5 Development of the healthy airway microbiota	34
1.1.6 The role of the microbiota on host immunity and airway disease	41
1.2 Investigating respiratory infections in children.....	42
1.3 Chronic suppurative lung disease	43
1.3.1 Cystic fibrosis.....	44
1.3.1.1 Genetics of CF	45
1.3.1.2 Pathophysiology.....	48
1.3.1.3 Infections and exacerbations in CF	52
1.3.1.4 The role of the microbiota in CF	57
1.3.1.5 Disturbances in the microbiota.....	61
1.3.1.6 The role of antibiotics	68

1.3.1.7	Genotype.....	69
1.3.1.8	The lung-gut axis.....	69
1.3.1.9	Manipulating the microbiota.....	70
1.3.2	Primary Ciliary Dyskinesia	70
1.3.2.1	Genetics and pathophysiology of PCD	71
1.3.2.2	Infections and the PCD microbiota.....	74
1.4	Hypotheses, aims and objectives.....	76
Chapter 2: Materials and methods		78
2.1	List of reagents, kits and equipment used	78
2.2	Study population.....	80
2.3	Clinical information collation.....	82
2.4	Sample collection and handling.....	86
2.4.1	Oropharyngeal swab collection.....	87
2.5	Bacterial DNA extraction	88
2.6	Quantitative PCR (qPCR) for quantification of bacterial load	90
2.7	16S rRNA gene PCR amplification assay	92
2.8	Amplicon purification.....	94
2.9	DNA quantification and equimolar pooling for library normalisation and repeat purification.....	96
2.10	Agarose gel purification of equimolar pool	97
2.11	Quantitative PCR (qPCR) for library quantification	99
2.12	16S rRNA gene sequencing using the Illumina MiSeq	100
2.13	Data analysis.....	104
2.13.1	Sequence analysis in QIIME.....	104
2.13.2	Microbial community analysis in R	107
2.13.2.1	Quality and contamination control checks	108

2.13.2.2	Pre-processing.....	109
2.13.2.3	Diversity analyses.....	110
2.13.2.4	OTU level comparisons	112
 Chapter 3: Comparison of the upper and lower airway microbiota in children with chronic lung diseases..... 113		
3.1	Introduction	113
3.2	Methods	113
3.2.1	Subjects.....	113
3.2.2	Sampling.....	114
3.2.2.1	Bronchoalveolar lavage samples (BALF).....	114
3.2.2.2	Bronchial brushings.....	115
3.2.3	16S rRNA gene library preparation and sequencing.....	116
3.2.4	Data analysis.....	117
3.3	Results.....	119
3.3.1	Patient demographics & sampling.....	119
3.3.2	Sequencing.....	123
3.3.3	Comparison of the upper and lower airway microbiota.....	131
3.3.4	Clinical variables and the upper and lower airway microbiota	141
3.4	Discussion.....	143
 Chapter 4: Initial development of the airway microbiota in infants with cystic fibrosis. 151		
4.1	Introduction	151
4.2	Methods	152
4.2.1	NBS study patient recruitment and follow-up	152
4.2.2	Sample collection & processing.....	153
4.2.3	Data analysis.....	153

4.3 Results.....	158
4.3.1 Patient demographics.....	158
4.3.2 Examining sequencing quality.....	165
4.3.3 Development of the airway microbiota in the first 2 years of life.....	168
4.3.3.1 Changes in bacterial load.....	168
4.3.3.2 Genera and OTU level changes	170
4.3.3.3 Changes in alpha diversity.....	178
4.3.3.4 Changes in beta-diversity.....	181
4.3.3.5 Perinatal predictors of community structure.....	181
4.3.3.6 The impact of the first exacerbation.....	184
4.3.3.7 The impact of the first growth of <i>P.aeruginosa</i>	189
4.3.3.8 The lower airway microbiota in infancy	195
4.4 Discussion	199
 Chapter 5: Comparison of the airway microbiota in Cystic Fibrosis and Primary	
Ciliary Dyskinesia	212
5.1 Introduction	212
5.2 Methods	213
5.2.1 Patient recruitment and follow-up.....	213
5.2.2 Clinical information.....	213
5.2.3 Sample collection & processing.....	214
5.2.4 Data analyses	215
5.3 Results.....	217
5.3.1 Patient demographics.....	217
5.3.2 Sampling frequency.....	222
5.3.3 Examining sequencing quality.....	227
5.3.4 Spontaneously expectorated sputum versus throat swabs.....	230

5.3.5	Development of the airway microbiota in children with CF	236
5.3.5.1	Genera and OTU level changes with age.....	236
5.3.5.2	Changes in community diversity with age.....	242
5.3.5.3	Clinical variables influencing community structure	243
5.3.5.4	Changes with a pulmonary exacerbation in children with CF	246
5.3.6	Development of the airway microbiota in children with PCD.....	250
5.3.6.1	Genera and OTU level changes with age.....	250
5.3.6.2	Changes in community diversity with age.....	256
5.3.6.3	Clinical variables influencing community structure	258
5.3.6.4	Changes at exacerbation	260
5.3.7	Comparison of the microbiota between CF and PCD	265
5.3.8	The impact of antibiotics on diversity of the CF microbiota	267
5.4	Discussion	267
Chapter 6:	General discussion	281
References.....		297
Appendices.....		324

List of figures

Figure 1.1:	Structure of the 16S rRNA gene	27
Figure 1.2:	Overview of physical, chemical and innate and adaptive immune defences which protect the airways from allergens, microorganisms and pollutants in inspired air	32
Figure 1.3:	Illustration of species richness (the number of different species) and evenness (the skew in species distribution)	37
Figure 1.4:	Illustration of the three factors thought to determine the airway microbiota: immigration, elimination and proliferation of community members	40
Figure 1.5:	The six classes of mutations in CFTR	46
Figure 1.6:	Illustration of the changing prevalence of “traditional” CF pathogens with age in UK	53
Figure 1.7:	Cross-sectional diagram of the ultrastructure of motile cilia	72
Figure 2.1:	Library preparation workflow for 16S rRNA gene sequencing	103
Figure 2.2:	Workflow diagram for data analyses	105
Figure 3.1:	Rarefaction curves with yellow lines denoting the number of OTUs sampled at 1,000 reads and 3,000 reads	118
Figure 3.2:	Sequencing depth for samples in upper and lower airway microbiota comparison	119

Figure 3.3:	Illustration of the combinations of upper and lower airway samples taken	122
Figure 3.4:	Examination of PCR negative controls, technical, sampling controls and all controls combined for contaminant OTUs	125
Figure 3.5:	Examination of PCR negative controls, technical, sampling controls and all controls following removal of suspected contaminants <i>Burkholderia</i> (OTU ID 1606), <i>Undibacterium</i> (OTU ID 1727) and <i>Ralstonia</i> (OTU ID 1703) to confirm removal.	126
Figure 3.6:	Non-metric multidimensional scaling (NMDS) plot to examine for sequencing consistency between each plate of samples	127
Figure 3.7:	NMDS plot to examine for batch effect	128
Figure 3.8:	Barplot of 27 genera present in the mock communities for each sequencing plate	129
Figure 3.9:	Comparing bacterial load (16S rRNA copy number per swab) between paired CS and TS	130
Figure 3.10:	Heatmap of the 50 most common OTUs present between upper (TS and CS) and lower airway samples (BALF and bronchial brushings)	133
Figure 3.11:	Individual patient barplots (N = 39) comparing paired lower airway samples and TS for the 50 most common OTUs.	134

Figure 3.12:	Alpha-diversity comparisons between lower airway samples and TS	136
Figure 3.13:	Alpha-diversity comparisons between lower airway samples and TS	137
Figure 3.14:	Non-metric multidimensional scaling (NMDS) plot comparing the unweighted UniFrac score between TS and lower airway samples	138
Figure 4.1a:	Histogram of sequencing depth for samples in NBS cohort.	155
Figure 4.1b:	Rarefaction curve calculated using Shannon Diversity Index measurements	156
Figure 4.2	Summary of samples obtained for infants diagnosed with CF on newborn screening before and after rarefaction to 600 reads.	157
Figure 4.3a:	Illustration of the number of monthly samples collected.	163
Figure 4.3b:	Illustration of sample frequency clustered into age ranges	164
Figure 4.4:	Barplot comparing mock communities run on each sequencing plate to assess for consistency between sequencing plates for NBS study	167
Figure 4.5:	Changes in bacterial load with age on TS obtained in the first 2 years of life	168
Figure 4.6:	Sankey plot illustrating the most common species	172

present in the airway microbiota in the first 2 years of life

Figure 4.7:	Changes in the relative abundance of five most common genera with age in infants with CF	173
Figure 4.8:	Heatmap illustrating changes in the 50 most common OTUs in the first 2 years in ascending age	174
Figure 4.9:	Example plot of an individual patient illustrating changes in relative abundance of OTUs against bacterial load (16S rRNA gene copies per swab), diversity changes and clinical variables	177
Figure 4.10:	Scatterplot of changes in Shannon diversity index with age	178
Figure 4.11:	Changes in alpha-diversity by age range illustrating an increasing trend in alpha diversity with age	180
Figure 4.12:	Changes in bacterial load (16S rRNA gene copies per throat swab) with exacerbations (N = 5)	185
Figure 4.13:	Changes in the relative abundance of the five most abundant genera with exacerbations	188
Figure 4.14:	Bacterial load in relation to <i>P. aeruginosa</i> growth and treatment.	190
Figure 4.15:	Changes in the relative abundance of the five most common genera in patients who grew <i>P. aeruginosa</i> in the first two years of life	191
Figure 4.16:	Individual patient barplots of the twenty most common	193

genera prior to *P. aeruginosa* growth (Pre-Pa), at the time of *P. aeruginosa* growth (Pa) and after *P. aeruginosa* growth (Post-Pa)

Figure 4.17:	Boxplots illustrating change in alpha diversity before (Pre-Pa), during (Pa) and after (Post-Pa) growth of <i>P. aeruginosa</i> on bacterial cultures	194
Figure 4.18:	Sankey plot showing the relative abundance of the most common genera in lower airway samples in infancy	197
Figure 4.19:	Comparing alpha diversity on lower airway samples between patients who grew <i>P. aeruginosa</i> with those who did not	198
Figure 5.1:	Justification for rarefaction to 1000 reads. Illustrating the rarefaction curve calculated using Shannon Diversity Index measurements	216
Figure 5.2	Summary of the number of samples in each study group (CF or PCD) before and after rarefaction to 1000 reads.	223
Figure 5.3a:	Frequency of samples taken per month of the study duration for patients with CF	224
Figure 5.3b:	Frequency of samples taken per month of the study duration for patients with PCD	225
Figure 5.4:	Sampling frequency grouped into four study timepoints: Baseline, 3-6 months, 9-12 months and 15-18 months.	226
Figure 5.5:	Contaminant identification & removal	228
Figure 5.6:	Stacked barplots comparing mock communities of each	229

	sequencing run to assess for consistency between sequencing runs	
Figure 5.7:	Heatmap comparing the relative abundance of OTUs in sputum and TS	232
Figure 5.8:	Example individual patients' barplots for expectorated sputum versus throat swab comparison.	233
Figure 5.9:	Comparing richness between sputum and TS	235
Figure 5.10:	Sankey plot illustrating the most common genera present in the airway microbiota in children with CF	239
Figure 5.11:	Changes in the relative abundance of the 5 most common genera with age throughout childhood in CF	240
Figure 5.12:	Example plot of an individual patient with CF illustrating changes in relative abundance of OTUs against spirometry measurements (FEV_1), diversity changes (Inverse Simpson's) and clinical variables (antibiotic use, organisms grown on clinical microbiology and presence of upper and lower airway symptoms)	241
Figure 5.13:	Changes in alpha-diversity by age range for CF	242
Figure 5.14:	Changes in the relative abundance of the 5 most abundant genera with exacerbations in CF	249
Figure 5.15:	Sankey plot illustrating the most common genera present in the airway microbiota in PCD throughout childhood	253
Figure 5.16:	Changes in the relative abundance of the 5 most common genera with age throughout childhood in PCD	254

Figure 5.17:	Example plot of an individual patient with PCD	255
Figure 5.18:	Changes in alpha-diversity by age range for PCD	257
Figure 5.19:	Boxplot showing difference in richness between patients with PCD with different ciliary ultrastructural defects	259
Figure 5.20:	Changes in the 5 most common genera with exacerbations, treatment and recovery for children with PCD	263
Figure 5.21:	Non-metric multidimensional scaling (NMDS) plot illustrating differences in beta diversity (shown here using Bray Curtis dissimilarity) for patients with PCD receiving IV antibiotics for treatment of an exacerbation at baseline ("B), exacerbation ("E"), treatment (at week 1 "T1" and week 2 "T2") and recovery ("R")	264
Figure 5.22:	Non-metric multidimensional scaling (NMDS) comparison of Bray Curtis dissimilarity children (aged 0.5– 17 years) with CF and PCD	266

List of tables

Table 1.1:	List of taxonomic ranks in descending order	25
Table 1.2:	Koch's postulates for identifying the microbial causes of infectious diseases from pure culture	25
Table 1.3:	Fredricks and Relman's revised postulates	29
Table 1.4:	Description of anaerobes that have previously been identified in CF patients on culture independent studies	67
Table 2.1:	Follow-up protocol for the three study groups in CLIMB study	80
Table 2.2:	Clinical information collected during study visits.	85
Table 3.1:	Summary of patient characteristics in the comparison of the upper and lower airway microbiota	121
Table 3.2:	Summary of sequencing quality statistics for the comparison of the upper and lower airway microbiota	123
Table 3.3:	Comparing the mean relative abundance (percent) of the most common or clinically important genera between TS and paired lower airway samples (BALF or bronchial brushing)	140
Table 3.4:	Beta diversity summaries of significant clinical variables influencing community structure	143
Table 4.1:	Patient demographics for NBS infants in the CLIMB study	161
Table 4.2:	Summary of sequencing quality statistics for the NBS cohort in the CLIMB study	165

Table 4.3:	Differences in bacterial load for clinical variables	169
Table 4.4:	List of OTUs showing significant ($P_{\text{adj}} < 0.05$) changes in relative abundance with age	176
Table 4.5:	Testing the influence of clinical variables on beta diversity	183
Table 4.6:	Changes in the relative abundance (%) of common genera between baseline and exacerbation samples	186
Table 5.1:	Patient demographics comparing children with CF and PCD	220
Table 5.2:	Organisms grown on clinical microbiology during study period for patients with CF and PCD	221
Table 5.3:	Beta diversity comparisons between TS and sputum for CF and PCD	236
Table 5.4:	Testing the influence of clinical variables on alpha and beta diversity for children with CF.	245
Table 5.5:	Changes in relative abundance of 10 most common genera between baseline and exacerbation samples in CF	247
Table 5.6:	Changes in the relative abundance of the 10 most common genera between baseline and exacerbation samples	261

List of abbreviations

ABC	ATP-binding cassette
ABPA	Allergic bronchopulmonary aspergillosis
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
ASL	Airway surface liquid
BALF	Bronchoalveolar lavage fluid
BCC	<i>Burkholderia cepacia</i> complex
BMI	Body mass index
BWA	Burrows-Wheeler Aligner
CF	Cystic fibrosis
CFRD	Cystic fibrosis related diabetes
CFTR	Cystic fibrosis transmembrane conductance regulator
CFU	Colony forming units
CI	Confidence interval
CLIMB	Chronic suppurative lung disease Longitudinal Investigation of the Airway Microbiota
CRP	C-reactive protein
CS	Cough swab
CSLD	Chronic suppurative lung disease
CT	Computed tomography
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
ENaC	Epithelial sodium channels

FDR	False discovery rate
FEV ₁	Forced expiratory volume in 1 second
FOB	Fibreoptic bronchoscopy
FVC	Forced vital capacity
FWER	Family-wise error rate
g	Gauge
GI	Gastrointestinal
GORD	Gastro-oesophageal reflux disease
<i>H. influenzae</i>	<i>Haemophilus influenzae</i>
HMP	Human Microbiome Project
HRCT	High resolution computed tomography
IL-8	Interleukin-8
IV	Intravenous
LCI	Lung clearance index
LMA	Laryngeal mask airway
LME	Lysing Matrix E tube
LRTIs	Lower respiratory tract infections
LSCS	Lower segment Caesarian section
MCC	Mucociliary clearance
μl	Microliters
μM	Micromolar
Molgen	Molecular Genetics and Genomics Group
MRSA	Methicillin resistant <i>Staphylococcus aureus</i>
MSSA	Methicillin sensitive <i>Staphylococcus aureus</i>
NBS	Newborn screening

NGS	Next generation sequencing
NHLI	National Heart and Lung Institute
NMDS	Non-metric multi-dimensional scaling
nNO	Nasal nitric oxide
NSD	Nucleotide spanning domain
NTM	Non-tuberculous mycobacteria
NTP	Nucleoside triphosphates
OTU	Operational taxonomic unit
Q30	Phred quality score > 30
<i>P. aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
PCD	Primary ciliary dyskinesia
PCL	Periciliary layer
PCR	Polymerase chain reaction
PERMANOVA	Permutational multivariate analysis of variance
PyNASt	Python Nearest Alignment Space Termination tool
PVL	Panton-Valentine Leucocidin
QIIME	Quantitative Insights Into Microbial Ecology
qPCR	Quantitative polymerase chain reaction
RBH	Royal Brompton Hospital
RML	Right middle lobe
rRNA	Ribosomal ribonucleic acid
RSV	Respiratory Syncytial Virus
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
SD	Standard deviation
SMG	<i>Streptococcus milleri</i> group

SOP	Standard operating procedure
SPRI	Solid phase reversible immobilisation
SVD	Spontaneous vaginal delivery
TNF- α	Tumour necrosis factor- α
T-RFLP	Terminal restriction fragment length polymorphism
TS	Throat swab
URT	Upper respiratory tract

Chapter 1: Introduction

1.1 The human microbiome

1.1.1 Lower respiratory tract infections

Lower respiratory tract infections (LRTIs) are a significant disease burden, affecting the lives of many children. In healthy children, acute LRTIs can be caused by a wide variety of pathogens including viruses, such as Respiratory Syncytial Virus (RSV), and bacteria, such as *Streptococcus pneumoniae* and *Mycoplasma pneumoniae*. In contrast, chronic respiratory infections are predominantly due to opportunistic organisms in children with chronic lung diseases. Chronic infections as well as acute flare-ups (pulmonary exacerbations, or 'lung attacks'¹) are the leading cause of morbidity and mortality in children with underlying airway disorders, such as cystic fibrosis (CF) and primary ciliary dyskinesia (PCD). Consequently better understanding and management of chronic airway infections is essential to improve prognosis.

1.1.2 Studying bacteria and the 16S rRNA bacterial gene

Bacteria are one of the most abundant life forms on Earth, with an estimated total of approximately $4-6 \times 10^{30}$ cells living in mixed populations². The study of communities of microorganisms is perhaps as old as the study of microbiology itself, beginning with Antonie van Leeuwenhoek's discovery of bacteria or "animalcules" of different forms in the dental plaque scraped from between his teeth in 1683³. However, it was not until 200 years later that the greatest advancements in our understanding of the pathogenic role of microorganisms

were made. Spearheaded by the creation of methods for the isolation of bacteria from pure culture by Robert Koch in the late 1800s, bacteria were classified into taxonomic groups based on appearance and nutritional requirements (taxonomic ranks given in Table 1.1). This then allowed the microbial causes of infectious diseases to be more clearly identified based on guidelines now known as “Koch’s postulates” (Table 1.2)⁴.

Table 1.1: List of taxonomic ranks in descending order.

Example given using genus *Pseudomonas*.

Taxonomic rank	Example
Kingdom	Bacteria
Phylum	Proteobacteria
Class	Gammaproteobacteria
Order	Pseudomonadales
Family	Pseudomonadaceae
Genus	<i>Pseudomonas</i>
Species	<i>Pseudomonas aeruginosa</i>

Table 1.2: Koch's postulates for identifying the microbial causes of infectious diseases from pure culture.

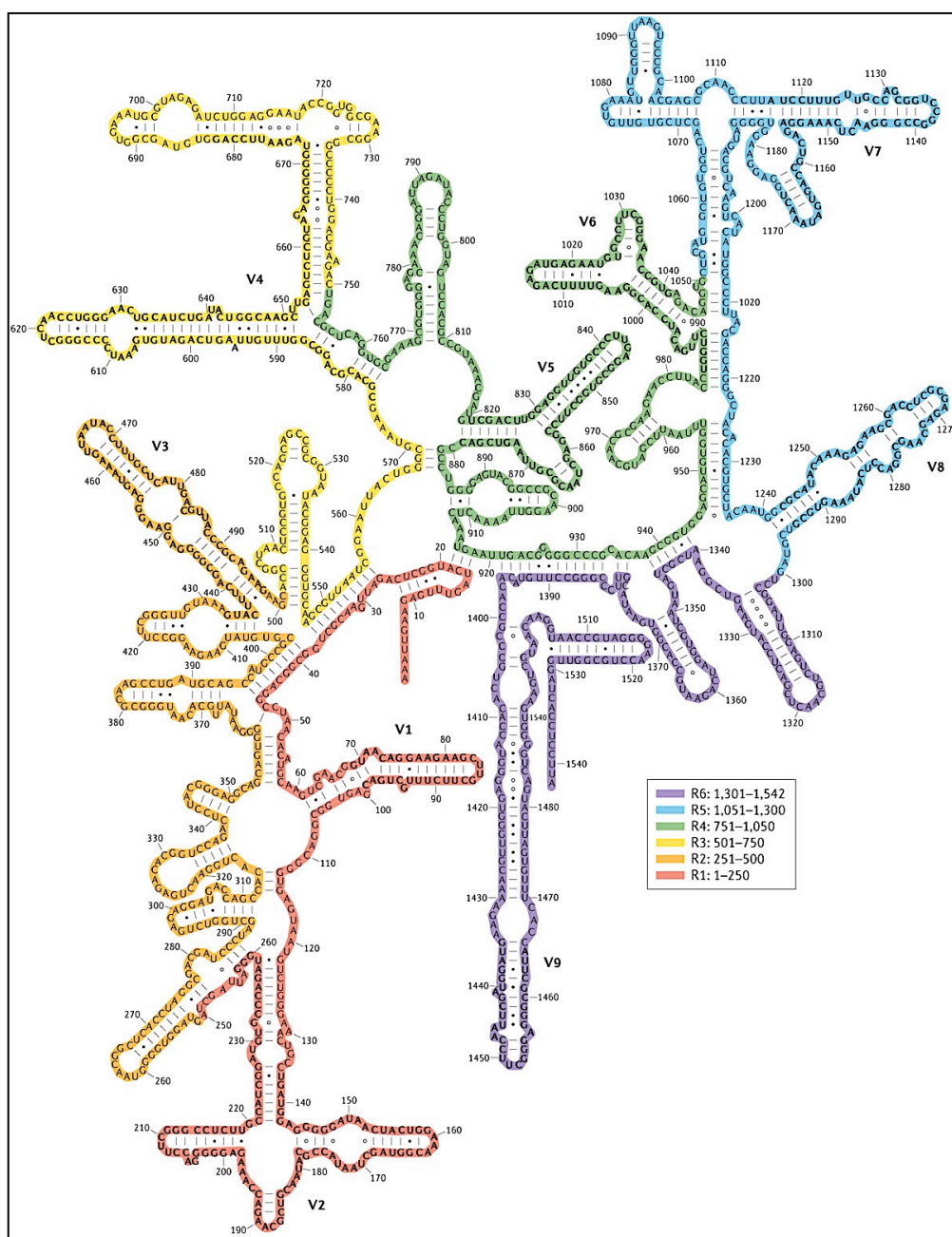
Koch’s postulates
The microorganism should: <ul style="list-style-type: none"> • Account for the pathological changes seen associated with the disease • Be able to be isolated in pure culture from a diseased organism • Cause disease when introduced into a healthy host

Whilst bacterial cultures have remained the mainstay of microbiological diagnosis for the past 150 years, they are not without their limitations. Many bacteria are difficult to culture *in vitro*. The successful growth of a microorganism requires knowledge of the optimum conditions needed (such as temperature, pH and oxygen tension) and multiple culture plates are often required for polymicrobial samples. The “great plate count anomaly”⁵ estimated that only 1%⁶ of viable bacterial which can be visualized microscopically can be grown using standard cultures^{7,8}. With enhanced, targeted phenotypic culture techniques, however, it has recently been demonstrated that over 60% of bacteria resident in the human gut are capable of being grown *in vitro*⁹. Nonetheless, microbial cultures cannot currently identify all human associated bacteria and may underestimate bacterial diversity within a community and potentially miss important pathogens.

In contrast, molecular gene-targeted microbial identification offers a much broader insight and has revolutionized our understanding of microbial communities. The 16S rRNA gene, which encodes part of the small 30S subunit of the bacterial ribosome, is an attractive target for culture-independent bacterial identification. As an essential housekeeping gene, it is ubiquitous in all prokaryotes and has changed little over time¹⁰. It contains highly conserved regions, suitable for annealing universal primers, and 9 variable regions, the sequence variation of which can be used to taxonomically group bacteria to at least the genus level^{11,12} and make estimations of entire microbial communities (Figure 1.1).

Figure 1.1: Structure of the 16S rRNA gene.

Example shown for *Escherichia coli* consisting of broadly conserved regions flanked by nine variable regions (V1 – V9). Reprinted with permission from Macmillan Publishers Ltd: Nature Publishing Group. Cite: Yarza P *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nature Reviews in Microbiology; 2014: 12(9): 635 – 645¹³. Copyright © 2014.



Initial attempts utilizing the 16S rRNA gene to characterise bacterial communities relied on community fingerprinting techniques, such as terminal restriction fragment length polymorphism (T-RFLP)^{14,15}. In the last 10 years, however, these have largely been superseded by next generation sequencing (NGS), such as Illumina MiSeq and the now discontinued Roche 454 pyrosequencing. With traditional Sanger sequencing, gene sequences are manually read on a gel with 4 lanes, one for each nucleotide. In contrast NGS performs many PCR reactions simultaneously with fluorescently labelled nucleotides, such that their sequences can be automatically read in real time. This allows large numbers of genes to be sequenced rapidly, cheaply and with less manpower^{16,17}. As a result of 16S rRNA gene sequencing, it has been realised that microbial communities throughout the body appear far more complex than previously appreciated, challenging our understanding of the role of microbes in disease processes.

With the increase in NGS technologies, the study of microbial communities or the “microbiota” in health and disease has expanded substantially. The term “microbiota” describes a community of microorganisms and has frequently been used synonymously with “microbiome”, which is defined as all the microorganisms in a given environment together with their genes and given interactions¹⁸. In addition, 16S rRNA gene sequencing has enabled the identification of novel bacteria. In 1980, *The Approved Lists of Bacterial Names* (<http://www.bacterio.net/-number.html#total>) listed 1,791 names of known bacterial species. By January 2017, this list of names had risen to 12,391, an increase of approximately 500%. Collectively these advancements provide a

unique opportunity to gain further insights into the role of bacterial communities in health and disease.

With greater discovery of bacteria using culture-independent techniques, Koch's postulates and a single pathogen view of disease causation are increasingly difficult to satisfy. Fredricks and Relman⁴ have therefore proposed a revised set of guidelines to aid in the identification of pathogenic bacteria (Table 1.3). Whilst an improvement, the revised postulates fail to account for some known disease associations such as that between Human Papillomavirus and cervical cancer⁴.

Table 1.3: Fredricks and Relman's revised postulates.

Based on identification of bacterial genes using culture-independent techniques⁴.

Fredricks and Relman's revised postulates
The nucleic acid sequences of a suspected microbial pathogen should be:
<ul style="list-style-type: none">• Present in most cases of an infectious disease• Preferentially found in diseased but not disease-free tissues• Absent or found in low abundance in disease-free tissues• Reduced or undetectable (in copy numbers) with disease resolution and rise with relapse• Consistent with a microorganism with known biological characteristics of that group of organisms• Confirmed with tissue-sequence correlates and visible microorganisms in areas of presumed disease• Predate disease or correlate with severity of disease (in copy numbers)• Reproducible

1.1.3 The microbiome in health and disease

The human body is inhabited by trillions of microbes, with bacterial cells originally postulated to outnumber human cells, although recent estimates accounting for the large number of non-nucleated red cells in the body suggest a revised ratio of 1:1¹⁹. These microbes have co-evolved with each other and their human host to create a “super-organism” of complex ecosystems. Bacterial communities have been found on epithelial surfaces throughout the body including the gut, skin^{20,21}, oral cavity²², nasopharynx^{23,24} and urogenital tract²⁵.

Some parts of the body have been extensively studied, e.g. the gut, where the microbiota has recognised important biological functions, such as aiding in digestion¹¹, developing immunotolerance to attenuate or abrogate autoimmune disease and providing “colonization resistance” to protect the host against invading pathogens²⁶. In contrast, the microbiota of other body habitats, such as the lungs, have only recently been discovered²⁷, are less well studied and their role in disease is unclear.

1.1.4 The airway microbiota in healthy individuals

The adult airway typically has an airflow of ~ 7.5 litres per minute²⁸ at rest and is constantly exposed to microorganisms, allergens and particulate matter. To protect against this constant exposure the airways have physical, mechanical and chemical barriers.

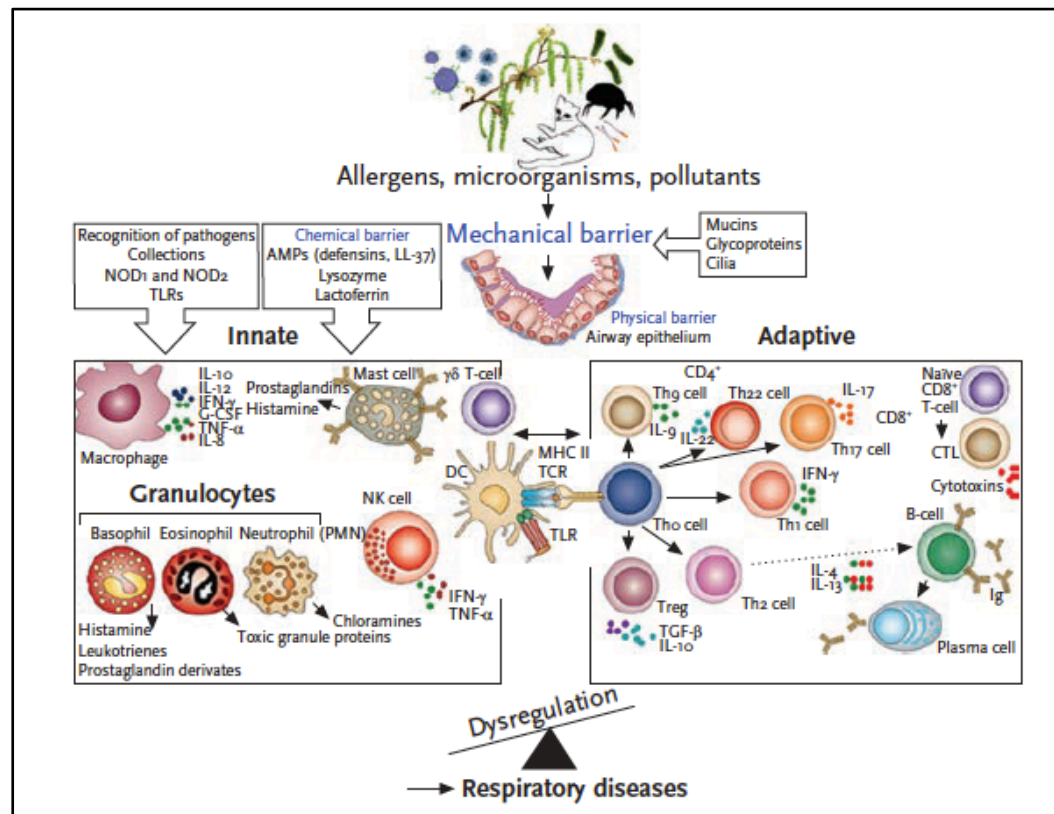
Beginning at the nose the airways are lined with pseudostratified epithelium, consisting of basal cells, goblet cells, serous cells and ciliated epithelial cells that collectively act as a physical barrier. Initial mechanical barriers, mucins and glycoproteins at the epithelial surface prevent invasion by microorganisms by trapping them in mucus, which can then be removed effectively by mucociliary clearance (MCC) through the coordinated action of motile cilia²⁹. Mucosal innate immune cells, such as macrophages, granulocytes (neutrophils, eosinophils, basophils) and natural killer cells protect against infection by opsonisation. In addition, toll-like receptors and dendritic cells present antigens for adaptive immune responses by T cells. In response to pathogen exposure, innate immune cells release antimicrobial peptides into the lumen, along with chemokines and cytokines (Figure 1.2)³⁰. When these mechanisms fail, bacterial infections can become established.

Figure 1.2: Overview of physical, chemical and innate and adaptive immune defences which protect the airways from allergens, microorganisms and pollutants in inspired air.

Reprinted with permission from the European Respiratory Society. Copyright © 2013. Cite: ERS Handbook of Paediatric Respiratory Medicine (page 20).

Editors: Eber E and Midulla F. First Edition 2013. doi:

10.1183/9781849840392-hbp01³⁰.



Until recently, the lower airways were thought to be sterile in health. Hence the lower airways were not sampled when one of the largest studies of the healthy human microbiome to date, the National Institute for Health Human Microbiome Project (HMP), was initiated. The HMP involved 300 healthy adults who were sampled across 18 body sites³¹. Using 16S rRNA gene PCR, however, the airways have been shown to contain 2,000 bacterial genomes per cm² of surface area sampled. Thus the lungs are not sterile in healthy individuals, as previously

thought, but harbour a significant number of microbes, questioning our understanding of the pathogenesis of LRTIs.

Anatomically, the airways can be divided into the upper airway (nasopharynx, oropharynx [throat, soft palate and tonsils] and larynx) and lower airway (trachea, bronchi, bronchioles and alveoli). There have been several attempts to characterise the organisms present in the healthy lower airway microbiota. These have been hampered in part by the difficulties in directly sampling the lower airways (e.g. the invasiveness of bronchoscopy), particularly in children. As a consequence, many such studies have included fewer than 20 participants^{27,32-34}. To my knowledge, the largest study to date included 64 healthy adult volunteers, 19 of whom were smokers³⁵. Only one study to date has examined the lower airway in children (N = 7) without chronic lung disease²⁷.

Despite this there are some similarities between studies. In particular, anaerobes, such as *Veillonella* spp. and *Prevotella* spp., have been found to be highly abundant in the healthy lower airway^{27,32-36}. Anaerobes are organisms which have traditionally been difficult to culture in the laboratory due to the oxygen-free conditions needed for growth. In addition, studies in adults have identified several recognised airway pathogens residing in the healthy lower airway, such as *Streptococcus* spp.³²⁻³⁶, *Haemophilus* spp.^{33,35} and *Neisseria* spp.^{32,36}, challenging the previous notion that bacterial newcomers lead to infections.

The upper airway microbiota, however, can be more easily accessed using non-invasive methods such as nasal swabs or washings and has consequently received greater attention. Comparisons of the microbiota along the respiratory tree have found that the nasopharyngeal microbiota is distinct from that of the oropharynx and lower airways and changes with age^{23,37}. In the first week of life, a study of 102 healthy infants found that the nasopharyngeal microbiota is dominated by the skin commensal *Staphylococcus aureus* (*S. aureus*) regardless of mode of delivery. Over the next 6 months, this is gradually replaced by *Moraxella catarrhalis*, *Corynebacterium* spp. and *Dolosigranulum pigrum*³⁸. By 18 months of age, a similar study of 96 healthy infants found a dominance of *Moraxella* spp., followed by *Haemophilus* spp., *Streptococcus* spp. and *Flavobacterium* spp.²⁴. In contrast, the oropharyngeal microbiota has been found to resemble the microbiota of the oral cavity and lower airways and is dominated by *Streptococcus* spp., followed by *Staphylococcus* spp., *Haemophilus* spp. and *Prevotella* spp.³⁹. Overall this suggests that a similar microbiota exists between the oropharynx and lower airway. The oropharynx may serve as a source of the lower airway microbiota by microaspiration^{32,35}, or the lower airway may contaminate the upper airway during coughing.

1.1.5 Development of the healthy airway microbiota

Previously it was thought that the *in utero* environment was sterile and that microbial colonisation began after birth upon first exposure to environmental microbes. It has been suggested, however, that the microbiota may begin to establish itself *in utero*, with reports of a unique microbiota in meconium^{40,41},

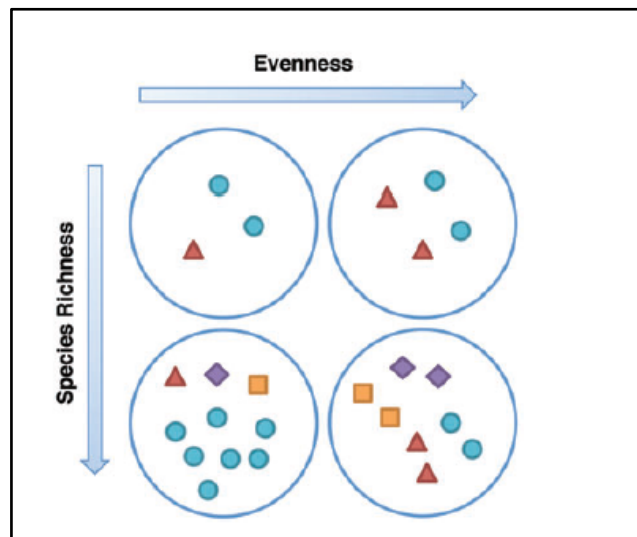
amniotic fluid⁴², umbilical cord blood⁴³ and placenta^{42,44,45}. For the latter, however, a study of six women was unable to distinguish a unique placental microbiota from that of contamination control swabs⁴⁶, suggesting that some of the observed profiles attributed to an in utero microbiota may in fact be spurious contaminants due to low biomass in these samples. Postnatal microbial communities have been detected in the upper airways of healthy infants from as early as 5 minutes of age⁴⁷. Communities have also been detected in tracheal aspirates immediately after birth in intubated term and preterm infants, even in those infants delivered by Caesarean section (LSCS) with intact membranes and no evidence of chorioamnionitis^{48,49}. The airway microbiota may therefore begin to be established very early in life, potentially before birth through transfer from the amniotic fluid or haematogenous spread from the placenta.

How the airway microbiota develops thereafter and its influence on health and disease is unclear. Some insights can be gained from studies of the microbiota in the gut, the most extensively studied niche in the human body, from which it appears that colonising populations within a community may have clear trajectories. In the infant gut, there is a general trend towards increasing biomass and successive diversification of the microbiota, with initial dominance of facultative anaerobes and later species replacement with obligate anaerobes as falling oxygen tensions within the colon confer a selective advantage. Shifts in community composition were seen with weaning between 4 – 6 months of age, highlighting the importance of dietary influences on informing the gut microbiota. By the end of the first 3-5 years of life a stable, healthy microbiome is established resembling that of adults^{50,51}.

There has been much interest in diversity of the microbiota and its potential influence on health and disease. Community diversity can be measured within subjects (alpha diversity), and between subjects (beta diversity). Alpha diversity is a measure of community “richness”, defined as the number of different species, and “evenness”, defined as the relative abundance of species (Figure 1.3). It is often expressed by either the Shannon Diversity Index or the Inverse Simpson’s Diversity Index¹⁸. It has been thought that in a diverse and balanced community, individual microorganisms are less likely to dominate and cause disease⁵². Therefore community diversity may be desirable in evading infections.

Figure 1.3: Illustration of species richness (the number of different species) and evenness (the skew in species distribution).

Different shapes and colours represent different organisms. With increasing richness (shown from top to bottom), there is an increase in the number of different species present. With increasing evenness (shown from left to right), there is a more balanced spread of organisms with no single organism able to dominate. Alpha-diversity, a composite measure of both species evenness and richness, increases with increasing species richness and evenness. Reprinted with permission from Oxford University Press. Copyright © 2013. Cite: Cox, M.J. *et al.* Sequencing the human microbiome in health and disease. *Human Molecular Genetics* 2013; 22(1): R88 – R94¹⁸.



There have been conflicting reports on the role of diversity in disease processes depending on the body site sampled. For example, in the gut low diversity of the microbiota has been linked to inflammatory bowel disease⁵³ and obesity⁵⁴. Similarly, low diversity has been found to confer susceptibility to *Salmonella* spp. induced gut inflammation in mice⁵⁵. In the vagina, the opposite appears to occur with high diversity found in patients with bacterial vaginosis⁵⁶. Given that different areas of the body harbour a unique microbiota³¹, it is possible that the

relationship of community diversity to the development of diseases may not be universal to every niche in the human body.

Studies of the development and diversity of the infant airway microbiota have not shown the same trend as that seen in the gut. Longitudinal studies of the healthy infant airway to date have focused on the nasopharyngeal microbiota. In one study, nasopharyngeal samples taken at intervals (7 days – 1 month) from birth until 6 months of age demonstrated an increase in diversity with age, particularly in the first 2 months of life³⁸. In contrast a study by the same research group of 60 healthy infants sampled at 6 -12 monthly intervals between 1.5 months and 24 months of age found no significant difference in diversity of the nasopharyngeal microbiota²³. Mika *et al.* however, found a small decrease in diversity with age in 47 infants sampled twice weekly from 5 weeks until 1 year of age³⁷. Conversely, a large study of the oropharyngeal microbiota in 134 infants with untreated episodic wheeze and 200 healthy controls from rural Ecuador found an increase in diversity in the first 2 years of life⁵⁷. Whilst the importance of community diversity remains unclear, these findings suggest that the greatest changes in the airway microbiota may occur in the first few months of life.

Early life environmental factors also appear to have an influence on the evolving microbiota and later health outcomes. These include age²³, environmental exposures, such as geographical location of the host⁵⁸, seasonality³⁷, smoking⁵⁹ and breastfeeding. *Corynebacterium* spp. and *Dolosigranulum* spp.- dominated nasopharyngeal microbiota have been found in breastfed infants in the first 2

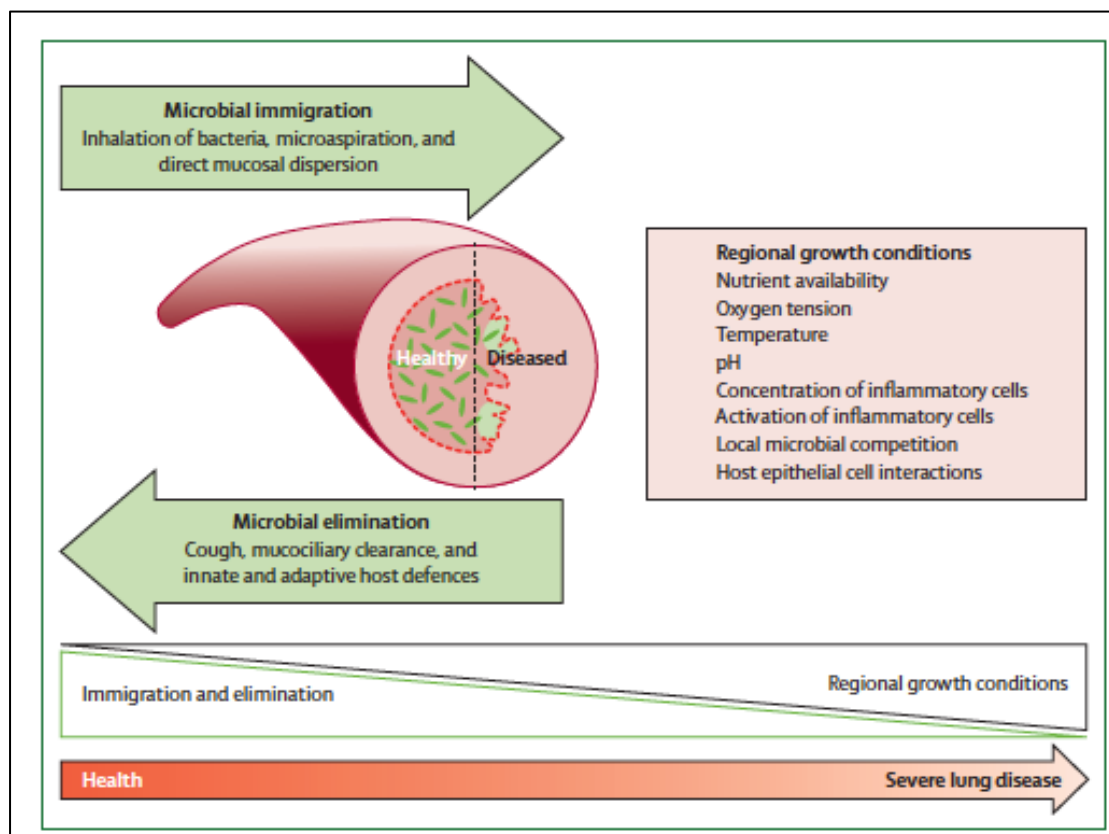
years of life and have been associated with a decreased frequency of parental reported upper respiratory tract infections²³. The role of mode of delivery on the development of the infant microbiota is controversial. By sampling multiple body sites, including the oral mucosa, nasopharynx and meconium, a study of four mother-infant pairs found that the neonatal microbiota closely resembled either the maternal vaginal or skin microbiota depending on whether the infant was delivered vaginally or by LSCS⁴⁷. Conversely, a study of eighty-one mother-infant pairs found only a small difference in the oral, nasal, skin and stool microbiota at birth and no difference at 6 weeks of age whatever the mode of delivery. However, significant expansion and diversification of each niche microbiota was seen⁶⁰. Thus there may be a window of opportunity in infancy when the foundations for the developing airway microbiota are being established to promote the development of a profile associated with health.

There are three essential factors which are thought to govern community assemblage: immigration, local proliferation and elimination^{61,62}. Immigration is an importance process for community diversification. The survival of organisms with a community is governed by a combination of environmental selective pressures (such as nutrient availability, oxygen tension, pH and temperature), host defence mechanisms and traits of the microbes themselves⁶¹, such that those organisms that are better adapted to survive in a given niche are able to outcompete their microbial counterparts and thrive. Niche environments, however, are not static and are susceptible to short and long term perturbations (by, for example, antibiotic exposure). Synergistic interactions between microbes are also important in this regard in permitting adaptations via

mutations and horizontal gene transfer (e.g. for antibiotic resistance) to allow community members to resist perturbations and continue to survive⁶¹. These processes are evident in chronic lung diseases in which changes within the niche, such as the increased production of nutrient-rich mucus, create an environment conducive to local growth of those bacteria well adapted to survive in these conditions. This facilitates local proliferation of community members and chronic infections to be established (Figure 1.4).

Figure 1.4: Illustration of the three factors thought to determine the airway microbiota: immigration, elimination and proliferation of community members.

Reprinted from The Lancet, 384 (9944), 691 – 702, Dickson R.P. *et al.*, The role of the microbiome in exacerbations of chronic lung diseases, Copyright © (2014), with permission from Elsevier⁶³.



1.1.6 The role of the microbiota on host immunity and airway disease

The relationship between colonising bacteria and allergic airways diseases has received much attention. In one study of 321 infants, bacterial colonisation of the hypopharynx with *Streptococcus pneumoniae*, *Moraxella catarrhalis*, or *Haemophilus influenzae* (*H. influenzae*) at 1 month of age was associated with an increased risk of recurrent wheeze and asthma at 5 years of age⁶⁴. In contrast, a study of 7,412 children found an inverse relationship between *Helicobacter pylori* infection and the development of asthma, allergic rhinitis, dermatitis and eczema⁶⁵. Thus early host infection patterns may affect later risk of allergic diseases.

It has been suggested that the infant microbiota plays a pivotal role in programming the evolving host immune system. Studies in germ free newborn mice have demonstrated an exaggerated, allergic airway inflammatory response following an immunological challenge, favouring a T-helper type-2 response^{66,67} with increased number and degranulation of eosinophils⁶⁶. This was significantly reduced upon re-colonisation with specific pathogen free bacteria⁶⁷. Exposing pregnant dams to antibiotics perinatally has been found to cause dysbiosis of the gut microbiota of neonatal mice with decreased circulating numbers of neutrophils and an increased risk of *Escherichia coli* and *Klebsiella pneumoniae* sepsis seen. This was reversed with transfer of a healthy gut microbiota⁶⁸. Overall, these studies suggest that the microbiota has a critical role in both the development of a regulated and protective lung immune system and avoidance of infections.

Consequently, the role of the microbiota in airway diseases has been questioned. Hilty *et al.*²⁷ found significant differences in the airway microbiota in a cross-sectional study comparing 8 healthy controls with 5 patients with COPD and 11 with asthma. Lung disease was found to be more polymicrobial than previously appreciated, with Proteobacteria strongly associated with disease. Differences in the airway microbiota have been seen between antibiotic and steroid-naïve infants with wheeze and healthy controls³⁹. In addition, the airway microbiota may influence treatment response, with one study demonstrating differences in the microbiota between corticosteroid sensitive and corticosteroid resistant asthmatics and controls⁶⁹. Thus alterations in the microbiota may contribute to disease processes through the breakdown of host-microbe symbiosis.

Overall, the airway microbiota represent a heterogeneous ecosystem in which there is a complex interplay between environmental exposures, host characteristics, defence mechanisms and interactions between different microbes along the respiratory tree. Dynamic changes in the microbiota may be associated with age, disease states and treatments. Thus, fully understanding the role of the airway microbiota in health and disease requires further longitudinal studies.

1.2 Investigating respiratory infections in children

Obtaining lower airway samples from children is often challenging. Most children cannot spontaneously expectorate sputum, particularly those under 5

years of age, and obtaining bronchoalveolar lavage fluid (BALF) is invasive and cannot be performed frequently for ethical reasons. Upper airway cultures of either throat swabs (TS) or cough swabs (CS) are therefore often used as surrogates for lower airway sampling. However, the reported sensitivity and specificity of upper airway cultures for specific pathogens is highly variable. For example, the sensitivity of TS for *Pseudomonas aeruginosa* (*P. aeruginosa*) ranges from 35.7– 71%^{70, 71}. For some organisms such as non-tuberculous mycobacteria (NTM), the use of CS is not recommended due to failure of NTM detection even in patients with proven chronic infection^{72, 73}. Whether this is due to problems inherent to the sample itself or to poor yields on culture is at present unknown.

Applying next generation sequencing to paired upper and lower respiratory samples will show whether upper airway samples can be used to reflect the lower airway microbiota. Several studies comparing oropharyngeal swabs with BALF and bronchial brushings have shown continuity of bacterial communities along the respiratory tree.^{27,32,74}. Although the studies have involved a small number of individuals with and without airway disease, they nonetheless provide evidence of the potential utility of TS; this will be tested as part of the work of this thesis.

1.3 Chronic suppurative lung disease

Chronic suppurative lung disease (CSLD) describes a group of conditions in which chronic cough productive of sputum, and chronic lower airway infection

and inflammation are the predominant features⁷⁵. The aetiologies are wide ranging, including congenital disorders, such as CF and PCD, primary and acquired immunodeficiencies, chronic infections (e.g. tuberculosis) and mechanical obstruction of the airways^{75,76}. Characteristically the disease course in CSLD is one of recurrent cycles of infection and inflammation, leading to progressive destruction of the bronchial walls, airflow obstruction and eventual bronchiectasis⁷⁷.

Although bronchiectasis is defined as “irreversible dilatation of peripheral airways”⁷⁵, there is evidence to suggest that early airway dilatation may be reversible in children^{78,79}. On serial high resolution computed tomography (HRCT) scans, 27% of children with non-CF bronchiectasis showed complete resolution and a further 36% showed improvement in airway dilatation with medical management (which included antibiotic treatment, airway clearance techniques and nutritional strategies)⁷⁸. CSLD may therefore represent a spectrum, where chronic productive cough progresses to initially reversible and later irreversible airway dilatation if untreated⁷⁷. Bacterial infections and neutrophilic infiltration (and especially free neutrophil elastase) have been implicated in the progression of bronchiectasis⁷⁹. Effective treatments for early disease states may therefore be key in halting disease progression.

1.3.1 Cystic fibrosis

CF is an autosomal recessive, multi-organ disorder affecting the lungs, pancreas, gastrointestinal and reproductive tracts among other systems. It is caused by

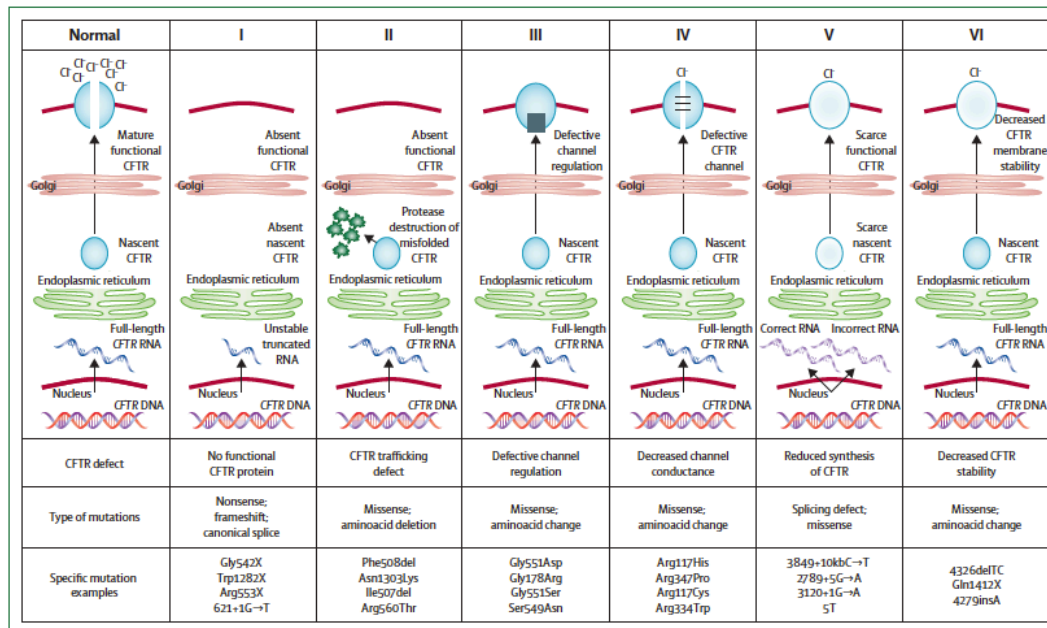
mutations in the cystic fibrosis transmembrane conductance regulator (CFTR), a multifunctional chloride channel present on many epithelial surfaces involved in the regulation of chloride, sodium and bicarbonate transport. The United Kingdom (UK) incidence of CF is ~1 in 2500⁸⁰ with a carrier frequency of ~1 in 25. It is one of the most common life-limiting genetic conditions worldwide, with a predicted life expectancy of 50 years for children born in the year 2000⁸¹. Over 90% of deaths are caused by the pulmonary manifestations of the disease⁸². In the UK, since 2007 the majority of children with CF have been diagnosed through newborn screening (NBS) at a median age of 3 weeks⁸³ and deaths in infancy are rare⁸¹.

1.3.1.1 Genetics of CF

There are over 2,000 known CFTR mutations (www.genet.sickkids.on.ca), 272 of which are confirmed to be 'disease producing' (<http://www.CFTR2.org>), with others as yet of uncertain significance. The 272 disease producing mutations can be grouped into 6 classes based on their known or predicted expression of CFTR at the apical surface of the epithelium. Each class is associated with varying degrees of disease severity, with Class I to III associated with a more severe phenotype than Class IV to VI (Figure 1.5).

Figure 1.5: The six classes of mutations in CFTR.

Reprinted from The Lancet, 388 (10059), 2519 – 2531, Elborn, J.S., Cystic fibrosis, Copyright (2014), with permission from Elsevier⁸⁴.



Class I mutations are largely nonsense mutations leading to a premature stop codon resulting in a shortened, unstable and non-functional protein with virtually no functional CFTR reaching the cell surface. These are severe mutations, an example of which is p.Gly524X⁸⁵. They affect up to 10% of the CF population worldwide.

Class II mutations result in errors in protein folding which prevent trafficking of CFTR after production. Degradation of the misfolded protein in the endoplasmic reticulum leads to failure of CFTR appearing at the apical membrane and more severe disease, although small amounts escape this process to reach the cell surface, where it shows properties of Class III and Class VI mutations described below. This group includes the most common CF gene defect, p.Phe508Del, in

which a 3 base pair deletion results in the absence of the amino acid phenylalanine at position 508 of the CFTR protein. Ninety percent of CF patients have at least one copy of pPhe508Del with homozygosity found in 50% of CF patients⁸⁶ and associated with more severe disease. In heterozygotes, disease severity is determined by the non-p.Phe508Del gene.

Class III mutations are often known as gating defects, as CFTR is folded appropriately allowing insertion into the cell membrane but fails to function correctly and rarely opens. This class includes p.Gly551Asp which affects approximately 3% of the population⁸⁵. Class III mutations have been the focus of much attention in recent years, as studies of the small molecule Ivacaftor, aimed at potentiating CFTR present at the cell surface, have shown significant improvements in lung function, weight and sweat chloride concentrations for patients with this otherwise severe mutation⁸⁷.

In Class IV mutations, there is reduced conductance of chloride through CFTR inserted at the cell surface. P.Arg117His is one the commonest mutations in this class and affects 0.5% of the worldwide population. Class V result from missense mutations which generate both abnormal and correctly spliced mRNA, leading to reduced amounts of functional CFTR. Both Class IV and V mutations are associated with milder disease and exocrine pancreatic sufficiency⁸⁵. Class VI mutations are rare with reduced amounts of functional CFTR because of a shortened half-life at the apical cell membrane⁸⁵ (Figure 1.5).

The relationship between genotype and phenotypic in individual patients with CF is complex. Variations exist in the degree of correlation between genotype and disease severity in different affected organs. For the exocrine pancreas, disease severity shows close correlation with genotype whereas there is poor correlation of genotype with lung disease with a wide spectrum of disease severity observed even in those patients homozygous for the p.Phe508Del mutation. Thus whilst genotype is important, this indicates that there are other factors which influence severity of lung disease in CF⁸⁸. Twin and sibling studies have indicated a possible role for modifier genes⁸⁹ and environmental factors, such as nutritional status⁹⁰, bacterial infections⁹¹, tobacco smoke exposure⁹² and socioeconomic status (independent of access to healthcare)⁹³, presenting potential modifiable targets for improving the health of patients with CF. Indeed, the increase in survival of CF patients as a result of greater disease surveillance and earlier more aggressive treatment, is testimony to the important role of environmental influences on disease progression.

1.3.1.2 Pathophysiology

CFTR is an apical membrane anion channel belonging to the ATP-binding cassette (ABC) transporter family. It consists of 5 domains: 2 membrane spanning domains which form the channel pore; 2 nucleotide spanning domains (NSD) which control ATP-driven channel gating, and a regulatory (R) domain unique to CFTR which governs channel activity⁹⁴. CFTR directly regulates chloride secretion through ATP hydrolysis by the NSDs⁹⁴ and phosphorylation of the R domain by protein kinase A, resulting in channel opening and chloride

efflux. In addition, CFTR is a bicarbonate transporter, although with an estimated efficiency of a quarter of that of chloride⁹⁵, and it has been hypothesised to negatively control epithelial sodium channels (ENaC) with loss of this regulation thought to lead to sodium and water hyperabsorption.

There have been several hypotheses as to how CFTR dysfunction leads to the clinical manifestations of CF. The current favoured one is the low volume hypothesis which suggests that CFTR mutations lead to airway surface liquid (ASL) dehydration. ASL is a thin layer of fluid which covers the airway epithelium and has two distinct components: an upper, high viscosity mucus layer containing high molecular weight mucins and an inner, low viscosity periciliary liquid layer (PCL). In the healthy airways, the depth of the PCL has an important role in regulating antimicrobial defences and MCC⁹⁶. The mucus layer traps inhaled particles whilst PCL aids mucus clearance by providing optimal lubrication and viscosity for effective ciliary beating. Maintenance of ASL volume is therefore essential for elimination of microbes.

Using CF cell culture models, ASL volume has been demonstrated to be reduced as a result of defective sodium and chloride transport and hyperabsorption of water, leading to increased mucus viscosity, collapse of cilia and impaired MCC⁹⁷. There is increased secretion of mucus, and once infection and inflammation ensue, high concentrations of neutrophil-derived DNA and actin drive further increases in mucus viscosity. In addition, defective cough clearance causes secretion retention in the lower airways thus creating a nidus for bacterial infection⁹⁸. In addition, as a result of abnormal bicarbonate transport, ASL pH is

reduced further creating an environment which impairs bacterial killing⁹⁹. Collectively, these changes allow for the selection of microbes in the CF airways that would normally fail to thrive in healthy airways.

Whilst the low volume hypothesis is currently favoured, evidence from CF cell and animal models has been conflicting with dog kidney cell lines demonstrating sodium hyperabsorption¹⁰⁰ which newborn pigs, ferrets and rats do not¹⁰¹. Furthermore a study of human tracheal xenografts from aborted fetuses implanted subcutaneously into immunodeficient mice found that the viscosity and hydration of foetal ASL did not differ between those with and without CF. The differences in ASL volume and hydration seen between animal models and *in vitro* cell models may in part be due to difficulties in recreating the dynamic environment of the lungs with static cell models. Using a cell model designed to mimic the phasic sheer stress of the lungs associated during normal breathing, one study demonstrated restoration of PCL volume and improved regulation of ion transport under phasic motion in CF cells. This process was exclusively ATP-mediated and susceptible to infective insults, such as viral infections, with loss of PCL height demonstrated with RSV infection¹⁰². This suggests that airway surface dehydration may be driven in part by environmental influences such as exposure to infections¹⁰³. Alternatively, CFTR dysfunction may impair microbial killing via a reduction in ASL pH. Using a pig model of CF, reduction in ASL pH led to reduced bacterial killing in the lungs, suggesting that airway pH plays a critical role in the early acquisition of airway infections in CF⁹⁹.

Aberrant innate airway immunity in CF further contributes to the propagation of airway infections. Airway neutrophilia has been demonstrated in children with CF as early as 4 months of age^{104,105} and in particular neutrophil elastase is thought to contribute to the development of CF airways disease, or at least be strongly associated with its development. The CF airways have been found to have an exaggerated pro-inflammatory response with secretion of cytokines, such as interleukin -8 (IL-8) and tumour necrosis factor- α (TNF- α), leading to increased neutrophil chemotaxis. Upon degranulation, azurophilic granules release neutrophil elastase which is thought to be central to the pathophysiology of CF. The primary role of neutrophil elastase is to degrade phagocytosed cells. Increased release of neutrophil elastase in CF airways however also leads to secondary damage both directly, through the degradation of structural proteins such as elastin and collagen, and indirectly, through inactivation of other components of the immune system, such as immunoglobulins and complement, thus interfering with opsonisation and bacterial killing. Furthermore neutrophil elastase cleaves complement, leading to release of C5a (a chemoattractant for neutrophils) creating a vicious cycle of infection and inflammation and consequently tissue destruction, bronchiectasis and eventual respiratory failure¹⁰⁶. Whilst it is thought that CF airway inflammation occurs in response to infection, there is increasing evidence that CFTR dysfunction may directly promote airway inflammation even in the absence of infection^{107,108}. Paradoxically, this excessive inflammatory response frequently fails to eliminate pathogens. The reasons for this are at present unknown.

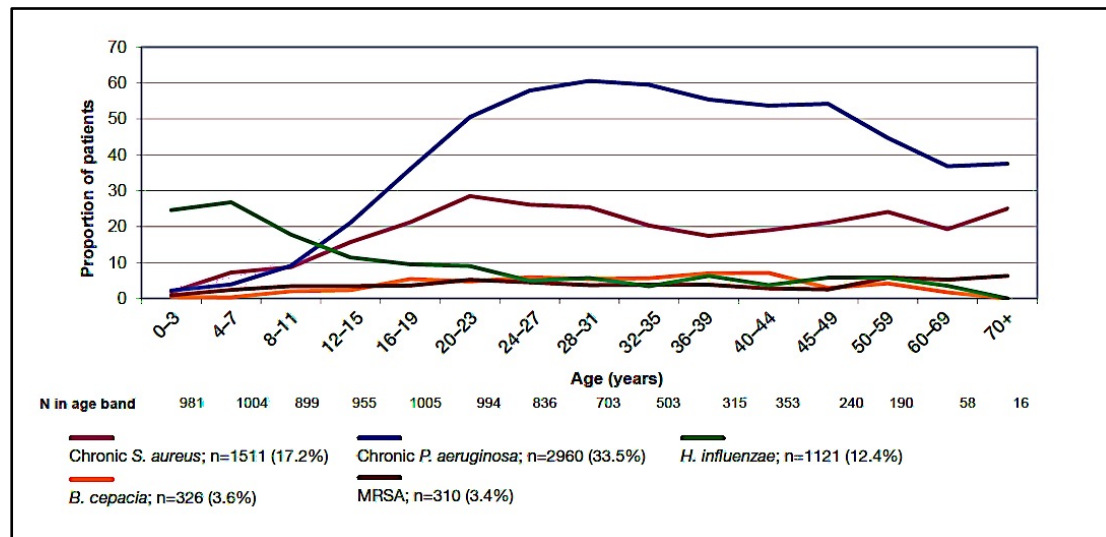
1.3.1.3 Infections and exacerbations in CF

The disease course in CF characteristically involves periods of stability punctuated by intermittent detrimental exacerbations¹⁰⁹. A quarter of patients do not recover their baseline lung function following an exacerbation¹¹⁰ with the greatest lung function decline being found in patients who have less than 6 months between exacerbations requiring hospitalization¹¹¹. More frequent admissions before 2 years of age have been associated with decreased forced expiratory volume in 1 second (FEV₁), decreased weight for age and presence of bronchiectasis on computed tomography (CT) by age 5 years¹¹². Clearly prompt recognition and treatment of exacerbations is important to prevent disease progression.

The pathophysiological bases for exacerbations are poorly understood. It was previously thought, based on culture data, that bacterial chronic infection of the lower airways in CF follows a stereotypical pattern with only very few opportunistic pathogens. *S. aureus* and *H. influenzae* are common early in life with later emergence of prognostically important Gram-negative organisms, such as *P. aeruginosa*, particularly the mucoid phenotype¹¹³, and *Burkholderia cepacia* complex (BCC)¹¹⁴(Figure 1.6). Pulmonary exacerbations were thought to reflect a new infection within the lower airway. However, in a study examining exacerbations in 211 patients, 70% of samples were culture negative for the usual CF pathogens despite suggestive clinical features which improved with antibiotic treatment¹¹⁵. Thus conventional culture may not accurately reflect changes occurring within the airways.

Figure 1.6: Illustration of the changing prevalence of “traditional” CF pathogens with age in UK.

Based on serial cross-sectional data from bacterial cultures taken from the UK Cystic Fibrosis Registry data in 2013. Reprinted with permission from UK Cystic Fibrosis Trust ¹¹⁶.



P. aeruginosa is the most significant organism in CF, affecting over 70% of patients in the first three years of life¹¹⁷. It is capable of undergoing a myriad of phenotypic changes that enable it to adapt and thrive in the CF lungs to form chronic infections that are difficult to treat. These phenotypic changes include alginate overproduction to form mucoid types which provide a protective layer from external insults, such as antibiotics: quorum sensing and biofilm formation¹¹⁸. The deleterious effects of *P. aeruginosa* on lung function are well documented, with a greater rate of decline seen with mucoid types than for non-mucoid types^{113,119,120}. For these reasons eradication of any growths of *P. aeruginosa* is always attempted, even in asymptomatic individuals¹²¹.

Whilst the prognostic significance of *P. aeruginosa* is well established, uncertainty exists over the importance of other “typical” CF pathogens. *S. aureus*, a common commensal of the nares, chronically infects 17.2% of patients with CF (Figure 1.6)¹¹⁶ and is frequently one of the first organisms isolated in children. It has several features which render it capable of causing severe infections, most notably the production the leucocytolytic toxin Panton-Valentine Leucocidin (PVL), biofilm formation and small colony variants which confer antibiotic resistance and have been associated with advanced disease¹²². In children, harmful effects of *S. aureus* infection on lung function¹²⁰ and lung inflammation, measured by increased neutrophil counts, neutrophil elastase and IL-8 in BALF¹²³ have been demonstrated. In contrast, a recent study in adults found lower rates of exacerbation and C-reactive protein (CRP) levels in patient with *S. aureus* in comparison with *P. aeruginosa*¹²⁴. Whilst this suggests that *S. aureus* may be associated with milder disease in adults, it is also possible that differences in virulence factors and antibiotic resistance patterns may account for differences in severity seen between patients with chronic *S. aureus* infections.

There has been a long-standing debate on the use of anti-staphylococcal prophylaxis in children with CF. In the UK all children are started on anti-staphylococcal antibiotics at diagnosis, usually with flucloxacillin or sometimes amoxicillin with clavulanic acid. This practice, however, is controversial. A previous study demonstrated reduced prevalence of *S. aureus* with the use of prophylactic cephalexin, a second-generation cephalosporin, but at a cost of an increased prevalence of *P. aeruginosa*¹²⁵. This however may be attributable to

the use of a cephalosporin, a class notoriously associated with selection of resistant organisms. If so, similar problems should not occur with targeted anti-staphylococcal treatment with narrow spectrum antibiotics such as flucloxacillin. There have been several attempts to resolve this debate, including a recent Cochrane review which proved inconclusive¹²⁶. The ongoing cystic fibrosis anti-staphylococcal antibiotic prophylaxis trial (CF START, www.cfstart.org.uk), in which children randomised to flucloxacillin prophylaxis are being compared with those receiving targeted antibiotics for growths of *S. aureus*, will hopefully shed further light on this dilemma.

Despite being highly prevalent in children, the prognostic significance of *H. influenzae* has received comparatively little study. *H. influenzae* is a Gram-negative bacterium which may be encapsulated, antigens for which are included in the UK vaccination schedule, and un-encapsulated, otherwise termed nontypeable¹²⁷. It is the nontypeable strain which is commonly found in CF affecting 12.4% of patients¹¹⁶. Studies examining trends in FEV₁ with acquisition of *H. influenzae* have found conflicting results. One study demonstrated up to 20% lower FEV₁ with any growth of *H. influenzae*¹²⁸ whilst another study found preservation of lung function with both acute and chronic infections¹¹⁹. Whether *H. influenzae* is harmful or protective in CF lung disease remains an unanswered question.

Several newer “emerging” pathogens have recently been identified on culture of CF specimens including *Stenotrophomonas maltophilia* and *Achromobacter xylosoxidans*. For many of these “emerging” pathogens, their significance in CF

disease progression is unclear^{127,129}, with the exception of BCC. BCC is of particular concern due to its transmissibility between patients and association with “cepacia syndrome”, which is characterized by necrotizing pneumonia and high fevers, septicaemia and rapid progression to respiratory failure and death. Certain subtypes (such as *Burkholderia cenocepacia*) carry a poorer prognosis than others (such as *Burkholderia multivorans*, which carries a similar prognosis to *P. aeruginosa*)¹¹⁴ and eradication attempts have shown moderate success at best, with one group reporting successful treatment in only 29% (4 out of 14) of patients with BCC¹³⁰. Thus infection control through patient isolation is essential to prevent the onset of this highly pathogenic group of organisms. However, accurate microbiological identification of BCC is frequently impeded by misdiagnosis on sputum culture. In one study, 11% of isolates identified as BCC on culture were found to have been misidentified on phenotypic identification and polymerase chain reaction (PCR). Conversely 36% of BCC were missed and frequently diagnosed as other non-fermenting Gram-negative organisms, such as *Stenotrophomonas* spp. and *Ralstonia* spp.¹³¹. Given the prognostic importance of BCC, obtaining an accurate microbiological diagnosis is of paramount importance.

Anaerobes (such as *Prevotella* spp. and *Veillonella* spp.) have only recently been recognised to be associated with the CF airways, largely as a result of culture-independent studies. Despite difficulties in anaerobic cultivation, anaerobes have been found to be highly abundant on anaerobic culture of CF specimens^{132,133}. Whether anaerobes are pathogenic or protective is unclear, with evidence of reduced inflammatory markers (CRP) and improved lung

clearance index (LCI) with high anaerobe colony counts¹³⁴, whilst others have found increased cytotoxicity and neutrophilia with *Prevotella intermedia*. Given their high abundance, greater understanding of the role of anaerobes is needed to help guide antibiotic strategies. Current conventional choices, other than meropenem, at present provide little anaerobic cover.

1.3.1.4 The role of the microbiota in CF

Molecular microbiological techniques have demonstrated more diverse microbial communities in the CF airways than previously recognised, with estimates of up to 43 different bacterial phyla and approximately 2,000 different taxa¹³⁵. Notably a large number of obligate and facultative anaerobes have been found to be present¹³⁶⁻¹³⁸. Many of these taxa have not previously been associated with CF and therefore presently little is known about their role in disease progression.

It is plausible that fluctuations in disease course typical of CF may reflect perturbations in the microbiota occurring between periods of stability and exacerbation. Several longitudinal studies of small numbers of adults with CF have highlighted the importance of the airway microbiota in disease prognosis. At a community level, an inverse relationship has been seen between diversity of the sputum microbiota, lung function and disease severity^{135,139-143}. It is thought that in a diverse and balanced community, individual species are less able to dominate and cause disease⁵². Loss of diversity has been associated with an increased relative abundance of more pathological and phylogenetically related

organisms belonging to the family *Pseudomonadaceae*¹³⁵, with conservation of total bacterial density¹³⁹. With advanced disease, the airway microbiota in CF has been shown to be resistant (defined as showing little change) and resilient (returning to its original state) to environmental changes. Little change has been seen with exacerbations or external influences, such as antibiotic administration^{139,142,144}, making treatment strategies challenging. Whether reduced diversity is a marker of end-stage disease or an effect of the frequent courses of antibiotics that many of these patients have received during exacerbations, has yet to be determined.

The results from these adult studies should be interpreted cautiously. There are several methodological differences making direct comparisons between studies difficult due to the different sample sizes, sampling frequency and periods of patient follow-up used. The larger studies are cross-sectional^{140,141,145}, and those which are longitudinal are often only in small subsets of patients^{139,142}. For example, the study by Zhao *et al.*¹³⁹, which has the longest period of follow-up of up to 9 years, involved only 6 patients, 3 “mild” and 3 “progressive”. Extrapolating observations from such small samples to the general CF population is consequently difficult. Making cross-sectional comparisons between patients does not account for potential within patient variability, which has shown to be significant in microbial systems throughout the body^{21,31,50,138,146,147}. This problem can be reduced with greater use of longitudinal studies in which the patient acts as their own control. Nonetheless, despite these differences, it is striking that all of these studies have drawn similar conclusions, namely that advanced CF lung disease is associated with

reduced species richness and microbial community diversity. What drives these changes is unknown but there is a clear need for longitudinal studies with frequent sampling in large samples of patients to better understand this phenomenon.

Study of the microbiota in children, prior to confounding by frequent antibiotic use and the onset of advanced disease, may therefore be more informative. It has been suggested that the airway microbiome in children with CF will show a similar trajectory towards successive diversification to that seen in the healthy infant gut. Parallels can be drawn between with the microbial progression seen in CF on culture data. A similar trend towards diversification and successive replacement of early infecting species, such as *S. aureus*, by species better adapted to the CF lung environment, such as *P. aeruginosa*, may occur. These result from selective pressures both endogenous to the CF airways themselves (such as pH) and exogenous (such as repeated antibiotic administration) creating an environment conducive to the subsequent emergence of antibiotic resistant organisms and new Gram-negative organisms with age.

Studies of the airway microbiota in children with CF however have shown conflicting results regarding the role of diversity in airways disease. One of the few longitudinal studies in children with CF examined changes in the respiratory microbiome in 7 infants diagnosed with CF on NBS. Pyrosequencing of DNA extracted from oropharyngeal swabs taken every 3 months up to 21 months of age revealed a pronounced increase in diversity over time¹⁴⁸. Similarly, a cross-sectional study of 269 patients with CF spanning a 60-year age range, from 4

years of age upwards, found that community diversity was greatest until 10 years of age but decreased thereafter with plateauing occurring at 25 years. Differences in diversity were mirrored by changes in lung function across all age groups¹⁴⁵. What may cause this apparent trend reversal in adolescence is unclear.

In contrast, longitudinal studies of the nasopharyngeal microbiota in infants have suggested that community composition is more important than overall diversity. No significant differences in overall diversity were seen between infants with CF and age-matched controls in the first year of life although significant changes were seen in community composition^{149,150}. In CF, there is an initial increase in *Staphylococcal* species^{149,150}, but this is gradually replaced by *Moraxella* spp. and later *Streptococcus* spp. by 3 months of age.¹⁴⁹ In controls, *Moraxella* spp., *Dolosigranulum* spp., *Haemophilus* spp. and *Prevotella* spp. have been found to have a greater relative abundance with differences between infants with CF and healthy controls observed from as early as 2 months of age¹⁵⁰. To my knowledge, all of the studies of the airway microbiota in children with CF to date have been conducted in infants who have not received anti-staphylococcal prophylaxis, which as highlighted earlier (Section 1.3.1.3) is commenced in virtually all UK infants at diagnosis. Therefore, a study of infants on anti-staphylococcal prophylaxis may reveal lower rates of *Staphylococcus* spp. in CF to those previously reported and help to elucidate key compositional changes in the airway microbiota between CF and controls as well as changes with the CF clinical course.

Whilst the nasopharyngeal microbiota has received greater study, oropharyngeal sampling is thought to be a more reliable surrogate of the lower airway microbiota when direct sampling of the lower airways is not feasible¹⁵¹. In the largest reported study of the oropharyngeal microbiota to date, 54 oropharyngeal samples were collected from 13 infants for up to 3 years¹⁵². Whilst no overall change in diversity was found with growth of *P. aeruginosa*, an increase in the relative abundance of *Salmonella* spp. was seen when *P. aeruginosa* was grown on bacterial culture. No significant change was seen with exacerbations possibly due to the small number of exacerbators (N = 5) and infrequency of sampling. Breastfeeding, however, was found to be protective, with a longer time to first pulmonary exacerbation or growth of *P. aeruginosa* in those infants who had received breastmilk in infancy. Hence, modifiable clinical variables may have an important impact on the airway microbiota and subsequent lung health. More longitudinal studies are needed coupling detailed clinical information with frequent sampling to better understand the progression of the childhood airway microbiome to identify strategies for maintaining lung health into adulthood.

1.3.1.5 Disturbances in the microbiota

In natural ecosystems, perturbations in stable communities are observed following events that alter a community either directly or indirectly by altering the host environment¹⁵³. These can be termed “disturbances”. Such disturbances may alter the balance of community members, with disappearance of some members and proliferation of others leading to successive waves of

change until a new equilibrium is established¹⁵⁴. It is possible that the changes seen in progressive CF disease reflect such adaptations in response to disturbances. Once an alternative community state is established, attempts to return to previous community diversity may be challenging⁵². Therefore interventions are required prior to major community shifts in order to maintain diversity although key to this is the identification of potential disturbances (such as changes in nutrient availability, microbial invasions and antibiotic administration).

Microbial “invasion”, defined as “the successful establishment of a non-native organism in a community”¹⁵³, is one potential cause of disturbances. Successful invasion is governed by current community assemblage and stability¹⁵³. Invasions may be a cause of at least some pulmonary exacerbations in CF¹⁵⁵. It is difficult to know which species are “non-native” in the CF airways given that many of those now being identified have previously not been associated with CF. Several studies however have demonstrated no significant changes in community diversity or total bacterial density at a species level either preceding or at the onset of an exacerbation^{139,142,143,156}, with only a few studies reporting changes in the relative abundance of specific pathogens^{136,157-160}. Thus the interactions of species rather than changes in specific individual species may be more significant in disease. It is worth noting that previous studies have not been able to discriminate changes occurring at a strain level, which is particularly important for some species, such as *P. aeruginosa*.

The poor prognosis associated with *P. aeruginosa* may relate to the interactions of *P. aeruginosa* with other community members. Klepac-Ceraj *et al.*⁷⁴ found that the airway communities of CF patients with *P. aeruginosa* present were less diverse than those of patients in which *P. aeruginosa* was absent. Given that *P. aeruginosa* may be present in BALF in infants as young as 5 months of age¹⁶¹, it seems unlikely that it is the mere presence or absence of *P. aeruginosa* which governs community diversity. The relative abundance of *P. aeruginosa*¹⁴² and its impact on other species in the microbiota through production of molecules inhibiting the growth of competing organisms¹⁴¹ have been implicated in governing community diversity and respiratory function decline. Community dynamics and overall community composition therefore appear to be of critical importance.

One interesting relationship is that of *P. aeruginosa* with anaerobes (such as *Veillonella* spp. and *Prevotella* spp.), which have been found in the CF airways in both children and adults^{14,137,138,162} either in equal or greater proportional abundance than *P. aeruginosa*¹³⁷. Anaerobes have previously been dismissed due to belief that they are purely oral commensal contaminants. However, a study using T-RFLP compared the microbiota in expectorated sputum and mouthwash samples from 19 adult CF patients. A difference of up to 74% was found in bands detected between the two sample types, with absence of the highest volume band from mouthwash samples compared with sputum. This suggests that CF sputum is not contaminated to a significant extent by oral microbiota and the presence of anaerobes represents true infection¹⁴. Furthermore, since bacterial RNA has been shown to be present within sputum samples, it appears that the

majority of these anaerobes are metabolically active and may be contributing to disease¹³⁸.

The contribution of anaerobes to CF disease progression appears to be genus specific (Table 1.4). *Prevotella* spp. and *Veillonella* spp. are the most frequent newly identified obligate anaerobes in the CF airways accounting for up to 83% and 54% of anaerobes respectively¹⁶³. They have been described as forming part of a “core” CF microbiota due to their presence in more than 75% of all CF samples¹⁴⁴. Both are Gram-negative anaerobes found predominantly in the oral cavity associated with periodontic disease and also colonise the human gut¹⁶⁴. They are also frequently found in healthy airways and commonly regarded as opportunistic bacteria, thought only to cause disease in immunocompromised states¹⁶³. In the CF airways, however, different genera appear to exhibit different effects on disease progression. *Veillonella* spp. has been found to have an inverse relationship with lung function¹³⁴ and the opposite trend has been seen with *Prevotella* spp.^{134,165}. The balance of different anaerobic genera in the microbiota may therefore be important in determining differences between stable and progressive disease.

Changes in the abundance of *Prevotella* spp. and *Veillonella* spp. have been observed both before an exacerbation and with antibiotic treatment¹⁴⁴, with *Prevotella* spp. in particular found in high concentrations at the time of exacerbation in one study¹⁶². However, Field *et al.*¹⁶² also found that levels remained high in samples taken during periods of stability, implying that an increase in *Prevotella* spp. alone cannot fully explain respiratory exacerbations.

Using a rat chronic infection lung model, oropharyngeal communities were shown to significantly increase the virulence of *P. aeruginosa* by modulation of *P. aeruginosa* gene expression through signalling by the quorum sensing molecules¹⁶⁶. In addition it has been argued that anaerobes may be responsible for early CF infections and create an environment favouring later infection by *P. aeruginosa*¹³⁷. Many anaerobes produce short chain fatty acids such as acetic acid. Not only have these been found to contribute to airway inflammation, by inducing release of the neutrophil chemo-attractant IL-8^{167,168}, but they also promote *P. aeruginosa* growth, when present in low concentrations¹⁶⁸. Current antimicrobial treatment strategies, which predominantly consist of anti-pseudomonal agents, may be more effective if earlier infections are targeted. There is some evidence to suggest that anti-pseudomonal treatments for exacerbations may have a more profound effect on reduction of non-pseudomonal species, such as *Prevotella* spp., within a community,^{162,169} and paradoxically cause a mean increase in the abundance of *P. aeruginosa*¹⁶⁹. There is a need to better understand the interactions between different organisms to ensure that treatment strategies favour community assemblages associated with health.

Another emerging group of pathogens within the CF microbiome is the *Streptococcus milleri* group (SMG), which includes the species *Streptococcus constellatus*, *Streptococcus intermedius* and *Streptococcus anginosus*. SMG are largely facultative anaerobes and common commensals in various body sites such as the gut and oropharynx¹⁵⁸. Streptococci are not conventional organisms described in the CF airways but have been recovered in increased numbers

during CF pulmonary exacerbations^{158,162} sometimes in combination with *Prevotella* spp.¹⁶². In one report, SMG were the dominant pathogens in almost 40% of pulmonary exacerbations¹⁵⁹. A correlation between SMG and co-infection with *P. aeruginosa* has also been found suggesting that a synergistic relationship between these pathogens can occur. *Streptococcus* spp. itself has also been associated with clinical stability and increased diversity¹⁵⁷, representing a core genus of both the respiratory and gut microbiome in stable infants¹⁴⁸. It has been suggested therefore that modest levels of *Streptococcus* spp. may be associated with increased diversity, whereas high levels cause the converse¹⁵⁷. It may however be that specific species are associated with health and others with disease. Further evaluation of this organism may provide valuable insights into diversity changes in the CF microbiome.

Table 1.4: Description of anaerobes that have previously been identified in CF patients on culture independent studies.

GI = gastrointestinal, CRP = C-reactive protein, FEV₁ = forced expiratory volume in 1 second.

Genus	Description	Habitat	Disease associations	Role in CF
<i>Prevotella</i> spp.	Obligate anaerobe. Gram negative. Rod shaped.	Oral cavity, GI tract and healthy airways.	Common in periodontic disease ¹⁶⁴ .	Positive correlation with lung function ^{134,165} . Relationship to exacerbations unclear: increased relative abundance in one study ¹⁶² and decreased in another ¹⁴⁴ .
<i>Veillonella</i> spp.	Obligate anaerobe. Gram negative. Cocci.	GI tract, oral cavity and healthy airways.	Rarely disease causing. Case reports of meningitis ¹⁷⁰ and periodontic disease ¹⁶⁴ .	Inverse relationship with lung function ¹³⁴ and increased in relative abundance at exacerbation ¹⁴⁴ .
<i>Gemella</i> spp.	Facultative anaerobe. Gram positive Diplococci.	Mucous membranes, particularly oral and upper GI tract.	Case reports of infective endocarditis ¹⁷¹ .	Increased relative abundance at exacerbation in one study ¹⁶⁰ .
<i>Granulicatella</i> spp.	Facultative anaerobe. Gram positive. Cocci, coccobacilli or rods.	Upper respiratory, GI and urogenital tracts.	Infective endocarditis, bacteraemia (including early onset neonatal sepsis) and septic arthritis ¹⁷² .	Associated with lower CRP, neutrophil elastase and higher FEV ₁ ¹⁷³ . Low relative abundance overall but > 30 times more prevalent in outpatients than inpatients ¹⁵⁷ .
<i>Streptococcus milleri</i> group (SMG).	Facultative anaerobes. Gram positive. Cocci	Oral cavity and GI tract.	Abscesses and sepsis ¹⁷⁴ .	Increased relative abundance at exacerbation in some studies ^{158,159,162} . One study reporting co-infection with <i>P. aeruginosa</i> ¹⁵⁷ .

1.3.1.6 The role of antibiotics

Although bacterial infection can be a cause of disturbance, it can also occur as a consequence thereof. Antibiotic use has been implicated as the primary driver of reduced diversity¹³⁹. Indeed those studies identifying an association between *P. aeruginosa* and community diversity have been confounded by concurrent antibiotic use⁷⁴. Changes in community richness and phylogenetic diversity have been observed both with short-term intravenous (IV) antibiotics^{133,142,143} and long-term oral and inhaled antibiotics⁷⁴. However, on cessation of antibiotic treatment communities have shown resilience¹⁴², returning to their pre-disturbance state. It has been suggested that repeated, frequent courses of antibiotics may cause sustained reductions in diversity¹³⁹. Those patients receiving frequent antibiotic courses however usually have more advanced disease thus confounding the association between antibiotic use and disease progression. At the Royal Brompton Hospital (RBH), some children with progressive disease are electively admitted for IV antibiotics every 3 months, irrespective of exacerbation status, in an attempt to slow disease progression. To distinguish between the effect of antibiotics and exacerbations on the CF microbiota, a comparison will be made in this thesis, between children who receive elective IV antibiotics with those treated during an exacerbation (usually with ceftazidime and tobramycin). If antibiotic use is associated with reductions in diversity, this may challenge current treatment strategies to a more judicious and better targeted use of antibiotics.

1.3.1.7 Genotype

Several studies have reported that both respiratory and gut microbiome composition are associated with homozygosity for the Phe508Del mutation^{74,135,175}. In the gut, homozygous Phe508Del patients were found to have an increased relative abundance of pathogenic bacteria such as particular strains of *Escherichia coli*, and a reduction in protective bacteria, such as *Bifidobacterium* spp. in comparison to heterozygous Phe508Del and non- Phe508Del patients¹⁷⁵. Furthermore, a study of the oropharyngeal microbiota in children found that community profiles were more similar in those patients who were either homozygous Phe508Del or heterozygous Phe508Del than non- Phe508Del patients⁷⁴. It is unclear whether this is related to specific genotypes, class mutations, or other confounders such as antibiotic use.

1.3.1.8 The lung-gut axis

There appears to be a relationship between the respiratory microbiome and that of the gut. In one study of 7 infants, comparison of stool and throat swab samples revealed similarities in community composition. Both were dominated by *Streptococcus* spp. and *Veillonella* spp. and interestingly the presence of some genera in the respiratory microbiome, such as *Enterococcus* spp., was preceded by their presence in the gut. Given that 25-50% of children with CF also have gastro-oesophageal reflux^{176,177}, microaspiration of gut microbes could be a potential route of respiratory colonisation. Moreover modification of the gut microbiome by oral probiotics has been shown to lead to significant reductions in pulmonary exacerbations and hospital admissions^{178 179}. This interesting

relationship between the gut and respiratory microbiome requires further evaluation.

1.3.1.9 Manipulating the microbiota

If there are changes in the microbiota between children with CF and controls, it may be possible to therapeutically manipulate the microbiota to maintain lung health. Indeed changes to the airway microbiota have been seen with treatments, such as Ivacaftor, which has been shown to lead to an increase in *Porphyrromonas* spp. and a decrease in *Streptococcus* spp¹⁶⁵. Evidence from other niches in the body suggests that transfer of a healthy microbiota to diseased individuals may restore the microbiota to a disease-free state. For example, faecal transplantation has shown remarkable effects in the treatment of refractory *Clostridium difficile* diarrhoea¹⁸⁰. Recently inoculation of the skin of newborn infants born by LSCS with their mother's vaginal microbiota has been shown to partially restore the microbiota to resemble that of infants delivered vaginally¹⁸¹. Such therapeutic strategies require further evaluation to determine the risks and benefits as well as investigation of whether similar strategies could be applied to the airway.

1.3.2 Primary Ciliary Dyskinesia

PCD is another inherited CSLD in which impaired mucociliary clearance results in recurrent upper and lower airway infections, with symptoms beginning soon after birth and leading to bronchiectasis. It is caused by congenital dysfunction of the motile cilia, microscopic hair-like structures that project from epithelial

cells in multiple areas of the body including the lungs, paranasal sinuses, middle ear and fallopian tubes. Cilia are also found on the embryonic node where they govern left-right asymmetry; failure of this leads to heterotaxic syndromes, present in around fifty percent of patients¹⁸². The classic triad of bronchiectasis, sinusitis and situs inversus is known as Kartagener's syndrome.

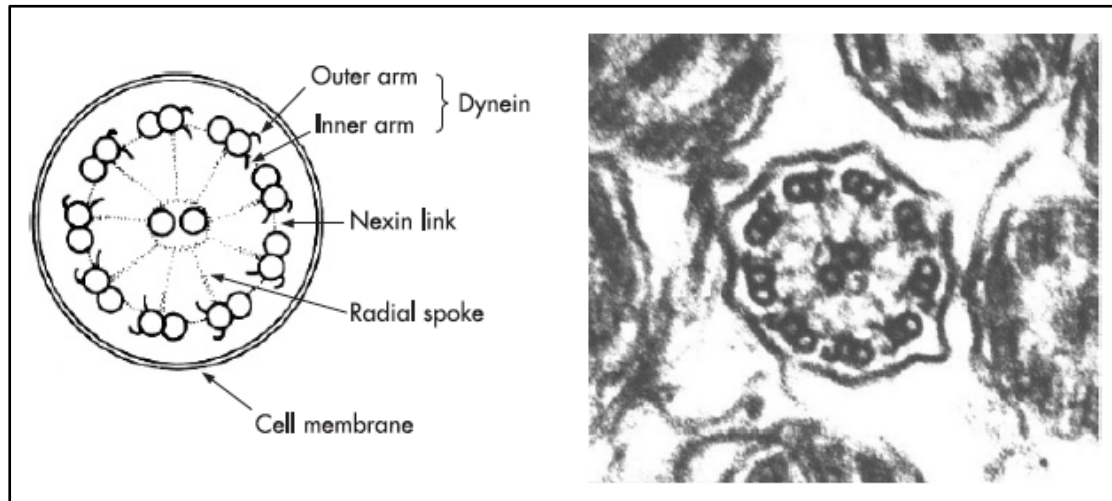
Many of the symptoms associated with PCD, such as chronic rhinitis and recurrent otitis media, are common to many minor childhood illnesses. Consequently the diagnosis is frequently late or missed. A European survey found that the median age of diagnosis is 5.3 years and one study reported that 70% of patients visited a physician more than fifty times before a diagnosis was made¹⁸³. Due to difficulties in confirming the diagnosis, the true burden of disease remains unknown with wide variation in prevalence reported from 1 in 4,000 to 1 in 40,000¹⁸².

1.3.2.1 Genetics and pathophysiology of PCD

PCD is due to mutations in genes encoding any one of the proteins forming the ciliary ultrastructure or its assembly. Motile cilia are composed of an outer ring of nine microtubule doublets joined by nexin links which connect to a central pair of single microtubules by radial spokes in a "9 + 2" arrangement. Inner and outer dynein arms on the microtubular doublets act as the powerhouse of the cilia and are responsible for coordinated ciliary beating (Figure 1.7)¹⁸².

Figure 1.7: Cross-sectional diagram of the ultrastructure of motile cilia.

Reprinted with permission of the BMJ Publishing Group. Copyright © 2007.
Cite: Bush A, *et al.* Primary ciliary dyskinesia: current state of the art. *Archives of Disease in Childhood* 2007; 92 (12): 1136 – 40¹⁸⁴.



Dynein arm defects are the most common ciliary ultrastructural defects causing PCD with lack of outer dynein arms alone or with absence of inner dynein arms accounting for 43% and 24% of cases respectively. Microtubular and central pair defects are less common¹⁸⁵. Patients with PCD have reduced (or uncoordinated) ciliary beat frequency and many patients also have identifiable ultrastructural defects on electron microscopy. Nevertheless approximately 26% of patients are reported to have normal ciliary ultrastructure¹⁸⁶, at least on conventional transmission electron microscopy, thus in the absence of identified ciliary abnormalities the diagnosis relies on combination of a typical phenotype together with supportive features such as low nasal nitric oxide levels, immunofluorescence of ciliary proteins or genetic testing^{187,188}.

PCD is a genetically heterogeneous disorder with predominantly autosomal recessive inheritance; approximately fifty gene defects have been identified to

date. The majority of the gene defects are nonsense, frameshift or splicing mutations¹⁸⁹. At present up to 65% of patients with PCD can be identified on genetic testing¹⁹⁰. Individual gene mutations correlate well with their respective ciliary ultrastructural defects. The most common mutations are: DNAH5 and DNAI1, which are responsible for 53%¹⁹¹ and 10%¹⁹² of outer dynein arm defects respectively and 30% of all PCD cases overall¹⁸⁹, and CCDC39 and CCDC40 mutations, which account for 69% of combined microtubular and inner dynein arm defects¹⁹³. The correlation between genotype and protein defect with clinical phenotype however is less well characterised, such that the clinical significance of specific defects is unclear.

Similar to CF, PCD is associated with neutrophilic airway inflammation. The inflammation has been found to be more pronounced than that seen in CF; higher IL-8¹⁹⁴ and total neutrophil counts, despite a lower bacterial load have been observed in PCD ^{194, 195}. Despite sharing a similar pathophysiology of impaired mucociliary clearance and airway chronic infection and neutrophilia, PCD rarely shows such a progressive course seen in CF with a reported drop in FEV₁ of 0.8% per year in PCD in comparison with 3.6% in CF¹⁹⁶. The reasons for this are unclear but important to understand to help identify novel treatment strategies for CF.

Although PCD has a more favourable prognosis than CF, PCD is not a mild disease. A third of children at diagnosis have been found to have an FEV₁ less than 80% predicted¹⁹⁷. With initiation of treatment, the majority of patients showed either stable or improved lung function. Marthin *et al.* however found

that despite initiation of treatment 34% of patients showed a progressive decline with a loss of 10% in FEV₁¹⁹⁷. Lobectomies have also been reported in approximately 8% of patients¹⁹⁶.

1.3.2.2 Infections and the PCD microbiota

Similar to CF, the disease course in PCD fluctuates between stability and infective exacerbations. Although research has been less intensive, similar deleterious effects of exacerbations have been found in PCD to that in CF with one study demonstrating that approximately a quarter of children with PCD fail to recover baseline lung function following an exacerbation¹⁹⁸.

It is possible that the differences seen in CF and PCD are related to differences in the respiratory microbiome. Worlitzsch *et al.* observed that hypoxic gradients exist in mucus from CF but not from PCD¹⁹⁹. If differences occur within the airway, this could favour particular bacteria and may account for differences in disease outcome. Indeed the lower airway microbiology of patients with PCD has both similarities and important differences to those with CF²⁰⁰. The most common organisms identified on microbial culture from patients with PCD are *H. influenzae* and *S. aureus*^{195,196,201}, organisms also prevalent in children with CF. However unlike CF, *Streptococcus pneumoniae*²⁰¹ is common in PCD. Whilst *P. aeruginosa* and NTM become more prevalent with age, their prevalence is lower than that seen in CF at 27% and 18% respectively¹⁹⁶.

There has been little reported about the PCD microbiota to date with only one cross-sectional study of 24 children and adults. This found that the PCD airway microbiota is polymicrobial, with 24 different genera observed per sample, including *Haemophilus* spp., *Streptococcus* spp., *Pseudomonas* spp., *Prevotella* spp. and *Porphyromonas* spp. No correlation was found between community diversity and clinical measures²⁰². There have been no reported longitudinal studies of the PCD microbiota. The Paediatric CF service at RBH sees over 320 children per year. The PCD service is one of only 3 diagnostic PCD centres in UK. As such, the paediatric population at RBH provide a unique opportunity to study a large cohort of children with CF and PCD longitudinally.

1.4 Hypotheses, aims and objectives

Hypotheses

- Upper airway samples would reflect the diversity and community composition of the lower airway microbiota.
- Changes in diversity and community membership of the airway microbiota will be seen with age and changing clinical status in children with the CSLDs, CF (including NBS) and PCD.
- The differences in prognosis between CF and PCD are associated with differences in the airway microbiota over time and in response to treatments, such as antibiotics.

Aims

To perform a two-part study of the airway microbiota in childhood:

- A prospective, cross-sectional study in children to assess whether upper airway samples (throat swabs (TS) and/or cough swabs (CS) could act as surrogates for samples obtained by fiberoptic bronchoscopy (FOB; BALF or bronchial brushings) in characterizing the lower airway microbiota for use in longitudinal sampling.
- A longitudinal study relating changes in the airway microbiota (such as diversity and community composition) as measured by upper airway samples with clinical features in:
 - Infants with CF identified on newborn screening
 - Children with CF compared with PCD, including pre- and post-antibiotic treatment in both groups.

Objectives

- To obtain throat swabs and cough swabs from children undergoing a clinically indicated bronchoscopy and compare them with lower airway samples (BALF and bronchial brushings) using molecular microbiology.
- To collect detailed longitudinal clinical information, including bacterial infection, antibiotic use and genotype, with changing clinical status to relate to changes in the microbiome.
- To recruit a prospective cohort of infants with CF diagnosed on newborn screening (NBS) and compare changes in molecular microbiology on throat swabs collected at three-month intervals for up to 2 years with age, clinical status and culture-based microbiology.
- To recruit prospective cohorts of children with CF and PCD and obtain throat swabs at three-month intervals for up to 2 years at times of clinical stability for molecular microbiology and to assess changes in the microbiota within CF and PCD with age and clinical status, and differences in the microbiota between CF and PCD.
- To obtain throat swabs for molecular microbiology at the start of antibiotic treatment for infective exacerbations of CF and PCD and at weekly intervals throughout the treatment periods, as well as at post-treatment clinical follow-up and compare these results to throat swabs collected at similar time points from children admitted for planned, routine intravenous antibiotics.

Chapter 2: Materials and methods

2.1 List of reagents, kits and equipment used

Reagents, Kits and Equipment	Manufacturer
For sample collection	
Sterilin cotton-tipped sterile swabs	Thermo Fisher Scientific, Waltham, USA
Sterilin polypropylene 30 ml universal container	Thermo Fisher Scientific, Waltham, USA
Cytology specimen brushes:	Olympus Keymed, UK
<ul style="list-style-type: none"> • 1.2 mm unsheathed • 2 mm sheathed 	
For DNA extraction	
FastDNA Spin Kit for Soil	MP Biomedicals, Solon, USA
Clear polypropylene centrifuge tubes (15 ml)	Corning Life Sciences, Tewksbury, USA
Lysing Matrix E tubes	MP Biomedicals, Solon, USA
Nanodrop 100 spectrophotometer	Thermo Fisher Scientific, Waltham, USA
Nunc CryoTubes (1.8 ml)	Thermo Fisher Scientific, Waltham, USA
Precellys 24 bead beater	Bertin Technologies, Montigny le Bretonneux, France
Heraeus Pico 21 microcentrifuge	Bertin Technologies, Montigny le Bretonneux, France
For qualitative PCR	
Q5 High Fidelity 2 x Mastermix <i>Taq</i> polymerase	New England BioLabs, Ipswich, USA
PCR grade water	MO BIO Laboratories, Carlsbad, USA
Primers	MO BIO Laboratories, Carlsbad, USA
MicroAmp Optical Adhesive Seal	Applied Biosystems/ Thermo Fisher Scientific, Waltham, USA

2.1: List of reagents, kits and equipment used continued

For gel electrophoresis	
Biotium Gel Red Nucleic Acid Stain	Bioline, London, UK
Molecular grade agarose 1% gel strength	Bioline, London, UK
100 bp DNA ladder	New England BioLabs, Ipswich, USA
GelRed dye	Biotium, Hayward, USA
For amplicon purification	
Round bottom 96 well plates	Corning Life Sciences, Tewksbury, USA
Plate seals for round bottom plates	Corning Life Sciences, Tewksbury, USA
Agencourt AMPure XP beads	Beckman Coulter, Brea, USA
Low EDTA TE buffer	Invitrogen/Thermo Fisher Scientific, Waltham, USA
Ambion 96 well magnetic stand	Invitrogen Life Sciences/ Thermo Fisher Scientific, Waltham, USA
For sequencing	
MiSeq reagent kit Version 2 (V2) containing:	Illumina, San Diego, CA, USA
<ul style="list-style-type: none"> • MiSeq reaction cartridge • Flow cell • Hybridisation buffer (HT1) • Incorporation buffer 	
Sodium hydroxide (NaOH) 10 M	Sigma-Aldrich, St. Louis, USA
PhiX Control v3	Illumina, San Diego, CA, USA
MiSeq benchtop sequencer	Illumina, San Diego, CA, USA
For quantitative PCR (qPCR)	
SYBR Green qPCR kit	Thermo Fisher Scientific, Waltham, USA
Viia 7 Real-Time PCR System	Applied Biosystems/ ThermoFisher Scientific, Waltham, USA

2.2 Study population

Registered patients, under the care of the Paediatric Respiratory Department at RBH, were recruited for a 2 part study: a cross sectional comparison of the upper and lower airway microbiota, and the CSLD Longitudinal Investigation of the MicroBiota (CLIMB study). Details of the study population recruited for the cross-sectional comparison of the upper and lower airway microbiota is provided in the corresponding results chapters (Chapter 3, Section 3.2.1).

The CLIMB study is a longitudinal study investigating changes in the microbiota in children with CF and PCD during periods of clinical stability, exacerbations (before, during and following an exacerbation at follow-up) and, where applicable, at clinically indicated bronchoscopy (Table 2.1).

Table 2.1: Follow-up protocol for the three study groups in CLIMB study.

Details include age at which recruited, frequency of follow-up for collection of non-invasive samples (Throat swabs [TS] and spontaneously expectorated sputum) and bronchoscopy samples, where applicable. NBS - newborn screened infants; PCD - Primary Ciliary Dyskinesia; CF - cystic fibrosis; IV - intravenous.

Group	Age	Frequency of sampling (TS or sputum)	Bronchoscopy samples
NBS	< 6 m	1- 3 m	3-5 m age
Neutrophilic airways disease:			
PCD	6 m – 16 yrs	3-6 m	Opportunistically
CF		2-3 m	
CF IV antibiotics			
Elective Exacerbation	6 m - 16 yrs	3 m	

The 3 main groups in this study were:

- Analysis of early changes in the microbiota in infants diagnosed on NBS.
- Comparison of neutrophilic airways disease (PCD and CF) in older children.
- Comparison of children who receive elective, 3 monthly admissions for IV antibiotics and an age-matched group who receive emergency IV antibiotics during exacerbations only.

Children were recruited and followed-up opportunistically during routine clinical appointments between December 2012 and March 2014 for up to 24 months duration. The inclusion criteria were: children who were registered patients at RBH; for CF, a diagnosis confirmed on the basis of NBS, or a combinations of typical clinical history, genotype and sweat chloride >60mmol/L in accord with standard guidelines²⁰³; for PCD, a doctor confirmed diagnosis based on a clinical phenotype, low nasal nitric oxide levels (nNO), electron microscopy of nasal ciliary brushings demonstrating abnormal ciliary beat frequency or specific abnormalities associated with PCD or 2 known PCD-causing gene mutations in *trans*, also in accord with current guidelines^{187,204} ; routine attendance at RBH for clinical appointments at least 3 monthly for CF and 6 monthly for PCD.

The exclusion criteria were: use of CFTR modulating medications, such as Ivacaftor (which became clinically available during this study), or inclusion in the Phase 2b gene therapy trial²⁰⁵. All patients newly diagnosed with CF are admitted to RBH for a 2-day educational visit to meet the multi-disciplinary

team. Patients were not recruited during this visit as it was deemed to be insensitive, but were instead recruited at their subsequent scheduled hospital visit.

Due to a lack of a universally accepted definition of pulmonary exacerbation, for this study it was defined as a change in symptoms necessitating an admission to hospital for IV antibiotics as diagnosed by a paediatrician independent of the study. Initiation of oral antibiotics was not included in the definition of pulmonary exacerbations in this study as the threshold for starting oral antibiotics can be highly variable, including a wide variety of indications such as viral upper respiratory tract infections and minor chest symptoms to more severe symptomatic and lung function decline.

Due to a lack of pilot data, a power calculation to determine the sample size needed to identify changes in the microbiota was not possible. As many children as possible were therefore recruited sequentially from all those who were receiving care predominantly at RBH.

2.3 Clinical information collation

Detailed clinical information was collected at the point of sampling. Information was obtained from the patient notes, electronic patient records or by direct questioning of the patient or their parents. The clinical variables (Table 2.2) collected were stored in a database, which was specifically designed for the study (B. Ahmed, Appendix A1). At each visit, presence of upper and lower

respiratory tract symptoms were recorded using a binary (yes or no) score. A clinical exacerbation was defined as a change in symptoms necessitating treatment with IV antibiotics as determined by an independent physician.

Gastro-oesophageal reflux disease (GORD) was defined either on positive pH study or clinical symptoms necessitating treatment with proton pump inhibitors or H₂ receptor antagonists. Spirometry was performed by children over 5 years of age supervised by the Paediatric Physiologist with values for FEV₁ and forced vital capacity (FVC) given in percentage predicted scores, adjusted for age, sex and height²⁰⁶. Body mass index (BMI) was calculated as the number of standard deviations (Z score) from the median using British 1990 growth reference charts (LMS growth - <http://www.healthforallchildren.com/shop-base/shop/software/lmsgrowth/>).

For *P. aeruginosa*, infection status at baseline was defined according to Lee *et al*²⁰⁷: “chronic” infection refers to growth of an organism in more than 50% of samples in the previous year; “intermittent” refers to growth of an organism in the previous year but in less than 50% of samples; “free” refers to growth of an organism previously but not in the past year, and “never” indicates an organism that has never been grown. Whilst these criteria were established for patients with CF, there are no equivalent criteria for growths of *P. aeruginosa* in PCD. Therefore, the same criteria were applied to both groups. For all other organisms and as no standardised criteria exist for defining chronic infections, previous growths were defined using a binary score: “yes”, if ever previously grown, and “no” if never previously grown. Positive cultures during the study

were recorded qualitatively as binary outcomes (“yes” for a positive culture and “no” for a negative culture). Disease complications, such as ABPA and CF-related diabetes (CFRD), were defined based on the diagnostic criteria listed in the RBH Clinical Guidelines on the Care of Children with CF (<http://www.rbht.nhs.uk/healthprofessionals/clinical-departments/cystic-fibrosis/clinical-cf-guidelines-care-of-children/>).

Table 2.2: Clinical information collected during study visits.

*Indicates clinical information collected at baseline visit only. † Infection status determined by clinical microbial cultures. Defined according to the criteria listed in Chapter 2, Section 2.3. GORD – gastro-oesophageal reflux disease; CFRD – CF-related diabetes; ABPA – allergic bronchopulmonary aspergillosis; CFTR – cystic fibrosis transmembrane conductance regulator; CT – computed tomography; NBS – newborn screened infants; IV – intravenous.

Demographics	Measurements	Previous results
<ul style="list-style-type: none"> Age Gender* 	<ul style="list-style-type: none"> Height (in cm) Weight (in kg) Spirometry (In children > 5 years age) 	<ul style="list-style-type: none"> CFTR Genotype* Any CT scan Any electron microscopy*
Past medical history	Infection status	Antibiotic use
Comorbidities/ Complications: <ul style="list-style-type: none"> GORD Pancreatic insufficiency CFRD Liver disease Haemoptysis Nasal polyposis ABPA 	Current symptoms: <ul style="list-style-type: none"> Upper respiratory Lower respiratory 	Antibiotic use*: <ul style="list-style-type: none"> Past 30 days Past 3 months
Birth history (in NBS group)*:	Infection status†:	Current antibiotics:
<ul style="list-style-type: none"> Mode of delivery Gestation Mode of feeding in infancy Meconium ileus 	Bacterial: <ul style="list-style-type: none"> <i>P. aeruginosa</i> <i>S. aureus</i> <i>H. influenzae</i> Non-tuberculous mycobacterium <i>Burkholderia cepacia</i> complex <i>Stenotrophomonas maltophilia</i> MRSA 	<ul style="list-style-type: none"> IV Oral Nebulised
Medication history		Prophylactic antibiotics
Current feeding route	Fungal: <ul style="list-style-type: none"> <i>Aspergillus fumigatus</i> <i>Candida</i> species 	Additional info at follow-up: Antibiotic use between visits: <ul style="list-style-type: none"> Antibiotic name Route Duration (days)

2.4 Sample collection and handling

The studies were approved by the RBH Biomedical Research Unit Advanced Lung Disease Biobank (NRES reference 10/H0504/9). The approval allows opportunistic collection of samples during scheduled clinical appointments at RBH. Informed written consent and age-appropriate assent was obtained from parents and children respectively and reconfirmed at each visit by a Paediatrician (B. Ahmed).

All samples were labelled with a unique sample identifier number at the point of collection. For the CLIMB study, paired clinical samples were sent to the Microbiology Department at RBH for fungal, bacteria and NTM culture as per routine care for patients with CSLD and processed in accordance with the Cystic Fibrosis Trust Guidelines²⁰⁸.

The temperature at which samples are stored may influence bacterial community structure. Cuthbertson *et al.*²⁰⁹ demonstrated significant changes in the microbiota when comparing sputum samples frozen immediately with those left at room temperature at varying intervals for up to 72 hours from 8 patients with CF. The greatest change in community structure occurred within the first hour that samples were left at room temperature. It is not known whether this observation applies only to sputum samples or all respiratory samples. Consequently all the samples obtained in this study were immediately frozen on dry ice prior to subsequent storage at -80°C. Samples remained frozen until DNA extraction.

2.4.1 Oropharyngeal swab collection

The majority (98%) of “oropharyngeal swabs” (cough swabs [CS] and throat swabs [TS]) were collected by a single operator (B. Ahmed), with the remainder collected either by a Paediatric Specialist Registrar or a Paediatric Physiotherapist. Oropharyngeal swabs were collected using cotton-tipped sterile swabs and a wooden tongue depressor.

Methods previously described for collecting TS include swabbing the posterior pharyngeal wall in anaesthetised patients⁷¹, the tonsils¹³⁵ or the soft palate³⁹, with reported validity of all techniques for identifying lower airway pathogens. For the longitudinal study, a method of TS sampling was required that could be repeatedly tolerated by awake children. TS were collected as described by Cardenas *et al.*³⁹, as this had been shown to be an effective method for distinguishing disease groups. TS were performed by first depressing the tongue with a wooden depressor and then carefully moving the swab back and forth in front of the uvula five times whilst applying even pressure. Care was taken not to touch any other surface to prevent contamination.

CS were collected by placing a swab into the posterior pharynx and asking the child to cough whilst carefully avoiding contact with the pharyngeal mucosa. CS were only collected in children old enough to cough on demand. Where both TS and CS were collected, TS was collected first. Technical swab controls were also taken from each batch of swabs by removing a sterile swab from its container, exposing to the air and then replacing it in the container and again placing on dry ice prior to freezing at -80°C.

All patients for the cross-sectional study were fasting at the time of sample collection. For patients in the longitudinal study, non-invasive, upper airway samples were collected at least 20 minutes after eating, in an attempt to limit the risk of vomiting and contamination of samples with food. Details of the collection methods for lower airway samples (bronchoalveolar lavage fluid [BALF] and bronchial brushings) are provided in Chapter 3, Section 3.2.2. Details of spontaneously expectorated sputum collection are provided in Chapter 5, Section 5.2.3.

2.5 Bacterial DNA extraction

DNA was extracted using the MP Bio FastDNA SPIN Kit for Soil according to the manufacturer's instructions but with minor modifications (as detailed below) for handling respiratory samples. The steps below describe the Standard Operating Procedure (SOP) in use in the Molecular Genetics and Genomics Group (Molgen) at NHLI. A microcentrifuge (Thermo Fisher Scientific, Waltham, USA) was used for all centrifugation steps. To limit microbial community composition changes during defrosting, all swab and bronchial brushing samples were kept frozen on dry ice until transference of swab or brush head to a Lysing Matrix E (LME) tube, to which 978 microlitres (μ l) of Sodium Phosphate Buffer had been added. For whole BALF, and in order to maximize cell yield, the sample was gently thawed on ice before splitting into 2 ml sterile microcentrifuge tubes followed by centrifugation at 21,000 $\times$ g for 30 minutes to pellet cellular matter. The resultant cell pellet was resuspended in sodium phosphate buffer and transferred to an LME tube. Supernatant was discarded. For sputum samples,

300 µl (or 1 weighed plug, if less than 300 µl of sputum was obtained) was defrosted and added directly to an LME tube.

Next 122 µl of MT buffer was added to each LME tube and samples were homogenised twice at a speed of 6,000 rpm for 30 seconds using the Precellys®24 homogeniser (Bertin Instruments, France). After the homogenization step, LME tubes were transferred to a benchtop microcentrifuge and centrifuged for 10 minutes at 14,000 x g in order to pellet cell debris. Supernatants were transferred to 2.0 ml sterile microcentrifuge tubes prior to addition of 250 µl of Protein Precipitation solution to each sample. As brushes and swabs can retain a percentage of the lysate, for these samples the brush or swab head was transferred to an autoclaved spin basket within a 2 ml microcentrifuge tube and centrifuged for a further minute at 14,000 x g. Any additional supernatant obtained was combined with the first and transferred to a 2.0 ml sterile microcentrifuge tube prior to addition of the 250 µl of Protein Precipitation Solution.

Samples were again centrifuged for 5 minutes at 14,000 x g to pellet the protein precipitate. Supernatant was removed and added to a 15 ml tube containing 1 ml of Binding Matrix. Samples were then rotated for 20 minutes to allow DNA binding to the matrix and then left upright for 3 minutes to allow the silica matrix to settle. 800 µl of supernatant was carefully removed, taking care not to disturb the settled binding matrix. The remainder of the protocol then followed the manufacturer's instructions, with inclusion of the manufacturer's optional 55°C for 5 minutes heating step to improve DNA yields (adaption introduced for

samples extracted after May 2013). Extracted DNA was transferred to a sterile 0.5 ml screw cap tube with 1 µl used to quantify DNA yield using a NanoDrop (Thermo Fisher Scientific, Waltham, USA). Samples were stored at -80°C until further use.

From September 2014 onwards, to control for the effects of contamination from DNA extraction kits, blank controls with no sample added were taken from each DNA extraction kit as recommended by Salter *et al.*²¹⁰. These were sequenced alongside patient samples and technical sampling controls for swabs, brushings and BALF.

2.6 16S rRNA Quantitative PCR (qPCR) for quantification of bacterial load

Bacterial load was quantified using the SYBR Fast qPCR Kit (KAPA BioSystems). A standard curve was constructed using a clone of *Vibrio natriegens* DSMZ 759 with 5 serial dilutions ranging from a concentration of 10⁴ to 10⁸ molecules/µl. Sample DNA was diluted 1:5 in PCR grade water to a total volume of 25 µl per sample. Custom, 16S rRNA gene primers were used (Illumina)²¹¹ targeting the V4 hypervariable region : forward primer S-D-Bact-0564-a-S-15 : 5' AYT GGG YDT AAA GNG 3', and reverse primer S-D-Bact-0785-b-A-18 : 5' TAC NVG GGT ATC TAA TCC 3'. These are the same primers and target region that were subsequently used for sequencing.

Sample reactions were set up in MicroAmp Fast-96 well reaction plates placed in a splash free support base to protect the SYBR green reagent from the light.

Bacterial DNA within samples was quantified in 15 µl triplicate reactions as follows:

- 7.5 µl of SYBR Fast qPCR Master Mix.
- 0.3 µl of each F and R primer (10 µM).
- 1.9 µl PCR grade water.
- 5 µl of template (bacterial standard, non-template control [PCR grade water] or sample DNA).

Plates were sealed with a MicroAMP Optical Adhesive Seal and qPCR reactions were run on the ViiA7 Real-Time PCR system (Applied Biosystems/ Thermo Fisher Scientific, Waltham, USA) with the following conditions:

- 90°C for 3 minutes
- 40 cycles of
 - 95°C for 20 seconds
 - 50°C for 30 seconds
 - 72°C for 30 seconds

Quantification was deemed reliable if a reaction efficiency of 60-70%, $R^2 > 0.998$ was attained and if the library dilutions amplified within the dynamic range of the assay. Bacterial DNA was quantified from the standard curve.

2.7 16S rRNA gene PCR amplification assay

The V4 hypervariable region of the 16S rRNA bacterial gene was amplified by PCR using custom, indexed primers (Illumina)²¹¹:

Forward primer S-D-Bact-0564-a-S-15 : 5' AYT GGG YDT AAA GNG 3'

Reverse primer S-D-Bact-0785-b-A-18 : 5' TAC NVG GGT ATC TAA TCC 3'

16S rRNA primer sequence was based on that recommended by Klindworth *et al.*²¹¹, the experimental design for dual-barcoding followed the recommendations in Kozich *et al.*²¹² and the adaptors were from Illumina. The primers consist of the following sequences:

- A P5 and P7 adaptor for hybridization to the flow cell during sequencing corresponding to the forward and reverse primer respectively.
- An 8bp index for multiplexing.
- A 10bp pad to prevent hairpin formation and increase the melting temperature to 65°C, to match that of the Illumina MiSeq sequencing primers.
- A 3bp linker sequence anti-complimentary to the 16S rRNA gene.
- A gene-specific primer²¹¹.

Quadruplicate 25 µl PCR reactions were performed for each extracted DNA sample. Reactions consisted of:

- 12.5 µl Q5 High Fidelity 2 x Master Mix *Taq* polymerase.
- 9.5 µl PCR grade water.
- 1 µl of each primer (1.5µM concentration).
- 1 µl of DNA template.

Preliminary work conducted within the Molgen comparing 7 types of Taq polymerase (Q5, Roche Taq, TaKaRa Premix Ex-Taq 2 X Mastermix (Takara Bio, Japan), Onetaq (NEB), Biotaq (Bioline, London, UK), AmpliTaq Gold® (Thermo Fisher Scientific, Waltham, USA) and Moltaq (Molzym Life Science, Germany) to optimise PCR conditions for 454 pyrosequencing showed that only Q5 consistently amplified DNA using barcoded primers with no evidence of contamination. Q5 was therefore chosen for all qualitative PCRs.

Ninety six well PCR plates were set up that in addition to DNA sample templates included: a mock community (Appendix A2) consisting of 27 bacterial genera common to the respiratory tract; a PCR negative control, where 1 µl of PCR grade water was added instead of DNA template, and at least 1 technical or DNA kit control added to each plate. Bacterial clones included in the mock community were: *Actinomyces odontolyticus*, *Bifidobacterium dentium*, *Burkholderia cepacia*, *Chlamydophila pneumoniae*, *Corynebacterium pseudodiphtheriticum*, *Enterococcus faecalis*, *Escherichia coli*, *Fusobacterium nucleatum*, *Granulicatella adiacens*, *Haemophilus influenzae*, *Haemophilus parainfluenzae*, *Leptotrichia buccalis*, *Moraxella catarrhalis*, *Mycobacterium psychotolerans*, *Mycobacterium bovis*, *Mycoplasma pneumoniae*, *Neisseria meningitidis*, *Neisseria flavescens*, *Norcardia farcinica*, *Pasteurella multocida*, *Pseudomonas fluorescens*, *Prevotella buccae*, *Rothia mucilaginosa*, *Salmonella enterica*, *Staphylococcus aureus*, *Streptococcus agalactiae*, *Streptococcus constellatus*, *Streptococcus infantis*, *Streptococcus mitis*, *Streptococcus parasanguinis*, *Streptococcus pneumoniae*, *Streptococcus pseudopneumoniae*, *Streptococcus pyogenes*, *Streptococcus sanguinis*, *Treponema denticola*, *Veillonella dispar*, *Vibrio* spp..

PCR cycling conditions were:

- One cycle of 95°C for 2 minutes.
- Thirty five cycles of
 - 95°C for 20 seconds
 - 50°C for 20 seconds
 - 72°C for 5 minutes

That PCR had been successful was confirmed by running aliquots of PCR products on 1.2% agarose gels according to methods described by Sambrook *et al.*²¹³. Samples were loaded using the loading dye provided in the MP Bio Fast DNA SPIN Kit for Soil and run alongside a 100 bp DNA ladder. First the positive and negative controls were run to confirm successful amplification and to examine for contamination. Following this, quadruplicate PCR reactions were pooled to a total volume of 100 µl per well in a 96 well round bottom plate and a second gel run to confirm successful sample amplification.

2.8 Amplicon purification

Next amplicon purification, using solid phase reversible immobilization (SPRI) magnetic beads (Agencourt Ampure XP), was performed to remove primer dimer and any remaining unused PCR reagents. DNA purification using Ampure XP is size and charge dependent. The size of DNA fragments removed is inversely proportional to the ratio of SPRI beads: larger fragments will be removed with a decreasing ratio of beads. For the V4 amplicon, a ratio of 0.7 µl of SPRI beads to 1 µl of amplicon was used.

First the Ampure XP solution was vortexed for 20 seconds to ensure the beads were fully resuspended. 70 μ L of Ampure XP was then added to 100 μ L of pooled amplicon in each well of the 96 well plate and mixed by pipetting up and down 20 times. The reaction mixture was left at room temperature for 10 minutes to allow binding of the DNA amplicons to the paramagnetic beads after which the plate was placed on a magnetic stand (Invitrogen Life Sciences/ Thermo Fisher Scientific, Waltham, USA) for 5 minutes. This resulted in separation of the DNA bound beads from the supernatant the latter then being removed and discarded, whilst taking care not to disturb the beads and thereby cause amplicon loss.

The plate was then removed from the magnetic stand and samples were washed twice in 100 μ L of fresh 80% ethanol. Between ethanol washes, the plate was placed back onto the magnetic stand, left for 1 minute and the supernatant again removed and discarded. The duration for which the beads were left at room temperature between ethanol washes was limited to prevent over drying and irreversible binding of DNA to the beads to ensure minimal loss of amplicon. Following the second ethanol wash, the plate was left on the magnetic stand at room temperature for 2 minutes to allow evaporation of excess ethanol without over drying. Samples were then eluted in 20 μ L of low EDTA: TE buffer (pH 8.0) and left at room temperature for 5 minutes. The plate was then replaced on the magnetic stand for 2 minutes and the supernatant containing the DNA was removed and retained in a new 96 well plate. The plate was sealed using a plate sealer and stored at 4°C.

2.9 DNA quantification and equimolar pooling for library normalisation and repeat purification

Post the first round of purification, amplicon concentrations were calculated using the Quant-iT PicoGreen dsDNA quantification assay. This assay uses a fluorescent nucleic acid stain to reliably quantify dsDNA. The relationship of the intensity of the fluorescence signal from the PicoGreen reagent to the quantity of dsDNA present is linear (according to the manufacturer's information). Quantification is calculated by comparison against a standard curve of control DNA of known concentration.

The standard curve (100 ng, 50 ng, 25 ng, 12.5 ng, 6.25 ng, 3.13 ng, 1.56 ng and 0 ng) was set up using DNA lambda (100 ng/ μ l) provided within the kit diluted in 1 x TE (pH 7.5). For the first standard, DNA was diluted 100 fold by adding 6 μ l of DNA to 564 μ l of 1 x TE (1ng/ μ l). For the next 6 standards, a 1:2 dilution was made by adding 300 μ l of the previous standard to 300 μ l of 1 x TE and vortexing for 10 seconds. A blank control containing 300 μ l of 1 x TE with no DNA added was used as the final standard. One hundred μ l of each standard was added in duplicate to a 96 well black fluorometer plate.

99 μ L of 1 x TE and 1 μ l of sample amplicon were added to each of the remaining 80 wells in the 96 well plate and mixed by pipetting up and down 4 times. 100 μ l of a 1:200 dilution of the PicoGreen fluorescent dye was added to each well in the 96 well plate to make a final reaction volume of 200 μ l and again mixed by pipetting up and down 4 times. DNA fluorescence was measured under a

microplate reader with samples excited at 480 nm and the fluorescent signal measured at 520 nm.

DNA quantification was considered to be reliable when the R^2 value of the standard curve was above 0.998. Amplicon concentration was calculated from the quadratic equation for the line of best fit for the standard curve ($y = ax + b$). 20 to 30 ng of each amplicon was added to an equimolar, normalised library pool. In instances where total amplicon concentration was less than 30 ng, the entire amplicon volume was added to maximise the chances of successful sequencing. A second round of Ampure XP purification was then performed on the equimolar pool. At this stage, the following modifications were included: a 2 ml microcentrifuge tube format was used (see workflow diagram Figure 2.1); 200 μ l of 80% ethanol was added at each wash step, and following the second ethanol wash the library pool was left to dry at room temperature until all the ethanol had evaporated to prevent loss of the sample during agarose gel extraction.

2.10 Agarose gel purification of equimolar pool

To ensure no remaining excess primer dimer was present in the equimolar pool, agarose gel purification was performed. The gel was prepared by dissolving 1.85 g of 1% agarose in 150 ml of 1 x TB buffer, to which 6 μ l of gel red was added, the gel poured and then left to set for 30 minutes. Once set, 10 μ l of loading dye was added to 30 μ l of library and loaded into the gel, with 5 μ l of 100 bp ladder added to each lane either side of the library with the gel run at 100V

for 45 minutes to allow separation of the amplicon from primer dimer. Bands were visualized using an ultra violet light box and the band corresponding to the expected size of the amplicon (200 bp) excised and placed in a 1.5 ml sterile microcentrifuge tube. The microcentrifuge tube was weighed before and after addition of the gel slice to determine the weight of the gel.

Gel extraction was performed using the QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. QG buffer was added at a volume of three times the weight of the gel slice and incubated at 50°C for up to 10 minutes until the gel had completely dissolved. Next one volume (equivalent to the weight of the gel slice) of isopropanol was added and the reaction mixture was transferred to a QIAquick spin column and centrifuged twice at 17,000 x g for 1 minute. To increase the yield of the DNA library obtained, the flow through was retained and placed back onto the spin column after the first centrifugation and centrifuged again after which the flow through was discarded. All subsequent centrifugation steps were at 14,000 x g. 500 µl of Buffer QG was then added, centrifuged for a further minute and the flow through discarded. 750 µl of Buffer PE was added and the column left to stand for 2-5 minutes to further increase yield. This was centrifuged twice for 1 minute, with the flow through discarded after each round. To elute the sample 30 µl of Buffer EB was added to the column followed by incubation at room temperature for 4 minutes. The column was then transferred into a clean 1.5 ml microcentrifuge tube and centrifuged for 1 minute and the supernatant containing the purified library retained.

To confirm primer dimer removal and assess library quality, library DNA fragments were analysed using the Agilent BioAnalyzer High Sensitivity DNA analysis kit according to the manufacturer's instructions.

2.11 Quantitative PCR (qPCR) for library quantification

Library quantification was performed using the Kapa library quantification kits for Illumina platforms (Kapa Biosystems), which contains a pre-prepared standard in a ten-fold dilution series, Kapa SYBR® FAST qPCR Master Mix (2X) and primers 1 and 2 targeting the Illumina adaptors.

The qPCR step is to ensure that the final library pool is at the optimum concentration for flow cell loading on the Illumina MiSeq. The dynamic range of the DNA standards provided ranges from 20 – 0.0002 pM. To ensure that the quantity of the sample library fell within this range and accurate quantification, 1 µl of the library pool was taken and diluted 1:1000 in 999 µl of PCR grade water. Three further 1:2 dilutions of the library pool were made to achieve library concentrations of 1:2000, 1:4000 and 1:8000. Each dilution of the standard, PCR grade water and dilutions of the library pool were added sequentially in triplicate to a MicroAmp Fast 96-well reaction plate.

Each 20 µl reaction consisted of:

- 12 µl of mastermix to which Illumina primers were added.
- 4 µl of PCR grade water.
- 4 µl of template (DNA standard or library dilution or PCR grade water).

PCR reactions were run on the ViiA7 Real-Time PCR system (Applied Biosystems/ Thermo Fisher Scientific, Waltham, USA). The cycling conditions were:

- 95°C for 5 minutes
- Then 35 cycles of
 - 95°C for 30 seconds
 - 60°C for 45 seconds
 - and melting at 65 – 95°C

Quantification was deemed reliable if a reaction efficiency of 90-110%, $R^2 > 0.998$ was attained and if the library dilutions amplified within the dynamic range of the assay. Using the standard curve and the average fragment length determined from the BioAnalyzer trace, the library concentration was calculated using the formula:

$$\text{Library concentration} = \text{qPCR concentration} \times (452 \text{ bp} / \text{average fragment length}) \times \text{dilution factor}.$$

2.12 16S rRNA gene sequencing using the Illumina MiSeq

The MiSeq Reagent Kit v2 (Illumina) was used to generate paired reads of length 250 bp. Gene sequencing using the Illumina MiSeq occurs on the surface of the flow cell, a thick glass slide consisting of lanes of oligonucleotides complementary to the P5 and P7 adaptor regions of the PCR primers. Single stranded DNA libraries hybridize to these oligonucleotides and a complementary

strand is cleaved by DNA polymerases. Bridge amplification is used to multiply the DNA library massively in parallel, resulting in millions of replicates bound strongly to the surface of the flow cell in clusters. The forward and reverse reads are then sequenced sequentially by hybridization to the corresponding sequencing primers. Base-calling occurs by adding fluorescent-labelled nucleoside triphosphates (NTPs) and capturing the image generated. The unique sequence of the index primer allows subsequent sample identification.

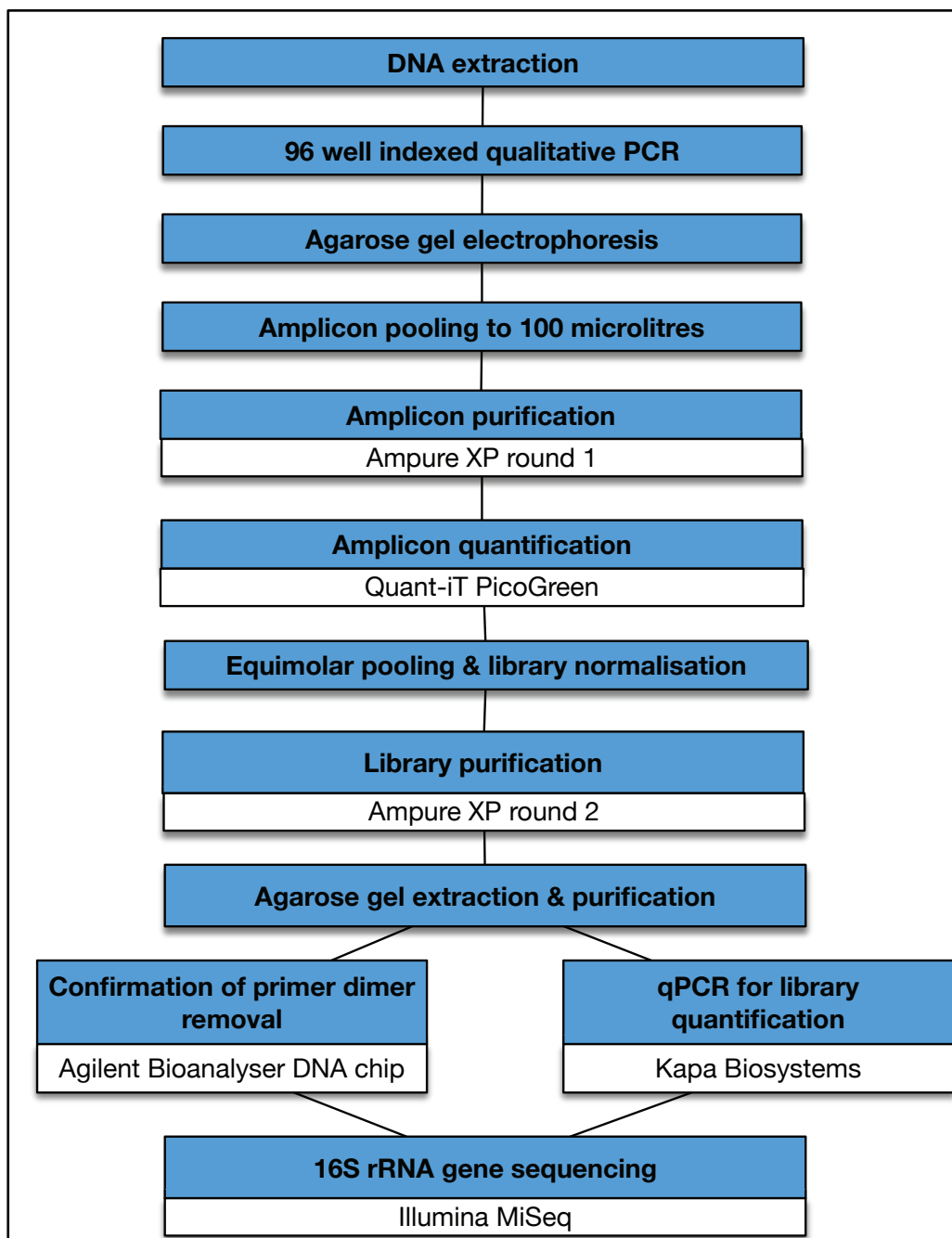
A control DNA library from the phage PhiX (Illumina) was added (as per Illumina protocol) during MiSeq reactions to artificially inflate genetic diversity for low diversity libraries. A PhiX spike is necessary because the optics of the MiSeq utilize two light channels: a red channel for A and C base calls, and a green channel for G and T. Both of these must be excited at each cycle for a successful sequencing run. A diverse set of clusters is therefore required. The desired library concentration and PhiX spike was determined from validation work done previously within the Molecular Genetics and Genomics Group (NHLLI, Imperial College unpublished data). Libraries were denatured with fresh 0.2 N NaOH, diluted to an 8 pM concentration using HT1 buffer (Illumina) and spiked with 20% of 8 pM Illumina PhiX control library. This was loaded onto the MiSeq cartridge.

4 µl of the 3 sequencing primers for read 1, the index and read 2 were added to the MiSeq cartridge. Libraries and primers were loaded onto the MiSeq cartridge using extra-long 200 µl gel loading tips and gently mixed. A mapping file containing sample ID and the sequences of the individual index primers was

loaded onto the MiSeq software and the reaction the run for 500 cycles with an approximate duration of 39 hours.

Upon completion of each run, quality of the sequencing run was assessed by: the total number of reads generated; the percentage of bases with a Phred quality score greater than 30 (Q30, ideally > 75%); cluster density (should be between 500 and 900 k/mm²), and the percentage of clusters passing filter (ideally > 70%). Each MiSeq run generated a FASTQ file containing the sequences generated per sample and a quality score per base, which was then analysed using Quantitative Insights Into Microbial Ecology (QIIME).

Figure 2.1: Library preparation workflow for 16S rRNA gene sequencing.



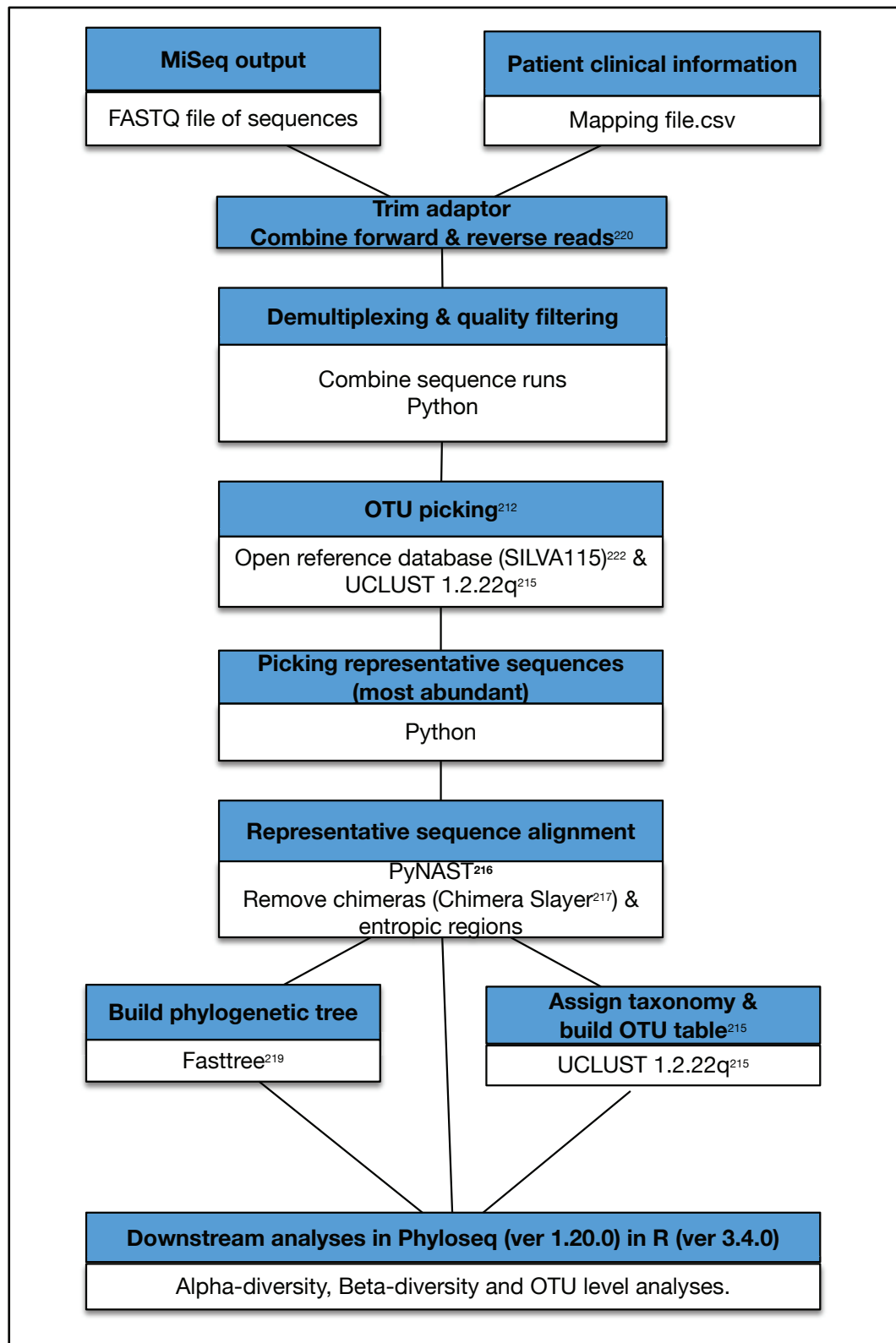
2.13 Data analysis

2.13.1 Sequence analysis in QIIME

Upstream analyses were performed using QIIME (Version 1.9.0)²¹⁴⁻²¹⁹ (workflow diagram shown in Figure 2.2). QIIME is an open-source multi-platform pipeline for microbial community analyses. To perform these analyses, QIIME requires two input files: the FASTQ file generated from the MiSeq and a mapping file containing the sample metadata. The mapping file is used to assign sequences to their original sample based on the unique barcode of the index primer. The minimum information required therefore is the sample identifier, barcode sequence, linker primer sequence and a description of the sample. Clinical variables may also be included for use in downstream analyses.

Figure 2.2: Workflow diagram for data analyses.

Upstream analyses were performed in QIIME (version 1.9.0)²¹⁴ and downstream analyses in phyloseq (version 1.20.0) in R (version 3.4.0).



QIIME is currently best designed for analysing single-end data. In order to process paired-end reads (the forward and reverse reads) it is necessary to combine them into a single read using the command `fastq-join`²²⁰. To do this the 8 bp forward and reverse barcodes were first combined into a single 16 bp unique identifier. To maintain sequence alignment, the ends of the sequences were trimmed and the forward and reverse reads then combined into a single read allowing a minimum overlap length of 200 bp and 10% misalignment. For any instances of mismatch, the base with the highest quality score was retained.

Next samples were de-multiplexed thereby assigning sequences back to their original sample. Following this, barcode and primer sequences were removed, as they were no longer required. Quality filtering was performed as follows. Phred scores for each base were assessed and any base with a score below the minimum quality threshold ($< Q30$) was removed. If 10 consecutive bases failed to meet the threshold, the read was truncated. The resulting sequence was discarded if $< 70\%$ of consecutive bases in the combined read were $\geq Q30$. During de-multiplexing, multiple sequencing runs containing samples from the same study group were combined into one dataset. Sequences generated from PhiX control library were removed after de-multiplexing using the Burrows-Wheeler Alignment tool (BWA)²²¹.

Taking the resulting sequences, operational taxonomic units (OTUs, representing clusters of similar 16S rRNA gene sequences) were next assigned using an open reference approach with sequences clustered against the SILVA ribosomal RNA gene database (Version 115)²²². Using this approach, sequences are matched

against the reference database; sequences which fail to be matched are added as *de novo*. Input sequences were initially pre-filtered against the reference set with a percent identity threshold of 60%. Sequences falling below this threshold likely represent sequencing errors and were therefore discarded. OTU picking was performed in UCLUST (Version 1.2.22q)²¹⁵ with a threshold of 97% sequence similarity.

The most abundant sequence was picked as a representative for the OTU cluster. Sequences were aligned to one another using the Python Nearest Alignment Space Termination (PyNAST) tool²¹⁶ to allow the removal of chimeric sequences using ChimeraSlayer²¹⁷, taxonomy was assigned using the RDP Classifier 2.2²¹⁸ and the SILVA ribosomal RNA gene database (Version 115)²²² and subsequent construction of a phylogenetic tree using FastTree (Version 2.1.3)²¹⁹. Species identity of the representative sequence was then assigned using UCLUST (Version 1.2.22q)²¹⁵ and the SILVA ribosomal RNA gene database (Version 115)²²². An OTU table was constructed detailing the OTUs present per sample. The OTU table, phylogenetic tree and representative sequences were uploaded into R for downstream analyses.

2.13.2 Microbial community analysis in R

Downstream analyses were performed using both R (Version 3.4.0 “You Stupid Darkness”) and Rstudio (Version 0.99.473) and the following packages:

- ape²²³ (version 4.1)– for analyses of phylogenetics and evolution.

- Biostrings²²⁴ (version 2.44.1)– to string objects representing biological sequences and matching algorithms.
- ggplot2²²⁵ (version 2.2.1)– a graphical package.
- gridExtra²²⁶ (version 2.2.1)– for constructing grid based plots.
- Multtest²²⁷ (version 2.32.0)– resampling-based multiple hypothesis testing.
- phyloseq²²⁸ (version 1.20.0)– for analysing microbiota data.
- plyr²²⁹ (version 1.8.4)– for splitting, applying and combining data.
- reshape2²³⁰ (version 1.4.2)– for reshaping data.
- stringr (version 1.2.0)– wrappers for common string operations.
- vegan²³¹ (version 2.4.3)– a community ecology package.

2.13.2.1 Quality and contamination control checks

Prior to further analyses, and as sample numbers meant a number of MiSeq runs were conducted, data quality was assessed to ensure sequences were of high quality and that quality was consistent across sequencing runs. To assess sequencing quality, barplots were constructed from the run quality statistics generated in QIIME allowing changes in quality scores along the length of the sequence using ggplot2 to be evaluated. To assess for batch effects, 2 plots were created: Non-metric multi-dimensional scaling (NMDS) ordinations to visualise differences in clustering between samples and barplots to allow comparison of the mock communities.

Contamination of samples can have a significant impact on data analyses, particularly for low biomass samples²¹⁰. PCR negative controls, technical controls and DNA kit control samples were therefore screened for contaminants. OTUs were considered to be contaminants if they were highly abundant within the controls and had either previously been reported as kit contaminants²¹⁰ or had not been reported in association with human disease. OTUs identified as contaminants were removed.

2.13.2.2 Pre-processing

Low abundance OTUs with less than 20 reads were removed to account for spurious reads. The sequencing depth varied by several orders of magnitude between samples. If the raw data was compared directly between samples due to variation in sequencing depth there was the potential risk of true biological trends being missed or wrongly identified. In order to remove this risk, prior to downstream analyses sample counts were standardized by rarefying to an even sequencing depth, termed “rarefaction level”. Rarefaction performs random subsampling of a population without replacement. It is the most commonly reported method of standardization but can result in drop out samples depending on the chosen rarefaction level. The higher the rarefaction level, the greater the number of samples removed; the lower the rarefaction level, the greater the number of OTUs that may be missed.

The rarefaction level was therefore chosen by examining rarefaction curves to determine the number of reads at which most samples reached an asymptote. At

this point, most of the OTUs are likely to have been captured and further sampling is unlikely to capture additional OTUs. Any sample with a total number of reads below the rarefaction level will be discarded. Therefore, a rarefaction level was chosen that would balance capturing the majority of OTUs, whilst still allowing sufficient samples to be retained for downstream analyses.

2.13.2.3 Diversity analyses

Differences in diversity were measured using the phyloseq²²⁸ and vegan²³¹ packages.

Alpha-diversity

Alpha-diversity was measured by species richness, evenness, the Shannon and Inverse Simpson's diversity indices. The Shannon diversity index quantifies the uncertainty in predicting species identity when randomly sampling from a community. The Inverse Simpson's diversity index quantifies the probability that two species randomly sampled from a community without replacement will be the same. The inverse index is used to provide a more intuitive score where the index increases as diversity increases.

To compare differences in alpha-diversity between samples, first the Shapiro-Wilk test for normality was used to determine the distribution of the data. For parametric variables, paired t-tests were used to compare 2 samples and ANOVA for > 2 samples. For non-parametric variables, the Wilcoxon signed rank test was used to compare 2 paired samples and the Kruskal-Wallis test for > 2

samples. Where significant differences were found, these were illustrated using box plots. Kendall rank correlation was used to compare two continuous variables for longitudinal data (e.g. changes in diversity with age). Further details of individual alpha-diversity tests used for each study group are given in the corresponding results chapters.

[Beta-diversity](#)

Beta-diversity measures between samples diversity. This was tested using the Bray-Curtis dissimilarity index²³², UniFrac²³³ and weighted UniFrac²³⁴ scores. The Bray-Curtis index measures dissimilarity between two samples by estimating the number of shared organisms from the total number of organisms present in both samples. The UniFrac score is a qualitative measure of the phylogenetic distance between two communities, with the weighted UniFrac score additionally accounting for the relative abundance of organisms within these samples. Quantitative measures of beta diversity, such as Bray Curtis dissimilarity and weighted UniFrac scores, provide information on community differences due to the relative abundance of species. The UniFrac score provides information on community differences driven by selective pressures on the community²³⁴. The results of these scores are provided in a numerical distance matrix, which can then be used for further analyses.

Non-metric multi-dimensional scaling (NMDS) plots were used to visualize trends in beta-diversity. NMDS uses the rank order abundance of OTUs based on a distance matrix to plot their position in multi-dimensional space, such that samples that are more closely related will cluster together. Statistical

comparisons of beta-diversity between samples were performed using the `adonis` function in `vegan`²³¹, which performs a permutational multivariate analysis of variance using distance matrices. The “`strata`” argument of this function, which assigns groups or “blocks” within which to constrain permutations, was used to account for the repeated measures study design in both the cross-sectional and longitudinal studies.

2.13.2.4 OTU level comparisons

Testing for differences between communities at OTU and genera levels was performed by the following methods. To visualise trends in community structure, heatmaps of the fifty most common OTUs and individual patients barplots were constructed using the `ggplot2`²²⁵ and `gridExtra`²²⁶ packages. To compare for differences in individual genera between samples, the “`Multtest`”²²⁷ package was used. This package performs multiple testing procedures (e.g. *t*-tests or correlations) whilst controlling for family-wise error rate (FWER) and false discovery rate (FDR). The specific testing procedure used with `Multtest` is described in the corresponding results chapters. Specific OTU and genera level analyses for each study population are also detailed in the corresponding results chapters.

Chapter 3: Comparison of the upper and lower airway microbiota in children with chronic lung diseases

3.1 Introduction

There are few paediatric longitudinal studies of the lung microbiota in part due to the difficulties in obtaining repeated lower airway samples in young children. As described in Chapter 1 (Section 1.2), TS or CS are often used as surrogates for lower airway samples but their use (particularly CS) in representing the lower airway microbiota in young children and infants has not yet been validated.

This first results chapter therefore focuses on a comparison of the upper and lower airway microbiota to assess whether upper airway samples (CS and/or TS) could act as surrogates for samples obtained by FOB (BALF or bronchial brushings) in characterizing the lower airway microbiota for use in longitudinal sampling. This was assessed by comparing the microbiota on paired upper and lower airway samples collected from children undergoing a clinically indicated FOB at RBH. It was hypothesised that upper airway samples would reflect the diversity and community composition of the lower airway microbiota.

3.2 Methods

3.2.1 Subjects

Children were recruited from all those undergoing a clinically indicated FOB at RBH between November 2012 and April 2013. Indications included the

investigation of recurrent LRTIs and the management of chronic lung diseases, such as CF, for which recent antibiotic courses may have been prescribed. Consequently, it was not feasible to control for recent antibiotic use and these children were not excluded. Information on antibiotic usage was captured from patient records and parent recollection.

3.2.2 Sampling

Samples were collected on the day of FOB. At least one upper airway (TS and CS in those children able to cough on demand to perform CS) and a paired lower airway sample (BALF or bronchial brushing) were collected from each child. Upper airway samples were collected and stored as described in Chapter 2, Section 2.4.1. Lower airway samples were collected as described below (Sections 3.2.2.1 and 3.2.2.2). All samples were placed immediately on dry ice and stored at -80°C prior to DNA extraction.

3.2.2.1 Bronchoalveolar lavage samples (BALF)

BALF was collected using a SOP based on the recommendations of the European Respiratory Society BAL Task Force²³⁵ with the child under a general anaesthetic. FOB was performed by a Paediatric Respiratory Consultant. The bronchoscope was passed into the airway either nasally or orally via a laryngeal mask airway (LMA). The choice to use an LMA was determined by the consultant anaesthetist depending on the indication for FOB (and the clinical importance of visualising the upper airways, which would be precluded by LMA) and mode of anaesthesia used (e.g. IV or gas induction). To avoid upper airway

contamination, no suction was used until the cords had been traversed. As far as possible to account for lobar differences in bacterial infection²³⁶, aliquots of 1 ml/kg of sterile 0.9% saline were instilled into 2 lobes and aspirated. Lobes were usually the right middle lobe (RML, 3 aliquots) and one other lobe, most commonly the lingula or the macroscopically most severely affected lobe (one aliquot). After aliquoting 5 ml of the BALF for routine clinical culture, virology and cytology, the remainder was pooled together in a sterile DNase/RNase free 15 ml tubes and homogeneity ensured by inversion of the container three times. The pooled sample was split into 4 ml aliquots for DNA extraction for 16S rRNA gene sequencing. Samples were stored in sterile DNase/RNase free 15 ml tubes. Technical controls were taken monthly for all FOB samples. For BALF, these were collected using the same methods in use in clinical practice for collecting “prewash” BALF by passing 4 mls of sterile 0.9% saline through the bronchoscope prior to the bronchoscopy being passed into the patient. Bacterial culture of BALF was performed as per standard clinical practice according to the Cystic Fibrosis Trust Guidelines²⁰⁸ in the Clinical Microbiology Department at RBH.

3.2.2.2 Bronchial brushings

Bronchial brushings were taken by using either a 1.2 mm diameter, unsheathed or 2 mm diameter, sheathed bronchial cytology brush (Olympus Keymed, UK) depending on the size of the bronchoscope used (dictated by the age and size of the child). Brushings were taken without resheathing as previous evidence has shown that this method increases cell yield more than two fold in comparison to

resheathed brushings in the context of bronchial carcinoma ²³⁷. To date no such comparison has been published in relation to microbiological diagnosis, despite there being a theoretical risk of contamination with upper airway pathogens by leaving the brush unsheathed. Immediately after the brush was removed, the tip was cut into a sterile DNase/RNase free 15 ml tube. Technical controls were taken by passing an unsheathed brush through the bronchoscope prior to the bronchoscope being used in the patient.

3.2.3 16S rRNA gene library preparation and sequencing

For DNA extraction of BALF, in order to maximize cell yield, the sample was gently thawed on ice and a maximum of 2 x 2 ml aliquots of BALF (median 3.2 mls, range 0.5 – 4 ml) were centrifuged at 21,000 x g in a microcentrifuge (Heraeus Pico 21 microcentrifuge, Thermo Fisher Scientific, USA) for 30 minutes to pellet cellular matter. The resultant cell pellet was resuspended in Sodium Phosphate Buffer and transferred to an LME tube (Chapter 2 Section 2.5). Supernatant was discarded. For DNA extraction from bronchial brushes, brush heads were kept frozen on dry ice until transference to a LME tube, to which 978 microlitres (µl) of Sodium Phosphate Buffer had been added. DNA extraction and library preparation for all samples collected was thereafter performed as described in Chapter 2 Sections 2.5 and Sections 2.7 – 2.10. With the exception of 10 samples, all samples from an individual patient were run on the same plate, limiting the impact of any batch effect on comparisons of upper and lower airway samples. The exceptions were: TS for patients Subject numbers 16, 44,

48 and 49; CS for Subject number 18; BALF for Subject numbers 24, 47, 55, and bronchial brushings for Subject numbers 34 and 45.

3.2.4 Data analysis

Due to a lack of pilot data describing the lower airway microbiota in children, an *a priori* power calculation could not be performed and sample size was opportunistic. Data analysis was performed as described in Chapter 2 Section 2.13 with the following additions. Samples were rarefied to 1000 reads prior to downstream processing. From examining rarefaction curves (Figure 3.1), this rarefaction level was felt to reflect a balance between capturing the majority of OTUs whilst still allowing sufficient samples to be retained for paired upper and lower airway comparisons as the majority of samples had < 5,000 reads (Figure 3.2).

Bland-Altman plots were constructed to analyse the agreement in alpha diversity (richness, evenness and Shannon diversity index) between samples. When performing PERMANOVA²³¹ to compare beta diversity between upper and lower airway samples, a blocked design was used to account for paired comparisons within the same patient. Spearman's rank correlation was used to test whether the most abundant organisms were similar between samples. Multtest was used to perform multiple t-tests with a Benjamini Hochberg correction to compare differences in the relative abundance of the fifty most common genera between upper and lower airway samples. A *P* value of less than 0.05 was considered to

be statistically significant. Sequence data has been uploaded to the European Nucleotide Archive (Accession number: PRJEB14074).

Figure 3.1: Rarefaction curves with yellow lines denoting the number of OTUs sampled at 1,000 reads and 3,000 reads.

This illustrates that an asymptote is reached by 1,000 reads. At this threshold, the majority of OTUs will have been sampled and little additional information will be obtained at higher rarefaction levels (e.g. 3,000 reads). Therefore a rarefaction level of 1,000 reads was chosen.

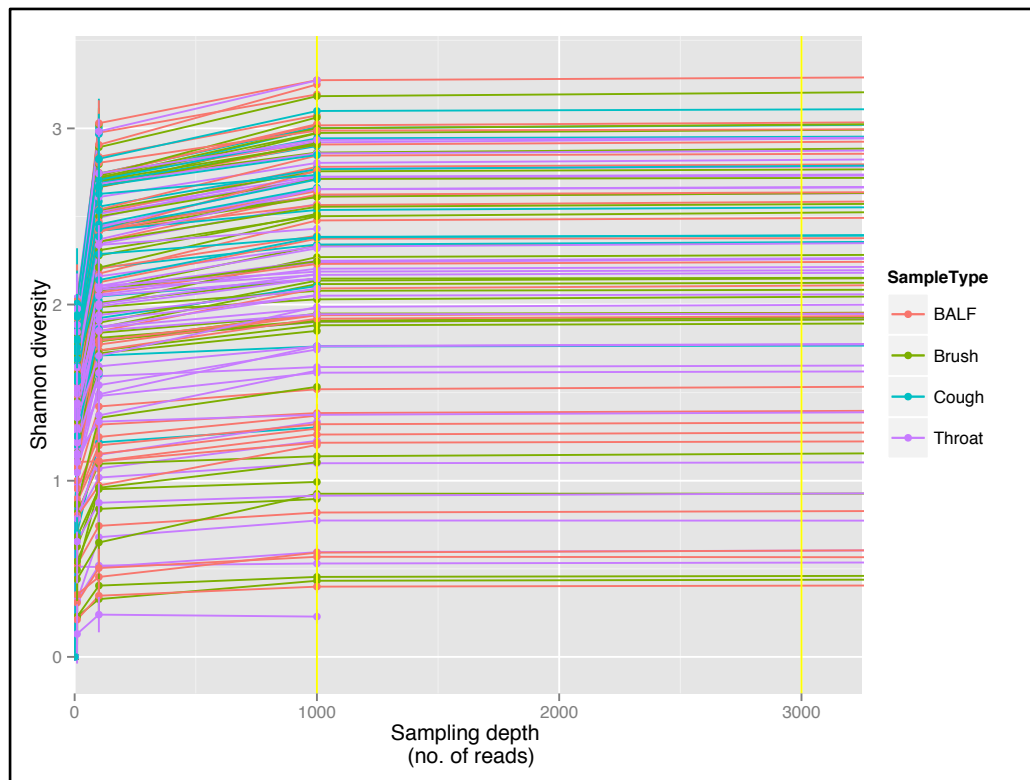
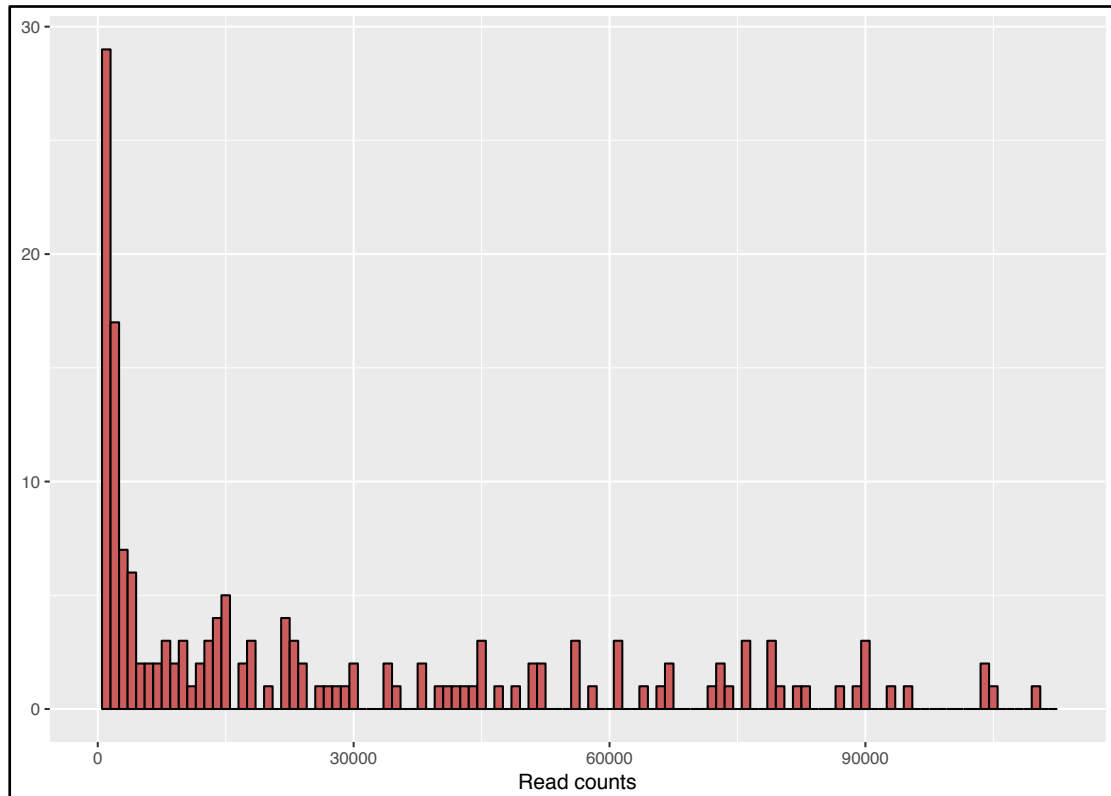


Figure 3.2: Sequencing depth for samples in upper and lower airway microbiota comparison.

This illustrates the distribution of read counts (sequencing depth) for both upper and lower airway samples. The percentage of samples at a given read count is shown. The majority of samples had < 5,000 reads.



3.3 Results

3.3.1 Patient demographics & sampling

Eighty-one children underwent a clinically indicated FOB during the study period of which sixty-five children (80%) were recruited. Forty-nine out of sixty five children (75%) were included in the study. Reasons for exclusion were: lack of paired upper and lower airway samples, either due to lack of consent for bronchial brushings and insufficient BALF volume obtained for both clinical microbiology and 16S rRNA gene sequencing thus no lower airway sample being obtained (N = 15), and lack of consent for publication of findings (N = 1).

Patient characteristics are detailed in Table 3.1. Indications for bronchoscopy included: CSLD (45%), namely CF and PCD, diagnosed on standard criteria^{238, 204}, and non-CSLD controls (55%), with the majority of the controls (24/27, 89%) being investigated for recurrent LRTIs. For those with CSLD, twenty (25%) were infants with CF diagnosed on newborn screening (NBS) undergoing a routine bronchoscopy at 3-5 months of age¹⁰⁵. The age distribution of patients was skewed with a median of 5.4 years (0.1 – 16.2 years). Thirty-nine patients (80%) were clinically stable at the time of bronchoscopy. Eighteen patients (37%) had received a treatment course of antibiotics in the previous 30 days, with 29% receiving antibiotics at the time of bronchoscopy (either intravenous [IV], oral, nebulised or a combination).

For 23 patients at least one of each sample type (CS, TS, BALF, bronchial brushing) was collected. In the remainder, at least one pair of upper and lower airway samples was collected allowing comparison (Figure 3.3). These included 47 TS, 37 CS, 42 BALF samples and 42 bronchial brushings. Twenty-six of the BALFs were culture positive. Organisms identified on BALF culture were: “Upper respiratory tract flora” (N = 13); *S. aureus* (N = 5); *H. influenzae* (N = 4); *Streptococcus pneumoniae* (N = 2); *Serratia marcescens* (N = 1), and *P. aeruginosa* (N = 1).

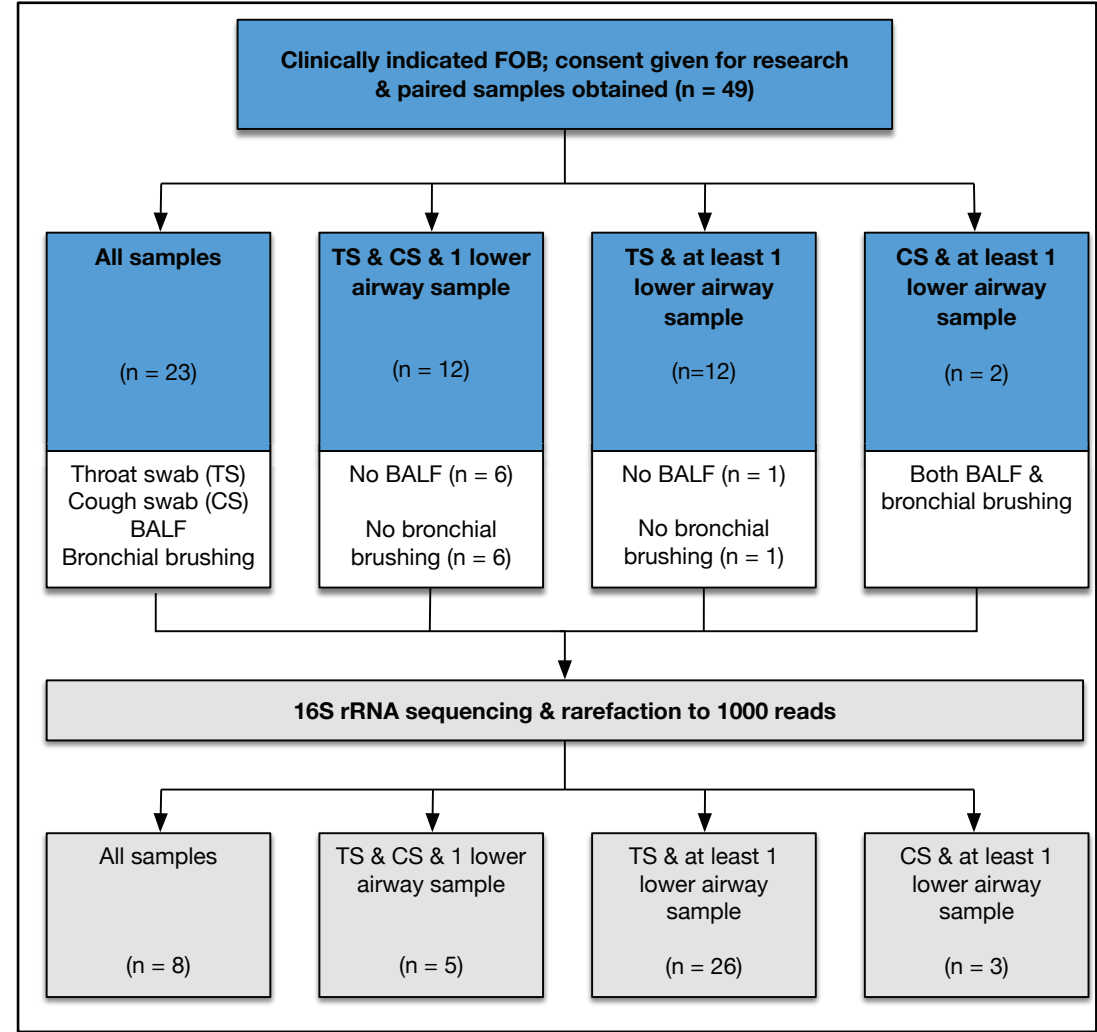
Table 3.1: Summary of patient characteristics in the comparison of the upper and lower airway microbiota (N = 49).

CSLD – chronic suppurative lung disease; CF – cystic fibrosis; NBS – newborn screened; PCD – Primary ciliary dyskinesia.

Demographic	Median (range) or n (%)
Age (years)	5.4 (0.1 – 16.2)
Gender (female)	29 (59)
Underlying pathology	
CSLD	22 (45)
• CF (Newborn screened [NBS] infants)	• 12 (25)
• CF (not NBS)	• 5 (10)
• PCD	• 5 (10)
Non-CSLD “controls”	27 (55)
• Recurrent LRTI	• 24 (49)
• Upper airway pathology	• 2 (4)
• Haemoptysis	• 1 (2)
Gastro-oesophageal reflux disease (GORD)	10 (20)
Current symptoms	
• Lower respiratory	• 10 (20)
• Upper respiratory	• 21 (42)
Antibiotic usage	
Current treatment course	14 (29)
• Intravenous (IV)	• 3 (6)
• Oral	• 8 (16)
• Nebulised	• 6 (12)
Oral prophylaxis	20 (41)
• Augmentin	• 2 (4)
• Azithromycin	• 9 (18)
• Flucloxacillin	• 9 (18)
Inhaled corticosteroids	12 (25)
Bronchoscopy route	
• Oral (Via laryngeal mask airway)	• 11 (22)
• Nasal	• 38 (78)

Figure 3.3: Illustration of the combinations of upper and lower airway samples taken.

In 47 patients, TS and at least 1 lower airway sample was collected for comparison. Samples for forty-two patients (of varying combinations i.e. all four samples types, TS and CS plus one lower airway sample etc.) remained after rarefying to 1,000 reads.



3.3.2 Sequencing

16S rRNA gene sequencing was performed on 168 extracted DNA samples. After combining the forward and reverse reads, de-multiplexing and quality filtering, a total of 5.97 million reads were obtained with an average of 31,117 reads per sample (range 118 – 192,812 reads). Run quality data is given in Table 3.2.

Table 3.2: Summary of sequencing quality statistics for the comparison of the upper and lower airway microbiota.

This illustrates the number of paired reads (in millions) generated by each sequencing run (“BAX1” and “BAX2”) and their associated quality statistics. Shown are: the percentage of bases with a Phred quality score greater than 30 (Q30); cluster density, and percentage of clusters passing filter (ideally > 70%).

Run no	No of reads (in millions, M)	> Q30 (%)	Cluster density (k/mm ²)	Clusters passing filter (%)
BAX1	7.19	82.8	433	68.8
BAX2	7.96	81.2	413	81

Technical and PCR negative controls were examined to identify potential contaminant OTUs, which were then removed (Figure 3.4 and 3.5). Contaminant OTUs included *Burkholderia* (OTU ID 1606), *Undibacterium* (OTU ID 1727) and *Ralstonia* (OTU ID 1703). The genera *Bradyrhizobium* spp., *Sediminibacterium* spp. and *Methylobacterium* spp. were also removed as these have previously been identified as common reagent contaminants²¹⁰. NMDS plots illustrated consistency between sequencing runs, with both mock communities (“mock”) clustering together and negative PCR controls (“negative”) and technical controls (“control BALF”) being more dispersed (Figure 3.6). A small significant batch

effect was found between the two sequencing runs using Bray Curtis dissimilarity and weighted UniFrac scores ($r^2 = 0.03$ and 0.01 respectively, $P = 0.002$), but not the unweighted UniFrac score ($r^2 = 0.01$, $P > 0.05$) (Figure 3.7). Examining the mock communities, all expected 27 genera were present from each sequencing plate confirming successful sequencing. For run number BAX2 (represented by “mock 2” in Figure 3.8), *Streptococcus* spp. appears to have been underrepresented with reduced relative abundance of each of the streptococcal species included in the mock community listed in Chapter 2 Section 2.7. However, as the batch effect is small and, with the exception of the 10 samples listed in Chapter 3 Section 3.2.3, all samples from the same patient were sequenced on the same plate and thus equally affected by batch effect, this was not adjusted for in downstream analyses.

Figure 3.4: Examination of PCR negative controls (N = 2), technical, sampling controls (N = 4) and all controls combined (N = 6) for contaminant OTUs.

This illustrates the reads counts for OTUs with > 100 reads in the control samples. OTUs with > 750 reads were considered to be contaminants and removed. These included: *Burkholderia* (OTU ID 1606), *Undibacterium* (OTU ID 1727) and *Ralstonia* (OTU ID 1703).

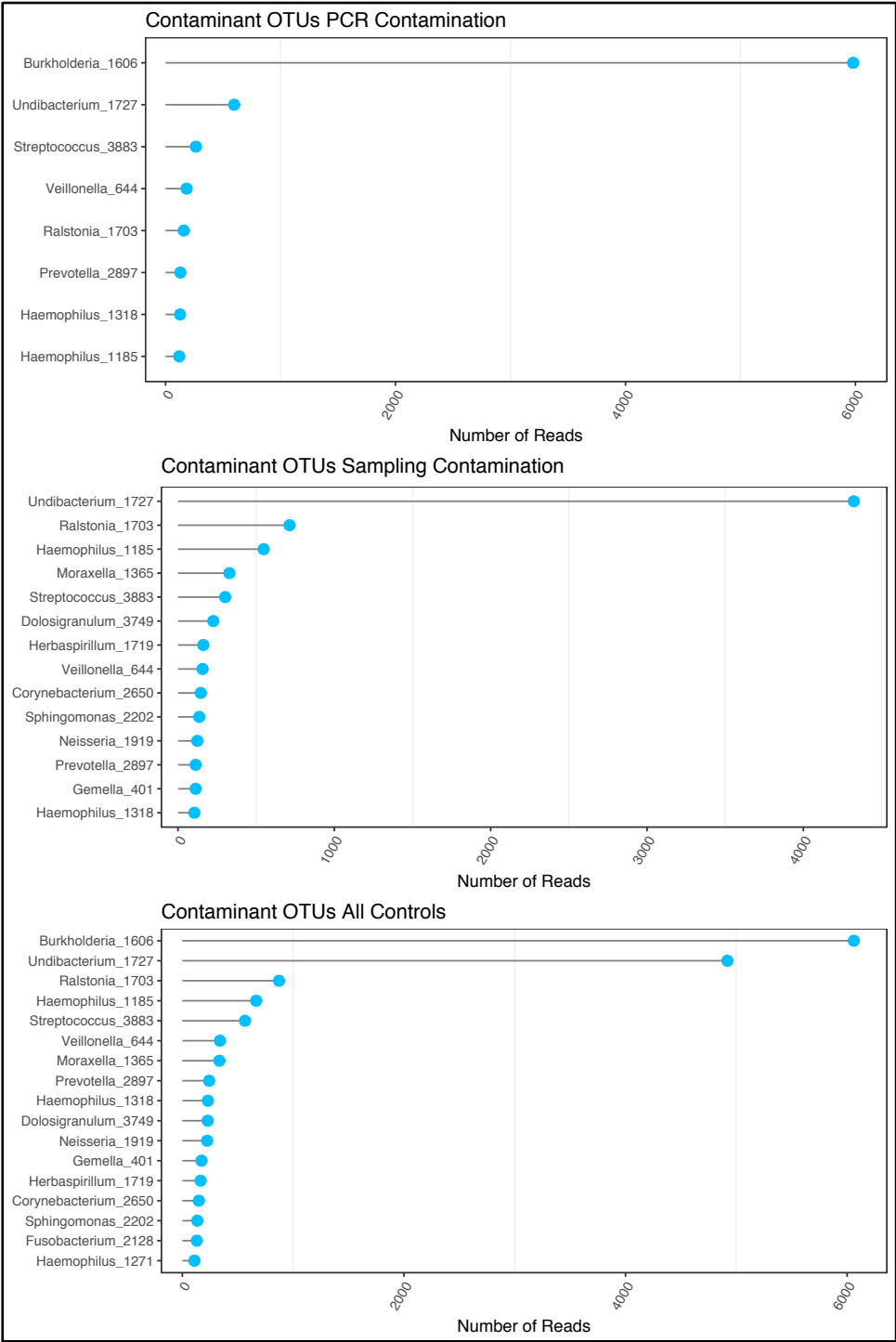


Figure 3.5: Examination of PCR negative controls, technical, sampling controls and all controls following removal of suspected contaminants *Burkholderia* (OTU ID 1606), *Undibacterium* (OTU ID 1727) and *Ralstonia* (OTU ID 1703) to confirm removal.

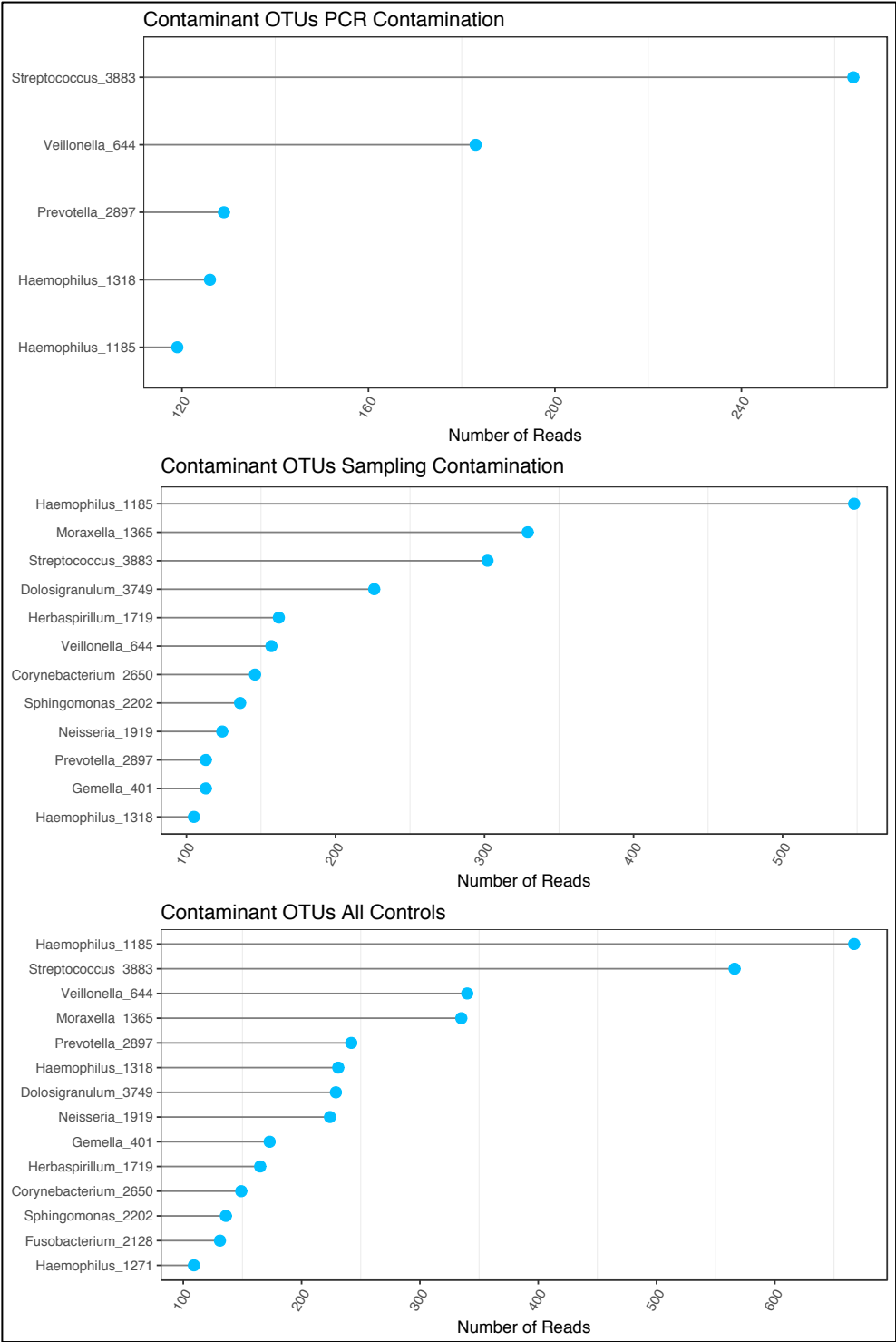


Figure 3.6: Non-metric multidimensional scaling (NMDS) plot to examine for sequencing consistency between each plate of samples.

This illustrates consistent sequencing between each plate of samples with mock communities (“mock”) clustering together and negative PCR controls (“negative”) and technical controls (“control BALF”) being more dispersed.

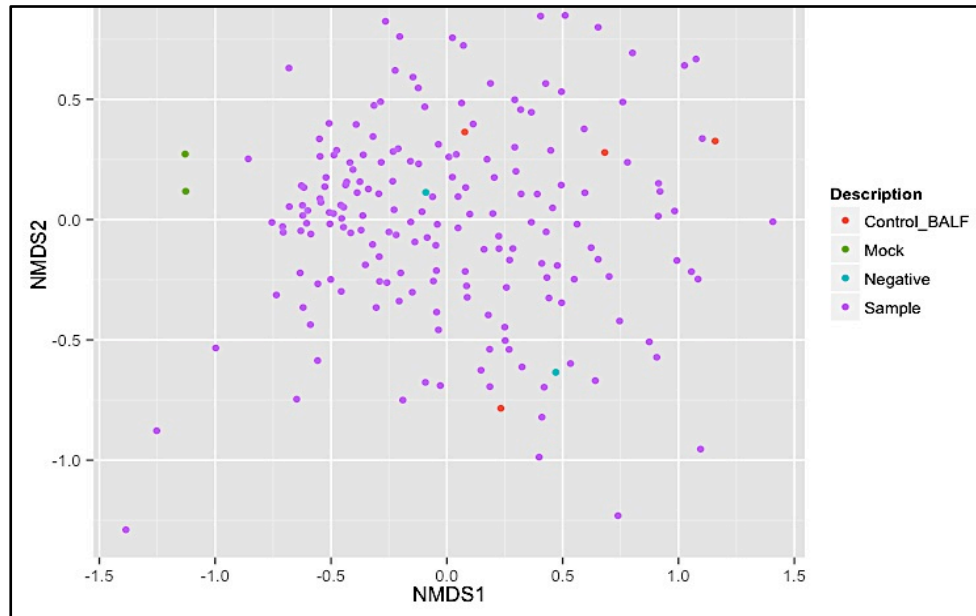


Figure 3.7: NMDS plot to examine for batch effect.

Samples coloured by sequencing plate number (plate 1 shown in dark blue and plate 2 shown in light blue). This demonstrates a small batch effect, with separation in clustering between samples sequenced on plate 1 and plate 2.

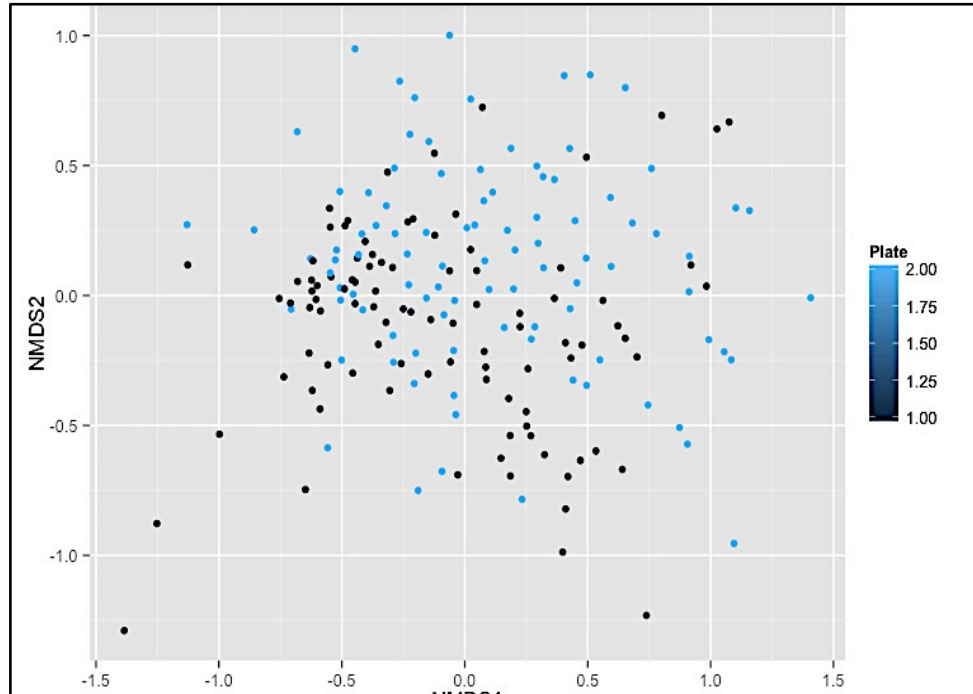
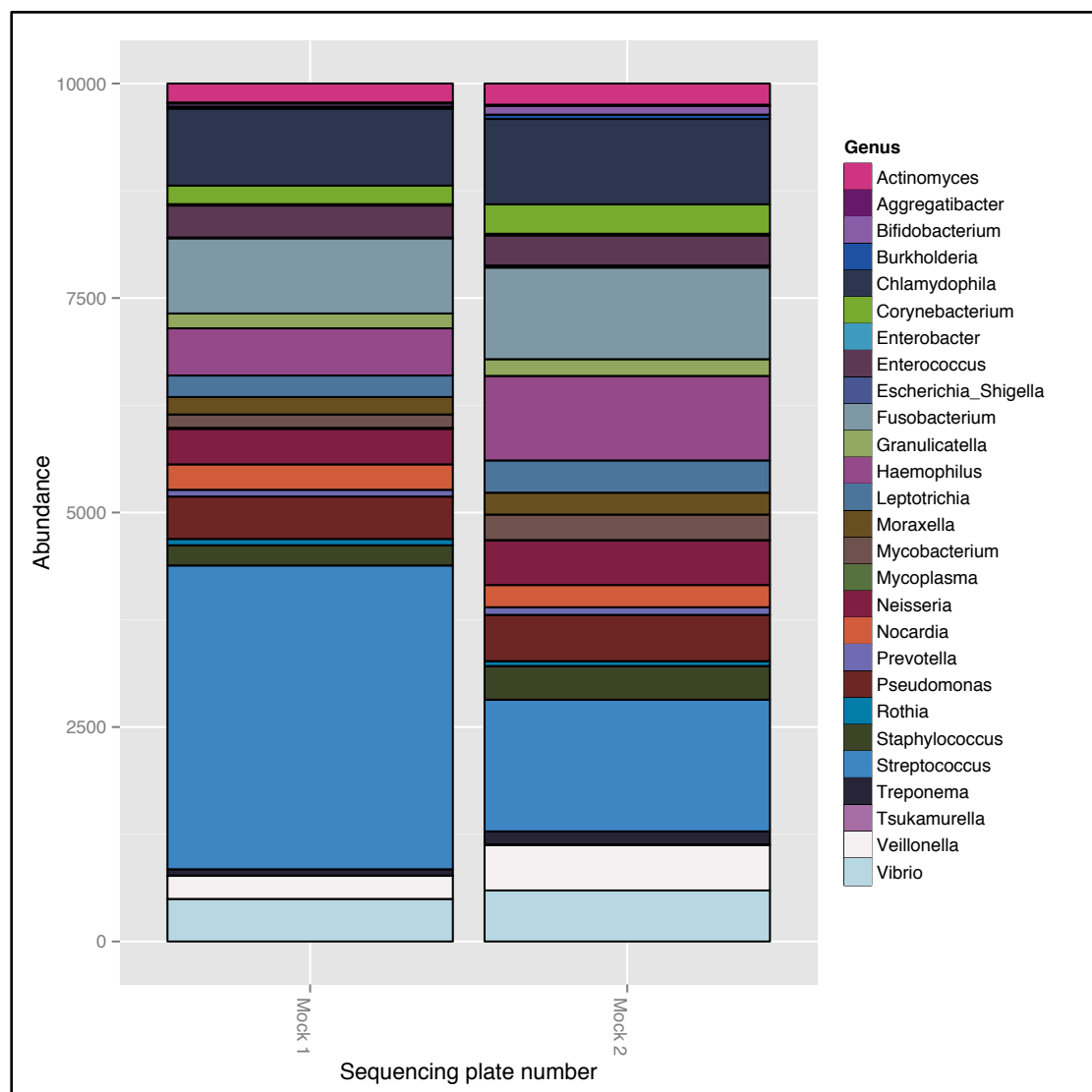


Figure 3.8: Barplot of 27 genera present in the mock communities for each sequencing plate.

This illustrates successful sequencing of the 27 genera expected in the mock community and a decreased relative abundance of *Streptococcus* spp. in Mock 2. “Mock 1” represents the mock community sequenced in sequencing run BAX1 and “Mock 2” represents the mock community sequenced in sequencing run BAX2. Full names of bacteria included in mock community are detailed in Chapter 2, Section 2.7.

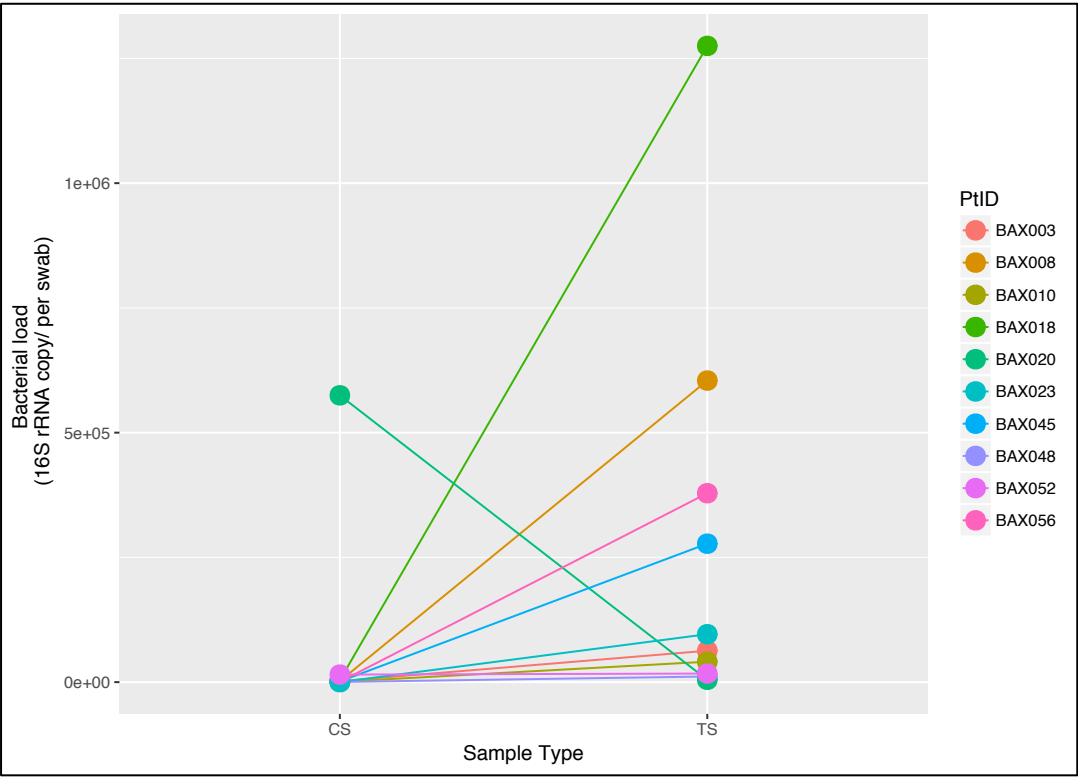


After rarefaction 34/42 (81%) of the BALFs, 36/42 (86%) bronchial brushings, 44/47 (94%) TS and 17/37 (46%) CS remained. Due to the large number of CS

removed at rarefaction, only 8 patients remained in whom all 4 sample types (TS, CS, BALF and bronchial brushing) were available for comparison. In contrast after rarefaction 39 patients had at least a TS and a paired lower airway sample remaining with 17 having a CS and paired lower airway sample (Figure 3.4). In ten patients, qPCR was performed comparing bacterial load (16S rRNA copy number per swab) between paired CS and TS. Bacterial load was significantly lower in CS than TS (median in CS = 1657.5 (range 68 - 574, 733), median in TS = 79,480 (range 4660 - 1,275,133), $W = 8$, $P = 0.049$) (Figure 3.9).

Figure 3.9: Comparing bacterial load (16S rRNA copy number per swab) between paired CS and TS.

This illustrates showing significantly lower biomass in CS ($N = 10$, $P = 0.049$).



3.3.3 Comparison of the upper and lower airway microbiota

For BALF and bronchial brushings, the three most common genera were identical being: *Haemophilus* spp. (23.6% and 24.6), *Streptococcus* spp. (20.3% and 20.5% of total reads respectively), and *Prevotella* spp. (6.5% and 7.9%). Similar genera were seen in TS but with *Streptococcus* spp. the most common (39.5%) followed by *Haemophilus* spp. (15.4%) and *Prevotella* spp. (8.7%). At an OTU level, there was no difference observed between upper and lower airway samples (Figure 3.10).

Alpha diversity was compared between paired BALF and bronchial brushings (N = 26). Overall, no difference was found in richness ($t_{(25)} = 0.563$, $P = 0.578$), evenness ($W = 244$, $P = 0.084$) or by the Shannon Diversity Index ($W = 230$, $P = 0.173$). In addition, there was no significant difference between BALF and bronchial brushings in relation to beta diversity using the Bray Curtis dissimilarity index ($P = 0.650$, $r^2 = 0.005$), unweighted UniFrac ($P = 0.114$, $r^2 = 0.02$) or the weighted UniFrac scores ($P = 0.255$, $r^2 = 0.008$). Therefore, in instances where both BALF and bronchial brushings were available for comparison with TS, only bronchial brushings were compared with TS.

Individual patient barplots demonstrated that the airway microbiota is highly individual, with similarities usually observed between TS and lower airway samples (Figure 3.11 and Appendix A3). *Streptococcus* spp., however, appear to have a higher relative abundance in TS than lower airway samples. Where large differences were present, dominance of an individual organism appears to lead to an uneven community. There was a trend for dominant organism in BALF to

be the same bacterium isolated by routine clinical culture of the same fluid (Appendix A3) As there is no standardised Bray Curtis dissimilarity threshold at which samples are considered to be highly dissimilar, from examining individual patient barplots, greater dissimilarity was seen between TS and lower airway samples with a score of ≥ 0.7 . There were twelve patients with Bray Curtis scores above this threshold. Of these, five had positive growths on culture of BALF specimens (excluding growths of “upper respiratory tract flora only”, N = 3). Organisms cultured were *Staphylococcus aureus* (N = 2), *Haemophilus influenzae* (N = 2) and *Serratia marcescens* (N = 1); the dominant genus in their corresponding sequenced sample was identical, namely *Staphylococcus* spp.; *Haemophilus* spp. and *Serratia* spp. respectively. Four patients had no growth on culture of BALF, of which three had a dominant organism on sequencing (*Streptococcus* spp., *Neisseria* spp. and *Moraxella* spp.).

Figure 3.10: Heatmap of the 50 most common OTUs present between upper (TS and CS) and lower airway samples (BALF and bronchial brushings).

Each square in the grid represents the relative abundance (using rarefied data) for an OTU for a given sample. This illustrates similarities in the relative abundances of OTUs between upper and lower airway samples.

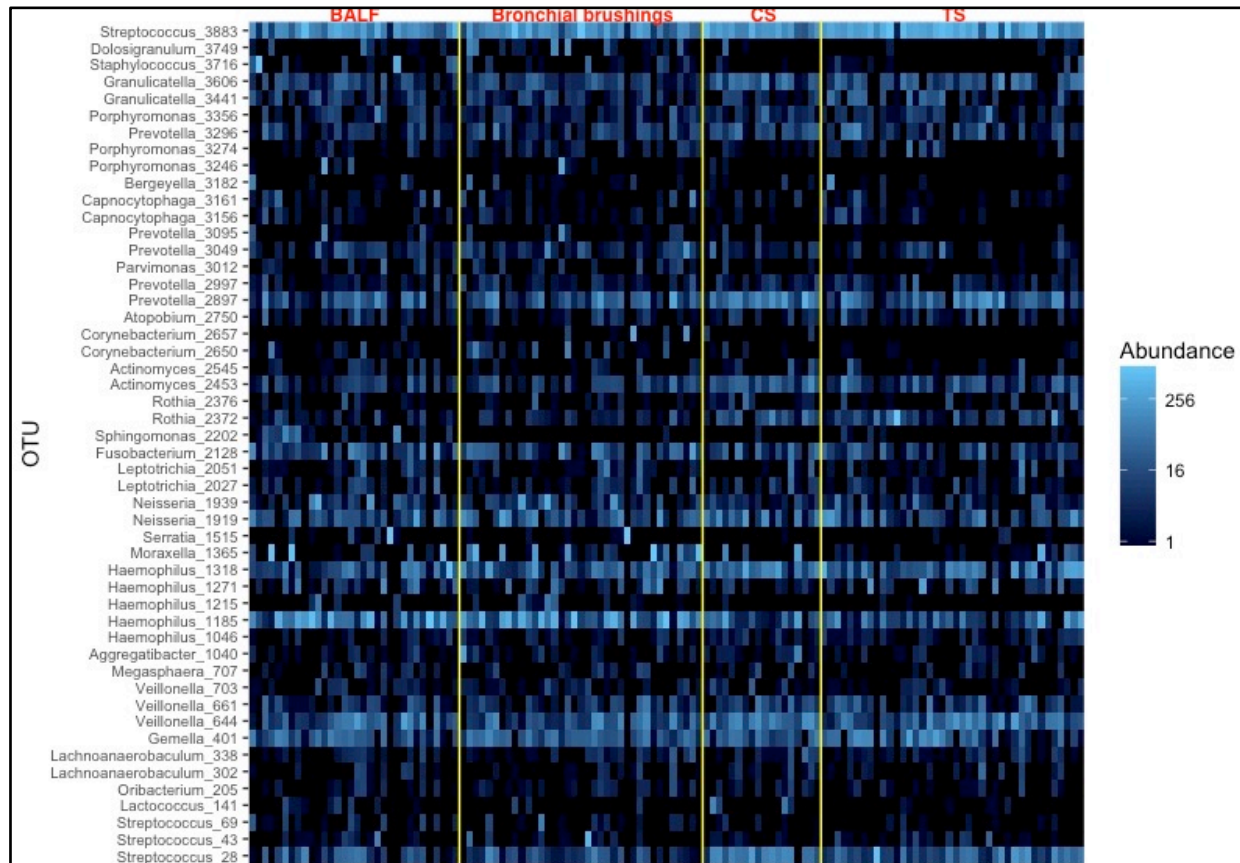
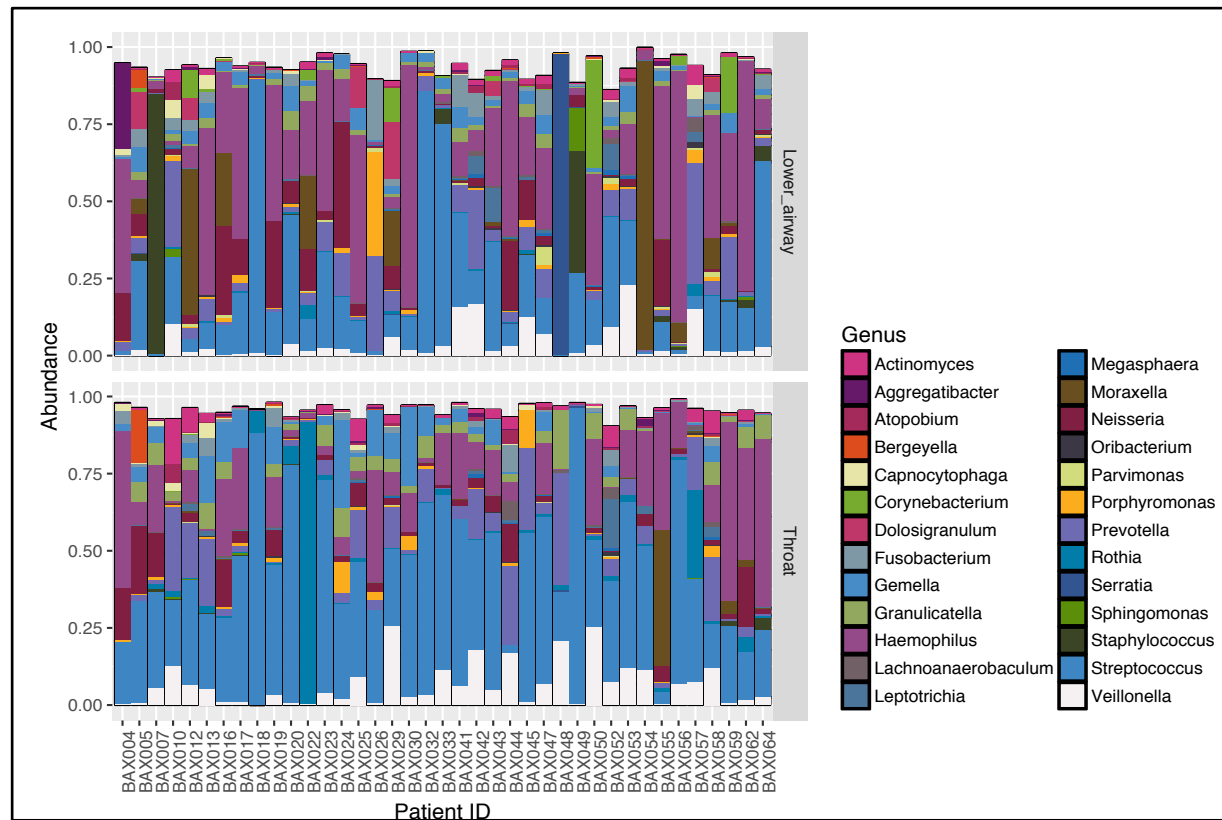


Figure 3.11: Individual patient barplots (N = 39) comparing paired lower airway samples and TS for the 50 most common OTUs.

The relative abundance of OTUs for each sample is shown with TS below and their corresponding lower airway sample above. Bars are of uneven heights due to the presence of low abundance “other” OTUs which have not been included in the plot. This illustrates that the airway microbiota is highly individual with variable degrees of similarity between TS and lower airway samples.



Comparing TS and lower airway samples (either BALF or bronchial brushings) (N = 39), no significant difference was found in richness, evenness or by the Shannon Diversity Index ($P > 0.05$, Figure 3.12). Bland-Altman plots of alpha diversity measurements showed somewhat less agreement between TS and lower airway samples with low evenness (Figure 3.13), again suggesting that there is less agreement between TS and lower airway samples in the presence of a dominant organism. The mean Shannon Diversity Index was 2.09 (standard deviation (SD) 0.64) for TS and 2.12 (SD 0.78) for lower airway samples. Using these observed mean Shannon Diversity values, a post-hoc power calculation was performed. This showed that a sample size of 44 would have 90% power to detect a difference in means of 0.3 assuming a standard deviation of differences of 0.6, using a paired t-test with a 0.05 two-sided significance level. A small significant difference ($P = 0.001$) was found in beta-diversity between TS and lower airway samples using the Bray Curtis, unweighted and weighted UniFrac scores ($r^2 = 0.06, 0.03$ and 0.06 respectively) (NMDS plot showing unweighted UniFrac comparison Figure 3.14).

Figure 3.12: Alpha-diversity comparisons between lower airway samples and TS.

No significant difference was seen in: (a) richness ($t_{(38)} = 1.6523$, $P = 0.107$, mean difference = 5.38, 95% Confidence Interval (CI) [-1.21 – 12.0]); (b) evenness ($W_{(38)} = 367$, $P = 0.756$, median difference -0.012, 95% CI [-0.070 – 0.051]), and (c) Shannon diversity index ($W_{(36)} = 384$, $P = 0.940$, median difference = -0.013, 95% CI [-0.263 – 0.276]). Different colours represent individual patients with lines connecting paired lower airway samples and TS.

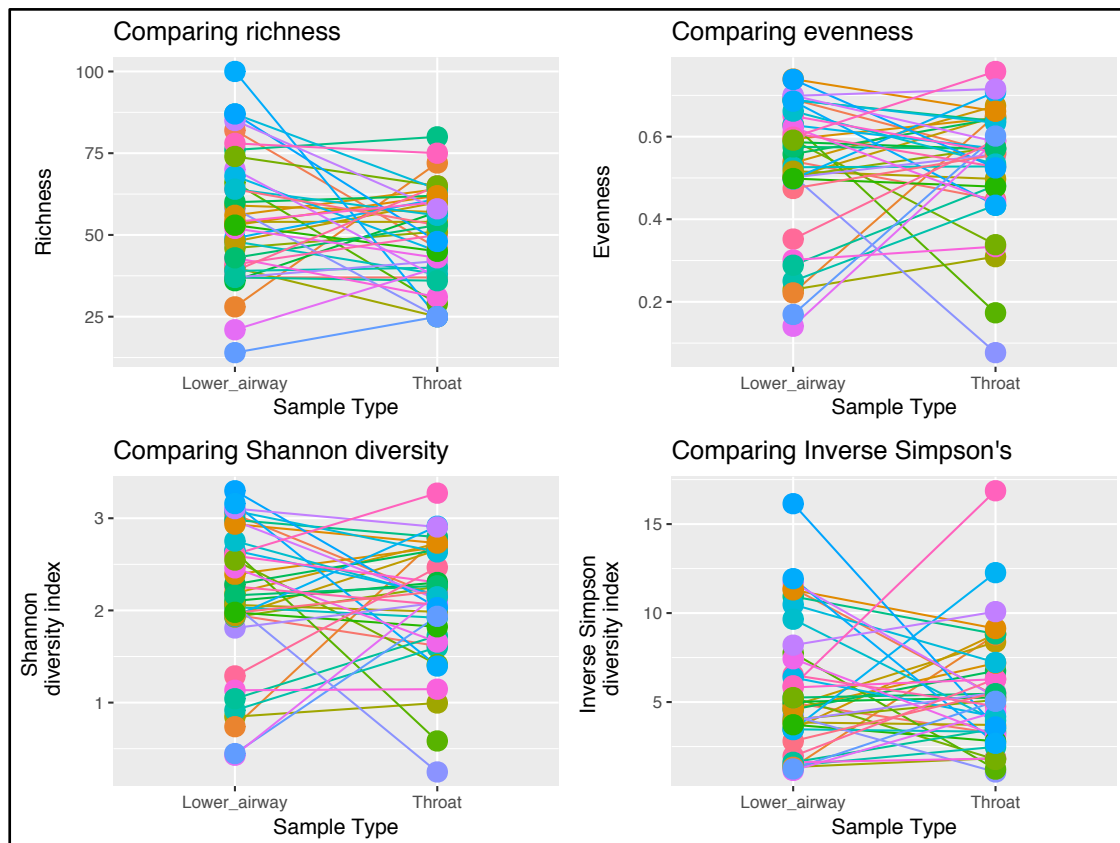


Figure 3.13: Bland Altman plots showing agreement between TS and lower airway samples in alpha diversity measurements.

Alpha diversity measurements illustrated are: (a) richness, (b) evenness and (c) Shannon Diversity Index. Agreement between TS and lower airway samples is shown by comparing the mean against the absolute difference in alpha diversity measurements between TS and lower airway samples. The blue lines indicate the mean and 95% limits of agreement. The majority of samples lie between these limits, demonstrating overall agreement, apart from at low levels of evenness and Shannon Diversity.

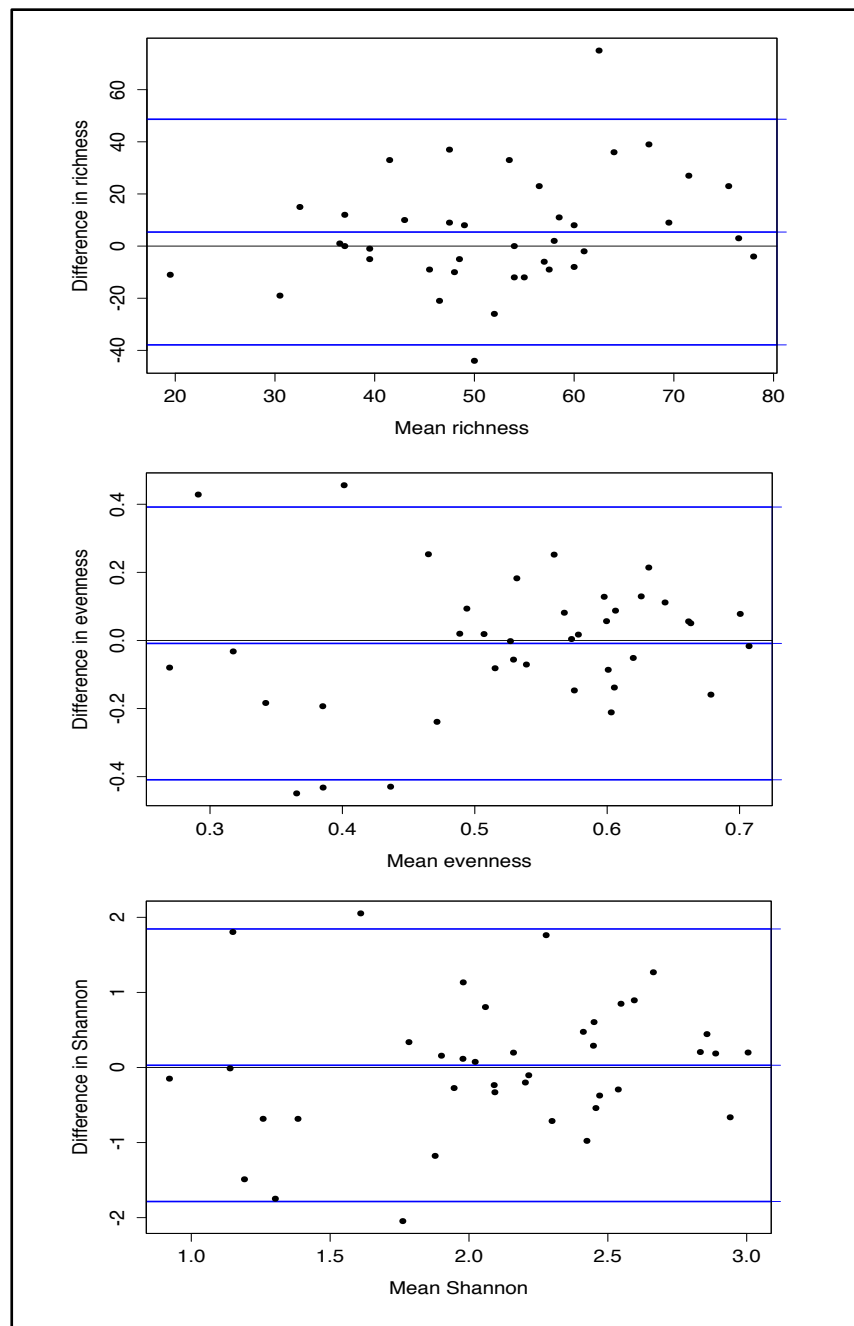
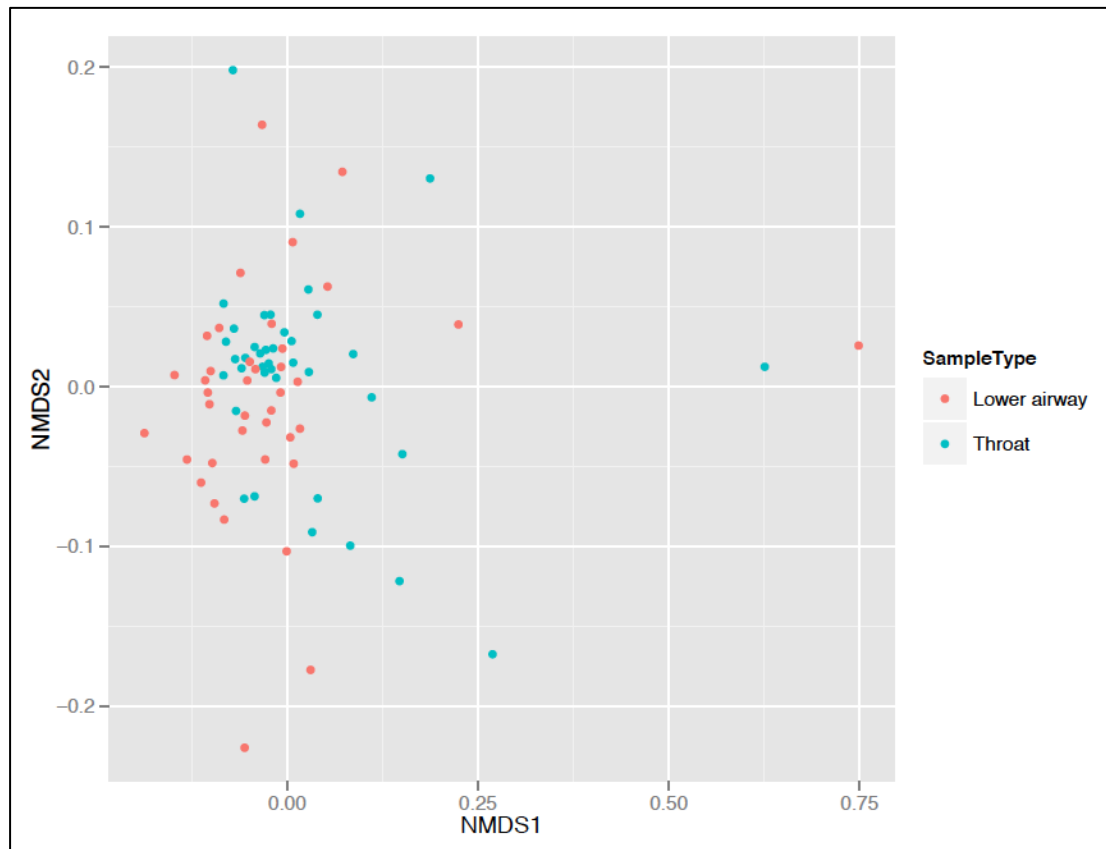


Figure 3.14: Non-metric multidimensional scaling (NMDS) plot comparing the unweighted UniFrac score between TS and lower airway samples.

This shows similarity in the clustering pattern between TS and lower airway samples.



After filtering to remove OTUs with less than a total of 20 reads, considering genera 80% (72 out of 90) were in common being present both in lower airway samples and TS, whilst 15.6% (14 out of 90) were unique to BALF and 4.4% (4 out of 90) were unique to TS. Those genera which were different between upper and lower airway samples were present in very low relative abundance (< 0.1%).

Spearman's rank correlation testing showed good correlation between the relative abundances of genera and OTUs present when comparing lower airway

samples with TS (genera level: $r = 0.776$, $P < 0.001$; OTU level: $r = 0.557$, $P < 0.001$). Using multiple paired t-tests with Bonferroni correction to compare the relative abundance of genera between TS and lower airway samples, a significantly higher relative abundance in *Streptococcus* spp. was seen ($t = -182$, $P < 0.0001$, $P_{\text{adj}} = 0.0009$) in TS. No other genus was significantly different between TS and lower airway samples ($P_{\text{adj}} > 0.05$) (Table 3.3).

Table 3.3. Comparing the mean relative abundance (percent) of the most common or clinically important genera between TS and paired lower airway samples (BALF or bronchial brushing).

Using multiple paired t-tests with Benjamini Hochberg correction, only *Streptococcus* spp. was significantly different ($P_{\text{adj}} = 0.0009$). TS – throat swab; NS - non-significant ($P_{\text{adj}} > 0.05$). SD - standard deviation.

Genus	Lower airway mean (SD)	TS mean (SD)	P_{adj}-value
<i>Haemophilus</i> spp.	25.2 (24.4)	15.6 (16.2)	NS
<i>Streptococcus</i> spp.	21.5 (22.2)	39.7 (22.8)	0.0009
<i>Prevotella</i> spp.	7.7 (9.9)	8.3 (10.0)	NS
<i>Neisseria</i> spp.	6.9 (10.1)	4.6 (6.3)	NS
<i>Moraxella</i> spp.	6.2 (17.4)	1.4 (7.0)	NS
<i>Veillonella</i> spp.	4.2 (5.6)	6.8 (7.1)	NS
<i>Staphylococcus</i> spp.	4.1 (16.3)	0.2 (0.7)	NS
<i>Gemella</i> spp.	3.0 (2.6)	5.3 (7.3)	NS
<i>Fusobacterium</i> spp.	2.6 (3.8)	1.3 (2.0)	NS
<i>Serratia</i> spp.	2.6 (15.9)	0.05 (0.1)	NS
<i>Corynebacterium</i> spp.	2.2 (6.6)	0.03 (0.1)	NS
<i>Granulicatella</i> spp.	1.7 (1.7)	4.2 (3.9)	NS
<i>Dolosigranulum</i> spp.	1.6 (4.2)	0.03 (0.1)	NS
<i>Porphyromonas</i> spp.	1.6 (5.4)	1.2 (2.5)	NS
<i>Leptotrichia</i> spp.	1.1 (2.7)	1.0 (2.6)	NS
<i>Rothia</i> spp.	0.5 (1.0)	4.5 (15.6)	NS
<i>Burkholderia</i> spp.	0.1 (0.1)	0.00 (0.02)	NS
<i>Pseudomonas</i> spp.	0.04 (0.1)	0.04 (0.1)	NS
<i>Ralstonia</i> spp.	0.04 (0.1)	0.01 (0.03)	NS

Although CS sequenced poorly, when successfully sequenced they also showed similarities in diversity with lower airway samples. The three most common genera in CS were similar to those seen in BALF, bronchial brushings and TS with *Streptococcus* spp. the most common (29.6% of reads), followed by *Haemophilus* spp. (13.6%) and *Prevotella* spp. (13.5%)(Figure 3.10).

Comparing CS with lower airway samples (N = 17), no significant difference was seen in richness ($t_{(16)} = -0.831$, $P = 0.418$), evenness ($W = 40$, $P = 0.089$) or the Shannon diversity index ($W = 48$, $P = 0.190$). A significant difference was seen in the Bray Curtis ($r^2 = 0.08$, $P = 0.002$) and weighted UniFrac scores ($r^2 = 0.06$, $P = 0.029$), but not in the unweighted UniFrac score ($r^2 = 0.03$, $P = 0.2$). Considering genera, 70.4% (57 out of 81) were common to both lower airway samples and CS whilst 19.6% (16 out of 81) were unique to lower airway samples and 10.0% (8 out of 81) were unique to CS (Figure 3.15). Spearman's rank correlation testing showed good correlation between the relative abundances of OTUs present when comparing lower airway samples with CS (genera level: $\rho = 0.795$, $P < 0.001$; species level: $\rho = 0.579$, $P < 0.001$).

3.3.4 Clinical variables and the upper and lower airway microbiota

As highlighted in the prior section, there were variable degrees of similarity between the upper and lower airway samples (Figure 3.11). To determine whether specific clinical features underpinned these differences, beta-diversity testing was performed using a PERMANOVA which included each of the variables listed in Table 3.1 as well as "Patient ID" (i.e. which patient was sampled) and

“Sample type” (11 variables included in the PERMANOVA in total). The influence of FOB route and underlying pathology were tested in separate PERMANOVAs using FOB samples or TS only respectively.

Two variables were found to exert a large influence on community structure: “Patient ID” had the largest influence accounting for 47.0 – 53.8% of the variance ($P = 0.001$), whilst the underlying pathology accounted for 18.7% – 22.5% ($P \leq 0.006$) (Table 3.4). All other variables, including “Sample Type”, had more minor influence. FOB route accounted for 5.43% - 8.16% of the variance ($P < 0.05$), with fewer *Corynebacterium* spp. (OTU 2657), *Moraxella* spp. (OTU 1365) and *Dolosigranulum* spp. (OTU 349) present in FOB performed nasally than orally. Other significant variables include the use of prophylactic antibiotics and nebulized antibiotics. Patient age was not a significant influence on community structure.

Table 3.4. Beta diversity summaries of significant clinical variables influencing community structure.

Adonis (PERMANOVA) results shown are for those variables which were statistically significant ($P < 0.05$) using the Bray Curtis dissimilarity score.

IV - intravenous. FOB - fiberoptic bronchoscopy.

Variable & diversity measurement	r²	P-value
Sample Type	0.06	0.001
Patient ID	0.54	0.001
Underlying pathology	0.20	0.001
Lower respiratory tract symptoms	0.02	0.003
Nebulised antibiotics	0.02	0.003
Current IV antibiotics	0.01	0.012
FOB route	0.08	0.001

3.4 Discussion

The aim of this chapter was to assess whether upper airway samples (CS and TS) could act as surrogates for characterising the lower airway microbiota and therefore be used for longitudinal sampling in children with chronic lung diseases. This was evaluated through a comparison of the airway microbiota on paired upper and lower airway samples collected from children undergoing a clinical FOB. In accord with the initial hypothesis, TS sequenced well and reflected the lower airway microbiota in many children with chronic lung diseases at a community level. When compared with lower airway samples

(either BALF or bronchial brushings), there was no significant difference in within sample alpha diversity. Whilst a significant difference in beta diversity was seen between these sample types, the degree of variation due to sample site and other clinical variables, such as antibiotic usage, was relatively small. This was consistent when testing whether similar organisms were present between samples (Bray Curtis dissimilarity) and when testing the relative abundance and phylogenetic relationships between these samples (unweighted and weighted UniFrac scores). For sample site it ranged from 2.95% (unweighted UniFrac) to 6.31% (Bray Curtis). *Streptococcus* spp. and *Haemophilus* spp. were two of the commonest genera in both TS and lower airways, with strong correlation seen in the relative abundance of genera between samples.

CS, however, sequenced poorly, with only 44% of CS sequencing well compared with 94% of TS. This is likely attributable to the limited biomass and recoverable microbial DNA from this type of sample. When successfully sequenced, whilst CS also showed similarities in diversity with lower airway samples, due to their poor yield on sequencing, these results suggest that CS should not be used to reflect the lower airway microbiota. With increasing use of molecular techniques in clinical microbiology, these results demonstrate that sampling method is an important consideration with the collection of TS advised in preference to CS for genomic based methods of bacterial detection.

By contrast, there was much greater variability between disease states (up to 22.5%) than between upper and lower airway samples. There were sufficient numbers of children for whom lower and upper airway samples were discordant

to preclude the use of TS for clinical decision making in individuals (and testing this was never the aim of this chapter). Nevertheless TS contain valuable research information that could be used to explore the development of the lower airway microbiota in groups of children with different diseases. For example, TS could be used to determine whether the more benign clinical course of PCD compared to that of CF relates to different temporal evolution of the lower airway microbiota.

At an intra-individual level, for both CS and TS there were instances where significant differences between upper and lower airway samples were observed. At the genus level, however, only *Streptococcus* spp. was significantly different between TS and lower airway samples, with a higher relative abundance in TS. Examining individual patient barplots, this appears to be consistent across all samples and therefore not related to the sequencing plate the sample was run on.

Dominance of an OTU in either the upper or lower airway sample that was not reflected in its companion sample also led to greater dissimilarity between TS and lower airway samples. For those samples showing the greatest dissimilarity, there was a trend for the dominant organism in BALF to be the same organism grown on bacterial cultures of the same fluid. This could suggest overgrowth of an organism in an individual sample, outcompeting other organisms, or low biomass in that sample allowing artificial dominance of an organism. Therefore, where samples are dominated by an individual genus, the results should be treated with caution and repeat sampling paired with culture dependent

microbiology may be advised. Upper airway samples may therefore fail to detect dominant pathogens in the lower airways and vice versa, limiting clinical use in an individual child; this should also be remembered in research studies.

Two main factors influenced the degree of variation in the airway microbiota. The patient sampled had the greatest influence (53.8% of variation, Bray Curtis dissimilarity score). This highlights how individual the microbiota is to each patient which is well-recognized^{54,147,239}. The underlying pathology accounted for up to 22.5% of variation, assessed using TS only. This suggests that disease differences can be detected utilizing TS and that TS may be useful in longitudinal studies of the airway microbiota comparing different patient groups. All other factors had a much more minor influence. Further work comparing TS with a lower airway sample (e.g. bronchial brushings alone) in larger numbers of patients from each disease group is needed to explore potential factors influencing the relationship between the upper and lower airway microbiota.

These findings in 49 children are similar to those of Charlson *et al.*³², who demonstrated similarity of the microbiota along the respiratory tree in 6 adults (including smokers), and the study of Boutin *et al.*²⁴⁰ who demonstrated similarities between TS and sputum samples in 20 adults and children with CF with a mean age of 16.1 years. Similarly, Marsh *et al.*¹⁵¹ compared upper airway samples (nasopharyngeal and oropharyngeal swabs) with BALF sampled from a single lobe and found upper airway samples were a reliable surrogate in 69% of children with either idiopathic bronchiectasis, protracted bacterial bronchitis or healthy airways. To my knowledge this is the first study comparing the

microbiota using both TS and CS and lower airways samples in young children with CSLD (CF or PCD) and non-CSLD controls (mainly recurrent LRTI). This is important since such children are less able to spontaneously expectorate. They are therefore the group in whom a reliable surrogate for lower airway sampling is particularly pertinent.

The current study does have limitations. It was not possible to perform an *a priori* power calculation so the study may be underpowered. Doing a *post-hoc* power calculation a sample size of 44 would have 90% power to detect a difference in the mean Shannon Diversity Index between TS and lower airway samples with a 0.05 two-sided significance level. The clinical relevance of a difference of 0.3 in Shannon diversity is unknown, however, making these results difficult to interpret. A significant difference in beta-diversity was demonstrated however, suggesting the study can detect differences between at least groups of patient samples. Nonetheless caution should be taken when interpreting the CS data in this study given the low sample size (N = 17) and their very low biomass, rendering them at increased risk of contamination by spurious OTUs.

The sequencing depth varied greatly between samples, ranging from 118 to 192,812 reads. To account for this, similar to other microbiota studies samples were rarefied to 1,000 reads to balance between capturing OTUs and retaining sufficient samples for paired comparisons²⁴¹.

Eighteen patients (37%) had received a course of antibiotics in the previous 30 days and 41% were prescribed prophylactic antibiotics. Whilst this could

potentially introduce a bias in our results by influencing both the upper and lower airway microbiota so that they show greater similarity, it would not have been ethically permissible to stop a clinically indicated treatment for the purposes of this study. In addition, 12% of patients were prescribed nebulised antibiotics inhaled through a mouthpiece at least in older children, the deposition of which may be thus different between the upper and lower airways, potentially driving greater differences. Due to the small numbers of patients who received nebulised antibiotics, however, it was not possible to divide these patients into subgroups to perform meaningful comparisons.

Ideally, all bronchoscopies would have been performed either via an LMA or endotracheal tube in order to minimize risk of contamination of the bronchoscope. The route however had to be determined by clinical considerations. Bronchoscopy route accounted for up to 8.16% variance in community structure, much less than the variance due to the individual patient (53.8%) or the underlying disease (22.5%). Greater abundance of organisms associated with the nasal passages such as *Corynebacterium* spp. was seen in samples where bronchoscopy was performed nasally. This may represent microaspiration of organisms from the upper airways or transfer of organisms from the nasal passages by the bronchoscope. Nasopharyngeal samples were not collected and have previously been found to be highly diverse in young children²⁴. Their inclusion in future studies would allow the relationship between upper and lower airway samples to be explored in greater detail.

Furthermore, BALF was only collected from two lung lobes (right middle lobe and either lingual or most diseased lobe) in this study. Given that lobar differences exist in bacterial distribution²³⁶, ideally all six lobes would have been sampled and reproducibility tested within each lobe sampled. The methods used in this study were in accordance with the European Respiratory Society BAL Task Force²³⁵ and to attempt to limit for lobar differences, samples from different lobes were pooled prior to processing. Further work sampling each lobe may help to ascertain whether differences between upper and lower airway samples may be due to lobar differences within the lung.

In summary, CS sequenced poorly and whilst those successfully sequenced were similar to lower airway samples, their use cannot be recommended for non-culture based microbiota studies. For TS, there was good correlation between the upper and lower airway microbiota in many children with chronic lung diseases at a community level and disease differences could be discerned on TS. This suggests that TS may be useful to study group differences in children with, for example, PCD and CF although larger studies comparing TS and lower airway samples within different disease groups are required to explore this relationship fully. At an intra-individual level, whilst several patients showed similarities between TS and lower airway samples, large differences existed for some patients. This was particularly true if a dominant organism was present in TS that was not present in its corresponding lower airway sample and vice versa. In some patients upper airway samples failed to detect a dominant lower airway pathogen which lower airway bacterial cultures may identify, also suggesting interpretation of TS in an individual in a clinical context should be cautious.

Further work is needed to determine in which individual patients TS are likely to be reliable lower airway surrogates and those in whom the combination of culture dependent and molecular microbiology using TS would be important. Overall, in accord with the initial hypothesis, TS but not CS are a useful surrogate in longitudinal studies of the lower airway microbiota for different disease groups. TS were therefore the chosen sample type in the longitudinal studies (Chapters 4 and 5) of the airway microbiota in children with CF diagnosed on NBS the comparison of CF and PCD, and the effects of pulmonary exacerbations and intravenous antibiotics.

Chapter 4: Initial development of the airway microbiota in infants with cystic fibrosis

4.1 Introduction

The role of the microbiota in early CF disease progression is unclear. In the UK, since 2007 the majority of children with CF have been diagnosed through NBS at a median age of 3 weeks⁸³ thus providing an opportunity to study the early changes in the microbiota. This chapter describes the analysis of the airway microbiota in infants diagnosed with CF on NBS recruited as part of the CLIMB study described in Chapter 2, section 2.2. The aim was to better describe how the airway microbiota develops over the first two years of life and the role of community diversity in infants diagnosed with CF on NBS, in particular prior to important clinical events, such as the first exacerbation requiring admission for IV antibiotics and first growth of *P. aeruginosa*. To assess this, a prospective cohort of infants with CF were recruited and TS collected at regular intervals for up to 2 years of age to compare changes in molecular microbiology with age, clinical status and culture-based microbiology. In addition, the results include a cross-sectional subgroup analysis of the lower airway microbiota in infants undergoing a routine clinical FOB between 3-5 months of age which were collected as part of the comparison of the upper and lower airway microbiota reported in Chapter 3. It is the first study to report study of the infant microbiota in CF using the Illumina MiSeq, which allows greater depth of sequencing with a lower error rate than with 454 pyrosequencing. As such, use of the MiSeq may be more informative in identifying changes in the microbiota.

4.2 Methods

4.2.1 NBS study patient recruitment and follow-up

Thirty patients were recruited from all those referred to RBH with a new diagnosis of CF on NBS as part of the CLIMB study. Full details of the CLIMB study are given in Chapter 2, Section 2.2. In the UK, all children are offered newborn bloodspot screening for CF, as well as sickle cell disease, congenital hypothyroidism and six inherited metabolic diseases at 5 days of age. If a diagnosis of CF is confirmed on NBS, infants and their families have a two-day educational visit, during which prophylactic antibiotics and multidisciplinary team management is commenced. Patients were recruited at their first scheduled clinical appointment following their educational visit. Based on previous RBH CSLD Department records, on average 20 patients per year are referred to RBH with a new diagnosis of CF on NBS. Therefore, the aim was to recruit as many of these as possible in an 18-month period. During this period 35 infants were diagnosed with CF of which 30 (86%) consented to participate in the study. Not wishing for TS to be collected was the most commonly reported reason by parents for not consenting for their child to participate in the study (N = 3).

The majority of infants were followed up at RBH every 2-3 weeks initially and subsequently every 2-3 months, with the exception of those opting to be managed predominantly at their local hospital. During the study period, infants

with CF underwent a routine, clinical FOB at approximately 3 months of age. No infant had a second FOB during the study period.

4.2.2 Sample collection & processing

The majority of TS, 97% (233 out of 241) were collected by myself with the remainder (8 swabs) collected either by a Paediatric Respiratory registrar (N = 7) or by the patient's parent (N = 1). TS and FOB samples were collected and processed as previously described in Chapter 2, Section 2.4 and Chapter 3, Section 3.2.2. Sample collection deviated from the SOP for 5 samples (1.7% of samples). Reasons for SOP deviation included: Oropharynx only swabbed 3 times due to patient distress (N = 3); sample collected by parent into incorrect tube (N = 1), and sample accidentally left at room temperature for 48 hours (N = 1). Samples which deviated from the SOP were discarded from analyses.

4.2.3 Data analysis

In addition to the analysis pipeline described in Chapter 2 Section 2.13, the following analyses were performed for the NBS cohort. Sankey plots were constructed using the riverplot package in R to illustrate the proportion (%) of each organism by phylogenetic rank within samples. Individual patient plots, illustrating barplots of the most common genera against changes in bacterial load, the Inverse Simpson's diversity index, culture results and presence or absence of lower airway symptoms were constructed using ggplot2²²⁵ and gridExtra²²⁶.

Kendall correlation testing was used to assess the relationship between age (in days) and changes in bacterial load and alpha diversity. In addition, mixed effects modelling was performed to assess the relationship between bacterial load or alpha diversity with age (grouped into 2 month age ranges until 1 year of age and 3 month age ranges between 1-2 years of age, reflecting more frequent routine follow up under 1 year of age) or clinical variables (such as growth of *P. aeruginosa*), using the lme4 (Linear Mixed-Effects Models using Eigen and S4, version 0.14.3)²⁴² and nmle (Linear and Nonlinear Mixed Effects Models, version 3.1-131)²⁴³ packages. Mixed effects models contain both fixed and random effects and are useful for exploring relationships between variables, particularly for studies with repeated measures, longitudinal design or instances where there may be missing data. They are therefore more applicable for analysing changes in diversity and genera level changes than a repeated measures ANOVA, which is less robust for handling missing data. Variance and *P* values were calculated using the MuMIn (Multi Modal Inference) package (version 1.15.6)²⁴⁴.

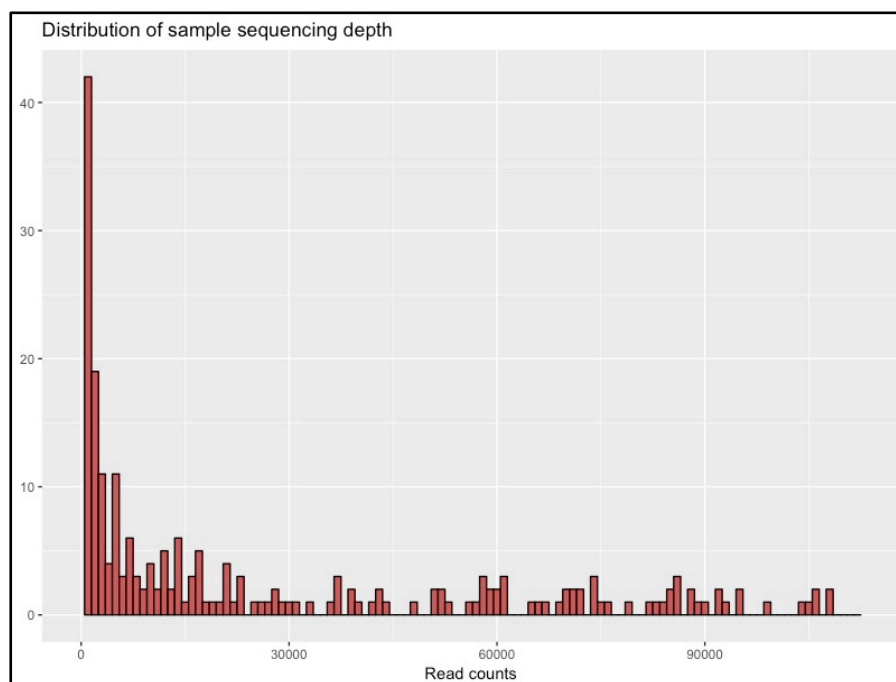
A blocked design was used when performing PERMANOVA²³¹ to compare beta diversity between paired samples longitudinally. Analyses were blocked either by participant study number (to analyse the influence of age on beta diversity) or by age range (to analyse the influence of clinical variables on beta diversity). Multiple correlation testing using Spearman's rank with a false discovery rate (FDR) correction using the multtest package and mixed effects models were used to assess genera and OTU level changes with age. A *P* value of less than 0.05 was considered to be statistically significant.

The sequencing depth (the number of sequences per sample) differed widely between samples (Figure 4.1a). For this analysis a rarefaction level of 600 reads was chosen. Examining rarefaction curves, the majority of samples had reached an asymptote at 600 reads (Figure 4.1b). At this level, 38 samples (14%) had a sequencing depth below 600 reads and were removed (Figure 4.2).

Figure 4.1: Justification for rarefaction to 600 reads.

(4.1a): Histogram of sequencing depth for samples in NBS cohort.

This illustrates the percentage of samples at a given sequencing depth (read count). Wide variation in sequencing depth is seen with the majority of samples having under 5000 reads.



(4.1b): Rarefaction curve calculated using Shannon Diversity Index measurements.

The asymptote of the rarefaction curve illustrates the sequencing threshold at which the majority of OTUs from the sample community are identified. The yellow lines depict the number of OTUs captured at 600, 1000 and 3000 reads. At 600 reads, the majority of samples have reached an asymptote.

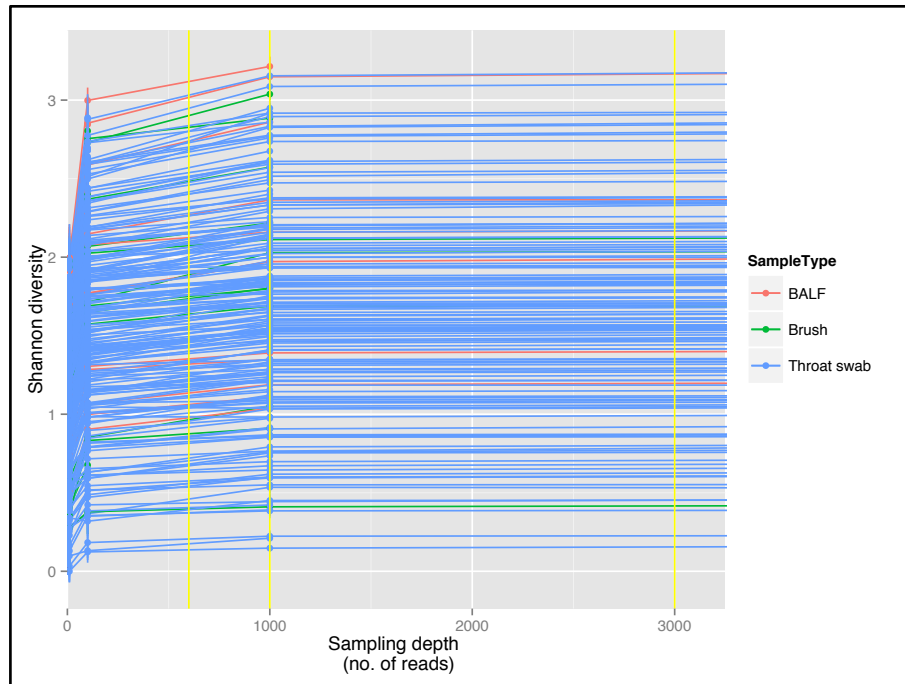
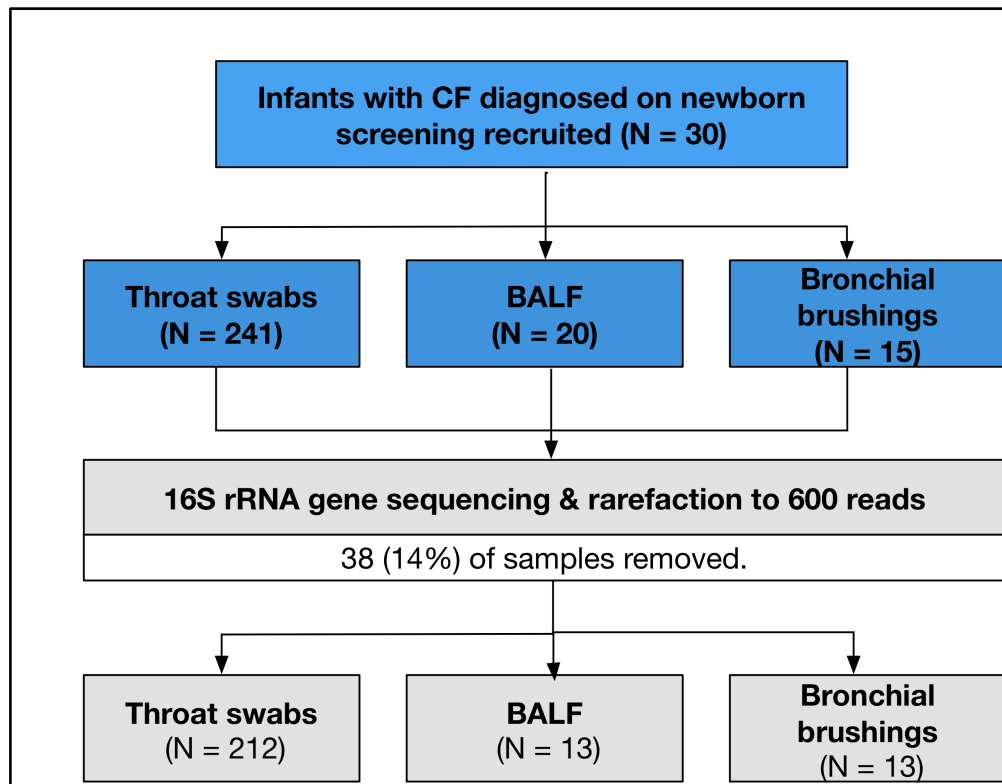


Figure 4.2: Summary of samples obtained for infants diagnosed with CF on newborn screening before and after rarefaction to 600 reads.



Random re-sampling of a population with replacement is an alternative reported method of standardisation. Using this method, species are randomly sampled from a population and diversity measures (e.g. richness and evenness) from the species sample are calculated. These species are then replaced back into the population sampled and the process repeated to a given number of iterations (e.g. 1,000 iterations). A matrix containing the mean, median, standard deviation and range of diversity scores is constructed for each sample, which can be used in statistical analyses. The advantage of this method is it does not require a minimum sequencing threshold, thus no samples are removed using this method, allowing more comparisons to be made between patients. A disadvantage is that rare species may fail to be sampled.

To ascertain whether random re-sampling with replacement (with 1,000 iterations) yields similar results to rarefaction (to 600 reads), alpha diversity results were compared between these two methods of standardisation. This showed that the results using random sampling with replacement were comparable with rarefaction. For example, a weak correlation was seen comparing changes in richness and age whether using rarefaction ($t = 0.181$, $P < 0.001$) or random re-sampling with replacement ($t = 0.171$, $P < 0.001$). Currently, however, this method cannot be used with the adonis function (PERMANOVA) in phyloseq, therefore random re-sampling with replacement was used with mixed effects modelling and rarefaction with adonis.

4.3 Results

4.3.1 Patient demographics

Thirty infants with CF were recruited, with a median age of 84 days (35 – 235 days) at recruitment. Patient characteristics are summarized in Table 4.1. Patients were followed up for a mean duration of 14 months (SD 5 months), with a median of 35 days (range 1 – 301 days) between each sample visit. Two patients (7%) were lost to follow up as they moved outside London at 9 and 11 months of age respectively. Three patients had predominantly local hospital care and therefore only 6 monthly or annual follow up at RBH.

All patients were born at term (> 37 weeks gestation) with the majority born vaginally (70%) and 53% receiving some breast milk in infancy (either

exclusively breastfed or in combination with formula feeds). Twelve patients (40%) were homozygous for pPhe508Del and seventeen (57%) patients were heterozygous for the mutation. In heterozygous patients, eleven (65%) of the non p.Phe508Del mutations were Class I – III mutations, which are associated with a more severe, pancreatic insufficient phenotype. The non p.Phe508Del mutations were: five patients with Class I mutations (p.Gly542X⁸⁴, c.1585-1G>A²⁴⁵ and c.489+1G>T²⁴⁵); four patients with Class II mutations (p.Leu206Trp²⁴⁶, p.Arg560Thr⁸⁴, and p.Gly85Glu¹⁰⁹); two patients with Class III mutations (p.Gly551Asp⁸⁴ and p.Arg352Gln²⁴⁷); three patients with Class IV mutations (p.Arg117His⁸⁴ and p.Arg334Trp²⁴⁶); one patient with a Class V mutation (c.3717+12191C>T²⁴⁵), and two patients for whom the mutation class was unknown.

The majority (73%) of patients were pancreatic insufficient. Eight patients (27%) had GORD. In accordance with UK CF Trust⁷² and RBH guidelines, all infants received prophylactic antibiotics during the study period; 25 patients (83%) had started antibiotic prophylaxis prior to recruitment (<http://www.rbht.nhs.uk/healthprofessionals/clinical-departments/cystic-fibrosis/clinical-cf-guidelines-care-of-children/>).

Seven patients (23%) had a pulmonary exacerbation requiring IV antibiotics (usually with ceftazidime and tobramycin) in the first 2 years of life at a mean age of 7 months (sd +/- 4 months). An analysis of the number of days of oral antibiotics per patient was not performed as this relied on retrospective parental reports, which many parents often struggled to recall confidently. *P. aeruginosa*

was the most commonly grown organism on bacterial cultures, with 7 (23%) patients *P. aeruginosa* positive at a mean age of 7 months (sd +/- 3 months). All growths of *P. aeruginosa* on culture were reported as "scanty growth" (equivalent to 10^3 CFU/ml). Given the prognostic significance, eradication was then attempted at RBH for any growth of *P. aeruginosa* with three weeks of oral ciprofloxacin and three months of colistin nebulisers. Six patients had a single growth of *P. aeruginosa* which was successfully eradicated with such treatment. One patient grew *P. aeruginosa* twice, which was successfully eradicated on both occasions. Three (43%) of the patients who grew *P. aeruginosa* also received IV antibiotics, one of whom received this as part of the Trial of Optimal Therapy for Pseudomonas Eradication in Cystic Fibrosis (TORPEDO-CF) study (<http://www.torpedo-cf.org.uk/>).

Table 4.1: Patient demographics for NBS infants in the CLIMB study.

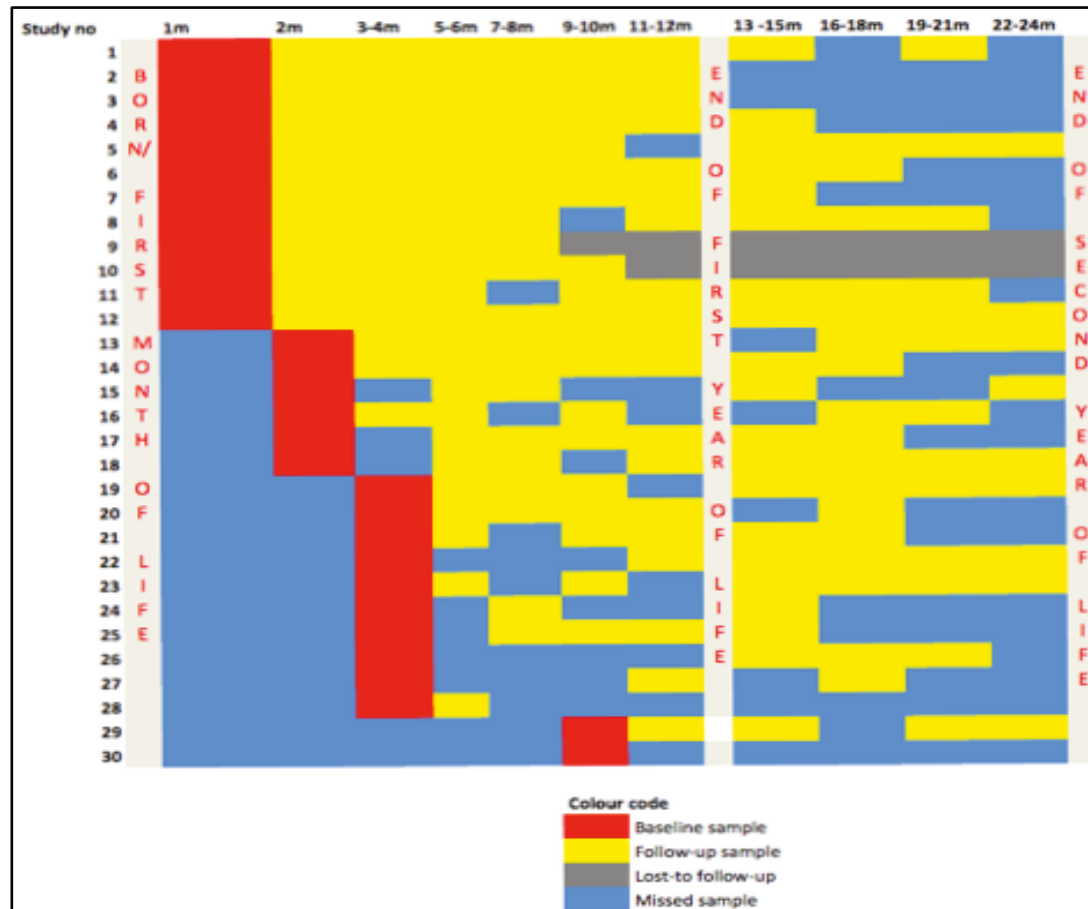
GORD - gastro-oesophageal reflux disease; SVD - spontaneous vaginal delivery;
LSCS - lower segment Caesarian section; TS - throat swab.

Demographic (N = 30)	Median (range) or N (%)
Age at recruitment (in days)	84 (35 - 235)
Gender (female)	17 (57%)
Genotype:	
Homozygous p.Phe508Del	12 (40%)
Heterozygous p.Phe508Del	17 (57%)
Other	1 (3%)
Extra-pulmonary features:	
Pancreatic insufficiency	22 (73%)
Meconium ileus	1 (3%)
GORD	8 (27%)
Birth history	
Mode of delivery:	
SVD	21 (70%)
LSCS	6 (20%)
Feeds in infancy:	
Breastfeeding (exclusively or in combination with formula)	16 (53%)
Exclusively formula fed	14 (47%)
Exacerbations & antibiotic use	
Exacerbations	7 (23%)
Prophylactic antibiotics at recruitment	25 (83%)
Bacterial growth on TS culture	
<i>P. aeruginosa</i>	7 (23%)
<i>S. aureus</i>	1 (3%)
<i>H. influenzae</i>	1 (3%)
"Upper respiratory tract flora"	1 (3%)

Sampling frequency for each individual patient is illustrated in Figure 4.3. It was not possible to collect samples for each patient at each desired timepoint due usually to patient non-attendance at routine clinical appointments but occasionally because of unavailability of the sample collector. Due to a greater frequency of scheduled clinical appointments, samples were collected more frequently in the first year of life (Section 4.2.2). Age ranges were therefore grouped into two and three month intervals during the first and second year of life respectively to account for missing timepoints and to allow the greatest number of comparisons between patients. Twelve (40%) children were successfully recruited at the first clinical appointment after their educational visit (between 5-9 weeks of age) and had multiple samples in the first two months of life. Samples collected at one and two months of age were not grouped into a two month age range as significant changes in diversity and community structure of the airway microbiota have been noted in healthy infants by two months of age, suggesting this may be a critical period in the development of the microbiota^{23,38}. Where a child had more than one sample in a given time period, the first sample taken in that range was used for comparisons between patients for analyses where age was analysed as a categorical variable; all samples at all timepoints were included for analyses where age was analysed as a continuous variable.

4.3b: Illustration of sample frequency clustered into age ranges.

Due to greater frequency of sampling in the first year of life, samples were clustered into 2-month age ranges until 1 year of life and 3 month age ranges thereafter to allow the greatest number of comparisons between patients.



4.3.2 Examining sequencing quality

Two hundred and ninety patient samples from thirty patients, with a median of ten samples per patient (range 1 – 15 samples) were sequenced on six plates alongside twelve PCR controls and eight kit controls (five technical controls from blank swabs and prewash BALF and three DNA extraction kit controls). 14,529,156 high quality reads were obtained after combining forward and reverse reads, demultiplexing and quality filtering, with a median of 14,891 reads per patient sample (range 88 – 272,268 reads). Run quality statistics are summarized in Table 4.2.

Table 4.2: Summary of sequencing quality statistics for the NBS cohort in the CLIMB study.

This illustrates the number of paired reads (in millions) generated by each sequencing run (“BAX1”, “BAX2”, “NBS1”, “NBS2”, “NBS3” and “NBS4”) and their associated quality statistics. Shown are: the percentage of bases with a Phred quality score greater than 30 (Q30), cluster density and percentage of clusters passing filter (ideally > 70%).

Sequencing run number	Total no of reads (million)	> Q30 (%)	Cluster Density (k/mm²)	Clusters passing filter (%)
BAX1	7.2	82.8	433	68.8
BAX2	8.0	81.2	413	81.0
NBS1	14.7	77.4	673	72.6
NBS2	11.4	81.3	489	76.1
NBS3	11.6	83.2	507	79.3
NBS4	12.5	80.1	547	74.9

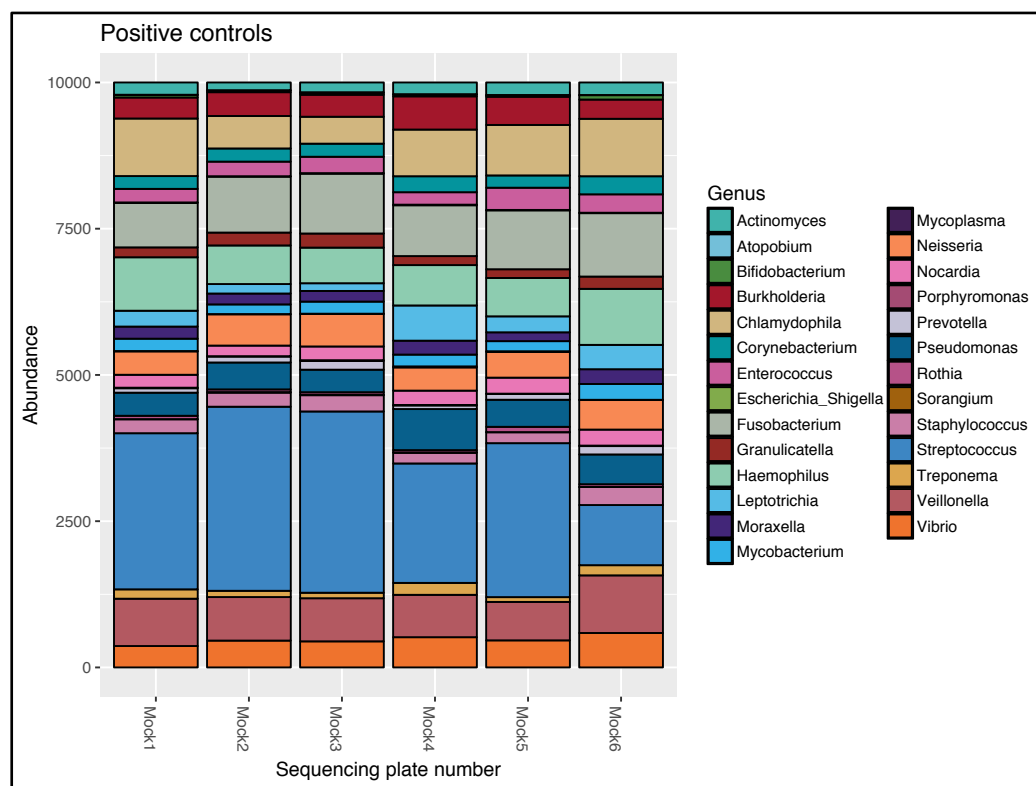
As per the quality control measures described in Chapter 2, Section 2.13.2.1. PCR negative, DNA extraction and technical controls were examined to identify

contaminants. OTUs were identified as contaminants if they had greater than 3,000 total reads in the controls or had been identified as common reagent contaminants in previous literature²¹⁰. The following were considered contaminants and removed: *Undibacterium* spp., *Comamonadaceae*, *Sediminibacterium* spp., *Methylobacterium* spp., *Planomicrobium* spp. and a *Burkholderia* OTU (ID 1931).

NMDS plots showed consistency between all runs. Upon PERMANOVA testing, a small batch effect between sequencing plates was however seen, accounting for up to 6% of variance in diversity (Bray Curtis $r^2 = 0.06$, unweighted UniFrac $r^2 = 0.05$, and weighted UniFrac $r^2 = 0.05$, all $P = 0.001$). Examining bar plots of the mock community from each sequencing plate, a reduction in *Streptococcus* spp. was seen in plate “BAX2” (shown as “Mock 6” in Figure 4.4). This reduction in relative abundance had been previously noted as the plate was involved in the cross sectional comparison of lower and upper airway microbiota (Chapter 3 Section 3.2.2). When plate BAX2 was excluded, the batch effect was small at $\leq 5\%$ (Bray Curtis $r^2 = 0.05$, $P = 0.002$; unweighted UniFrac $r^2 = 0.05$, $P = 0.001$, and weighted UniFrac $r^2 = 0.04$, $P = 0.001$). Eleven TS and eleven BALF samples were sequenced on plate “BAX2”. As the batch effect was small, these samples were not excluded from further analyses.

Figure 4.4: Barplot comparing mock communities run on each sequencing plate to assess for consistency between sequencing plates for NBS study.

Illustrated is the relative abundance of the 27 genera included in the mock community as listed in Chapter 2, Section 2.7. Each barplot corresponds to a sequencing plate on which samples from infants with CF in the NBS cohort were sequenced. Plates “Mock 5” and “Mock 6” correspond to plates “BAX1” and “BAX2” respectively, which were involved in the cross sectional comparison of lower and upper airway microbiota. This shows a reduction in *Streptococcus* spp. in “Mock 6” (BAX2) but consistent sequencing of the mock community for all other sequencing plates.



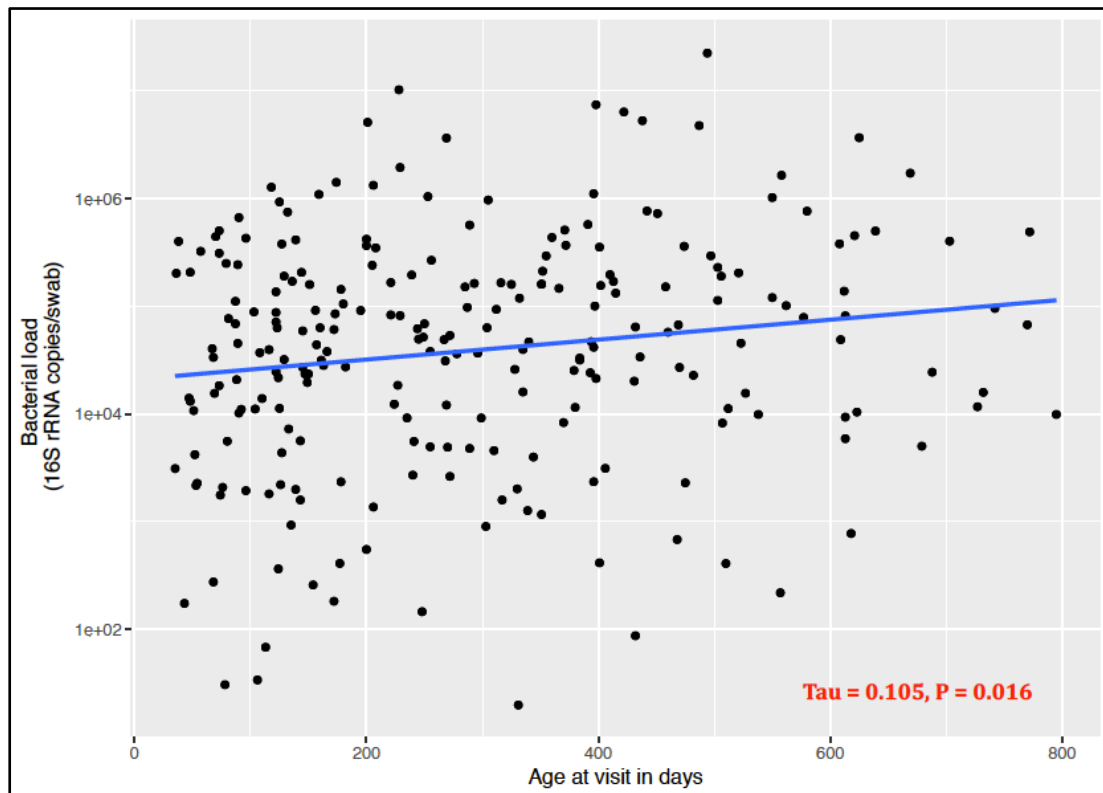
4.3.3 Development of the airway microbiota in the first 2 years of life

4.3.3.1 Changes in bacterial load

For the thirty patients, a median of 48,877 copies of the 16S rRNA gene were obtained per swab (range 20 – 22,461,921 copies per swab). Comparing changes in bacterial load with age in days, there was a small but statistically significant increase in bacterial load over the first 2 years of life ($T = 0.105$, $P = 0.016$) (Figure 4.5). However, using a mixed effects model, there was no significant difference in bacterial load with age for paired samples by age ranges (e.g. 3-4m, 5-6m etc.).

Figure 4.5: Changes in bacterial load with age on TS obtained in the first 2 years of life.

This demonstrates a weak positive correlation (Kendall's Tau = 0.105, $P = 0.016$) between age (in days) and bacterial load (16S rRNA copies per swab) in the first 2 years of life.



There was no difference in bacterial load between: infants who were born by SVD or LSCS; infants who received any breast milk versus those who were exclusively formula fed; those who were homozygous pPhe508Del and heterozygous pPhe508Del, and those who were or were not pancreatic insufficient (all $P > 0.05$) (Table 4.3).

Table 4.3: Differences in bacterial load for clinical variables.

Comparisons shown in the mean (SD) in bacterial load (16S rRNA copy number per swab) in the presence or absence of the clinical variables listed below. All differences were non-significant (NS, $P > 0.05$). SVD – spontaneous vaginal delivery.

Clinical variable	Yes	No	<i>P</i> -value
Gender (female)	614,000 (2,442,000)	285,000 (723,000)	NS
Genotype (homozygous)	614,000 (2,544,000)	344,000 (1,089,000)	NS
Mode of delivery (SVD)	444,500 (1,930,000)	506,000 (1,582,000)	NS
Breastfeeding	375,000 (1,135,000)	562,000 (2,487,000)	NS
Pancreatic insufficiency	525,000 (2,089,000)	249,000 (539,000)	NS

4.3.3.2 Genus and OTU level changes

At the genus level, *Streptococcus* was the most common genus present in the first 2 years of life in infants with CF, representing over half the bacterial community (55% of reads). Other common genera included: *Haemophilus* (12.5%), *Veillonella* (7.4%), *Neisseria* (5.6% of reads), *Granulicatella* (3.8% of reads), *Gemella* (3.1%) and *Prevotella* (2.9% of reads). *Staphylococcus* had low relative abundance (0.1% of reads) and *Pseudomonas* was not a common genus (< 0.01% of reads). The most common genera present in the first 2 years of life are illustrated in Figure 4.6.

Mixed effects models (Section 4.2.4) were used to test changes in relative abundance of the five most common genera over the first 2 years of life from baseline with patient study number as a random effect and age range as the fixed effect. This allowed paired, longitudinal comparisons in the relative abundance of genera to be performed.

Streptococcus spp. and *Haemophilus* spp. began with a similar relative abundance but diverged thereafter (Figure 4.7). *Streptococcus* spp. increased in relative abundance in the first 7-8 months of life ($t_{(172)} = 3.42$, $P < 0.001$) before reaching an asymptote, with no statistically significant difference in the relative abundance of *Streptococcus* spp. at 2 years of age to that in the first 2 months of life ($t_{(172)} = 1.09$, $P = 0.279$). In contrast, *Haemophilus* spp. decreased significantly in relative abundance from 3 to 9 months of age (at 9-10 months, $t_{(172)} = -3.77$, $P < 0.001$) and thereafter plateaued, although their relative

abundance at 2 years of age was still significantly lower than in the first 2 months of life ($t_{(172)} = -2.20$, $P = 0.029$) (Figure 4.9). Similarly, *Veillonella* spp. decreased rapidly at 2 months of age ($t_{(172)} = -4.36$, $P < 0.001$) and then remained at a similar relative abundance until 2 years of age ($t_{(172)} = -1.67$, $P = 0.097$), whereas the relative abundance of *Neisseria* spp. fluctuated and *Granulicatella* spp. only showed a significant increase in relative abundance after the first year of life ($t_{(172)} = 1.65$, $P = 0.013$ at 19-21m of age).

Staphylococcus spp. and *Pseudomonas* spp. were very low in relative abundance in the first 2 years of life, with a significant increase seen at the end of the second year for *Staphylococcus* spp. ($t_{(172)} = 2.94$, $P = 0.004$) but not for *Pseudomonas* spp. ($t_{(172)} = 1.63$, $P = 0.105$), although the levels of both remained very low.

Figure 4.6: Sankey plot illustrating the most common species present in the airway microbiota in the first 2 years of life.

Analysed by combining all TS collected during the study period following rarefaction to 600 reads. The plot illustrates the hierarchy of classification by the eight major taxonomic ranks (Phylum, Class, Order, Family, Genus and Species) and the relative abundance (%) of different species. . The width of the ribbon represents the proportional abundance of that species within the community. The most common genera have been highlighted in colour. This illustrates that *Streptococcus* is the most common genus, accounting for 55% of reads. Other common genera include: *Haemophilus*, *Veillonella*, *Neisseria*, *Granulicatella*, *Gemella* and *Prevotella*.

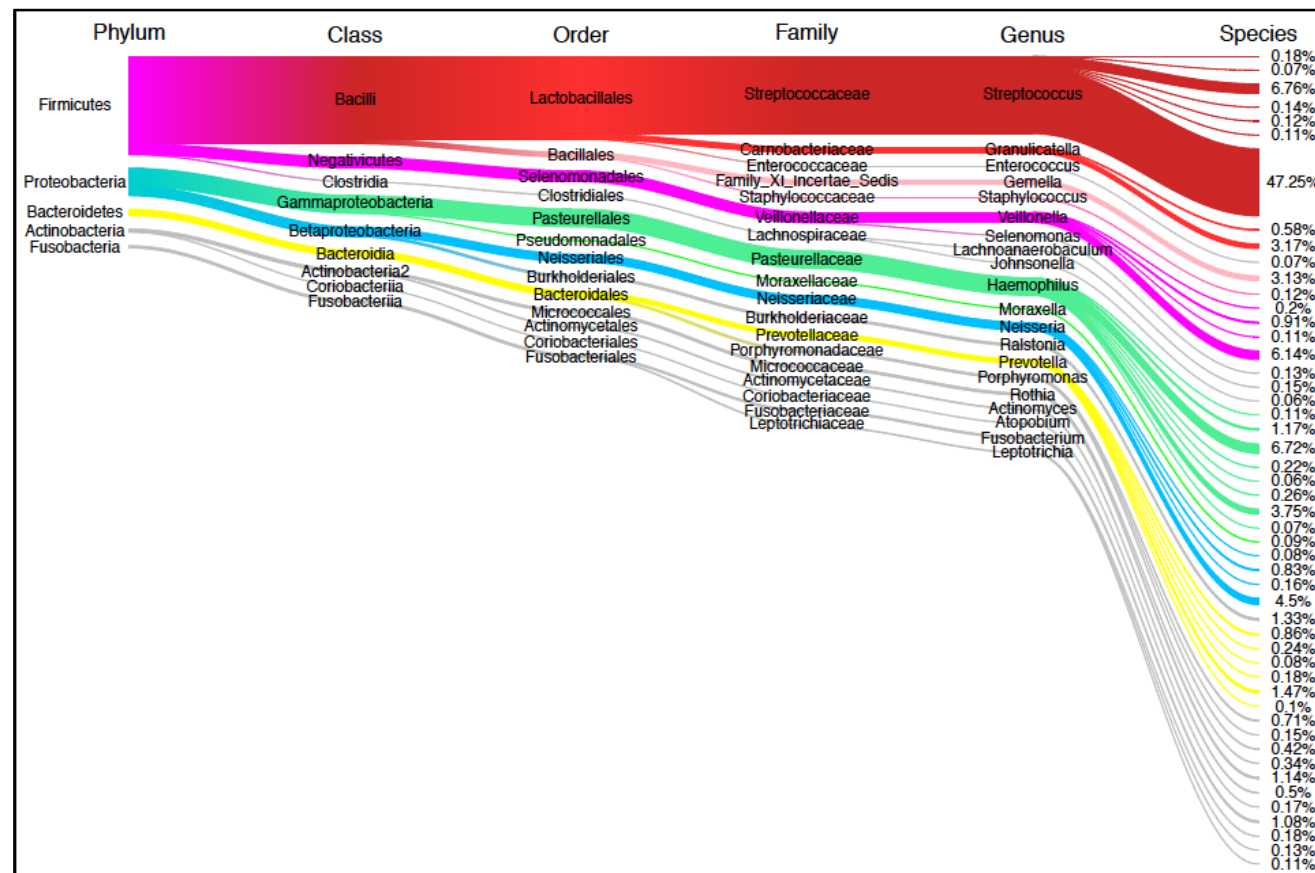


Figure 4.7: Changes in the relative abundance of five most common genera with age in infants with CF.

This figure shows the mean relative abundance of *Streptococcus* spp., *Haemophilus* spp., *Neisseria* spp., *Veillonella* spp. and *Granulicatella* spp. measured in number of reads from samples rarefied to 600 reads. The error bars represent the standard deviation. This illustrates an inverse relationship in the relative abundance of *Streptococcus* spp. and *Haemophilus* spp. in the first two years of life: *Streptococcus* spp. increased in the first 7-8 months of life ($t_{(172)} = 3.42, P < 0.001$) before reaching an asymptote, whereas, *Haemophilus* spp. decreased from 3 to 9 months of age (at 9-10 months, $t_{(172)} = -3.77, P < 0.001$) and then plateaued.

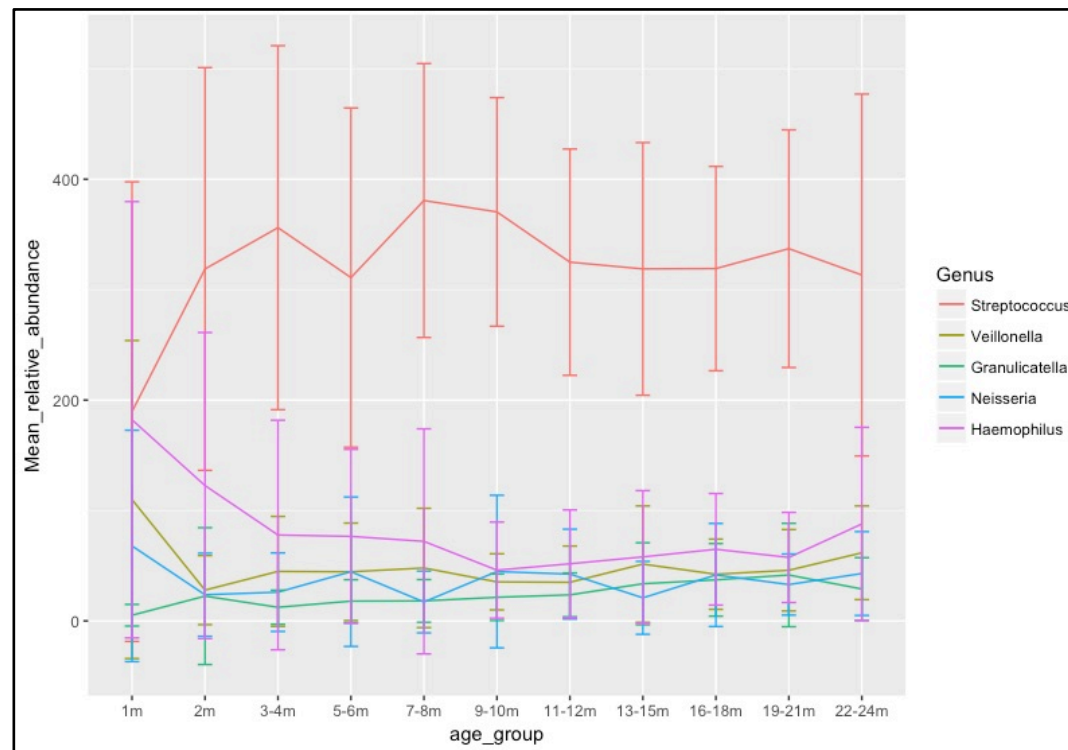
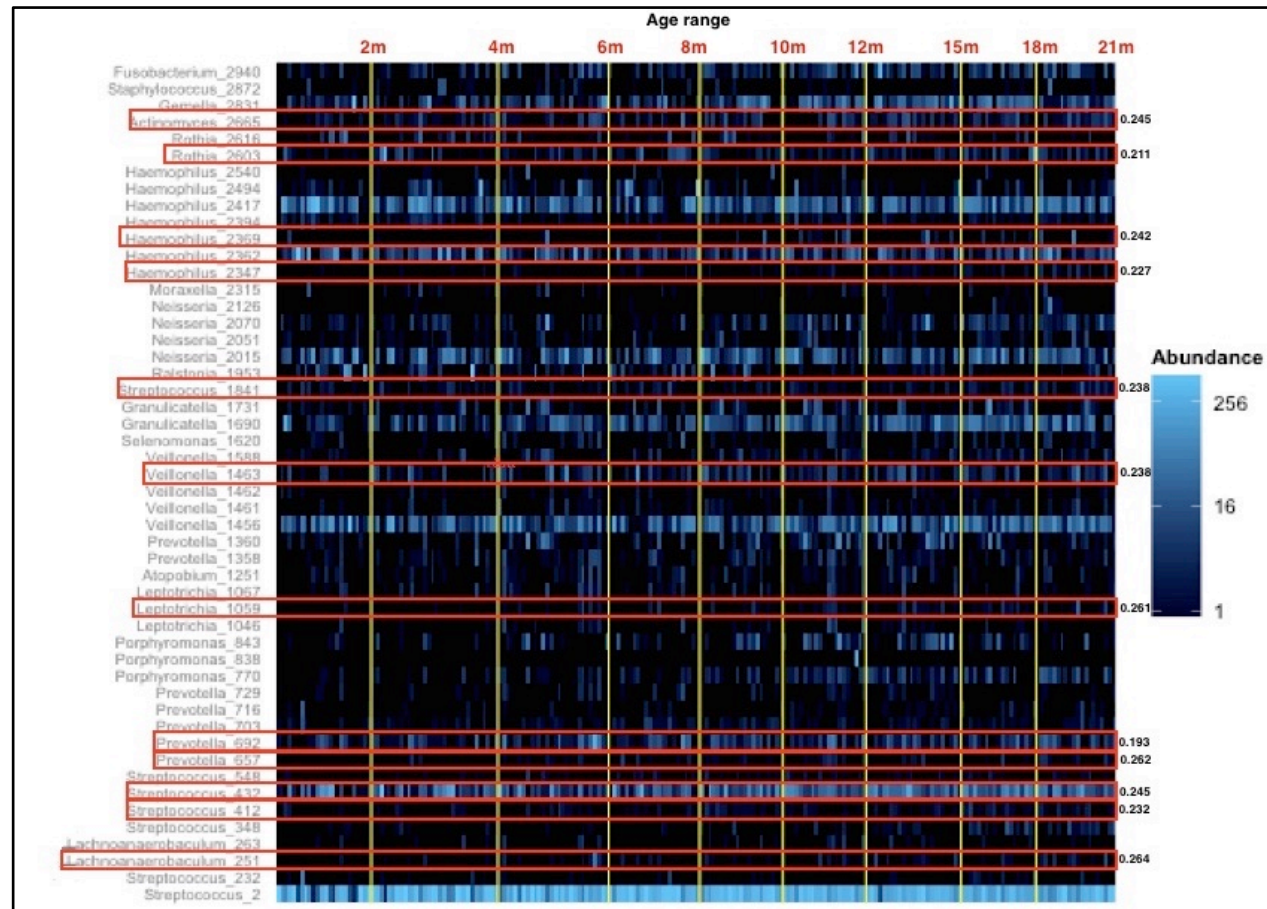


Figure 4.8: Heatmap illustrating changes in the 50 most common OTUs in the first 2 years in ascending age.

Red bars highlight those OTUs which increased in relative abundance with age on Spearman's rank correlation testing, with the r value given on the right hand side of the bar. All correlations were weak ($r < 0.3$).



At the OTU level, multiple Spearman rank correlation testing with FDR correction was performed to assess to changes in the relative abundance of all OTUs with age. Several OTUs showed a small ($r \leq 0.245$) but significant ($P_{\text{adj}} < 0.05$) increase in relative abundance with age. Several OTUs belonging to four of the most common genera showed an increase in relative abundance in the first 2 years and including *Streptococcus* (OTU ID 412, 432 and 1841), *Haemophilus* (OTU ID 2347 and 2369), as well as individual OTUs belonging to the genera *Neisseria* (OTU ID 2124) and *Veillonella* (OTU ID 1463). In addition, several less common OTUs showed an increase in relative abundance with age, including *Alysiella* (OTU ID 2166) (from the Neisseriaceae family), *Enterococcus* (OTU ID 1712), *Lactobacillus* (OTU ID 1586), *Prevotella* OTU ID 657 and 692) and *Rothia* (OTU ID 2603) (Table 4.4). One species of *Streptococcus* (OTU ID 1897) showed a decrease in relative abundance. Overall changes in the fifty most common OTUs over the first 2 years of life are shown in the heatmap in Figure 4.8.

From examining individual patient barplots, whilst the airway microbiota appeared to be highly individual, *Streptococcus* spp. was abundant throughout all patients at all timepoints. No consistent changes were seen with changes in symptoms or positive growth on bacterial cultures (Figure 4.9 & Appendix A4).

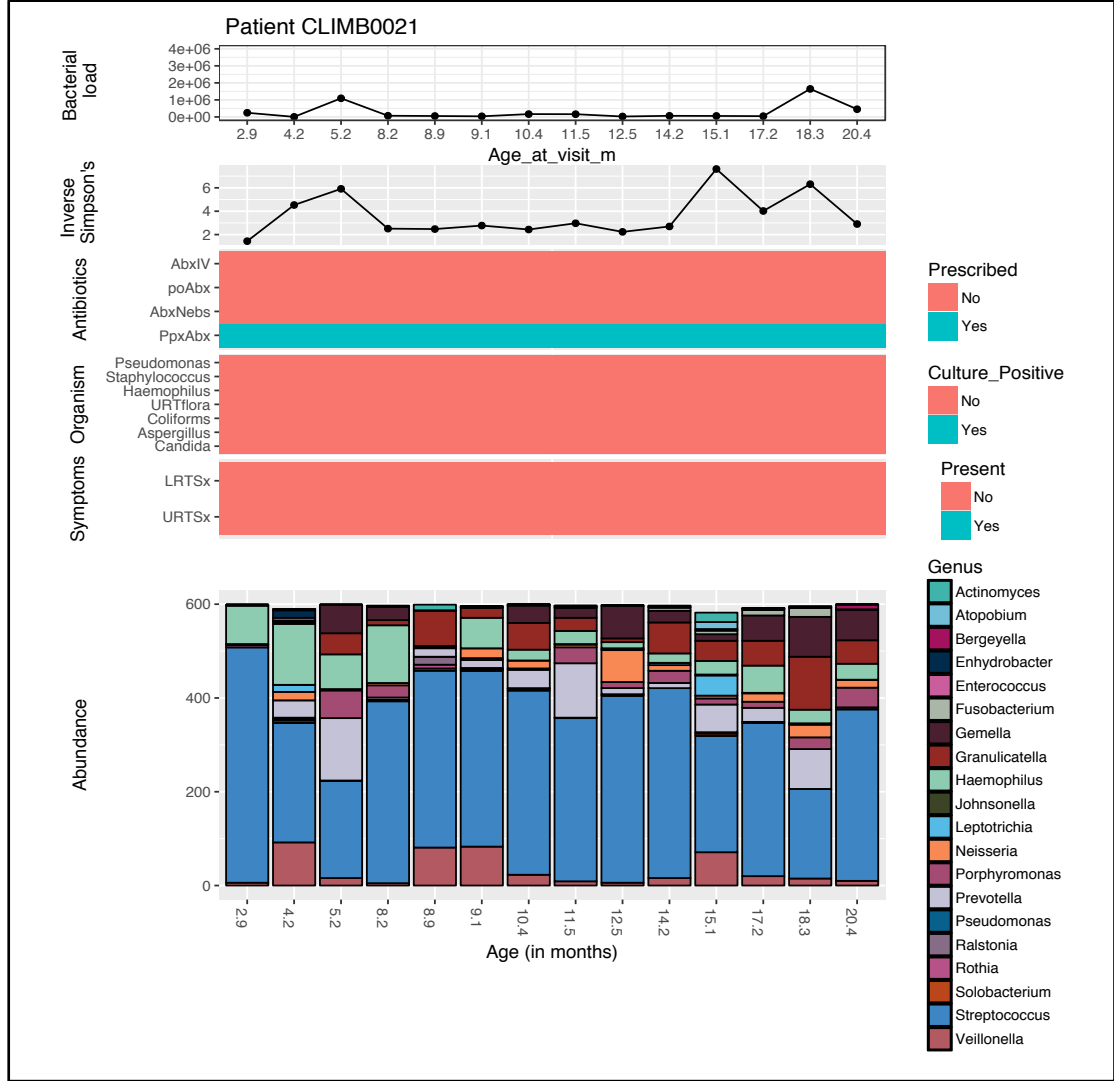
Table 4.4: List of OTUs showing significant ($P_{\text{adj}} < 0.05$) changes in relative abundance with age.

Assessed using Spearman rank multiple correlation testing with FDR correction. Several species of *Streptococcus*, *Haemophilus* and *Prevotella*, as well as other individual species showed a small ($r < 0.3$) but significant positive correlation with age in the first two years of life.

OTU name	r value	Adjusted P-value
<i>Actinomyces</i> (OTU ID 2665)	0.25	< 0.001
<i>Alysiella</i> (OTU ID 2166)	0.23	0.006
<i>Atopobium</i> (OTU ID 1256)	0.21	0.013
<i>Capnocytophaga</i> (OTU ID 812)	0.20	0.016
<i>Enterococcus</i> (OTU ID 1712)	0.20	0.014
<i>Haemophilus</i> (OTU ID 2347)	0.23	0.006
<i>Haemophilus</i> (OTU ID 2369)	0.24	0.003
<i>Lachnoanaerobaculum</i> (OTU ID 251)	0.26	< 0.001
<i>Lactobacillus</i> (OTU ID 1586)	0.19	0.025
<i>Lautropia</i> (OTU ID 2253)	0.26	< 0.001
<i>Leptotrichia</i> (OTU ID 1059)	0.26	< 0.001
<i>Neisseria</i> (OTU ID 2124)	0.19	0.025
<i>Prevotella</i> (OTU ID 657)	0.26	< 0.001
<i>Prevotella</i> (OTU ID 692)	0.19	0.020
<i>Rothia</i> (OTU ID 2603)	0.21	0.013
<i>Streptococcus</i> (OTU ID 412)	0.23	0.006
<i>Streptococcus</i> (OTU ID 432)	0.25	0.002
<i>Streptococcus</i> (OTU ID 1841)	0.24	0.004
<i>Streptococcus</i> (OTU ID 1897)	-0.19	0.016
<i>Veillonella</i> (OTU ID 1463)	0.24	0.006

Figure 4.9: Example plot of an individual patient illustrating changes in relative abundance of OTUs against bacterial load (16S rRNA copies per swab), diversity changes and clinical variables.

Despite frequent sampling during an 18-month follow-up period, little change was seen in bacterial load, the Inverse Simpson’s diversity index or community structure (shown by the barplot) with changes in symptoms, growth of *P. aeruginosa* or antibiotic treatment (oral or IV). Other individual patient barplots (Appendix A4) similarly showed a lack of identifiable biomarkers associated with changing clinical status.

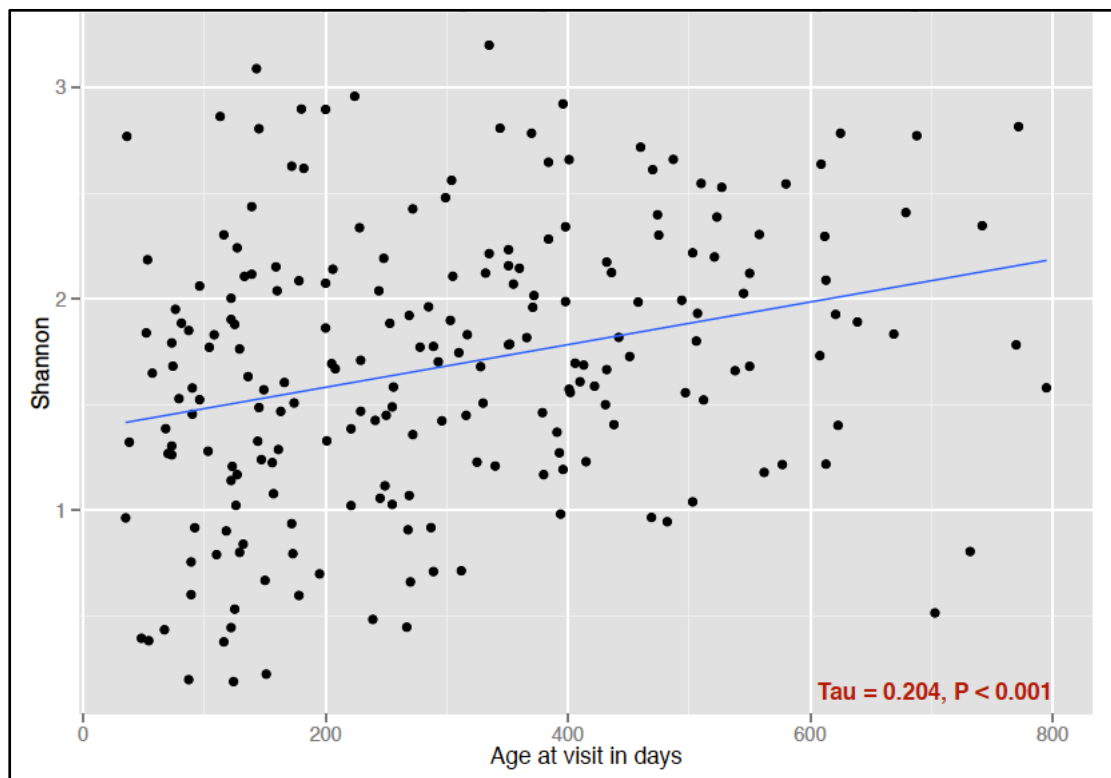


4.3.3.3 Changes in alpha diversity

A weak positive correlation was seen between richness (Tau = 0.181), evenness (Tau = 0.196), the Shannon (Tau = 0.204) and Inverse Simpson's diversity indices (Tau = 0.183) ($P < 0.001$) and age. This suggests that there was a small increase in the number of different organisms present and their spread within the community during the first 2 years of life (Figure 4.10).

Figure 4.10: Scatterplot of changes in Shannon diversity index with age.

This illustrates a weak positive correlation (Kendall's Tau = 0.204, $P < 0.001$) between the Shannon Diversity Index and age (in days).

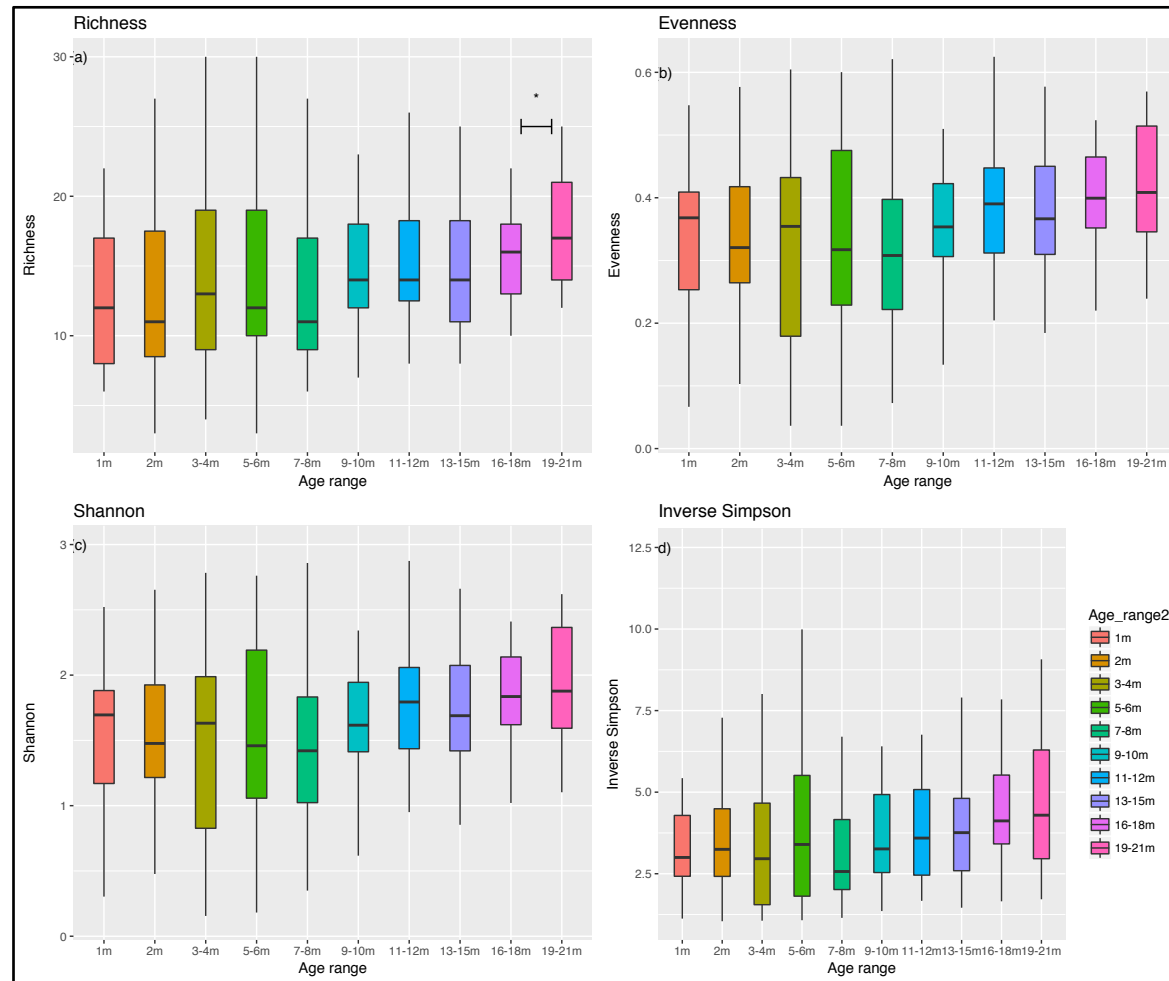


Mixed effects models were used to test the relationship between alpha diversity and age grouped into 2-3 month age ranges. However, this showed there was no significant association between alpha diversity (richness, evenness, Shannon and Inverse Simpson's diversity index) and increasing age ($P > 0.05$) (Figure 4.11).

This suggests that the age of the patient does not influence the number of different organisms (measured by richness) and the spread of these organisms (measured by evenness) within a community in the first 2 years of life. There was no significant difference in alpha diversity with upper or lower respiratory tract symptoms ($P > 0.05$).

Figure 4.11: Changes in alpha-diversity by age range illustrating an increasing trend in alpha diversity with age.

Using a mixed effects model, no significant association between alpha diversity and age ($P > 0.05$).



4.3.3.4 Changes in beta-diversity

Beta diversity measures species turnover and nestedness within a community²⁴⁸ and was analysed to assess the influence of age on community structure using the adonis function in R. A blocked design was used, constrained by patient study number to account for the repeated measures study design. A significant influence of patient age on beta diversity was demonstrated using the Bray Curtis ($r^2 = 0.15$, $P = 0.03$), unweighted UniFrac ($r^2 = 0.16$, $P = 0.027$) and weighted UniFrac ($r^2 = 0.17$, $P = 0.04$) scores. This suggests that age has a significant impact on community structure. The specific organisms present, their phylogenetic relationships and relative abundance (as reflected in each of the above scores respectively) differed with increasing age during the first 2 years of life in infants with CF.

4.3.3.5 Perinatal predictors of community structure

Alpha and beta diversity testing were repeated to determine whether any clinical variables influence community structure. Evenness, the Shannon diversity index and Inverse Simpson's diversity index but not richness were significantly higher in patients delivered by LSCS than SVD (Evenness median 14.5 [range 6-21] in LSCS and 13 [3 – 32] in SVD, $P = 0.045$; Shannon diversity index median 0.4 [0.1 – 0.6] in LSCS and 0.345 [0.02 – 0.63] in SVD, $P = 0.045$, and Inverse Simpsons median 3.7 [1.1 – 11.4] in LSCS and 3.2 [1.04 – 12.3] in SVD, $P = 0.04$). There was also a significantly lower Inverse Simpson's diversity index in those patients who had GORD (median 2.8 [range 1.1- 10.8]) than those who did not (median 3.5 [range 1.0 – 12.3]) ($P = 0.04$).

A PERMANOVA was used to test the influence of multiple clinical variables simultaneously on community composition measured by beta diversity. Overall, whilst several clinical variables had a significant influence on beta diversity, the degree of variance attributed to any of the clinical variables tested was small. Mode of delivery (SVD or LSCS) exerted the greatest influence, accounting for up to 3% of variance in community structure (Weighted UniFrac, $P = 0.015$). Other significant variables on beta diversity testing included: growth of *P. aeruginosa* during the study period; genotype (homozygous pPhe508Del vs. heterozygous pPhe508Del/other); pancreatic insufficiency; the presence of upper respiratory tract symptoms, and receiving oral or nebulised antibiotics (Table 4.5).

Table 4.5: Testing the influence of clinical variables on beta diversity.

Listed below are the clinical variables that exerted a significant influence on beta-diversity as measured using a permutational multivariate ANOVA (PERMANOVA).

Clinical variable	Diversity measure	r ²	P-value
Mode of delivery	Bray Curtis	0.02	0.009
	Weighted UniFrac	0.03	0.026
Homozygous	Bray Curtis	0.02	0.004
p.Phe508Del	UniFrac	0.01	0.001
Pancreatic insufficiency	UniFrac	0.01	0.011
Upper respiratory tract symptoms (yes/no)	Bray Curtis	0.01	0.027
	Weighted UniFrac	0.02	0.013
Any <i>P. aeruginosa</i> grown	Bray Curtis	0.01	0.015
	Weighted UniFrac	0.02	0.015
Nebulised antibiotics	Bray Curtis	0.01	0.039
	UniFrac	0.01	0.017
Oral antibiotics	Bray Curtis	0.01	0.006
	UniFrac	0.01	0.031

4.3.3.6 The impact of the first exacerbation

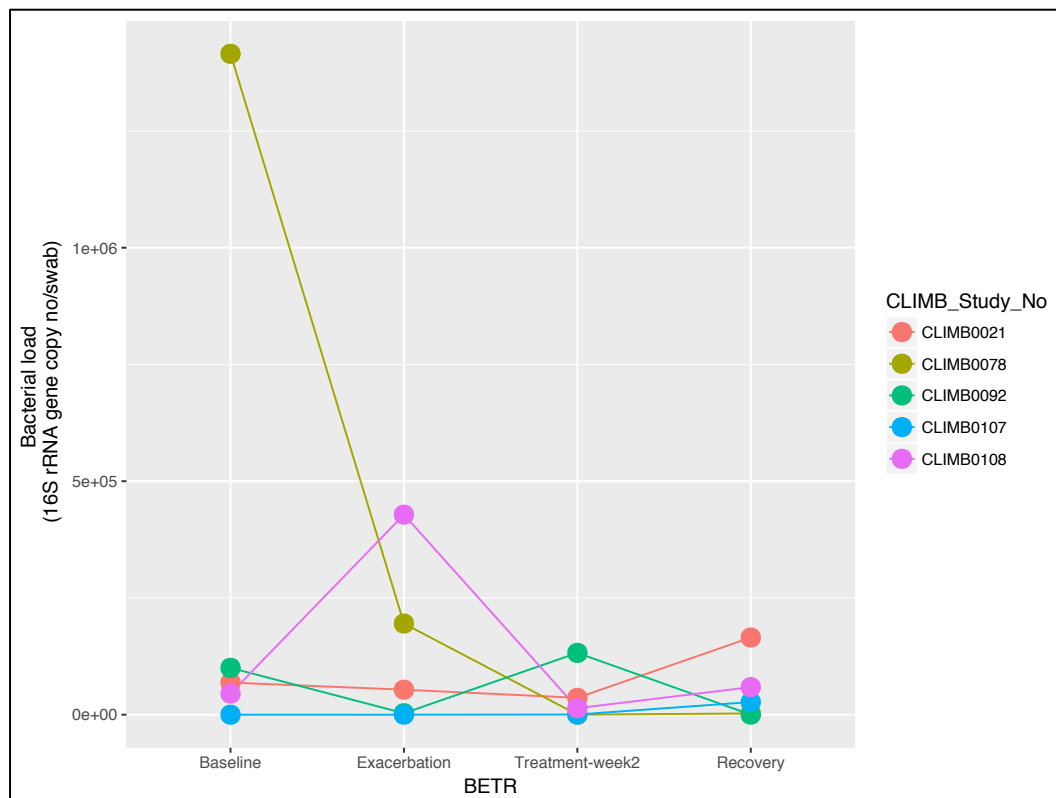
Seven patients experienced an exacerbation (definition see Chapter 2, Section 2.3) during the study period. Five of the patients had sequential samples collected at:

- Baseline (B) - within the month prior to starting IV antibiotics;
- Exacerbation (E) - within 48 hours of starting IV antibiotics;
- Treatment (T) - between days 10-14 of IV antibiotics;
- Recovery (R) - within 1 month of completing IV antibiotics.

Due to the small sample size, an exploratory analysis of the changes at exacerbation was performed. One patient (CLIMB0078) showed a markedly higher bacterial load at baseline than the other patients who had an exacerbation. This sample was collected by a helpful parent, who collected the TS into a tube containing culture medium, as is used in clinical practice (Figure 4.12). The sample collection methods for this sample therefore deviated from the TS collection SOP and this sample was excluded from statistical analyses. Overall, there was no significant difference in bacterial load (copies of 16S rRNA gene per throat swab) between baseline, exacerbation, treatment and recovery samples using mixed effects modelling ($P > 0.05$).

Figure 4.12: Changes in bacterial load (16S rRNA copies per throat swab) with exacerbations (N = 5).

Illustrated for each individual patient at 4 timepoints: “Baseline” - within 1 month prior to exacerbation; “Exacerbation” - within the first 48 hours of admission for IV antibiotics; “Treatment – week 2” – between days 10-14 of IV antibiotic treatment; “Recovery” – within 1 month of cessation of IV antibiotic treatment. No significant change in bacterial load between these timepoints was observed ($P > 0.05$).



At the genus level, whilst *Streptococcus* remained the most common genus at exacerbation, its relative abundance decreased from 77.4% at baseline to 70.6% at exacerbation for all *Streptococcus* combined. Whereas *Haemophilus* was the second most abundant genus at baseline its relative abundance decreased at exacerbation, from 10.3% at baseline to only 1.0% of reads at exacerbation. Other changes occurring between baseline and exacerbation samples: *Granulicatella* increased from 1.7% to 9.1%; *Veillonella* increased from 3.1% to

8.8%, and *Prevotella* and *Gemella* remained constant from 1.0% and 2.5% to 1.8% and 2.8% respectively (Table 4.6).

Table 4.6: Changes in the relative abundance (%) of common genera between baseline and exacerbation samples.

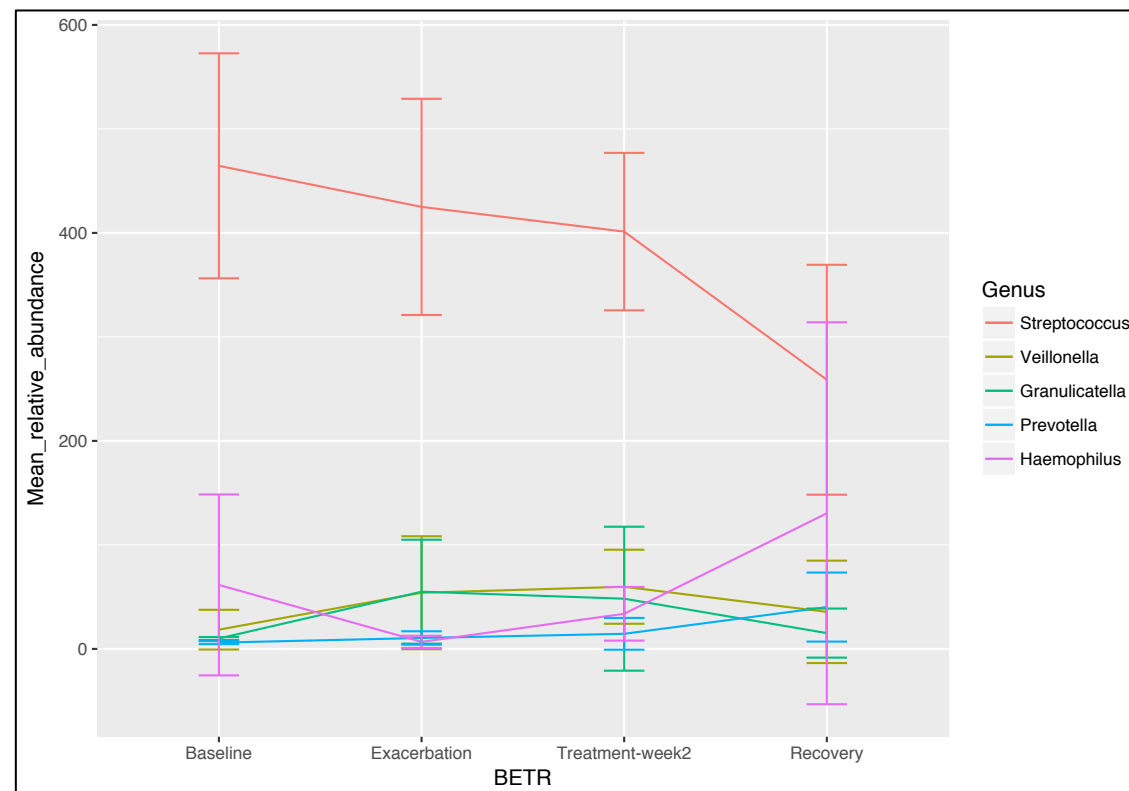
Genus	Relative abundance at baseline (%)	Relative abundance at exacerbation (%)	P-value
<i>Streptococcus</i>	77.4	70.6	0.662
<i>Haemophilus</i>	10.3	1.0	0.096
<i>Veillonella</i>	3.1	8.8	0.079
<i>Granulicatella</i>	1.7	9.1	0.288
<i>Gemella</i>	2.5	2.8	0.931
<i>Prevotella</i>	1.0	1.8	0.819

Examining trends in the 5 most common genera at baseline, exacerbation, during treatment and at recovery, the relative abundance of *Streptococcus* spp. decreased from baseline to exacerbation, treatment and recovery, showing a shaper rate of decline between treatment and recovery samples. In contrast *Haemophilus* spp. decreased from baseline to exacerbation but recovered thereafter at treatment and recovery samples. *Granulicatella* spp. and *Veillonella* spp. showed similar but less pronounced trends, increasing slightly at exacerbation, but decreasing thereafter. *Prevotella* spp. showed little change in relative abundance throughout (Figure 4.13). However, there was no significant difference in the relative abundance of the five most common genera between baseline and exacerbation samples using a mixed effects model ($P > 0.05$). Similarly, there was no significant difference at an OTU level using multiple correlation testing using Spearman rank with FDR correction ($P_{\text{adj}} < 0.05$).

No difference was seen in alpha diversity between baseline, exacerbation, treatment and recovery samples using mixed effects modelling ($P > 0.05$) or Beta diversity (PERMANOVA, Bray Curtis $P = 0.089$, unweighted UniFrac $P = 0.629$, Weighted UniFrac $P = 0.127$).

Figure 4.13: Changes in the relative abundance of the five most abundant genera with exacerbations.

The figure illustrates changes in the mean relative abundance (shown as the total number of reads following rarefaction to 1000 reads) of *Streptococcus* spp., *Haemophilus* spp., *Granulicatella* spp., *Veillonella* spp. and *Prevotella* spp. at baseline (1 month prior to starting IV antibiotics), exacerbation (within 48 hours of starting IV antibiotics), treatment (at days 10-14 of IV antibiotics) and recovery (within 1 month of stopping IV antibiotics). All changes were non-significant ($P > 0.05$).



4.3.3.7 The impact of the first growth of *P.aeruginosa*

Seven patients grew *P. aeruginosa* during the study period. All growths were reported as “scanty” (equivalent to 10^3 colony forming units). From those, paired samples prior to, during and after *P. aeruginosa* growth were obtained from 6 patients. These samples were described as:

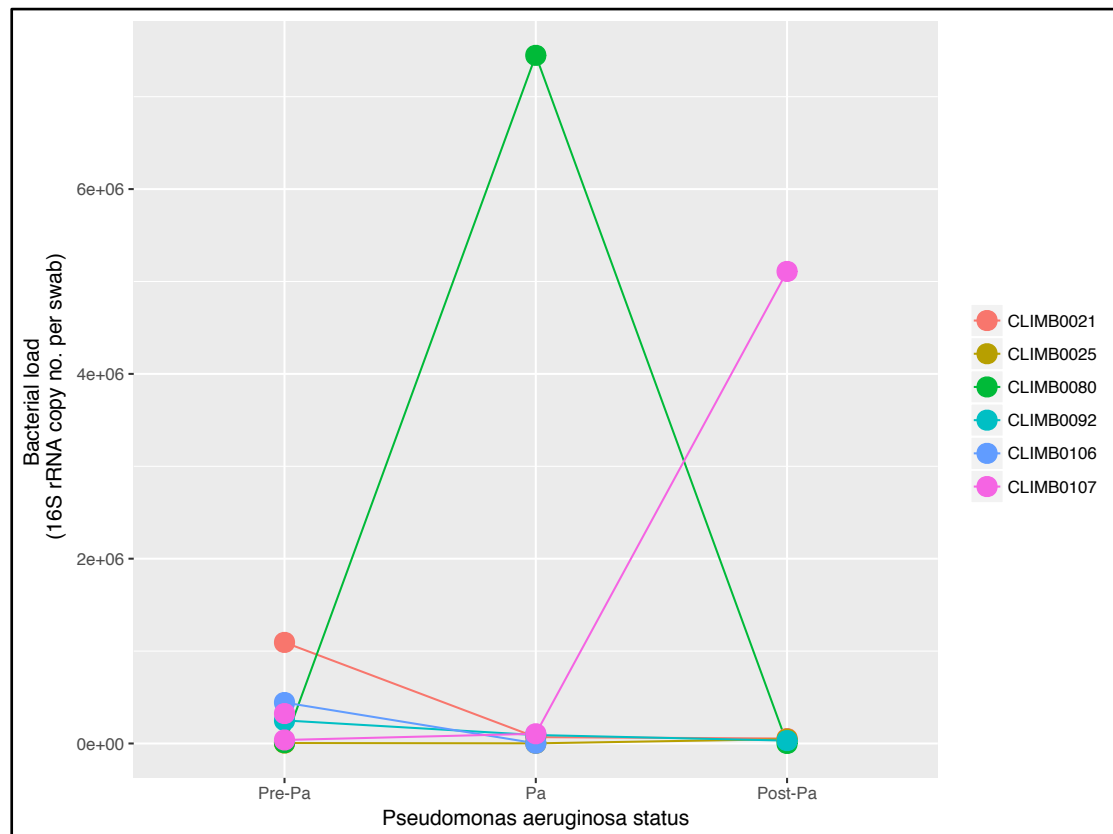
- Pre-Pa – to represent the last sample taken prior to *P. aeruginosa* growth.
- Pa – to represent the sample collected at the time *P. aeruginosa* was grown on bacterial culture in a clinical sample.
- Post-Pa – to represent the sample collected after treatment for *P. aeruginosa*, following successful eradication when bacterial cultures were negative.

Samples taken at the time of bacterial growth of *P.aeruginosa* showed a low relative abundance (< 1%) of *Pseudomonas* spp. on 16S rRNA gene sequencing.

Overall, there was no significant difference in total bacterial load between samples prior to, during or after growth of *P. aeruginosa* ($H = 0.009$, $P = 0.923$) (Figure 4.14). One patient (CLIMB0080) showed an increase in total bacterial load at the time of *P. aeruginosa* growth and another patient (CLIMB0107) showed a marked increase in bacterial load after *P. aeruginosa* treatment. There were no identifiable differences between these samples and other samples collected from patients who grew *P. aeruginosa*.

Figure 4.14: Bacterial load in relation to *P. aeruginosa* growth and treatment.

The figure illustrates that there was no significant change in bacterial load. Samples were taken at: “Pre-Pa”, reflecting the last throat swab taken prior to *P. aeruginosa* being isolated on clinical cultures; “Pa” to reflect the sample taken at the time of *P. aeruginosa* growth; “Post-Pa”, to reflect the first sample taken in which there was no growth of *P. aeruginosa* following its isolation.

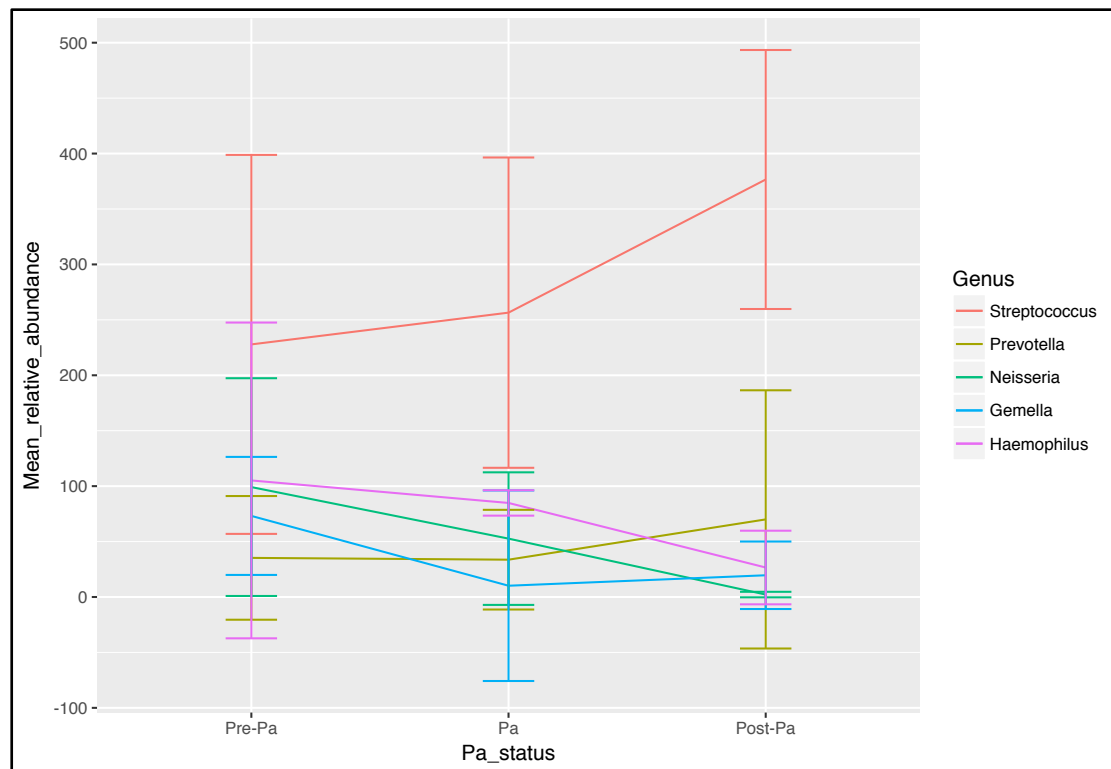


Due to the small number of patients, an exploratory analysis of changes in Pre-Pa, Pa and Post-Pa samples was performed. Examining trends in the most common genera, *Streptococcus* appeared to increase in relative abundance from before *P. aeruginosa* growth to when *P. aeruginosa* was isolated and continued to rise after. *Prevotella* showed a similar trend, whereas *Haemophilus* and *Neisseria* appeared to decrease throughout and *Gemella* appeared to decrease at growth of *P. aeruginosa* then began to recover thereafter (Figure 4.15).

Figure 4.15: Changes in the relative abundance of the five most common genera in patients who grew *P. aeruginosa* in the first two years of life.

Samples were taken at: “Pre-Pa”, reflecting the last throat swab taken prior to *P. aeruginosa* being isolated on clinical cultures; “Pa” to reflect the sample taken at the time of *P. aeruginosa* growth; “Post-Pa”, to reflect the first sample taken in which there was no growth of *P. aeruginosa* following it’s isolation.

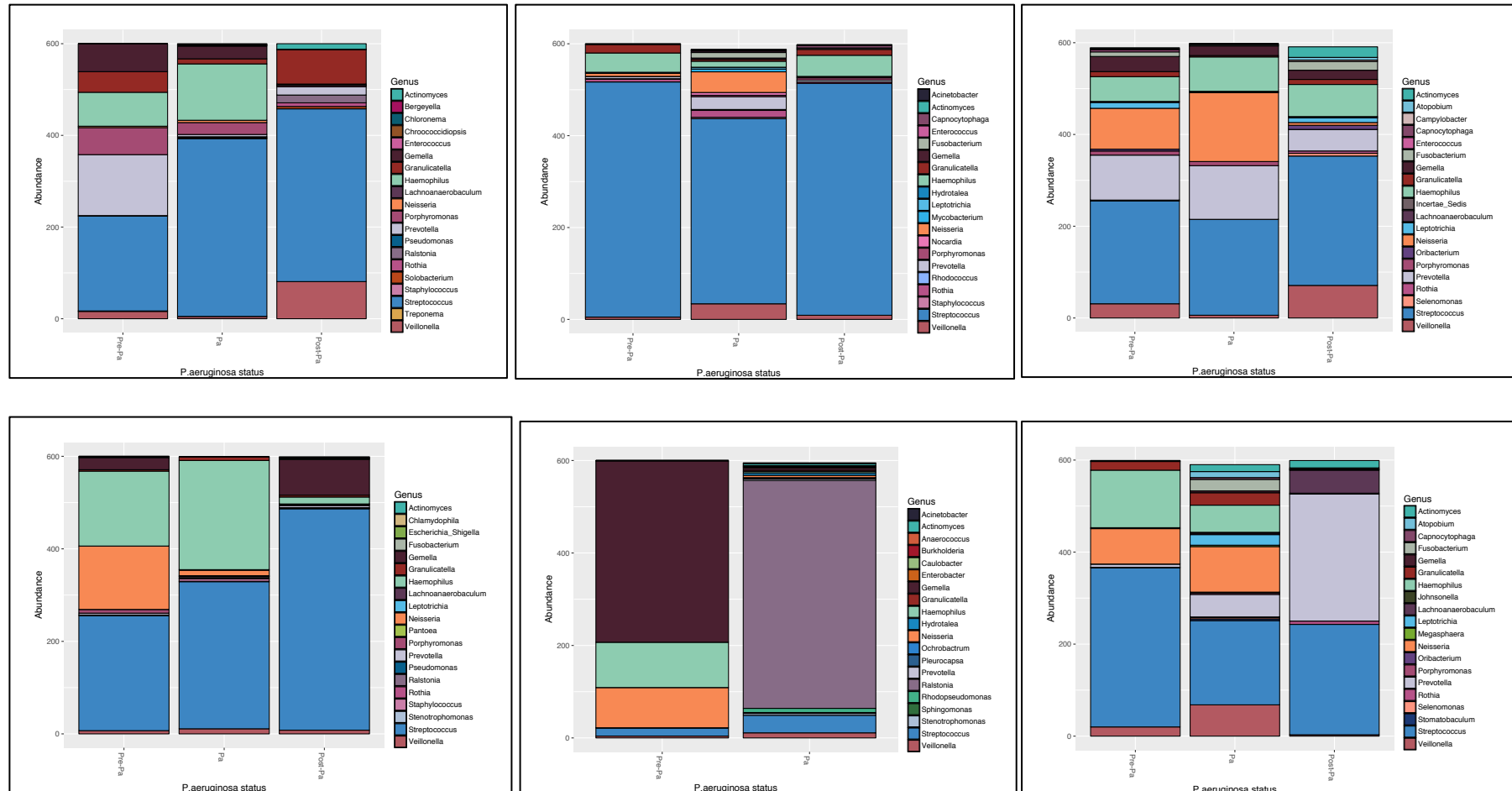
Streptococcus increased in relative abundance in *P. aeruginosa* and Post-Pa samples whilst *Haemophilus* showed the opposite trend. All patients contributed equally to these changes. All changes were non-significant ($P > 0.05$).



In Pre-Pa samples, the most common OTUs were: *Streptococcus* (OTU ID 2), accounting for 36.3% of reads; *Neisseria* (OTU ID 2015) 14.2%; *Gemella* (OTU ID 2831 12.1%; *Haemophilus* (OTU ID 2362 accounting for 7.48% of reads and OTU ID 2417 7.29%) and *Prevotella* (OTU ID 1360) 3.79%. In Pa samples, *Streptococcus* (OTU ID 2) remained the most common OTU, increasing to 40.36% of reads. Other common OTUs in Pa samples were *Haemophilus* (OTU ID 2362) 9.33%, *Neisseria* (OTU ID 2070 and OTU ID 2015 accounting for 3.64% of reads

each) and *Prevotella* (OTU ID 1360) 3.00%. *Pseudomonas* spp. was not one of the fifty most common OTUs in either Pre-Pa or Pa samples (Figure 4.16). However, there was no significant difference in relative abundance for any OTU between Pre-Pa, Pa and Post-Pa samples using either mixed effects modelling or multiple correlation testing using Spearman's rank with FDR correction ($P_{adj} > 0.05$).

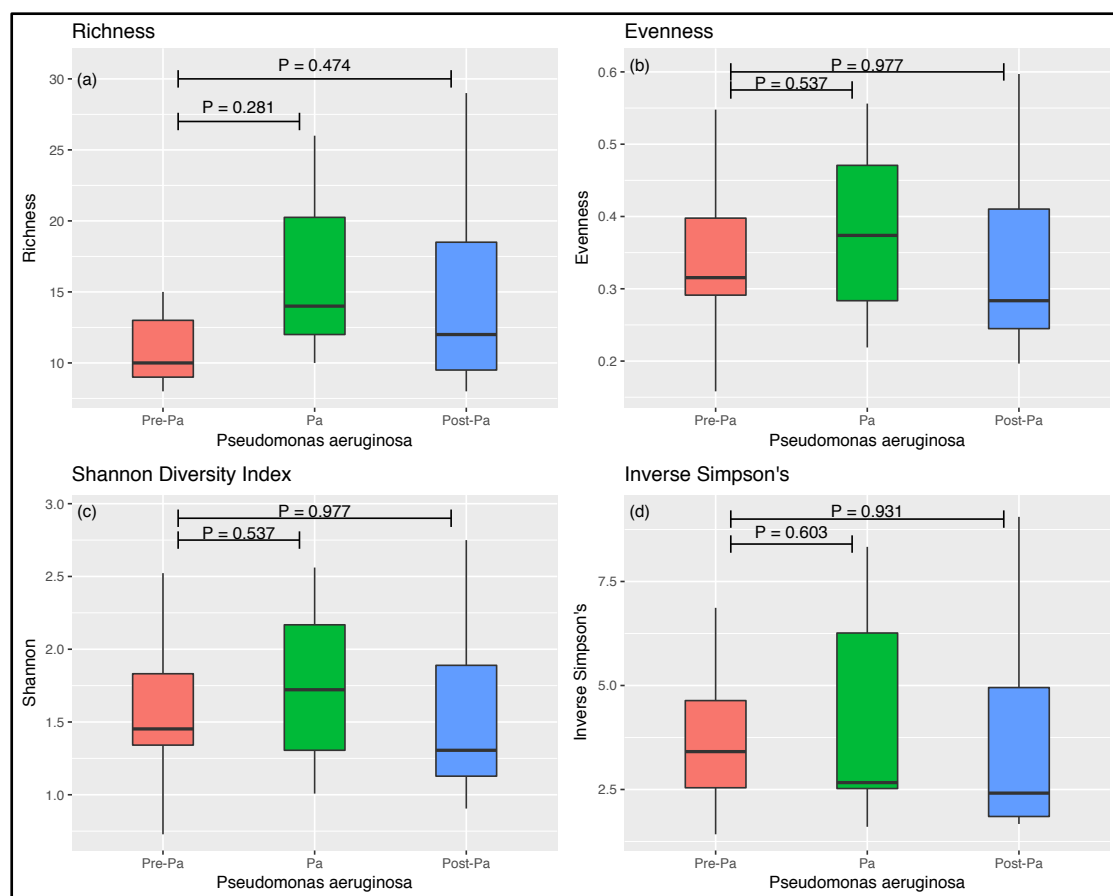
Figure 4.16: Individual patient barplots of the twenty most common genera prior to *P. aeruginosa* growth (Pre-Pa, shown on the left of each plot), at the time of *P. aeruginosa* growth (Pa, middle) and after *P. aeruginosa* growth (Post-Pa, right).



For alpha diversity, there was no significant change with growth of *P. aeruginosa* (Figure 4.17). For beta diversity, a significant difference was seen between Pre-Pa, Pa and Post-Pa using Bray Curtis ($r^2 = 0.135$, $P = 0.036$) and weighted UniFrac ($r^2 = 0.153$, $P = 0.025$) but not the unweighted UniFrac score ($r^2 = 0.153$, $P = 0.068$).

Figure 4.17: Boxplots illustrating change in alpha diversity before (Pre-Pa), during (Pa) and after (Post-Pa) growth of *P. aeruginosa* on bacterial cultures.

Measured by (a) richness, (b) evenness, (c) the Shannon Diversity Index and (d) the Inverse Simpson's diversity index. Mixed effects models were used to test for differences in each of the above alpha diversity metrics with growth and treatment of *P. aeruginosa*. There were no significant differences in alpha diversity at any of these timepoints ($P > 0.05$).



4.3.3.8 The lower airway microbiota in infancy

To determine whether diversity of the microbiota is prognostically important, alpha diversity was compared on lower airway samples between: (a) infants prescribed IV antibiotics during the study period and those who did not require IV antibiotics; (b) infants who subsequently grew *P. aeruginosa* (N = 5) and those who did not. Twenty-two patients provided twenty-five samples (7 patients in whom BALF was obtained only, 12 with bronchial brushings only and 3 patients in whom both samples were obtained) at a median age of 4 months (range 1-8 months). The results in Chapter 3 illustrated that BALF and bronchial brushings are reasonably equivalent in their representation of the lower airway microbiota. Therefore, for patients in whom both BALF and bronchial brushings were collected (N = 3), bronchial brushings were analysed in preference to BALF.

There have been few descriptions of the lower airway microbiota. Figure 4.18 illustrates the most common genera present in the lower airway microbiota at 4 months of age. Similar to TS, *Streptococcus* and *Haemophilus* remained the most common genera in lower airway samples, representing 27.92% and 22.39% of total reads respectively. *Staphylococcus*, a much less common genus in upper airway samples in the first 2 years of life (Section 4.3.3.2) was the third most abundant genus in lower airway samples at 4 months of age, representing 11.46% of reads. Other common organisms were similar to those seen longitudinally in upper airway samples: *Veillonella* (5.94%), *Neisseria* (4.01%), *Prevotella* (4.48%) and *Granulicatella* (3.19%) (Figure 4.18).

There was no significant difference between infants based on need for IV antibiotics (Evenness $t = 0.342$, $P = 0.762$; Richness $t = 0.129$, $P = 0.909$; Shannon Diversity Index $t = 0.248$, $P = 0.825$; Inverse Simpson's Diversity Index $W = 27$, $P = 0.793$). Comparing infants who grew *P. aeruginosa* with those who were free of *P. aeruginosa* growth, a significant decrease in richness was seen in infants who subsequently grew *P. aeruginosa* ($t = 3.11$, $P = 0.008$). There was no significant difference in evenness ($t=1.55$, $P = 0.170$), the Shannon Diversity Index ($t = 2.11$, $P = 0.068$) and the Inverse Simpson's Diversity Index ($W = 52$, $P = 0.130$) (Figure 4.19), although it is difficult to assess this definitively due to the low sample size.

Figure 4.18: Sankey plot showing the relative abundance of the most common genera in lower airway samples in infancy (median age 4 months, N = 22). Whilst *Streptococcus* remained the most prevalent genus in lower airway samples, its relative abundance is less in comparison with TS. Other common genera included: *Haemophilus*, *Staphylococcus*, *Veillonella*, *Neisseria*, *Prevotella* and *Granulicatella*.

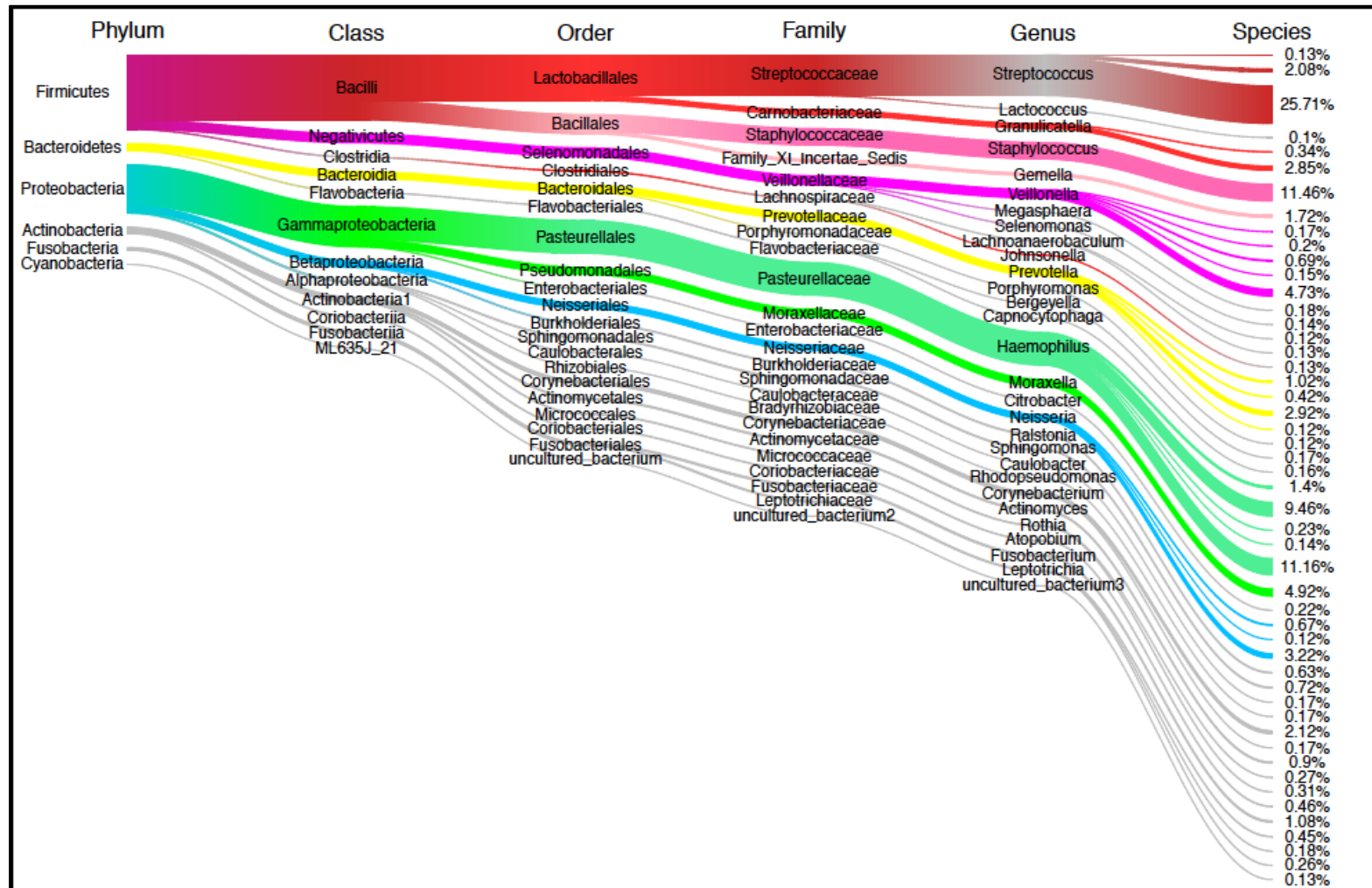
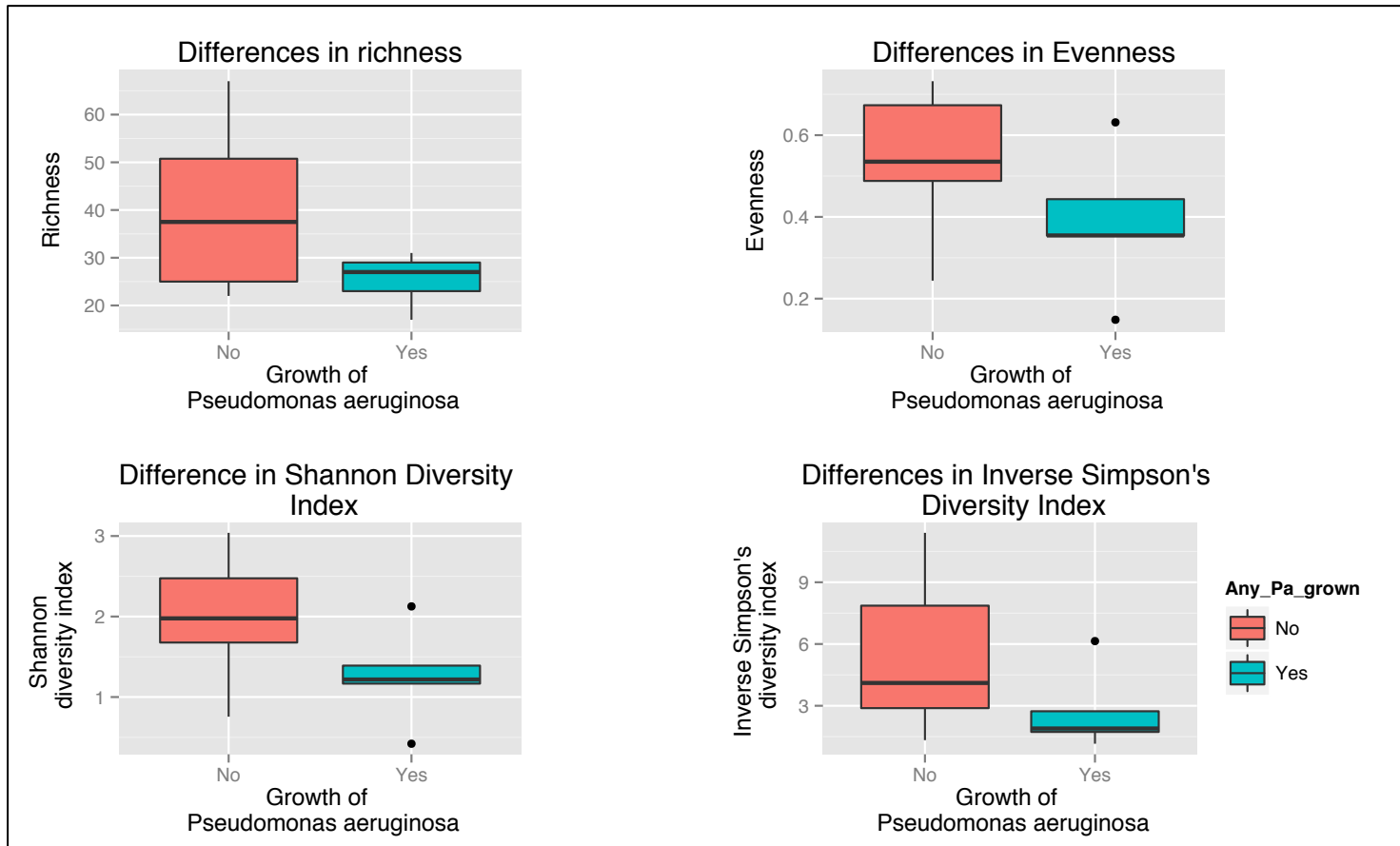


Figure 4.19: Comparing alpha diversity on lower airway samples between patients who grew *P. aeruginosa* with those who did not.

There was significantly lower richness in patients who grew *P. aeruginosa* ($t = 3.11$, $P = 0.008$) but not in evenness, the Shannon Diversity Index and the Inverse Simpson's Diversity Index.



4.4 Discussion

The aim of this chapter was to describe and better understand how the infant microbiota in CF develops over the first two years of life, in particular prior to the first exacerbation requiring admission for IV antibiotics and first growth of *P. aeruginosa* and the role of community diversity in these processes. This was achieved by recruiting a prospective cohort of infants diagnosed with CF on NBS (N = 30) and collecting TS at median frequency of 5 weeks between samples over the first two years of life and comparing changes in the microbiota with age, clinical status and culture-based microbiology. It is the largest study of the oropharyngeal microbiota to date in infants with CF with more frequent sampling than previously reported. This has resulted in a greater opportunity to identify changes in the microbiota with a sampling method that has shown good correlation with the lower airway microbiota (Chapter 3).

Previous studies in adults have suggested that diversity of the airway microbiota may be an important determinant of disease progression^{139,140}. The results of this study have shown that the airway microbiota in infants with CF is highly individual and gradually evolves over the first 2 years of life (Sections 4.3.3.1 – 4.3.3.4). In contradiction to the study hypothesis, only a small positive correlation was seen between age and both bacterial load and alpha diversity of the airway microbiota, measured by richness (the number of different organisms present), evenness (the spread of those organisms), the Shannon Diversity Index and the Inverse Simpson's Diversity Index (which give a composite score of both richness and evenness). In contrast, a more pronounced effect of age was seen

on beta diversity, a measure of community nestedness and turnover between samples. Up to 17.2% of variance in the unweighted UniFrac score was attributed to the age of the patient. In addition, there was no significant difference in diversity between infants who experienced a pulmonary exacerbation or grew *P. aeruginosa* and those who did not in a cross-section of lower airway samples taken in infants aged 3-5 months (Section 4.3.3.8); the relationship between the two, specifically whether a drop in diversity is cause or effect or coincident to *P. aeruginosa* isolation, needs further study. This suggests that examining changes in the organisms present and their relative abundance (which are measured by the unweighted UniFrac score) in the first 2 years of life may be more informative than focussing on changes in alpha diversity, as has been reported in several previous studies¹³⁹.

At the genus level, *Streptococcus* was the most common genus present in oropharynx in the first 2 years of life (Section 4.3.3.2), accounting for 55.0% of total reads; this is not what would have been expected from conventional culture results. *Haemophilus* was the second most common genus, accounting for 12.5% of reads. Both *Streptococcus* and *Haemophilus* were also the most common genera on BALF, reconfirming the findings of Chapter 3 that the oropharyngeal microbiota shows good correlation with that of the lower airway. Despite their high detection rates on 16S rRNA gene sequencing, *Streptococcus* spp. and *Haemophilus* spp. were not common organisms detected on bacterial culture of paired clinical samples (see Table 4.2). *H. influenzae* was identified in only one patient in this study, and no streptococcal species were identified on culture. These findings support those of Mahboubi *et al.*²⁴⁹, in which concordance

between bacterial culture and pyrosequencing of the V3- V5 hypervariable region of the 16S rRNA gene was assessed for 945 sputum samples collected from 132 patients CF. Mahboubi *et al.* found that despite detection of *Streptococcus* spp. in 88.7% of samples using pyrosequencing, *Streptococcus* spp. was only identified on culture in 1.6% of samples. Similarly, *Haemophilus* spp. was identified in 37.2% of samples using pyrosequencing and only 4.8% of samples using bacterial cultures. In contrast, “typical” CF organisms, such as *P. aeruginosa*, showed good concordance between culture and sequencing results. This further highlights that bacterial culture may fail to identify organisms present in the airway microbiota, even when they are highly abundant. A further explanation of these findings may be that many Streptococcal species, such as those belonging to the *Streptococcus milleri* group, are often dismissed as “normal oral” or “upper respiratory tract flora”, even when detected. As only one participant grew “upper respiratory tract flora” during this present study, this is less likely to be applicable to the results presented in this Chapter.

Over the first 2 years of life, trends in the relative abundance of *Streptococcus* spp. and *Haemophilus* spp. appear to have an inverse relationship: as the relative abundance of *Streptococcus* spp. increases in the first 7-9 months of life, the relative abundance of *Haemophilus* spp. decreases over the first 9-11 months of life (Section 4.3.3.2). Thus *Streptococcus* spp. and *Haemophilus* spp. may be competitors and have an antagonistic relationship in the airways of infants with CF, although further data are required to confirm this.

In vitro, *Streptococcus pneumoniae* outcompetes *H. influenzae* when co-cultured, particularly, in low pH environments²⁵⁰. Studies using a pig model of CF lung disease have demonstrated that airway surface liquid (ASL) is more acidic in CF due to defective bicarbonate transport as a result of absent or defective cystic fibrosis transmembrane regulator protein than in those with normal CFTR function. Reduced ASL pH was associated with reduced bacterial killing, suggesting that airway pH plays a critical role in the ability of the CF airways to evade infections⁹⁹. It is possible therefore, that in the airways of infants with CF, low ASL pH creates an environment which favours *Streptococcus* spp. growth, creating one mechanism by which it may outcompete *Haemophilus* spp. *Streptococcus* spp. are also found in high abundance in the airways of healthy infants²³ and their role in the CF airways is uncertain, with one study suggesting that it is a core genus in both the respiratory and gut microbiota in infants with CF¹⁴⁸. However, a recent study of 770 adolescents with CF found that *H. influenzae* exerts a protective effect on lung function, with preservation of lung function seen in adolescents with CF both following the acquisition and continuous infection with *H. influenzae*¹¹⁹. In addition, a study of patients with non-CF bronchiectasis showed that patients with *Haemophilus* spp. dominated communities experienced fewer pulmonary exacerbations²⁵¹. Therefore, therapeutic strategies which aim to maintain *Haemophilus* spp. levels, either through the use of specific antibiotics to target other community members and avoid *Haemophilus* spp. killing, or by maintaining airway surface pH, may be beneficial in maintaining lung health in infants with CF. However, the alternative explanation, that *Haemophilus* spp. may merely be less harmful, needs to be excluded before this therapeutic strategy is utilised.

Staphylococcus spp. and *Pseudomonas* spp., two of the organisms more commonly identified on bacterial cultures in children, had very low relative abundance in the microbiota in the first 2 years of life. Whilst the relative abundance of *Pseudomonas* spp. shows little change in the first 2 years of life, *Staphylococcus* spp. shows a significant increase in relative abundance at the end of the second year of life. However it should be noted that all children in this study received prophylactic anti-staphylococcal antibiotics (either flucloxacillin [N = 23] or Co-Amoxiclav [N = 7]) throughout the study period, in accordance with UK CF Trust Guidelines⁷². It is possible therefore that the low levels of *Staphylococcus* spp. seen reflect the suppression of *Staphylococcus* spp. growth as a consequence of the prophylactic antibiotics, the effects of which on the airway microbiota are currently unknown. The increase seen at two years of age may reflect decreasing adherence to flucloxacillin prophylaxis with increasing age.

Several studies of adults have found a high number of obligate and facultative anaerobes present in the airways of patients with CF^{14,144} which increase with age¹⁶⁷. Similarly in this study several obligate and facultative anaerobes were found in the airway microbiota of infants with CF, the most abundant of which was *Veillonella* spp. accounting for 7.36% of reads during the first 2 years of life (Section 4.3.3.2). With increasing age the relative abundance of *Veillonella* spp. was found to decrease whilst less abundant anaerobes, particularly *Prevotella* spp., showed a small but significant increase with age. The role of anaerobes in CF airways disease is currently unclear. Anaerobes have however been found to be present in the airways of healthy adults²⁷ and infants²³. Furthermore, a study

in adults with CF reported a correlation between increased levels of anaerobes and increased FEV₁ and decreased inflammation, measured by sputum neutrophil elastase and serum CRP¹⁷³. In contrast, increased levels of some anaerobes, such as *Gemella* spp.¹⁶⁰, *Veillonella* spp. and *Prevotella* spp.¹⁴⁴ have been found during pulmonary exacerbations with a recent study suggesting that *Prevotella* spp. may indirectly shield *P. aeruginosa* from antibiotic killing through Extended-Spectrum-Beta-Lactamase production²⁵². From the results in this study, it is possible that in the first few months of life the airways of infants with CF contain some protective anaerobes which are gradually replaced by genera, such as *Prevotella* spp., which individually may not be pathogenic but however promote the survival of pathogens, such as *P. aeruginosa*, once present. Further work is required to fully determine the role of anaerobes in CF airway disease and whether the addition of antibiotics targeting anaerobes may be beneficial in preventing disease progression.

Pulmonary exacerbations have been shown to be an important determinant of later FEV₁ in children with CF¹¹². During the 2-year follow-up period of this study, 7 (23%) infants experienced a pulmonary exacerbation. Due to a lack of a universally accepted definition of pulmonary exacerbation, this was defined as a change in symptoms necessitating an admission to hospital for IV antibiotics. Using this definition, however, limited the analysis of changes at exacerbations to only those patients with the most severe exacerbation.

To determine the changes in the microbiota at exacerbation, samples were compared at 4 timepoints: baseline, exacerbation, treatment and recovery.

Overall, there was no significant change in bacterial load, alpha or beta diversity between these timepoints, suggesting that changes in diversity are not relevant to pulmonary exacerbations (Section 4.3.3.6). At a genus level, the relative abundance of *Streptococcus* spp. was largely static at exacerbation (77.4% at baseline and 70.6% at exacerbation). Decreases were seen thereafter, with a sharp rate of decline seen between treatment and recovery samples; the prognostic significance of this decline merits further study. *Haemophilus* spp. showed mostly the opposite trend: a decrease from 12.5% at baseline to 0.84% of reads at exacerbation but recovery thereafter. Whilst these changes did not reach statistical significance (and it is acknowledged that this part of the study was underpowered), these trends suggest that *Streptococcus* spp. and *Haemophilus* spp. may have an antagonistic relationship such that *Haemophilus* spp. occupies the space left in the niche by *Streptococcus* spp. with antibiotic treatment. Little change was seen in the relative abundance of anaerobes at exacerbation. This may be due to the small numbers of patients who met the definition of pulmonary exacerbation used in this study. In addition, one limitation of this study is that, the definition for pulmonary exacerbations used may have missed changes in patients with milder symptoms who were treated with oral antibiotics but may still have had significant lung function decline²⁵³. Use of oral antibiotics was not included for two reasons. Firstly, because in current clinical practice there is a low threshold for initiating oral antibiotic treatment for even mild respiratory symptoms. Secondly, because this relied on retrospective parental reports of oral antibiotic usage in which there was a lack of confidence.

However, for each patient, individual barplots tracking changes in the microbiota with clinical variables (change in symptoms, microbial growth on culture and antibiotic use etc.) were plotted (Appendix A4). These did not identify a consistent bacterial biomarker associated with changes in upper and lower respiratory tract symptoms necessitating oral or IV antibiotics. However, it is important to note that the sample size in this study is small (N = 5 with serial samples). A longitudinal study with a larger sample size is therefore needed to explore further the changes occurring at pulmonary exacerbation.

Streptococci have previously been suggested as a biomarker for pulmonary exacerbations with one study reporting SMG to be the dominant pathogens in almost 40% of pulmonary exacerbations¹⁵⁹ with treatment targeted at *Streptococcus* spp. showing improvements in patients previously unresponsive to IV antibiotics in one case series¹⁵⁸. Given that the *Streptococcus* genus has itself been associated with clinical stability and diversity¹⁵⁷, its relationship to pulmonary exacerbations remains unclear. One limitation of the present study is that organisms could not be characterised at a species level. It is plausible that some species of *Streptococcus* spp., such as SMG, may be more pathogenic than others and may be associated with exacerbations. It has also been suggested that there may be a tipping point, at which levels of *Streptococcus* spp. lead to lung disease¹⁵⁷. Thus regular microbiological surveillance, combining both culture data and 16S rRNA gene sequencing, may be beneficial to identify changes in the relative abundance of *Streptococcus* spp. early so that targeted treatment towards reducing *Streptococcus* spp. can be considered. Whether changes in the relative abundance of *Streptococcus* spp. in TS are reflective of lower airway

changes in the microbiota is uncertain as illustrated in Chapter 3. Further study of the changes in the airway microbiota in childhood, which ideally would include PCR of specific *Streptococcus* genes to aid species discrimination and repeated lower airway sampling, with clinical correlations may help to further our understanding of the role of *Streptococcus* spp. in exacerbations in CF.

Seven infants grew *P. aeruginosa* for the first time during the study period. In all of these cases, the growth of *P. aeruginosa* was reported to be scanty. The relevance of scanty upper airway growths of *P. aeruginosa* to lower airway disease is unclear as frequently upper and lower airway cultures for

*P. aeruginosa*⁹⁸ do not correlate. To investigate whether changes in the microbiota occur prior to first *P. aeruginosa* growth, at the time of growth and following treatment for *P. aeruginosa*, serial samples at each of these timepoints were compared between patients (Section 4.3.3.7). Whilst there was no significant difference in bacterial load or alpha diversity, between samples difference in beta diversity was seen. This suggests that there are differences in community structure with growth and treatment of *P. aeruginosa*. Surprisingly, *Pseudomonas* spp. was an uncommon genus present in the samples taken from patients at the time of *P. aeruginosa* growth with a relative abundance of less than 0.1%. It is unlikely that this represents an error in sequencing, because *Pseudomonas* spp. was readily sequenced from the mock communities included in each sequencing plate. This could, however, reflect the high sensitivity of bacterial cultures for isolating *P. aeruginosa*, even when it is present in small numbers, the success of clinical microbiology protocols in actively seeking to identify *P. aeruginosa* when present, or the ability of *P. aeruginosa* to outgrow

other organisms when co-cultured, emphasising the benefit of obtaining results from both bacterial culture and 16S rRNA gene sequencing when analysing the airway microbiota.

At each timepoint in this study, patient sampling was paired with collection of detailed clinical data, including variables such as birth history, feeding history, genotype, antibiotic use and presence of upper and lower respiratory tract symptoms. These revealed that several clinical variables, such mode of delivery, exert a small but significant influence on the airway microbiota in the first 2 years of life (Section 4.3.3.5). In contrast to previous studies, alpha diversity was found to be higher in those delivered by LSCS than those delivered by SVD. This seems counterintuitive as SVD exposes the newborn infant to more microbes than LSCS. In this present study, there was no distinction made between emergency (in labour delivery following rupture of membranes) and elective (pre-onset of labour and rupture of membranes) LSCS due to the small total number of patients born by LSCS ($N = 6$). Future work comparing larger cohorts of patients born either by SVD, elective or emergency LSCS may reveal a different trend that was not possible in the current study design. In addition, whilst the results of this study suggest that mode of delivery at birth may influence community structure, they should be treated with caution, due to the very uneven sample sizes, as 70% of infants ($N = 21$) in this study were delivered by SVD.

There are several important limitations in this study. Firstly, whilst the number of participants is larger than previously reported studies of the oropharyngeal

microbiota, the number in this study is still small. During the 18-month recruitment period, the family of almost every infant with CF consented to take part in research. Thus, within the lifetime of this PhD, obtaining a greater sample size was not realistically feasible. Nevertheless despite the small sample size, changes in the microbiota with age were seen and the additional findings highlight that a further multi-centre study is warranted.

Another limitation of this study is the lack of a healthy control group. It was not feasible to include a healthy control group due to practical difficulties in recruiting healthy babies and ensuring regular, frequent attendances to RBH for TS collection during this PhD. Consequently it is difficult to know what changes seen in this study are related to CF disease processes or may be normal changes occurring in infants during the first 2 years of life. Studies of the nasopharyngeal microbiota in infants with CF compared with healthy controls however suggest that there are differences in the microbiota between infants with CF and healthy infants from as early as 2 months of age with higher abundances of *S. aureus* and *Streptococcus mitis* in CF and increased *H. influenzae* and *Moraxella* spp. in healthy controls¹⁴⁹. It is plausible therefore that differences would also be seen in the oropharyngeal microbiota between infants with CF and healthy controls. Whilst infants with PCD could be considered as a comparator group, the average age at diagnosis in PCD is higher at 5.3 years¹⁸³, thus precluding a direct comparison between these groups. In addition, all of the CF children in this study received anti-staphylococcal prophylaxis which may have impacted community structure. This prophylaxis is not a policy adopted worldwide and the benefit of antibiotic prophylaxis in children with CF is still under debate.

Therefore, a future study comparing longitudinal changes in the microbiota between infants with CF on antibiotic prophylaxis with those receiving no prophylaxis, such as in the CF START study (www.cfstart.org.uk), would provide real insights into the role of antibiotic prophylaxis in airway infections in CF.

Ideally the participants in this study would have been recruited as soon as possible following diagnosis, ideally within the first 1-2 months of life. Whilst many families are admitted for 2 days following diagnosis for education, it was felt to be insensitive to recruit participants during this highly emotional experience. Participants were therefore recruited at their next routine clinical appointment following their educational visit. This resulted in variability in the age of recruitment. In addition, patient sampling was performed opportunistically during routine clinical appointments by a single operator (B. Ahmed). Whilst this ensured consistent sampling techniques, this sometimes resulted in missed samples at a given timepoint, a limitation of the present study.

In conclusion, in accordance with the study aims, this chapter reports the largest longitudinal study of the oropharyngeal microbiota in NBS infants to date, with more frequent follow-up and one of the longest periods of follow-up. Overall, there was little change in bacterial load, the number of different organisms and the spread of those organisms with increasing age, pulmonary exacerbations or first growth of *P. aeruginosa*. This contradicts the study hypothesis that changes in the microbiota would be seen with clinical status in the first two years of life. In contrast, the changes seen in community structure with age, measured by beta diversity and specific genus level changes, provide some evidence to support the

hypothesis. An inverse relationship in the relative abundance of *Streptococcus* spp. and *Haemophilus* spp. was seen both with age and pulmonary exacerbation suggesting a potentially competitive relationship between these organisms within the niche of the CF infant airways. This suggests that the interactions between organisms in the airway microbiota might be important in exacerbations in infants with CF. The airway microbiota seen at 2 years of age does not reflect that seen reported in CF adults indicating that at 2 years of life the airway microbiota in CF is still being established. Longitudinal study of older children with CF is therefore needed to better understand the relationship between the airway microbiota and clinical outcomes in CF. This will therefore be explored in the next chapter.

Chapter 5: Comparison of the airway microbiota in Cystic Fibrosis and Primary Ciliary Dyskinesia

5.1 Introduction

The aim of this chapter was to assess whether changes in the airway microbiota occur with age and changing clinical status in children with CF and PCD who were recruited as part of the CLIMB study described in Chapter 2, section 2.2. CF and PCD share a similar pathology of airway infection and neutrophilia and respectively secondary and primary impairment of mucociliary clearance¹⁹⁴. The prognosis however differs, with a slower rate of decline of lung function and longer survival in PCD¹⁹⁷. The airway microbiota in PCD has been little studied but differences between it and that of CF may relate to the differences in prognosis for the two diseases. If such differences in the airway microbiota exist between patients with CF and those with PCD, this could identify new therapeutic strategies that convert CF from a severe to a milder airway disease. Thus a second aim was to assess whether the differences in prognosis between CF and PCD are associated with differences in the airway microbiota over time and in response to treatments, such as antibiotics. The objective was to recruit prospective cohorts of children with CF (excluding NBS) or PCD and collect TS at three month intervals for up to 2 years at times to describe the longitudinal changes in the airway microbiota in CF and PCD with clinical features, including pre- and post- IV antibiotic treatment. In addition, the differences in the microbiota between CF and PCD were evaluated.

5.2 Methods

5.2.1 Patient recruitment and follow-up

Children with CF and PCD were recruited as part of the CLIMB study. Full details of the CLIMB study are given in Chapter 2, Section 2.2. At the beginning of the study (December 2012), 330 children with CF and 150 children with PCD received care under the Paediatric Respiratory Department at RBH. The majority of these children received shared care at RBH and their local hospital and therefore had fewer appointments at RBH. Based on records from the Paediatric Respiratory Department, 90 children with CF and 50 children with PCD were seen exclusively at RBH at the beginning of the study, including fifteen children with more progressive CF disease who were listed for elective admissions every 3 months for IV antibiotics, irrespective of clinical status. The aim therefore was to recruit as many of these children as possible. Diagnostic and inclusion criteria for CF and PCD are described in Chapter 2, Section 2.2). Generally, children with CF were seen routinely every 2-3 months, whereas children with PCD were seen every 6 months (see Chapter 2 Section 2.2). Samples were collected opportunistically during routine clinical appointments.

5.2.2 Clinical information

Clinical information (such as infection status, lung function etc.) was collected as described previously in Chapter 2, Section 2.3 and is given in Chapter 2, Table 2.2. Median change in FEV₁% values during the study period was calculated by calculating the difference between the first and last sample taken divided by the duration of follow-up for each individual patient.

As previously (Chapter 2, Section 2.3), the definition of a pulmonary exacerbation was defined as a deterioration in clinical status necessitating IV antibiotics, as judged by a paediatrician independent of the study. For some patients, a two-week course of IV antibiotics was initiated in hospital and then continued at home. Due to uncertainty in the reliability of retrospective parental reports regarding duration of oral antibiotic administration, oral treatment courses of antibiotics were not included in the analyses. Presence of bronchiectasis was determined by a Radiologist on CT scans performed for clinical reasons.

5.2.3 Sample collection & processing

TS were collected from patients during routine clinical appointments (outpatient and inpatient) as described in Chapter 2, Section 2.2. In addition, spontaneously expectorated sputum was collected when available. Patients were asked to expectorate sputum into two sterile universal containers. The first container was sent for routine clinical microbiology and kept at room temperature until transfer to the Clinical Microbiology laboratory at RBH. The second container was immediately frozen on dry ice and stored at -80°C for culture independent analysis. The majority of samples (393 samples [96%]) were collected by a single operator (B. Ahmed). Samples were processed as previously described in Chapter 2, Section 2.4.

For DNA extraction, whole sputum samples were defrosted and 300 µl was aliquoted and added directly to a LME tube (Chapter 2, Section 2.5). DNA extraction and library processing for 16S rRNA gene sequencing followed the methods described in Chapter 2, Sections 2.5 – 2.12.

5.2.4 Data analyses

From sequencing data obtained for each sample, forward and reverse reads were combined and quality filtered as previously described in Chapter 2, Section 2.13.1. Examining rarefaction curves, an asymptote is reached by 1,000 reads, suggesting at this level an accurate representation of the microbiota is obtained (Figure 5.1) without compromising the number of samples retained following rarefaction. Samples were therefore rarefied to 1,000 reads prior to downstream analyses to standardise sequencing depth. Four hundred and nine samples were collected; 76 samples (18.6%) had a read count below this threshold and were therefore removed at rarefaction.

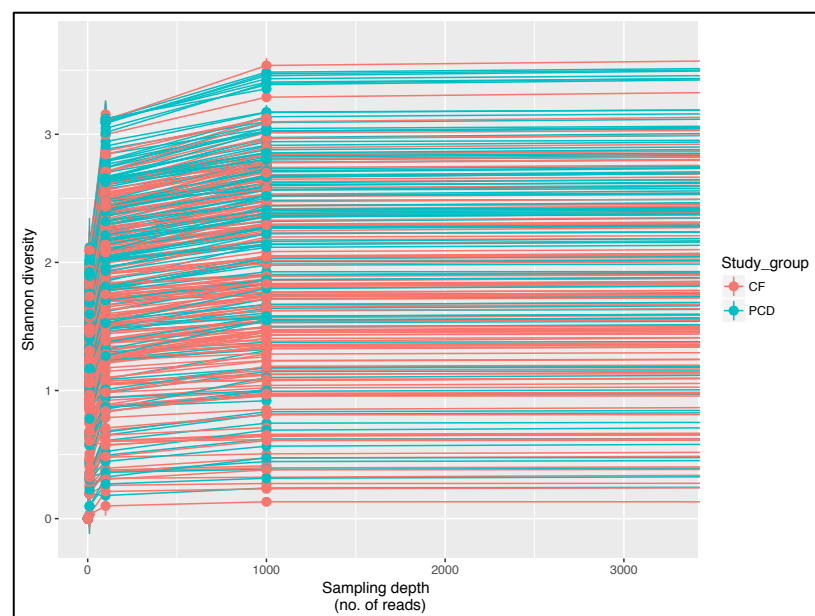
Downstream data analyses were performed in phyloseq in R (Chapter 2, Section 2.13.2). Longitudinal changes in the relative abundance of OTUs were compared both within and between disease groups. For OTU level analyses, Sankey plots were constructed as previously described in Chapter 4, Section 4.2.4. Individual patient plots, illustrating barplots of the most common OTUs against changes in FEV₁ % (for children aged over 5 years), the Inverse Simpson's diversity index, culture results and presence or absence of lower airway symptoms were constructed using ggplot2 and gridExtra. Genera and OTU level differences as

well as alpha diversity testing to assess the relationship between alpha diversity and age and clinical variables were assessed as previously described in Chapter 4, Section 4.2.3.

For beta-diversity testing, to allow comparisons between paired samples longitudinally, a block study design was used when using the adonis function using the strata argument of this function. Analyses were blocked either by participant study number (to analyse the influence of age on beta diversity) or by age range (to analyse the influence of clinical variables on beta diversity). A *P*-value of less than 0.05 was considered to be statistically significant.

Figure 5.1: Justification for rarefaction to 1000 reads. Illustrating the rarefaction curve calculated using Shannon Diversity Index measurements.

The asymptote of the rarefaction curve illustrates the sequencing threshold at which the majority of OTUs from the sample community are identified. At 1,000 reads, an asymptote was reached, providing justification for rarefaction to this level.



5.3 Results

5.3.1 Patient demographics

For the longitudinal comparison of the airway microbiota in CF versus PCD, 31 children (age matched) were studied in each diagnostic category. If more than one child with CF matched to a child with PCD, the CF child who was closest in age to the PCD child was selected. Within the CF group there were seven children who received elective IV antibiotics whilst in the PCD group there were three. Baseline demographics are shown in Table 5.1.

The mean age of participants was 9.3 years (sd 4.8 years) for CF and 9.9 years (sd 4.6 years) for PCD. Eighteen (58%) children with CF and 20 (65%) children with PCD were female. Patients were followed-up for a shorter period than 2 years as initially intended, with a mean duration of 1.1 years in each group (SD 0.46 years and 0.48 years for CF and PCD respectively). There were no significant differences in BMI-Z scores, FEV₁% at baseline, median FEV₁% change during the study period or FVC% between CF and PCD (Table 5.1).

The majority of children with CF were either homozygous for the p.Phe508Del mutation (55%) or heterozygous (29%). Mutation class information could only be found for three of the patients who did not have a p.Phe508Del mutation; two of these patients had Class I mutations (one who was p.Arg1162X homozygous and one who was heterozygous for p.Gly542X and 1717-1G>A) and one who had a complex mutation which had properties of Class II, III and VI mutations (p.Asn1303Lys). Twenty-eight (90%) patients with CF were pancreatic insufficient, two (6%) had CF related diabetes (CFRD), five (16%) had CF liver

disease, one (3%) had allergic bronchopulmonary aspergillosis (ABPA) and four (13%) had nasal polyposis. Twenty two (81%) children with PCD had abnormalities in ciliary structure with the remainder diagnosed with PCD based on clinical phenotype, low nasal nitric oxide (nNO), abnormal ciliary beat frequency or pattern abnormalities on light microscopy or confirmation of disease causing mutations (Table 5.1).

There were significantly more children with GORD (definition in Chapter 2, Section 2.3) in CF than in PCD (42% and 3% respectively, $P = 0.0003$) as diagnosed by the treating clinician. Ten patients with CF had CT chest performed prior to recruitment, of whom eight had bronchiectasis and five received elective IV antibiotics every 3 months and therefore were patients with worse disease. Eleven patients with PCD had a CT chest performed prior to or at recruitment, of which 10 showed bronchiectasis. No patients with CF or PCD had CTs performed during the follow-up period.

There was a significant difference in *P. aeruginosa* infection status at baseline. Twenty-five (81%) PCD patients had never grown *P. aeruginosa* whereas twenty-five (81%) CF patients had grown *P. aeruginosa* prior to recruitment, with six patients (19%) with CF and no patients with PCD chronically infected with this organism. There were several significant differences in baseline microbiology between CF and PCD. More patients with PCD had grown *H. influenzae* and *Streptococcus pneumoniae* prior to recruitment; more patients with CF had grown methicillin sensitive *S. aureus* (MSSA), non-tuberculous mycobacteria (NTM) and *Aspergillus fumigatus* prior to recruitment. Organisms

grown during the study period are listed in Table 5.2. Five organisms had different rates of positive culture between CF and PCD during the study:

P. aeruginosa ($P = 0.03$) and *Aspergillus fumigatus* ($P = 0.001$) were grown more frequently in CF than PCD, and *Streptococcus pneumoniae* ($P = 0.01$),

H. influenzae ($P < 0.0001$) and “upper respiratory tract flora” ($P < 0.0001$) were grown more frequently in PCD than CF.

Over half the patients in each disease group were admitted for IV antibiotics during the study period with eighteen patients (58%) with CF admitted for either emergency IV antibiotics for an exacerbation (35%) or elective 3 monthly IV antibiotics (23%). Seventeen patients (55%) with PCD were admitted for either emergency IVs (45%) or elective IVs (10%).

Table 5.1: Patient demographics comparing children with CF and PCD.

Values are given as mean (SD) or number (percentage, %). Significant differences between groups ($P < 0.05$) are shown in bold. NS - non-significant ($P > 0.05$); BMI - body mass index; MSSA - Methicillin sensitive *S.aureus*; MRSA - Methicillin resistant *S.aureus*; BCC - *Burkholderia cepacia* complex; NTM - non-tuberculous mycobacteria; GORD – gastro-oesophageal reflux disease.

Demographic	CF	PCD	P-value
No. of participants	31	31	N/A
Age (years)	9.3 (4.8)	9.9 (4.6)	NS
Gender (female)	18 (58%)	20 (65%)	NS
BMI (z-score)	-0.222 (0.969)	-0.149(1.29)	NS
Baseline FEV1 (%)	76.5 (16.9)	79.3 (17.9)	NS
Median FEV1% change (range)	2.6 (-18.5 – 138.7)	0 (-34.5 - 146.9)	NS
Baseline FVC (%)	91.8 (19.8)	82.0 (16.0)	NS
Genotype		N/A	N/A
p.Phe508Del homozygous	17 (55%)		
p.Phe508Del heterozygous	8 (26%)		
Other	6 (19%)		
Ciliary structure	N/A		N/A
Inner & outer dynein arm defect		7 (23%)	
Outer dynein arm defect		9 (29%)	
Inner dynein arm & other		3 (10%)	
Other		3 (10%)	
Normal electron microscopy		9 (29%)	
Pancreatic insufficient	28 (90%)	N/A	N/A
GORD	13 (42%)	1 (3%)	0.0003
Clinical microbiology			
<i>P. aeruginosa</i> infection status			< 0.0001
Chronic	6 (19%)	0 (0%)	
Intermittent	6 (19%)	2 (6%)	
Free	13 (42%)	4 (13%)	
Never	6 (19%)	25 (81%)	
Previous organisms grown:			
<i>H. influenzae</i>	3 (10%)	17 (55%)	< 0.0001
MSSA	14 (45%)	6 (19%)	0.030
MRSA	3 (10%)	1 (3%)	NS
<i>Streptococcus pneumoniae</i>	0 (0%)	10 (32%)	0.001
BCC	1 (3%)	0 (0%)	NS
NTM	6 (19%)	0 (0%)	0.012
<i>Stenotrophomonas maltophilia</i>	3 (10%)	0 (0%)	NS
<i>Serratia marcescens</i>	1 (3%)	0 (0%)	NS
<i>Aspergillus fumigatus</i>	11 (35%)	0 (0%)	0.0003
Exacerbations & treatments			
Inhaled steroid use	11 (35%)	8 (26%)	NS
Prophylactic antibiotics	22 (71%)	15 (48%)	NS
Exacerbations IV antibiotics	11 (35%)	14 (45%)	
Elective IV patients	7 (23%)	3 (10%)	

Table 5.2: Organisms grown on clinical microbiology during study period for patients with CF and PCD.

MSSA - Methicillin sensitive *S. aureus*; MRSA - Methicillin resistant *S. aureus*;

NTM - Non-tuberculous mycobacterium; URT - upper respiratory tract.

Significant differences ($P < 0.05$) are shown in bold.

Organism name	CF (No. of patients, %)	PCD (No. of patients, %)	P-value
<i>P. aeruginosa</i>	10 (32%)	3 (10%)	0.03
Mucoid <i>P. aeruginosa</i>	3 (10%)	0 (0%)	NS
MSSA	4 (13%)	4 (13%)	N/A
MRSA	1 (3%)	0 (0%)	NS
<i>Streptococcus pneumoniae</i>	0 (0%)	6 (19%)	0.01
<i>H. influenzae</i>	0 (0%)	14 (45%)	< 0.0001
<i>Stenotrophomonas maltophilia</i>	1 (3%)	1 (3%)	N/A
<i>Serratia marcescens</i>	2 (6%)	0 (0%)	NS
NTM	2 (6%)	0 (0%)	NS
“URT flora”	1 (3%)	23 (74%)	< 0.0001
<i>Moraxella catarrhalis</i>	0 (0%)	1 (3%)	NS
Coliforms	1 (3%)	1 (3%)	N/A
<i>Aspergillus fumigatus</i>	9 (29%)	0 (0%)	0.001

5.3.2 Sampling frequency

A median of seven samples per patient (range 2- 23 samples) were collected from CF patients with a median interval of 42 days between samples (range 1- 238 days). A median of five samples per patient (range 2 – 15 samples) were collected from PCD patients with a median interval of 61 days between samples (range 1 – 403 days) (Figure 5.2). Many patients with PCD had less frequent follow-up than patients with CF so frequency of sampling was significantly higher in CF than in PCD ($P = 0.004$). Sampling frequency for each group per study month is shown in Figure 5.3. Whilst it was planned for TS to be collected at each sample visit, this proved not feasible for all patients either due to non-attendance at routine clinical appointments or because of unavailability of the sample collector. PCD patients who were able to expectorate were often less keen to have TS taken in addition to sputum. Consequently more sputum samples ($N = 91$) than TS ($N = 65$) were collected for PCD patients, with the reverse being the case for CF patients (TS: $N = 171$; sputum: $N = 82$) (Figure 5.2).

To account for the more frequent sampling in children with CF than those with PCD and allow the greatest number of comparisons between the study groups, sampling time points were combined into 3-month age ranges (baseline, 3- 6 months, 9 – 12 month, 15 – 18 months) (Figure 5.4). Where a patient had more than one sample in a given time period, the first sample taken in that range was used for comparisons between patients for analyses where age was analysed as a categorical variable; all samples at all timepoints were included for analyses where age was analysed as a continuous variable.

Figure 5.2: Summary of the number of samples in each study group (CF or PCD) before and after rarefaction to 1000 reads.

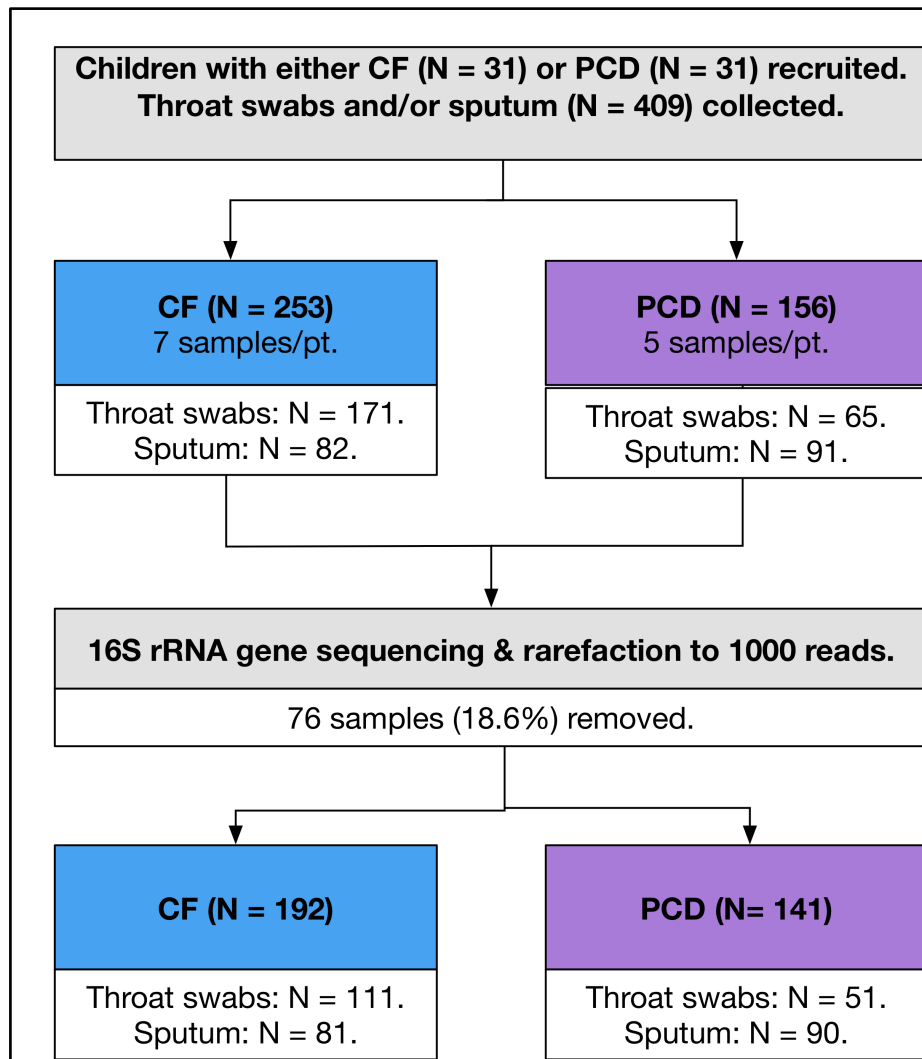
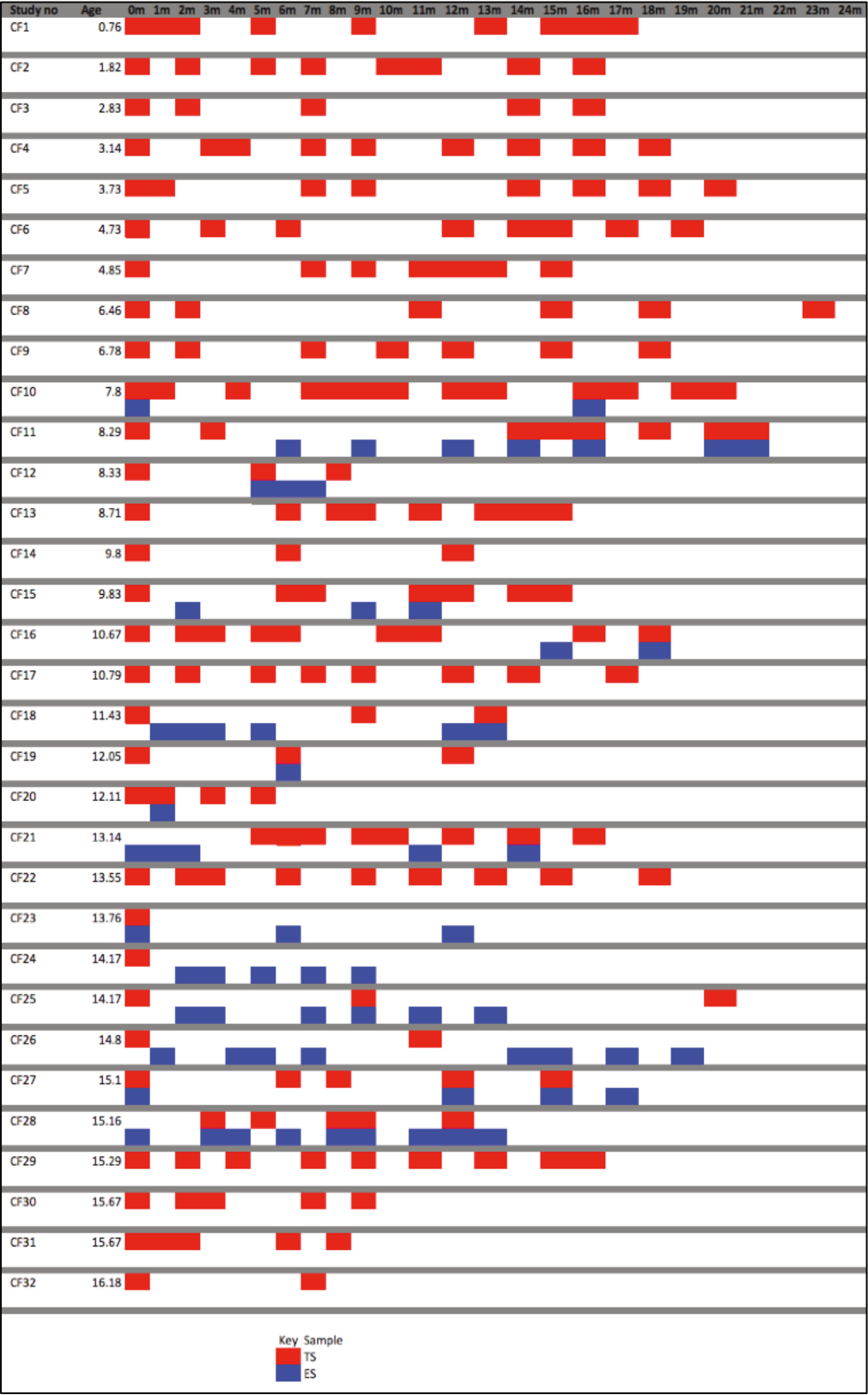


Figure 5.3: Frequency of samples taken per month of the study duration for (a) Patients with CF and (b) Patients with PCD organised in ascending age.
 Red indicates a timepoint at which TS was collected; blue indicates sputum (ES) collected.



(b)

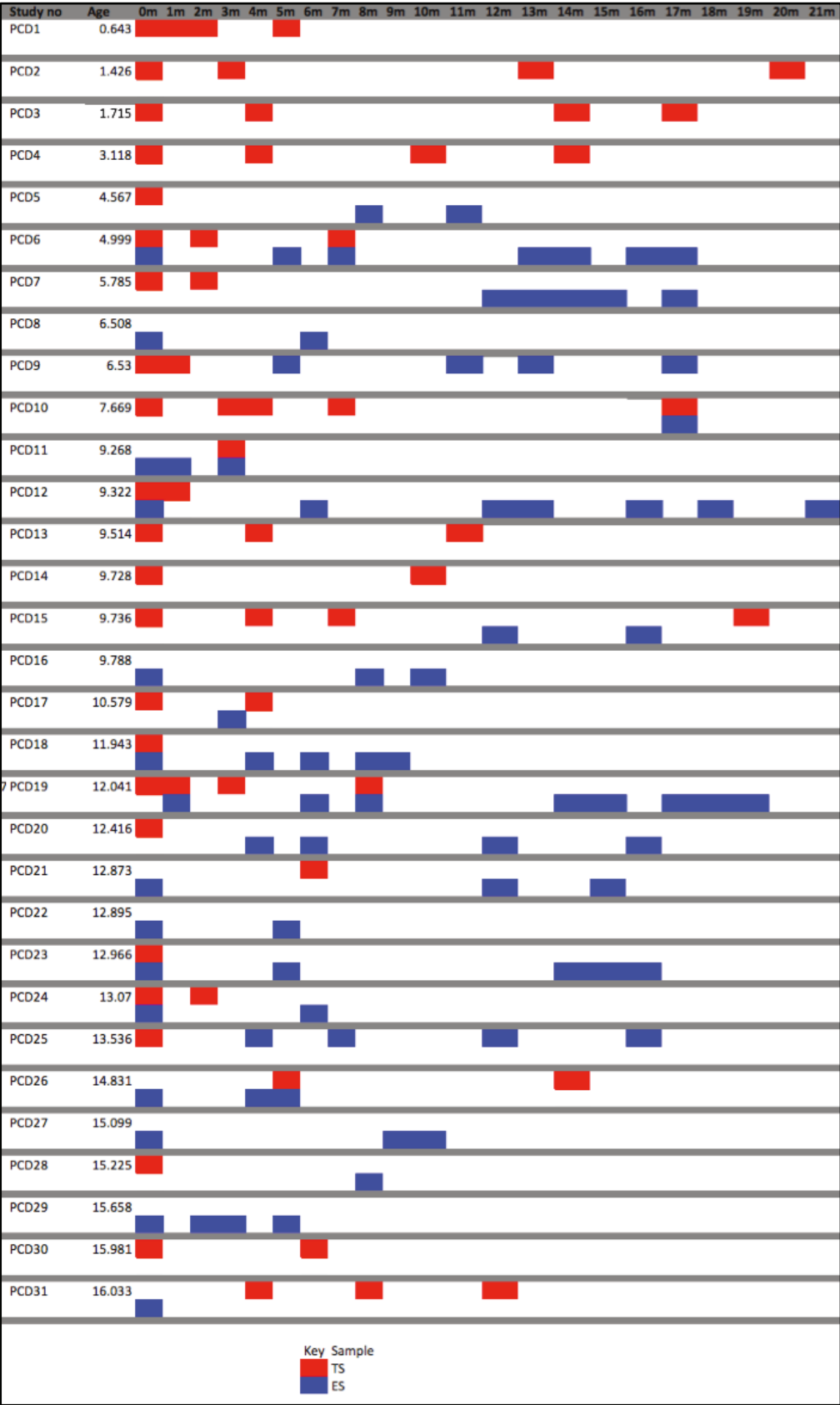
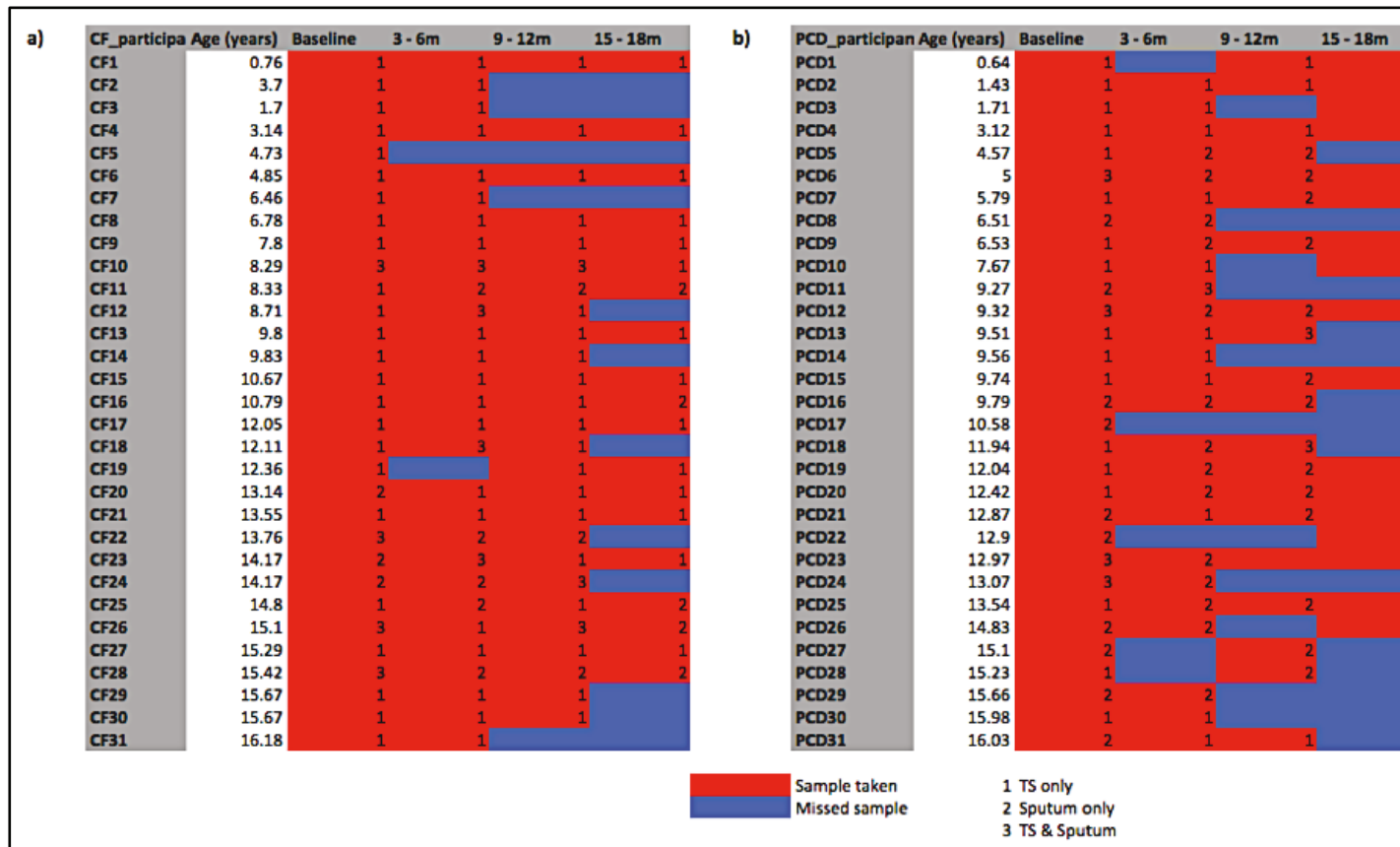


Figure 5.4: Sampling frequency grouped into four study timepoints: Baseline, 3-6 months, 9-12 months and 15-18 months.

(a) for patients with CF and (b) for patients with PCD. Red indicates a timepoint at which a sample was obtained; blue indicates a missed sample. “1” indicates a timepoint where TS was collected only; “2” where sputum was collected only; “3” where both TS and sputum were collected.



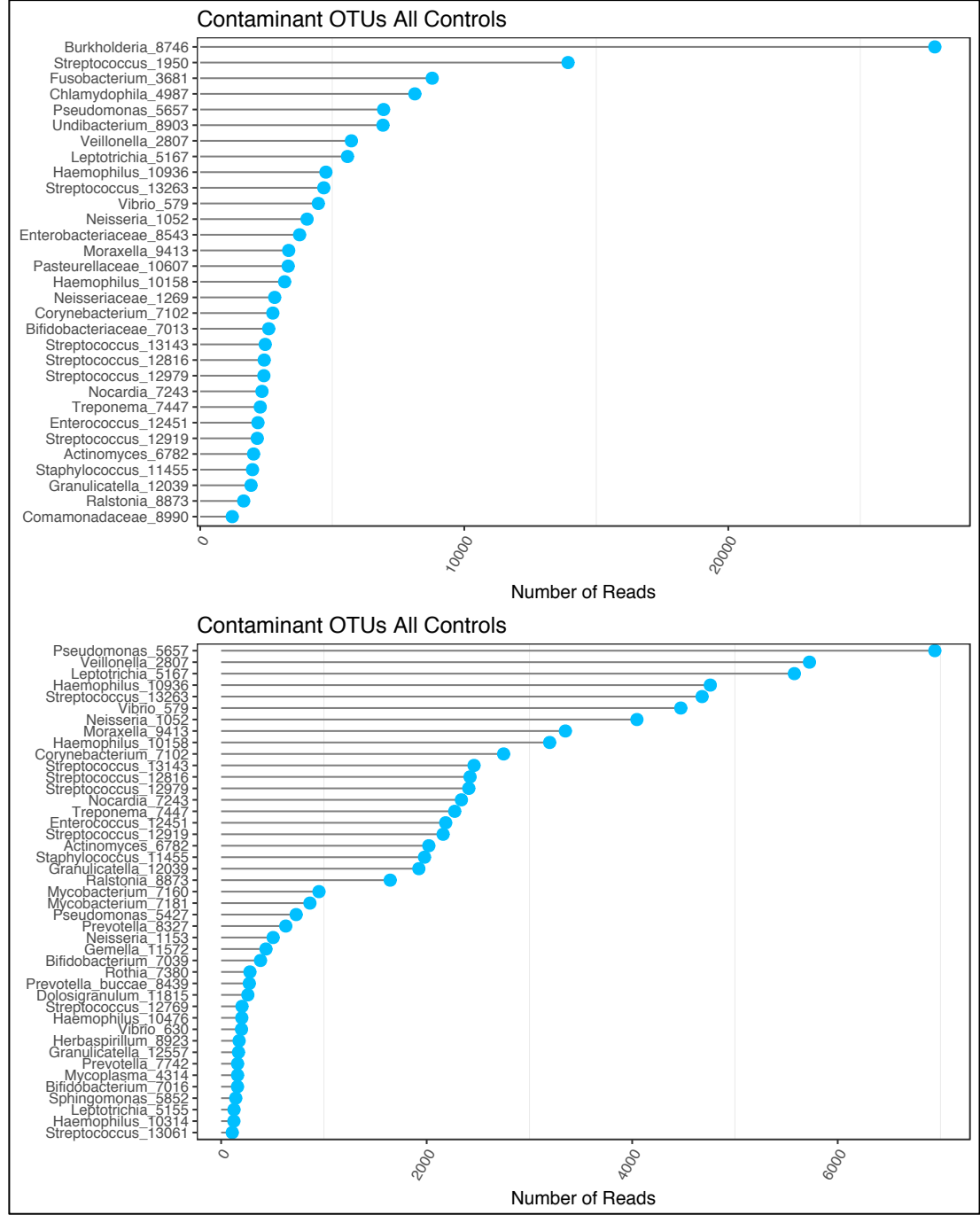
5.3.3 Examining sequencing quality

Four hundred and nine patient samples from the 62 patients were sequenced on 10 plates alongside 20 PCR controls (10 mock communities and PCR negative controls each) and 13 kit controls (4 technical controls from blank swabs and 9 DNA extraction kit controls). 35,923,028 reads were obtained after combining forward and reverse reads, de-multiplexing, quality filtering and removal of PhiX, with a median of 18,324 reads obtained per patient sample (range 82 - 348,657 reads). Examining run quality statistics, all the sequencing runs achieved the required thresholds.

Upon examining PCR negative, DNA extraction and technical controls, the following were identified as contaminants based on the previous literature²¹⁰ and were therefore removed: *Undibacterium* spp., *Comamonadaceae*, *Chlamydophila* (OTU ID 4987) and *Burkholderia* (OTU ID 8746) (Figure 5.5).

Figure 5.5: Contaminant identification & removal.

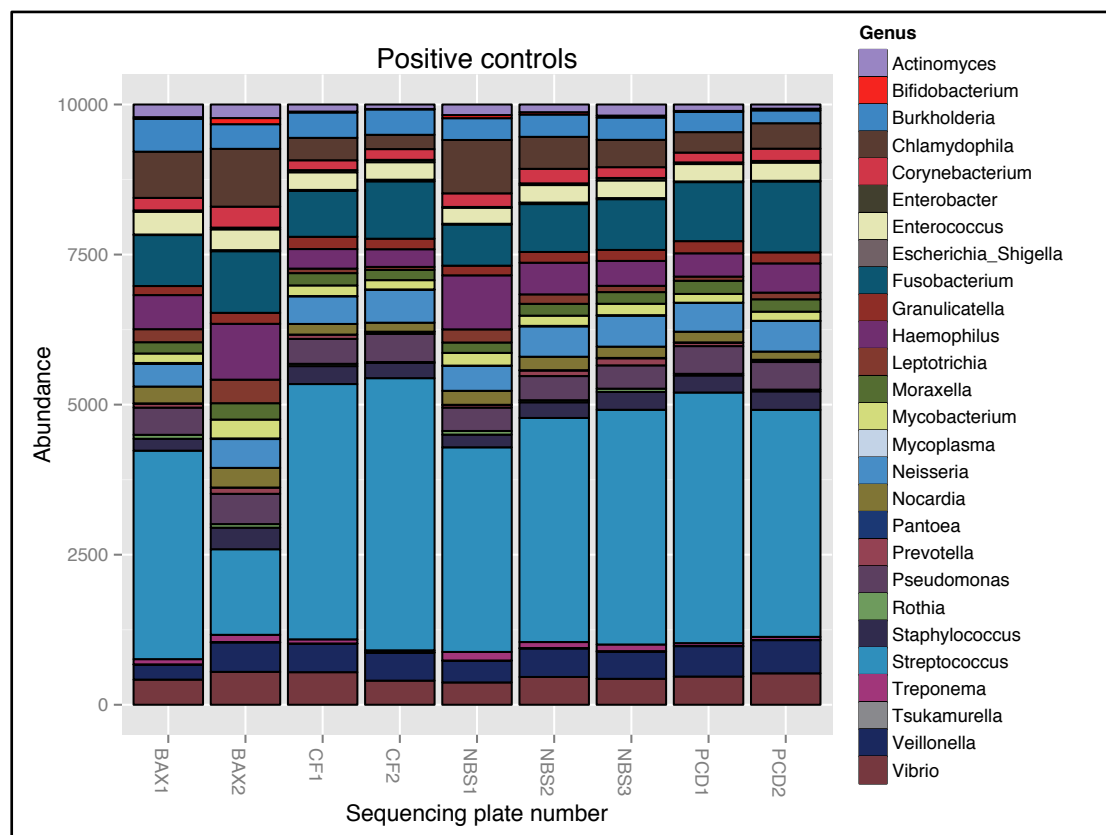
This illustrates the most common OTUs present within PCR negative, technical sampling and DNA extraction kit control samples. The top and bottom images show the OTUs present before and after contaminant removal respectively. Contaminants removed were: *Undibacterium spp.*, Comamonadaceae, a *Chlamydomphila* OTU (ID 4987).



NMDS plots showed consistency between all runs. A small significant batch effect was seen using the unweighted UniFrac score ($r^2 = 0.07$, $P = 0.003$) but not with the Bray Curtis ($r^2 = 0.15$, $P = 0.052$) or the weighted UniFrac scores ($r^2 = 0.04$, $P = 0.855$). When examining barplots of the mock community, a reduction in *Streptococcus* spp. can be seen in BAX2, as noted previously in Chapter 3 (Section 3.3.2). The other sequencing plates showed similarity between the mock communities (Figure 5.6).

Figure 5.6: Stacked barplots comparing mock communities of each sequencing run to assess for consistency between sequencing runs.

As previously identified (Chapter 3 Section 3.3.2) a reduction in *Streptococcus* spp. in the mock community of run BAX2 exists but for all other sequencing runs the mock community is consistent.



5.3.4 Spontaneously expectorated sputum versus throat swabs

As highlighted earlier (Section 5.3.2) for some time points only sputum was collected with no corresponding TS. In the early comparison of the upper and lower airway microbiota (Chapter 3), sputum was only available for two of the patients in the study group (mostly because of their young age) and consequently it had not been feasible to conduct a comparison of the microbiota between sputum and lower airway samples.

In this chapter, for the 62 patients being studied 16 had paired sputum and TS samples collected at the same clinical appointment (9 patients with PCD and 7 with CF). These paired samples were therefore investigated to determine whether sputum and TS could both be used equivalently to assess longitudinal changes in the airway microbiota. Comparing all TS combined and all sputum combined, the four most common genera were found to be identical and were: *Streptococcus* (47.5% and 27.1% of total reads respectively), *Veillonella* (13.6% and 18.8%), *Prevotella* (7.0% and 11.0%) and *Neisseria* (5.6% and 3.9%). *Pseudomonas* was the fifth most common genus in sputum (3.9%) but found to be less abundant in TS (0.5%). Examining a heatmap of the fifty most common OTUs between individual TS and sputum (Figure 5.6), however, a difference in the relative abundance of *Pseudomonas* was not apparent. In four sputum and six TS samples *Pseudomonas* was one of the fifty most common OTUs present with a similar relative abundance in these samples, with the exception of one sputum sample, which had a higher relative abundance of *Pseudomonas* (Figure 5.8).

Individual patient barplots demonstrated similarities between TS and sputum for the majority of patients (75%, 12 out of 16) (see Figure 5.8 as an example; individual patient barplots shown in the Appendix A5). A minority (25%, 4 out of 16) of patients showed higher relative abundances of *Pseudomonas* (N = 2), *Staphylococcus* (N = 1) and *Moraxella* (N = 1) in sputum than in TS. Using Spearman's rank to compare samples at a genera level, there is strong correlation between the relative abundance of genera present between TS and sputum ($r = 0.75$, $P < 0.0001$). Similarly, there was no significant difference seen in the relative abundance of genera between paired sputum and TS using multiple paired t-tests with Benjamini Hochberg correction ($P_{\text{adj}} > 0.05$).

Figure 5.7: Heatmap comparing the relative abundance of OTUs in sputum and TS.

Samples organised by sample type and ascending patient study number. *Pseudomonas* spp. is highlighted in red to illustrate the similarities in relative abundance of *Pseudomonas* spp. in sputum and TS samples, with the exception of one sputum sample (emphasised with a red box), which had a higher relative abundance of *Pseudomonas* spp.

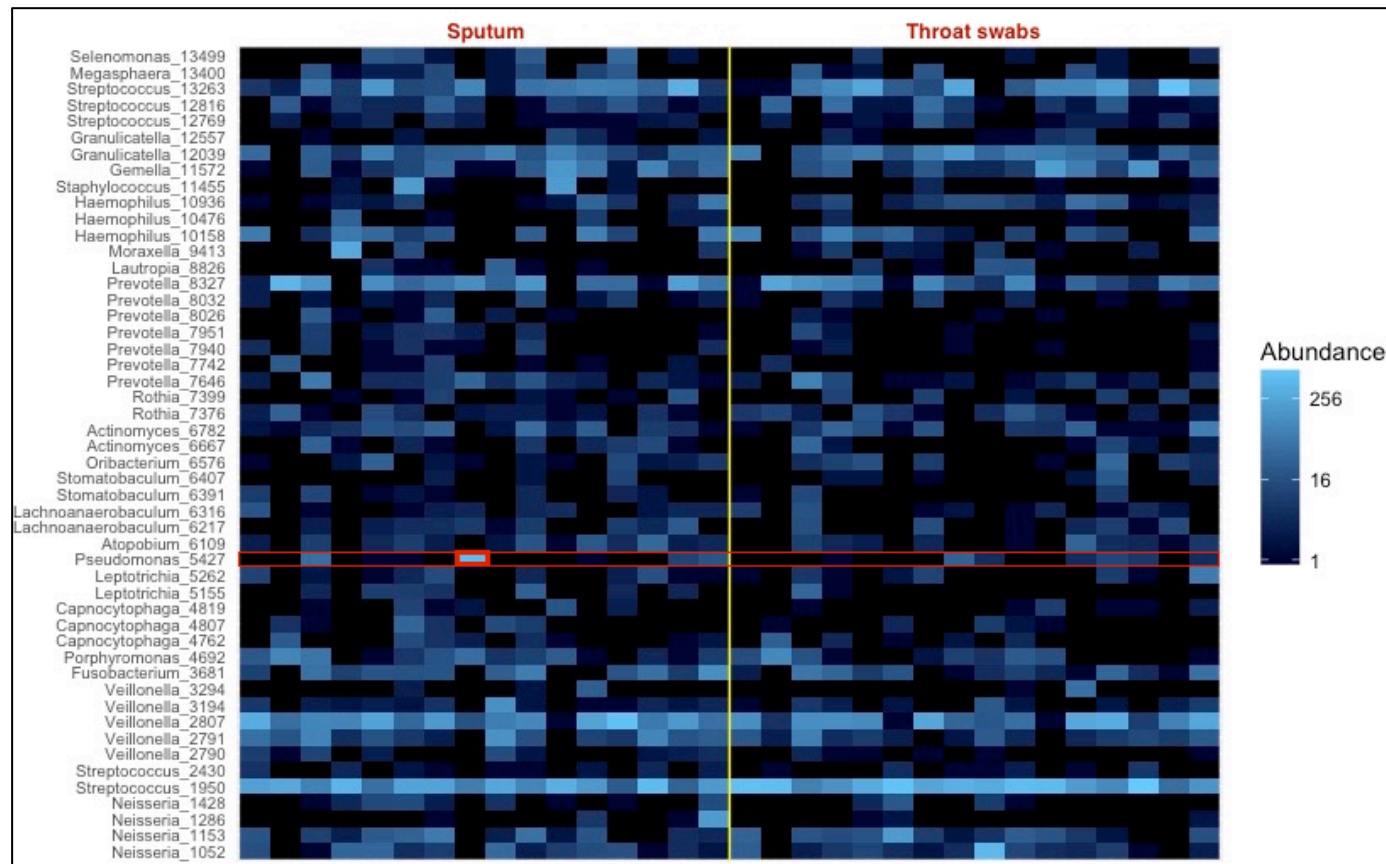
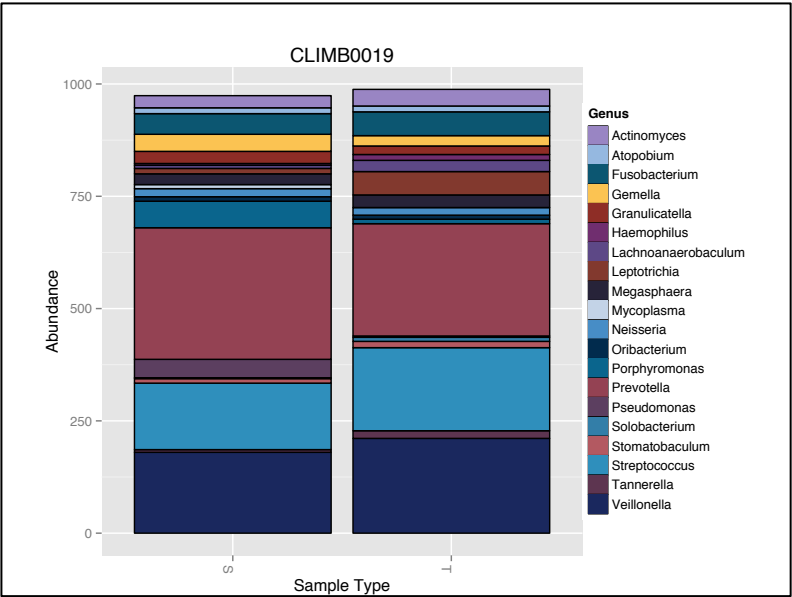


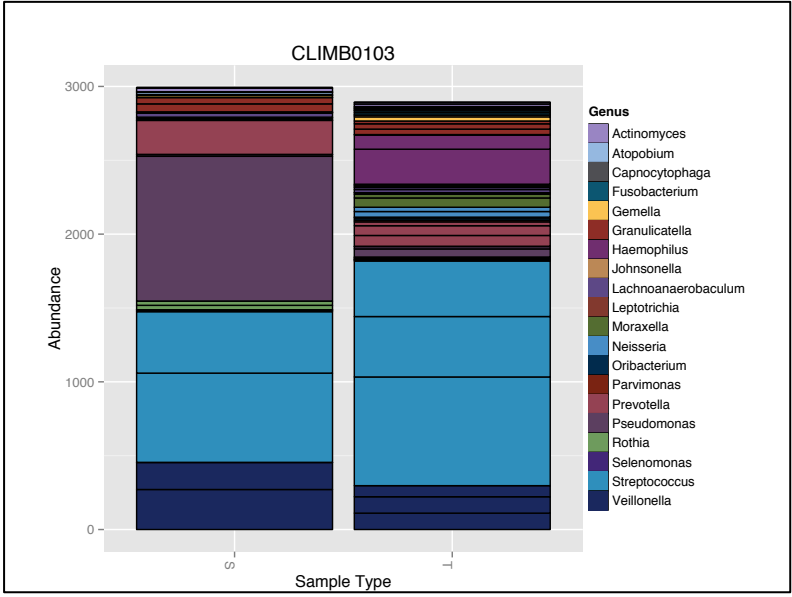
Figure 5.8: Example individual patients' barplots for expectorated sputum versus throat swab comparison.

- (a) Example of a patient in whom sputum and TS appeared similar. Below
(b) Example of a patient in whom *Pseudomonas* spp. have a higher relative abundance in sputum than in TS.

(a)



(b)



No significant difference in alpha diversity (richness, evenness, Shannon or the Inverse Simpson's diversity index) was seen between TS and sputum for patients with CF ($P > 0.05$). Whilst richness was significantly lower in TS than in sputum for patients with PCD ($P = 0.011$), no significant difference was seen in evenness, Shannon diversity or the Inverse Simpson's diversity index ($P > 0.05$) (Figure 5.9).

There were no significant differences between TS and sputum for either CF or PCD in relation to beta diversity using the Bray Curtis dissimilarity, unweighted UniFrac or weighted UniFrac scores (Table 5.3).

Figure 5.9: Comparing richness between sputum and TS.

CF samples are shown above and PCD below demonstrating no significant difference between samples for CF but a significantly lower richness in TS in comparison with sputum for PCD.

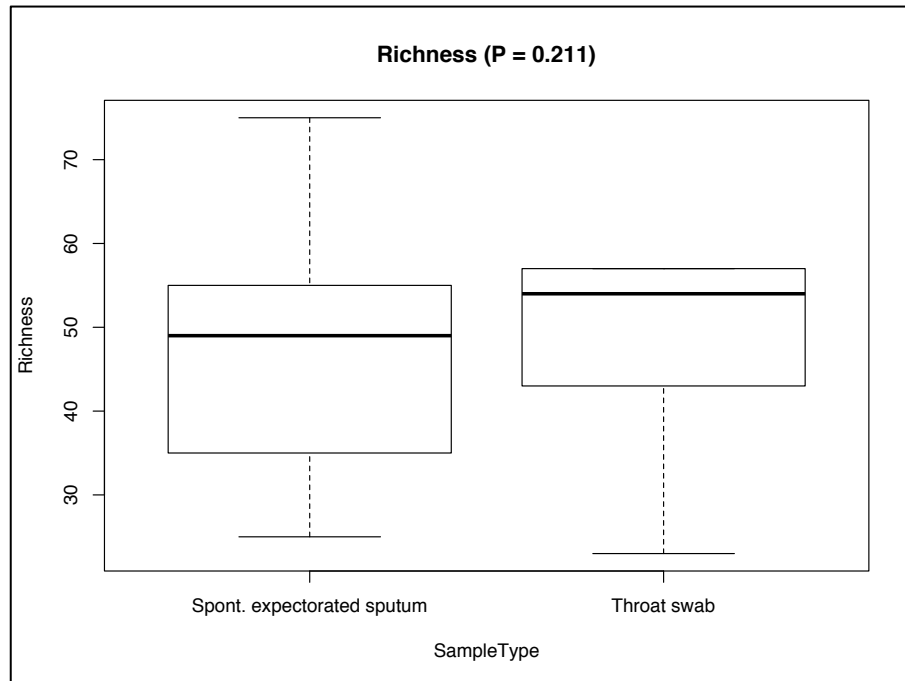
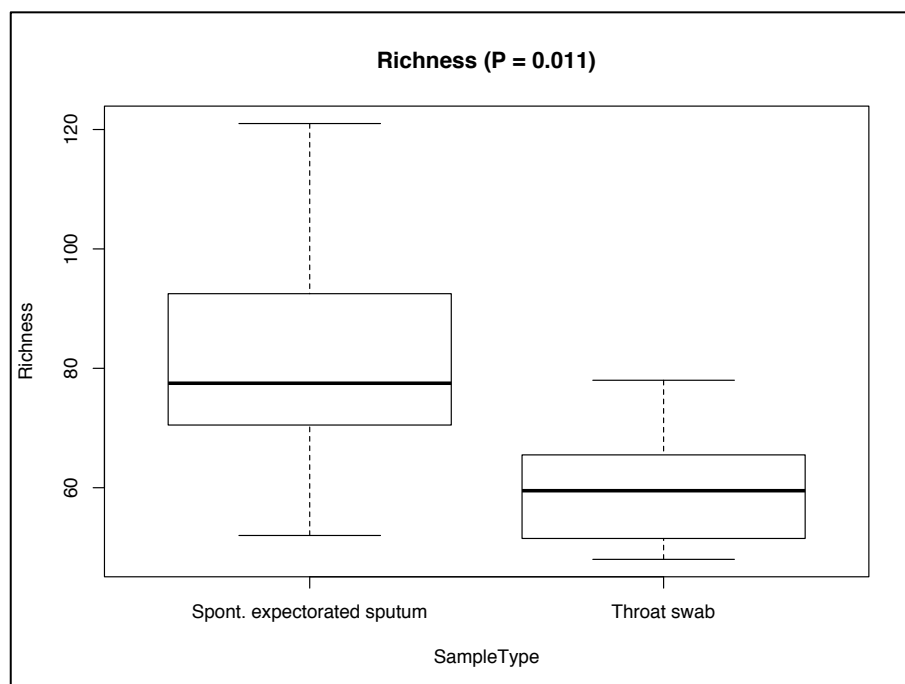


Table 5.3: Beta diversity comparisons between TS and sputum for CF and PCD.

Study group	Beta diversity metric	r ²	P-value
CF	Bray Curtis	0.11	0.419
	UniFrac	0.06	0.991
	Weighted UniFrac	0.13	0.309
PCD	Bray Curtis	0.04	0.780
	UniFrac	0.07	0.326
	Weighted UniFrac	0.04	0.869

Overall, for patients with CF and PCD, TS and sputum showed good correlation in relation to the OTUs identified and similarities in diversity and community structure, as illustrated by alpha and beta diversity analyses respectively. Therefore, in the subsequent analyses in this chapter both TS and sputum have been used to compare patients with CF with those with PCD longitudinally. TS was analysed, where available, and sputum if no TS sample had been collected at that timepoint.

5.3.5 Development of the airway microbiota in children with CF

5.3.5.1 Genus and OTU level changes with age

For the thirty-one CF patients, at genus level *Streptococcus* was the most common genus present throughout childhood until 17 years of age, representing over half the bacterial community (55.0% of total reads). Other common genera included: *Pseudomonas* (8.1%), *Veillonella* (8.1%), *Granulicatella* (4.2%), *Gemella* (4.2%), *Neisseria* (3.8%), *Prevotella* (3.3%), *Haemophilus* (3.2%), *Rothia* (1.45%), *Staphylococcus* (1.1%), and *Fusobacterium* (1.1%) (Figure 5.9). There was poor

concordance (32%) between *Pseudomonas* spp. identification on 16S rRNA gene sequencing and *P. aeruginosa* growth on bacterial cultures. Poor concordance was seen for all other organisms. Individual patient barplots illustrating changes in the microbiota coupled with clinical microbiology results are shown in Appendix A6.

Mixed effects models were used to test changes in the relative abundance of the five most common genera with increasing age from baseline with patient study number as a random effect and age range as the fixed effect. Examining trends in the 5 most common genera, *Streptococcus* was the most abundant genus throughout childhood but showed a non-significant decrease in relative abundance with age and reached its lowest relative abundance at 14 years of age ($t_{(124)} = -1.90$, $P = 0.060$) but showed an increasing trend thereafter. *Pseudomonas* was the second most abundant genus throughout childhood remaining constant until 9 years of age but then fluctuated in relative abundance thereafter. Similarly the relative abundance of *Gemella* and *Veillonella* fluctuated throughout childhood, whilst the relative abundance of *Granulicatella* showed little change. None of these changes however reached statistical significance ($P > 0.05$) (Figure 5.11). From examining individual patient barplots (Appendix A6), whilst the airway microbiota appeared to be highly individual, *Streptococcus* were abundant throughout all patients at all timepoints. There were no consistent changes in genera seen with changes in symptoms or positive growth on bacterial cultures (example shown in Figure 5.12).

To assess changes in the relative abundance of all OTUs with age multiple correlation testing using Spearman rank with FDR correction was performed. This revealed two OTUs to have a small ($r \leq 0.3$) but significant ($P_{\text{adj}} < 0.05$) change in relative abundance with age: *Haemophilus* (OTU ID 10158) decreasing with age ($r = -0.286$, $P_{\text{adj}} = 0.003$) and *Actinomyces* (OTU ID 6782) increasing with age ($r = 0.226$, $P_{\text{adj}} = 0.01$).

Figure 5.10: Sankey plot illustrating the most common genera present in the airway microbiota in children with CF.

Analysed by combining all TS and sputum collected during the study period. Number of patients = 31. The width of the ribbon represents the proportional abundance of that species within the community. This plot reveals that *Streptococcus* was the most abundant taxon (55.0% of total reads).

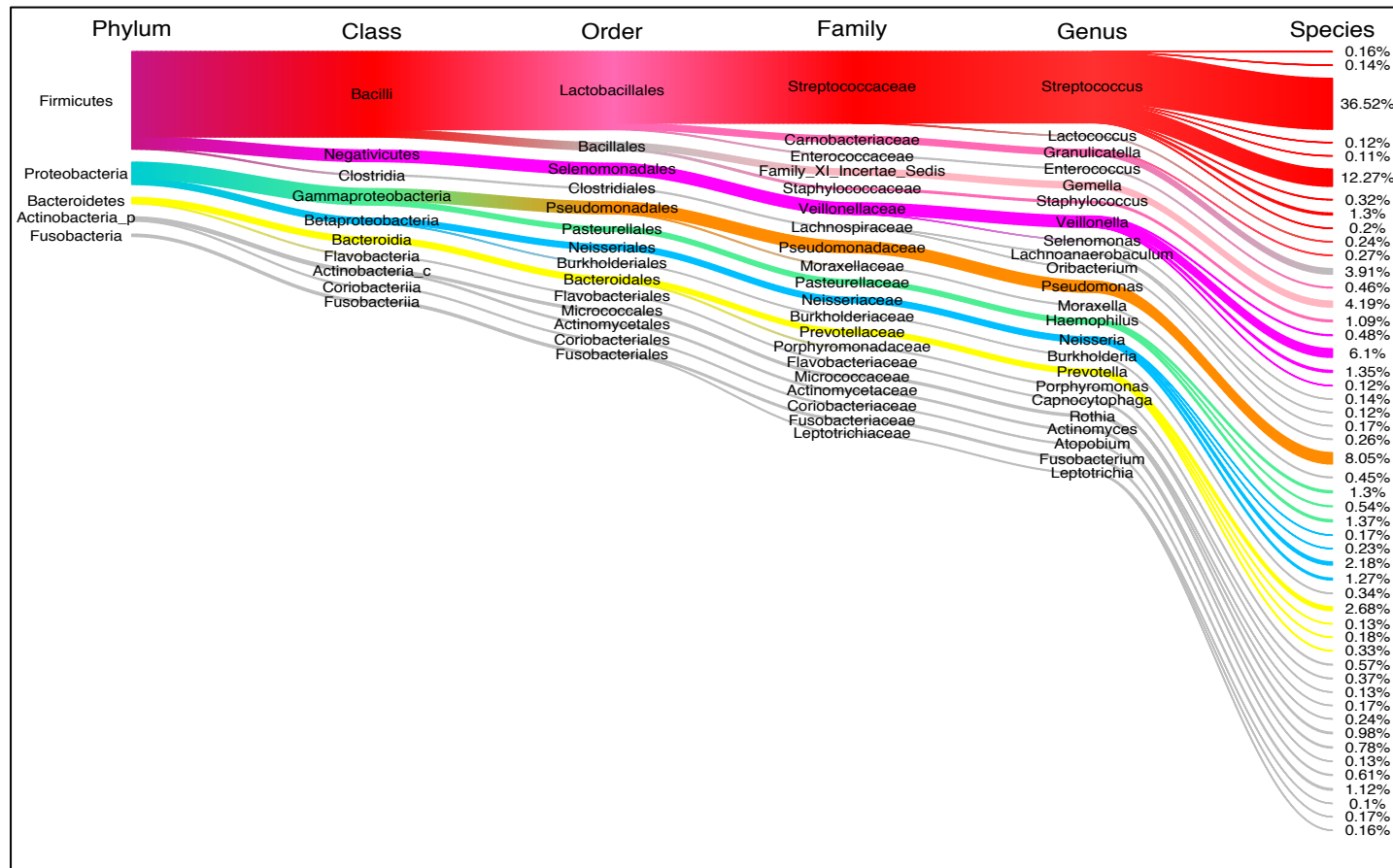


Figure 5.11: Changes in the relative abundance of the 5 most common genera with age throughout childhood in CF.

This figure shows trends in the mean relative abundance with standard error bars of *Streptococcus* spp., *Pseudomonas* spp., *Veillonella* spp., *Granulicatella* spp. and *Gemella* spp. with increasing age until 17 years of age. All changes were non-significant ($P > 0.05$).

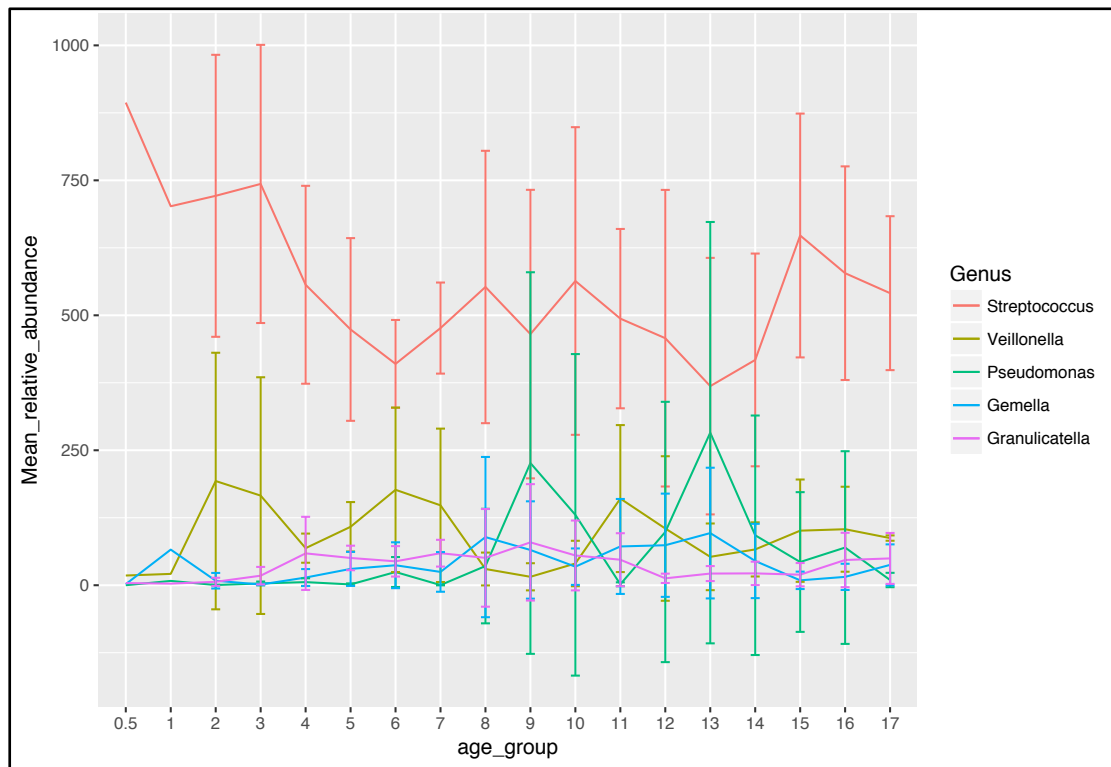
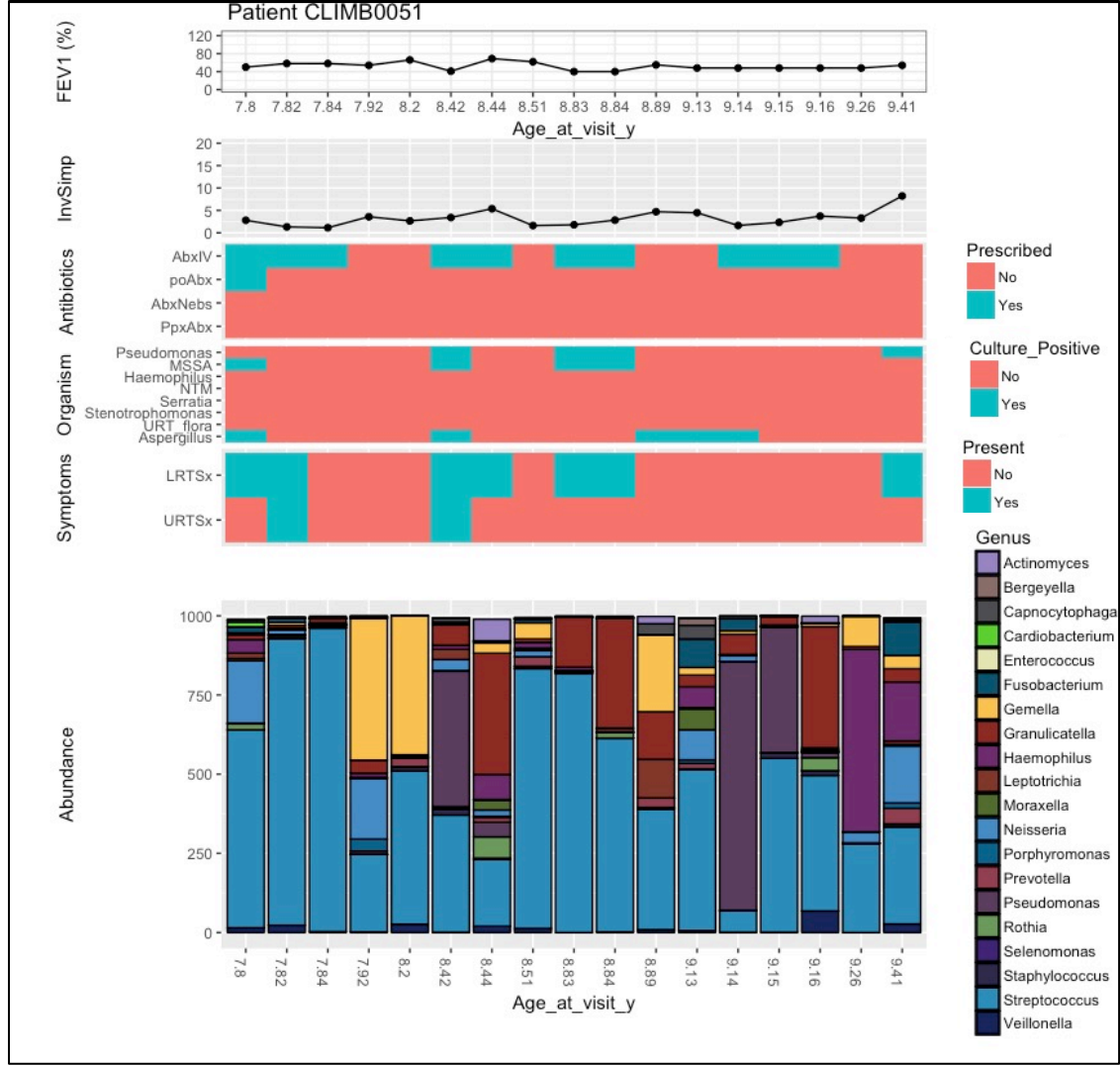


Figure 5.12: Example plot of an individual patient with CF illustrating changes in relative abundance of OTUs against spirometry measurements (FEV₁), diversity changes (Inverse Simpson's) and clinical variables (antibiotic use, organisms grown on clinical microbiology and presence of upper and lower airway symptoms).

The example shown is for a patient who was admitted for elective IV antibiotics every 3 months. Despite frequent sampling, no consistent trends were seen with changes in FEV₁ % predicted, intravenous antibiotics, positive culture for *P. aeruginosa* and presence of symptoms.

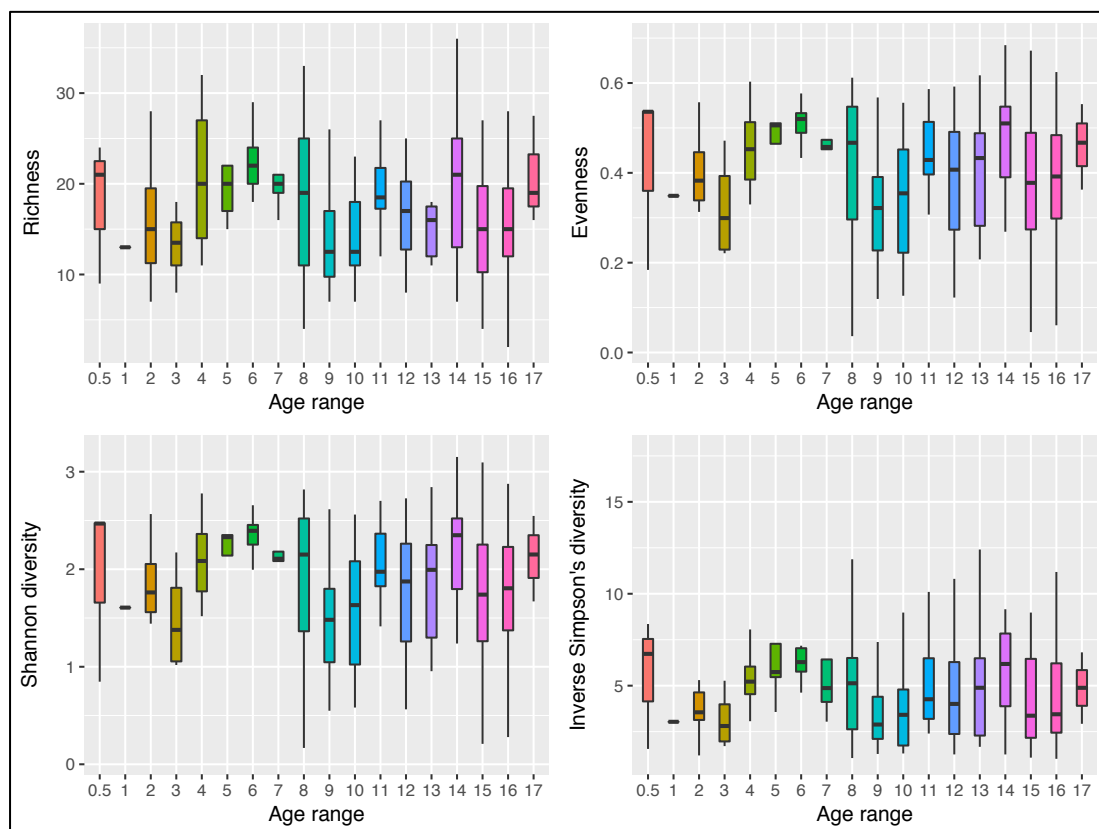


5.3.5.2 Changes in community diversity with age

Overall, there was no significant correlation between age (from 0.5 years – 17 years) and richness, evenness, Shannon or Inverse Simpson's diversity index ($P > 0.05$). Mixed effects models also confirmed there was no relationship between alpha diversity and age in years (Figure 5.13). Differences in beta diversity with age were assessed using the same approach as described in Chapter 4, Section 4.2.4. A blocked design was used constrained by patient study number to account for the repeated measures study design. No significant influence of patient age was seen on community structure (using Bray Curtis dissimilarity, unweighted UniFrac and weighted UniFrac, $P > 0.05$). This suggests that the age of the patient does not appear to have an influence on community structure.

Figure 5.13: Changes in alpha-diversity by age range for CF.

All results non-significant ($P > 0.05$).



5.3.5.3 Clinical variables influencing community structure

Alpha and beta diversity testing were repeated to determine whether any clinical variables (Table 5.1) influence community structure again using a blocked design but constrained by age. Significant differences ($P < 0.05$) in alpha diversity were found for the following variables: *P. aeruginosa* infection status (“chronic”, “intermittent”, “free” or “never”, as defined in Chapter 2, Section 52.3), and genotype (homozygous p.Phe508Del, heterozygous or other). Alpha diversity (richness, evenness, Shannon diversity and the Inverse Simpson’s diversity index) was higher in patients who were intermittently infected with *P. aeruginosa* or free of *P. aeruginosa* infection in comparison to those who had chronic *P. aeruginosa* or had never grown *P. aeruginosa*. Due to lack of confidence in parental recall of oral antibiotic use, it was not possible to determine whether children who had grown *P. aeruginosa* had received more courses of antibiotics than those who did not. Richness was higher in patients who were heterozygous p.Phe508Del than those who were homozygous or had other mutations (Table 5.4).

Overall, whilst several clinical variables have a significant influence on community structure (measured by beta diversity), the degree of variance attributed to any of the clinical variables tested was small ($< 7\%$). Genotype and *P. aeruginosa* infection status at baseline exerted the greatest influence on community structure, accounting for up to 5.8% and 6.9% of variance in community structure respectively (Bray Curtis dissimilarity, $P = 0.001$) (Table 5.4). Several other variables were also found to have a significant but small influence ($< 2\%$) on community structure and therefore are of uncertain

biological significance. These included: spirometry at baseline (FEV₁ %predicted or FVC %predicted) and CF liver disease. Due to only two patients with heterozygous p.Phe508Del mutations having a milder phenotype (Class IV and V mutations) mutation class was not compared between patients.

Table 5.4: Testing the influence of clinical variables on alpha and beta diversity for children with CF.

Listed below are the clinical variables that exerted a significant influence ($P < 0.05$) on beta diversity as measured using a permutational multivariate ANOVA (PERMANOVA). Only those variables with $r^2 > 0.02$ on beta diversity testing are included. CFRD – cystic fibrosis related diabetes; GORD – gastro-oesophageal reflux disease.

Clinical variable	Alpha diversity metric	Median (range)	P value
<i>P. aeruginosa</i> infection:			
• Chronic	Inverse Simpson's*	3.2 (1.0 – 16.1)	0.007
• Intermittent		5.0 (1.1 – 14.0)	
• Free		5.1 (1.1 – 17.8)	
• Never		3.2 (1.2 – 10.8)	
Genotype:			
• Homozygous	Richness	17 (4 - 37)	0.0007
• Heterozygous		19 (7 – 36)	
• Other		13 (2 – 34)	
Clinical variable	Beta diversity metric	r ²	P value
Genotype	Bray Curtis	0.06	0.001
	UniFrac	0.04	0.011
	Weighted UniFrac	0.06	0.003
<i>P. aeruginosa</i> infection	Bray Curtis	0.07	0.001
CFRD (N = 2)	Bray Curtis	0.06	0.001
	Weighted UniFrac	0.04	0.007
Gender	Bray Curtis	0.02	0.043
	Weighted UniFrac	0.03	0.015
GORD (N = 13)	Weighted UniFrac	0.02	0.006

*Indicates that all the alpha diversity metrics (richness, evenness, Shannon and Inverse Simpson's diversity index) were significant. Only Inverse Simpson's has been quoted.

5.3.5.4 Changes with a pulmonary exacerbation in children with CF

Eleven patients experienced a pulmonary exacerbation during the study period. Of those eleven, seven had samples collected at each of B, E, T, R timepoints, as defined in Chapter 4, Section 4.3.3.6.

Whilst *Streptococcus* remained the most abundant genus at exacerbation, its relative abundance increased from 37.4% at baseline to 54.5% at exacerbation. Several other genera showed a small increase from baseline to exacerbation, including: *Pseudomonas* (from 13.1% to 15.9%) and *Granulicatella* (from 1.7% to 4.1%). Three common genera halved in their relative abundance from baseline to exacerbation: *Staphylococcus* (from 12.2% to 6.0%), *Veillonella* (from 8.0% to 3.5%) and *Haemophilus* (from 7.6% to 3.4%). All other genera showed less than 1% change between baseline and exacerbation (Table 5.5).

Table 5.5: Changes in relative abundance of 10 most common genera between baseline and exacerbation samples in CF.

Genus	Relative abundance at baseline (%)	Relative abundance at exacerbation (%)	P-value
<i>Streptococcus</i>	37.4	54.5	0.270
<i>Pseudomonas</i>	13.1	15.9	0.873
<i>Staphylococcus</i>	12.2	6.0	0.534
<i>Veillonella</i>	8.0	3.5	0.209
<i>Haemophilus</i>	7.6	3.4	0.398
<i>Prevotella</i>	3.1	2.1	0.570
<i>Neisseria</i>	2.3	1.8	0.691
<i>Granulicatella</i>	1.7	4.1	0.419
<i>Gemella</i>	1.3	2.2	0.490
<i>Fusobacterium</i>	1.3	0.3	0.170

Examining trends in the five most common genera at baseline, exacerbation, treatment (weeks 1 and 2) and recovery, the relative abundance of *Streptococcus* spp. showed a non-significant increase from baseline to exacerbation and increased further with treatment and at recovery. All other changes in genera were small. The relative abundance of *Pseudomonas* spp. showed an increasing trend at exacerbation but decreased with treatment and remained low at recovery. The relative abundance of *Staphylococcus* spp. and *Haemophilus* spp. showed a similar relationship, with a gradual decrease between B, E and T timepoints but with small increase in *Haemophilus* spp. in recovery samples. The relative abundance of *Veillonella* spp. showed an increasing trend at exacerbation but fell with treatment before rising again at recovery (Figure 5.14). Using mixed effects models and multiple correlation testing using Spearman rank with FDR correction, no significant difference however was

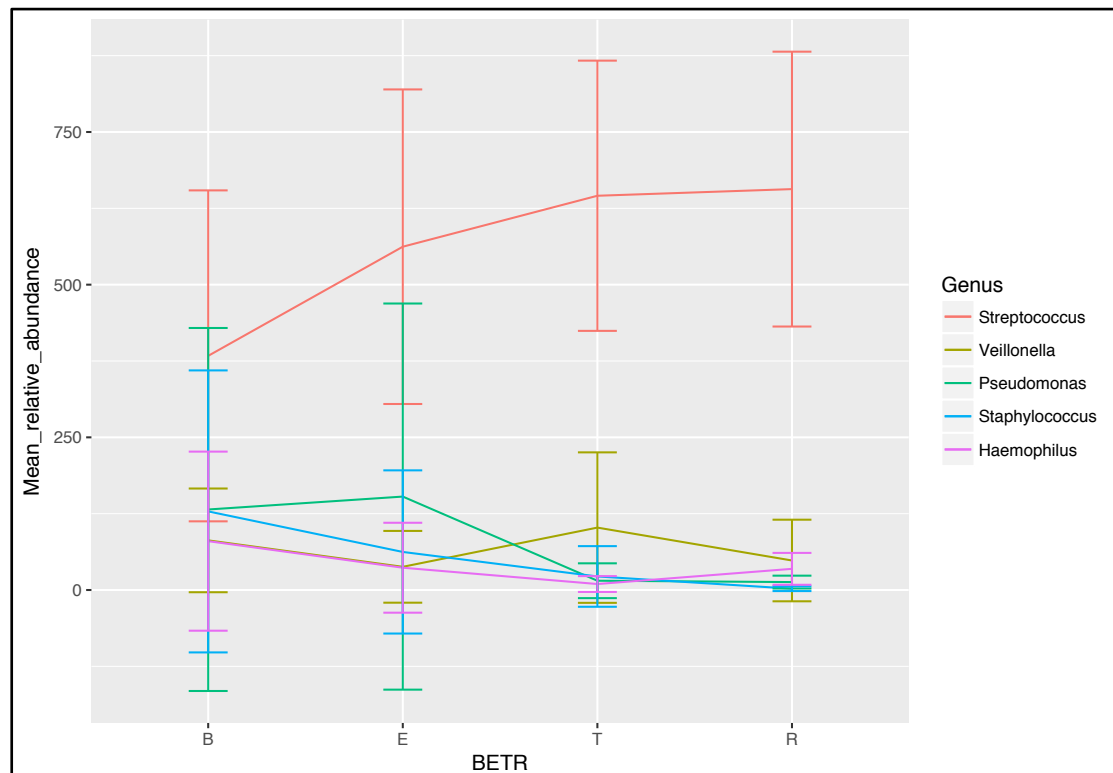
found in the relative abundance of any specific genera or OTUs respectively between baseline, exacerbation, treatment or recovery samples ($P_{\text{adj}} > 0.05$).

Similarly, no significant difference was seen in alpha diversity between baseline, exacerbation, treatment and recovery samples using mixed effects modelling ($P > 0.05$) or beta diversity (PERMANOVA, Bray Curtis $P = 0.457$, unweighted UniFrac $P = 0.870$, Weighted UniFrac $P = 0.977$).

There was no significant difference in FEV₁ (% predicted) between B, E, T and R timepoints (Kruskal-Wallis, $H = 38$, $P = 0.582$). Therefore, an assessment of differences between patients who recover lung function following treatment for an exacerbation and those who do not could not be made.

Figure 5.14: Changes in the relative abundance of the 5 most abundant genera with exacerbations in CF.

The figure illustrates changes in the relative abundance with standard error bars of *Streptococcus* spp., *Pseudomonas* spp., *Haemophilus* spp., *Veillonella* spp. and *Staphylococcus* spp. at baseline (1 month prior to starting IV antibiotics), exacerbation (within 48 hours of starting IV antibiotics), treatment (at days 10-14 of IV antibiotics) and recovery (within 1 month of stopping IV antibiotics).



5.3.6 Development of the airway microbiota in children with PCD.

5.3.6.1 Genus and OTU level changes with age

Examination of the data for the 31 PCD patients, revealed that at the genus level *Streptococcus* was the most common genus present throughout childhood until 17 years of age but represented just under one third of the bacterial community (31.3% of total reads). This is in contrast to CF where *Streptococcus* represented just over half of the bacterial community. *Haemophilus* were more prevalent in PCD than in CF representing one fifth of the community (20.5%) in contrast to 3.2% in CF. *Pseudomonas* was less prevalent in PCD than CF (1.2% and 8.1% respectively). The relative abundance of other common genera was similar between CF and PCD and included: *Veillonella* (10.4%), *Neisseria* (6.3%), *Prevotella* (5.2%), *Granulicatella* (4.0%), *Moraxella* (3.3%), *Gemella* (2.8%), *Fusobacterium* (2.3%) and *Pseudomonas* (1.2%). The most common genera present in children with PCD are illustrated in Figure 5.15.

Whilst *Streptococcus* spp. and *Haemophilus* spp. were also the two most common taxa identified on bacterial cultures, in contrast to the results of 16S rRNA gene sequencing, *Haemophilus* spp. was the most common organism cultured (55%) and *Streptococcus* spp. were the second most common organisms (32%) (Table 5.1). *Staphylococcus* spp. were not one of the fifty most abundant OTUs identified on sequencing and was only cultured in 13% of patients. Similar to CF, anaerobes (e.g. *Veillonella* spp. and *Neisseria* spp.) were not isolated by bacterial cultures for PCD patients. Individual patient barplots illustrating changes in the microbiota coupled with clinical microbiology results are shown in Appendix A7.

As before, mixed effects models were used to test changes in the relative abundance of the 5 most common genera with increasing age from baseline with patient study number as a random effect and age range as the fixed effect. Examining trends in the five most common genera, *Streptococcus* had a high relative abundance in early childhood whereas the relative abundance of *Haemophilus* was low. With age, changes in the relative abundance of *Streptococcus* and *Haemophilus* showed an inverse relationship: *Streptococcus* decreased in relative abundance until 5 years of age ($t_{(84)} = -2.61, P = 0.011$) whilst *Haemophilus* increased ($t_{(82)} = 1.90, P = 0.060$); *Streptococcus* then showed an increasing trend in relative abundance until 9 years of age, whilst *Haemophilus* showed a decreasing trend although neither of these were significant ($P = 0.222$ and 0.912 respectively). *Streptococcus* decreased in relative abundance between the ages of 10 and 17 years ($t_{(84)} = -2.46, P = 0.016$ at 17 years of age) whilst *Haemophilus* increased ($t_{(82)} = 2.35, P = 0.021$ at 17 years of age). This contrasts with CF where *Streptococcus* decreased in relative abundance until 14 years of age and increased thereafter and *Haemophilus* was notably not one of the five most common genera (Figure 5.16).

The relative abundance of *Neisseria* spp. fluctuated from 9 years of age onwards and showed a significant increase at around 12 years of age ($t_{(82)} = 2.10, P = 0.039$). No significant changes were seen in the relative abundance of *Prevotella* spp. and *Veillonella* spp. (Figure 5.16). Examining individual patient barplots, no consistent changes were seen in genera with age or clinical status (example shown in Figure 5.17). At the OTU level, applying multiple correlation testing

with FDR correction there was no significant difference in the relative abundance of OTUs with age.

Figure 5.15: Sankey plot illustrating the most common genera present in the airway microbiota in PCD throughout childhood.

The width of the ribbon represents the proportional abundance of that species within the community.

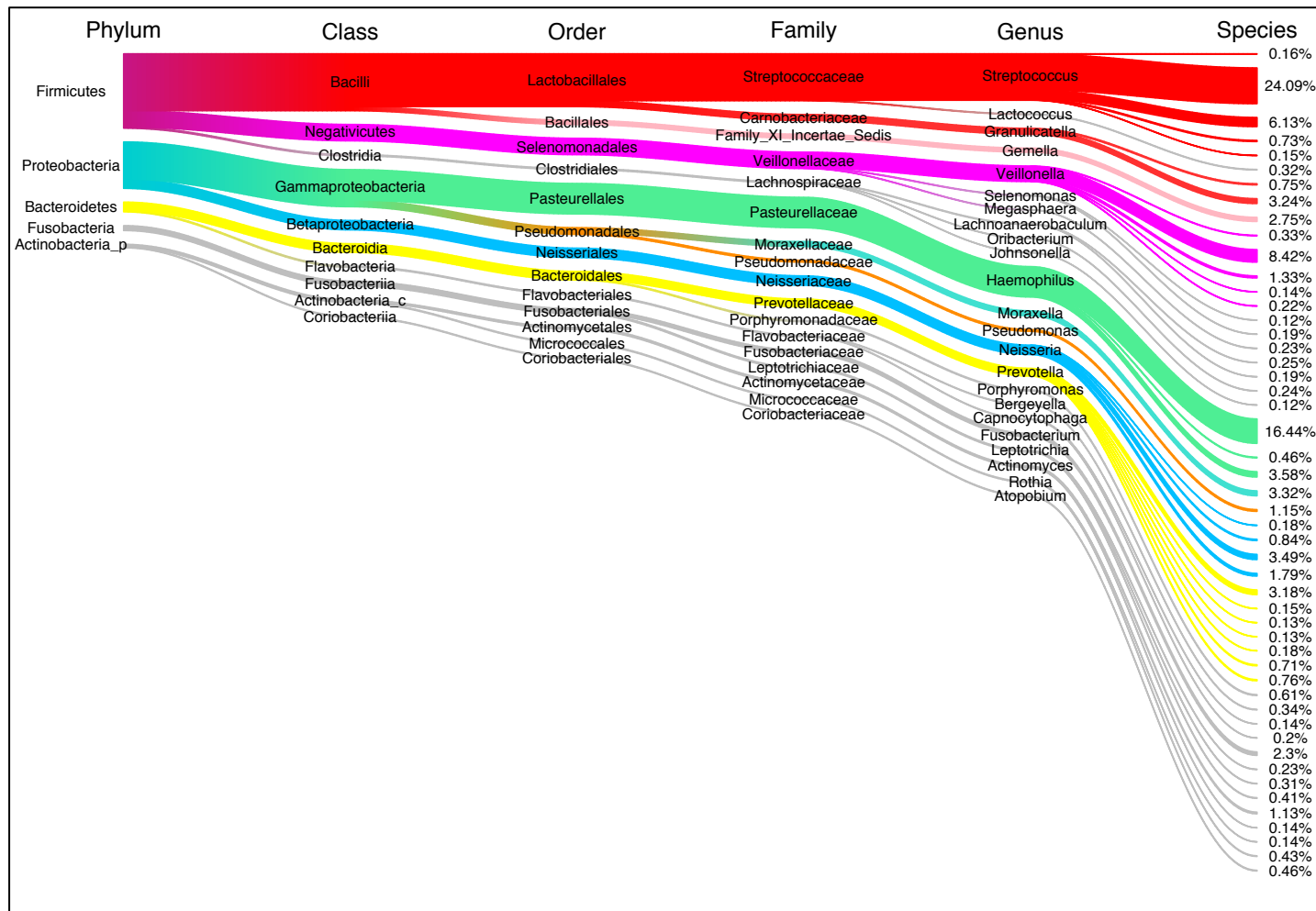


Figure 5.16: Changes in the relative abundance of the 5 most common genera with age throughout childhood in PCD.

This figure shows trends in the mean relative abundance with standard error bars of *Streptococcus* spp., *Haemophilus* spp., *Veillonella* spp., *Granulicatella* spp. and *Neisseria* spp. with increasing age.

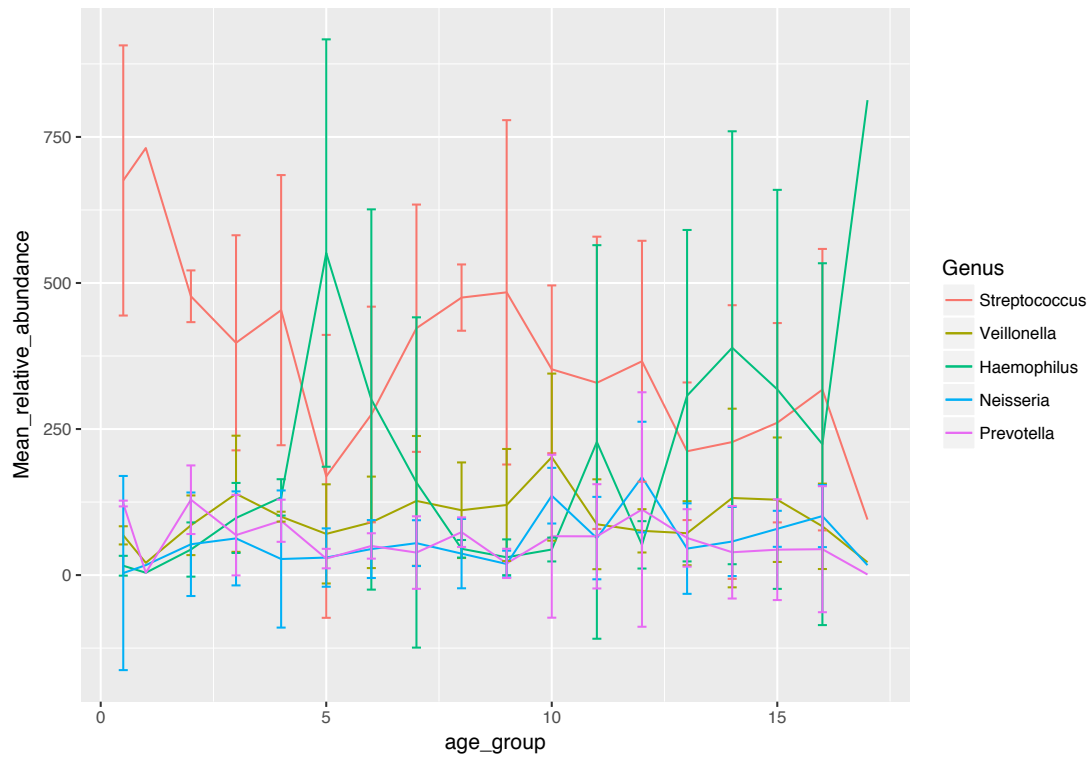
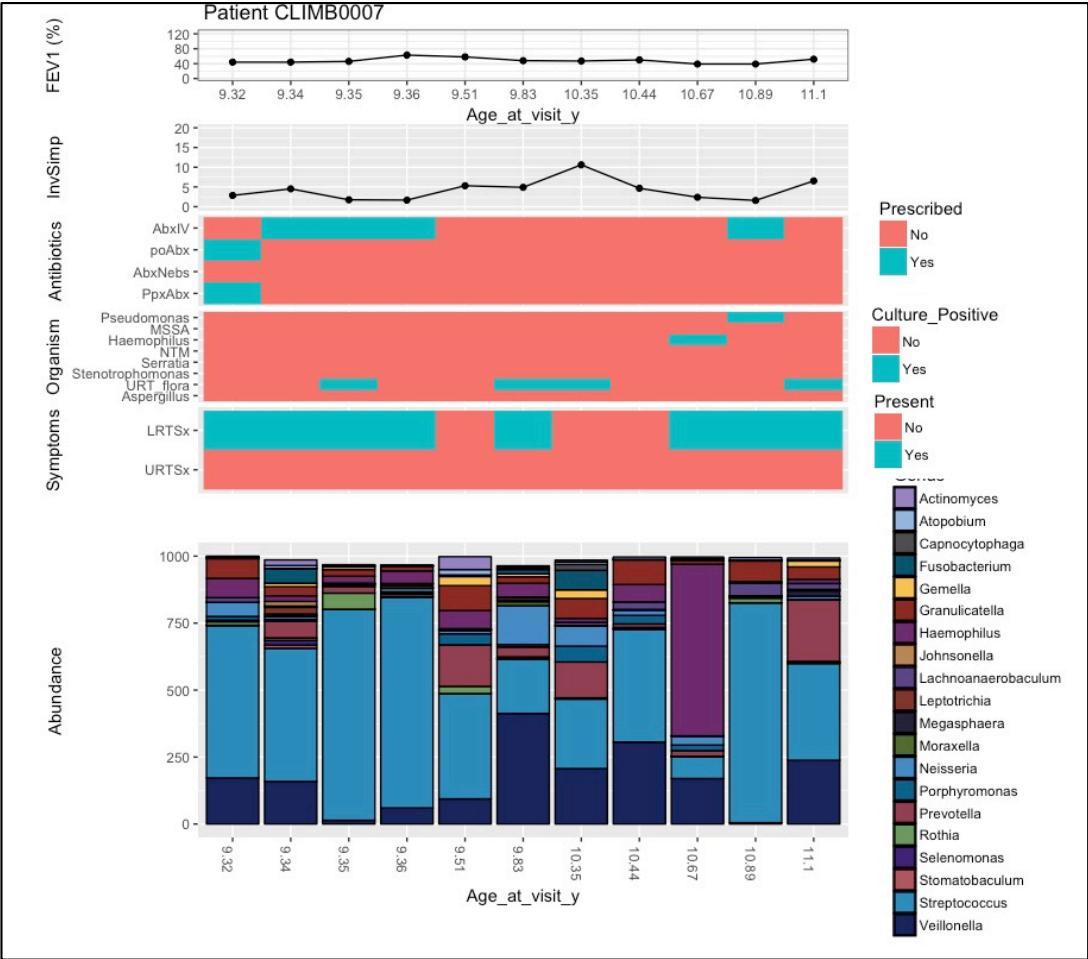


Figure 5.17: Example plot of an individual patient with PCD.

This illustrates changes in relative abundance of OTUs against spirometry measurements (FEV₁ % predicted), diversity changes (Inverse Simpson's) and clinical variables (antibiotic use, organisms grown on clinical microbiology and presence of upper and lower airway symptoms). Overall, no consistent changes in the microbiota were seen with clinical variables.

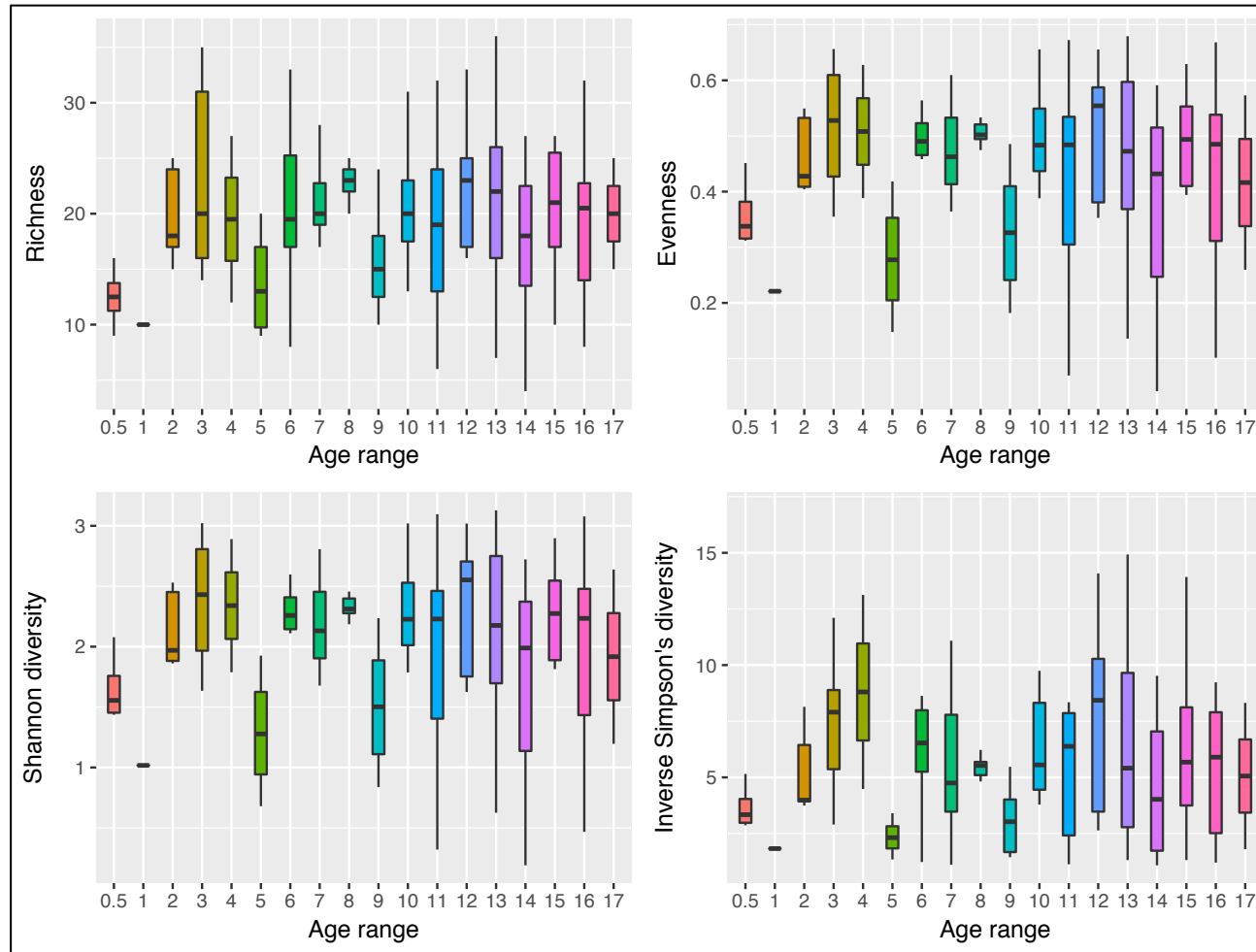


5.3.6.2 Changes in community diversity with age

No significant correlation was seen with increasing age with richness, evenness, Shannon or the Inverse Simpson's diversity index ($\tau = 0.014$, P 's 0.447, 0.661, 0.663 and 0.806 respectively). Mixed effects models were used to test the relationship between alpha diversity and age in years (tested as a categorical variable). In contrast to CF, there was a significant increase in richness until adolescence in patients with PCD ($t_{(96)} = 2.36$, $P = 0.02$ at 13 years of age) with a non-significant decrease seen thereafter ($t_{(96)} = 1.27$, $P = 0.207$). There were no significant changes with age for evenness, Shannon or the Inverse Simpson's diversity index ($P > 0.05$) (Figure 5.18). Similar to CF, no significant influence of patient age was seen on community structure using the Bray Curtis, unweighted UniFrac and weighted UniFrac ($P > 0.05$).

Figure 5.18: Changes in alpha-diversity by age range for PCD.

Only richness showed a significant change until 13 years of age ($t_{(96)} = 2.36, P = 0.02$). All other changes are non-significant ($P > 0.05$).



5.3.6.3 Clinical variables influencing community structure

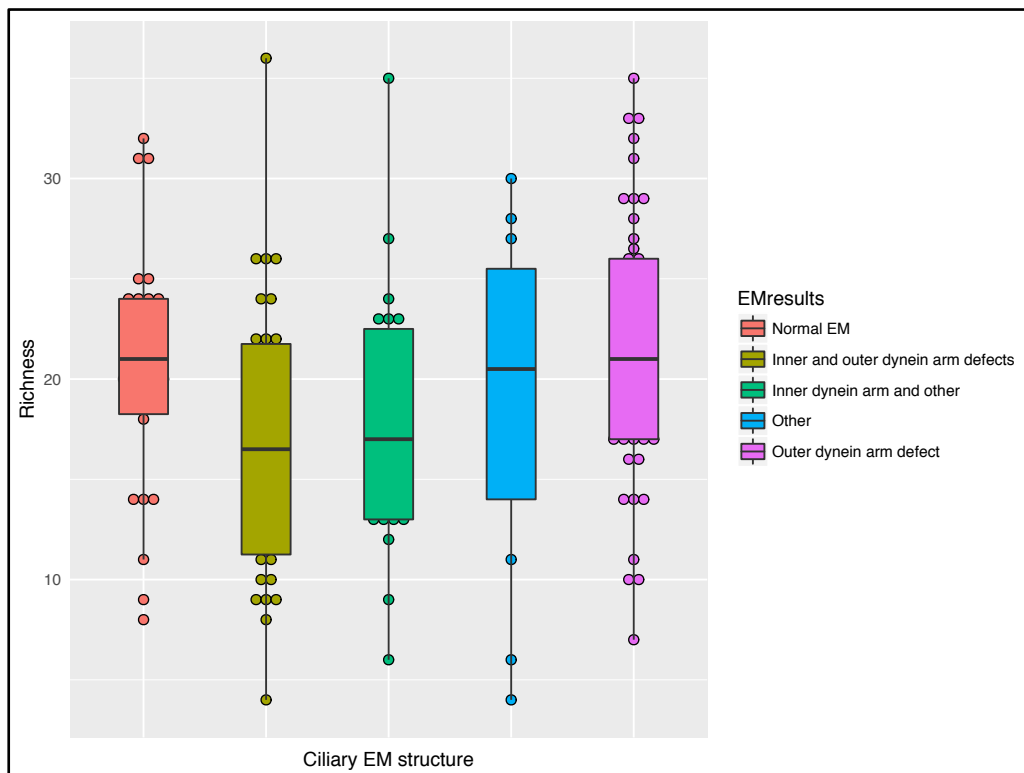
Alpha and beta diversity testing was repeated to identify whether any clinical variables influenced the PCD microbiota. Alpha diversity was tested using t-tests or Wilcoxon signed-rank test for parametric and non-parametric binary variables and ANOVAs and Kruskal-Wallis test similarly for variables with multiple values. Beta diversity testing was performed using a PERMANOVA. In contrast to CF, only two clinical variables influenced either alpha or beta diversity: ciliary structure and gender. Ciliary structure (classified as “inner and outer dynein arm defect”, “outer dynein arm defect”, “other” [e.g. radial spoke or microtubule defects], “inner dynein arm defect and other” and “normal”) had the largest influence on community structure, accounting for up to 8.3% of the variance in community structure (Bray Curtis, $r^2 = 0.05$, $P = 0.042$, UniFrac $r^2 = 0.04$, $P = 0.06$ and weighted UniFrac $r^2 = 0.08$, $P = 0.001$). Comparing alpha diversity, there was a significant difference in richness, evenness, Shannon diversity and Inverse Simpson’s diversity indices with different ciliary ultrastructural defects ($P < 0.05$). However, on further analysis, this difference was only evident between patients with combined inner and outer dynein arm defects (richness mean 16.9 ([d 6.7]) and patients with isolated outer dynein arms defects (richness mean 21.5 [SD 6.4]) only (Tukey HSD, $P_{\text{adj}} = 0.017$) (Figure 5.19).

Gender also had a significant although much smaller influence on community structure (using Bray Curtis dissimilarity and unweighted UniFrac scores, $r^2 = 0.02$ and 0.02 respectively, $P = 0.001$). There was however no significant

influence of gender on alpha diversity (richness, evenness, Shannon or Inverse Simpson's diversity indices, $P > 0.05$).

Figure 5.19: Boxplot showing difference in richness between patients with PCD with different ciliary ultrastructural defects.

Ciliary ultrastructural defects on electron microscopy (EM) ("EM results") were classified as: "Normal EM", "Inner and outer dynein arm defects", "Inner dynein arm and "other" defects (e.g. radial spoke and microtubular defects) and "other". A significant difference was seen between those with combined inner and outer dynein arm defects and those with isolated outer dynein arm defects only (Tukey HSD, $P_{\text{adj}} = 0.017$).



5.3.6.4 Changes at exacerbation

Fourteen patients with PCD had an exacerbation requiring admission for IV antibiotics during the study, twelve of whom had sequential samples at baseline, exacerbation, treatment (week 1 of treatment = T1, week 2 of treatment = T2) and recovery. At the genus level, in the month prior to exacerbation, *Haemophilus* spp. was the most common genus with a relative abundance of 32% of total reads, which decreased at exacerbation to 13.2%. In contrast to CF, *Streptococcus* was the second most abundant genus but changed little in relative abundance from 27.0% at baseline to 28.8% at exacerbation. With the exception of *Pseudomonas*, *Prevotella* and *Neisseria* (between baseline and exacerbation: *Pseudomonas* decreased from 7.6% to 2.8%; *Prevotella* increased from 2.3% to 7.7%; *Neisseria* increased from 4.1% to 6.0%), (Table 5.6) all other genera showed less than 1% change between baseline and exacerbation.

Table 5.6: Changes in the relative abundance of the 10 most common genera between baseline and exacerbation samples.

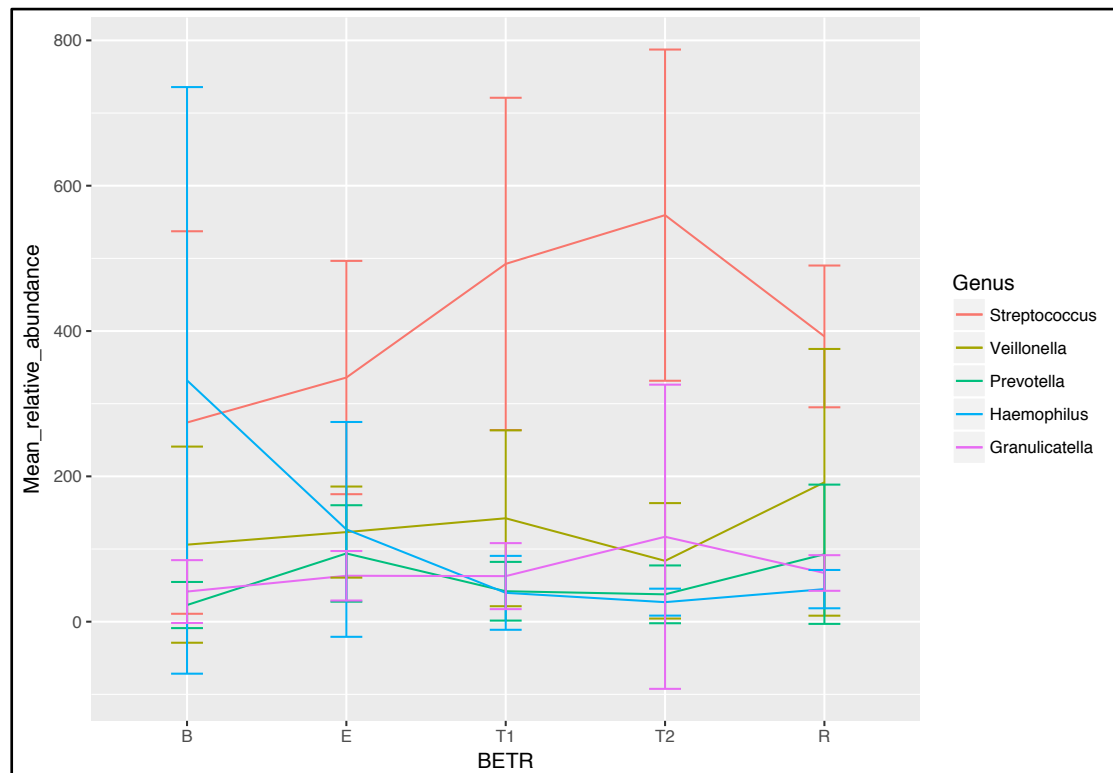
Genus	Relative abundance at baseline (%)	Relative abundance at exacerbation (%)	P-value
<i>Haemophilus</i>	32.0	13.2	0.025
<i>Streptococcus</i>	27.0	28.8	0.769
<i>Veillonella</i>	10.3	10.0	0.819
<i>Pseudomonas</i>	7.6	2.8	0.452
<i>Neisseria</i>	4.1	6.0	0.433
<i>Granulicatella</i>	3.9	4.7	0.767
<i>Gemella</i>	2.5	3.1	0.161
<i>Prevotella</i>	2.3	7.7	0.018
<i>Actinomyces</i>	1.7	1.0	0.842
<i>Fusobacterium</i>	1.5	1.6	0.745

Using mixed effects models to examine trends in the five most common genera at baseline, exacerbation, treatment (both weeks 1 and 2; T1 and T2) and recovery, *Streptococcus* spp. and *Haemophilus* spp. again showed opposite trends. *Streptococcus* spp. increased significantly at week 1 ($t_{(29)} = 2.44$, $P = 0.021$) and week 2 ($t_{(29)} = 2.45$, $P = 0.02$) of treatment and then fell within 1 month after treatment (non-significant, $P = 0.402$). This was in contrast to the trend seen in CF where the relative abundance of *Streptococcus* spp. continued to increase in the recovery sample. *Haemophilus* spp. decreased significantly in relative abundance at exacerbation ($t_{(29)} = -2.36$, $P = 0.025$) and with treatment ($t_{(29)} = -3.20$, $P = 0.003$ after 1 week of treatment and $t_{(29)} = -2.97$, $P = 0.006$ after 2 weeks of treatment) but increased within 1 month after treatment ($t_{(29)} = -2.64$, $P = 0.013$). *Prevotella* spp. showed a significant increase at exacerbation ($t_{(29)} =$

2.49, $P = 0.018$) but showed no change in relative abundance from baseline with treatment or recovery. *Veillonella* spp. and *Granulicatella* spp. showed no significant changes (Figure 5.20). At the OTU level, using multiple correlation testing using Spearman rank with FDR correction, no significant difference was found in the relative abundance of any OTUs between baseline, exacerbation, treatment or recovery samples ($P_{\text{adj}} > 0.05$).

Figure 5.20: Changes in the 5 most common genera with exacerbations, treatment and recovery for children with PCD.

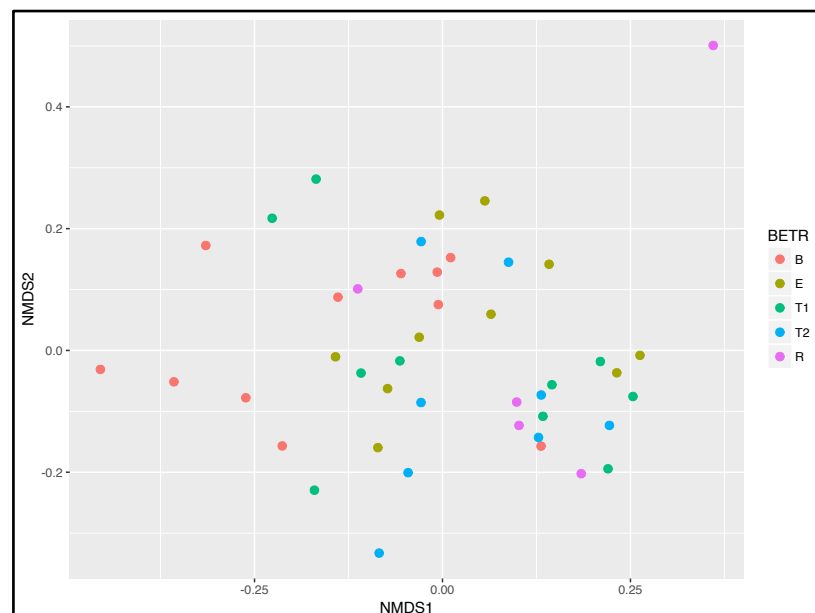
The figure illustrates changes in the relative abundance of *Streptococcus* spp., *Haemophilus* spp., *Granulicatella* spp., *Veillonella* spp. and *Prevotella* spp. at baseline (1 month prior to starting IV antibiotics), exacerbation (within 48 hours of starting IV antibiotics), treatment (during week 1 (T1) and week 2 (T2)) and recovery (within 1 month of stopping IV antibiotics).



Using mixed effects models, there was a significant increase in richness ($t_{(29)} = 2.4, P = 0.022$), evenness ($t_{(29)} = 3.12, P = 0.004$), Shannon diversity ($t_{(29)} = 3.12, P = 0.004$) and Inverse Simpson's diversity index ($t_{(29)} = 2.31, P = 0.028$) at exacerbation in comparison with baseline. No significant changes were seen in alpha diversity with either treatment or recovery. For beta diversity, a significant difference in community structure using Bray Curtis dissimilarity ($r^2 = 0.17, P = 0.001$) and unweighted UniFrac score ($r^2 = 0.10, P = 0.007$) but not weighted UniFrac ($P = 0.226$). There are no post-hoc tests currently for the

adonis (PERMANOVA) function in the Vegan package in R, thus it was not possible to statistically test where this difference lies between the five timepoints. From visual examining the NMDS plots, clustering of baseline samples appears to differ from exacerbation, treatment and recovery samples (Figure 5.21 example NMDS plot: Bray Curtis dissimilarity). Similar to CF, there was no significant difference in FEV₁ (% predicted) between B, E, T and R timepoints (Kruskal-Wallis, $H = 31$, $P = 0.415$). Therefore, an assessment of differences between patients who recover lung function following treatment for an exacerbation and those who do not could not be made.

Figure 5.21: Non-metric multidimensional scaling (NMDS) plot illustrating differences in beta diversity (shown here using Bray Curtis dissimilarity) for patients with PCD receiving IV antibiotics for treatment of an exacerbation at baseline (“B”), exacerbation (“E”), treatment (at week 1 “T1” and week 2 “T2”) and recovery (“R”).



5.3.7 Comparison of the microbiota between CF and PCD

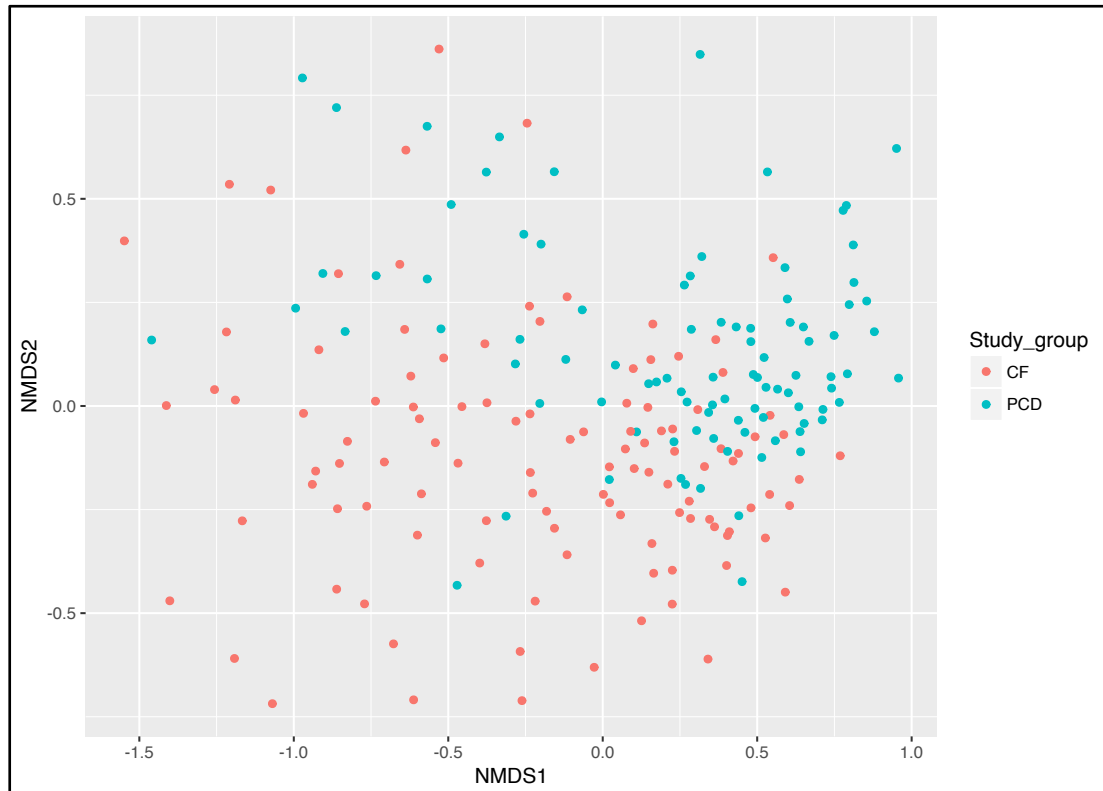
Comparing the relative abundance of the most common genera between CF and PCD (all samples and all timepoints), there was a significant difference in the relative abundance of *Haemophilus* spp. only, which was six fold higher in PCD (mean 0.23 [sd 0.05]) than CF (mean 0.04 [sd 0.02], $P < 0.0001$, $P_{\text{adj}} 0.0005$).

Overall, alpha diversity was significantly higher in PCD than CF (all samples and all timepoints) as measured by species richness (mean 19.6 [sd 6.9] and 17.1 [sd 6.2] in PCD and CF respectively, $P = 0.007$), evenness (median 0.46 [range 0.041 – 0.681] and 0.42 [range 0.036 – 0.686] in PCD and CF respectively, $P = 0.040$) and Shannon diversity (median 2.12 [range 0.190 – 3.137] and 1.94 [range 0.167 – 3.158] in PCD and CF respectively, $P = 0.04$) but not Inverse Simpson's diversity index (median 4.94 [range 1.1 – 17.1] and 4.53 [range 1.06 – 16.3] in PCD and CF respectively, $P = 0.072$).

Differences in beta diversity were measured by comparing Bray Curtis dissimilarity, unweighted UniFrac and weighted UniFrac scores between children with CF and PCD. Overall, when comparing all samples at all timepoints, there was a significant influence of study group (CF or PCD) on community structure, with between 1.3% (weighted UniFrac, $P = 0.013$) and 5.5% (Bray Curtis, $P = 0.001$) of variance explained by study group (Figure 5.22).

Figure 5.22: Non-metric multidimensional scaling (NMDS) comparison of Bray Curtis dissimilarity in children (aged 0.5– 17 years) with CF and PCD.

CF is shown in red and PCD is shown in blue. This demonstrates differences in clustering between patients with CF and PCD ($r^2 = 0.06$, $P = 0.001$).



5.3.8 The impact of antibiotics on diversity of the CF microbiota

Studies of the CF microbiota in adults have suggested reduced diversity is associated with end stage disease and have implicated frequent antibiotic use as the primary driver of reduced diversity¹³⁹. As described in the RBH CF Clinical Guidelines, children with more severe lung disease are admitted every 3 months for IV antibiotics irrespective of whether they are experiencing an exacerbation at the time of admission. Looking at patients recruited into the study there were seven who received 3 monthly IV antibiotics. Of these, five patients had both elective admissions and emergency IV admissions for treatment of an exacerbation. Although numbers were small the effect of antibiotics on the airway microbiota in CF elective versus emergency IV admissions was compared within these five patients at BETR timepoints as previously described in Chapter 4, Section 4.3.3.6.

Overall, there was no significant difference in either alpha (richness, evenness, Shannon or Inverse Simpson's diversity indices) using mixed effects modelling or beta diversity (Bray Curtis dissimilarity, unweighted UniFrac or weighted UniFrac scores) at any timepoint (B, E, T and R) using PERMANOVA when comparing elective and emergency IV admissions ($P > 0.05$). However, in view of the small numbers, these data must be interpreted with care.

5.4 Discussion

The aim of this chapter was to determine the influence of the airway microbiota on clinical features in children with CF and PCD and whether differences in

prognosis between CF and PCD are related to differences in the airway microbiota. Accordingly, sixty-two age-matched children with CF and PCD (N = 31 in each group) were recruited and TS or sputum collected every 2-3 months for an average of 1.1 years to assess the relationship between longitudinal changes in diversity and community composition and clinical features, particularly pre- and post-IV antibiotics. It is the first study to report on the longitudinal changes in the airway microbiota children with CF of all ages (0 – 17 years) and the first such longitudinal study in children with PCD. At the time of writing this thesis, the current literature at present does not contain any exploration of the role of the airway microbiota in prognostic differences between CF and PCD.

Previous studies in CF have suggested that diversity of the airway microbiota has an inverse relationship with disease progression^{139,140}, with a cross-sectional study of age-stratified patients with CF suggesting that there is an increase in diversity until adolescence and a decrease thereafter^{135,145}. The results of this present study have shown that the airway microbiota is highly individual in children with CF. There was no significant difference with age in either alpha diversity (measured by richness, evenness, Shannon Diversity Index and Inverse Simpson's Diversity Index) or beta diversity (measuring community nestedness and turnover). In addition, there was no significant difference in either alpha or beta diversity with spirometry (FEV₁ % predicted or FVC% predicted). Similarly, almost no significant changes in alpha or beta diversity were seen with age or spirometry on in children with PCD, with the exception of species richness which was found increase with age until 13 years of age using a mixed effects model.

Overall the data and results suggest that diversity of the airway microbiota is not an important determinant of disease progression in children with CF or PCD.

Comparing diversity between children with CF and those with PCD, alpha diversity was higher in PCD and beta diversity was also significantly different between CF and PCD when comparing all ages combined. Given that PCD is associated with a milder phenotype than CF, this could support the suggestion that increased diversity of the airway microbiota is associated with a more favourable prognosis although it contradicts the trends described with lung function above. This may be however because disease severity in this study with regard to PCD and CF was relatively mild (mean FEV₁ (% predicted) in CF: 76.5 [SD 16.9]; in PCD: 79.3 [17.9]), reducing the ability to detect a relationship with disease severity which would have been possible if very severe patients were included. Furthermore, there was no significant difference in FEV₁ between CF and PCD. This is a surprise finding considering that PCD is associated with a milder phenotype and could suggest a selection bias in either the group of CF or PCD children recruited in this study, which could not be excluded. Thus, in accordance with the aims of this study, whilst a difference in the microbiota between CF and PCD could be demonstrated, whether this difference relates to differences in disease prognosis could not be fully determined.

A further possible explanation for these findings is that FEV₁ may not be a sensitive marker for tracking lung disease in children, as it often falls late with lung disease in CF²⁵⁵. Whilst protocols for infant lung function testing have been developed by some centres, currently the protocol cannot be reliably performed

in children under 5 years of age and therefore significant early changes may be missed. In contrast, lung clearance index (LCI), a measure of ventilation inhomogeneity measured by multi-breath washout testing, has been shown to track changes in lung disease and can be performed in preschool children²⁵⁶. Thus tracking change in LCI with changes in airway microbiota diversity may be more revealing. Unfortunately insufficient children in this study had LCI performed as part of routine clinical care to permit this comparison. Consequently from the results in this study it is not currently possible to conclude whether diversity is beneficial.

The results in this study contrast with the findings of Zhao *et al.*¹³⁹ who observed decreased alpha diversity with “progressive” lung disease over a nine year follow-up period in adult patients with CF. In the study by Zhao *et al.*¹³⁹, “progressive” lung disease was defined by the average decline in FEV₁ per year of study calculated by calculating the difference in FEV₁ between the first and last sample collected in the study divided by the study period. Applying Zhao *et al.*’s methods for calculating lung function change in the dataset in this present study, there was no relationship found between FEV₁ % change and diversity of the microbiota. There are several reasons why the results in this study may differ from those of Zhao *et al.*¹³⁹. Firstly, the patients in the study by Zhao *et al.*¹³⁹ were older (aged between 18 and 30 years) than the patients in the present study. Furthermore, Zhao *et al.* only found an association between FEV₁ and diversity in three progressive patients and not the three stable patients also recruited in their study. In addition, no standardised threshold exists for defining “stable” and “progressive” FEV₁ decline, thus the use of arbitrary cut-offs may

lead to misleading results. Caution should therefore be taken in generalising the findings in three CF patients to an entire population.

Whilst several other studies in adults have found a similar relationship between diversity and lung function to that of Zhao *et al.*¹³⁹, the studies have likewise only involved a small number of patients (e.g. N = 4)¹⁴³. As such, the present study described in this current chapter has involved a substantially larger sample size of 31 patients and whilst still small it is better powered to reflect true differences.

At the genus level, *Streptococcus* was the most common genus in children with CF, representing 55% of total reads. *Pseudomonas* was the second most common genus (8.1%) and was present at a higher relative abundance than that seen in early CF in the NBS cohort reported in Chapter 4. In the NBS cohort *Pseudomonas* was not a common genus. This is consistent with traditional culture data from the UK CF Registry that has shown that *Pseudomonas* is more prevalent in later childhood. Similar to several previous studies of the CF airway microbiota and the results from the NBS cohort reported in Chapter 4, several anaerobes were prevalent in the airway microbiota in children with CF including *Prevotella* spp., *Veillonella* spp., *Granulicatella* spp., *Gemella* spp. and *Fusobacterium* spp.

Overall, with the exception of *Pseudomonas* spp., 16S rRNA gene sequencing and bacterial culture results were often discrepant particularly for *Streptococcus* spp. and anaerobes. The findings of this Chapter are similar to those of Mahboubi

*et al.*²⁴⁹ (Chapter 4, Section 4.4 for discussion of Mahboubi *et al.* findings) and the findings in Chapter 4 for children with CF on NBS. Only 3% of children with CF included in this study grew “upper respiratory tract flora” on culture, therefore it is unlikely that *Streptococcus* spp. have been dismissed as belonging to this group. Whilst bacterial cultures showed greater similarity with 16S rRNA gene sequencing in patients with PCD discrepancies were nonetheless still seen with *H. influenzae* being the most common organism cultured (45%) and *Streptococcus pneumoniae* the second most abundant whilst anaerobes failed to be cultured. In this study, however, 16S rRNA gene sequencing was only able to discern organisms to the genus level. Thus it is plausible that the discrepancies between culture and sequencing results may reflect different species being identified on sequencing to those identified on culture. Anaerobes, such as *Prevotella* spp. and *Veillonella* spp., are notoriously difficult to grow in culture hence the low detection of these organisms on bacterial cultures may reflect the limitations of the clinical microbiology techniques during the study period. Successful detection of a large number of anaerobes in CF has been reported with strict anaerobic bacterial culture techniques¹³². Use of these techniques may demonstrate better concordance between microbiota profiles and bacterial culture results.

Whilst *Streptococcus* was the most common genus in patients with PCD, its relative abundance was lower than that seen in CF, representing just under one third of total reads. In contrast, *Haemophilus* spp. were significantly more abundant (21%) and *Pseudomonas* spp. were less abundant (non-significant) in PCD (10%) than in CF (3.2%). Comparing these results with the results in

Chapter 4 of NBS infants with CF, NBS and PCD appear to have greater similarities than NBS with older children with CF, with a greater abundance of *Haemophilus* spp. and a low abundance of *Pseudomonas* spp. seen in NBS and PCD whereas the reverse is true for older children with CF. Similarly when comparing longitudinal changes in the five most abundant genera, PCD showed similar trends to NBS infants with an inverse relationship seen in the relative abundance of *Streptococcus* spp. and *Haemophilus* spp. with age. CF, however, showed a non-significant reduction in the relative abundance of *Streptococcus* spp. until 13 years of age and an increase thereafter only. Overall this perhaps suggests that the microbiota in PCD, a milder form of CSLD than CF, is similar to that of the early microbiota in CF (NBS) and that perhaps a switch occurs in CF from a less pathogenic to a more pathogenic community composition and this does not occur for PCD, at least in childhood. Whether this switch is of direct pathological significance or a marker of another underlying disease process is unclear. Elucidating what may cause such a switch may help to identify therapeutic targets for maintaining lung health.

Differences in the relative abundance of *Haemophilus* spp. and *Pseudomonas* spp. could be one potential cause of this switch. Similar to the findings in this study, an age-stratified cross sectional study of children and adults with CF found that *Haemophilus* spp. are highly abundant in early childhood and are replaced by *Pseudomonas* spp. with age¹³⁵. A cross-sectional study analysing the microbiota of 60 patients with non-CF bronchiectasis found a strong competitive effect between *Haemophilus* spp. and *Pseudomonas* spp., such that when *Haemophilus* spp. were the dominant species *Pseudomonas* spp. were either absent or present

in very low abundance and vice versa. *Haemophilus* spp. dominated microbiota profiles were associated with a greater prevalence of aerobes, such as *Neisseria* spp., and *Pseudomonas* spp. dominated profiles were associated with a greater prevalence of anaerobes, such as *Prevotella* spp., thought to be related to production of thickened mucus and reduced oxygen concentrations by *Pseudomonas* spp.²⁵⁷ It is plausible therefore that the airway microbiota in early CF is largely influenced by the high abundance of *Haemophilus* spp., leading to a milder phenotype, which is later replaced by *Pseudomonas* spp., leading to a more severe phenotype. Interventions aimed at promoting *Haemophilus* spp. to aid it to outcompete *Pseudomonas* spp. may therefore improve prognosis for patients with CF, although this is speculative as the role of *Haemophilus* spp. in CF has not yet been determined and it may not be beneficial. *In vivo* competition experiments using established CF models, such as the CF pig, are needed to explore this potential relationship further.

During the study period, eleven patients with CF experienced a pulmonary exacerbation requiring admission for IV antibiotics. Of these eleven, 7 patients had samples taken in the month prior to exacerbation (baseline), at admission for IV antibiotics (exacerbation), during treatment with IV antibiotics (at week 1 = T1 and week 2=T2) and within 1 month of stopping IV antibiotics (recovery). Overall, no significant difference was seen at the genus level or in alpha or beta diversity between these timepoints. Similar to several previous studies of the airway microbiota^{139,143,152} this suggests that the microbiota does not change significantly during an exacerbation. However, the number of patients with CF experiencing an exacerbation who were analysed in this study was small. Larger

longitudinal studies are needed to fully ascertain the changes occurring in patients with CF at exacerbation.

Fourteen patients with PCD experienced an exacerbation, twelve of whom had samples taken at baseline, exacerbation, treatment and recovery. Comparing these timepoints, in contrast to CF, species richness, evenness, Shannon and Inverse Simpson's diversity indices showed a significant increase at exacerbation from baseline but not between exacerbation and treatment or recovery samples. Coupled with the significant changes in beta diversity seen in the Bray Curtis dissimilarity and unweighted UniFrac scores, this could suggest that community composition shifts occur at exacerbation for patients with PCD but this is not seen in CF, although it is important to note that the numbers under study are small. Comparing genera, *Haemophilus* was found to significantly decrease at exacerbation whereas *Prevotella* spp. increased, suggesting that reductions in *Haemophilus* spp. may create a niche in which more pathogenic organisms can flourish. Furthermore, treatment strategies for exacerbations in both CF and PCD do not currently target anaerobes, such as *Prevotella* spp. If *Prevotella* spp. are increased at exacerbations, including anaerobic cover (such as metronidazole) to current treatment for exacerbations may improve clinical outcomes. Randomised clinical trials to assess the potential benefits and harms of such a treatment strategy are needed. Insights may be gained to inform current antibiotic strategies from the ongoing CF Microbiome-determined Antimicrobial Therapy Trial in Exacerbations: Results Stratified trial ("CFMATTERS", <https://www.cfmatters.eu/>) in which 252 adult patients with CF who have stable lung disease and are chronically infected with *P. aeruginosa*

have been randomised to receive either standard treatment at exacerbations with IV ceftazidime and tobramycin or “microbiota-driven” treatment, in which an additional antibiotic is given based on the dominant organism in the patient’s microbiota to compare differences in FEV₁.

During this study, detailed clinical information was collected concurrently with patient samples. This permitted the relationship between both modifiable clinical variables (such as antibiotic use) and non-modifiable variables (such as genotype) and the microbiota to be explored. Whilst several clinical variables were found to have a statistically significant influence on both alpha and beta diversity, the degree of variance in the microbiota attributed to any one variable was invariably small (often < 5%) and therefore unable to inform changes in current treatment strategies.

To ascertain the influence of antibiotics, differences in the microbiota between emergency IV (during an exacerbation) and elective IV admissions (3 monthly, not during an exacerbation) were compared in a subgroup of five patients who had at least one emergency and one elective admission each. The aim of this was to determine whether previously reported effects of antibiotic treatment in driving reductions in diversity of the microbiota¹³⁹ would still be seen when not confounded by intrinsic changes of exacerbations. Comparison were made within patients rather than between patients who had been admitted for emergency IVs vs. elective IVs, as was originally planned due to difficulties in age and sex matching. Similar to previous studies, this showed no significant difference at either baseline, admission, with treatment or at recovery in either

alpha or beta diversity^{144,160}. However, given that there were only five patients included in this part of the study, it is likely that there was insufficient statistical power to detect a difference. Future studies comparing the microbiota in large samples of children with CF randomised to either elective 3 monthly IV antibiotics or IV antibiotics during exacerbations only would be more informative, particularly as the diagnosis of exacerbations improve with more objective, biomarker driven diagnosis^{258,259}.

The study reported here has several strengths. It is to date the largest longitudinal study to date of both the CF and PCD microbiota in childhood. The groups were well matched in terms of age, gender, lung function and BMI. Patients were sampled frequently with a median of 42 days between samples in CF and 61 days in PCD, allowing meaningful comparisons and a more detailed investigation of the microbiota to be made.

However, one limitation of this study is that different sampling methods (TS and sputum) were used at different timepoints. Ideally the same sample type would have been collected at each timepoint from each patient. Unfortunately this proved impossible as sometimes patients refused to have TS taken. Unfortunately too few (N = 2) of the patients included in the comparison of the upper and lower airway microbiota (Chapter 3) were able to produce sputum to allow for the correlation between sputum and lower airway samples to be assessed. More sputum samples were collected from PCD patients than CF. Thus, it is plausible that the differences seen in the microbiota between these two groups may reflect differences between these two sample types. A comparison

however of paired TS and sputum from 16 patients in this study did not demonstrate a significant difference at the genus level or in alpha or beta diversity. This would support the conclusion that the differences seen between children with CF and PCD are true disease differences and not due to differences in sampling.

A further limitation is that samples were obtained opportunistically when patients attended for routine clinical appointments. As a result, there were no pre-specified timepoints at which samples were obtained and sampling frequency was variable between patients (Figures 5.2 and 5.3) often with missing samples at a given timepoint for each patient. Performing a statistical comparison of the longitudinal changes in the microbiota between CF and PCD was therefore not possible during the time of my doctoral studies but will be the focus of future work.

Due to time constraints, it was not possible to complete sample processing for all the samples collected in the CLIMB study. This includes performing quantitative PCR (qPCR) to measure changes in bacterial load with changes in the microbiota and clinical status in children with CF and PCD, as was reported for NBS children in Chapter 4. Nonetheless the results of this study demonstrated that there is an increase in richness until adolescence (13 years of age) but not for any other alpha diversity metric in patients with PCD. This suggests that the number of different organisms increases until adolescence for children with PCD but not the spread of these organisms (evenness) and hence overall diversity is unchanged. Coupling this finding with changes in bacterial load may provide further useful

insights to help determine whether bacterial load increases as the number of different organisms (richness) increases. This would allow it to be determined whether the microbiota is expanding overall or whether established organisms are being successively replaced by new organisms within the niche.

In conclusion, in accordance with the study aims, children with CF and PCD were recruited for one of the first longitudinal studies of the airway microbiota in these disease groups to date. A further aim of this study had been to compare changes in the microbiota between CF and PCD longitudinally to investigate the role of the airway microbiota in prognostic differences between these two groups. Unfortunately such a statistical analysis was not possible and only a descriptive study was performed. Overall, in contradiction to the study hypothesis, there were no significant changes in diversity with increasing age or FEV₁ % predicted in either children with CF or PCD. Similarly, there were no significant changes in the microbiota seen with exacerbations in CF. Community composition, however, did differ between CF and PCD, with changes throughout childhood in PCD with patterns similar to that seen in NBS infants with CF (Chapter 4). Given that both PCD and NBS infants have reportedly milder lung disease than older children with CF, this could suggest a potential switch from a potentially less pathogenic community composition in infancy to a more pathogenic community later in childhood. Changes in the relative abundance of *Haemophilus* spp. (which was more prevalent in PCD) and *Pseudomonas* spp. (which was more prevalent in CF) may contribute to this switch. An increase in *Prevotella* spp. and a decrease in *Haemophilus* spp. was seen at exacerbation in PCD, suggesting that anaerobes may have a role in exacerbations. Further work

is needed to further clarify the competitive relationship between *Haemophilus* spp., *Pseudomonas* spp. and *Prevotella* spp. as well as randomised controlled trials to ascertain the risks and benefits of including anaerobic cover into current treatment strategies for exacerbations.

Chapter 6: General discussion

Recurrent airway infection and its complications are the leading cause of death in CF but remain poorly understood. Whilst longitudinal studies of the airway microbiota in adults have suggested an important role of diversity of the microbiota in disease progression¹³⁹, there have been few such longitudinal studies in early CF disease in children. This is in part due to uncertainty regarding the reliability of upper airway samples in representing the lower airway microbiota and the impracticalities of repeatedly performing FOB in children, which requires a general anaesthetic.

The aim of this PhD was to perform a two-part study of the airway microbiota in childhood to assess the role of the airway microbiota in children with CSLDs. In the first part, a prospective, cross-sectional study comparing the microbiota on upper (CS and TS) and lower airway samples (BALF and bronchial brushings) was performed to determine whether upper airway samples could act as surrogates for characterising the lower airway microbiota in longitudinal studies. It was hypothesised that upper airway samples would reflect the diversity and community composition of the lower airway microbiota. The results of this were used to inform the sample collection in the **CSLD Longitudinal Investigation of the Microbiota (CLIMB study)** in the second part of this study, which related changes in the airway microbiota with clinical features in infants with CF identified on NBS and older children with CF compared with PCD.

The key findings of the cross sectional study were:

- TS sequenced whereas CS sequenced poorly, precluding their use for molecular microbiology (Chapter 3, Section 3.3.2).
- There was good correlation between TS and lower airway samples at a community level (community diversity and membership) but not at an individual level (Chapter 3, Section 3.3.3). Differences in the microbiota were identified in different diseases on TS (Chapter 3, Section 3.3.4).
- There was a trend in samples showing the greatest discordance for this to be driven by dominance of a genus in either the lower airway sample or TS. The dominant genus on BALF was frequently from the same genus identified on clinical culture of the same fluid (Chapter 3, Section 3.3.3).

In accord with the study hypothesis, this suggests that TS reflect the lower airway microbiota at a community level and can distinguish different disease groups but show an inconsistent relationship at an individual level limiting their use in clinical decision making in an individual child. In particular, caution should be observed when a dominant organism is identified on TS as these samples may be poor surrogates for the lower airway microbiota. This has important implications in the interpretation of TS for longitudinal studies. Where samples are dominated by an individual genus, repeat sampling paired with culture dependent microbiology to confirm findings may be advisable or alternative sampling methods, such as BALF, should be considered. Furthermore, this highlights the importance of combining bacterial cultures with molecular microbiology to obtain a more comprehensive insight into the lower airway microbiota.

These findings are unsurprising as a recent study has demonstrated similar results in infants with CF²⁶⁰ (N = 17). To my knowledge, this thesis reports one of the largest studies to date (N = 49) comparing the upper and lower airway microbiota in children with a variety of chronic lung diseases (e.g. CF, PCD and non-CSLD controls) of all ages (0.2 – 16.1 years). It is the first study to compare CS with the lower airway microbiota and demonstrate their limited use.

In contrast to previous studies, nasopharyngeal sampling was not included in this study due to several previous studies demonstrating that nasopharyngeal swabs have a distinct microbiota profile to that of the lower airways in both adults and children^{151,260}. In addition, the role of induced sputum, performed by ultrasonic nebulisation of hypertonic saline, as a surrogate for lower airway sampling in non-productive patients was not assessed. This was due to difficulties in obtaining induced sputum in young children, and the time-consuming nature of the procedure, although there have been recently reported improved success rates in this group²⁶¹⁻²⁶³. Nevertheless, a study including induced sputum is needed to better ascertain whether it is better than TS in reflecting the lower airway microbiota in young children.

Overall, the findings of this study suggest that TS can be used to follow the development of the lower airway microbiota at a community level in different diseases. Thus the aim of this part of the study, to determine whether TS could be used as surrogates for lower airway sampling in longitudinal studies of the airway microbiota was met. TS were therefore used as the predominant sampling method in the CLIMB study.

The aim of the CLIMB study was to better understand the role of the airway microbiota in disease progression in children with CSLDs, in the long term to try to identify strategies for maintaining lung health into adulthood. This was achieved by performing a longitudinal study relating changes in the airway microbiota with age and clinical features in: (a) infants with CF identified on NBS and (b) children with CF compared with PCD, including pre- and post- antibiotic treatment. Prospective cohorts of infants with CF diagnosed on NBS (N = 30), older children with CF (N = 31) or PCD (N = 31) and TS or sputum collected at regular intervals (median approximately 1 month for NBS and 2 months for CF and PCD) for up to two years. This included an assessment of the changes occurring during prognostically important clinical events, namely with the first growth of *P. aeruginosa*, at the first pulmonary exacerbation in the NBS cohort and subsequent exacerbations in older children with CSLDs.

The key findings of the CLIMB study were:

- A small change only in diversity was seen in the first two years of life (Chapter 4, Section 4.3.3.3), with no change in diversity with age thereafter in either CF or PCD (Chapter 5, Sections 5.3.5.2 and 5.3.6.2) or with important clinical events, such as pulmonary exacerbations (Chapter 4, Section 4.3.3.6) or the first growth of *P. aeruginosa* (Chapter 4, Section 4.3.3.7). Overall diversity was higher in PCD than in CF (Chapter 5, Section 5.3.7).
- *Streptococcus* was the most common genus in children of all ages with CF or PCD (Chapter 4, Section 4.3.3.2 and Chapter 5, Sections 5.3.5.1 and 5.3.6.1) although its relative abundance was higher in CF than PCD. In CF

in the first 7-9 months of life, the relative abundance of *Streptococcus* spp. increased and thereafter showed a decreasing trend with increasing age.

- *Haemophilus* was the second most common genus in NBS infants (Chapter 4, Section 4.3.3.2) and PCD (Chapter 5, Section 5.3.6.1) and *Pseudomonas* was the second most common genus in older children with CF (Chapter 5, Section 5.3.5.1). Neither *Staphylococcus* spp. nor *Pseudomonas* spp. were common in NBS infants.
- An inverse relationship was seen between the relative abundance of *Streptococcus* spp. and *Haemophilus* spp. with age in NBS and PCD but not older children with CF (Chapter 4, Section 4.3.3.2 and Chapter 5, Section 5.3.6.1).
- IV antibiotic treatment for pulmonary exacerbations had no significant impact on diversity of the airway microbiota in NBS, CF or PCD (Chapter 4, Section 4.3.3.6, Chapter 5, Sections 5.3.5.4, 5.3.6.4 and 5.3.8).

These findings contradict the first part of the study hypothesis that changes in diversity would be seen with age and clinical status in children with CF and PCD. This was an unexpected finding, given that two age-stratified cross-sectional studies of the airway microbiota, one of which included a large number of patients (N = 269)¹⁴⁵, suggested that diversity increases rapidly in early childhood^{135,145}. Whilst this may be due to differences in study design (longitudinal vs. cross-sectional), it is also plausible that diversity is reduced overall in infants with CSLDs in comparison to infants with healthy airways. Thus, a limitation of this study is that there is no comparator healthy control group to determine whether the microbiota profile found is specific to children

with CSLDs due to the challenges in recruiting large numbers of healthy children for regular hospital attendance for sampling of the airway microbiota. An alternative would have been to recruit children with mild protracted bacterial bronchitis to compare children with CSLDs with mild airways disease, as these children are already under frequent follow-up and sampling at RBH but this was not feasible within the time of this PhD. The findings of this study, however, suggest that diversity of the airway microbiota may not be an important determinant of disease outcomes at least in children with CSLD.

However, examining relationships at the genus level may be more informative. That *Streptococcus* was the most common genus in both NBS and older children with CF was a surprise finding because *Streptococcus* spp. are not traditionally associated with the CF airways. Despite being more traditionally associated with PCD, the relative abundance of *Streptococcus* spp. in PCD was lower than in CF. PCD is associated with a less severe phenotype and taken with the trend towards an increased relative abundance at exacerbation in CF, this could suggest a potentially pathogenic role for *Streptococcus* spp. in children with CSLD. The *Streptococcus milleri* group (SMG) has previously been implicated in causing exacerbations in CF^{158,159}. A limitation of 16S rRNA gene sequencing is that it is often only able to identify bacteria to the genus level. Thus species level information was not available to ascertain whether the streptococci in this study belonged to SMG or another species of *Streptococcus*. Given that *Streptococcus* was the most common genus in all three of the groups studied and decreased in relative abundance with age in CF, it is plausible that species replacement of potentially beneficial (or at least neutral) species with more pathogenic ones

occur with increasing age. Overall, the role of *Streptococcus* spp. in disease progression requires further study.

A novel finding of this study was that an inverse relationship might exist between *Streptococcus* spp. and *Haemophilus* spp. with age in both NBS infants and PCD but not older children with CF. This leads to the hypothesis that there is competitive inhibition between *Streptococcus* spp. and *Haemophilus* spp. in children with CSLDs. *In vitro*, *Streptococcus pneumoniae* outcompetes *H. influenzae* when co-cultured, particularly, in low pH environments²⁵⁰. It is possible therefore that the low pH environment of the CF airways favours *Streptococcus* spp. growth. Thus *Streptococcus* spp. and *Haemophilus* spp. may be competitors and have an antagonistic relationship in the airways of infants with CF, although further data are required to confirm this.

Whether *Haemophilus* spp. are beneficial or pathogenic in CSLDs is unclear. Comparing the most common genera present between NBS infants and older children with CF, there appears to be an apparent replacement of *Haemophilus* spp. in early CF with *Pseudomonas* spp. in later CF. This suggests that there may be competition between *Haemophilus* spp. and *Pseudomonas* spp. as has previously been suggested in a cross-sectional study of patients with non-CF bronchiectasis²⁵⁷. In addition, similar to NBS, a high relative abundance of *Haemophilus* spp. was seen in PCD. If a high relative abundance of *Haemophilus* spp. is associated with a milder disease course, this may support a potentially beneficial role for *Haemophilus* spp. Interventions aimed at promoting *Haemophilus* spp., such as through the use of probiotics, may therefore improve

prognosis for patients with CF. However, the alternative explanation, that *Haemophilus* spp. may merely be less harmful, needs to be excluded before this therapeutic strategy is utilised. Overall, the changes in genera seen in children with CSLDs support the second part of the study hypothesis that changes in community membership would be seen with age, clinical status and between CF and PCD. Whether these changes are pathogenic or reflective of other disease processes is undetermined.

Collectively, these findings demonstrate that greater similarity exists between the airway microbiota in NBS CF and PCD than with older children with CF. This suggests that a switch occurs in childhood CF from a milder to a more severe phenotype that does not happen in PCD (although I cannot exclude that this may happen in adult life), potentially driven by changes in the relative abundance of *Haemophilus* spp. and *Pseudomonas* spp. Further longitudinal study to determine when and how such a switch may occur may help to identify new therapeutic targets to promote a profile associated with a milder phenotype for children with CSLDs.

The low relative abundance of *Staphylococcus* spp. seen in NBS infants in this study differs from longitudinal studies of the nasopharyngeal microbiota in infants with CF in USA and Europe^{10,11}. A possible explanation for these differences is that all the infants with CF in this study received prophylactic anti-staphylococcal antibiotics, in accordance with then current UK CF Trust guidelines⁷², which may have suppressed changes in the microbiota and confounded trends in the natural history of the microbiota. The use of antibiotic

prophylaxis is not a universally accepted practice internationally due to concerns regarding previous reports of increased rates of *P. aeruginosa* with administration of prophylaxis particularly with use of a broad-spectrum cephalosporin¹²⁵. However, *Pseudomonas* spp. showed similarly low abundance to *Staphylococcus* spp. in the cohort of NBS infants reported here. This leads to the hypothesis that anti-staphylococcal prophylaxis does not increase rates of *Pseudomonas* spp. detection, a hypothesis which is currently being investigated in the CF anti-staphylococcal antibiotic prophylaxis trial (CF START, <http://www.cfstart.org.uk/>).

The lack of significant identifiable changes in the microbiota that precede the onset of either the first pulmonary exacerbation or the first growth of *P. aeruginosa* could suggest that virulence factors of the bacteria may drive these events despite no overall change in their relative abundance. Alternatively, other microbes may contribute significantly to both these two events in children with CF, such as viruses and fungi. Rhinovirus is of particular interest given *in vitro* models demonstrating that rhinovirus infection leads to the release of planktonic *P. aeruginosa* from biofilms, leading to a more proinflammatory state²⁶⁴. To my knowledge, to date there have been no reported studies combining investigation of the microbiome, respiratory virome and mycobiome in infants with CF. In part this is due to challenges in selecting appropriate target gene regions and primer design for fungi and no marker genes for viruses. This has been the focus of an ongoing PhD project by Dr Imogen Felton at the National Heart and Lung Institute. Further work in larger cohorts of NBS infants with a longer period of follow-up incorporating virome and mycobiome analysis could

be highly beneficial in identifying microbiological biomarkers and novel treatments for important clinical events in CF, such as exacerbations.

In this PhD, the influence of antibiotics on the airway microbiota was also explored by: (a) examining the changes in the microbiota with treatment for pulmonary exacerbations with IV antibiotics, and (b) comparing admissions for emergency IV antibiotics during an exacerbation with elective admissions for 3-monthly IV antibiotics when patients were stable in a group of children with progressive CF. Overall, no significant difference was seen with treatment for exacerbations in NBS, CF or PCD or in diversity between elective and emergency admissions.

This finding has several implications. Firstly, it is plausible that standard IV antibiotic treatment in CF (predominantly with ceftazidime and tobramycin) fails to target potential pathogens in the CF airways, such as anaerobes. This is important in the context of the high relative abundance of *Streptococcus* spp. in each of these disease groups as highlighted above, many of which are facultative anaerobes. In particular, ceftazidime has been found to have low²⁶⁵ or intermediate activity²⁶⁶ against SMG, which may be important contributors to airway disease in CF and are currently unabated. This warrants further exploration of antibiotic strategies which provide greater anaerobic cover. Furthermore, these findings contradict those of Zhao *et al.*¹³⁹, in which antibiotic use was implicated as the primary driver for reduced diversity associated with end stage disease in adults with CF. However, the numbers in each of these groups was small, thus it is plausible that the study was underpowered to detect

a difference with IV antibiotics. Larger studies are therefore needed to determine the effect of antibiotics on the airway microbiota, which could not be conclusively determined with this study.

A common limitation in each part of this study was that there was predominantly one sample operator throughout. Although this is a significant strength as it allowed for consistency in sampling technique, it frequently led to scheduling conflicts between patients being followed up opportunistically and missed samples. Having more sample collectors would not only have allowed more samples to be collected but may have allowed more time for completion of the laboratory work of the samples collected from the 56 children with CF (excluding NBS) recruited as part of the CLIMB study and better powered comparisons to be made.

However, this is one of the largest longitudinal studies of the oropharyngeal microbiota in NBS infants to date with more frequent sampling than previously reported. This allowed greater opportunity to identify changes in the microbiota with a sampling method that has shown better correlation with the lower airway microbiota than previous similar studies using nasopharyngeal sampling^{260,267}. It is the first longitudinal study of the airway microbiota in children with CF and PCD of all ages (0 -17 years) with changing clinical status thus permitting a comparison between these two conditions which share a similar pathophysiology but different prognosis, with PCD generally associated with a more favourable prognosis than CF. Thus, it is the first study to suggest that the microbiota in PCD is similar to that of the early microbiota in CF (NBS) and that a

switch may occur in CF from a less to a more pathogenic community composition that does not occur in PCD, for reasons which are unknown. Whilst differences in the relative abundance of specific genera with age have been demonstrated in this study in children with CSLD, whether these organisms are biologically active and contributing to disease remains unanswered.

Overall, the CLIMB study demonstrates that diversity of the microbiota may not be a determinant of disease severity in children with CSLD, namely CF and PCD. *Streptococcus* spp. are highly abundant in the airways of children with CSLD and an inverse relationship was seen in the relative abundances of *Streptococcus* spp. and *Haemophilus* spp. in NBS and PCD but not in CF. This suggests that a switch occurs in CF from a milder to a more severe phenotype that does not happen in PCD, with changes in the relative abundance of *Haemophilus* spp. and *Pseudomonas* spp. implicated as one potential driver of this. Further longitudinal study to determine when and how such a switch may occur may help to identify new therapeutic targets to promote a profile associated with a milder phenotype for children with CSLDs.

Future work

In the comparison of the upper and lower airway microbiota, the role of induced sputum as a surrogate for lower airway sampling in non-productive patients was not assessed. This needs to be addressed by performing a comparative study of induced sputum with TS and bronchial brushings taken from children undergoing a clinically indicated bronchoscopy for 16S rRNA gene sequencing to determine which is the optimum surrogate for the lower airway microbiota.

Induced sputum could therefore be considered as an alternative longitudinal sampling method in young children, but would only be employed if it offered substantial advantages over TS, given the time and effort needed to induce sputum.

Whether *Streptococcus* spp. and *Haemophilus* spp. are protective or pathogenic in CF is as yet unknown, as is the relationship between *Haemophilus* spp. and *Pseudomonas* spp. The findings in this thesis are exploratory and as such, it was not possible to determine the interactions between organisms in the microbiota. Future *in vivo* competition experiments using established animal models of CF, such as the newborn CF pig, are needed to explore this potential relationship further. Stoltz *et al.*²⁶⁸ demonstrated that, in contrast to wild type pigs, upon challenge with an aerosolised inoculum of *S. aureus* the airways of newborn CF pigs failed to eradicate *S. aureus* (determined from direct culture of lung tissue), suggesting that bacterial killing is impaired in CF pigs within a few hours of birth. In the experiment by Stoltz *et al.*²⁶⁸, *S. aureus* was chosen as it is a common pathogen in early CF disease. However, a similar experiment should be performed by nebulising each of the genera highlighted as potentially important in the CLIMB study, such as *Streptococcus*, or using bacterial cocktails of *Haemophilus* spp. with either *Streptococcus* spp. or *Pseudomonas* spp. If a pathogenic effect of either *Streptococcus* spp. or *Haemophilus* spp. was found using animal models, this may warrant re-evaluation of current antibiotic treatments in CF. *Streptococcus* specific PCR for speciation may also allow further insights into the role of different streptococci in early CF lung disease.

Overall, inclusion of specific PCRs to determine the species level changes occurring in the microbiota would be of great benefit.

The human microbiota is highly individual. Thus, performing large, longitudinal studies is important to better understand the relationship between the microbiota and health and disease. In total, 56 children with CF (not including NBS) were recruited for the CLIMB study. However, during this PhD, processing (16S rRNA gene sequencing and qPCR) of all the samples collected (approximately 1200 in total for the CLIMB study) could not be completed. Including the remaining 25 patients with CF for whom samples were collected would provide better statistical power to enable detection of changes with associated with clinical status, such as exacerbations and is a planned area for future work.

An interesting relationship, which was not explored in this study, is that of the gut and airway microbiota. A previous longitudinal study of paired stool and oropharyngeal samples from 7 infants obtained over the first 21 months of life with CF found that the presence of some genera in the airway microbiota, such as *Enterococcus*, was preceded by their presence in the gut¹⁴⁸. The mechanism by which this may occur is unclear but if such a relationship is shown to exist, this may have important implications as it suggests that orally administered treatments, such as pre- and probiotics, may be successful in modulating the airway microbiota in the treatment of CF airway disease. In addition, many children with CF receive a high fat diet, which may impact the gut microbiota²⁶⁹. Early life influences during the preschool years (< 5 years of age) appear to be

particularly important in the development of later chronic airway diseases²⁷⁰. Larger longitudinal studies are therefore needed comparing the TS and stool microbiota in NBS CF infants followed for up to 5 years are needed to further explore this relationship. This may be best achieved with a multi-centre study of NBS infants to provide adequate statistical power.

As research into the role of the airway microbiota in children with chronic lung diseases expands, there is a greater need to understand not only which bacteria are present in the airways of children with CSLDs but also their activity, both in relation to each other, other microbes (such as viruses and fungi) and their human host. Thus longitudinal studies combining metagenomic, metabolomic and proteomic platforms with changes in tissue inflammation and clinical features in children with CSLD could provide the greatest insight yet into the functional role of the microbiota. Through this, microbial targets could be identified for intervention studies aimed at modulating the airway microbiota, either indirectly through the use of oral probiotics, or directly through nebulisation of antibiotic cocktails for microbiota transfer.

Conclusions

This study has demonstrated that TS provide sufficient information to be used as surrogates for sampling the lower airway microbiota in longitudinal studies. In the CLIMB study, *Streptococcus* spp. were found to be highly abundant in children with CF and PCD of all ages. Similarities in the microbiota were seen between NBS and PCD, with in particular an inverse relationship seen between *Streptococcus* spp. and *Haemophilus* spp. This suggests that a switch occurs in CF

during childhood from a milder to a more pathological community composition. Further longitudinal work is needed to determine the importance of these findings in guiding therapeutic strategies in children with CSLDs.

References

1. FitzGerald JM. Targeting lung attacks. *Thorax*. 2011;66(5):365-366.
2. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*. 1998;95(12):6578-6583.
3. Leuwenhoek AV. An Extract of a Letter from Mr. Anth. Van Leuwenhoek, concerning Animalcules Found on the Teeth; Of the Scaleyness of the Skin, &c. *Philosophical Transactions*. January 1, 1693 1693;17(192-206):646-649.
4. Fredricks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clinical microbiology reviews*. Jan 1996;9(1):18-33.
5. Staley JT, Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual review of microbiology*. 1985;39:321-346.
6. Huang YJ, Marsland BJ, Bunyavanich S, *et al*. The Microbiome in Allergic Disease: Current Understanding and Future Opportunities - 2017 PRACTALL Document of the American Academy of Allergy, Asthma & Immunology and the European Academy of Allergy and Clinical Immunology. *The Journal of allergy and clinical immunology*. Feb 28 2017.
7. Hengstmann U, Chin KJ, Janssen PH, Liesack W. Comparative phylogenetic assignment of environmental sequences of genes encoding 16S rRNA and numerically abundant culturable bacteria from an anoxic rice paddy soil. *Applied and environmental microbiology*. Nov 1999;65(11):5050-5058.
8. Chin KJ, Hahn D, Hengstmann U, Liesack W, Janssen PH. Characterization and identification of numerically abundant culturable bacteria from the anoxic bulk soil of rice paddy microcosms. *Applied and environmental microbiology*. Nov 1999;65(11):5042-5049.

9. Browne HP, Forster SC, Anonye BO, *et al.* Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature*. 2016;533(7604):543-546.
10. Janda JM, Abbott SL. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. *Journal of clinical microbiology*. 07/11 2007;45(9):2761-2764.
11. Sim K, Powell E, Shaw AG, McClure Z, Bangham M, Kroll JS. The neonatal gastrointestinal microbiota: the foundation of future health? *Archives of Disease in Childhood - Fetal and Neonatal Edition*. 2013;98(4):F362-F364.
12. Ronaghi M. Pyrosequencing sheds light on DNA sequencing. *Genome research*. 2001;11(3 - 11).
13. Yarza P, Yilmaz P, Pruesse E, *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Micro*. 2014;12(9):635-645.
14. Rogers GB, Carroll MP, Serisier DJ, *et al.* Use of 16S rRNA gene profiling by terminal restriction fragment length polymorphism analysis to compare bacterial communities in sputum and mouthwash samples from patients with cystic fibrosis. *Journal of clinical microbiology*. Jul 2006;44(7):2601-2604.
15. Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Bruce KD. characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16S ribosomal DNA terminal restriction fragment length polymorphism profiling. *Journal of clinical microbiology*. Nov 2004;42(11):5176-5183.
16. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333-351.
17. Liu L, Li Y, Li S, *et al.* Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*. 2012;2012:11.
18. Cox MJ, Cookson WO, Moffatt MF. Sequencing the human microbiome in health and disease. *Human molecular genetics*. Oct 15 2013;22(R1):R88-94.

19. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology*. 2016;14(8):e1002533.
20. Grice EA, Kong HH, Renaud G, *et al*. A diversity profile of the human skin microbiota. *Genome research*. Jul 2008;18(7):1043-1050.
21. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Forensic identification using skin bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*. Apr 6 2010;107(14):6477-6481.
22. Wade WG. The oral microbiome in health and disease. *Pharmacological research*. Mar 2013;69(1):137-143.
23. Biesbroek G, Tsvitvadze E, Sanders EA, *et al*. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *American journal of respiratory and critical care medicine*. Dec 1 2014;190(11):1283-1292.
24. Bogaert D, Keijser B, Huse S, *et al*. Variability and Diversity of Nasopharyngeal Microbiota in Children: A Metagenomic Analysis. *PloS one*. 2011;6(2):e17035.
25. Ravel J, Gajer P, Abdo Z, *et al*. Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America*. Mar 15 2011;108 Suppl 1:4680-4687.
26. Byrd AL, Segre JA. Adapting Koch's postulates. *Science*. 2016;351(6270):224-226.
27. Hilty M, Burke C, Pedro H, *et al*. Disordered Microbial Communities in Asthmatic Airways. *PloS one*. 2010;5(1):e8578.
28. Petersson J, Glenny RW. Gas exchange and ventilation-perfusion relationships in the lung. *European Respiratory Journal*. 2014;44(4):1023-1041.
29. Ganesan S, Comstock AT, Sajjan US. Barrier function of airway tract epithelium. *Tissue Barriers*. 2013;1(4):e24997.
30. The European Respiratory Society. *Paediatric Respiratory Medicine*. First ed: The European Respiratory Society; 2013.

31. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 06/14/print 2012;486(7402):207-214.
32. Charlson ES, Bittinger K, Haas AR, *et al*. Topographical continuity of bacterial populations in the healthy human respiratory tract. *American journal of respiratory and critical care medicine*. Oct 15 2011;184(8):957-963.
33. Erb-Downward JR, Thompson DL, Han MK, *et al*. Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PloS one*. Feb 22 2011;6(2):e16384.
34. Einarsson GG, Comer DM, McIlreavey L, *et al*. Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers. *Thorax*. Sep 2016;71(9):795-803.
35. Morris A, Beck JM, Schloss PD, *et al*. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *American journal of respiratory and critical care medicine*. May 15 2013;187(10):1067-1075.
36. Bassis CM, Erb-Downward JR, Dickson RP, *et al*. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *mBio*. Mar 03 2015;6(2):e00037.
37. Mika M, Mack I, Korten I, *et al*. Dynamics of the nasal microbiota in infancy: a prospective cohort study. *The Journal of allergy and clinical immunology*. Apr 2015;135(4):905-912 e911.
38. Bosch AATM, Levin E, van Houten MA, *et al*. Development of Upper Respiratory Tract Microbiota in Infancy is Affected by Mode of Delivery. *EBioMedicine*. 2016;9:336-345.
39. Cardenas PA, Cooper PJ, Cox MJ, *et al*. Upper airways microbiota in antibiotic-naïve wheezing and healthy infants from the tropics of rural Ecuador. *PloS one*. 2012;7(10):e46803.
40. Mshvildadze M, Neu J, Shuster J, Theriaque D, Li N, Mai V. Intestinal microbial ecology in premature infants assessed with non-culture-based techniques. *The Journal of pediatrics*. Jan 2010;156(1):20-25.

41. Gosalbes MJ, Llop S, Valles Y, Moya A, Ballester F, Francino MP. Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*. Feb 2013;43(2):198-211.
42. Collado MC, Rautava S, Aakko J, Isolauri E, Salminen S. Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Scientific reports*. 2016;6:23129.
43. Jimenez E, Fernandez L, Marin ML, *et al*. Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. *Current microbiology*. Oct 2005;51(4):270-274.
44. Stout MJ, Conlon B, Landeau M, *et al*. Identification of intracellular bacteria in the basal plate of the human placenta in term and preterm gestations. *American journal of obstetrics and gynecology*. 2013;208(3):226.e221-226.e227.
45. Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. *Science translational medicine*. May 21 2014;6(237):237ra265.
46. Lauder AP, Roche AM, Sherrill-Mix S, *et al*. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome*. June 23 2016;4(1):29.
47. Dominguez-Bello MG, Costello EK, Contreras M, *et al*. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*. Jun 29 2010;107(26):11971-11975.
48. Lal CV, Travers C, Aghai ZH, *et al*. The Airway Microbiome at Birth. *Scientific reports*. 2016;6:31023.
49. Lohmann P, Luna RA, Hollister EB, *et al*. The airway microbiome of intubated premature infants: characteristics and changes that predict the development of bronchopulmonary dysplasia. *Pediatric research*. Sep 2014;76(3):294-301.

50. Palmer C, Bil EM, DiGuilio DB, Relman DA, Brown PO. Development of the Human Infant Intestinal Microbiota. *PLoS Biology*. 2007;5(7):e177.
51. Koenig JE, Spor A, Scalfone N, *et al*. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*. . 2011;108(Suppl 1):4578 - 4585.
52. VanDevanter DR, LiPuma JJ. Microbial diversity in the cystic fibrosis airways: where is thy sting? *Future microbiology*. Jul 2012;7(7):801-803.
53. Qin J, Li R, Raes J, *et al*. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. Mar 04 2010;464(7285):59-65.
54. Turnbaugh PJ, Hamady M, Yatsunenko T, *et al*. A core gut microbiome in obese and lean twins. *Nature*. Jan 22 2009;457(7228):480-484.
55. Stecher B, Chaffron S, Kappeli R, *et al*. Like will to like: abundances of closely related species can predict susceptibility to intestinal colonization by pathogenic and commensal bacteria. *PLoS pathogens*. Jan 2010;6(1):e1000711.
56. Fredricks DN, Fiedler TL, Marrazzo JM. Molecular identification of bacteria associated with bacterial vaginosis. *The New England journal of medicine*. Nov 03 2005;353(18):1899-1911.
57. Cardenas PA. Upper airways microbiota profiling in a case/control study between wheezing and healthy children from the tropics of Ecuador. London: National Heart and Lung Institute, Imperial College London; 2015.
58. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. *Genome medicine*. Apr 27 2016;8(1):51.
59. Lim MY, Yoon HS, Rho M, *et al*. Analysis of the association between host genetics, smoking, and sputum microbiota in healthy humans. *Scientific reports*. Mar 31 2016;6:23745.
60. Chu DM, Ma J, Prince AL, Antony KM, Seferovic MD, Aagaard KM. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nature medicine*. 2017;23(3):314-326.

61. Costello EK, Stagaman K, Dethlefsen L, Bohannon BJ, Relman DA. The application of ecological theory toward an understanding of the human microbiome. *Science*. Jun 08 2012;336(6086):1255-1262.
62. Dickson RP, Huffnagle GB. The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease. *PLoS pathogens*. Jul 2015;11(7):e1004923.
63. Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. *Lancet*. Aug 23 2014;384(9944):691-702.
64. Bisgaard H, Hermansen MN, Buchvald F, *et al*. Childhood Asthma after Bacterial Colonization of the Airway in Neonates. *New England Journal of Medicine*. 2007;357(15):1487-1495.
65. Chen Y, Blaser MJ. *Helicobacter pylori* colonization is inversely associated with childhood asthma. *The Journal of infectious diseases*. Aug 15 2008;198(4):553-560.
66. Gollwitzer ES, Saglani S, Trompette A, *et al*. Lung microbiota promotes tolerance to allergens in neonates via PD-L1. *Nature medicine*. Jun 2014;20(6):642-647.
67. Herbst T, Sichelstiel A, Schar C, *et al*. Dysregulation of allergic airway inflammation in the absence of microbial colonization. *American journal of respiratory and critical care medicine*. Jul 15 2011;184(2):198-205.
68. Deshmukh HS, Liu Y, Menkiti OR, *et al*. The microbiota regulates neutrophil homeostasis and host resistance to *Escherichia coli* K1 sepsis in neonatal mice. *Nature medicine*. May 2014;20(5):524-530.
69. Goleva E, Jackson LP, Harris JK, *et al*. The Effects of Airway Microbiome on Corticosteroid Responsiveness in Asthma. *American journal of respiratory and critical care medicine*. 2013;188(10):1193 - 1201.
70. Jung A, Kleinau I, Schonian G, *et al*. Sequential genotyping of *Pseudomonas aeruginosa* from upper and lower airways of cystic fibrosis patients. *European Respiratory Journal*. 2002;20(6):1457-1463.
71. Armstrong DS, Grimwood K, Carlin JB, Carzino R, Olinsky A, Phelan PD. Bronchoalveolar lavage or oropharyngeal cultures to identify lower

- respiratory pathogens in infants with cystic fibrosis. *Pediatric pulmonology*. May 1996;21(5):267-275.
72. Cystic Fibrosis Trust. Standards for the clinical care of children and adults with cystic fibrosis in the UK. Second edition. ed2011.
 73. Ahmed B, Balfour-Lynn IM, Alshafi K. Cough swabs should not be used to isolate non-tuberculous mycobacteria in children with cystic fibrosis. *Archives of disease in childhood*. Sep 2012;97(9):854-855.
 74. Klepac-Ceraj V, Lemon KP, Martin TR, *et al*. Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and *Pseudomonas aeruginosa*. *Environmental microbiology*. May 2010;12(5):1293-1303.
 75. Chang AB, Grimwood K, Maguire G, King PT, Morris PS, Torzillo PJ. Management of bronchiectasis and chronic suppurative lung disease in indigenous children and adults from rural and remote Australian communities. *The Medical journal of Australia*. Oct 6 2008;189(7):386-393.
 76. Li AM, Sonnappa S, Lex C, *et al*. Non-CF bronchiectasis: does knowing the aetiology lead to changes in management? *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*. Jul 2005;26(1):8-14.
 77. Goyal V, Grimwood K, Marchant J, Masters IB, Chang AB. Pediatric bronchiectasis: No longer an orphan disease. *Pediatric pulmonology*. May 2016;51(5):450-469.
 78. Gaillard EA, Carty H, Heaf D, Smyth RL. Reversible bronchial dilatation in children: comparison of serial high-resolution computer tomography scans of the lungs. *European journal of radiology*. Sep 2003;47(3):215-220.
 79. Mott LS, Park J, Murray CP, *et al*. Progression of early structural lung disease in young children with cystic fibrosis assessed using CT. *Thorax*. Jun 2012;67(6):509-516.
 80. Dodge JA, Morison S, Lewis PA, *et al*. Incidence, population, and survival of cystic fibrosis in the UK, 1968-95. UK Cystic Fibrosis Survey Management Committee. *Archives of disease in childhood*. Dec 1997;77(6):493-496.

81. Dodge JA, Lewis PA, Stanton M, Wilsher J. Cystic fibrosis mortality and survival in the UK: 1947-2003. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*. Mar 2007;29(3):522-526.
82. Cystic Fibrosis Foundation Patient Registry 1994 Annual Data Report. Bethesda, MD, USA: Cystic Fibrosis Foundation;1995.
83. Lim MT, Wallis C, Price JF, *et al*. Diagnosis of cystic fibrosis in London and South East England before and after the introduction of newborn screening. *Archives of disease in childhood*. Mar 2014;99(3):197-202.
84. Elborn JS. Cystic fibrosis. *Lancet*. Nov 19 2016;388(10059):2519-2531.
85. Rogan MP, Stoltz DA, Hornick DB. Cystic fibrosis transmembrane conductance regulator intracellular processing, trafficking, and opportunities for mutation-specific treatment. *Chest*. Jun 2011;139(6):1480-1490.
86. Boyle MP, De Boeck K. A new era in the treatment of cystic fibrosis: correction of the underlying CFTR defect. *The Lancet. Respiratory medicine*. Apr 2013;1(2):158-163.
87. Ramsey BW, Davies J, McElvaney NG, *et al*. A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. *The New England journal of medicine*. Nov 03 2011;365(18):1663-1672.
88. Cutting GR. Modifier genes in Mendelian disorders: the example of cystic fibrosis. *Annals of the New York Academy of Sciences*. Dec 2010;1214:57-69.
89. Vanscoy LL, Blackman SM, Collaco JM, *et al*. Heritability of lung disease severity in cystic fibrosis. *American journal of respiratory and critical care medicine*. May 15 2007;175(10):1036-1043.
90. Steinkamp G, von der Hardt H. Improvement of nutritional status and lung function after long-term nocturnal gastrostomy feedings in cystic fibrosis. *The Journal of pediatrics*. 124(2):244-249.
91. Parad RB, Gerard CJ, Zurakowski D, Nichols DP, Pier GB. Pulmonary Outcome in Cystic Fibrosis Is Influenced Primarily by Mucoid *Pseudomonas aeruginosa* Infection and Immune Status and Only Modestly by Genotype. *Infection and immunity*. 1999;67(9):4744-4750.

92. Rubin BK. Exposure of Children with Cystic Fibrosis to Environmental Tobacco Smoke. *New England Journal of Medicine*. 1990;323(12):782-788.
93. Schechter MS, Shelton BJ, Margolis PA, Fitzsimmons SC. The association of socioeconomic status with outcomes in cystic fibrosis patients in the United States. *American journal of respiratory and critical care medicine*. May 2001;163(6):1331-1337.
94. Sheppard DN, Welsh MJ. Structure and function of the CFTR chloride channel. *Physiological reviews*. Jan 1999;79(1 Suppl):S23-45.
95. Poulsen JH, Fischer H, Illek B, Machen TE. Bicarbonate conductance and pH regulatory capability of cystic fibrosis transmembrane conductance regulator. *Proceedings of the National Academy of Sciences of the United States of America*. Jun 07 1994;91(12):5340-5344.
96. Haq IJ, Gray MA, Garnett JP, Ward C, Brodlie M. Airway surface liquid homeostasis in cystic fibrosis: pathophysiology and therapeutic targets. *Thorax*. Mar 2016;71(3):284-287.
97. Matsui H, Grubb BR, Tarran R, et al. Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell*. Dec 23 1998;95(7):1005-1015.
98. Ahmed B, Bush A, Davies JC. How to use: bacterial cultures in diagnosing lower respiratory tract infections in cystic fibrosis. *Archives of disease in childhood. Education and practice edition* 2014; 99(5): 181-7.
99. Pezzulo AA, Tang XX, Hoegger MJ, *et al*. Reduced airway surface pH impairs bacterial killing in the porcine cystic fibrosis lung. *Nature*. Jul 04 2012;487(7405):109-113.
100. Stutts M, Canessa C, Olsen J, et al. CFTR as a cAMP-dependent regulator of sodium channels. *Science*. 1995;269(5225):847-850.
101. Stoltz DA, Meyerholz DK, Welsh MJ. Origins of cystic fibrosis lung disease. *The New England journal of medicine*. Jan 22 2015;372(4):351-362.
102. Tarran R, Button B, Picher M, et al. Normal and cystic fibrosis airway surface liquid homeostasis. The effects of phasic shear stress and viral infections. *The Journal of biological chemistry*. Oct 21 2005;280(42):35751-35759.

- 103.** Bacconnais S, Tirouvanziam R, Zahm JM, *et al.* Ion composition and rheology of airway liquid from cystic fibrosis fetal tracheal xenografts. American journal of respiratory cell and molecular biology. Apr 1999;20(4):605-611.
- 104.** Thursfield RM, Bush A, Alton EW, Davies JC. Airway inflammation is present by 4 months in CF infants diagnosed on newborn screening. Thorax. 2012;67:A40-A41.
- 105.** Stafler P, Davies JC, Balfour-Lynn IM, Rosenthal M, Bush A. Bronchoscopy in cystic fibrosis infants diagnosed by newborn screening. Pediatric pulmonology. Jul 2011;46(7):696-700.
- 106.** Downey DG, Bell SC, Elborn JS. Neutrophils in cystic fibrosis. Thorax. Jan 2009;64(1):81-88.
- 107.** Cohen-Cymberknoh M, Kerem E, Ferkol T, Elizur A. Airway inflammation in cystic fibrosis: molecular mechanisms and clinical implications. Thorax. Dec 2013;68(12):1157-1162.
- 108.** Cantin AM, Hartl D, Konstan MW, Chmiel JF. Inflammation in cystic fibrosis lung disease: Pathogenesis and therapy. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society. Jul 2015;14(4):419-430.
- 109.** O'Sullivan BP, Freedman SD. Cystic fibrosis. Lancet. May 30 2009;373(9678):1891-1904.
- 110.** Sanders DB, Bittner RC, Rosenfeld M, Hoffman LR, Redding GJ, Goss CH. Failure to recover to baseline pulmonary function after cystic fibrosis pulmonary exacerbation. American journal of respiratory and critical care medicine. Sep 01 2010;182(5):627-632.
- 111.** Waters V, Stanojevic S, Atenafu EG, *et al.* Effect of pulmonary exacerbations on long-term lung function decline in cystic fibrosis. The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology. Jul 2012;40(1):61-66.
- 112.** Byrnes CA, Vidmar S, Cheney JL, *et al.* Prospective evaluation of respiratory exacerbations in children with cystic fibrosis from newborn screening to 5 years of age. Thorax. Jan 23 2013.

113. Henry RL, Mellis CM, Petrovic L. Mucoïd *Pseudomonas aeruginosa* is a marker of poor survival in cystic fibrosis. *Pediatric pulmonology*. Mar 1992;12(3):158-161.
114. Jones AM, Dodd ME, Govan JR, *et al.* *Burkholderia cenocepacia* and *Burkholderia multivorans*: influence on survival in cystic fibrosis. *Thorax*. Nov 2004;59(11):948-951.
115. Zemanick ET, Wagner BD, Harris JK, Wagener JS, Accurso FJ, Sagel SD. Pulmonary exacerbations in cystic fibrosis with negative bacterial cultures. *Pediatric pulmonology*. Jun 2010;45(6):569-577.
116. UK Cystic Fibrosis Registry. Annual Data Report. . 2013;
<http://www.cysticfibrosis.org.uk/media/598466/annual-data-report-2013-jul14.pdf>.
117. Burns JL, Gibson RL, McNamara S, *et al.* Longitudinal assessment of *Pseudomonas aeruginosa* in young children with cystic fibrosis. *Journal of Infectious Diseases*. 2001;183(3):444-452.
118. Sommer LM, Alanin MC, Marvig RL, *et al.* Bacterial evolution in PCD and CF patients follows the same mutational steps. *Scientific reports*. 2016;6:28732.
119. Hector A, Kirn T, Ralhan A, *et al.* Microbial colonization and lung function in adolescents with cystic fibrosis. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*. May 2016;15(3):340-349.
120. Pillarisetti N, Williamson E, Linnane B, *et al.* Infection, inflammation, and lung function decline in infants with cystic fibrosis. *American journal of respiratory and critical care medicine*. Jul 01 2011;184(1):75-81.
121. Smith WD, E. Bardin, L. Cameron, *et al.* Current and future therapies for *Pseudomonas aeruginosa* infection in patients with cystic fibrosis. . *FEMS Microbiol Lett* 2017; fnx121. .
122. Junge S, Gorlich D, den Reijer M, *et al.* Factors Associated with Worse Lung Function in Cystic Fibrosis Patients with Persistent *Staphylococcus aureus*. *PloS one*. 2016;11(11):e0166220.
123. Sagel SD, Gibson RL, Emerson J, *et al.* Impact of *Pseudomonas* and *Staphylococcus* infection on inflammation and clinical status in young

- children with cystic fibrosis. *The Journal of pediatrics*. Feb 2009;154(2):183-188.
124. Ahlgren HG, Benedetti A, Landry JS, *et al*. Clinical outcomes associated with *Staphylococcus aureus* and *Pseudomonas aeruginosa* airway infections in adult cystic fibrosis patients. *BMC pulmonary medicine*. 2015;15:67.
 125. Stutman HR, Lieberman JM, Nussbaum E, Marks MI. Antibiotic prophylaxis in infants and young children with cystic fibrosis: a randomized controlled trial. *The Journal of pediatrics*. Mar 2002;140(3):299-305.
 126. Smyth AR, Rosenfeld M. Prophylactic anti-staphylococcal antibiotics for cystic fibrosis. *Cochrane database of systematic reviews (Online)*. Apr 18 2017;4:CD001912.
 127. Gibson RL, Burns JL, Ramsey BW. Pathophysiology and management of pulmonary infections in cystic fibrosis. *American journal of respiratory and critical care medicine*. Oct 15 2003;168(8):918-951.
 128. VandenBranden SL, McMullen A, Schechter MS, *et al*. Lung function decline from adolescence to young adulthood in cystic fibrosis. *Pediatric pulmonology*. 2012;47(2):135-143.
 129. Davies JC, Rubin BK. Emerging and unusual gram-negative infections in cystic fibrosis. *Seminars in respiratory and critical care medicine*. Jun 2007;28(3):312-321.
 130. Horsley A, Webb K, Bright-Thomas R, Govan J, Jones A. Can early *Burkholderia cepacia* complex infection in cystic fibrosis be eradicated with antibiotic therapy? *Frontiers in cellular and infection microbiology*. 2011;1:18.
 131. McMenamin JD, Zacccone TM, Coenye T, Vandamme P, LiPuma JJ. Misidentification of *Burkholderia cepacia* in US cystic fibrosis treatment centers: an analysis of 1,051 recent sputum isolates. *Chest*. Jun 2000;117(6):1661-1665.
 132. Tunney MM, Field TR, Moriarty TF, *et al*. Detection of Anaerobic Bacteria in High Numbers in Sputum from Patients with Cystic Fibrosis. *American journal of respiratory and critical care medicine*. 2008;177(9):995-1001.

133. Tunney MM, Klem ER, Fodor AA, *et al.* Use of culture and molecular analysis to determine the effect of antibiotic treatment on microbial community diversity and abundance during exacerbation in patients with cystic fibrosis. *Thorax*. Jul 2011;66(7):579-584.
134. O'Neill K, Bradley JM, Johnston E, *et al.* Reduced bacterial colony count of anaerobic bacteria is associated with a worsening in lung clearance index and inflammation in cystic fibrosis. *PloS one*. 2015;10(5):e0126980.
135. Cox MJ, Allgaier M, Taylor B, *et al.* Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PloS one*. 2010;5(6):e11044.
136. van der Gast CJ, Walker AW, Stressmann FA, *et al.* Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *The ISME journal*. May 2011;5(5):780-791.
137. Tunney MM, Field TR, Moriarty TF, *et al.* Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. *American journal of respiratory and critical care medicine*. May 1 2008;177(9):995-1001.
138. Rogers GB, Carroll MP, Serisier DJ, *et al.* Bacterial activity in cystic fibrosis lung infections. *Respiratory research*. 2005;6:49.
139. Zhao J, Schloss PD, Kalikin LM, *et al.* Decade-long bacterial community dynamics in cystic fibrosis airways. *Proceedings of the National Academy of Sciences of the United States of America*. Apr 10 2012;109(15):5809-5814.
140. Delhaes L, Monchy S, Frealle E, *et al.* The airway microbiota in cystic fibrosis: a complex fungal and bacterial community--implications for therapeutic management. *PloS one*. 2012;7(4):e36313.
141. Hunter RC, Klepac-Ceraj V, Lorenzi MM, Grotzinger H, Martin TR, Newman DK. Phenazine content in the cystic fibrosis respiratory tract negatively correlates with lung function and microbial complexity. *American journal of respiratory cell and molecular biology*. Dec 2012;47(6):738-745.
142. Stressmann FA, Rogers GB, van der Gast CJ, *et al.* Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial

- communities infecting the adult cystic fibrosis lung show stability and resilience. *Thorax*. Oct 2012;67(10):867-873.
143. Fodor AA, Klem ER, Gilpin DF, *et al*. The adult cystic fibrosis airway microbiota is stable over time and infection type, and highly resilient to antibiotic treatment of exacerbations. *PloS one*. 2012;7(9):e45001.
 144. Cuthbertson L, Rogers GB, Walker AW, *et al*. Respiratory microbiota resistance and resilience to pulmonary exacerbation and subsequent antimicrobial intervention. *The ISME journal*. May 2016;10(5):1081-1091.
 145. Coburn B, Wang PW, Diaz Caballero J, *et al*. Lung microbiota across age and disease stage in cystic fibrosis. *Scientific reports*. 2015;5:10241.
 146. Harris JK, De Groote MA, Sagel SD, *et al*. Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis. *Proceedings of the National Academy of Sciences of the United States of America*. Dec 18 2007;104(51):20529-20533.
 147. Utter DR, Mark Welch JL, Borisy GG. Individuality, Stability, and Variability of the Plaque Microbiome. *Frontiers in Microbiology*. 2016;7:564.
 148. Madan JC, Koestler DC, Stanton BA, *et al*. Serial analysis of the gut and respiratory microbiome in cystic fibrosis in infancy: interaction between intestinal and respiratory tracts and impact of nutritional exposures. *mBio*. 2012;3(4).
 149. Prevaes SM, de Winter-de Groot KM, Janssens HM, *et al*. Development of the Nasopharyngeal Microbiota in Infants with Cystic Fibrosis. *American journal of respiratory and critical care medicine*. Oct 22 2015.
 150. Mika M, Korten I, Qi W, *et al*. The nasal microbiota in infants with cystic fibrosis in the first year of life: a prospective cohort study. *The Lancet. Respiratory medicine*. May 11 2016.
 151. Marsh RL, Kaestli M, Chang AB, *et al*. The microbiota in bronchoalveolar lavage from young children with chronic lung disease includes taxa present in both the oropharynx and nasopharynx. *Microbiome*. 2016;4(1):37.

152. Hoen AG, Li J, Moulton LA, et al. Associations between Gut Microbial Colonization in Early Life and Respiratory Outcomes in Cystic Fibrosis. *The Journal of pediatrics*. Jul 2015;167(1):138-147 e131-133.
153. Shade A, Peter H, Allison SD, *et al*. Fundamentals of microbial community resistance and resilience. *Front Microbiol*. 2012;3:417.
154. Peterfreund GL, Vandivier LE, Sinha R, *et al*. Succession in the gut microbiome following antibiotic and antibody therapies for clostridium difficile. . *PloS one*. 2012;7(10):e46966.
155. Goss CH, Burns JL. Exacerbations in cystic fibrosis. 1: Epidemiology and pathogenesis. *Thorax*. Apr 2007;62(4):360-367.
156. Stressmann FA, Rogers GB, Marsh P, *et al*. Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*. Sep 2011;10(5):357-365.
157. Filkins LM, Hampton TH, Gifford AH, *et al*. Prevalence of streptococci and increased polymicrobial diversity associated with cystic fibrosis patient stability. *Journal of bacteriology*. Sep 2012;194(17):4709-4717.
158. Parkins MD, Sibley CD, Surette MG, Rabin HR. The *Streptococcus milleri* group--an unrecognized cause of disease in cystic fibrosis: a case series and literature review. *Pediatric pulmonology*. May 2008;43(5):490-497.
159. Sibley CD, Parkins MD, Rabin HR, Duan K, Norgaard JC, Surette MG. A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. *Proceedings of the National Academy of Sciences of the United States of America*. Sep 30 2008;105(39):15070-15075.
160. Carmody LA, Zhao J, Schloss PD, *et al*. Changes in cystic fibrosis airway microbiota at pulmonary exacerbation. *Annals of the American Thoracic Society*. Jun 2013;10(3):179-187.
161. Hilliard TN, Sukhani S, Francis J, *et al*. Bronchoscopy following diagnosis with cystic fibrosis. *Archives of disease in childhood*. Oct 2007;92(10):898-899.
162. Field TR, Sibley CD, Parkins MD, Rabin HR, Surette MG. The genus *Prevotella* in cystic fibrosis airways. *Anaerobe*. Aug 2010;16(4):337-344.

163. Sherrard LJ, Bell SC, Tunney MM. The role of anaerobic bacteria in the cystic fibrosis airway. *Current opinion in pulmonary medicine*. Jul 21 2016.
164. Takahashi N. Oral Microbiome Metabolism: From "Who Are They?" to "What Are They Doing?". *Journal of dental research*. Dec 2015;94(12):1628-1637.
165. Bernarde C, Keravec M, Mounier J, *et al*. Impact of the CFTR-potentiator ivacaftor on airway microbiota in cystic fibrosis patients carrying a G551D mutation. *PloS one*. 2015;10(4):e0124124.
166. Duan K, Dammel C, Stein J, Rabin H, Surette MG. Modulation of *Pseudomonas aeruginosa* gene expression by host microflora through interspecies communication. *Molecular Microbiology*. 2003;50(5):1477-1491.
167. Mirkovic B, Murray MA, Lavelle GM, *et al*. The Role of Short-Chain Fatty Acids, Produced by Anaerobic Bacteria, in the Cystic Fibrosis Airway. *American journal of respiratory and critical care medicine*. Dec 1 2015;192(11):1314-1324.
168. Ghorbani P, Santhakumar P, Hu Q, *et al*. Short-chain fatty acids affect cystic fibrosis airway inflammation and bacterial growth. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*. Oct 2015;46(4):1033-1045.
169. Daniels TW, Rogers GB, Stressmann FA, *et al*. Impact of antibiotic treatment for pulmonary exacerbations on bacterial diversity in cystic fibrosis. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*. Jan 2013;12(1):22-28.
170. Bhatti MA, Frank MO. *Veillonella parvula* Meningitis: Case Report and Review of *Veillonella* Infections. *Clinical Infectious Diseases*. 2000;31(3):839-840.
171. Mosquera JD, Zabalza M, Laniero M, Blanco JR. Endocarditis due to *Gemella haemolysans* in a patient with hemochromatosis. *Clinical Microbiology and Infection*. 6(10):566-568.

172. Cargill JS, Scott KS, Gascoyne-Binzi D, Sandoe JA. Granulicatella infection: diagnosis and management. *Journal of medical microbiology*. Jun 2012;61(Pt 6):755-761.
173. Zemanick ET, Harris JK, Wagner BD, *et al.* Inflammation and Airway Microbiota during Cystic Fibrosis Pulmonary Exacerbations. *PloS one*. 2013;8(4):e62917.
174. Asam D, Spellerberg B. Molecular pathogenicity of *Streptococcus anginosus*. *Molecular oral microbiology*. Aug 2014;29(4):145-155.
175. Schippa S, Iebba V, Santangelo F, *et al.* Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Allelic Variants Relate to Shifts in Faecal Microbiota of Cystic Fibrosis Patients. *PloS one*. 2013;8(4):e61176.
176. Malfroot A, Dab I. New insights on gastro-oesophageal reflux in cystic fibrosis by longitudinal follow up. *Archives of Disease in Childhood*. 1991;66:1339 - 1345.
177. Brodzicki J, Trawinska M, Korzon M. Frequency, consequences and pharmacological treatment of gastrooesophageal reflux in children with cystic fibrosis. *Medical Science Monitor*. 2002;8:CR529 - CR537.
178. Bruzzese E, Raia V, Spagnuolo MI, *et al.* Effect of *Lactobacillus GG* supplementation on pulmonary exacerbations in patients with cystic fibrosis: a pilot study. *Clinical nutrition (Edinburgh, Scotland)*. Jun 2007;26(3):322-328.
179. Weiss B, Bujanover Y, Yahav Y, Vilozni D, Fireman E, Efrati O. Probiotic supplementation affects pulmonary exacerbations in patients with cystic fibrosis: a pilot study. *Pediatric pulmonology*. Jun 2010;45(6):536-540.
180. van Nood E, Vrieze A, Nieuwdorp M, *et al.* Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *The New England journal of medicine*. Jan 31 2013;368(5):407-415.
181. Dominguez-Bello MG, De Jesus-Laboy KM, Shen N, *et al.* Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nature medicine*. Mar 2016;22(3):250-253.
182. Lucas JS, Burgess A, Mitchison HM, Moya E, Williamson M, Hogg C. Diagnosis and management of primary ciliary dyskinesia. *Archives of disease in childhood*. Sep 2014;99(9):850-856.

183. Sommer JU, Schafer K, Omran H, *et al.* ENT manifestations in patients with primary ciliary dyskinesia: prevalence and significance of otorhinolaryngologic co-morbidities. *European archives of oto-rhino-laryngology : official journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS) : affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery*. Mar 2011;268(3):383-388.
184. Bush A, Chodhari R, Collins N, *et al.* Primary ciliary dyskinesia: current state of the art. *Archives of disease in childhood*. Dec 2007;92(12):1136-1140.
185. Shoemark A, Dixon M, Corrin B, Dewar A. Twenty-year review of quantitative transmission electron microscopy for the diagnosis of primary ciliary dyskinesia. *Journal of clinical pathology*. Mar 2012;65(3):267-271.
186. Kouis P, Yiallourous PK, Middleton N, Evans JS, Kyriacou K, Papatheodorou SI. Prevalence of primary ciliary dyskinesia in consecutive referrals of suspect cases and the transmission electron microscopy detection rate: a systematic review and meta-analysis. *Pediatric research*. Mar 2017;81(3):398-405.
187. Lucas JS, Barbato A, Collins SA, *et al.* European Respiratory Society guidelines for the diagnosis of primary ciliary dyskinesia. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*. Jan 2017;49(1).
188. Shoemark A, Frost E, Dixon M, *et al.* Accuracy of Immunofluorescence in the Diagnosis of Primary Ciliary Dyskinesia. *American journal of respiratory and critical care medicine*. Feb 15 2017.
189. Horani A, Ferkol TW, Dutcher SK, Brody SL. Genetics and biology of primary ciliary dyskinesia. *Paediatric respiratory reviews*. Mar 2016;18:18-24.
190. Knowles MR, Daniels LA, Davis SD, Zariwala MA, Leigh MW. Primary ciliary dyskinesia. Recent advances in diagnostics, genetics, and characterization of clinical disease. *American journal of respiratory and critical care medicine*. Oct 15 2013;188(8):913-922.

- 191.** Hornef N, Olbrich H, Horvath J, *et al.* DNAH5 Mutations Are a Common Cause of Primary Ciliary Dyskinesia with Outer Dynein Arm Defects. American journal of respiratory and critical care medicine. 2006;174(2):120-126.
- 192.** Zariwala MA, Leigh MW, Ceppa F, *et al.* Mutations of DNAI1 in primary ciliary dyskinesia: evidence of founder effect in a common mutation. American journal of respiratory and critical care medicine. Oct 15 2006;174(8):858-866.
- 193.** Antony D, Becker-Heck A, Zariwala MA, *et al.* Mutations in CCDC39 and CCDC40 are the major cause of primary ciliary dyskinesia with axonemal disorganization and absent inner dynein arms. Human mutation. Mar 2013;34(3):462-472.
- 194.** Bush A, Payne D, Pike S, Jenkins G, Henke MO, Rubin BK. Mucus properties in children with primary ciliary dyskinesia: comparison with cystic fibrosis. Chest. Jan 2006;129(1):118-123.
- 195.** Ratjen F, Waters V, Klingel M, *et al.* Changes in airway inflammation during pulmonary exacerbations in patients with cystic fibrosis and primary ciliary dyskinesia. The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology. Mar 2016;47(3):829-836.
- 196.** Noone PG, Leigh MW, Sannuti A, *et al.* Primary ciliary dyskinesia: diagnostic and phenotypic features. American journal of respiratory and critical care medicine. Feb 15 2004;169(4):459-467.
- 197.** Marthin JK, Petersen N, Skovgaard LT, Nielsen KG. Lung Function in Patients with Primary Ciliary Dyskinesia. American journal of respiratory and critical care medicine. 2010;181(11):1262-1268.
- 198.** Sunther M, Bush A, Hogg C, McCann L, Carr SB. Recovery of baseline lung function after pulmonary exacerbation in children with primary ciliary dyskinesia. Pediatric pulmonology. Dec 2016;51(12):1362-1366.
- 199.** Worlitzsch D, Tarran R, Ulrich M, *et al.* Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients. Journal of Clinical Investigation. 2002;109(3):317-325.

200. van der Gast CJ, Cuthbertson L, Rogers GB, *et al.* Three Clinically Distinct Chronic Pediatric Airway Infections Share a Common Core Microbiota. *Annals of the American Thoracic Society*. Mar 5 2014.
201. Maglione M, Bush A, Nielsen KG, *et al.* Multicenter analysis of body mass index, lung function, and sputum microbiology in primary ciliary dyskinesia. *Pediatric pulmonology*. Dec 2014;49(12):1243-1250.
202. Rogers GB, Carroll MP, Zain NM, *et al.* Complexity, temporal stability, and clinical correlates of airway bacterial community composition in primary ciliary dyskinesia. *Journal of clinical microbiology*. Dec 2013;51(12):4029-4035.
203. Farrell PM, White TB, Ren CL, *et al.* Diagnosis of Cystic Fibrosis: Consensus Guidelines from the Cystic Fibrosis Foundation. *The Journal of pediatrics*. Feb 2017;181S:S4-S15 e11.
204. Bush A, Hogg C. Primary ciliary dyskinesia: recent advances in epidemiology, diagnosis, management and relationship with the expanding spectrum of ciliopathy. *Expert review of respiratory medicine*. Dec 2012;6(6):663-682.
205. Alton EW, Armstrong DK, Ashby D, *et al.* Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial. *The Lancet. Respiratory medicine*. Sep 2015;3(9):684-691.
206. Rosenthal M, Bain SH, Cramer D, *et al.* Lung function in white children aged 4 to 19 years: I--Spirometry. *Thorax*. 1993;48(8):794-802.
207. Lee TWR, Brownlee KG, Conway SP, Denton M, Littlewood JM. Evaluation of a new definition for chronic *Pseudomonas aeruginosa* infection in cystic fibrosis patients. *Journal of Cystic Fibrosis*. 2003;2(1):29-34.
208. Cystic Fibrosis Trust. Standards of Care: Laboratory standards for processing microbiological samples from people with cystic fibrosis. London, UK. September 2010.
209. Cuthbertson L, Rogers GB, Walker AW, *et al.* Time between collection and storage significantly influences bacterial sequence composition in sputum samples from cystic fibrosis respiratory infections. *Journal of clinical microbiology*. Aug 2014;52(8):3011-3016.

210. Salter SJ, Cox MJ, Turek EM, *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 2014;12:87.
211. 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.* Jan 7 2013;41(1):e1.
212. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology.* Sep 2013;79(17):5112-5120.
213. Sambrook J, Fritsch EF, Maniatis T. Gel electrophoresis of DNA. *Molecular cloning: a laboratory manual.* Second Edition. ed: Cold Spring Harbour Laboratory; 1989.
214. Caporaso JG, Kuczynski J, Stombaugh J, *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature methods.* May 2010;7(5):335-336.
215. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England).* Oct 1 2010;26(19):2460-2461.
216. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics (Oxford, England).* Jan 15 2010;26(2):266-267.
217. Haas BJ, Gevers D, Earl AM, *et al.* Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome research.* Mar 2011;21(3):494-504.
218. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and environmental microbiology.* 2007;73(16):5261-5267.
219. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS one.* 2010;5(3):e9490.
220. Aronesty E. ea-utils: Command-line tools for processing biological sequencing data. *Expression Analysis, Durham, NC.* 2011.

221. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* (Oxford, England). Mar 1 2010;26(5):589-595.
222. Quast C, Pruesse E, Yilmaz P, *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*. January 1, 2013 2013;41(D1):D590-D596.
223. Paradis E, Claude J, K S. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* (Oxford, England). 2004;20:289 - 290.
224. H P, P A, R G, S D. Biostrings: String objects representing biological sequences, and matching algorithms.
225. Wickham H. *ggplot2: elegant graphics for data analysis*. Springer, New York. 2009.
226. Auguie B. gridExtra: Miscellaneous Functions for "Grid" Graphics. . 2016; <https://CRAN.R-project.org/package=gridExtra>.
227. Pollard KS, Dudoit S, van der Laan MJ. Multiple Testing Procedures: R multtest Package and Applications to Genomics. In: Springer, ed. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. 2005.
228. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one*. 2013;8(4):e61217.
229. Wickham H. The Split_Apply_Combine Strategy for Data Analysis. *Journal of Statistical Software*. 2011;40(1):1 - 29.
230. Wickham H. Reshaping Data with the reshape Package. *Journal of Statistical Software*. 2007;21(12):1 - 20.
231. Dixon P. VEGAN, a package for R functions for community ecology. *Journal of Vegetation Science*. 2003;14:927 - 930.
232. Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*. 1957;27(4):325-349.
233. Lozupone C, Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and environmental microbiology*. Dec 2005;71(12):8228-8235.

234. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and Qualitative β Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Applied and environmental microbiology*. 2007;73(5):1576-1585.
235. de Blic J, Midulla F, Barbato A, *et al.* Bronchoalveolar lavage in children. ERS Task Force on bronchoalveolar lavage in children. *European Respiratory Society. The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*. Jan 2000;15(1):217-231.
236. Gutierrez JP, Grimwood K, Armstrong DS, *et al.* Interlobar differences in bronchoalveolar lavage fluid from children with cystic fibrosis. *European Respiratory Journal*. 2001;17:281 - 286.
237. Parker RL, Haesaert SP, Kovnat DM, Bachus B, Snider GL. Bronchial brushing in bronchogenic carcinoma. Factors influencing cellular yield and diagnostic accuracy. *Chest*. Mar 1977;71(3):341-345.
238. Farrell PM, Rosenstein BJ, White TB, *et al.* Guidelines for diagnosis of cystic fibrosis in newborns through older adults: Cystic Fibrosis Foundation consensus report. *The Journal of pediatrics*. Aug 2008;153(2):S4-S14.
239. Oh J, Byrd AL, Deming C, *et al.* Biogeography and individuality shape function in the human skin metagenome. *Nature*. 2014;514(7520):59-64.
240. Boutin S, Graeber SY, Weitnauer M, *et al.* Comparison of Microbiomes from Different Niches of Upper and Lower Airways in Children and Adolescents with Cystic Fibrosis. *PloS one*. 2015;10(1):e0116029.
241. Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N, Knight R. Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nature methods*. Oct 2010;7(10):813-819.
242. Douglas Bates, Martin Mächler, Ben Bolker, Walker. S. Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software*. 2015;67(1):1-48.
243. Jose Pinheiro, Douglas Bates, Saikat DebRoy , Deepayan Sarkar, Team. RC. nlme: Linear and Nonlinear Mixed Effects Models. 2017; <https://CRAN.R-project.org/package=nlme>.

244. Bartoń K. MuMIn: Multi-Model Inference. 2016; <https://CRAN.R-project.org/package=MuMIn>.
245. Dupuis A, Keenan K, Ooi CY, *et al*. Prevalence of meconium ileus marks the severity of mutations of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene. *Genetics in medicine : official journal of the American College of Medical Genetics*. Apr 2016;18(4):333-340.
246. Clain J, Lehmann-Che J, Dugueperoux I, *et al*. Misprocessing of the CFTR protein leads to mild cystic fibrosis phenotype. *Human mutation*. Apr 2005;25(4):360-371.
247. Veit G, Avramescu RG, Chiang AN, *et al*. From CFTR biology toward combinatorial pharmacotherapy: expanded classification of cystic fibrosis mutations. *Molecular biology of the cell*. Feb 01 2016;27(3):424-433.
248. Whittaker RH. Evolution and Measurement of Species Diversity. *Taxon*. 1972;21(2/3):213-251.
249. Mahboubi MA, Carmody LA, Foster BK, Kalikin LM, VanDevanter DR, LiPuma JJ. Culture-Based and Culture-Independent Bacteriologic Analysis of Cystic Fibrosis Respiratory Specimens. *Journal of clinical microbiology*. Mar 2016;54(3):613-619.
250. Tikhomirova A, Trappetti C, Paton JC, Kidd SP. The outcome of *H. influenzae* and *S. pneumoniae* inter-species interactions depends on pH, nutrient availability and growth phase. *International journal of medical microbiology : IJMM*. Dec 2015;305(8):881-892.
251. Rogers GB, Zain NM, Bruce KD, *et al*. A novel microbiota stratification system predicts future exacerbations in bronchiectasis. *Annals of the American Thoracic Society*. May 2014;11(4):496-503.
252. Sherrard LJ, McGrath SJ, McIlreavey L, *et al*. Production of extended-spectrum beta-lactamases and the potential indirect pathogenic role of *Prevotella* isolates from the cystic fibrosis respiratory microbiota. *International journal of antimicrobial agents*. Feb 2016;47(2):140-145.
253. Stanojevic S, McDonald A, Waters V, *et al*. Effect of pulmonary exacerbations treated with oral antibiotics on clinical outcomes in cystic fibrosis. *Thorax*. 2016.

254. Oakley LL, Renfrew MJ, Kurinczuk JJ, Quigley MA. Factors associated with breastfeeding in England: an analysis by primary care trust. *BMJ Open*. 2013;3(6).
255. Davies JC, Cunningham S, Alton EFW, Innes JA. Lung clearance index in CF: a sensitive marker of lung disease severity. *Thorax*. 2008;63(2):96-97.
256. Stanojevic S, Davis SD, Retsch-Bogart G, *et al*. Progression of Lung Disease in Preschool Patients with Cystic Fibrosis. *American journal of respiratory and critical care medicine*. May 01 2017;195(9):1216-1225.
257. Rogers GB, van der Gast CJ, Serisier DJ. Predominant pathogen competition and core microbiota divergence in chronic airway infection. *The ISME journal*. Jan 2015;9(1):217-225.
258. Quon BS, Dai DL, Hollander Z, *et al*. Discovery of novel plasma protein biomarkers to predict imminent cystic fibrosis pulmonary exacerbations using multiple reaction monitoring mass spectrometry. *Thorax*. Mar 2016;71(3):216-222.
259. Nick JA, Sanders LA, Ickes B, *et al*. Blood mRNA biomarkers for detection of treatment response in acute pulmonary exacerbations of cystic fibrosis. *Thorax*. Oct 2013;68(10):929-937.
260. Prevaes SM, de Steenhuijsen P, Pijpers WA, de Winter-de Groot KM, *et al*. Concordance between upper and lower airway microbiota in infants with cystic fibrosis. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*. Mar 2017;49(3).
261. Mussaffi H, Fireman EM, Mei-Zahav M, Prais D, Blau H. Induced sputum in the very young: a new key to infection and inflammation. *Chest*. Jan 2008;133(1):176-182.
262. Blau H, Linnane B, Carzino R, *et al*. Induced sputum compared to bronchoalveolar lavage in young, non-expectorating cystic fibrosis children. *Journal of Cystic Fibrosis*. 2014;13(1):106-110.
263. Jochmann A, Artusio L, Robson K, *et al*. Infection and inflammation in induced sputum from preschool children with chronic airways diseases. *Pediatric pulmonology*. Aug 2016;51(8):778-786.

- 264.** Chattoraj SS, Ganesan S, Jones AM, *et al.* Rhinovirus infection liberates planktonic bacteria from biofilm and increases chemokine responses in cystic fibrosis airway epithelial cells. *Thorax*. 2011;66(4):333-339.
- 265.** Aracil B, Gomez-Garces JL, Alos JL. A study of susceptibility of 100 clinical isolates belonging to the *Streptococcus milleri* group to 16 cephalosporins. *The Journal of antimicrobial chemotherapy*. Mar 1999;43(3):399-402.
- 266.** Bancescu G, Skaug N, Dumitriu S, Bancescu A. Antimicrobial susceptibility of some streptococci strains of anginosus group isolated from oral and maxillofacial infections. *Roumanian archives of microbiology and immunology*. Jan-Mar 1999;58(1):57-63.
- 267.** Brown PS, Pope CE, Marsh RL, *et al.* Directly sampling the lung of a young child with cystic fibrosis reveals diverse microbiota. *Annals of the American Thoracic Society*. Sep 2014;11(7):1049-1055.
- 268.** Stoltz DA, Meyerholz DK, Pezzulo AA, *et al.* Cystic fibrosis pigs develop lung disease and exhibit defective bacterial eradication at birth. *Science translational medicine*. Apr 28 2010;2(29):29ra31.
- 269.** Murphy EA, Velazquez KT, Herbert KM. Influence of High-Fat-Diet on Gut Microbiota: A Driving Force for Chronic Disease Risk. *Current opinion in clinical nutrition and metabolic care*. 2015;18(5):515-520.
- 270.** Kuehni CE, Strippoli MP, Low N, Silverman M. Asthma in young south Asian women living in the United Kingdom: the importance of early life. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*. Jan 2007;37(1):47-53.

Appendices

A1: Image taken of the CLIMB study database for collation of patient data

BAX study no	BAX001	Encounter type	Bronchoscopy
Hospital no:	08/19706	Genotype 1:	pPhe508Del
Date of Birth:	16/10/2000	Genotype 2:	Arg11His
gender	FEMALE	CT scan results:	N/A
Ethnicity:	British Caucasian	EM results:	N/A
Study group	3: Elective Abx		
Date of encounter:	26/11/2012		

Research studies	Birth history	Feeds	Family history	Lung function	Complications	Bacteria	Fungal/ Viral	Clinical status/ antibiotics	Antifungals/ Other meds	Samples	F/u
------------------	---------------	-------	----------------	---------------	---------------	----------	---------------	------------------------------	-------------------------	---------	-----

Clinical status Score LRT symptoms LRT symptoms ☒ URT symptoms ☐	Intravenous antibiotics IV Abx ☒ IV Abx duration (days) 8 No IV Abx last yr: 3 Name of IV antibiotic 1: Ceftazidime Name of IV antibiotic 2: Tobramycin	Oral antibiotics Preceding po Abx ☐ Current po Abx ☐ No Abx last 30 days: 22 No Abx last 3 months: 22 Prophylactic Abx ☒ Preceding po Abx name: N/A Current po Abx name: N/A Prophylactic Abx name: Flucloxacillin	Inhaled antibiotics Inh Abx: Colomycin
--	---	---	--

A2: Mock community composition

DSM Number	Species
7288	<i>Burkholderia cepacia</i>
19748	<i>Chlamydophila pneumoniae</i>
20478	<i>Enterococcus faecalis</i>
30083	<i>Escherichia coli</i>
15643	<i>Fusobacterium nucleatum subsp. nucleatum</i>
4690	<i>Haemophilus influenzae</i>
1135	<i>Leptotrichia buccalis</i>
9143	<i>Moraxella catarrhalis</i>
2291	<i>Mycoplasma pneumoniae</i>
10036	<i>Neisseria meningitidis</i>
43665	<i>Nocardia farcinica</i>
16031	<i>Pasturella multocida subsp. multocida</i>
50090	<i>Pseudomonas fluorescens</i>
17058	<i>Salmonella enterica subsp. enterica</i>
20231	<i>Staphylococcus aureus</i>
2134	<i>Streptococcus agalactiae</i>
20575	<i>Streptococcus constellatus subsp. constellatus</i>
12492	<i>Streptococcus infantis</i>
6778	<i>Streptococcus parasanguinis</i>
20566	<i>Streptococcus pneumoniae</i>
18670	<i>Streptococcus pseudopneumoniae</i>
20565	<i>Streptococcus pyogenes</i>
20567	<i>Streptococcus sanguinis</i>
14222	<i>Treponema denticola</i>
44697	<i>Mycobacterium psychrotolerans</i>
43990	<i>Mycobacterium. bovis</i>
19120	<i>Actinomyces odontolyticus</i>
20436	<i>Bifidobacterium dentium</i>
44287	<i>Corynebacterium adiacens</i>
9848	<i>Granulicatella adiacens</i>
8978	<i>Haemophilus parainfluenzae</i>
17633	<i>Neisseria flavescens</i>
19025	<i>Prevotella buccae</i>
20746	<i>Rothia mucilaginosa</i>
12643	<i>Streptococcus mitis</i>
20735	<i>Veillonella dispar</i>

A3: E Appendix for use in conjunction with Chapter 3: Individual patient barplots comparing the upper and lower airway microbiota.

A4: E Appendix for use in conjunction with Chapter 4: Individual patient barplots for infants with cystic fibrosis (CF) diagnosed on newborn screening (NBS).

A5: E Appendix for use in conjunction with Chapter 5: Individual patient barplots comparing the airway microbiota on throat swabs (TS) and sputum samples in children with cystic fibrosis (CF) or primary ciliary dyskinesia (PCD).

A6: E Appendix for use in conjunction with Chapter 5: Individual patient barplots for older children with cystic fibrosis (CF).

A7: E Appendix for use in conjunction with Chapter 5: Individual patient barplots for children with Primary Ciliary Dyskinesia (PCD).

A8: Abstracts arising from thesis

Comparison of the upper and lower airway microbiota in children

Ahmed B, Cox MJ, Cookson WOC, Davies JC, Moffatt MF, Bush A.

- Oral presentation, UK CF Microbiology Consortium, Liverpool, November 2014
- Poster presentation, British Thoracic Society Winter Meeting, London, December 2014

Early development of the airway microbiota in infants with CF.

Ahmed B, Cox MJ, Cuthbertson L, James P, Cookson WOC, Davies JC, Moffatt MF, Bush A.

- Poster presentation, North American CF Conference, Orlando, October 2016
- Oral presentation, UK CF Microbiology Consortium, Liverpool, November 2016

A9: Prizes awarded during PhD

- Highly Commended Poster Prize for poster on “Comparison of the Upper and Lower Airway Microbiota in Children” at the 8th Annual NHLI Postgraduate Research Student Presentation Day, Imperial College London, July 2014
- Winner of MPHrp Team Award at the Fourth National Institute for Health Research (NIHR) Experimental Medicine Research Training Camp, July 2013

How to use: bacterial cultures in diagnosing lower respiratory tract infections in cystic fibrosis

Bushra Ahmed,¹ Andrew Bush,¹ Jane C Davies^{1,2}

¹Department of Respiratory Paediatrics, Royal Brompton Hospital, London, UK
²Department of Gene Therapy, Imperial College London, London, UK

Correspondence to
Dr Bushra Ahmed, Department of Respiratory Paediatrics, Royal Brompton Hospital, Sydney Street, London SW3 6NP, UK; b.ahmed@rbht.nhs.uk

Accepted 15 November 2013

ABSTRACT

Respiratory infections are the leading cause of morbidity and mortality in cystic fibrosis. Certain bacteria, such as *Pseudomonas aeruginosa*, are associated with a worse clinical outcome than others, but can be completely eradicated if identified and treated early. The diagnosis of lower respiratory tract infections can be challenging in the non-expectorating patient, in whom upper airway samples, such as cough swabs, are a surrogate for lower airway sampling. However, the results of these often do not fit with the clinical picture, presenting a management dilemma. Frequently, clinicians are faced with a negative culture result in a progressively symptomatic patient and vice versa. When judging the clinical significance of a positive upper airway culture result in an asymptomatic patient, it is important to consider the prognostic significance of the organism cultured. Given that the reported sensitivity of upper airway swabs (which includes throat swabs) is variable, ranging from 35.7% to 71% for *Pseudomonas aeruginosa*, 50% to 86% for *Staphylococcus aureus* and 11% to 92% for *Haemophilus influenza*, upper airway samples may fail to identify lower airway infections. Therefore, in symptomatic children, a repeatedly negative upper airway swab should not be considered as reassuring, and alternative sampling methods, such as induced sputum or bronchoalveolar lavage, should be considered. Here we use some examples of common scenarios to illustrate how best to use bacterial cultures to aid management decisions in cystic fibrosis.

INTRODUCTION

Recurrent bacterial infections are the leading cause of morbidity and mortality in cystic fibrosis (CF), accounting for over 80% of deaths.¹ Early detection of infections is vital. Newborn screening allows most children to be diagnosed soon after

birth providing a previously inaccessible window of opportunity for microbiological surveillance. The nutritional benefits of early diagnosis are clear and studies have also reported improvements in lung function, *Pseudomonas aeruginosa* acquisition rates and survival.² Conversely chronic infection with *P aeruginosa* is associated with increased morbidity and mortality.³

It is important to detect lower airway infection accurately and promptly. CF Trust Guidelines⁴ recommends regular culture of expectorated sputum. In adults this is usually straightforward. Sputum is convenient, has a yield comparable with the 'gold standard' bronchoalveolar lavage (BAL),⁵ and has a high negative predictive value (NPV) and positive predictive value (PPV), up to 100%.⁶

More challenging is the non-expectorating patient, in particular when asymptomatic. Unfortunately, many patients, particularly young children, and also a proportion of adults, are unable to expectorate and cough swabs are performed. However, these have poor sensitivity and specificity and can lead to undertreatment or inappropriate antibiotic therapy.

This review focuses on the strengths and limitations of non-invasive airway sampling in CF, and when alternatives, such as sputum induction or BAL should be considered.

PHYSIOLOGICAL BACKGROUND: AIRWAY INFLAMMATION AND INFECTION

CF is due to a defect in the gene encoding the CF transmembrane conductance regulator protein, leading to either absence or lack of function of CF transmembrane conductance regulator, defective regulation of ion transport and airway surface dehydration. The current favoured hypothesis is

To cite: Ahmed B, Bush A, Davies JC. *Arch Dis Child Educ Pract Ed* Published Online First: [please include Day Month Year]
doi:10.1136/archdischild-2012-303408

Interpretations

that reduced airway surface liquid leads to reduced ciliary function, increased mucus viscosity and decreased mucociliary clearance.⁷ There is mucus hypersecretion, and once infection and inflammation are present, mucus viscosity is further increased by neutrophil-derived DNA and actin. Cough clearance is also impaired, resulting in the retention of secretions in the lower airways, where they provide a nidus for bacterial infection.

Early in life, bacteria such as *Staphylococcus aureus* predominate. With increasing age, *P aeruginosa* emerges as the most common and significant pathogen (figure 1).⁸ Persistence in the airway and a number of environmental triggers lead to mutations in *P aeruginosa*, overproduction of alginate and a mucoid phenotype. In addition, quorum sensing initiates biofilm formation, which enhances resistance to antibiotics.⁹ Although traditionally only a few organisms have been associated with CF airway infection, culture-independent techniques demonstrate greater diversity of microbes although the clinical significance of this is currently unclear.^{10 11}

TECHNOLOGICAL BACKGROUND

Lower airway cultures can be obtained by (A) spontaneous expectoration of sputum, (B) cough swab or plate, (C) induced sputum or (D) direct access to the lower airway (usually with bronchoscopy). Some clinics also measure serum antipseudomonal antibodies. Spontaneously expectorated sputum samples are relatively straightforward and will not be discussed further.

Cough swabs are performed by placing a swab into the posterior pharynx and asking the child to cough (or stimulating cough in a very young child); the aim is to avoid direct contact with the pharyngeal mucosa to prevent upper airway contamination. With older, compliant children, this is easy, although in a well child with a dry cough, yield may be low. Difficulties arise in younger children. The ideal timing of the test is debatable, with some studies showing increased sensitivity if oropharyngeal cultures are collected after physiotherapy¹² or nebulised hypertonic saline.¹³ Throat swabs directly sampling from the upper airway are preferred in USA. Swabs should not be taken immediately after a child has been fed, due to the risk of gagging and potential contamination. Cough plates, taken by asking a child to cough directly onto a plate of culture medium, are sometimes used as an alternative, although repeated samples with specific agar plates are required. Their sensitivity is however debatable, with a recent study of 95 non-expectorating children demonstrating a pathogen isolation rate of only 8% with cough plates in comparison with 18.2% with cough swabs.¹⁴

Sputum induction is performed by ultrasonic nebulisation of hypertonic saline; even in healthy individuals, some 'sputum' can often be expectorated and the procedure is, in general, well tolerated although it may

cause bronchoconstriction. However, it is time consuming, with a median duration of 49 min, and expensive, with an additional cost of US\$150 in comparison with conventional techniques, thus limiting its use within the outpatient clinic.¹⁵ It is not performed routinely in most clinics, although some report high success rates even in small babies.¹⁶ Sputum induction, if successful, could obviate the need for BAL. Studies are underway to address this issue systematically.

BAL, often said to be the gold standard, has disadvantages; most importantly, it requires sedation or a general anaesthetic and cannot be performed frequently. There are regional differences in pathogens, so unless the entire bronchial tree is sampled, which is impractical, organisms may be missed.^{17 18}

The temperature at which samples are stored can impact culture results. The optimum storage temperature would prevent overgrowth of organisms, but not lead to bacterial death. Ideally, samples would either be processed immediately or frozen. However, in clinical practice, particularly for samples collected in the community, storage at room temperature or refrigeration at 4°C is more practical. There is a lack of data comparing the effect of storage temperature on culture results in CF. Studies in bronchiectasis confirmed a negative impact of refrigeration of sputum compared with storage at room temperature for up to 24 h, with a 10-fold reduction in the number of pathogens present in 23% and 8% of samples, respectively.¹⁹ Conversely, May *et al*²⁰ showed a 50% decrease in isolation of *Haemophilus influenza* in samples posted to the laboratory and arriving 24–36 h later. The CF Trust⁴ recommends as a compromise that samples be processed as soon as possible, and to use refrigeration where there is likely to be a delay of more than a few hours, and to interpret posted samples with caution.

The laboratory handling of samples can pose difficulties, and only experienced laboratories should be used. Mucolytics, such as dithiothreitol, can help in the culture of bacteria; mechanical homogenisation with sterile glass beads is preferred by other laboratories. The culture of 'polymicrobial samples' is difficult, because of bacterial overgrowth, particularly by mucoid *P aeruginosa*, and different pathogens will have different growth requirements, so it is important to use selective media^{4 21} especially for *Burkholderia cepacia* complex organisms. Culture duration is variable according to pathogen, with up to 5 days required for *B cepacia* complex.⁴

The use of *P aeruginosa* serology is controversial; the rationale is based on the premise that children will seroconvert during infection and revert once infection is eradicated. If true, serology would provide an attractive option but the data is variable, in part relating to differences in the nature of the sample (BAL vs sputum or oropharyngeal cultures), antigen assayed and cut-off values used (table 1).

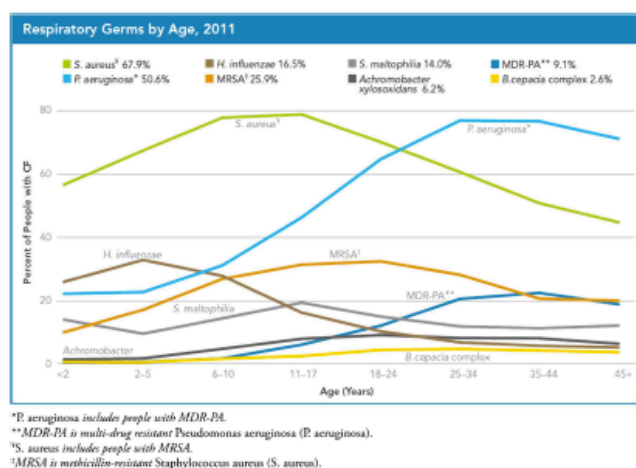


Figure 1 Changes in prevalence of specific bacteria with age.⁸ Reproduced with the permission of the Cystic Fibrosis Foundation.

INDICATIONS AND LIMITATIONS

In the chronic management of children with CF, how often should bacterial cultures be taken? Do regular surveillance cultures improve the rate of detection of bacterial infections?

Current, non-evidence based guidelines recommend culture of sputum or cough swab at every clinic visit (every 2–3 months) and, in addition, during exacerbations.^{21–22} There are no studies comparing outcomes of surveillance microbiology with culture only when patients are symptomatic. The rationale behind frequent surveillance is that significant bacterial infection is common in the absence of symptoms,²³ with more than 50% of patients with *P. aeruginosa* positive cultures reported to be asymptomatic,²⁴ and that early attempts at eradication are likely to be more successful. Without regular surveillance, the detection of prognostically important pathogens may otherwise be missed.

In an asymptomatic child with CF, does a positive surveillance cough swab mandate treatment with antibiotics?

This is a common scenario, and the interpretation of the positive cough swab depends on the clinical context. The literature is (A) largely based on throat rather than cough swabs and (B) divided on issues of sensitivity and specificity. Cough swabs have usually been compared with sputum, whereas throat swabs have been compared with BAL. Importantly, the PPV of upper airway cultures can vary greatly for individual organisms, with reported ranges of 44% to 83.3% for *P. aeruginosa*, 33% to 64% for *S. aureus* and 50% to 81% for *H. influenzae* (see table 2), although this is predominantly for throat swabs. Hence, the accuracy of a positive upper airway culture to predict lower airway infections is questionable. There have been no randomised controlled trials of treating versus 'ignoring' positive cultures in this context.

Table 1 Value of serology in the early identification of *Pseudomonas aeruginosa* infections

Authors	Subjects' age and number	Comparison sample	Antigens assayed	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Douglas <i>et al</i> ²⁵	0.1–5.6 years 76 patients	BAL	MAG* Exotoxin A	53 93	82 43	26 19	94 98
Hayes <i>et al</i> ²⁶	0–6 years† 69 patients	Sputum Oropharyngeal cultures	Cell lysate Exotoxin A Elastase	52 49 36	90 90 81	84 84 65	65 64 55
Ratjen <i>et al</i> ²⁷	1–52 years 375 patients		Alkaline protease Exotoxin A Elastase	85.4 72.0 76.2	97.5 97.5 97.5	96.6 95.9 96.2	88.8 80.8 83.2

*MAG—Multiple *Pseudomonas aeruginosa* antigens.

†Patients followed up from diagnosis on newborn screening over 6 years.

BAL, bronchoalveolar lavage; NPV, negative predictive value; PPV, positive predictive value.

Table 2 The value of upper airway cultures in cystic fibrosis

Authors	Swab type	Comparison sample	Age range (mean)	Pathogen	Sensitivity	Specificity	PPV	NPV
Ramsey <i>et al</i> ²⁵	Throat	BAL	4 months–25 years (8.2 years)	<i>Pseudomonas aeruginosa</i> <i>Staphylococcus aureus</i> <i>Haemophilus influenza</i>	56% 54% 11%	93 100% 87%	NS	NS
Armstrong <i>et al</i> ²⁸	Throat	BAL	1–52 months (17 months)	<i>Pseudomonas aeruginosa</i> <i>Staphylococcus aureus</i> <i>Haemophilus influenza</i>	71% 86% 92%	93% 61% 92%	57% 33% 50%	96% 95% 99%
Rosenfeld <i>et al</i> ⁴⁸	Throat	BAL	1<18 months 2>18 months	<i>Pseudomonas aeruginosa</i> <i>Staphylococcus aureus</i> <i>Haemophilus influenza</i>	1. 44% 2. 68% 1. 80% 2. 73% 1. 81% 2. 55%	1. 95% 2. 94% 1. 70% 2. 71% 1. 95% 2. 92%	1. 44% 2. 76% 1. 64% 2. 59% 1. 81% 2. 50%	1. 95% 2. 91% 1. 88% 2. 82% 1. 95% 2. 93%
Burns <i>et al</i> ³⁴	Throat	BAL	2.5–15.5 months (11.4 months)	<i>Pseudomonas aeruginosa</i>	NS	NS	69%	95%
Equi <i>et al</i> ²⁹	Cough	Sputum	Adults	NS	34%	100%	100%	21%
Jung <i>et al</i> ⁶	Throat	BAL	5.2–34.2 years (14.2 years)	<i>Pseudomonas aeruginosa</i>	35.7	96.2	83.3	73.5
Kabra <i>et al</i> ¹²	1. Cough 2. Throat	Sputum	6.37–8.5 years (7.2 years)	<i>Pseudomonas aeruginosa</i> <i>Staphylococcus aureus</i>	1. 42% 2. 40% 1. 50% 2. 57%	1. 100% 2. 99% 1. 99% 2. 99%	NS	NS
Maiya <i>et al</i> ⁴⁹	Cough a. Moist b. Dry	Sputum	8–16 years	NS	a. 35% b. 35%	NS	NS	NS
Chavasse <i>et al</i> ⁵⁰	Cough	Sputum	NS	NS	41%	NS	NS	NS

BAL, bronchoalveolar lavage; NPV, negative predictive value; NS, not stated; PPV, positive predictive value.

The following details of the clinical context need to be considered: *What is the pathogen that has been identified and what is its prognostic significance? Has this child grown this pathogen before? Is this a new isolate or does it reflect chronic infection? Is the laboratory able to give some idea of the number of organisms isolated? Some common clinical scenarios are discussed in more detail below.*

A surveillance cough swab positive for *P aeruginosa* $\times 10^{-3}$ cfu/mL in an infant with previously negative cultures

Ramsey *et al*²⁵ have reported that quantitative oropharyngeal cultures are neither more sensitive nor specific than qualitative cultures for *P aeruginosa*, *S aureus* and *H influenza*. Given the adverse prognosis associated with *P aeruginosa*, treatment would be initiated by most specialists after even a single upper airway culture in a previously uninfected child, irrespective of quantity of organisms present. Good practice would be to attempt a second culture prior to initiating eradication treatment, although this may not always be practical. No trials have yet defined the optimal eradication regimen; success rates of around 80% have been achieved with combinations of oral ciprofloxacin/colomycin and with nebulised tobramycin. The UK is currently conducting a randomised controlled trial to compare oral or intravenous antipseudomonals (both arms also receive nebulised colomycin).²⁶

Should attempts at eradication fail, guidelines recommend that regular inhaled antipseudomonal antibiotics (eg, colomycin) are continued for chronic *P aeruginosa* growths, with a 2 week course of oral ciprofloxacin reserved for early use with increased symptoms.²¹

Recurrent growth of *Staphylococcus aureus* on surveillance cough swabs despite flucloxacillin prophylaxis

Despite optimal management, some patients with CF become chronically infected with bacterial pathogens. The range of PPVs for methicillin-sensitive *S aureus* (MSSA) on oropharyngeal cultures is the lowest compared with *P aeruginosa* and *H influenza*, therefore this result is even less reliable than for other organisms (table 2). In this scenario, with a well child who is already receiving a prophylactic agent, attention should be focussed on determining adherence to prescribed medications, a particular problem in the teenage years. Although guidelines recommend treatment with oral antibiotics for initial growths of MSSA,²¹ for persistent infection, aggressive use of antibiotics may achieve little and carries with it the potential for side effects. The long term consequences of MSSA in CF are generally regarded as less severe than those of organisms, such as *P aeruginosa*,²⁷ which would require continued treatment, as described above. Further courses of antistaphylococcal agents can be considered should symptoms develop.

To our knowledge, there have been no reports on the predictive values of cough swabs for the detection of methicillin-resistant *S aureus* (MRSA) specifically.

The effect of MRSA on morbidity in CF is debateable. Two large observational studies have reported conflicting results. Dasenbrook *et al*²⁸ observed an increased rate of decline in forced expiratory volume in 1 s of 0.5% predicted per year in 1732 patients with new persistent MRSA versus patients without MRSA. In contrast Sawicki *et al*²⁹ reported a decline in lung function beginning prior to acquisition of MRSA in 593 new MRSA cases, suggesting that morbidity associated with MRSA lung infections may be a marker of disease severity rather than MRSA status itself. Similarly, Hubert *et al*²⁷ suggested that MRSA may be more harmful only when there is co-infection with *P aeruginosa*, thus several confounders could contribute to the effect of MRSA on lung function decline. Given that a 1.27 times increased risk of death has been reported with MRSA infection, and no increased risk of death for patients in whom eradication is achieved within 1 year,³⁰ eradication of MRSA seems beneficial and should be attempted. In the absence of randomised controlled trials to inform treatment strategies,³¹ guidelines recommend treatment of first isolates or return of MRSA following previous successful treatment with topical treatment and either combined oral rifampicin and fusidic acid or nebulised vancomycin or a combination of the three. Children with chronic infections may benefit from prolonged oral antibiotics until free from MRSA.²¹

A surveillance cough swab positive for *B cepacia* complex

Accurate identification of *B cepacia* complex is essential. Certain subtypes (eg, *Burkholderia cenocepacia*, *Burkholderia dolosa*) carry a poorer prognosis than others (eg, *Burkholderia multivorans*).³² There are also infection control implications; these organisms can be highly transmissible and patients harbouring them need to be strictly isolated.³³ While awaiting full identification, isolation precautions are mandatory. The PPV of upper airway cultures increases with repeated testing.³⁴ Further samples should be sent to the reference laboratory. Some will turn out to be another gram negative organism, often not present on repeated culture and of uncertain significance. If *B cepacia* complex is confirmed, guidelines recommend that eradication should be attempted,³³ although the reported success of eradication regimens are moderate at best, with one group reporting successful treatment in only 4/14 (29%) of patients with *B cepacia* complex.³⁵ A recent Cochrane review did not identify any randomised controlled trials to inform the choice of eradication strategies³⁶ thus the choice of agents should be determined by sensitivities.³³

A surveillance cough swab positive for *Achromobacter xylosoxidans* on two consecutive cough swabs taken 3 months apart

Pathogens not previously associated with CF are being isolated with increased frequency. It is difficult to know when to treat organisms such as *Achromobacter xylosoxidans* and *Stenotrophomonas maltophilia*.

Interpretations

Box 1 Questions to consider with a symptomatic child with negative cough swabs

- ▶ Has the cough swab been performed by an adequately trained and experienced person?
- ▶ Was it processed in a laboratory experienced with cystic fibrosis samples?
- ▶ Does this child have a history of previous lower airway infection which could be used to help guide treatment on the balance of probability?
- ▶ Is the child so unwell that empirical treatment should be started in any case?
- ▶ Does the clinical status of the child merit an invasive investigation, such as bronchoscopy?
- ▶ Is there evidence of an intercurrent viral infection (for example, a positive NPA) that might predispose the child to secondary bacterial infections?

Current evidence available suggests that many of these rarer organisms may not be associated with clinical deterioration,³⁷ thus guidelines²¹ do not recommend treatment unless patients are symptomatic. In this case, treatment choice should be guided by sensitivities. There is also uncertainty over non-aeruginosa *Pseudomonas* species, which are treated aggressively in some centres and not at all in others. An exception to this is *Pandora apista*, which is sometimes misidentified as *B. cepacia* complex,³⁷ and has been associated with clinical deterioration.²¹ Guidelines recommend treatment where this is isolated.²¹

In a symptomatic child with CF, does a negative cough swab exclude a bacterial lower airway infection?

If a symptomatic child with CF (the patient) has a negative cough swab (the test), how confident can we be that this reflects the absence of lower airway infection (the outcome)? As discussed above, the literature is unclear. Although some studies have reported high NPVs of up to 99%,³⁸ these are mostly for throat swabs. A study from our centre³⁹ confirmed that a negative cough swab should not be reassuring and that further attempts should be made to obtain lower airway secretions, for example, by BAL or sputum induction. This is particularly true for certain organisms, such as non-tuberculous mycobacteria (NTM); in six patients known to have positive sputum samples for NTM, all cough swabs were negative,⁴⁰ thus cough swabs do not detect NTM. Whether treatment is started depends on the clinical context. Questions that may be asked to guide decision making are outlined in box 1.

In a child with CF, who is undergoing *P. aeruginosa* eradication on the basis of positive BAL cultures, does a negative cough swab taken at the end of treatment indicate that eradication has been successful?

This is very difficult; there is no convincing evidence. Studies in which infections have been diagnosed on

BAL cultures have focussed on repeat BAL samples for confirmation of eradication.⁴¹ With early and aggressive treatment, successful eradication of *P. aeruginosa* can be achieved, with one study reporting clearance of even mucoid *P. aeruginosa* in 67/116 children for more than 1 year.⁴² However, given cough swabs have a low NPV the cough swab result should not be relied on in isolation to confirm this. Particular caution should be taken with patients receiving inhaled antipseudomonal antibiotics for *P. aeruginosa* eradication, as these may lead to false negative results. Supporting information should be sought including:

Were there any clinical features that could guide us, for example, symptoms improving or lung function returning to baseline? If so, and cough swabs were negative at the end of eradication, therapy is stopped with careful ongoing monitoring. *Were there features to suggest chronic infection, such as the presence of a mucoid *P. aeruginosa*?* A paediatric CF specialist would be cautious and continue chronic suppressive treatment beyond 3 months. *Should BAL be repeated to confirm eradication?* Perhaps, but given the invasive nature of the intervention and the possibility of sampling errors, most clinicians (and parents) would be reluctant to do this; if the child was not doing well clinically, BAL or sputum induction should be considered.

Test your knowledge

1. A 6-year-old girl with cystic fibrosis presents with features suggestive of respiratory infection. Which of the following statements is true?
 - A. Expecterated sputum is likely to be easily obtained.
 - B. A negative cough swab rules out infection and treatment is not required.
 - C. If she is unable to expectorate, induced sputum samples may be a suitable alternative.
2. A cough swab shows a growth of *Pseudomonas aeruginosa* $\times 10^{-3}$ for the first time. Which of the following statements is true?
 - A. If the child is asymptomatic, this is a likely contaminant and does not warrant treatment.
 - B. Only significant growths of *Pseudomonas aeruginosa* (of $> \times 10^5$) should be treated, irrespective of symptoms.
 - C. Any growth of *Pseudomonas aeruginosa* should be treated.
3. Which of the following is untrue of bronchoalveolar lavage?
 - A. A limitation is the ability to sample only small areas of the lower airway.
 - B. It is a useful sample for assessing inflammation.
 - C. It should not be performed repeatedly as a routine.

In children who cannot expectorate sputum, is BAL superior to oropharyngeal cultures for monitoring changes in the respiratory flora?

A large randomised controlled trial attempted to answer this question. Surveillance was either by standard methods, such as oropharyngeal cultures, or with BAL; the latter was performed before 6 months of age, during exacerbations and following eradication for suspected *P aeruginosa* infection. The study found no difference in rates of *P aeruginosa* or any other end point between the two groups.⁴¹ Considering the risks associated with BAL, current evidence does not support the use of routine, regular BAL surveillance.

FUTURE RESEARCH

Molecular techniques

Non-culture based, molecular techniques are promising. PCR has been used for the early detection of *P aeruginosa*, although there have conflicting reports regarding its sensitivity and, where available, it should therefore be used as an adjunct to conventional culture.⁴³ Sequencing of the 16S rRNA gene from the bacterial genome is perhaps the most exciting, as it has the potential to allow surveillance of entire bacterial communities, often comprising vast numbers of disparate bacteria. It remains to be determined which of these organisms are most important and whether in fact some of them may be clinically beneficial.

Non-invasive, breath-based detection systems

Bacteria produce an array of exoproducts, including volatile organic compounds. Such signals, if specific to an individual organism, could provide a valuable screening tool. Various markers have been proposed, such as 2-pentylfuran as a marker of infection with *Aspergillus fumigatus* and hydrogen cyanide as a marker of *P aeruginosa*.⁴⁴ The technology and clinical utility of these tests are currently under investigation.

Clinical bottom line

- ▶ Cough swabs are useful in surveying non-expectorating, asymptomatic patients, where more invasive testing would not be justified; their sensitivity and specificity is however suboptimal. A negative cough swab should not be taken as reassuring in an unwell child; induced sputum or bronchoalveolar lavage may be needed.
- ▶ Eradication of *Pseudomonas aeruginosa* is successful in the majority of cases, provided the infection is detected promptly, so early identification is crucial.
- ▶ Response to antibiotics may correlate poorly with in vitro sensitivities of cultured organisms.

Answers to the quiz

- Question 1: C.
Question 2: C.
Question 3: B.

Contributors BA, AB and JCD contributed equally to writing and editing the manuscript.

Competing interests None.

Provenance and peer review Commissioned; externally peer reviewed.

REFERENCES

- 1 O'Sullivan BJ, Freedman SD. Cystic fibrosis. *Lancet* 2009;373:1891–904.
- 2 Dijk FN, McKay K, Barzi F, *et al*. Improved survival in cystic fibrosis patients diagnosed by newborn screening compared to a historical cohort from the same centre. *Arch Dis Child* 2011;96:1118–23.
- 3 Stuart B, Lin JH, Mogayzel PJ Jr. Early eradication of *Pseudomonas aeruginosa* in patients with cystic fibrosis. *Paediatr Respir Rev* 2010;11:177–84.
- 4 Cystic Fibrosis Trust Microbiology Standards Working Group. *Standards of care: laboratory standards for processing microbiological samples from people with cystic fibrosis*. London, UK: Cystic Fibrosis Trust, 2010.
- 5 Aaron SD, Kottachchi D, Ferris WJ, *et al*. Sputum versus bronchoscopy for diagnosis of *Pseudomonas aeruginosa* biofilms in cystic fibrosis. *Eur Respir J* 2004;24:631–7.
- 6 Jung A, Kleinau I, Schonian G, *et al*. Sequential genotyping of *Pseudomonas aeruginosa* from upper and lower airways of cystic fibrosis patients. *Eur Respir J* 2002;20:1457–63.
- 7 Matsui H, Grubb BR, Tarran R, *et al*. Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell* 1998;95:1005–15.
- 8 Cystic Fibrosis Foundation Patient Registry 2011. Annual Data Report. Bethesda, Maryland, USA.
- 9 Williams HD, Davies JC. Basic science for the chest physician: *Pseudomonas aeruginosa* and the cystic fibrosis airway. *Thorax* 2012;67:465–7.
- 10 Delhaes L, Monchy S, Frealle E, *et al*. The airway microbiota in cystic fibrosis: a complex fungal and bacterial community—implications for therapeutic management. *PLoS One* 2012;7:e36313.
- 11 Zhao J, Schloss PD, Kalikin LM, *et al*. Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc Natl Acad Sci USA* 2012;109:5809–14.
- 12 Kabra SK, Alok A, Kapil A, *et al*. Can throat swab after physiotherapy replace sputum for identification of microbial pathogens in children with cystic fibrosis? *Indian J Pediatr* 2004;71:21–3.
- 13 Ho SA, Ball R, Morrison LJ, *et al*. Clinical value of obtaining sputum and cough swab samples following inhaled hypertonic saline in children with cystic fibrosis. *Pediatr Pulmonol* 2004;38:82–7.
- 14 Byrne NM, O'Brien C, Spencer DA. The efficacy of cough plates in the identification of bacterial pathogens in children with cystic fibrosis. *Arch Dis Child* 2013;98:768–71.

Interpretations

- 15 Al-Saleh S, Dell SD, Grasemann H, *et al*. Sputum induction in routine clinical care of children with cystic fibrosis. *J Pediatr* 2010;157:1006–11.e1.
- 16 Mussaffi H, Fireman EM, Mei-Zahav M, *et al*. Induced sputum in the very young: a new key to infection and inflammation. *Chest* 2008;133:176–82.
- 17 Gilchrist JP, Salamat S, Clayton S, *et al*. Bronchoalveolar lavage in children with cystic fibrosis: how many lobes should be sampled? *Arch Dis Child* 2011;96:215–17.
- 18 Gutierrez JP, Grimwood K, Armstrong DS, *et al*. Interlobar differences in bronchoalveolar lavage fluid from children with cystic fibrosis. *Eur Respir J* 2001;17:281–6.
- 19 Pye A, Hill SL, Bharadwa P, *et al*. Effect of storage and postage on recovery and quantitation of bacteria in sputum samples. *J Clin Pathol* 2008;61:352–4.
- 20 May JR, Delves DM. The survival of haemophilus influenzae and pneumococci in specimens of sputum sent to the laboratory by post. *J Clin Pathol* 1964;17:254–6.
- 21 The Cystic Fibrosis Trust's Antibiotic Working Group. *Standards of care: antibiotic treatment for cystic fibrosis*. London, UK: Cystic Fibrosis Trust, 2009.
- 22 <http://www.rchbht.nhs.uk/healthprofessionals/clinical-departments/paediatrics/childrencf/>
- 23 Hilliard TN, Sukhani S, Francis J, *et al*. Bronchoscopy following diagnosis with cystic fibrosis. *Arch Dis Child* 2007;92:898–9.
- 24 Douglas TA, Brennan S, Gard S, *et al*. Acquisition and eradication of *P. aeruginosa* in young children with cystic fibrosis. *Eur Respir J* 2009;33:305–11.
- 25 Ramsey BW, Wentz KR, Smith AL, *et al*. Predictive value of oropharyngeal cultures in identifying lower airway bacteria in cystic fibrosis patients. *Am Rev Respir Dis* 1991;144:331–7.
- 26 <http://www.torpedo-cf.org.uk/>
- 27 Hubert D, Reglier-Poupet H, Sermet-Gaudelus I, *et al*. Association between *Staphylococcus aureus* alone or combined with *Pseudomonas aeruginosa* and the clinical condition of patients with cystic fibrosis. *J Cyst Fibros* 2013;12:497–503.
- 28 Dasenbrook EC, Merlo CA, Diener-West M, *et al*. Persistent methicillin-resistant *Staphylococcus aureus* and rate of FEV1 decline in cystic fibrosis. *Am J Respir Crit Care Med* 2008;178:814–21.
- 29 Sawicki GS, Rasouliyan L, Pasta DJ, *et al*. The impact of incident methicillin resistant *Staphylococcus aureus* detection on pulmonary function in cystic fibrosis. *Pediatr Pulmonol* 2008;43:1117–23.
- 30 Dasenbrook EC, Checkley W, Merlo CA, *et al*. Association between respiratory tract methicillin-resistant *Staphylococcus aureus* and survival in cystic fibrosis. *JAMA* 2010;303:2386–92.
- 31 Lo DK, Hurley MN, Muhlebach MS, *et al*. Interventions for the eradication of methicillin-resistant *Staphylococcus aureus* (MRSA) in people with cystic fibrosis. *Cochrane Database Syst Rev* 2013;2:CD009650.
- 32 Jones AM, Dodd ME, Govan JR, *et al*. *Burkholderia cenocepacia* and *Burkholderia multivorans*: influence on survival in cystic fibrosis. *Thorax* 2004;59:948–51.
- 33 The *Burkholderia Cepacia* Complex. Suggestions for Prevention and Infection Control. Report of the UK Cystic Fibrosis Trust Infection Control Group. London, UK, September 2004.
- 34 Burns JL, Gibson RL, McNamara S, *et al*. Longitudinal assessment of *Pseudomonas aeruginosa* in young children with cystic fibrosis. *J Infect Dis* 2001;183:444–52.
- 35 Horsley A, Webb K, Bright-Thomas R, *et al*. Can early *Burkholderia cepacia* complex infection in cystic fibrosis be eradicated with antibiotic therapy? *Front Cell Infect Microbiol* 2011;1:18.
- 36 Horsley A, Jones AM. Antibiotic treatment for *Burkholderia cepacia* complex in people with cystic fibrosis experiencing a pulmonary exacerbation. *Cochrane Database Syst Rev* 2012;10:CD009529.
- 37 Davies JC, Bilton D. Bugs, biofilms, and resistance in cystic fibrosis. *Respir Care* 2009;54:628–40.
- 38 Armstrong DS, Grimwood K, Carlin JB, *et al*. Bronchoalveolar lavage or oropharyngeal cultures to identify lower respiratory pathogens in infants with cystic fibrosis. *Pediatr Pulmonol* 1996;21:267–75.
- 39 Equi AC, Pike SE, Davies JC, *et al*. Use of cough swabs in a cystic fibrosis clinic. *Arch Dis Child* 2001;85:438–39.
- 40 Ahmed B, Balfour-Lynn IM, Alshafi K. Cough swabs should not be used to isolate non-tuberculous mycobacteria in children with cystic fibrosis. *Arch Dis Child* 2012;97:854–5.
- 41 Wainwright CE, Vidmar S, Armstrong DS, *et al*. Effect of bronchoalveolar lavage-directed therapy on *Pseudomonas aeruginosa* infection and structural lung injury in children with cystic fibrosis: a randomized trial. *JAMA* 2011;306:163–71.
- 42 McPherson H, Rosenthal M, Bush A. Can mucoid *Pseudomonas aeruginosa* be eradicated in children with cystic fibrosis? *Pediatr Pulmonol* 2010;45:566–8.
- 43 Deschaght P, Schelstraete P, Lopes dos Santos Santiago G, *et al*. Comparison of culture and qPCR for the detection of *Pseudomonas aeruginosa* in not chronically infected cystic fibrosis patients. *BMC Microbiol* 2010;10:245.
- 44 Enderby B, Smith D, Carroll W, *et al*. Hydrogen cyanide as a biomarker for *Pseudomonas aeruginosa* in the breath of children with cystic fibrosis. *Pediatr Pulmonol* 2009;44:142–7.
- 45 Douglas TA, Brennan S, Berry L, *et al*. Value of serology in predicting *Pseudomonas aeruginosa* infection in young children with cystic fibrosis. *Thorax* 2010;65:985–90.
- 46 Hayes D Jr, Farrell PM, Li Z, *et al*. *Pseudomonas aeruginosa* serological analysis in young children with cystic fibrosis diagnosed through newborn screening. *Pediatr Pulmonol* 2010;45:55–61.
- 47 Ratjen F, Walter H, Haug M, *et al*. Diagnostic value of serum antibodies in early *Pseudomonas aeruginosa* infection in cystic fibrosis patients. *Pediatr Pulmonol* 2007;42:249–55.
- 48 Rosenfeld M, Emerson J, Accurso F, *et al*. Diagnostic accuracy of oropharyngeal cultures in infants and young children with cystic fibrosis. *Pediatr Pulmonol* 1999;28:321–8.
- 49 Maiya S, Desai M, Baruah A, *et al*. Cough plate versus cough swab in patients with cystic fibrosis: a pilot study. *Arch Dis Child* 2004;89:577–9.
- 50 Chavasse RJ, Cordle R, Petkar H. Cough plates for microbiological surveillance in cystic fibrosis. *Arch Dis Child* 2007;92:279.

A11: Reprint permissions for Figures 1.1 – 1.7



Thank you for your order!

Dear Dr. Bushra Ahmed,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee: Dr. Bushra Ahmed
Order Date: Jul 17, 2017
Order Number: 4151420003263
Publication: Nature Reviews Microbiology
Title: Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences
Type of Use: reuse in a dissertation / thesis
Order Total: 0.00 GBP

View or print complete [details](#) of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

How was your experience? Fill out this [survey](#) to let us know.

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
<https://myaccount.copyright.com>



RightsLink®



ERS

Kay Sharpe
European Respiratory Society Permissions Editor
D +44 114 267 28 61

ERS 442 Glossop Road Sheffield S10 2PX United Kingdom
Central T +44 114 267 28 60 F +44 114 266 50 64 Wersnet.org



-----Original Message-----

From: Ahmed, Bushra [mailto:b.ahmed12@imperial.ac.uk]

Sent: 17 July 2017 16:57

To: Kay Sharpe

Subject: Permission to reproduce image

Dear ERS,

I would like to request your permission to reproduce an image from the Paediatric Respiratory Medicine ERS Handbook for use in my PhD thesis please? The image is on page 20 of the first edition of the book and it entitled, "Figure 1: Overview of the initiation and interaction of the innate and adaptive immune system."

The title of my thesis is "Longitudinal Investigation of the Airway Microbiota in Children with Chronic Suppurative Lung Disease" and I would like to include the image in both an electronic and printed version of my thesis. I would be happy to provide any further information you require.

If this is ok, please let me know if there is a format you would like me to use when listing where I reproduced the image from in my thesis.

Best wishes,

Bushra Ahmed

--

Bushra Ahmed

From: **permissions** permissions@ersj.org.uk
Subject: RE: Permission to reproduce image
Date: 18 July 2017 at 08:45
To: Ahmed, Bushra b.ahmed12@imperial.ac.uk



Dear Bushra,

Thank you for your enquiry.

Permission is granted for the material to be reproduced for your thesis in accordance with ERS copyright policy and full acknowledgement must be given— see below

Notes, Terms & Conditions (where applicable)

"Green" Open Access and Author Archiving: Authors who do not wish to pay for the ERJ Open option will still have their manuscripts made free to access via the ERJ online archive following the journal's 18-month embargo period; after this embargo period, authors also have licence to deposit their manuscripts in an institutional (or other) repository for public archiving, provided the following requirements are met:

- 1) The final, peer-reviewed, author-submitted version that was accepted for publication is used (before copy-editing and publication).
- 2) A permanent link is provided to the version of the article published in the ERJ, through the dx.doi.org platform. For example, if your manuscript has the DOI 10.1183/09031936.00123412, then the link you provide must be dx.doi.org/10.1183/09031936.00123412
- 3) The repository on which the manuscript is deposited is not used for systematic distribution or commercial sales purposes.
- 4) The following required archiving statement appears on the title page of the archived manuscript: "This is an author-submitted, peer-reviewed version of a manuscript that has been accepted for publication in the European Respiratory Journal, prior to copy-editing, formatting and typesetting. This version of the manuscript may not be duplicated or reproduced without prior permission from the copyright owner, the European Respiratory Society. The publisher is not responsible or liable for any errors or omissions in this version of the manuscript or in any version derived from it by any other parties. The final, copy-edited, published article, which is the version of record, is available without a subscription 18 months after the date of issue publication."

Authors of articles published under one of the Creative Commons licences (this includes all articles in ERJ Open Research, European Respiratory Review, Breathe, ERS Monograph, ERS Handbook series) retain further rights to share, reuse or adapt their manuscript. The extent of these rights depends on the specific Creative Commons licence used. Please consult the relevant section of the online instructions for authors. These publications are copyrighted material and must not be copied, reproduced, transferred, distributed, leased, licensed, placed in a storage retrieval system or publicly performed or used in any way except as specifically permitted in writing by the publishers (European Respiratory Society), as allowed under the terms and conditions of which it was purchased or as strictly permitted by applicable copyright law. Any unauthorised distribution or use of this text may be a direct infringement of the publisher's rights and those responsible may be liable in law accordingly. Copyright remains with European Respiratory Society©



Thank you for your order!

Dear Dr. Bushra Ahmed,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Dr. Bushra Ahmed
Order Date:	Aug 19, 2017
Order Number:	4172581374508
Publication:	Human Molecular Genetics
Title:	Sequencing the human microbiome in health and disease
Type of Use:	Thesis/Dissertation
Order Total:	0.00 GBP

View or print complete [details](#) of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

How was your experience? Fill out this [survey](#) to let us know.

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
<https://myaccount.copyright.com>



RightsLink®

THE LANCET

"In patients with first-time myocardial infarction, all combinations of aspirin, clopidogrel, and vitamin K antagonists are associated with increased risk of non-fatal and fatal bleeding, compared with aspirin alone."

Aspirin	Clopidogrel	Vitamin K	Aspirin + Clopidogrel	Aspirin + Vitamin K	Clopidogrel + Vitamin K	Aspirin + Clopidogrel + Vitamin K
100 mg daily	75 mg daily	10 mg daily	100 mg daily + 75 mg daily	100 mg daily + 10 mg daily	75 mg daily + 10 mg daily	100 mg daily + 75 mg daily + 10 mg daily

Thank you for your order!

Dear Dr. Bushra Ahmed,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee: Dr. Bushra Ahmed
Order Date: Aug 19, 2017
Order Number: 4172710535999
Publication: The Lancet
Title: The role of the microbiome in exacerbations of chronic lung diseases
Type of Use: reuse in a thesis/dissertation
Order Total: 0.00 GBP

View or print complete [details](#) of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

How was your experience? Fill out this [survey](#) to let us know.

Tel: +1-855-239-3415 / +1-978-646-2777
customer@copyright.com
<https://myaccount.copyright.com>



RightsLink®



Thank you for your order!

Dear Dr. Bushra Ahmed,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee: Dr. Bushra Ahmed
Order Date: Jul 17, 2017
Order Number: 4151461077194
Publication: The Lancet Respiratory Medicine
Title: A new era in the treatment of cystic fibrosis: correction of the underlying CFTR defect
Type of Use: reuse in a thesis/dissertation
Order Total: 0.00 GBP

View or print complete [details](#) of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

How was your experience? Fill out this [survey](#) to let us know.

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
<https://myaccount.copyright.com>



RightsLink®

From: Chloe Ainsley Chloe.Ainsley@cysticfibrosis.org.uk
Subject: RE: Image reproduction permission
Date: 19 July 2017 at 12:51
To: b.ahmed12@imperial.ac.uk
Cc: Rebecca Cosgriff Rebecca.Cosgriff@cysticfibrosis.org.uk, Kieran Earlam Kieran.Earlam@cysticfibrosis.org.uk



Hi Bushra,

Thank you for your email. It is fine for you to use the image 1.17 from our Annual Data Report 2013 for your PhD thesis.

When crediting the image can you ensure that you write our name in full 'Cystic Fibrosis Trust' please.

Kind regards,

Chloe

Chloe Ainsley

Design Content Producer

DD: 020 3795 1503

cysticfibrosis.org.uk

Cystic Fibrosis Trust, One Aldgate, London EC3N 1RE

From: Ahmed, Bushra [<mailto:b.ahmed12@imperial.ac.uk>]

Sent: 17 July 2017 19:44

To: Enquiries <Enquiries@cysticfibrosis.org.uk>

Subject: Image reproduction permission

Dear CF Trust,

I would like to request your permission to reproduce an image 1.17 entitled "Lung Infections in 2013" from your Annual data report 2013 please? I would like to use it in a PhD thesis I am writing entitled "Longitudinal Investigation of the Airway Microbiota in Children with Chronic Suppurative Lung Disease", which I plan to submit in September 2017 in both electronic and hard format. I would be happy to provide any further information you need.

If this is ok, please let me know if there is a particular style you would like me to use when crediting the image to the CF Trust.

Best wishes,

Bushra Ahmed

We're fighting for Life Unlimited by cystic fibrosis. [Join us.](#)

Registered as a charity in England and Wales (1079049) and in Scotland (SC040196). A company limited by guarantee, registered in England and Wales number 3880213.



Thank you for your order!

Dear Dr. Bushra Ahmed,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Dr. Bushra Ahmed
Order Date:	Aug 20, 2017
Order Number:	4172880706807
Publication:	Archives of Disease in Childhood
Title:	Primary ciliary dyskinesia: current state of the art
Type of Use:	Dissertation/Thesis
Order Total:	0.00 GBP

View or print complete [details](#) of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

How was your experience? Fill out this [survey](#) to let us know.

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
<https://myaccount.copyright.com>



RightsLink®